The Classical Cadence as a Closing Schema: Learning, Memory, & Perception

David Robert William Sears

Music Theory Area
Department of Music Research
Schulich School of Music
McGill University
Montreal, Canada

September 2016

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

For Shannon

Abstract

How do we know when a passage of music is coming to an end? According to scholars in the learning sciences, much of what we understand about the conventionalized ending formulæ characterizing music of the classical style—what scholars have for centuries called *cadences*—is acquired implicitly over the course of many years. This dissertation considers the validity of this claim using corpus-analytic and experimental techniques, and drawing from theories of implicit (statistical) learning, schema theory, and expectation.

Due to the cross-disciplinary nature of the research paradigm, the dissertation is divided into three parts to present theoretical, computational, and experimental approaches. Part I reviews contemporary accounts of the classical cadence articulated in the "New *Formenlehre*" tradition and then outlines the theories that account for the acquisition and mental representation of the most common cadence types associated with the late-eighteenth-century repertories of Haydn, Mozart, and Beethoven, paying particular attention to the cadence typology presented in William E. Caplin's *Classical Form*. Following Robert Gjerdingen's schema-theoretic approach, I argue that listeners who are familiar with classical music have internalized these cadence types as a flexible network of interrelated mental representations, or *rival closing schemata*.

To support this view, Part II re-examines the classical cadence using statistical modeling procedures that might simulate the learning mechanisms underlying human cognition. The corpus consists of symbolic representations of 50 sonata-form expositions from Haydn's string

quartets (Opp. 17–76), along with an annotated collection of 245 exemplars of the five cadence categories that achieve, or promise, cadential arrival: perfect authentic, imperfect authentic, half, deceptive, and evaded. Models for the representation (the *multiple-viewpoint* framework), discovery (non-contiguous *n*-grams), classification (the *ratio* model, the *neighbor-joining* method, and *nearest neighbor analysis*), and prediction (a finite-context model called *IDyOM*) of cadences demonstrate the psychological reality of cadential schemata.

Finally, Part III extends the findings from Part II in five experimental studies. In Experiments 1–2 I asked participants to provide completion ratings for cadences heard both in and out of context to examine the roles played by syntactic and rhetorical parameters in models of cadential strength. Next, Experiments 3–5 consider the link between expectancy and cadential closure, using both explicit (retrospective and continuous ratings) and implicit (reaction-time) response methods. Taken together, the reported findings provide converging evidence in support of the view that category systems for the classical cadence are psychologically relevant if they mirror the structure of attributes encountered in a given repertory that listeners are likely to learn and remember.

Résumé

Comment savons-nous que, dans une musique, un passage touche à sa fin? Selon les spécialistes en science de l'apprentissage, une grande partie de ce que nous comprenons comme étant des formules conventionnelles de clôture caractéristiques de la musique du style classique—ce que les chercheurs ont appelé pendant des siècles *cadences*—est acquis implicitement au cours de nombreuses années. La présente thèse évalue la validité de cette affirmation en utilisant à la fois des techniques d'analyse de corpus ainsi que des techniques expérimentales, tout en s'inspirant des théories de l'apprentissage implicite, la théorie des schémas ainsi que la théorie des attentes.

En raison de la nature interdisciplinaire du paradigme de recherche, la thèse est divisée en trois parties présentant respectivement les approches théoriques, informatiques, et expérimentales. La première partie présente tout d'abord les discussions contemporaines de la cadence classique telle que pensée dans la tradition "Nouvelle *Formenlehre*". S'ensuit la description des théories prenant en compte l'acquisition et la représentation mentale des types de cadences les plus courantes associées au répertoire de Haydn, Mozart et Beethoven datant de la fin du dix-huitième siècle. Une attention particulière est accordée à la typologie de cadence présentée par William E. Caplin dans son traité *Classical Form*. En prenant pour point de départ l'approche théorique des schémas de Robert Gjerdingen, je soutiens que les auditeurs familiers avec la musique classique ont intériorisé ces types de cadences en tant que réseau flexible de représentations mentales interdépendantes, ou *schémas de clôture rivaux*.

Afin de renforcer ce point de vue, la seconde partie réinterroge la cadence classique via l'utilisation de procédures de modélisation statistiques qui pourraient simuler les mécanismes d'apprentissage qui sous-tendent la cognition humaine. Le corpus est composé de représentations symboliques de 50 expositions de forme sonate en provenance de quatuors à cordes de Haydn (Opp. 17–76), ainsi que d'une collection annotée de 245 exemplaires des cinq catégories de cadence atteignant ou promettant l'arrivée cadentielle: authentique parfaite, authentique imparfaite, demi, rompue, et évitée. Les modèles pour la représentation (le cadre de *points de vue multiple*), la découverte (*n*-grammes non contigus), la classification (le modèle de *ratio*, la méthode *d'agglomération des voisins*, et *l'analyse du plus proche voisin*), et la prédiction (un modèle de contexte fini appelé *IDyOM*) des cadences démontrent la réalité psychologique de schémas cadentiels.

Enfin, la dernière partie approfondit les conclusions de la partie II par le biais de cinq études expérimentales. Pour les expériences 1 et 2, dans le but d'examiner les rôles joués par les paramètres syntaxiques et rhétoriques dans les modèles de force cadentielle, j'ai demandé aux participants de fournir des évaluations de degré de clôture de cadences entendues, certaines étant données dans leur contexte musical, d'autres non. Ensuite, les expériences 3 à 5 considèrent le lien entre l'attente et la clôture cadentielle, en utilisant les méthodes d'intervention explicites (évaluations rétrospectives et continues) et implicites (temps de réponse). Dans l'ensemble, les résultats rapportés fournissent des preuves convergentes soutenant l'hypothèse selon laquelle les systèmes de catégories pour la cadence classique sont psychologiquement pertinents si ils reflètent la structure des attributs rencontrés dans un répertoire donné que les auditeurs sont susceptibles d'apprendre et de se souvenir.

Acknowledgments

It is a great pleasure to acknowledge the many wonderful people who contributed to this thesis. First and foremost, I am deeply indebted to my supervisors, William E. Caplin and Stephen McAdams. Professor Caplin's theory of formal functions informs much of my thinking about the cadence concept, but it was in the workshop environment he cultivated in his graduate seminars that I learned to test my own assumptions against evidence, draw inferences from musical examples, and remember the simple (albeit tacit) mantra that an analyst never strays too far from the musical materials. On a project so wide-ranging as this one, I am also truly fortunate to have met Professor McAdams at the earliest stage of my graduate studies. He invites his students to play in an enormous, interdisciplinary sandbox where scholarship is inherently collaborative and no research question is too big or too small. In his laboratory, I was given the greatest possible freedom to pursue my own scholarly interests so long as I strived to meet the same high standards of evidence he demands for his own work. I also owe each of my supervisors a debt of gratitude for providing me with a sounding board and an editorial eye throughout the writing process, both of which have been absolutely invaluable, and by which I have improved as a writer and a thinker. It should be evident that any errors in this thesis are entirely my own.

I managed to rope a number of people into helping me with various aspects of the project, many of whom also work(ed) in the Music Perception and Cognition Laboratory. My thanks to Yinan Cao, Chelsea Douglas, and David Weigl for help with the mixed effects models in Chapter 8; to Kai Siedenburg and Cecilia Taher for many useful discussions about similarity and clustering, much of which was put to use in Chapter 5; to Lorraine Chuen, Hauke Egermann, Meghan Goodchild, Mattson Ogg, Dieu-Ly Tran, Konstantinos Trochidis, and Finn Upham for supporting my diversion into psychophysiology; and to Sven-Amin Lembke and Song Hui Chon for so often finding simple solutions to my Matlab problems. I am also grateful to Caitlin Martinkus, Mikaela Miller, Markus Neuwirth, Jason Noble, Jacob Spitzer, and Michel Vallières for inspiring many of the ideas that appear in this thesis. Finally, a special thanks to Bennett Smith for programming the software for all of the experiments in Part III; his expertise, kindness, and dry wit were a welcome reprieve from the Monday blues.

This thesis borrows freely from a few open-source computational models whose developers deserve special mention. Marcus Pearce was kind enough to invite a music theorist with little experience in machine learning algorithms to spend two months at the School of Electronic Engineering and Computer Science at Queen Mary University of London. The application and extension of his finite-context model in Chapter 6 would not have been possible without his patience and collaborative joie de vivre. I am also grateful to Tom Collins for taking the time to acquaint me with Marc Leman's *echoic memory* model and Petr Janata's *tonal space* model, both of which appear in Chapter 8. Also, my thanks to René Rusch, Jonathan Wild, and Nicole Biamonte for their guidance and encouragement about a number of issues, both philosophical and professional, and for demonstrating how interdisciplinary and inclusive the field of music theory can be. Finally, a special thank you to my thesis examiners, Robert Gjerdingen, Nicole Biamonte, and Christine Beckett, for their valuable input.

Lastly, I would like to thank my friends and family for embodying so many of the traits I have tried to emulate. As a scientist, my father is fearless in the face of life's big questions; his curiosity, passion, and industry remain a continuous source of inspiration. My mother helms the H.M.S. Sears and keeps each of us on course with love and support. I will always cherish the memories of our evening math sessions at the dinner table, where I learned to approach difficult problems with patience and quiet determination. Finally, I would like to thank Shannon, my best friend and confidante. You are brilliant, kind, and funny, you remind me not to take things too seriously, and your strength and resolve have no rival. You have been my lodestar in these stormy seas, so I dedicate this thesis to you.

Contribution of Authors

The document is formatted as a *monograph* dissertation, but its core chapters comprise research articles formatted for publication in scholarly journals, which have either already been published or are in preparation for submission.

- Chapter 2: **Sears, David**. "The Perception of Cadential Closure." In *What is a Cadence? Theoretical and Analytical Perspectives on Cadences in the Classical Repertoire*, edited by Markus Neuwirth and Pieter Bergé, 251–283. Leuven: Leuven University Press, 2016.
- Chapter 4: **Sears, David**. "Beneath the Surface: Pattern Discovery Using Non-Contiguous *N*-grams." Working Paper, Schulich School of Music, Centre for Research in Music, Media, and Technology, McGill University, Montreal, QC, 2016.
- Chapter 5: **Sears, David**. "The Classical Cadence Typology: Classification Using Phylogenetic Trees." Working Paper, Schulich School of Music, Centre for Research in Music, Media, and Technology, McGill University, Montreal, QC, 2016.
- Chapter 6: **Sears, David**, Marcus Pearce, William E. Caplin, and Stephen McAdams. "Simulating Expectations for Tonal Cadences Using Finite-Context Models." Working Paper, Schulich School of Music, Centre for Research in Music, Media, and Technology, McGill University, Montreal, QC, 2016.

- Chapter 7: **Sears, David**, William E. Caplin, and Stephen McAdams. "Perceiving the Classical Cadence." *Music Perception* 31, no. 5 (2014): 397–417.

- Chapter 8: **Sears, David**, Marcus Pearce, Jacob Spitzer, William E. Caplin, and Stephen McAdams. "Expectations for Tonal Cadences: Sensory and Cognitive Priming Effects." Working Paper, Schulich School of Music, Centre for Research in Music, Media, and Technology, McGill University, Montreal, QC, 2016.

William Caplin and Stephen McAdams served as the thesis supervisors. In addition to providing guidance on all aspects of the project (analytical techniques, interpretation of experimental results, etc.), both supervisors served as co-authors for the creation of the Haydn Corpus in Part II and the experimental studies in Part III. Professor Caplin reviewed all analytical annotations from the Haydn Corpus and the stimulus sets used in the experimental studies. Professor McAdams contributed to the design and analysis of each experiment. Both supervisors also financed the research with remuneration of participants.

Dr. Marcus Pearce of Queen Mary University of London provided the finite-context model described in Chapter 6 (called *IDyOM*), which predicts the next event in a monophonic musical stimulus by acquiring knowledge through unsupervised statistical learning of sequential structure. He oversaw the analysis of the Haydn Corpus and offered technical training about the model, which appears in Chapters 6 and 8. Jacob Spitzer completed an undergraduate research project examining expectations for cadences, and under my supervision helped to design, conduct, and analyze Experiments 3 and 4 in Chapter 8.

As principal author, I am responsible for all aspects of the dissertation. I created the Haydn Corpus in Part II and the stimulus sets in Part III, extended *IDyOM* to analyze vertical simultaneities within polyphonic textures (i.e., chords), and designed and analyzed all experiments reported in Part III.

Contents

List of Examples

List of Figures

List of Tables

Introduction

...while the patterns of music do not seem to be objectified by most members of most cultures, including our own, they are thoroughly learned. It apparently makes no difference that most of us cannot make sharp definitions of consonance and dissonance, or speak with real knowledge of the perfect cadence; we recognize what is consonant and what is dissonant in our music, and we have learned our music patterns well enough to know when the closing measures of a composition are brought to a satisfying or to an "unfinished" end. We learn what kinds of sounds are satisfactorily fitted into our music without necessarily having any technical knowledge about it; music structure is carried subliminally and, since it is not objectified in most individual cases, it is resistant to change.

Alan P. Meriam

Pattern discovery is an essential task in many fields, but particularly so in that branch of criticism concerned with the theory and analysis of musical form. According to Herbert A. Simon and Richard K. Sumner, "one of the purposes of analyzing musical structure and form is to discover the patterns that are explicit or implicit in musical works."[1] This seems especially true of the highly stereotyped harmonic and melodic formulæ appearing at the ends of phrases, themes, and larger sections in the instrumental repertories of the late eighteenth century—what music theorists and composers call *cadences*.

By way of example, consider the first eight measures from the third movement of Mozart's

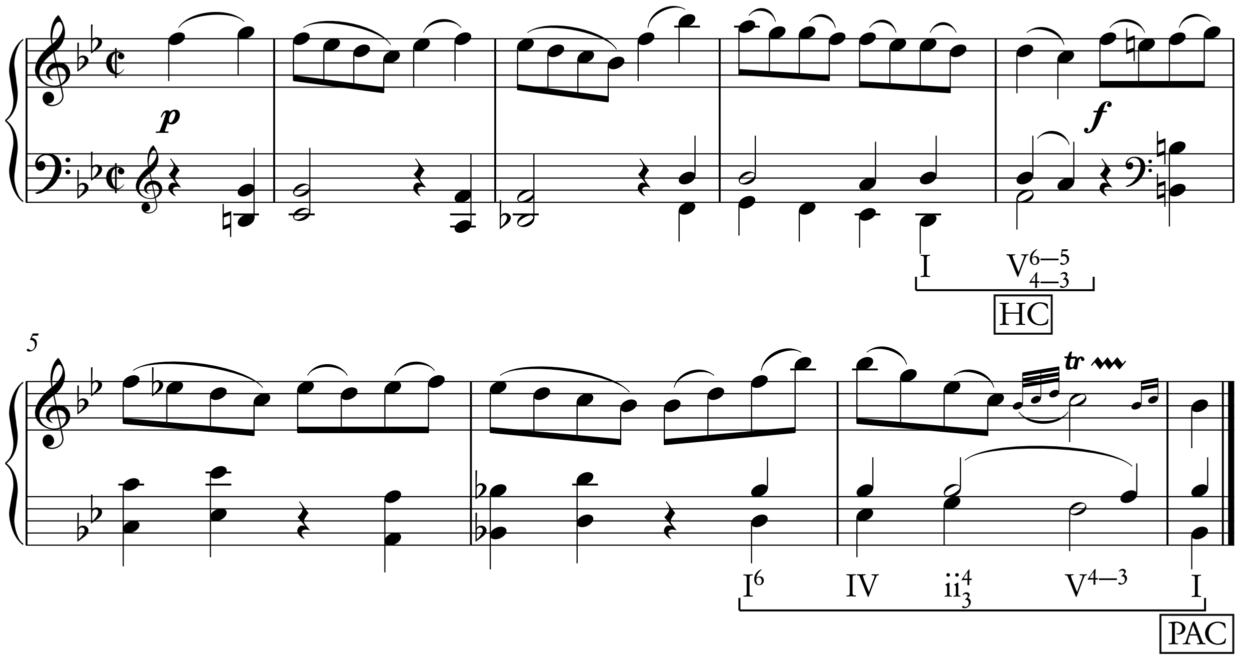
[1]Herbert A. Simon and Richard K. Sumner, "Pattern in Music," in *Machine Models of Music*, ed. Stephan M. Schwanauer and David A. Levitt (Cambridge, MA: MIT Press, 1993), 83.

Piano Sonata in B-flat, K. 281, shown in Example 1. With the exception of its somewhat unusual opening harmony (a tonicization of ii), this passage is in many respects a fairly conventional beginning in Mozart's keyboard style. Two phrases subdivide the opening eight measures: an antecedent phrase in mm. 1–4 followed by a consequent phrase in mm. 5–8. Although the second phrase largely repeats the first, a few compositional devices help to distinguish the consequent from the antecedent: the change from a *piano* to a *forte* dynamic, the subtle embellishments of the right-hand melody, and the expansion of register in the left hand. Yet, the most significant difference appears at the end of each phrase. The antecedent concludes with a dominant harmony in root position, a stable albeit active sonority whose metrical placement and expanded duration serve to reinforce the perception of ending. Theorists have termed this recurrent pattern a *half cadence*. The ending of the consequent, and of the theme as a whole, features harmonic motion from a root-position dominant to a root-position tonic at the downbeat of m. 8, as well as the arrival of the melody on the first scale degree, characteristics that exemplify a closing formula theorists have termed a *perfect authentic cadence*.

Although we tend to theorize little as to how passages like this one begin, we have a great deal to say about how they end. The highly conventionalized nature of these endings has prompted theorists to describe and explain the compositional procedures involved in articulating cadences. In the words of Jonathan Dunsby, the cadence concept is in fact "one of the few consistently patterned aspects of musical structure."[2] But why all this fuss about (cadential) patterns? Why is the discovery and classification of cadences germane to the interests of theorists, or indeed, anyone?

In the epigraph with which I began, anthropologist Alan P. Meriam makes a basic assumption about the "patterns of music," cadential or otherwise: they are "thoroughly learned," "carried

[2]Jonathan Dunsby, "Schoenberg on Cadence," *Journal of the Arnold Schoenberg Institute* 4, no. 1 (1980): 43.



Example 1: Mozart, Piano Sonata No. 3, K. 281, iii, mm. 1–8.

subliminally," and "resistant to change."[3] Thus for Meriam, "perfect cadences" are relevant to theorists, composers, or any other group of listeners because they are *learned* and *remembered*.[4] Psychologist Jay Dowling seems to agree:

> [The standard IV–V–I cadence] illustrates the type of musical structure that is implicitly comprehended by the brain when we listen to music. The average listener cannot explicitly verbalize the harmonic relationships just described but is nevertheless quite aware when the harmonies arrive at stable points of rest and when they need further resolution. It is this type of implicit knowledge that we wish to understand and explain.[5]

In other words, the classical cadence is the ideal pattern by which to test the hypothesis that

[3] Alan P. Meriam, *The Anthropology of Music* (Evanston, IL: Northwestern University Press, 1964), 297.

[4] A few decades earlier, fellow anthropologist Melville J. Herskovits explained that "the peculiar value of studying music... is that, even more than other aspects of culture, its patterns tend to lodge on the unconscious level ("Patterns of Negro Music," *Illinois State Academy of Science* 34 [1941]: 19).

[5] W. Jay Dowling and Dane L. Harwood, *Music Cognition* (New York: Academic, 1986), 18.

knowledge of music is thoroughly learned, carried subliminally, and resistant to change.

This dissertation is divided into three parts to present theoretical, computational, and experimental approaches to the classical cadence. Part I proceeds from the general to the particular, beginning with the concepts of closure and stability in music discourse, and then narrowing the theoretical and stylistic purview to the cadences and other closing formulæ characterizing classical music. Chapter 1 considers how definitions of closure engage theories of learning, memory, and perception. In my view, the phenomenal experience of closure depends on the mental representation of hierarchical organizational systems like tonality and meter where certain events are more permanent (or stable) in memory and facilitate processing during perception. To extend these claims to the recurrent temporal patterns appearing at the ends of phrases and themes in the classical style, Chapter 2 reviews contemporary approaches to the classical cadence and then outlines the theories that account for the acquisition and mental representation of the most common cadence types associated with the late-eighteenth-century repertories of Haydn, Mozart, and Beethoven, paying particular attention to the cadence typology presented in William E. Caplin's *Classical Form*. Following Robert Gjerdingen's schema-theoretic approach, I argue that listeners who are familiar with classical music have internalized these cadence types as a flexible network of interrelated mental representations, or *rival closing schemata*.

To support this view, Part II presents a corpus study of the classical cadence that uses statistical modeling procedures to simulate the learning mechanisms underlying human cognition for the purposes of pattern discovery, classification, and prediction. Chapter 3 presents the *Haydn Corpus*, which consists of symbolic representations of 50 sonata-form expositions from Haydn's string quartets (Opp. 17–76), along with an annotated collection of 245 exemplars of the five cadence categories that achieve, or promise, cadential arrival: perfect authentic, imperfect authentic, half, deceptive, and evaded. Using the *multiple-viewpoint* framework developed by

Darrell Conklin, I encode irreducible attributes of the musical surface like chromatic pitch, note onset time (in beats), metric position, key, and mode. From these basic types I then derive a number of other viewpoints to represent the "core" events of the classical cadence: the chromatic scale degrees, melodic intervals, and contours of the outer parts, a coefficient representing the strength of the metric position, and a vertical sonority, presented as a combination of vertical interval classes or chromatic scale degrees. Armed with this corpus, Chapter 4 employs an *n*-gram approach to determine whether cadences and other closing formulæ are the most recurrent patterns in the Haydn Corpus. Next, Chapter 5 classifies the cadence collection on the basis of the features they share using a family of techniques for similarity estimation and clustering pioneered (or inspired) by psychologist Amos Tversky. Finally, Chapter 6 applies a finite-context model developed by Marcus Pearce called the *Information Dynamics of Music* model (or IDyOM) to determine whether expectancy formation, fulfillment, and violation contribute to the perceived closing strength of the cadence categories in Caplin's typology.

To consider whether the findings from Part II relate to the schematic knowledge of contemporary listeners, Part III examines the influence of musical expertise on the perception and cognition of cadential closure in Mozart's keyboard sonatas using the many methods of inference developed in the experimental sciences. In Experiments 1–2 in Chapter 7, I asked participants to provide completion ratings for cadences heard both in and out of context to examine the roles played by syntactic and rhetorical parameters in models of cadential strength. Next, Experiments 3–5 in Chapter 8 consider the link between expectancy and cadential closure using retrospective ratings, continuous ratings, and an implicit reaction-time task based on the priming paradigm. Across all five studies, the reported findings support the view that the classical cadence is the quintessential phrase-level event schema, a perfect distillation of the features characterizing the classical style.

Part I

The Concept of Closure

Chapter 1

Closure and Stability in Tonal Music

> *A similar rushing forth of participating energies ... takes place at the cadence and here too, in the jostling, space gets tight, so that the individual, regardless of how important he may be, has to be satisfied with a fraction of the space available... This is surely the psychology of the close.*
>
> Arnold Schoenberg

There is likely not a scholarly text in the history of music that does not touch upon the concept of *closure*. One finds references to its various word forms (e.g., *close*, *closed*, *closing*, etc.) or its many cognates (e.g., Fr. *cadence*, Ger. *Schluss*, Lat. *clausula*, etc.) in some of the earliest analytical writings about music, and the idea still has considerable currency among contemporary scholars. This is perhaps because closure remains a lodestar for musical organization, the discovery of which remains the central task of many music theorists. As Mark Anson-Cartwright points out,[1] some writers have even drawn inspiration from studies of closure in other art forms, where critics like Frank Kermode,[2] Barbara Herrnstein Smith,[3] and

[1]Mark Anson-Cartwright, "Concepts of Closure in Tonal Music: A Critical Study," *Theory and Practice* 32 (2007): 1.

[2]Frank Kermode, *The Sense of an Ending: Studies in the Theory of Fiction* (London: Oxford University Press, 1968).

[3]Barbara H. Smith, *Poetic Closure: A Study of How Poems End* (Chicago: The University of Chicago Press,

Rudolf Arnheim have been particularly influential.[4] The concept of closure thus looms large for many scholars, both because it encapsulates so much about our experience of artistic works and because it invites cross-fertilization between art forms and their associated disciplines.

One reason for the persistence of the term closure in analytical writing is that, as a linguistic metaphor, it maps potentially opaque aspects of musical structure onto kinaesthetic features of typical human behaviors, such as bodily movement through space or the physical manipulation of objects. Linguists and philosophers call these linguistic mappings *image schemas*, "recurring, dynamic pattern[s] of our perceptual interactions and motor programs that [give] coherence and structure to our experience."[5] Janna Saslaw has argued, for example, that musical events are *closed* in the sense that, like the edges of a container, they form a boundary separating one passage from another.[6] The appeal of the container schema is thus that it offers a linguistic shorthand for describing seemingly ineffable experiences like boundary perception in terms of everyday kinaesthetic behaviors like closing a jar.[7]

Nevertheless, closure has become something of an umbrella term in analytical discourse. Its usage in present-day scholarship admits multiple meanings beyond those expressed by the original container schema, and it often operates on a vast continuum of time spans, from the

---

1968). See, for example, Kofi Agawu, "Concepts of Closure and Chopin's Opus 28," *Music Theory Spectrum* 9 (1987): 1–17; Patrick McCreless, "The Hermeneutic Sentence and Other Literary Models of Tonal Closure," *Indiana Theory Review* 12 (1991): 35–73. Ironically, Smith herself cites Leonard B. Meyer's discussion of closure in his first book, *Emotion and Meaning in Music*, as an important influence on her theory of poetic closure (*Poetic Closure: A Study of How Poems End*, 1–37).

[4] Rudolf Arnheim, *Art and Visual Perception: A Psychology of the Creative Eye* (Berkeley: University of California Press, 1954). See, for example, Steve Larson, "On Rudolf Arnheim's Contribution to Music Theory," *The Journal of Aesthetic Education* 27, no. 4 (1993): 97–104; Steve Larson, *Musical Forces: Motion, Metaphor, and Meaning in Music* (Bloomington and Indianapolis: Indiana University Press, 2012).

[5] Mark Johnson, *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason* (Chicago: The University of Chicago Press, 1987), xiv.

[6] Janna Saslaw, "Forces, Containers, and Paths: The Role of Body-Derived Image Schemas in the Conceptualization of Music," *Journal of Music Theory* 40, no. 2 (1996): 222.

[7] For a discussion of image schemas and their relevance to music, see Saslaw, "Forces, Containers, and Paths: The Role of Body-Derived Image Schemas in the Conceptualization of Music"; Lawrence M. Zbikowski, *Conceptualizing Music: Cognitive Structure, Theory, and Analysis* (Oxford: Oxford University Press, 2002), 63–95.

two-note motive to the multi-movement work. In cases where closure seems too general and imprecise, modifiers like 'tonal,' 'formal,' 'syntactic,' and 'rhetorical' are also commonplace, thereby restricting its purview to specific musical parameters or repertories. And yet since a definitive account of the term would deserve its own volume,[8] and since this thesis concerns a particular species of closing pattern scholars call *cadence*, §1.1 will merely reiterate what others have said about the most common definitions of closure in music discourse. §1.2 then considers how those definitions might trespass on theories of learning, memory, and perception by abandoning the concept of closure—which rarely appears in the writings of experimental psychologists—in favor of *stability*, a term that often appears in its stead.

§1.1 Definitions of Closure in Music Discourse

The concept of closure in music discourse often refers to one or more of the following four definitions:

i. *the segmentation of the musical surface into motives, phrases, and sections.*

This definition makes two assumptions about musical organization: (1) that a composition segments into discrete "chunks," such that closing events precede temporal *boundaries*; and (2) that this segmentation process organizes these segments hierarchically, such that segments at one level of the structural hierarchy—say, for example, sections—nest (or subsume) segments at lower levels—phrases and motives.[9] In most theories of segmentation, this process applies recursively, phrases nesting within sections, motives nesting within phrases, and so on. Fred Lerdahl and

[8] For a review of the concept of closure in music analysis, see Anson-Cartwright, "Concepts of Closure"; Janet Joichi, "Closure, Context, and Hierarchical Grouping in Music: A Theoretical and Empirical Investigation" (PhD Dissertation, Northwestern University, 2006), 66–85; Crystal Peebles, "The Role of Segmentation and Expectation in the Perception of Closure" (Dissertation, Florida State University, 2011), 6–84.

[9] Bob Snyder's description of closure in terms of segmental groups is a case in point (*Music and Memory: An Introduction* [Cambridge, MA: The MIT Press, 2000], 59–67).

Ray Jackendoff have called this organization *grouping structure*,[10] but since psychologists also employ the term 'grouping' to refer to the *simultaneous* and *sequential* processes by which listeners organize the incoming auditory stimulus into continuous streams of sound events,[11] we might instead call it *segmental grouping*.

Unfortunately, the relationship between segmental grouping and closure is not entirely clear, since the former process is not always synonymous with the latter concept in music discourse. According to musicologist Leonard B. Meyer, closure "is not simply cessation—silence. It involves conclusion—almost in the syllogistic sense that the conclusion or completion is implicit in the premises, in the earlier phases of the musical motion."[12] The caesura following a dominant seventh chord may elicit a decisive segment boundary, for example, but the failure to resolve that dominant will extend the tension and imply further continuation. Thus, for Meyer, the perception of closure depends in part on the cessation of expectations following the terminal event(s) of a musical process. Nevertheless, this is not to say that segmentation and closure represent distinct processes. Crystal Peebles has suggested, for example, that segmentation is a prerequisite for closure. In her view, closure is "the feeling of finality that occurs at the anticipated end of a musical segment."[13]

Implicit in these criticisms is another definition of closure, one that associates a closing event with the "conclusion," "completion," or "anticipated end" of a temporal process:

ii. *the terminus (or completion) of a temporal process.*

This definition is perhaps the most pervasive in contemporary scholarship. It assumes first, that closure characterizes musical events, rather than segment boundaries, where an 'event' could

[10]Fred Lerdahl and Ray Jackendoff, *A Generative Theory of Tonal Music* (Cambridge, MA: The MIT Press, 1983), 8.

[11]Albert S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (Cambridge, MA: MIT Press, 1990).

[12]Leonard B. Meyer, *Emotion and Meaning in Music* (Chicago: University of Chicago Press, 1956), 130.

[13]Peebles, "The Role of Segmentation and Expectation in the Perception of Closure," 1.

refer to an isolated note or chord, or to longer time spans in the structural hierarchy, such as motives and phrases; and second, that closing events *follow* other events in the musical process; to borrow an expression from William E. Caplin, they "must *end* something."[14] In other words, closing events serve a specific *temporal function*; they complete a goal-directed process.

If we assume that the mechanisms eliciting segment boundaries differ in some way from those effecting temporal functionality, an event conveying an initiating or medial function could precede a segment boundary. From this point of view, grouping structure—at least as it was intended by Lerdahl and Jackendoff—is divorced from temporal function, from the perception of beginnings, middles, and ends. And just as I related segmental grouping in definition i to the hierarchical organization of musical works, so too can we apply definition ii to multiple levels of the structural hierarchy. A beginning at one level—say, for example, a composition's opening theme—may subsume beginnings, middles, and ends at lower levels—its phrases, motives, and so on. To be sure, theories of temporality articulated in the "New *Formenlehre*" tradition take precisely this approach.[15] Caplin suggests, for example, that "closure in general involves bringing to completion some process implicating one or more modes of musical organization at a given structural level of a work," with cadences serving to end local-to-middle-ground levels of musical organization—the levels of the phrase and theme.[16]

[14]William E. Caplin, "The Classical Cadence: Conceptions and Misconceptions," *Journal of the American Musicological Society* 57, no. 1 (2004): 56.

[15]Markus Neuwirth coined the term "New *Formenlehre*" tradition to refer to the revival of interest in theories of musical form over the last few decades ("Recomposed Recapitulations in the Sonata-Form Movements of Joseph Haydn and his Contemporaries" [PhD Dissertation, Leuven University, 2013], 9), perhaps best exemplified in recent volumes by William E. Caplin, James Hepokoski and Warren Darcy, and Janet Schmalfeldt. See, for example, William E. Caplin, *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven* (New York: Oxford University Press, 1998); William E. Caplin, James Hepokoski, and James Webster, *Musical Form, Forms, & Formenlehre: Three Methodological Reflections*, ed. Pieter Bergé (Leuven: Leuven University Press, 2009); James Hepokoski and Warren Darcy, *Elements of Sonata Theory: Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata* (New York: Oxford University Press, 2006); Janet Schmalfeldt, *In the Process of Becoming: Analytic and Philosophical Perspectives on Form in Early Nineteenth-Century Music* (New York: Oxford University Press, 2011).

[16]Caplin, "The Classical Cadence," 56.

To consider the articulation of closure at more global levels, however, authors typically rely on another definition, one of *return*:

iii. *a return to baseline, homeostasis, or equilibrium.*

Meyer has called this definition the "law of return," and it again applies to both local and global levels of musical organization. Put simply, definition iii assumes that closure results from the *recurrence* of events following some intermediary process, resulting in a ternary formal design at a given level of the structural hierarchy (e.g., ABA). Metaphors of motion are also commonplace here, since closing events are said to return the composition from a state of tension and conflict to one of equilibrium, stability, or resolution. Meyer argues, for example, that "a feeling of harmonic completeness arises when the music returns to the harmonic base from which it began."[17] Critics often appeal to this definition at more global levels of musical organization, however, such as at the "double return" of the main theme and home key in the recapitulation of sonata-form movements.[18] In this context, recurrence is assumed to resolve a tonal or thematic conflict brought about by the presence of intermediary and contrasting materials;[19] it closes (or ends) the large-scale tonal and thematic motion.

This view of closure is most readily associated with Schenker's mature theory of tonal structure, but it might also describe more recent architectonic theories, such as the generative theory of Lerdahl and Jackendoff. In either case, the principle of tonal return assumes that the same principles apply continuously at every level of a composition's structural hierarchy. Following Schenker, Charles Rosen has suggested, for example, that the modulation to a contrasting key area in sonata form is "essentially a dissonance raised to a higher plane, that

[17] Meyer, *Emotion and Meaning in Music*, 150.

[18] James Webster, "Binary Variants of Sonata Form in Early Haydn Instrumental Music," in *Internationaler Joseph Haydn Kongress Wien 1982*, ed. Eva Badura-Skoda (Munich: Henle, 1986), 127.

[19] Charles Rosen, *The Classical Style: Haydn, Mozart, Beethoven* (New York: Norton, 1972), 120.

of the total structure."[20] According to this view, hierarchical organization is therefore limited to superordinate or subordinate relations between events on the musical surface; there are no coordinate (A↔B) or proordinate (A→B) relations between events either within or between levels.[21] As such, closure at both local and global levels of musical organization is characterized entirely by the superordinate status of the closing event within the structural hierarchy.

Each of the definitions presented thus far has been more or less ambivalent about the *experience* of closure during music listening. Surely some of the above assumptions are more psychologically relevant than others. Thus, one final definition deserves mention, one that situates the concept of closure within the psychological effects it presumably engenders:

iv. *the perception of rest, repose, quiescence, or stability; an arrival.*

Applications of closure in music discourse routinely demonstrate a certain psychologizing impulse.[22] Appeals to the "feeling" or "sense" of finality or resolution are especially common. These descriptions also continuously feature terms associated with phenomenal experience, such as relaxation, satisfaction, rest, repose, quiescence, and stability.

Given the discussion thus far, it should be evident that the above definitions are not mutually exclusive. For example, Anson-Cartwright's definition of closure as "that condition of rest or finality which a piece or movement attains at the moment of structural (tonal) resolution" resonates with definitions ii and iii, while his suggestion that listeners with exposure to a set of stylistic conventions will "feel" tonal closure during the act of listening recalls definition

[20] Ibid., 26.

[21] In the simplest sense, musical *organization* refers to the relationships between events on the musical surface, be they notes, chords, motives, phrases, or any other coherent 'units' of that organization. The connections between these events might represent temporally ordered (or *proordinate*) relations, such as the V–I progression in tonal harmony, unordered (or *coordinate*) relations, such as the members of a major triad, or in hierarchically arranged systems, superordinate/subordinate relations, such as the prolongation of a given harmony through other (subordinate) harmonies (e.g., I–V^4_3–I^6). For a discussion of these relational types, see George Mandler, "Organization, Memory, and Mental Structures," in *Memory Organization and Structure*, ed. C. Richard Puff (New York, NY: Academic Press, 1979), 303–319.

[22] Chapter 2 links this impulse to discussions of cadence, specifically (see §2.2).

iv.[23] It is also worth noting that these definitions are by no means exhaustive. A number of other issues appear frequently in discussions of closure that I have not mentioned here: the status of closure as both time-point and time-span, the potential for certain parameters to afford hierarchic structuring (e.g., pitch and temporal duration) when others do not (e.g., dynamics and tempo), and the strength of closure provide ready examples. But since each of these issues resurfaces in later chapters, I will instead turn my attention to the psychological mechanisms underlying the phenomenal experience of closure, particularly as it pertains to the cadences and other mid-level closing patterns associated with the instrumental repertories of the late eighteenth century: music that today exemplifies the *classical style*.

My goal in the next section is to explain how various psychological processes might relate to the experience of closure for the melodic and harmonic events appearing at the moment of *cadential arrival*, a term designated by Caplin to refer to the terminal event(s) of a cadence, such as the onset of a root-position tonic harmony at the end of a perfect authentic cadence. But since I noted in the introduction that the concept of closure appears infrequently in music psychology, I will prefer the term *stability*, which often appears in its stead. Thus, I begin by clarifying what psychological stability might mean for individual events like notes and chords, leaving a discussion of the processes underlying the perception of cadences and other recurrent closing patterns for Chapter 2.

§1.2 Stability: The Sensory-Cognitive Continuum

Among music theorists and psychologists, closure often accompanies stability. Meyer writes, for example, that closure is "the arrival at relative stability,"[24] while psychologists Jamshed

23 Anson-Cartwright, "Concepts of Closure," 3.

24 Leonard B. Meyer, *Explaining Music: Essays and Explorations* (Berkeley: University of California Press, 1973), 81.

J. Bharucha and Carol Krumhansl explain that stable events "are perceived as more final and serve as better completions" compared to unstable ones.[25] But whereas definitions of closure typically allude to the temporal functionality of "closing" events (see definition ii above), the concept of stability appeals to the hierarchical systems that govern musical organization. This is to say that stability is a property (or characteristic) of a system where certain events are more central (or *stable*) than others. Meyer's definition of tonality is a case in point:

> The term "tonality" refers to the relationships existing between tones or tonal spheres within the context of a particular style system...; some of the tones of the system are active. They tend to move toward the more stable points in the system—the structural or substantive tones.
>
> But activity and rest are relative terms because tonal systems are generally hierarchical: tones which are active tendency tones on one level may be focal substantive tones on another level and vice versa. Thus in the major mode in Western music the tonic tone is the tone of ultimate rest toward which all other tones tend to move. On the next higher level the third and fifth of the scale, though active melodic tones relative to the tonic, join the tonic as structural tones; and all the other tones, whether diatonic or chromatic, tend toward one of these. Going still further in the system, the full complement of diatonic tones are structural focal points relative to the chromatic notes between them. And, finally, as we have seen, any of these twelve chromatic notes may be taken as substantive relative to slight expressive deviations from their normal pitches.[26]

For Krumhansl and Bharucha, the "stable points" characterized by Meyer as "focal" or "inactive"

[25] Jamshed J. Bharucha and Carol L. Krumhansl, "The Representation of Harmonic Structure in Music: Hierarchies of Stability as a Function of Context," *Cognition* 13 (1983): 83.

[26] Meyer, *Emotion and Meaning in Music*, 214–215.

became synonymous with a host of other terms, such as structural significance, resistance to change, priority, and resolution,[27] all of which reflect—to a greater or lesser degree—the hierarchical system of dynamic, temporal relations Krumhansl would call the *tonal hierarchy*.[28]

But how might organizational systems like tonality influence listeners? If the temporal relations characterizing these systems played no role during listening, the hierarchies of stability that are "out there" in the music—on which the perception of closure seems to depend—would presumably collapse. For our purposes, this means that *psychological* stability is a property of the *mental organization* of musical materials, rather than of the musical materials themselves. Or put another way, these systems are not "objective properties of the music" referring to a "fixed, internal structure,"[29] but psychological *effects* resulting from how listeners organize sensory stimuli in general, without which tonality and meter—as products of human creativity—would not exist.

Meyer makes this distinction between objective and subjective organization explicit in an earlier passage of *Emotion and Meaning in Music*.

> Our opinion or feeling as to the completeness of a given stimulus is a product of the natural modes of mental organization. These function both within the framework of what is given in the style and within the sound terms established in the particular work. In other words, the mind, governed by the law of Prägnanz, is continually striving for completeness, stability, and rest.[30]

Setting aside Meyer's allusion to Gestalt theory,[31] his description of the perception of closure

---

[27] Carol L. Krumhansl, *Cognitive Foundations of Musical Pitch* (New York, NY: Oxford University Press, 1990), 19.

[28] I review the experimental studies demonstrating this hierarchy in §7.1. I also replicate this hierarchy in §4.1.

[29] Brian Hyer, "Tonality," in *Cambridge History of Western Music Theory*, ed. Thomas Christensen (Cambridge: Cambridge University Press, 2002), 727.

[30] Meyer, *Emotion and Meaning in Music*, 128.

[31] The central point of the law of Prägnanz is that "psychological organization will always be as 'good' as

according to the "natural modes of mental organization" and its relation to "what is given in the style" assumes that the organized objects and events encountered in the external environment are at least partially isomorphic with their *internal* (or *mental*) *representations*. Psychologist Arthur Reber has called this position *representational realism*, and according to Stephen McAdams, it is likely "one of the major contentions of the cognitive approach to psychology."[32] In reference to tonal music, for example, McAdams notes that these representations might correspond "to the perception or imagination of a major chord, to the connecting of a sequence of clarinet notes into a melody, or to the expectation of a tonic chord to resolve the tension state evoked by a dominant seventh chord."[33] In each case, the symbolic representation in the musical materials corresponds in some meaningful way with its manner of representation in the minds of listeners.[34]

---

the prevailing conditions allow. In this definition the term 'good' is undefined. It embraces such properties as regularity, symmetry, simplicity and others..." (Kurt Koffka, *Principles of Gestalt Psychology* [New York: Harcourt, Brace & Co., 1935], 303). For Meyer's discussion of the law of Prägnanz and its relevance to his theory of expectation, see *Emotion and Meaning in Music*, 83–127.

[32]Stephen McAdams, "Music: A Science of the Mind?," *Contemporary Music Review* 2 (1987): 18. Roger Shepard argues for a similar principle that he calls *psychophysical complementarity*, which states that the mental processes of animals have evolved to be complementary with the structure of the surrounding world ("Psychophysical Complementarity," in *Perceptual Organization*, ed. Michael Kubovy and James R. Pomerantz [Hillsdale, N.J.: Erlbaum, 1981], 279–341). Albert Bregman notes, for example, that the physical world allows an object to be rotated without changing its shape, so the mind must have mechanisms for rotating its representation of objects without changing their shapes as well (*Auditory Scene Analysis: The Perceptual Organization of Sound*, 39).

[33]McAdams, "Music: A Science of Mind?," 19.

[34]Cognitive scientist Zenon Pylyshyn explains, "to be in a certain representational state is to have a certain symbolic expression in some part of memory" (*Computation and Cognition* [Cambridge, MA: MIT Press, 1985], 29). Reber cautions, however, that the procedures (or *rule systems*) underlying the organization of a stimulus by an experimenter do not necessarily correspond with those employed by listeners, despite the similarity between the structure of the stimulus display and its manner of mental representation. "It is likely that insights into the nature of mental representation will be garnered from careful examination of the structured nature of the stimulus display, but it is also likely that the operative description of the array may not correspond with the initial formal characterizations that were used in its construction. In short, we know with surety what the rules were that we used to construct the displays. But we cannot know with anything like the same confidence either the rules that characterize the display nor those that are induced by our subjects. Careful examinations of their behavior will provide some understanding of the rule systems they have induced, and the basic principle of representational realism suggests that we regard this characterization as the one best suited for the stimulus displays themselves (*Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious* [Oxford: Oxford University Press, 1993], 118).

In this context, the notion of psychological stability for organizational systems like tonality has profound implications for learning, memory, and perception. Presuming listeners acquire mental representations of major triads, clarinet melodies, and dominant-to-tonic progressions during music listening, the concept of psychological stability might also reflect the permanence of those representations in memory and their influence on subsequent processing. One might argue, for example, that listeners are more likely to expect more stable events in the tonal system because they are better remembered. Some three decades after the publication of Meyer's book, this was, in fact, precisely the view Krumhansl proposed.

> Human cognitive and perceptual systems invest certain elements with special status: these elements are given priority in processing, are most stable in memory, and are more important for linguistic descriptions.[35]

For Krumhansl, an organizational system like tonality is psychologically relevant if its most central events are more stable in memory and facilitate processing during perception. To return to Meyer's definition of tonality, this processing advantage explains why certain (unstable) tones are active while other (stable) tones are inactive or restful: stable events are resistant to change, continuation, or further implication because they are better remembered, and thus easier to process. In short, stable events are more *expected* than their unstable counterparts.[36]

But how do these systems emerge in the minds of listeners? This question revives a very old argument about the influence of previous knowledge on the organization of sensory experience and the formation of expectations. On one side, the *nativists* assume that the capacity to comprehend complex, "rule-governed" syntactic structures like natural language or tonal music

[35] Krumhansl, *Cognitive Foundations of Musical Pitch*, 18.

[36] Bharucha and Krumhansl argue, for example, that "in the predominant Western idiom and in many forms of Eastern music as well, beginnings and endings of segments are usually marked by stable musical events. The listener familiar with this idiom has expectations as to when unstable events should resolve to more stable events, and the composer works with these expectations" ("The Representation of Harmonic Structure in Music," 94).

is given a priori, which is to say that the brain inherits the necessary sensory and cognitive mechanisms to perceive and produce such structures in everyday life. Perhaps more importantly, this position also assumes that the biological constraints placed on one genre or style period also necessarily apply to all of the others. As a result, acoustic and psychoacoustic (or *physicalist*) accounts assume that musical concepts like consonance and dissonance, harmony, voice leading, and tonality represent cross-cultural universals.

On the other side, the *empiricists* claim that the external environment engages general induction routines in the cortex that operate across modalities and across stimulus forms. With sufficient exposure, these routines learn and remember the underlying structure of the stimulus environment and then guide the sensory apparatus to seek out and find similar structures.[37] In our case, this would mean that knowledge of the organizational systems governing classical music should differ from person to person, resulting in a range of perceptual experiences for phenomena like stability and closure. What is more, tonality itself would no longer serve as a monolith in experimental research, since the temporal relations characterizing that system would vary from one style to another. In other words, if the empiricist view is correct, the concept of tonality as it is intended by experimental psychologists would instead reflect a plurality of potentially overlapping tonal systems, each governed by a given repertory or style period and each depending on the previous experiences of a given individual listener.

The distinction between nature and nurture remains quite popular in the cognitive sciences, though the tendency towards monism is not always useful. For example, the capacity for three-dimensional vision develops early and rapidly, but the influence of previous knowledge

[37]Reber, *Implicit Learning and Tacit Knowledge*, 156. Even the *empiricist* position just outlined assumes that the cortex begins with a certain kind of structure (i.e., the general induction routines) already in place, which Reber calls *process nativism*. Reber contrasts this view with a *content nativism* of the kind supported by scholars like Noam Chomsky and Jerry Fodor, which assumes that the very content of mind, the content of encapsulated cognitive modules, is laid down in the genes (Ibid., 149).

on visual perception is now widely accepted.[38] In light of the plethora of reported effects for both sensory and cognitive processes in numerous aspects of human auditory processing, it therefore seems unreasonable to assume that only one of these processes could account for the psychological effects associated with organizational systems like tonality. Nevertheless, this dual framework remains quite useful as a classification tool, since so many explanations—both past and present—divide the mind into its sensory and cognitive components. And since both accounts continue to receive attention in the scholarly community, it is helpful to retain this division for the sections that follow.

1.2.1 Sensory Principles

To account for hierarchical organizational systems like tonality and meter, as well as more general musical concepts like melody and voice leading, consonance and dissonance, and harmony, sensory explanations take as their starting point the mechanisms responsible for organizing complex auditory environments into, say, barking dogs, backfiring cars, and whirring blenders. In this context, the concept of stability is an emergent property of a more general organizational principle we might call *coherence*, which refers to the mechanisms by which acoustic components or events cohere as a single entity.[39] For nativists, the concept of coherence is a one-size-fits-all approach that yields increasingly complex organizations—from tones, to chords, to keys—by assuming that stable events organize (or cohere) in the auditory periphery

[38] Reber, *Implicit Learning and Tacit Knowledge*, 154.

[39] Stephen McAdams, "The Auditory Image: A Metaphor for Musical and Psychological Research on Auditory Organization," in *Cognitive Processes in the Perception of Art*, ed. W. Ray Crozier and Antony J. Chapman (Amsterdam: North-Holland, 1984), 291.

more readily than unstable events.[40] But how does this organization process take place?[41]

The vibrations produced by sound sources encountered in the external environment radiate out into the air as sound pressure waves. These waves combine linearly in the atmosphere—bouncing off some objects and getting partially absorbed by others—before reaching the ears as a complex mixture of acoustic energy.[42] This mixture may feature a single auditory event, which consists of a set of simultaneous frequency components emanating from the time-limited vibrations of one sound source, or a complex of events and their frequency components emanating from many sources. To infer the underlying structure of such a mixture, the human auditory system must group the various components into their constituent sound sources and then track those sources over time, a process psychologist Albert Bregman calls *auditory scene analysis* (ASA).[43]

ASA begins with the perceptual fusion or concurrent grouping of frequency components into auditory events. To recover the spectral and temporal characteristics of the various components of a complex sound, the sensory representation depends on a dual coding scheme in the auditory periphery.[44] On the one hand, the *tonotopic* (or *place*) scheme refers to the

[40] The human auditory system is divided into two subsystems: the peripheral auditory system (i.e., the outer, middle, and inner ear), and the central auditory system (i.e., the primary auditory cortex). For a review of the auditory periphery, see Graeme K. Yates, "Cochlear Structure and Function," in *Hearing*, ed. Brian C. J. Moore (San Diego: Academic Press, 1995), 41–74; Jan Schnupp, Israel Nelken, and Andrew King, *Auditory Neuroscience: Making Sense of Sound* (Cambridge, MA: The MIT Press, 2011), 51–92.

[41] Certainly an organizational system like meter also plays an important role in the stability of individual events in the classical style, for which sensory accounts also exist. See, for example, Olivia Ladinig et al., "Probing Attentive and Preattentive Emergent Meter in Adult Listeners Without Extensive Music Training," *Music Perception* 26, no. 4 (2009): 377–386; Edward W. Large and Caroline Palmer, "Perceiving Temporal Regularity in Music," *Cognitive Science* 26, no. 1 (2002): 1–37; Dirk Jan Povel, "Internal Representation of Simple Temporal Patterns," *Journal of Experimental Psychology: Human Perception and Performance* 7, no. 1 (1981): 3–18. Since I consider rhythm and meter in greater detail in Chapter 3, and since I am specifically concerned here with the stability of the sorts of events that appear at the moment of cadential arrival in classical music, I have restricted this discussion to physicalist accounts of pitch (i.e., melodic, harmonic, or tonal stability).

[42] Stephen McAdams and Carolyn Drake, "Auditory Perception and Cognition," in *Stevens' Handbook of Experimental Psychology*, ed. Hal Pashler and Steven Yantis, vol. 1: Sensation and Perception (New York: Wiley, 2002), 397.

[43] Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*.

[44] McAdams and Drake, "Auditory Perception and Cognition," 399.

organization of the hair cells along the basilar membrane, which act as filters akin to a Fourier analysis by responding maximally to different frequencies.[45] The hair cells responsible for a certain range of these frequencies then transduce the corresponding physical vibrations into nerve impulses. On the other hand, the *temporal* scheme refers to the process by which the basilar membrane registers small differences in the periodicities of these components by time-locking the pattern of neuron firings to certain phases of the wave form.[46] Together, these coding schemes recover the spectral and temporal characteristics of each component with remarkable precision.

The perceptual fusion of frequency components into a single auditory event—or the segregation of components into two or more events—depends on a host of cues, with the coincidence of more than one cue reinforcing the potential for fusion. In brief, components are likely to fuse if they are related by a common period (*harmonicity*), feature synchronous onsets (*onset synchrony*), originate from a common position in space (*spatial position*), or demonstrate similar intensity or frequency behavior over time (*modulation coherence*). Of this list, harmonicity and onset synchrony are arguably the most central.[47]

In the case of harmonicity, listeners typically associate complex periodic sounds with a fundamental frequency (or *virtual pitch*), though the sound itself consists of a number of component frequencies (or *spectral pitches*) related at integer multiples of the fundamental.[48] Concurrent grouping thus depends on the harmonicity of the various components, which is to

[45] Phil N. Johnson-Laird, Olivia E. Kang, and Yuan Chang Leong, "On Musical Dissonance," *Music Perception* 30, no. 1 (2012): 19. For an introduction to the history of the place theory of perception in the writings of physiologist Hermann von Helmholtz, see Benjamin Steege, *Helmholtz and the Modern Listener* (Cambridge: Cambridge University Press, 2012), 43–79.

[46] McAdams and Drake, "Auditory Perception and Cognition," 399; Brian C. J. Moore, "Pitch Perception," in *An Introduction to the Psychology of Hearing*, 2nd ed. (New York: Academic Press, 1982), 115–149.

[47] McAdams and Drake, "Auditory Perception and Cognition."

[48] Ernst Terhardt, "Pitch, Consonance, and Harmony," *Journal of the Acoustical Society of America* 55 (1974): 1061–1069.

say that harmonic sounds cohere more readily than non-harmonic sounds.[49] Brian C. J. Moore and his co-authors have demonstrated, for example, that listeners can hear out a harmonic component from the complex tone if it is mistuned by as little as 2% of its nominal frequency.[50] This mistuned harmonic can also affect the perceived frequency of the virtual pitch by pulling it in the direction of the mistuning, suggesting that the auditory periphery possesses a harmonic template or a time-domain coincidence processor that integrates those components featuring integer multiples of a fundamental.[51] What is more, the auditory system is extremely sensitive to small temporal asynchronies among the frequency components.[52] Rudolf A. Rasch found, for example, that a single frequency component from a complex tone becomes audible with an asynchrony as small as 35 ms.[53]

Once fused, the auditory periphery can extract a number of auditory attributes for a given event, such as loudness, pitch, and timbre.[54] In the domain of pitch, for example, the clarity or salience of the percept varies as a function of the fundamental frequency of the complex tone, which psychoacousticians describe using terms like *tonality*,[55] *tonalness*,[56] and *toneness*.[57] In a model of pitch that simulates the toneness of a pitch percept, Ernst Terhardt and his co-authors demonstrated that sensitivity was most acute in the *spectral dominance region*, which is centered

---

[49] David Huron, "Tone and Voice: A Derivation of the Rules of Voice-Leading from Perceptual Principles," *Music Perception* 19, no. 1 (2001): 7.

[50] Brian C. J. Moore, Robert W. Peters, and Brian R. Glasberg, "Thresholds for the Detection of Inharmonicity in Complex Tones," *Journal of the Acoustical Society of America* 77, no. 5 (1985): 1861–1867.

[51] McAdams and Drake, "Auditory Perception and Cognition," 400.

[52] Ibid., 403.

[53] Rudolf A. Rasch, "The Perception of Simultaneous Notes such as in Polyphonic Music," *Acustica* 40 (1978): 21–33.

[54] McAdams explains that auditory attributes like loudness and pitch are emergent properties of the grouping process, which is to say that ASA precedes the perception of these attributes ("Music: A Science of Mind?," 43; , "Recognition of Sounds Sources and Events," in *Thinking in Sound: The Cognitive Psychology of Human Audition*, ed. Stephen McAdams and Emmanuel Bigand [Oxford: Oxford University Press, 1993], 146–198).

[55] ANSI, *Psychoacoustical Terminology* (New York: American National Standards Institute, 1973).

[56] Richard Parncutt, *Harmony: A Psychoacoustical Approach* (Berlin: Springer-Verlag, 1989).

[57] Huron, "Tone and Voice."

near 300 Hz for complex tones—roughly D4 above middle C.[58] Thus, harmonic sounds also cohere more readily if the virtual pitch appears in this region.

In sum, concurrent grouping processes play a fundamental role in the formation of auditory events, particularly for the sorts of complex periodic sounds we find in music.[59] But are these presumably low-level processes laid down in the genes, or learned over the course of development? For Bregman, ASA depends to some degree on both explanations, but since the structure and function of the auditory periphery is likely innate, evolutionary inheritance is a necessary starting point.

> ... if you think of the physical world as having a "grammar" (the physical laws that are responsible for the sensory impressions that we receive), then each human must be equipped either with mechanisms capable of learning about many of these laws from examples or with a mechanism whose genetic program has been developed once and for all by the species as a result of billions of parallel experiments over the course of history, where the lives of the members of the species and its ancestors represent the successes and the lives of countless extinct families the failures. To me, evolution seems more plausible than learning as a mechanism for acquiring at least a general capability to segregate sounds. Additional learning-based mechanisms could then refine the ability of the perceiver in more specific environments.[60]

Over the past two decades, a number of studies have supported this view by identifying

[58]Ernst Terhardt, Gerhard Stoll, and Manfred Seewann, "Algorithm for Extraction of Pitch and Pitch Salience from Complex Tonal Signals," *Journal of the Acoustical Society of America* 71, no. 3 (1982): 679–688; Huron, "Tone and Voice," 7–8.

[59]In addition to concurrent grouping processes, ASA also concerns those *sequential* grouping processes that connect events into auditory streams, and those *segmental* grouping processes that segment (or "chunk") the event streams into auditory units like motives and phrases. For a review of these grouping processes in ASA, see Diana Deutsch, "Grouping Mechanisms in Music," in *The Psychology of Music*, 3rd ed., ed. Diana Deutsch (New York: Academic Press, 2013), 195–249; McAdams and Drake, "Auditory Perception and Cognition," 407–418.

[60]Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, 39–40.

concurrent or sequential grouping processes in newborn infants[61] as well as non-human animals, like macaque monkeys,[62] european starlings,[63] and even goldfish.[64] According to Bregman, this evidence suggests that the grouping processes underlying ASA "deal with a world in which the sound of a particular source is important for survival."[65]

In light of my earlier discussion of stability and expectation, it seems reasonable to assume that listeners process complex periodic events more quickly than non-periodic events because the former cohere more readily in the auditory periphery than the latter, particularly if those events fall within the spectral dominance region identified by Terhardt and his co-authors. Nevertheless, this assumption says very little about the hierarchies of stability described by Meyer and Krumhansl. Do the grouping processes underlying auditory event formation apply beyond the level of the note to event complexes like dyads and triads, or to harmonic progressions of the sort found in cadential contexts?

To extend the grouping processes governing ASA to more sophisticated musical objects like chords, melodies, and fused composite timbres, McAdams represents the combined aspects of a set of auditory impressions as an *auditory image*, a psychological representation of a sound entity that exhibits an internal coherence in its acoustic behavior.[66] In his view, the most

---

[61]Stephen McAdams and Josiane Bertoncini, "Organization and Discrimination of Repeating Sound Sequences By Newborn Infants," *Journal of the Acoustical Society of America* 102, no. 5 (1997): 2945–2953; István Winkler et al., "Newborn Infants Can Organize the Auditory World," *Proceedings of the National Academy of Sciences* 100, no. 20 (2003): 11812–11815.

[62]Yonatan I. Fishman et al., "Neural Correlates of Auditory Stream Segregation in Primary Auditory Cortex of the Awake Monkey," *Hearing Research* 151, nos. 1–2 (2001): 167–187.

[63]Stewart H. Hulse, Scott A. MacDougall-Shackleton, and Amy B. Wisniewski, "Auditory Scene Analysis By Songbirds: Stream Segregation of Birdsong by European Starlings (*Sturnus vulgaris*)," *Journal of Comparative Psychology* 111, no. 1 (1997): 3–13.

[64]Richard R. Fay, "Auditory Stream Segregation in Goldfish (*Carassius auratus*)," *Hearing Research* 1230, nos. 1–2 (1998): 69–76.

[65]Albert S. Bregman, "Progress in Understanding Auditory Scene Analysis," *Music Perception* 33, no. 1 (2015): 17.

[66]Stephen McAdams, "Spectral Fusion, Spectral Parsing and the Formation of Auditory Images" (PhD Dissertation, Stanford University, 1984); McAdams, "The Auditory Image: A Metaphor for Musical and Psychological Research on Auditory Organization."

powerful aspect of this concept is that it allows for the hierarchical organization of more complex perceptual structures from less complex structures by appealing to the same grouping processes at each level of the hierarchy.[67] He explains,

> It would appear that what is derived as a perceptual quality after one level of auditory organization may become an "element" contributing to grouping decisions at a higher level of organization. For example, the separation of several series of harmonic frequencies and the subsequent perceptual fusion of each one gives rise to a distinct pitch for each series. At the next level of organization, some of these pitches may be grouped according to the musical context as members of a common chord whose quality of being consonant or dissonant arises from their being considered as a group.[68]

According to this view, the concept of coherence applies in some general way to *any* set of auditory events, so long as they can be "perceived as a whole, as a single image."[69] For hierarchical auditory images like vertical sonorities, for example, the perceptual difference between one level (e.g., tones) and another (e.g., chords) is a matter of degree rather than kind.[70] This would imply that any of the factors that promote perceptual fusion or concurrent grouping also apply for vertical sonorities,[71] though Bregman also points out that the effects of these factors diminish as we ascend the organizational hierarchy.

> The fusion in chord perception is not as strong as the fusion of the partials of a single note or as weak as in the perception of unrelated sounds, but falls somewhere

[67] McAdams, "The Auditory Image: A Metaphor for Musical and Psychological Research on Auditory Organization," 291.

[68] McAdams, "Music: A Science of Mind?," 43.

[69] McAdams, "The Auditory Image: A Metaphor for Musical and Psychological Research on Auditory Organization," 291.

[70] David Huron, Review of *Harmony: A Psychoacoustical Approach*, by Richard Parncutt, *Psychology of Music* 19, no. 2 (1991): 219.

[71] Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, 509.

> between the two extremes. Perhaps we ought to call it vertical grouping rather than fusion. This vertical grouping contributes to the experience of chord perception in which the chord, rather than the individual tones, becomes the acoustic object that we hear. Thus, consonant pitch relations may be one of the bases for the vertical links between parts that hold the piece of music together.[72]

The earliest explanations for "vertical grouping" derived consonant pitch relations from simple frequency ratios or the partials of a complex tone (the *Klang*, or *corps sonore*),[73] but physiologist Hermann von Helmholtz provided the first systematic attempt to link the auditory periphery to vertical sonorities and the rules governing their succession. Armed with anatomical evidence disclosed by the use of the compound microscope,[74] Helmholtz theorized that the perception of consonance is a sensory response caused by (1) the absence of rapid beating (oscillations in amplitude) created by interactions of adjacent partials, which Terhardt would later call *sensory consonance*; and (2) the correspondence of partials between two or more tones, which amounts to an extension of Bregman's harmonicity principle.[75]

Psychoacoustic studies published over the last century have validated many of Helmholtz's

---

[72] Ibid., 495–496.

[73] Appealing to the frequency ratio between the fundamentals of two tones dates back to antiquity. Using a single-stringed instrument called a monochord, the ancient Greeks defined a gamut of pitches by specifying the numeric ratio between the length of the string segment that produces one pitch and that of the segment that produces some other. They observed that simple ratios like $\frac{2}{1}$ and $\frac{3}{2}$ produce more beautiful, euphonius, or consonant intervals than complex ratios like $\frac{16}{15}$ (Jan Herlinger, "Medieval Canonics," in *The Cambridge History of Western Music Theory*, ed. Thomas Christensen [Cambridge: Cambridge University Press, 2002], 168). More sophisticated explanations based on the acoustic properties of individual tones first appeared in the later writings of the French theorist and composer Jean-Philippe Rameau, who attempted to derive the principles of tonal harmony from the overtone series (or *corps sonore*). For a discussion of the influence of acoustics on Rameau's theory of harmony, see Nathan John Martin, "Rameau and Rousseau: Harmony and History in the Age of Reason" (PhD Dissertation, McGill University, 2008), 19–74.

[74] Burdette Green and David Butler, "From Acoustics to Tonpsychologie," in *The Cambridge History of Western Music Theory*, ed. Thomas Christensen (Cambridge: Cambridge University Press, 2002), 259.

[75] Hermann L. F. Helmholtz, *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*, 2nd ed., trans. Alexander J. Ellis (London: Longmans, Green, 1877/1885), 185–196; Terhardt, "Pitch, Consonance, and Harmony"; Ernst Terhardt, "The Concept of Musical Consonance: A Link between Music and Psychoacoustics," *Music Perception* 1, no. 3 (1984): 276–295.

findings with respect to the sensory consonance and harmonicity of dyads and triads. Studies of sensory consonance essentially began with a series of experiments by Reinier Plomp and Willem J. M. Levelt. Following the seminal discovery of *critical bands*, which represent the bands of auditory nerve fibers along the basilar membrane that respond selectively to a small range of frequencies,[76] Plomp and Levelt determined that the sensory dissonance (or *roughness*) caused by two adjacent tones peaks when the frequency difference equals around 25% of the critical bandwidth.[77] Subsequent algorithms by William Hutchinson and Leon Knopoff extended these findings to triads and replicated several music-theoretic intuitions about the consonance of intervals found in Western common practice, leading them to suggest that musical consonance results in part from the absence of beating among partials, just as Helmholtz predicted.[78]

Nevertheless, sensory explanations for musical consonance suffer from several limitations. First, they generally fail to predict the relative dissonance of the most common chord types. For ratings of "harmoniousness" of individual chords, for example, the following rank order of increasing dissonance obtains: major < minor < diminished < augmented.[79] Models of roughness based on acoustic beating predict that the augmented triad should be more consonant than the diminished triad, however.[80] Second, for both sine tones and complex periodic tones, models of roughness tend to produce higher consonance ratings as the size of the interval increases. As a consequence, an interval like the major seventh receives higher consonance ratings than the perfect fifth, a finding that clearly contradicts music-theoretic predictions.[81] To

[76]In other words, the critical bandwidth represents the range of frequencies over which the basilar membrane fails to resolve simultaneous tones or partials.

[77]Reinier Plomp and Willem J. M. Levelt, "Tonal Consonance and Critical Bandwidth," *Journal of the Acoustical Society of America* 38, no. 4 (1965): 548–560.

[78]William Hutchinson and Leon Knopoff, "The Acoustic Component of Western Consonance," *Interface* 7, no. 1 (1978): 1–29; William Hutchinson and Leon Knopoff, "The Significance of the Acoustic Component of Consonance in Western Triads," *Journal of Musicological Research* 3, nos. 1-2 (1979): 5–22.

[79]Norman D. Cook and Takashi X. Fujisawa, "The Psychophysics of Harmony Perception: Harmony is a Three-Tone Phenomenon," *Empirical Music Review* 1 (2006): 106–126.

[80]Johnson-Laird, Kang, and Leong, "On Musical Dissonance," 21.

[81]Terhardt, "The Concept of Musical Consonance: A Link between Music and Psychoacoustics," 280–281.

be sure, the German philosopher-psychologist Carl Stumpf criticized the Helmholtzian view of sensory consonance for precisely this reason. In Stumpf's view, tonal consonance is a top-down, psychological response to the tonal fusion (*Tonverschmelzung*) of simultaneous tones—the phenomenon of two tones blending to the extent that they are perceived as "unitary."[82]

To resolve these issues, many psychoacousticians—including Helmholtz himself—have suggested that musical consonance reflects two components: (1) the sensory consonance of dyads and other vertical sonorities resulting from biological constraints of the auditory periphery, and (2) tacit knowledge acquired during exposure to the principles of tonal music.[83] In a recent study examining the interaction between musical experience and presumably low-level sensory factors like beating and harmonicity, Josh H. McDermott and his co-authors found that harmonicity measures were positively correlated with musical training, leading them to suggest that previous experience enhances an innate bias for harmonic sounds. To be sure, the assumption that listeners *learn* to process and even prefer harmonic sounds across repeated exposures appears frequently in the experimental literature.[84] The well-known virtual pitch algorithms designed by Terhardt and Parncutt—which extend the harmonicity principle to vertical sonorities like dyads and triads—are noteworthy in this regard. Following a long tradition dating back to Rameau, they assumed that extracting the fundamental frequency of a complex tone (the *corps sonore*) is equivalent to determining the root of a chord (the *basse fondamentale*). But in their view, the principles governing tonal harmony result from perceptual *familiarity* with the complex periodic sounds associated with mammalian vocalizations. Terhardt explains,

> The whole learning process postulated by [Helmholtz], in which an individual

[82]For a brief introduction to Stumpf's *Tonpsychologie*, see Green and Butler, "From Acoustics to Tonpsychologie," 263–265.

[83]Helmholtz, *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*; Johnson-Laird, Kang, and Leong, "On Musical Dissonance"; Terhardt, "The Concept of Musical Consonance: A Link between Music and Psychoacoustics."

[84]See, for example, Josh H. McDermott et al., "Indifference to Dissonance in Native Amazonians Reveals Cultural Variation in Music Perception," *Nature* 535 (2016): 547–550.

> acquires familiarity with the basic musical intervals by aural analysis of harmonic complex tones, will with highest probability take place as an essential part of the perception of speech, as speech is the most significant auditory signal to humans. The purpose of that learning process, according to virtual-pitch theory, is to enable the auditory system to extract virtual pitch from any complex tone, even when some harmonics (in particular the fundamental) are not present. It is thus a sort of byproduct that in that learning process a sense of harmonic intervals (octave, fifth, etc.) is acquired. The biological significance of that sense lies in its being a means to extract virtual pitch from voiced speech sounds. The harmonic pitch intervals are just "acquired system parameters."[85]

The view that harmonicity and other concurrent grouping processes depend on top-down mechanisms related to learning and memory is consistent with McAdams' description of auditory imagery. In his view, "any listener still carries into the musical situation all of the 'perceptual baggage' acquired from ordinary in-the-world perceiving."[86] As a consequence, the mechanisms by which listeners cohere auditory events into auditory images reflect both bottom-up and top-down processes; listeners may "have a biological predisposition" to cohere some event complexes more readily than others, but "what is coherent psychologically may evolve with specialized musical experience."[87]

Thus, while it seems reasonable to suggest that the stability of events and event complexes like notes or triads depends at least to some degree on the biological constraints of sensory processing, these sorts of explanations still fail to account for the pattern of observations encountered in experimental studies of consonance and dissonance, stability, and so forth.

[85] Terhardt, "The Concept of Musical Consonance: A Link between Music and Psychoacoustics," 288.

[86] McAdams, "The Auditory Image: A Metaphor for Musical and Psychological Research on Auditory Organization," 290.

[87] McAdams, "Music: A Science of Mind?," 39.

One might argue, for example, that perfect intervals like the octave appear frequently in many musical styles because they reflect concurrent grouping principles like harmonicity and sensory consonance that trick the auditory periphery into partially fusing their constituent events. But it is also important to note that not all of the world's musical scales reflect the concept of octave equivalence.[88] What is more, in studies that require participants to judge the similarity of isolated tone-pairs, psychologists Edward M. Burns and W. Dixon Ward point out that musical training influences similarity judgments for octave intervals. Since "musicians tend to show significant octave effects" when nonmusicians show "little or none,"[89] they conclude that octave equivalence might be learned.[90] In short, it is conceivable that top-down mechanisms related to learning and memory at least partly determine the stability or coherence of individual events like notes and chords even at the lowest levels of sensory organization.

This dual-component view of stability is perhaps best exemplified in a series of studies examining the perception of tonal melodic closure in the early history of music psychology. Following Stumpf's theory of tonal fusion, which revived the Pythagorean explanation for the consonance of dyads in terms of simple frequency ratios, Theodor Lipps and Max F. Meyer proposed the "law of the power of 2" (sometimes called the *Lipps-Meyer Law*), which states that one of any two tones from a diatonic scale is more final if its number in the frequency ratio is a power of two. Meyer justified this view by appealing to the overtone series; a melody, he claimed, "moves from overtones to the fundamental tone."[91] Nearly two decades later, his student Paul Farnsworth demonstrated in a series of experimental studies that the repetition of

[88]Edward M. Burns and W. Dixon Ward point out that these scales typically appear in cultures that are either pre-instrumental or that use inharmonic instruments of the xylophone type ("Intervals, Scales, and Tuning," in *The Psychology of Music*, ed. Diana Deutsch [New York: Academic Press, 1982], 258).

[89]Edward M. Burns, "Intervals, Scales, and Tuning," in *The Psychology of Music*, 2nd, ed. Diana Deutsch (New York: Academic Press, 1998), 253.

[90]Burns and Ward, "Intervals, Scales, and Tuning," 264.

[91]Max Friedrich Meyer, "Experimental Studies in the Psychology of Music," *The American Journal of Psychology* 14, nos. 3/4 (1903): 193.

certain sequences over the course of the experimental session influences the perceived finality of the terminal note.[92] This training effect led Farnsworth to propose an alternative account based on the "habit principle," in which the discrimination of finality depends on the *association* of a ratio symbol with the terminal position in a tonal succession.[93] In fact, W. Van Dyke Bingham criticized the Lipps-Meyer Law soon after its publication for precisely this reason. For Bingham, the laws of consonance and dissonance—on which the Lipps-Meyer Law was presumably based—represent just one component of the sensory-cognitive apparatus, the other being predicated on the formation of top-down associations over the course of exposure. The conclusion of his dissertation brilliantly captures the distinction between nature and nurture.

> The operation of two main forces must be distinguished—one of them sensory, the other associative. The first of these, the phenomenon of consonance, is native and doubtless has its basis in the relatively simple action of the sensory apparatus in responding to auditory stimuli which are more or less similar—are, indeed, in a measure identical. But although the basis for consonance inheres in the inborn structure of the nervous system and the acoustical properties of vibrating bodies, nevertheless it is a commonplace of musical history and observation that these same native tendencies are subject to tremendous modification in the course of experience. One race, one age hears as consonant intervals which another age or race has never learned to tolerate; and within the history of individuals it is easily observable that consonance and dissonance are merely relative terms whose denotation shifts with growing experience. Moreover the whole complex group of phenomena we call tonality bears witness to the power of association to amplify

[92] Paul R. Farnsworth, "The Effect of Repetition on Ending Preferences in Melodies," *The American Journal of Psychology* 37, no. 1 (1926): 116–122.

[93] Paul R. Farnsworth, "Ending Preferences in Two Musical Situations," *The American Journal of Psychology* 37, no. 2 (1926): 238; Farnsworth, "The effect of repetition on ending preferences in melodies"; Paul R. Farnsworth, "A Modification of the Lipps-Meyer Law," *Journal of Experimental Psychology: General* 9, no. 3 (1926): 253–258.

and organize these native feelings.[94]

1.2.2 Cognitive Principles

Throughout the previous section I made repeated reference to organizational systems based on pitch and temporal duration, two of the most important auditory attributes in perception that serve as form-bearing dimensions in many compositional styles,[95] providing the basis for musical concepts like consonance and dissonance, melody and voice leading, harmony, tonality, and meter. In my view, these systems depend to some degree on the biological constraints of the auditory periphery, by which listeners form auditory images for the kinds of complex periodic sounds encountered in many musical traditions. But since these constraints generally fail to account for more sophisticated aspects of musical organization, such as the consonance and stability of event complexes like dyads and triads, we might also assume that the hierarchies of stability characterizing classical music reflect cognitive processes related to memory and learning.

Memory is an enormous concept in experimental psychology. Since I mention stores (e.g., short-term, long-term) and types (e.g., declarative, procedural) of various sorts in subsequent chapters, it will be useful to review these notions briefly here. Richard C. Atkinson and Richard M. Shiffrin published what remains the canonical multi-store model, which divides memory into three interrelated stores that vary according to the duration of retention: the *sensory register* (or *sensory memory*), *short-term* (or *working*) *memory*,[96] and *long-term memory*. Sensory memory is an ultra short-term memory buffer for the immediate storage of sensory information that

[94] W. Van Dyke Bingham, *Studies in Melody* (Baltimore: Review Publishing Company, 1910), 87.

[95] Stephen McAdams, "Psychological Constraints on Form-Bearing Dimensions in Music," *Contemporary Music Review* 4 (1989): 181–198.

[96] For Atkinson and Shriffin, working memory and short-term memory are synonymous, but a few years after the publication of the Atkinson-Shriffin model, Alan Baddeley and Graham Hitch proposed an alternative multi-store model that separated the two concepts ("Working Memory," in *The Psychology of Learning and Memory*, ed. Gordon H. Bower, vol. 8 [New York: Academic Press, 1974], 47–89).

continuously decays until it is lost.[97] In the presence of competing information, the duration of the store varies depending on the sensory modality; for vision (which is sometimes called *iconic memory*), Atkinson and Shiffin suggest estimates of no more than a second,[98] but for audition (which Ulric Neisser termed *echoic memory*),[99] the upper limit is closer to four seconds.[100] Thus, for our purposes, auditory sensory (or echoic) memory is responsible for the representation of auditory images like tones and chords, by which listeners register attributes like loudness and pitch. Like sensory memory, short-term memory (STM) decays and then disappears, but the temporal duration of the decay is considerably longer, perhaps as long as ten to twelve seconds.[101] The capacity of short-term memory is not boundless, however. In this case, the upper limit proposed by George A. Miller of 7 ± 2 events is a good rule of thumb.[102] According to Atkinson and Shriffin, STM receives selected inputs from sensory memory (i.e., from the bottom up) and from long-term memory (i.e., from the top down). Thus listeners in possession of long-term mental representations for vertical sonorities like dyads and triads might transfer this information to STM to seek out similar structures. Finally, long-term memory (LTM) is "a fairly permanent repository,"[103] which is to say that it decays very little over time.

The transfer process from one store to the other entails mechanisms like conscious attention during perception and rehearsal of information in STM. Atkinson and Shriffin also suggest that the transfer process between LTM and STM is bi-directional in that long-term mental representations for the kinds of objects and events encountered in everyday life may trigger the

[97] Richard C. Atkinson and Richard M. Shiffrin, "Human Memory: A Proposed System and Its Control Processes," in *The Psychology of Learning and Motivation*, ed. Kenneth Wartenbee Spence and Janet T. Spence, vol. 2 (New York: Academic Press, 1968), 90.

[98] Ibid., 92.

[99] Ulric Neisser, *Cognitive Psychology* (New York, NY: Taylor & Francis, 1967/2014), 190.

[100] Christopher J. Darwin and Michael T. Turvey, "An Auditory Analogue of the Sperling Partial Report Procedure: Evidence for Brief Auditory Storage," *Cognitive Psychology* 3 (1972): 255–267.

[101] Snyder, *Music and Memory*, 50.

[102] George Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychological Review* 63 (1956): 81–97.

[103] Atkinson and Shiffrin, "Human Memory: A Proposed System and Its Control Processes," 91.

organization and recognition of those objects and events in STM. It is also important to note that Atkinson and Shriffin limit top-down information transfers to STM, which in our case means that echoic memory is excluded from higher-level memory processes. Many researchers have criticized this view, however, since a number of studies over the past few decades have demonstrated top-down effects for even the most primitive aspects of sensory organization.[104] Thus mental representations from LTM and STM for organizational structures like chords, clarinet melodies, or composite timbres are likely to influence the formation of auditory images in echoic memory. As Neisser puts it, "in perceiving, the immediate past and the remote past are brought to bear upon the present."[105]

How might the processes related to STM constrain the emergence of organizational systems like tonality and meter, in which certain events are more stable than others? McAdams points out that the sensory-cognitive apparatus enhances the encoding and recognition of pitch and temporal duration—the auditory attributes on which these systems depend—by organizing values along these dimensions into a relatively small number of perceptually discrete, proportionally related categories.[106] Much of the world's music relies on musical scales, for example, which typically feature discrete and discriminable steps, octave equivalence, and a moderate number of degrees within the octave.[107] Without these sorts of constraints, listeners would be far less likely to learn and remember the patterns that characterize a given composition or repertory. What is more, the relative fixity of the categories along these dimensions allows

[104] Bregman refers to the influence of top-down knowledge on ASA grouping principles as "schema-based" processing effects (*Auditory Scene Analysis: The Perceptual Organization of Sound*, 395–454).

[105] Ulric Neisser, *Cognition and Reality* (San Francisco: W. H. Freeman and Company, 1976), 14. Despite the plethora of effects associated with each store, contemporary research continues to dispute the necessity of compartmentalizing memory in this way. According to Reber, for example, a singular memory process may underlie each of the various stores (*Implicit Learning and Tacit Knowledge*, 80–81).

[106] McAdams, "Psychological Constraints on Form-Bearing Dimensions in Music," 181; see also Leonard Meyer, "A Universe of Universals," *The Journal of Musicology* 16, no. 1 (1998): 8.

[107] Dowling and Harwood, *Music Cognition*, 91.

listeners to encode the relations among values within a given pattern.[108] As a consequence, transpositions of patterns in pitch or duration that preserve the relations between categories are more easily recognized during perception than those that do not.[109]

But what about the influence of LTM on lower memory processes, or on perception itself? Unfortunately, the storage metaphor is not particularly helpful in this case because it fails to capture the biological purpose of memory. Following the cognitive revolution, associationist theories like schema theory, connectionism, and predictive coding have suggested that the mind builds "mental models" of the external environment whose purpose is to predict the future. From this point of view, memory exists "not to allow us to relish past successes or regret past failures, but to allow us to repeat our successes and avoid future failures." In short, its purpose is "not *recall* but *preparation*."[110] British psychologist Kenneth Craik summarizes this view thusly:

> If the organism carries a 'small-scale model' of external reality and its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer and more competent manner to the emergencies which face it.[111]

But what form do our mental representations take, and how do we acquire them? We might further divide LTM into two types: explicit (or declarative) memory and implicit (or procedural) memory. Explicit memory denotes the representation of facts and events that

[108] McAdams explains that "our ability to recognize transposed or accelerated musical patterns testifies to the psychological reality of relational encoding" ("Psychological Constraints on Form-Bearing Dimensions in Music," 185).

[109] The auditory periphery plays some part in explaining why these auditory attributes serve as form-bearing dimensions and not others, but the point here is that the constraints placed on STM must also influence the construction of scales, the temporal duration of motives and phrases, and many other aspects of musical organization.

[110] David Huron, *Sweet Anticipation: Music and the Psychology of Expectation* (Cambridge, MA: MIT Press, 2006), 219.

[111] Kenneth Craik, *The Nature of Explanation* (Cambridge: Cambridge University Press, 1943), 61.

require conscious recall, such as the age of the Earth or the date of one's birthday. For these sorts of representations, the transfer from STM to LTM typically requires an explicit learning strategy based on conscious attention and repeated rehearsal. Implicit memory refers to our unconscious representations of how things might go, such as playing chess or conversing with friends. For these representations, the acquisition process is assumed to be based on an implicit learning strategy that *abstracts* the pattern of co-occurring features an object or event might share with previously encountered exemplars. This latter type of memory is presumably what governs our mental representations for hierarchical organizational systems like natural language or tonal music.

To justify this view, psychologist Arthur Reber proposed a domain-general learning mechanism to explain how humans learn their native language(s) during early development.[112] Unlike many of his nativist peers who believed that knowledge is not "acquired" at all but given a priori, Reber assumed that much of the knowledge necessary to comprehend complex, seemingly "rule-governed" syntactic structures is acquired tacitly, beneath conscious awareness. He referred to this acquisition process as *implicit learning* (IL), arguing that it "takes place largely independently of conscious attempts to learn and largely in the absence of explicit knowledge about what was required."[113] According to Reber, IL represents the default mode of knowledge acquisition, one subserved by procedural memory centers that are phylogenetically older than those underlying declarative memory.[114] Thus in his view, the acquisition of explicit, reflective, and declarative knowledge represents just the tip of the iceberg, with implicit, nonreflective, and procedural learning residing beneath the conscious surface.[115]

[112] Arthur S. Reber, "Implicit Learning of Artificial Grammars" (MA Thesis, Brown University, 1965); Arthur S. Reber, "Implicit Learning of Artificial Grammars," *Journal of Verbal Learning and Verbal Behavior* 6 (1967): 855–863.

[113] Reber, *Implicit Learning and Tacit Knowledge*, 5.

[114] Larry R. Squire, "Mechanisms of Memory," *Science* 232, no. 4758 (1986): 1615.

[115] Reber conceives of consciousness as a late arrival on the evolutionary scene. It is, in his words, "an emergent aspect of a complex brain" (*Implicit Learning and Tacit Knowledge*, 86).

But just how does one learn implicitly? Stripped down to its bare bones, IL is simply the detection of statistical regularities in the environment.[116] For Reber, a mental representation for a major triad is abstracted over time from the exemplars of that category that we encounter in classical music because the essential characteristics that apparently define the category co-occur with remarkable frequency. Reber calls this the *abstractive view* of knowledge acquisition,[117] and it pertains to all sorts of contexts. McAdams suggests, for example, that organizational systems like tonality and meter represent *abstract knowledge structures* because they embody a system of relations that apply across a large number of exemplars. In his view, these sorts of structures "serve to establish the relative stability or salience relations among the values along a given dimension. This domain is perhaps the most important for the consideration of form-bearing capacity because it is clear that if a system of habitual relations among values along a dimension cannot be learned, the power of that dimension as a structuring force would be severely compromised."[118] In other words, the stability of isolated events like notes and chords in tonal music is primarily *learned*.

To demonstrate how we might learn the statistical regularities governing these systems, IL studies typically use arbitrary and unfamiliar stimulus domains with complex, rule-governed, idiosyncratic structures. In the first IL study, Reber employed a finite-context, Markovian artificial grammar, which consists of a small alphabet of letters and a distribution of probabilities governing the transitions between them. In his case, the grammar produced a finite number of grammatical strings from three to eight letters in length. The basic procedure for IL experiments has two components: (1) an acquisition phase, during which participants memorize a subset of the grammatical strings the grammar can produce; and (2) a testing phase, during which participants determine the grammatical status of a series of new strings, some of which

[116]For this reason, it also sometimes called *implicit statistical learning*.
[117]Reber, *Implicit Learning and Tacit Knowledge*, 120–121.
[118]McAdams, "Psychological Constraints on Form-Bearing Dimensions in Music," 183.

were generated by the artificial grammar, and some not. Reber demonstrated over several experiments that participants could reliably distinguish grammatical from non-grammatical strings, suggesting they had become sensitive to the constraints of the grammar just by exposure to exemplary strings.[119]

Over the past five decades, researchers have demonstrated IL effects for sequences of letters,[120] abstract visual shapes,[121] and animal pictures.[122] In the music domain, the evidence suggests that listeners can implicitly learn familiar[123] and unfamiliar musical systems,[124] serial music,[125] and synthesized instrumental timbres.[126] A few studies have also generalized these findings to melodic and harmonic contexts using artificial grammars and probabilistic learning paradigms. In both cases, the results indicate that listeners can acquire the statistical regularities governing novel melodic or harmonic grammars during short-term exposure.[127] The size of this learning effect also varies as a function of the complexity of the grammar[128] and the length

[119] Reber, "Implicit Learning of Artificial Grammars."

[120] For a review of language acquisition using IL paradigms, see Jenny R. Saffran, "Statistical Language Learning: Mechanisms and Constraints," *Current Directions in Psychological Science* 12, no. 4 (2003): 110–114.

[121] József Fiser and Richard N. Aslin, "Statistical Learning of Higher-Order Temporal Structure from Visual Shape Sequences," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28, no. 3 (2002): 458–467.

[122] Jenny R. Saffran et al., "Dog is a Dog is a Dog: Infant Rule Learning is Not Specific to Language," *Cognition* 105, no. 3 (2007): 669–680.

[123] Jenny R. Saffran et al., "Statistical Learning of Tone Sequences By Human Infants and Adults," *Cognition* 70 (1999): 27–52.

[124] Psyche Loui, "Humans Rapidly Learn Grammatical Structure in a New Musical Scale," *Music Perception* 27, no. 5 (2010): 377–388.

[125] Zoltán Dienes and Christopher Longuet-Higgins, "Can Musical Transformations Be Implicitly Learned?," *Cognitive Science* 28 (2004): 531–558.

[126] Barbara Tillmann and Stephen McAdams, "Implicit Learning of Musical Timbre Sequences: Statistical Regularities Confronted with Acoustical (Dis)Similarities," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30, no. 5 (2004): 1131–1142.

[127] Erin Jonaitis and Jenny Saffran, "Learning Harmony: The Role of Serial Statistics," *Cognitive Science* 33 (2009): 951–968; Martin Rohrmeier and Ian Cross, "Tacit Tonality: Implicit Learning of Context-Free Harmonic Structure," in *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music*, ed. Jukka Louhivuori et al. (Jyväskylä: University of Jyväskylä, 2009), 443–452; Martin Rohrmeier, Patrick Rebuschat, and Ian Cross, "Incidental and Online Learning of Melodic Structure," *Consciousness and Cognition* 20 (2011): 214–222.

[128] Rohrmeier and Cross, "Tacit Tonality: Implicit Learning of Context-Free Harmonic Structure."

of the exposure phase,[129] which is to say that stimuli produced by less complex grammars or presented over a longer exposure phase tend to increase the participants' ability to detect non-grammatical sequences.

In sum, the evidence is overwhelming that participants can induce the statistical structure inherent in complex (musical) stimuli over a relatively short period of time. For our purposes, this means that listeners learn more complex hierarchical organizational systems like tonality and meter over the course of exposure if their most stable events appear more frequently in the classical style. Bharucha and Krumhansl would seem to agree, suggesting that stable tones "appear more frequently, in prominent positions, and with rhythmic stress."[130] Once learned, our statistical knowledge of these organizational systems "is thought to impose its organization on all subsequent musical experiences,"[131] allowing us to build mental models of our immediate experiences and predict future outcomes.

§1.3 Conclusions

I noted in §1.1 that definitions of closure often trespass on theories of learning, memory, and perception. To the degree that they reflect the perception of closure, theories of segmentation, formal function, and the law of return—as well as associated psychological effects like stability and rest—are not just overlapping magisteria in the study of musical experience, representing distinct but complementary areas of inquiry and potentially separable psychological processes. In my view, these theories—and the processes underlying them—result from a basic function of the human mind: that of prediction. Among music scholars, this view was first crystallized by Meyer, but the resurgence of associationist theories in the cognitive sciences over the past

[129] Jonaitis and Saffran, "Learning Harmony."
[130] Bharucha and Krumhansl, "The Representation of Harmonic Structure in Music," 64.
[131] Krumhansl, *Cognitive Foundations of Musical Pitch*, 284.

few decades has placed the study of the brain's predictive mechanisms at the forefront of contemporary scholarship. This is not to say that our tendency to continuously expect certain outcomes over others explains *every* aspect of closure; not all of the compositional procedures associated with a theory of closure will feature prominently in the minds of listeners. Rather, my point is that those aspects of closure that *do* correspond with the reception of musical works—such as segmental grouping, temporal functionality, and the perception of stability—reflect the brain's predictive processing strategies, where minimizing uncertainty about future events is a biological imperative.

But how do we generate expectations during music listening? I suggested in §1.2 that psychologists appeal to the term *stability* to refer to hierarchical systems of relations like tonality and meter under which certain events are more stable in memory and facilitate processing during perception. These "hierarchies of stability" depend on bottom-up processes of sensory organization and top-down processes related to memory and learning. In other words, listeners are more likely to learn and remember events and event complexes like tones and chords if they (1) cohere more readily in the auditory periphery, and (2) appear more frequently in a composition or repertory.

But what about recurrent temporal patterns like cadences? Do sensory principles play some part in the formation of these sorts of structures in tonal organization? Recall that for Bregman, the factors governing primitive sensory organization diminish as we ascend the organizational hierarchy. Event complexes like dyads and triads may reflect emergent properties of sensory organization in some general way, for example, but beyond this level—say, for cadences and other formulaic closing patterns—sensory accounts are increasingly unlikely. As Bregman puts it, these sorts of patterns "are specific to musical styles and are based on cognitive schemas, not primitive perceptual organization."[132] Thus, Chapter 2 offers a critique and discussion

[132] Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, 496.

of the cadence concept in the "New *Formenlehre*" tradition, and then appeals to theories of schematic organization articulated in the cognitive sciences to explain the acquisition and mental representation of cadences and other closing patterns.

Chapter 2

The Classical Cadence as a Closing Schema

The mind has become so habituated to the form of the ordinary perfect cadence that in a movement of highly emotional character it comes rather like a platitude.

William S. Rockstro

In the previous chapter I suggested that a sonority is psychologically *stable* if it 1) exhibits an internal coherence in its acoustic behaviour, and 2) appears frequently in a given style. The former relates to our capacity to organize concurrent sound events into a coherent auditory image, the latter to the likelihood that we will store that image in long-term memory. Putting it simply, sonorities that are maximally coherent and highly prevalent are more likely to serve as endings. But how do composers approach stable sonorities? And how do listeners know when the end is near?

In the classical style, stable events do not simply appear at random. Rather, the available compositional options for the events preceding stable sonorities were fairly constrained, resulting in the appearance of a limited number of highly stereotyped harmonic and melodic formulæ at the ends of phrases, themes, and larger sections—what theorists and composers have for centuries called *cadences*. To be sure, our preoccupation with applying the cadence concept to a dizzying variety of styles is a testament to the prevalence of conventionalized phrase endings in both Western and non-Western musics. Yet despite its foundational position in the *Formenlehre* tradition over the last few centuries,[1] the classical cadence continues to receive attention and undergo refinement in the scholarly community. The recent revival of interest in theories of musical form has prompted a number of studies that reconsider previously accepted explanations of how composers articulate cadences in the classical period,[2] that classify instances in which cadential arrival fails to materialize,[3] and that situate the concept of cadence within a broader understanding of both tonal and formal closure.[4] But despite such intense theoretical scrutiny, it remains unclear whether cadential patterns are represented in long-term memory, how they are perceived during music listening, and how the various constituent features of cadences contribute to the experience of closure.

In this chapter, I will suggest that the finality attributed to a given sonority is not only determined by its inherent psychological stability within the classical style, but also by the degree to which it is implied by parameters that appear in *prospect*. Indeed, my claim here is that listeners who are familiar with classical music have internalized the most common cadence types as a flexible network of interrelated mental representations, or what Robert Gjerdingen has called *rival event schemata*.[5] During music listening, the activation of this network in the prospective stage—and of the individual closing schemata contained within—results in the formation of expectations for the terminal event of the cadence, the fulfilment of which

[1] For a review of the various "doctrines of form" in the history of music theory, see Scott Burnham, "Form," in *The Cambridge History of Western Music Theory*, ed. Thomas Christensen (Cambridge: Cambridge University Press, 2002), 880–906.

[2] Caplin, "The Classical Cadence."

[3] Caplin, *Classical Form*, 101–111; Robert Hatten, "Interpreting Deception in Music," *In Theory Only* 12, no. 5 (1992): 31–50; Hepokoski and Darcy, *Elements of Sonata Theory*, 150–179; Janet Schmalfeldt, "Cadential Processes: The Evaded Cadence and the 'One More Time' Technique," *Journal of Musicological Research* 12, no. 1 (1992): 1–52.

[4] Anson-Cartwright, "Concepts of Closure."

[5] Robert O. Gjerdingen, *A Classic Turn of Phrase: Music and the Psychology of Convention* (Philadelphia: University of Pennsylvania Press, 1988), 62.

determines the perceptual boundary both for the schema itself, and perhaps more importantly, for the larger phrase-structural process that subsumes it.

The acquisition of this network of closing schemata is therefore essential to the perception of closure at local and middleground levels of musical organization. But how do listeners *acquire* the network? It should be evident that the preceding discussion makes an explicit statistical assumption about the nature of learning: namely, that the acquisition of mental representations results from the frequency of occurrence of patterns in the classical style. In short, a pattern that appears frequently is more likely to serve as a schema. Thus, by examining a large number of cadences from a relatively narrow, historically limited corpus, my goal here is to provide empirical evidence for the kinds of closing patterns that listeners may learn implicitly.

I begin in §2.1 by reviewing contemporary accounts of the classical cadence articulated in the "New *Formenlehre*" tradition, which identify the most common cadential types according to (1) their essential surface characteristics; and (2) the temporal context in which they appear. In §2.2, I outline the theories that explain the acquisition and mental representation of cadences and other frequently-occurring closing patterns, drawing particularly from research on expectation, implicit statistical learning, and schema theory. Finally, §2.3 applies Gjerdingen's schema-theoretic approach to the cadence typologies articulated in the *Formenlehre* tradition to consider whether listeners with sufficient exposure to classical music have internalized the most common cadence types as a flexible network of *rival closing schemata*.

§2.1 The Classical Cadence

2.1.1 Definitions

Of the many conventional figures and recurrent patterns that characterize the classical period, the cadence is generally regarded as a central feature.[6] In a style marked by "points of arrival, on every scale of magnitude, from the figure to the complete movement,"[7] cadential formulæ flourished in eighteenth-century compositional practice by serving to "mark the breathing places in the music, establish the tonality, and render coherent the formal structure," thereby cementing their position "throughout the entire period of common harmonic practice."[8] And yet, like many of the concepts in circulation in music scholarship (e.g., tonality, harmony, phrase, meter), the classical cadence has been extremely resistant to definition. To sort through the profusion of terms associated with cadence, Ann Blombach surveyed definitions in eighty-one textbooks distributed around a median publication date of 1970.[9] Shown in Figure 2.1, Blombach demonstrated that the cadence is associated most frequently with harmonic elements, though a number of other requisite components also feature, such as a characteristic melody or rhythm, the presence of a rest or pause, and so on.

From the elements listed in Figure 2.1, it appears that most definitions implicitly characterize the cadence as a *time span*, which consists of a conventionalized harmonic progression, and in

[6] Caplin, "The Classical Cadence," 51-52.

[7] Leonard G. Ratner, *Classic Music: Expression, Form, and Style* (New York: Schirmer Books, 1980), 33.

[8] Walter Piston, *Harmony*, 3rd ed. (New York: W. W. Norton & Company, 1962), 108. The view that a cadence establishes the tonality is widely held in music theory. In *Classical Form*, for example, William Caplin distinguishes those progressions that characterize a cadence at the end of a theme from those *prolongational* or *sequential* progressions that appear earlier in the theme. He writes, "the tonality itself is not made certain until its principal harmonic functions are articulated in a sufficiently powerful manner. It is thus the role of a *cadential* progression to confirm a tonal center as such" (*Classical Form*, 70; my italics).

[9] Ann Blombach, "Phrase and Cadence: A Study of Terminology and Definition," *Journal of Music Theory Pedagogy* 1 (1987): 229.

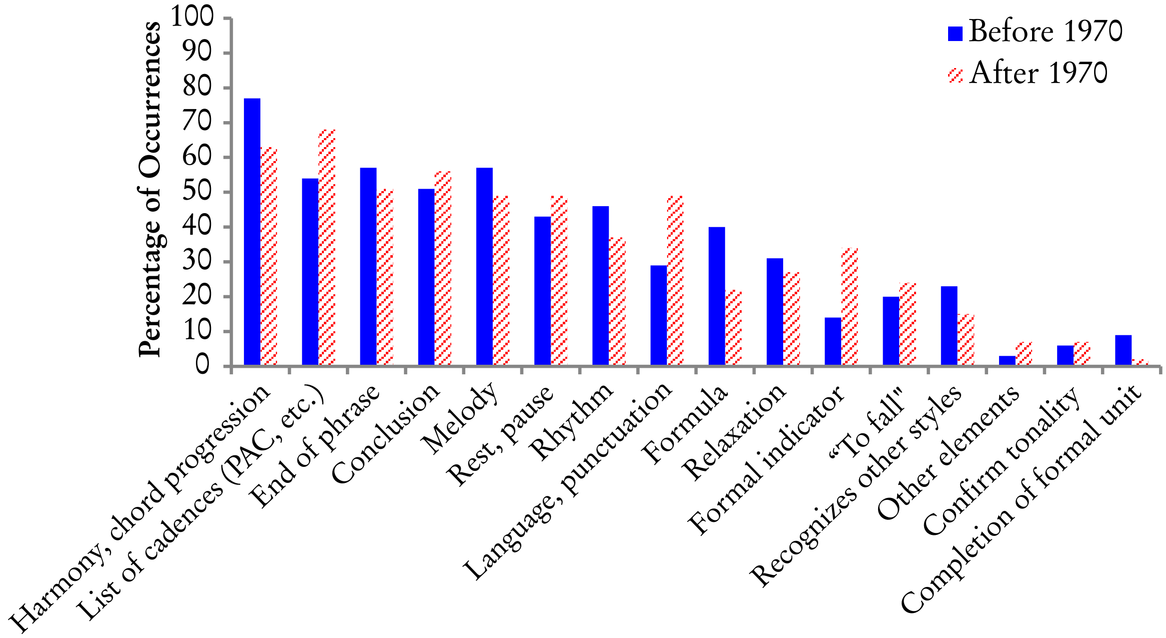


Figure 2.1: Bar plot of the percentage of occurrences for elements of cadence definitions that appear in textbooks published before (blue) and after (red diagonal) 1970. Reproduced from Table 1 of Blombach's, "Phrase and Cadence" (1987), 227.

some instances, a 'falling' melody.[10] In over half of the textbooks surveyed, these harmonic-melodic formulæ are also classified into a compendium of cadence types, with the degree of finality associated with each type sometimes leading to comparisons with punctuation in language.[11] But as this list also demonstrates, many scholars view the cadence as a "point of arrival,"[12] or *time point*, which marks the conclusion of an ongoing phrase-structural process

[10]Cadence derives from the latin *caderer*, "to fall."

[11]Walter Piston's description of the authentic cadence as "a full stop or period in punctuation," and the half cadence as "a comma, indicating a partial stop in an unfinished statement," is but one example from twentieth-century pedagogical texts (*Harmony*, 63). Analogies to punctuation have a long history in music theory. Gjerdingen's revival of Francesco Galeazzi's graded series of four cadence types is one noteworthy example. It consists of the complete cadence (period), the half cadence (colon), the resolution of an inverted dominant seventh to a root-position tonic (comma), and finally the melodic evasion of a complete cadence, which has no analogue (*Music in the Galant Style: Being an Essay on Various Schemata Characteristic of Eighteenth-Century Music* [New York: Oxford University Press, 2007], 156-157). Caplin has suggested that the analogy is ultimately misleading, however, asserting that "punctuation may be a visual sign of syntax [in written language] but is not a real source of syntax. A phrase or sentence achieves a degree of syntactical closure not by ending with any given punctuation mark, but by word meanings, inflections, and ordering" ("The Classical Cadence," 104).

[12]Ratner, *Classic Music*, 33.

(e.g., the end of a phrase), and which is often characterized as a moment of rest, quiescence, relaxation, or repose. Thus a cadence is simultaneously understood as time-span and time-point,[13] the former relating to its essential features (cadence as *formula*), the latter to the presumed boundary it precedes (cadence as *ending*).

The desire to characterize endings beyond those located in eighteenth-century compositional practice has led many to champion a more inclusive approach to the cadence concept, one that abandons the time-span view of cadence to accommodate those repertories for which harmony and tonality are not the primary agents of musical structure.[14] Definitions of cadences for atonal repertories as "temporary or permanent point[s] of repose imposed by *any* element or combination of elements" are thus fairly commonplace.[15] Blombach's definition is a case in point:

> A cadence is any musical element or combination of musical elements, including silence, that indicates relative relaxation or relative conclusion in music. ("Conclusion" is intended in the sense of "destination of ideas," as opposed to merely stopping with no indication of finality or direction.)[16]

Despite the breadth of application these definitions provide, the overwhelming majority of scholars have elected instead to restrict discussions of cadence to relatively narrow, stylistically unified repertories for which pitch centricity is a fundamental organizing principle, in some instances even limiting their investigations to the works of individual composers.[17] By

---

13 Caplin, "The Classical Cadence," 77–81.

14 Donald Johns, "On the Nature of Cadence," in *Music in Performance and Society: Essays in Honor of Roland Jackson*, ed. John Koegel and Malcolm S. Cole (Warren, MI: Harmonie Park, 1997), 396.

15 ibid. (my italics).

16 Blombach, "Phrase and Cadence," 229.

17 To provide just three examples from recent scholarship, see Jennifer Bain, "Theorizing the Cadence in the Music of Machaut," *Journal of Music Theory* 47, no. 2 (2003): 325–362; Roland Jackson, "Gesualdo's Cadences: Innovation Set Against Convention," in *Musicologia humana: Studies in Honor of Warren and Ursula Kirkendale* (Firenze, Italy: Leo S. Olschki, 1994), 275–289; Nathan John Martin and Julie Pedneault-Deslauriers, "The

narrowing the purview to particular time periods, the result has been "a more precise and focused conception" of the many ending formulæ present in a given repertory, allowing analysts to "make more subtle distinctions among a wide variety of harmonic, rhythmic, and formal phenomena."[18] Indeed, the revival of interest in theories of musical form over the last few decades—perhaps best exemplified in the recent volumes by William Caplin, James Hepokoski and Warren Darcy, and Janet Schmalfeldt—has led to considerable refinement of the cadential types associated with the instrumental music of Haydn, Mozart, and Beethoven.[19]

2.1.2 The Cadential Types: Essential Characteristics

For few other topics in music scholarship are neologisms more welcome than for the cadence types associated with classical music. To be sure, although the "high classical style" refers to a fairly limited period of music history (ca. 1770–1810), the compendium of cadential terms for the music of this period is still enormous. This fact reflects profound disagreement, appearing both in contemporaneous treatises and in recent research, as to the impact of a large number of musical parameters (e.g., melody, harmony, the presence and treatment of dissonance, inversion, texture, dynamics, tempo, timbre, and orchestration) for the identification of endings of varying strengths and for various levels of the structural hierarchy. Thus coming to a consensus as to the procedure by which closing formulæ may be identified and categorized remains a tremendous challenge.

Of the many scholars currently associated with the "New *Formenlehre*" tradition, William Caplin has presented perhaps the most comprehensive account of the classical cadence to date. Following the principles of form introduced by Arnold Schoenberg in *Fundamentals of Musical*

Mozartean Half Cadence," in *What Is a cadence? Theoretical and Analytical Perspectives on Cadences in the Classical Repertoire*, ed. Markus Neuwirth and Pieter Bergé (Leuven: Leuven University Press, 2015), 185–214.

[18] Caplin, "The Classical Cadence," 52.

[19] Caplin, *Classical Form*; Hepokoski and Darcy, *Elements of Sonata Theory*; Schmalfeldt, *In the Process of Becoming*.

Composition, and later developed by Schoenberg's student, Erwin Ratz, in *Einführung in die musikalische Formenlehre*, Caplin positions the cadence concept within a general theory of *formal functions*, which attempts to differentiate how musical segments at various levels of the structural hierarchy—ideas, phrases, themes, and so on—express their temporality as beginnings, middles, and ends.[20] In Caplin's form-functional approach, the cadence is understood as a temporal end, but it functions within his theory as both time point and time span. He thus distinguishes the moment of *cadential arrival*, which serves to "mark the structural end of a thematic region," from the time span leading up to and including cadential arrival that "communicates to the listener that 'the cadence' is forthcoming," which he terms *cadential function*.[21] And although he suggests that a number of features signal the impending cadential arrival within cadential function through "specific, harmonic, melodic, rhythmic, and textural devices,"[22] the cadential progression plays the most decisive role in his theory.[23]

His definitions of cadential arrival and cadential function thus depend entirely on features of the underlying harmonic progression. The moment of cadential arrival, for example, is identified neither at the onset of melodic resolution, nor at a melodic or harmonic event preceding a decisive segment boundary, but rather at the time point where the final *harmony* of the cadential progression first appears. Nor are the boundaries of cadential function determined by a decisive change in texture or a sudden leap in the melody, but instead by the initial and terminal harmonies of the cadential progression.[24] He writes,

[20] Caplin, Hepokoski, and Webster, *Musical Form, Forms, and Formenlehre*, 23.
[21] Caplin, *Classical Form*, 43.
[22] Caplin, "The Classical Cadence," 77.
[23] ibid., 56. Janet Schmalfeldt similarly refers to the cadential progression as the "supreme signal for thematic closure." Schmalfeldt, "Cadential Processes," 1.
[24] Mark Richards has recently questioned this view, asserting that the initiation of cadential function is signaled by melodic-textural changes in addition to a cadential progression ("Closure in Classical Themes: The Role of Melody and Texture in Cadences, Closural Function, and the Separated Cadence," *Intersections: Canadian Journal of Music* 31, no. 1 [2010]: 25). Caplin's response to this potential criticism has been to distinguish the parameters that contribute to the grouping structure from those that determine formal functionality. In Caplin's theory, the initiation of a temporal function is determined *entirely* by harmony and tonality, whereas a number of parameters,

> Even if the implied cadential arrival fails to materialize—owing to deception, evasion, or abandonment—we can still identify a passage of music whose formal function is cadential. Sometimes the cadential function is relatively compressed, as is the case especially with a simple half cadence ending a four-measure antecedent phrase. At other times, the cadential function is considerably expansive, such as in subordinate-theme areas where the confirmation of the new key requires powerful expression. But no matter what the length of the cadential function, its boundaries are essentially defined by the underlying cadential progression.[25]

Shown in Table 2.1, Caplin classifies every possible cadential category according to two fundamental types: those for which the goal of the cadential progression is tonic (the perfect authentic cadence and its variants), and those for which the goal is dominant (the half cadence and its variants). To the perfect authentic (PAC) and half (HC) cadence categories he also adds the imperfect authentic cadence (IAC), a melodic variant of the PAC. He refers to these three cadential categories as "the only genuine cadences in music in the classical style" because they are the only categories that can achieve thematic closure.[26]

Whereas imperfect authentic and half cadences remain categorically distinct from the perfect authentic cadence in Caplin's theory, the deceptive, evaded, and abandoned cadence categories generally do not, as they initially promise an authentic cadence, yet fundamentally deviate from the cadential progression, thus failing to achieve authentic cadential closure at the expected moment of cadential arrival. The deceptive cadence leaves harmonic closure somewhat open by closing with a non-tonic harmony, usually vi, but the melodic line resolves to a stable scale-degree at cadential arrival, thereby providing a provisional sense of ending for the ongoing

both syntactic and rhetorical, affect segmental grouping ("William Caplin Responds," *Intersections: Canadian Journal of Music* 31, no. 1 [2010]: 70).

[25] Caplin, "The Classical Cadence," 77.

[26] Caplin, *Classical Form*, 43.

Types	Categories	Essential Characteristics	Grouping Functions
	Perfect Authentic	V – I $\hat{1}$	↰, ↰↱
	Imperfect Authentic	V – I $\hat{3}$ or $\hat{5}$	↰, ↰↱
I	*Deceptive**	V – ?, Typically vi $\hat{1}$ or $\hat{3}$	↰, ↰↱
	*Evaded**	V – ? ?, Typically $\hat{5}$	↱
	*Abandoned**	Cadential Progression ?	↔
V	*Half*	? – V $\hat{5}$, $\hat{7}$, or $\hat{2}$	↰, ↰↱
	*Dominant Arrival**	Inverted V, 7^{th} $\hat{4}$, $\hat{5}$, $\hat{7}$, or $\hat{2}$	↰, ↰↱

Table 2.1: The cadential types and categories, along with the harmonic and melodic characteristics and the potential grouping functions for each category. Categories marked with an asterisk are failed cadences. Arrows denote the following grouping functions: ↰ *ending*, ↱ *beginning*, ↰↱ *elision*, ↔ *middle*.

thematic process. In addition to the submediant, a number of other harmonies may also appear, such as iv^6, I^6, or even a dissonant sonority like vii$^6_5$/V.[27]

The evaded cadence is characterized by a sudden interruption in the projected resolution of the melodic line; instead of resolving to $\hat{1}$, the melody leaps up, often to $\hat{5}$, thereby replacing

[27] Edward Latham has proposed that the deceptive cadence should be limited to consonant triads, however, reserving diminished and dominant-seventh chords for the evaded category because of the forward momentum generated by their dissonant character ("Drei Nebensonnen: Forte's Linear-Motivic Analysis, Korngold's Die Tote Stadt, and Schubert's Winterreise as Visions of Closure," *Gamut* 2, no. 1 [2009]: 313). For more information on the physiognomy of the deceptive cadence and the history of the concept, see Markus Neuwirth, "*Fuggir la cadenza*, or the Art of Avoiding Cadential Closure: Physiognomy and Functions of Deceptive Cadences in the Classical Repertoire," in *What Is a Cadence? Theoretical and Analytical Perspectives on Cadences in the Classical Repertoire*, ed. Markus Neuwirth and Pieter Bergé (Leuven: Leuven University Press, 2015), 117–156.

the expected ending with material that clearly initiates the subsequent process. Caplin notes, however, that cadential evasion may be characterized by a disruption in any number of parameters, such as melody, texture, dynamic, and register, so long as they counter the perception of a structural end at cadential arrival.[28] Thus, the evaded cadence projects no sense of ending whatsoever, as the events at the expected moment of cadential arrival, which should group backward by ending the preceding thematic process, instead group forward by initiating the subsequent process.

Like the deceptive and evaded cadence categories, the abandoned cadence deviates from the authentic cadential progression. But whereas in the former categories the cadential dominant remains essentially intact, in the abandoned cadence the composer "abandons" the progression before reaching the projected moment of cadential arrival, either by inverting the cadential dominant or by omitting the dominant entirely.[29] His notion of cadential abandonment—indeed, his entire theory of cadence—is thus predicated on the recognition of a root-position dominant as the *sine qua non* of the classical cadence.[30] For that reason, he also regards a half-cadential progression whose final dominant is inverted at cadential arrival as a deviation of the half cadence category, which he calls a *dominant arrival*.[31] Yet unlike the deviations of the authentic cadence, the dominant arrival may also refer to progressions whose final dominant contains a dissonant seventh, for in Caplin's view, the dominant would be too unstable to serve as a satisfactory cadential goal.[32]

---

[28] Caplin, *Classical Form*, 106.

[29] ibid., 107. Danuta Mirka refers to a similar situation in her discussion of *absent cadences* in the second movement of Haydn's Symphony No. 64 in A major ("Absent Cadences," *Eighteenth Century Music* 9, no. 2 [2012]: 213–235).

[30] "A central tenet of my concept of cadence is the requirement that dominant harmony occur exclusively in root position prior to the moment of arrival (or, in the case of a half cadence, just at the moment of arrival). So essential is this harmonic condition that if the dominant first appears inverted (say V^6_5) or becomes inverted after initially being in root position, then either no sense of cadence will be projected or else a potentially cadential situation fails to be fully realized as such" (Caplin, "The Classical Cadence," 70).

[31] Caplin, *Classical Form*, 79.

[32] Caplin additionally refers to a dominant arrival that appears before the section's "structural" end as a

In Caplin's theory, a passage therefore attains cadential status if it consists of certain *essential characteristics* relating to harmony and melody. Few scholars have questioned the important role accorded to harmonic and melodic content in establishing cadential closure, particularly for music from the classical style.[33] To be sure, Leonard Meyer referred to harmony and melody as among the *primary* or *syntactic* features of tonal music because they can be "segmented into perceptually discrete, proportionally related stimuli," thereby affording the capacity for hierarchic structuring, an attribute many scholars have claimed is essential both to perception and memory.[34] Those attributes that cannot be segmented into perceptually discrete relationships—dynamics, tempo, sonority, timbre, and so on—he referred to as *secondary* and *statistical*. Citing Barbara Herrnstein Smith's examination of the mechanisms engendering closure in poetry, Kofi Agawu also proposed a number of similar dichotomies to distinguish the harmonic or melodic aspects of a work from those he deemed ornamental, such as syntax-semantics, structure-rhetoric, and form-expression.[35]

Following Meyer and Agawu, Caplin thus differentiates among the cadential categories common in music-theoretical discourse—perfect authentic, imperfect authentic, deceptive, and so on—strictly according to the syntactic parameters specific to each category. He writes,

> In its syntactical aspect, a given cadence represents a particular cadential *type* on the basis of its harmonic-melodic content exclusively. In its rhetorical aspect, that

premature dominant arrival (*Classical Form*, 81). Poundie Burstein has recently questioned the utility of the distinction between half cadence and dominant arrival, however, suggesting that the claim that a HC must be a dominant triad in root position is a useful preference rule, but shouldn't be regarded as axiomatic. Burstein also points out that to regard inverted dominant endings as non-cadential deviations departs from theoretical traditions stretching back to the eighteenth century ("The Half Cadence and Other Such Slippery Events," *Music Theory Spectrum* 36, no. 2 [2014]: 209–213).

[33] Many scholars have questioned the subordinate status traditionally accorded to non-syntactic parameters in the perception of closure, however, in particular for music following the classical period. For an example, see Ann Hyland's analysis of the first movement of Schubert's D. 46 ("Rhetorical Closure in the First Movement of Schubert's Quartet in C Major: A Dialogue with Deformation," *Music Analysis* 28, no. 1 [2009]: 120–123).

[34] Leonard B. Meyer, *The Spheres of Music* (Chicago: The University of Chicago Press, 2000), 286. See also McAdams, "Psychological Constraints on Form-Bearing Dimensions in Music."

[35] On the rhetorical aspects of closure, see Agawu, "Concepts of Closure and Chopin's Opus 28," 3–5.

> cadence has a unique compositional realization entailing the entire range of musical parameters, including rhythm, meter, texture, intensity, and instrumentation... When characterizing cadential strength and weakness, it is important that we distinguish between the syntactical and rhetorical aspects.[36]

In the case of the authentic cadence, for example, the dominant and tonic harmonies of the cadential progression must be in root position, "their most stable form,"[37] and the tonic must support $\hat{1}$ (PAC), or $\hat{3}$ or $\hat{5}$ (IAC), in the soprano voice (see Table 2.1). If the dominant is inverted or the soprano does not present one of the above scale degrees at cadential arrival, the passage in question is, by *definition*, not an authentic cadence.

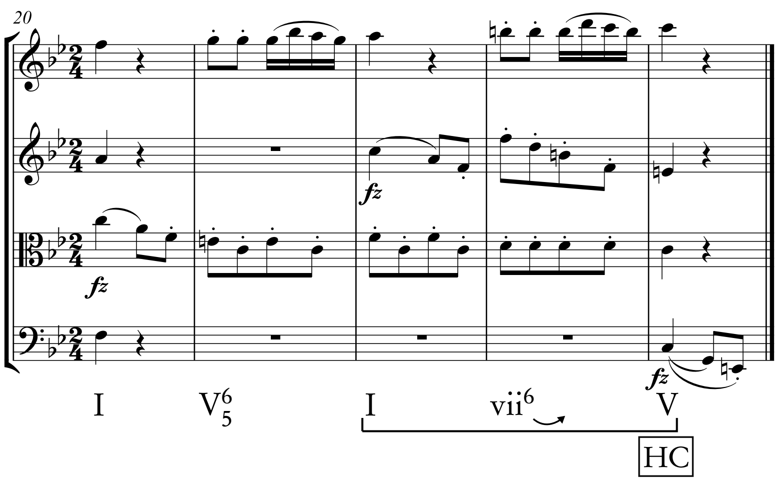
Thus, for a great many scholars in the "New *Formenlehre*" tradition, the resolution of an inverted dominant to a root-position tonic fails to meet the requirements of an *authentic* cadence. Indeed, a passage containing such a progression would not even constitute a cadence *sui generis*.[38] To justify this claim, researchers sometimes cite the apparent "strength" of the ascending fourth/descending fifth progression within the tonal system, asserting for example that the descending fifth motion of the V–I progression must be exposed in the bass "so that the sense of a strong harmonic progression can be projected most powerfully."[39] Or they argue that

[36] Caplin, "The Classical Cadence," 107.

[37] Caplin, *Classical Form*, 27.

[38] See, for example, note 30.

[39] Caplin, *Classical Form*, 27. Caplin's characterization of the authentic cadence as a "strong progression" recalls Schoenberg's discussion of the V–I progression in *Harmonielehre*. "...the strongest movement of the root is the leap of a fourth upward, because this movement seems to correspond to a tendency of the tone... The tone that was previously the principal tone, the root, becomes in the second chord a dependent tone, the fifth. More generally, the bass tone of the second chord is a higher category, a higher power, for it contains the first, the tone that itself was the root. In [C major in] the triad on G the *g* is sovereign. A progression that evokes this situation, which so to speak, sets a king over a prince, can only be a strong progression. But the *c* not only subjugates the root, it forces the other chord components as well to conform to its requirements; and the new chord contains, apart from the vanquished former root, nothing that recalls the former government. It contains, apart from that one, nothing but new tones. One can justifiably assume that progressions which produce similar situations are equally strong, or nearly so." He clarifies this point further in a footnote: "So long as a bass tone is not the root, its sole drive is to become just that. Once it is the root, then it has a different goal: to lose itself in, to become part of a higher entity" (*Harmonielehre*, trans. Roy Carter [Berkeley and Los Angeles, CA: University of California Press, 1978], 115–116).



Example 2.1: Haydn, String Quartet in B-flat, Op. 50/1, iv, mm. 20–24.

comparatively weaker progressions should be regarded as contrapuntal or prolongational, rather than as "weakened" cadential progressions,[40] because they serve only to sustain an individual harmony, rather than to confirm a tonal center as such.

Despite their apparent unsuitability in tonic cadences, these allegedly "prolongational" progressions nevertheless surface quite frequently in cadential progressions for which the goal harmony is dominant. Of the cadence categories listed in Table 2.1, the essential characteristics of the half cadence category are certainly less strict; a potential HC need only present a progression for which the penultimate and ultimate sonorities feature a change of root (e.g., I–V, I^6–ii^6–V, iv^6–V, vii^6/V–V, etc.), and with the ultimate sonority consisting of a root-position dominant triad that supports any chord member in the soprano.

According to these criteria, the identification of a half cadence in Example 2.1 would seem uncontroversial.[41] This rather abbreviated two-chord cadential progression consists of a temporary tonicization of the dominant in m. 23, one whose resolution in the following measure recalls a voice-leading formula that appears quite frequently in Renaissance contrapuntal

[40]Piston asserts, for example, that the imperfect cadence may refer to a cadential progression whose dominant is inverted. (*Harmony*, 110).

[41]Caplin identifies a half cadence at this moment in (*Classical Form*, 130).

practice. In his entry on "Cadence," published in *Grove's Dictionary* in 1908, William S. Rockstro referred to this close as the *clausula vera*, or "true close," a remnant of the old discant-tenor framework in which the tenor descends a whole step and the melody ascends a half step.[42] And yet this passage, which serves as a viable cadential progression in the context of a half cadence, would nevertheless be deemed too weak to serve as a cadential progression in any context in which the tonic serves as the goal harmony. But why should the progression in Example 2.1 confirm a tonal center in the context of a half cadence when it fails to do so for passages whose goal harmony is tonic?[43]

We could certainly expand our compendium to include tonic cadences that feature inverted-dominant resolutions, but both Caplin and Schmalfeldt have further argued that such passages are extraordinarily rare in the classical style.[44] Indeed, it is precisely because the closing formula identified in Example 2.1 appears frequently at the ends of phrases and themes that these authors have elected to ignore the prolongational status of the progression and apply the half cadence concept more flexibly. But if a prolongational progression like this one were to appear at tonic endings, Caplin writes that "such a phrase should, in principle, be regularly usable in formal locations that would bring thematic units to a close. But throughout the eighteenth century, it is rare to find such cases of formal units closed by harmonic progressions whose dominant is

[42] William S. Rockstro, "Cadence," in *Grove's Dictionary of Music and Musicians*, ed. J. A. Fuller Maitland, vol. 1 (London, 1908-1910), 435. To my knowledge, contemporary usage of Rockstro's term originated in Gjerdingen, *Music in the Galant style*, 164. Following Nathan Martin and Julie Pedneault-Deslauriers, I also classify the clausula vera within the expanding half cadence type in §3.4.3. For a discussion of the half cadence types in Mozart's keyboard sonatas, see Martin and Pedneault-Deslauriers, "The Mozartean Half Cadence."

[43] To take this point still further, would a half-cadential progression featuring ascending fourth/descending fifth motion in the bass (e.g., V^7/V–V, which Caplin terms a *reinterpreted* HC) represent a stronger ending than one featuring a stepwise ascending bassline (e.g., V^6_5/V–V)? In contemporary *Formenlehren*, the two are syntactically equivalent in a half-cadential context, but the latter progression is treated as prolongational and therefore non-cadential in the context of an authentic cadence.

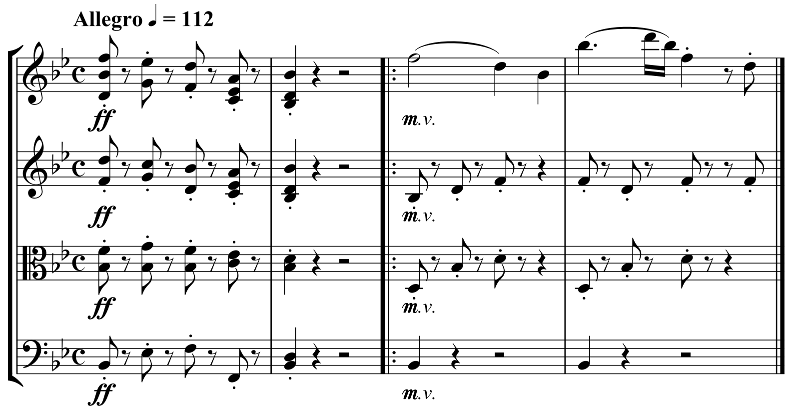
[44] William E. Caplin, "Teaching Classical Form: Strict Categories versus Flexible Analyses," *Tijdschrift voor Muziektheorie (Dutch Journal of Music Theory)* 18, no. 3 (2013): 119–135. Following Francesco Galeazzi's graded series of four cadence types (see note 11), Gjerdingen refers to root-position tonic cadences featuring inverted-dominant resolutions as *commas* (*Music in the Galant style*, 156).

inverted..."[45] Schmalfeldt agrees, acknowledging that while such passages "establish points of repose by effecting the closure of ideas and phrases, they rarely serve to articulate the large-scale harmonic closure of a complete thematic-contrapuntal process."[46]

It would seem then, that scholars justify the inclusion (or exclusion) of various cadence categories on purely empirical grounds, only drawing conclusions about the nature of certain essential characteristics for each category *after* examining a representative corpus of conventional formulæ that appear at the ends of phrases and themes. In other words, classifying the cadence categories according to their essential characteristics before identifying the temporal context in which they appear puts the cart before the horse. Lerdahl and Jackendoff make this point explicit in *A Generative Theory of Tonal Music*, for their concept of cadence depends entirely on an initial interpretation of the grouping structure. They define cadences as "signs or conventional formulas that mark and articulate the ends of groups from phrase levels to the most global levels of musical structure," and they assert that their theory "can 'find' these signs when they occur at the ends of groups, and label them as functioning cadentially for all the levels of grouping that terminate with them."[47] If these "signs" do not appear at the ends of groups, in Lerdahl and Jackendoff's view they are not cadences, since a cadence "by definition articulates the end of a group."[48]

The appeal to temporal context is thus a common refrain in the *Formenlehre* tradition, appearing in both contemporaneous treatises and modern scholarship. Ann Blombach has claimed, for example, that to attain cadential status, the mere presence of a cadential formula is not sufficient *unless* it appears at a phrase ending.[49] Instances in which cadential formulæ appear in other temporal contexts—e.g., at a theme's beginning, or following its end—are also

[45] Caplin, "The Classical Cadence," 73.
[46] Schmalfeldt, "Cadential Processes," 43.
[47] Lerdahl and Jackendoff, *A Generative Theory of Tonal Music*, 134.
[48] Ibid., 28.
[49] Blombach, "Phrase and Cadence," 232.



Example 2.2: Haydn, String Quartet in B-flat, Op. 71/1, i, mm. 1–4.

frequently dismissed as non-cadential.[50] The opening passage from the first movement of Haydn's string quartet in B-flat, Op. 71 illustrates this point. Shown in Example 2.2, Haydn opens the movement with a textbook perfect authentic cadence.[51] If these measures were to appear instead at the *end* of a thematic unit, scholars would likely have little trouble attributing cadential status to a passage that otherwise clearly exhibits *cadential content*.[52] Nevertheless, this passage would not function as a syntactical cadence for a number of scholars because, to quote Rockstro, "it has not any actual significance of the kind implied by a cadence, but only when it occurs at the end of a period or phrase of some sort."[53]

A few scholars have instead elected to privilege the time-point conception of cadential closure by classifying ending formulæ according to the grouping criteria established at cadential arrival. In "Cadential Processes: The Evaded Cadence and the 'One More Time' Technique,"

[50]In a recent article on mid-section cadences in Haydn's sonata form movements, Burstein appeals to Koch's *ruhepunkte des Geistes* to determine whether the surrounding context warrants the identification of a cadence. "In each case, one should take into account not only the rhythmic and melodic features of the phrase ending, but also the surrounding context: to paraphrase Koch, "feeling" must assist us in ascertaining whether specific resting points mark the end of complete or incomplete sections of the whole" ("Mid-section Cadences in Haydn's Sonata-Form Movements," *Studia Musicologica* 51, nos. 1-2 [2010]: 102).

[51]This compositional procedure is not limited to the first movement of Op. 71. See also the thematic introductions from the opening movements of Op. 33/5, Op. 50/6, and Op. 74/1.

[52]Caplin, "The Classical Cadence," 110.

[53]Rockstro, "Cadence," 438.

Schmalfeldt distinguishes between three types of syntactic closure: (1) *distinct closure*, in which the goal event closes a preceding process; (2) *elision*, in which the goal event both closes the preceding process and initiates the subsequent process; and (3) *evasion*, in which the goal event provides no ending whatsoever, and instead serves only to initiate the subsequent process.[54] She then situates the cadential categories, classified according to their essential characteristics—PAC, IAC, HC, and so on—within these three types.[55]

As should be evident from the above designations, Schmalfeldt essentially characterizes cadential arrival according to its temporal function, with the final harmony of the cadential progression serving either as a *beginning*, an *end*, or an elision of the two. In fact, this approach is fairly consonant with Caplin's theory of formal functions, but Schmalfeldt's designations reside much closer to the musical surface, pertaining not to the level of ideas or fragments as Caplin conceives them, but to the level of individual harmonic events.[56] I have therefore renamed Schmalfeldt's closure types to conform more readily with Caplin's formal-functional terms and included a fourth type to characterize an event at the projected moment of cadential arrival that instead appears in the middle of an ongoing process, resulting in the following four grouping functions: ↰ ending, ↱ beginning, ↰↱ elision, and ↔ middle. In Table 2.1 I have associated these grouping functions with each cadence category in Caplin's typology.

But which characteristics articulate a grouping function? Again, Caplin distinguishes between the various temporal functions—including the cadence categories that make up cadential function—strictly according to essential characteristics relating to harmony and melody. This approach certainly facilitates the identification of conventional formulæ, whose memorability

[54] In their theory, Lerdahl and Jackendoff apply these terms somewhat differently. They refer to elision as *overlap*, in which an event or sequence of events is shared by two adjoining groups; and they call evasion *elision*, in which the moment of ending is replaced with a beginning (*A Generative Theory of Tonal Music*, 55–62).

[55] She does not discuss the abandoned cadence or the dominant arrival in her theory, however.

[56] In Caplin's theory, the *idea unit* represents the smallest formal function, and it is typically comprised of two real measures. Caplin also points out that in the simplest cases, a single harmony supports the idea unit, but he concedes that an idea can be harmonized by several distinct chords (*Classical Form*, 41).

rests upon the recurrence of harmonic and melodic characteristics, but it fails to adequately represent all of the parameters that articulate segment boundaries at various levels of the structural hierarchy. To be sure, Caplin restricts the cadence concept to the articulation of "formal" boundaries at middleground levels of musical organization (i.e., at the levels of the phrase and theme), leaving aside "other modes of closure" that might be used to bring motives and other short ideas to an end.[57] Thus, although they often operate together, a grouping boundary (or rhythmic stop, as Caplin puts it) is "conceptually (and perceptually) distinct" from a formal end.[58]

Hepokoski and Darcy make a similar claim with regard to their concept of *essential expositional closure* (EEC), the moment when the subordinate theme area of a sonata-form movement attains the first satisfactory PAC in the new key that proceeds onward to differing material. In *Sonata Theory*, the EEC is the most important generic and tonal goal of the exposition, and they emphasize that it should be "conceptually privileged" because it closes the thematic materials in the subordinate theme and marks the securing of the secondary key.[59] Nevertheless, they note that despite its structural significance, the EEC should not be *perceptually* privileged for the simple reason that rhetorically stronger cadences very often occur in the closing section as reinforcement work. They write,

> One should not determine an EEC on the basis of what one imagines an EEC should 'feel' like in terms of force or unassailably conclusive implication. [...] The first PAC closing the essential exposition is primarily an attainment of an important generic requirement—nothing more and nothing less.[60]

And yet, as I will argue in §2.3, the grouping structure often plays a significant role in

[57] Caplin, "The Classical Cadence," 60.
[58] Caplin, *Classical Form*, 51.
[59] Hepokoski and Darcy, *Elements of Sonata Theory*, 123.
[60] Ibid., 124.

distinguishing the cadence categories listed in Table 2.1. Poundie Burstein asserts, for example, that "a half-cadential ending does not exist independently of features of texture, rhythm, and form that combine to demarcate it," and that "for a dominant harmony to be a convincing and syntactically proper close, various features must coordinate with one another."[61] Thus in his view, harmonic progressions should not receive so privileged a position in the identification of half cadences, since "the harmonies considered abstractly cannot assert a half cadence."[62] William Rothstein would seem to agree, suggesting that "elements of closure, beyond the harmonic cadence, may involve the register of the bass or melody, the presence or absence of some important melodic tone (either in the passage leading to the cadence or in the cadence itself), the presence or absence of a subdominant-type harmony in the cadential progression, or any of a number of other factors."[63]

Those scholars intent on establishing a psychological approach to the perception of closure have elected to cast a wider net, examining how syntactic and rhetorical parameters may interact to effect segmental grouping at multiple levels of the structural hierarchy. Schoenberg referred to this process as *the psychology of the close*;[64] Meyer called it *parametric congruence*. I quote the latter.

> Closure—the arrival at relative stability—is a result of the action and interaction among the several parameters of music. Because melody, rhythm, harmony, texture, timbre, and dynamics are relatively independent variables, some may act to create closure at a particular point in a work, while others are mobile and on-going. To the extent that the parameters can act together in the articulation of closure or, alternatively, in creating instability and mobility, they may be said to move

[61] Burstein, "The Half Cadence," 206.
[62] Ibid.
[63] William Rothstein, *Phrase Rhythm in Tonal Music* (New York: Schirmer Books, 1989), 116.
[64] Arnold Schoenberg, *The Musical Idea and the Logic, Technique, and Art of Its Presentation*, trans. Patricia Carpenter and Severine Neff (New York: Columbia University Press, 1995), 172–173.

> congruently. Conversely, when some parameters foster closure while others remain open, the parameters are said to be *noncongruent*.[65]

According to this view, parametric noncongruence may obtain for every category in Caplin's typology. In the case of the deceptive cadence, "rhythm and melody act to articulate closure, but harmony remains open and mobile."[66] For the half cadence, a potentially active, unstable dominant achieves the status of a cadential goal by means of metrical, textural, and rhythmic reinforcement.[67] Even for the authentic cadences, for which the principles that determine syntactic closure are said to be unambiguous, features related to rhythm, texture, and dynamics could nevertheless weaken or obscure the segment boundary; indeed, there are numerous instances in the classical style in which they do.

But if we seek to identify those ending formulæ that listeners are likely to learn and remember, we would do well to distinguish an investigation of the essential characteristics that might designate mental representations for the cadence categories listed in Table 2.1 from an examination of the numerous parameters—both at and following cadential arrival—that determine the strength of the segment boundary after the fact. In §2.2, I privilege the time-span view of the classical cadence as a conventional formula, one whose frequency of occurrence and temporal context allow listeners to generate expectations for the impending cadential arrival, setting aside the many issues surrounding segmental grouping and cadential strength until §2.3.

[65] Meyer, *Explaining Music*, 81.
[66] Ibid.
[67] Ibid., 85.

§2.2 Cadences as Mental Representations

2.2.1 The Psychologizing Impulse: Expecting Cadences

While the systematic study of listening behaviour normally remains outside the sandbox of music theory, scholars nonetheless remain highly sensitive to the potential effects of ending formulæ on listeners. Descriptions of cadential arrival as a moment of rest, finality, or repose—terms that abound in the history of theory—imply that closure is inherently *felt* during the act of listening. Heinrich Christoph Koch, a theorist and contemporary of Haydn and Mozart known today for his contributions to the theory of musical form, referred to such moments as "resting points of the spirit," noting that "only feeling can determine both the places where resting points occur in the melody and also the nature of these resting points."[68]

To account for the perception of closure in Western tonal music, many scholars have appealed to theories of expectation. Jonathan Dunsby has noted, for example, that in Schoenberg's view, the experience of closure for a given cadential formula is only satisfying to the extent that it fulfils a stylistic expectation.[69]

> [To close a piece of music] I believe we simply fulfil the requirements recognized by the conventions and the sense of form of the era, requirements which, in defining the possibilities, evoke expectation and thereby guarantee a satisfying close.[70]

Yet despite incipient references to tension, surprise, and expectancy in theories of cadence and closure, a psychological approach to musical expectations remained out of reach until the

[68]Heinrich Christoph Koch, *Introductory Essay on Musical Composition*, trans. Nancy Kovaleff Baker (New Haven, CT: Yale University press, 1983), 1–3. By referring to the "spirit," Koch thus implies that closure is not an external property of the sounding stimulus (e.g., a note or chord within a particular tonal context, a decisive textural or rhythmic break, etc.), but instead is determined internally by the listener.

[69]Dunsby, "Schoenberg on Cadence," 125.

[70]Schoenberg, *Harmonielehre*, 127.

mid-twentieth century with the publication of Leonard Meyer's dissertation and subsequent first book, *Emotion and Meaning in Music*.[71] For the theorists associated with what we now call the "Penn School,"[72] Meyer's dissertation provided the impetus for a research agenda predicated upon the study and analysis of the listening experience.[73] And of the many features associated with Western tonal music, Meyer considered the classical cadence the quintessential compositional device for eliciting specific expectations. Following the first musical example from *Emotion and Meaning in Music* (see Example 2.3), Meyer writes,

> In Western music of the eighteenth century, for example, we expect a specific chord, namely, the tonic (C major), to follow this sequence of harmonies [. . .]. Furthermore, the consequent chord is expected to arrive at a particular time, i.e., on the first beat of the next measure. Of course, the consequent which is actually forthcoming, though it must be possible within the style, need not be the one which was specifically expected. Nor is it necessary that the consequent arrive at the expected time. It may arrive too soon or it may be delayed. But no matter which of these forms the consequent actually takes, the crucial point to be noted is that the ultimate and particular effect of the total pattern is clearly conditioned by the specificity of the original expectation.[74]

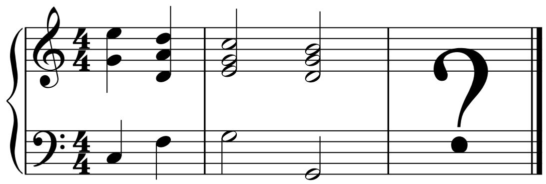
Hence, a cadence, or more precisely, the material preceding cadential arrival, elicits very definite

[71] Leonard B. Meyer, "Emotion and Meaning in Music" (PhD Dissertation, The University of Chicago, 1954); Meyer, *Emotion and Meaning in Music*.

[72] "... after relocating to the University of Pennsylvania in the 1970s, [Leonard Meyer] founded the so-called "Penn" school of music theory, with its strong focus on listeners" (Robert O. Gjerdingen, "The Psychology of Music," in *The Cambridge History of Western Music Theory*, ed. Thomas Christensen [Cambridge: Cambridge University Press, 2002], 976). Among Meyer's students, Gjerdingen counts Eugene Narmour, Justin London, and himself.

[73] According to Anna Tirovolas and Daniel Levitin, *Emotion and Meaning in Music* is the second most cited book in the journal *Music Perception* ("Analysis of Empirical Articles in Music Perception," *Music Perception* 29, no. 1 [2011]: 32).

[74] Meyer, *Emotion and Meaning in Music*, 25–26.



Example 2.3: A cadential progression, reproduced from Example 1 of Meyer's *Emotion and Meaning in Music* (1956), 25.

expectations concerning the melodic scale-degree, the harmony, and the metric position of the goal event. The moment of cadential arrival, on the contrary, elicits no further expectations with respect to these parameters.[75] This absence of expectancy following cadential arrival led Eugene Narmour to describe cadential arrival as a nonimplicative context,[76] or, to use Elizabeth Margulis's expression, as an event that suppresses expectancy.[77] These authors might therefore suggest that terms like rest, finality, and repose result from a desire to characterize the cessation of expectancy following cadential arrival. David Huron summarizes this point nicely: "When there cease to be expectations about what may happen next, it makes sense for brains to experience a sense of the loss of forward continuation—a loss of momentum, of will, determination or goal. In short, it makes sense for brains to experience a sense of repose or quiescence whenever the implications cease."[78]

[75]Depending on the formal or generic context, however, this point does not preclude the possibility that we may generate expectations for the initiation of subsequently new processes following a cadential goal, such as the prolongation of dominant harmony following a half cadence or the onset of codettas following a perfect authentic cadence. The possible types of expectation experienced during music listening are wide-ranging, but in this instance I employ the term 'expectation' rather narrowly to refer to schematic expectations for the musical parameters that characterize a cadential goal. For a discussion of the diverse applications of the term in music discourse, see Elizabeth Hellmuth Margulis, "Surprise and Listening Ahead: Analytic Engagements with Musical Tendencies," *Music Theory Spectrum* 29, no. 2 (2007): 197–217.

[76]Eugene Narmour, *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model* (Chicago, IL: University of Chicago Press, 1990), 102.

[77]Elizabeth Hellmuth Margulis, "Melodic Expectation: A Discussion and Model" (PhD Dissertation, Columbia University, 2003), 263. If I may echo Margulis's sentiment, Narmour's concept of closure is complex, but his claim is essentially that non-closural events, such as the leading tone, elicit intense and very specific implications, while closural events, such as the moment of cadential arrival, suppress further implications (*The analysis and cognition of basic melodic structures*, 102).

[78]Huron, *Sweet Anticipation*, 157.

But in the context of Example 2.2, to suggest that the harmonic progression in m. 1 elicits specific expectations for tonic harmony does not presume that listeners are familiar with the work itself. On the contrary, Jamshed Bharucha and others make a clear distinction between *schematic* and *veridical* expectancies, where schematic expectations represent long-term stylistic knowledge resulting from extensive exposure with a corpus of music, whereas veridical expectations refer to explicit prior knowledge about how a particular work goes.[79] What is more, Margulis has suggested that the schematic category itself may represent more than one variety of expectation.

> Specifically, schematic expectations inhabit a continuum from relatively deep to relatively shallow, where depth relates to availability for direct access (from little to much availability), susceptibility to change through exposure (from little to much susceptibility), and scope of application (from more universal to more limited). Examples of increasingly shallow schematic expectations might be: expectations for closure; expectations for cadential closure in tonal music; expectations for common cadence types in music from the classical period; and expectations for common cadence figures in the music of Mozart, where these expectations are increasingly available for access, increasingly susceptible to change through exposure to new pieces within the relevant repertoire, and increasingly limited in scope.[80]

Depending on the criteria for cadential identification, schematic expectations for cadential closure would necessarily inhabit much of Margulis' continuum, with the specificity of the definition corresponding to the depth of the schematic expectation. Take the perfect authentic cadence as one example. Conceived generally as a V–I progression, the PAC persisted throughout

[79] Jamshed J. Bharucha, "Music Cognition and Perceptual Facilitation: A Connectionist Framework," *Music Perception* 5, no. 1 (1987): 4. For further discussion of the expectation types (e.g., conscious, dynamic, etc.), see Huron, *Sweet Anticipation*, 219–238.

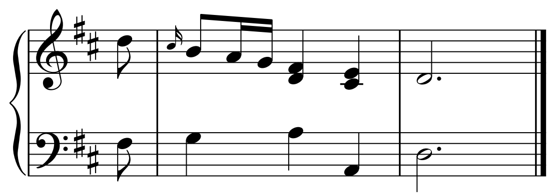
[80] Elizabeth Hellmuth Margulis, "A Model of Melodic Expectation," *Music Perception* 22, no. 4 (2005): 666.

much of the history of Western music, and thus, expectations for the resolution of such a progression would lie on the deep side of the spectrum, where they would be relatively unconscious, susceptible to very little change through exposure to other musical styles, and fairly broad in their scope of application. But in a short article published just after the second world war, Charles Cudworth observed a particular subspecies of the PAC that he termed the *Cadence Galante*,[81] but which we now call (after Gjerdingen's nomenclature) the *Cudworth Cadence*,[82] in which a complete authentic cadential progression supports an octave stepwise descent in the melody (see Example 2.4). Given the greater specificity of its essential characteristics, it is unsurprising that the Cudworth cadence was confined to a relatively narrow period of roughly eighty years at the beginning of the eighteenth century. And thus despite its frequent appearance in the instrumental works of composers like Pergolesi, Scarlatti, and Haydn, expectations for this Galant mannerism would necessarily lie on the shallow side of the schematic spectrum. Gjerdingen articulates this point clearly in *Music in the Galant Style*:

> Because it is based on the statistics of music heard, and on human learning, memory, and other cognitive abilities, schema theory does not insist on either a canonical set of schemata (certain patterns and no others), or a canonical set of relationships. It is descriptive rather than prescriptive. It accepts that connoisseurs may be able to recognize a large number of highly differentiated schemata with quite specific implications, while neophytes may apply a coarser, all-purpose set with very general implications.[83]

But how do such expectations form? As this quotation demonstrates, the most frequent answer given by theorists and psychologists infers a causal relationship between a statistically

[81] Charles L. Cudworth, "Cadence galante: The Story of a Cliché," *The Monthly Musical Record* 79 (1949): 176–178.
[82] Gjerdingen, *Music in the Galant style*, 146.
[83] Ibid., 377–378.



Example 2.4: The Cudworth Cadence, reproduced from Example 1 of Cudworth's "Cadence Galante" (1949), 176.

probable event and an expected event. Because cadences appear frequently and their underlying harmonic and melodic characteristics remain fairly consistent, listeners learn over repeated exposure to expect these endings. A theory of expectation is therefore appealing to the study of cadence because it provides a direct, causal link between events on the musical surface and the schematic knowledge of the listening subject.[84]

2.2.2 Schema Theory: Remembering Cadences

For the theorists of the Penn School, understanding music is not a matter of dictionary definitions, of knowing the various rules of musical syntax and grammar explicitly, but is instead a matter of acquiring the appropriate knowledge of a musical style by virtue of exposure over the course of one's life.[85] Thus, theoretical terms associated with classical music—perfect authentic cadences, augmented 6th chords, applied dominants, and the like—are perhaps better understood, to use Vasili Byros' expression, as declarative objectifications of procedural or tacit knowledge.[86] Putting it simply, in Meyer's view, knowledge of a style, and of the replicated patterns that characterize that style, is usually *tacit*.[87]

In fact, Meyer's insistence on tacit or procedural knowledge to explain complex musical

[84]Byros makes precisely this point in connection with the *Le-Sol-Fi-Sol* schema ("Meyer's Anvil: Revisiting the Schema Concept," *Music Analysis* 31, no. 3 [2012]: 280–281). I attempt to provide experimental evidence of schematic expectations for cadential closure in Chapters 6 and 8.

[85]Meyer, *Emotion and Meaning in Music*, 60–62.

[86]Byros, "Meyer's Anvil," 282.

[87]Leonard B. Meyer, *Style and Music: Theory, History, and Ideology* (Philadelphia, PA: University of Philadelphia Press, 1989), 10.

behaviors is consistent with theories of learning advanced in cognitive psychology over the last several decades. As I discussed in Chapter 1, psychologist Arthur Reber proposed a domain-general learning mechanism called *implicit learning* to explain how humans learn their native language(s) during early development.[88] In his view, listeners with exposure to complex syntactic structures like natural language or tonal music acquire mental representations for the most recurrent patterns over time because the characteristics that apparently define those patterns co-occur with remarkable frequency. As a result, Reber assumes that listeners form mental representations that are at least partially isomorphic with any encountered instance. He calls this the *abstractive view* of knowledge acquisition, but in studies of category formation, patterns learned in this way are often called *prototypes*.

So the formation of schematic expectations for the end of a passage, and thus, the perception of closure, depends in part on our ability to abstract the correlational structure of classical music and internalize the most common harmonic-melodic prototypes associated with cadential closure. It is worth noting, moreover, that it is precisely this ability that explains the remarkable diversity of such patterns in both Western and non-Western musics more generally. Although sensory principles have some part to play in determining the psychological stability of an individual harmonic-melodic event, the replication of quite distinct closing *patterns* in Joplin's rags, Du Fay's chansons, or indeed, Haydn's string quartets, testifies to their role as learned conventions within a particular historical and cultural context. Meyer makes precisely this point in connection with the primary (or syntactic) parameters that articulate closure.

> In any realm of human activity, the degree of "indifference" or interchangeability of convention is variable. In music, for instance, closure can be a result of the syntactic stipulation of the primary parameters (melody, harmony, rhythm) and/or the gradual abatement of the secondary parameters (descending pitches, lower

[88] Reber, "Implicit Learning of Artificial Grammars"; Reber, "Implicit Learning of Artificial Grammars."

> dynamics, slowing activity, less complex or dense textures). But the indifference levels of these closural means are not the same. The indifference level of the secondary parameters that create what Leonard Ratner has called the "dynamic curve" is very low... The closure created by a hierarchy of pitch relationships, however, has a considerable degree of indifference. Tonal systems, in other words, are unquestionably conventions. Witness the variability of closural progressions in Western art music from Machaut to, say, Prokofiev—not to mention the profusion of closing patterns in the music of other cultures.[89]

All of this is to say that these "learned conventions," or *schemas*, form the basis of our tacit knowledge for music of the classical style. But what constitutes a schema? Most psychologists note the first modern usage of the term in the work of British psychologist Frederic Bartlett.[90] Rather than representing memory as a storehouse containing traces of every past object or event encountered in everyday life,[91] Bartlett suggested that the human mind maintains a few organized settings of past reactions or experiences that serve as templates or models, against which present experiences could be compared. According to Bartlett, these schemata are "active, developing patterns" whose units are serially organized, not simply as individual members coming one after the other, but as a unitary mass.[92]

Bartlett's largely abstractive view of mental representations informed later research on pattern recognition, and was also influential in Eleanor Rosch's work on category formation,

---

[89]Leonard B. Meyer, "Nature, Nurture, and Convention: The Cadential Six-Four Progression," in *The Spheres of Music* (Chicago: The University of Chicago Press, 2000), 228–229.

[90]Frederic C. Bartlett, *Remembering: A Study in Experimental and Social Psychology* (London, United Kingdom: Cambridge University Press, 1932). David Rumelhart also acknowledges Kant's use of the term in *Critique of Pure Reason* ("The Representation of Knowledge in Memory," in *Schooling and the Acquisition of Knowledge*, ed. Richard C. Anderson, Rand J. Spiro, and William E. Mongague [Hillsdale, NJ: Erlbaum, 1977], 100–101).

[91]Reber calls this the "distributive (exemplar)" view, in which a stimulus is coded and stored, not on the basis of patterns and regularities among features, but as a kind of "raw" instance (*Implicit Learning and Tacit Knowledge*, 121–122).

[92]Bartlett, *Remembering*, 201.

but a genuine schema theory as it was envisioned by Bartlett, one which represented schemata as active, goal-directed processes that constantly inform and revise perception, would not be proposed again until the mid 1970s with the work of Marvin Minsky, Roger Schank and Robert Abelson, and David Rumelhart.[93] By that point in time, studies in pattern recognition had already indicated that participants could recognize prototypical shapes (e.g., square, circle) despite pronounced distortions, leading researchers to suggest that prototypes result from central tendencies among a cluster of features.[94] Rosch later extended these findings to the study of natural categories for domains like color, line orientation, and shape, demonstrating over a series of experiments that participants tend to define a category as a set of variations around its most clear cases.[95]

According to Rosch, classification systems like the one presented in Table 2.1 are not arbitrary products of historical accident, but rather the result of psychological principles of categorization.[96] In her view, category systems established in a particular historical or cultural context provide maximum information with least cognitive effort while mapping the perceived stimulus structure as closely as possible. Contrary to the *Aristotelian* or *classical* view often embraced within the *Formenlehre* tradition, in which categories consist of a set of "criterial" or essential features, Rosch further claimed that categories tend to be defined in terms of prototypes

[93] Marvin Minsky, "A Framework for Representing Knowledge," in *The Psychology of Computer Vision*, ed. P. H. Winston (New York: McGraw-Hill, 1975), 211–277; Roger C. Schank and Robert P. Abelson, *Scripts, Plans, Goals and Understanding: An Enquiry into Human Knowledge Structures* (Hillsdale, NJ: Erlbaum, 1977); David E. Rumelhart, "Notes on a Schema for Stories," in *Representation and Understanding: Studies in Cognitive Science*, ed. Daniel G. Bobrow and Allan M. Collins (New York: Academic Press, 1975), 211–236. Rumelhart retained the term *schema* in his theory, while Minsky and Schank and Abelson called their concepts *frame* and *script*, respectively, but Alan Baddeley suggests all three studies could be broadly regarded as examples of schema theories (*Human Memory: Theory and Practice* [Hillsdale, NJ: Erlbaum, 1990], 336).

[94] Fred Attneave, "Transfer of Experience with a Class-Schema to Identification-Learning of Patterns and Shapes," *Journal of Experimental Psychology* 54, no. 2 (1957): 81.

[95] Eleanor R. Heider, "Universals in Color Naming and Memory," *Journal of Experimental Psychology* 93, no. 1 (1972): 10–20; Eleanor H. Rosch, "Natural Categories," *Cognitive Psychology* 4 (1973): 328–350; Eleanor Rosch, "Cognitive Reference Points," *Cognitive Psychology* 7 (1975): 532–547.

[96] Eleanor Rosch, "Principles of Categorization," in *Cognition and Categorization*, ed. E. Rosch and B. B. Lloyd (Hillsdale, NJ: Erlbaum, 1978), 251.

that contain the attributes most representative of items inside and least representative of items outside the category.[97] A category is therefore best understood as a network of overlapping attributes, where its members are prototypical to the extent that they bear a *family resemblance* to—have attributes in common with—other members of the category.[98]

Gjerdingen has crystallized this point of view by visualizing category membership on a target, with the prototype at the bull's-eye. One could assess the degree of inclusiveness of a potential category member by comparing it to the cluster of attributes characterizing the category itself: the more attributes it shares, the more prototypical the member, and thus, the closer to the bull's-eye.[99] For a category whose attributes consist of letters from the English alphabet, for example, items of the form AD, BCD, CE, and ACD would bear a family resemblance, since each item has at least one attribute in common with one or more items. But note here that since none of the four items share any one attribute, the *Aristotelian* or *classical* model would fail to classify these items into the same category.

From a probabilistic point of view, a *classical* category is comprised of at least one attribute that appears in every member, so the probability P of that attribute's occurrence in any new member is always 1. For category members determined by family resemblance, however, the representativeness (or *cue validity*) of even its most representative attribute is potentially less than 1.[100] For the examples above, C and D would receive the highest cue validities because they appear in three of the four members ($P = .75$), followed by A, which appears twice ($P = .5$), and finally B and E, which appear just once ($P = .25$).

By basing category membership on the notion of family resemblance, Rosch's account exemplifies the probabilistic approach to category formation adopted by cognitive psychologists

[97] Ibid., 253–254.

[98] Eleanor Rosch and Carolyn B. Mervis, "Family Resemblances: Studies in the Internal Structure of Categories," *Cognitive Psychology* 7 (1975): 575.

[99] Gjerdingen, *A Classic Turn of Phrase*, 94.

[100] Rosch and Mervis, "Family Resemblances."

over the last half century, which Michael Posner has subsequently nicknamed the "Roschian revolution."[101] Rosch has been quick to point out, however, that to speak of a prototype for natural semantic categories is simply a convenient grammatical fiction, since prototypes for natural categories are rarely found in the real world.[102] According to her view, category membership should not be determined using "criterial features" because the prototype itself does not exist.[103] Instead, any particular instance can be judged along a continuum of prototypicality using the cluster of co-occurring attributes that are most representative (or prototypical) of the items inside and least representative (or prototypical) of the items outside the category. To be sure, an essential difference between the formation of categories in cognitive psychology on the one hand, and the identification of cadences in the *Formenlehre* tradition on the other, is precisely that no single attribute defines a given category. There are no *essential characteristics* for chairs, refrigerators, or stationwagons in Rosch's view because such categories almost invariably include members for which there is no common attribute.

Nevertheless, to abstract the prototype for a given cadence category would not be equivalent to abstracting the prototype for a shoe or a dalmation. The use of the term 'category' is, in this instance, something of a misnomer. According to Jean Mandler, categorical organization refers to the cognitive structures, hierarchically arranged, that govern our understanding of the relationships among superordinate (car, Porsche), subordinate (Porsche, car), and coordinate classes (Porsche, Lamborghini).[104] But whereas categorical organization concerns the level of inclusiveness of the category, schematic organization refers to the internal spatial or temporal layout of the individual structure and its manner of activation during perception. Unlike

[101] Michael I. Posner, "Empirical Studies of Prototypes," in *Noun Classes and Categorization*, ed. Colette Craig (Amsterdam: John Benjamin, 1986), 54.

[102] Rosch, "Principles of Categorization," 263.

[103] Rosch and Mervis, "Family Resemblances," 574.

[104] Jean M. Mandler, "Categorical and Schematic Organization in Memory," in *Memory Organization and Structure*, ed. C. Richard Puff (New York, NY: Academic Press, 1979), 262.

categories, the structure of a schema is not based on class membership, but on the connection of its individual attributes on the basis of contiguities that have been experienced in space or time. Modern schema theorists refer to these two schema types as *scene* and *event*, respectively.[105]

Conceived as a temporally organized representation of common sequences of events, an event schema is thus an inherently dynamic concept that bridges the perceptual and the cognitive. According to Rumelhart, it should not be viewed as a static definition consisting of a set of more or less probable features, but rather "as a *procedure* whose function is to determine whether, and to what degree, it accounts for the pattern of observations."[106] A schema is thus a kind of mental script that when activated, affords listeners the capacity to expect potential continuations and seek out new information from the incoming environment.

It is probably more fitting, then, to imagine an event schema for a cadence as roughly equivalent to a schema for eating out at a restaurant.[107] It consists of a set of temporally organized events that can be filled in any given instance by values that have greater or lesser degrees of probability of occurrence attached to them.[108] In the restaurant schema, the events may consist of arriving at the restaurant, being seated at a table, ordering, eating, and paying the bill. In the cadence schema, the events may consist of a sequence of harmonies, or a particular bass-soprano counterpoint, or perhaps some combination of the two. Indeed, as part of its specification, a schema contains the network of interrelationships among its constituent events.[109] Thus, determining whether a given pattern will activate a particular schema is an inherently probabilistic question, where the typical values of each event within the schema

---

[105]Ibid., 260.

[106]David E. Rumelhart, "Schemata: The Building Blocks of Cognition," in *Theoretical Issues in Reading Comprehension*, ed. Rand J. Spiro, Bertram C. Bruce, and William F. Brewer (Hillsdale, NJ: Erlbaum, 1980), 39 (my italics).

[107]The restaurant script is a well-known example from Schank and Abelson of a 'script-like' schema (*Scripts, Plans, Goals and Understanding*, 42–46).

[108]Mandler, "Categorical and Schematic Organization," 263.

[109]Rumelhart and Ortony, "The Representation of Knowledge in Memory," 101.

will serve as *default values* that constrain our expectations for possible continuations. And as Rumelhart notes, these constraints are not all-or-none, requiring that certain events within the schema have a fixed range of values; they are merely specifications of the *normal* range of values for each event in the schema.[110]

In sum, a 'schema' theory is not just a theory about the representation of knowledge in memory; it is also a theory about how that knowledge is used. To be sure, the allure of schema theory as it was intended by Bartlett is that at its foundations it places the temporal dimension of perceptual experience front and center. As a consequence, cognitive psychologists often appeal to schema theory to explain temporal experiences like reading, storytelling, and music listening. In these contexts, schemata abstracted from previous experience engender expectations about events that are likely to occur *during* perception, thereby orienting attentional processes to important aspects of the stimulus. As psychologist Jamshed Bharucha explains, "events are thus expected, implied, erroneously judged to have occurred and rendered more consonant, to the extent that their mental representations have been activated in anticipation of their occurrence."[111]

Following the emergence of schema theory in the mid 1970s, musical applications in experimental psychology largely centered around mental representations of tonal materials, either in reference to scales,[112] or to the stability relations characterizing tonality more generally—what Carol Krumhansl termed the "tonal hierarchy."[113] Modifiers like 'scale' and 'tonal' for musical schemata were thus somewhat commonplace in the 1980s,[114] with cadential formulæ receiving

[110] Rumelhart, "Schemata," 36.

[111] Bharucha, "Music Cognition and Perceptual Facilitation," 3.

[112] "Tonal scales constitute one of the most durable families of perceptual-motor schemata that have been observed in psychology" (W. Jay Dowling, "Scale and Contour: Two Components of a Theory of Memory for Melodies," *Psychological Review* 85, no. 4 [1978]: 345).

[113] Carol L. Krumhansl and Edward Kessler, "Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys," *Psychological Review* 89, no. 4 (1982): 334–368.

[114] For an example of "scale schemas," see Dowling and Harwood, *Music Cognition*, 124–152. For an example of "tonal schemas," see Jamshed J. Bharucha, "Anchoring Effects in Music: The Resolution of Dissonance," *Cognitive*

much less attention. It would take another decade—following the schema theories of Leonard Meyer, Eugene Narmour, and especially Robert Gjerdingen—for experimental psychologists to suggest that listeners with sufficient exposure to tonal music might possess long-term schematic representations for cadential formulæ.[115]

In the hands of the Penn School, examples of the schema concept and its various antecedents obtained greater specificity, applying less to tonal organization as a whole, and more to the highly replicated patterns characterizing a given style, either at the phrase level (e.g., cadences),[116] or at more global levels of musical organization (e.g., sonata form).[117] To be sure, with the publication of his second monograph, *Music, the Arts, and Ideas*, Meyer's view that listeners abstract recurrent temporal patterns as complex mental categories anticipated modern schema theory by nearly a decade:

> We perceive, comprehend, and remember our experiences—musical or other—in terms of more or less learned schematic types. Particular experiences and objects are comprehended and remembered as deriving from, and deviating from, schemata which serve as methods for "encoding" and remembering large amounts of information easily and efficiently."[118]

Meyer's "changing-note archetype" is perhaps the most well-known example of a phrase-level schema from the writings of the Penn School, in which the harmonic progression I–V–V–I supports the melodic pattern $\hat{1}$–$\hat{2}$–$\hat{7}$–$\hat{1}$;[119] Gjerdingen would devote the entirety of his first

Psychology 16, no. 4 (1984): 485–518; Jamshed J. Bharucha, "Event Hierarchies, Tonal Hierarchies, and Assimilation: A Reply to Deutsch and Dowling," *Journal of Experimental Psychology: General* 113, no. 3 (1984): 421–425.

[115] For example, see Burton Rosner and Eugene Narmour, "Harmonic Closure: Music Theory and Perception," *Music Perception* 9, no. 4 (1992): 383–412.

[116] Leonard B. Meyer, *Music, the Arts, and Ideas* (Chicago: The University of Chicago Press, 1967), 287–289; Meyer, *Explaining Music*, 213; Meyer, *The Spheres of Music*, 231; Gjerdingen, *A Classic Turn of Phrase*, 34.

[117] Gjerdingen, *A Classic Turn of Phrase*, 100–104; David Temperley, Review of *Music in the Galant Style*, by Robert O. Gjerdingen, *Journal of Music Theory* 50, no. 2 (2006): 286.

[118] Meyer, *Music, the arts, and ideas*, 287.

[119] Meyer, *Explaining Music*, 191.

monograph to this "classic turn of phrase,"[120] in his later writings naming it the "Meyer schema" in honor of his mentor.[121] Despite some disagreement as to the role played by presumably innate physiological and psychological constants in the formation of categories, both Meyer and Narmour essentially characterized recurrent temporal patterns as stylistic norms abstracted from experience.[122] Thus, in some senses, terms like *archetype*, *style form*, and *style structure* represent incipient schema theories in their own right.

Following schema theory's appearance in cognitive psychology, 'schema' gradually succeeded the lexicon of associated terms in the music theory community,[123] with Gjerdingen's various articles and monographs playing an especially important role. Unlike his predecessors, Gjerdingen adopted a corpus-analytic framework to examine the musical schemata characterizing the instrumental repertories of the eighteenth century, thereby bringing into sharper focus the complex mental categories first proposed by Meyer and Narmour. He explains,

> Applications of schema theory to music, as developed by Leonard B. Meyer, Eugene Narmour, and myself, focus on a listener's evolving interactions with a stream of musical events... The actual complexity of those real-time interactions, in which a listener attempts to relate each new sensation to learned regularities and remembered exemplars, may go far beyond verbal description. But if one focuses only on highly probable, idealized successions of events, then it may be possible to present at least an outline of a modern schema theory as it relates to musical patterns.[124]

Thus, in what follows I apply Gjerdingen's schema-theoretic approach to the cadence typologies

[120] Gjerdingen, *A Classic Turn of Phrase*. For Gjerdingen, the most common changing-note archetype featured $\hat{1}$–$\hat{2}$–$\hat{4}$–$\hat{3}$ in the melody and $\hat{1}$–$\hat{2}$–$\hat{7}$–$\hat{1}$ in the bass.

[121] Gjerdingen, *Music in the Galant style*, 111–128.

[122] Meyer, *Explaining Music*, 213.

[123] By the end of his career, Meyer used *archetype* and *schema* interchangeably to refer to a "replicated pattern class" (*The Spheres of Music*, 157).

[124] Gjerdingen, *Music in the Galant style*, 373.

articulated in the *Formenlehre* tradition to consider whether listeners with sufficient exposure to classical music have internalized the most common cadence types as a flexible network of interrelated mental representations, or *rival closing schemata*.

§2.3 The Cadence Typology: Rival Closing Schemata

2.3.1 Scale-Degree Schemata

Despite the breadth of the term's application for patterns at both local and global levels of musical organization, Gjerdingen generally reserves the schema concept for patterns whose initial and terminal events take place within the limits of short-term memory (i.e., around 6 seconds).[125] Schemata as Gjerdingen conceives them are thus "mid-size" parsings of the musical structure residing at the phrase level.[126] In his view, each schema is a cluster of constituent features and featural correlations whose manner of representation determines the historical specificity of the pattern.[127] He notes, for example, that standard music notation overspecifies a prototype's constituent features, since our mental representation of a given category is likely in no particular key or meter and may be quite general regarding the spacing of the voices, their timbres, and so on.[128] On the other hand, prototypes defined only by harmonic progressions fail to specify features relating to melodic organization that undoubtedly play an important role in perception and memory, resulting in rather coarse-grained schemata that likely transcend the period of interest.[129] Thus, Gjerdingen's representation scheme occupies a middle ground, representing each prototype as a complex correlation of melodic, contrapuntal, harmonic, and

[125] Gjerdingen, *A Classic Turn of Phrase*, 64. Simply put, listeners are far less likely to integrate sequential events into a stable pattern if they exceed this interval (Justin London, *Hearing in Time: Psychological Aspects of Musical Meter* [New York: Oxford University Press, 2004], 27).

[126] Gjerdingen, *A Classic Turn of Phrase*, 266; Gjerdingen, *Music in the Galant style*, 21.

[127] Robert O. Gjerdingen, "Courtly Behaviors," *Music Perception* 13, no. 3 (1996): 381.

[128] Gjerdingen, *Music in the Galant style*, 453.

[129] Gjerdingen, "Courtly Behaviors," 380.

metrical features.[130]

Each prototype represents individual events as gray lozenges containing a metric position (strong, weak, blank), a scale degree in the melody, a link to the next core melodic tone (up, down, same), a sonority (presented using figured bass notation), a scale degree in the bass, and a link to the next core bass tone. In other words, each prototype represents those features listeners are most likely to learn and remember while preserving the two-voice framework characterizing what Leonard Ratner once called the "classic texture."[131] Given his devotion to contrapuntal organization over harmonic progression and the emphasis he places on scale-degree patterns specifically, we might prefer the expression *scale-degree schemas*.[132]

Figure 2.2 presents Gjerdingen's *Do-Re-Mi* schema, which consists of a stepwise ascent from $\hat{1}$ to $\hat{3}$ in the melody and a lower-neighbor motion prolonging $\hat{1}$ in the bass. At first glance, this representation scheme conjures forth homorhythmic realizations of the prototype, but Gjerdingen distinguishes here between the lozenge-like *events* represented in Figure 2.2 and their corresponding *stages*, which refer to the temporal unfolding of those events within a given exemplar of the schema. In other words, the events of the *Do-Re-Mi* function for Gjerdingen either as points of reference or signs of punctuation that may be elaborated or prolonged within the stages of the schema.[133] In this regard, Gjerdingen's *event* and *stage* recall Narmour's *style form* and *style structure*, the former representing those seemingly time-independent patterns that recur with statistically significant frequency (e.g., triads and seventh chords), the latter ascribing

130 Gjerdingen, *A Classic Turn of Phrase*, 132.

131 "Classic texture maintained a tradition that was established in the early baroque style, circa 1600—*the polarity of treble and bass*. The treble carried the leading melodic line, supported by a bass part that set the harmony and provided rhythmic punctuation; middle voices completed the texture with chord tones" (Ratner, *Classic Music*, 108). In a review of *Music in the Galant Style*, David Temperley notes that there is a Schenkerian aspect to Gjerdingen's approach, in that his schemata consist only of "core" tones of the outer-voice melodies. In some senses, Gjerdingen's schemata are essentially middle-ground voice-leading patterns (Review of Music in the Galant Style, 288).

132 This is also the term Temperley prefers (Ibid., 278).

133 Gjerdingen, *Music in the Galant style*, 21–22.

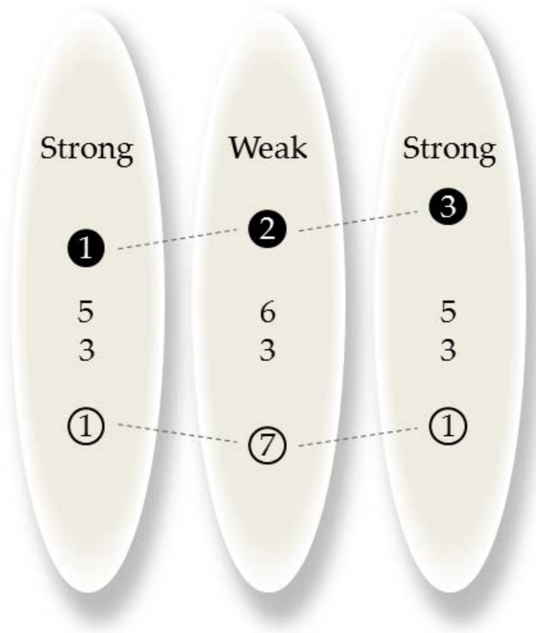


Figure 2.2: The *Do-Re-Mi* schema prototype, represented using Gjerdingen's notation (2007).

time-dependent function to those patterns, representing how they behave in real music.[134]

In many respects, Gjerdingen's schemata and Caplin's cadence categories have a great deal in common. Both theories view the compendium of schematic or categorical types as relatively short *time spans* consisting of recurrent temporal patterns, and both theories define these types according to their most representative features. That is, both theories identify schematic or cadential characteristics on the basis of their frequency of occurrence (and/or co-occurrence) within the style.

And yet, as I mentioned in the previous section, many scholars within the *Formenlehre* tradition also determine the cadential status of a given exemplar by appealing to temporal context. A cadence, they might argue, is a point of arrival, or *time point*, which marks the conclusion of an ongoing phrase-structural process. In other words, a cadence is a recurrent

[134]See especially chapter 11 of Eugene Narmour, *Beyond Schenkerism* (Chicago: The University of Chicago Press, 1977), and Gjerdingen, *A Classic Turn of Phrase*, 41–42.

pattern that serves a specific *temporal function* within a phrase or theme; it is an *end* of some sort. According to this point of view, mental representations for the cadence categories from Caplin's typology might also contain a functional attribute characterizing the temporal context in which their corresponding exemplars appear. To be sure, as both Rumelhart and Rosch point out, categories are very often characterized by the function of the objects or events they contain. Birds, for example, have perceptual attributes like beaks and feathers, and functional attributes like foraging and flight.[135]

Although Gjerdingen suggests that "formal location, especially in a late eighteenth- or early nineteenth-century repertory, has a wider range of variation and lower index of criticality than does basic harmonic progression or melodic design,"[136] the appeal to temporal context is still evident in all of the schemata he describes. From a small sample of some three hundred pairings of schemata, for example, he demonstrates the range of possible schema successions, noting that successful composers possessed "strategic knowledge of how to arrange the schemata to achieve certain aesthetic effects and to fulfill the requirements of particular moments in the course of specific musical genres."[137] The *Romanesca*, *Do-Re-Mi*, *Meyer*, and *Sol-Fa-Mi* all serve as "opening gambits," for example, while the *Prinner* functions as an especially common "riposte." Using his theory, one could traverse the permitted successions of a prototypical composition, imagining each schema as a sign-post along the formal trajectory: the *Indugio* is a "teasing delay" before a half cadence, the *Quiescenza* prolongs tonic harmony following an especially important cadence, the *Fonte* appears after the double bar in minuets or short movements, and so forth.

From this perspective, a cadence is not just any schema; it is a *closing* schema. Certainly

[135] Again, from the Roschian perspective, it is worth noting that none of these attributes defines the category 'bird.' Many extinct non-bird dinosaur species had feathers and beaks, for example, while extant bird species like ostriches and penguins do not fly. Certainly some of these characteristics are more representative of the category than others—all birds have feathers, but not all birds fly—but no attribute perfectly distinguishes birds from other categories.

[136] Robert O. Gjerdingen, "Defining a Prototypical Utterance," *Psychomusicology* 10, no. 2 (1991): 136.

[137] Gjerdingen, *Music in the Galant style*, 373.

the exemplars of the perfect authentic cadence—like the exemplars of any other schema in Gjerdingen's theory—do not *always* appear at the corresponding temporal position in a phrase or theme, as in Example 2.2; in such cases, the representativeness or criticality of the attribute decreases with each counterexample. But the point here is that exemplars of closing schemata appear with sufficient frequency at the ends of phrases and themes to justify the inclusion of functional attributes like temporal context among the network of perceptual attributes characterizing the schema. Gjerdingen makes precisely this point in relation to the *Fonte*, noting that the many relational features encapsulated in his representation scheme "contribute to defining the *fonte* prototype, as does its typical location immediately following the repeat sign."[138] Thus, if the *Romanesca* is a particularly frequent opening gesture, we can say at the very least that the schemas reflected in Caplin's typology serve to close off schema successions, with the terminal event of the cadence determining the perceptual boundary both for the schema itself, and more importantly, for the chain of schemata it follows.

Gjerdingen's scale-degree schemata and Caplin's cadence categories differ in another respect. Recall from the earlier section that for many theorists in the *Formenlehre* tradition, the final events of the cadence constitute its essential characteristics. A pattern that does not include V–I in root position and $\hat{1}$ in the soprano at the cadential arrival is not a perfect authentic cadence, they might argue. And yet for schema theorists, no feature defines a given category. In their view, a category represents the complex network of features shared by its members, with some features appearing in more members than others. It would seem that the *Aristotelian* or *classical* model and the *Roschian* or *probabilistic* approach are irreconcilable.

But in fact, for temporal formulæ like cadences, the apparent necessity of a category's final events is a common mathematical property of temporal sequences in general, one which mathematicians suggest is consistent with a *Markov process*. Put simply, if we imagine a system

[138] Gjerdingen, "Defining a Prototypical Utterance," 136.

that produces a sequence of symbols according to certain probabilities, called a *stochastic process*, a *Markov process* is a special case in which the probability of each event depends only on the previous events in the sequence.[139] And as Meyer explains, a common property of Markov processes is that the probabilities increase as the sequence unfolds.

> If music is a Markoff process, it would appear that as a musical event (be it a phrase, a theme, or a whole work) unfolds and the probability of a particular conclusion increases, uncertainty, information, and meaning will necessarily decrease. And in a closed physical system where the Markoff process operates this is just what does occur—probability tends to increase.[140]

Accordingly, one could surmise that the criticality or representativeness of each event increases as a schema unfolds, but particularly so for *closing* schemata, where the final event of the schema also serves as the final event of a much larger phrase-structural process. Thus, for an opening gambit like the *Do-Re-Mi*, the terminal event *Mi* is presumably more critical or representative of the schema than the initial events *Do* and *Re*, but probably less critical or representative than the final event of a closing schema like the perfect authentic cadence. Again, if music is a Markov process, uncertainty is "built into" the *Do-Re-Mi* because it appears at the beginning of a piece or section, where the relationships between attributes like pitch, rhythm, and meter have yet to be established, but this *systemic uncertainty* naturally decreases as the piece or section unfolds,[141] with the patterns appearing at its end serving as the most predictable, probabilistic, specifically envisaged formulæ in all of tonal music.[142]

[139] Claude Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal* 27, no. 3 (1948): 385.

[140] Leonard B. Meyer, "Meaning in Music and Information Theory," *The Journal of Aesthetics and Art Criticism* 15, no. 4 (1957): 419.

[141] Ibid.

[142] Meyer, *Emotion and Meaning in Music*, 50; Huron, *Sweet Anticipation*, 154.

Thus, it should come as no surprise that the essentialist view appears so frequently in the *Formenlehre* tradition, since listeners are far more likely to *expect* terminal events—both for the schemas themselves, and for the larger phrase-structural processes that subsume them—if they are highly probable, indeed *formulaic*. From this point of view, schematic expectations amount to probabilistic inferences, whereby frequently (co-)occurring events on the musical surface activate a given schema from memory, allowing listeners to generate expectations for the most probable continuations in prospect. The subsequent realization of those expectations in retrospect contributes to the perception of closure, and perhaps more importantly, reinforces the schema in memory.[143]

2.3.2 Cadential Strength

To this point I have said very little about the closural *strength* of the terminal events of the closing schema. If, as Schoenberg suggests, cadential formulæ are satisfying to the extent that they fulfill a stylistic expectation,[144] the perceived strength of the phrase-structural boundary might correspond with the strength of the expectations it generates. In other words, from the probabilistic view of category formation just described, the strength and specificity of our schematic expectations formed in prospect and their subsequent realization in retrospect contributes to the perception of cadential strength, where the most expected (i.e., probable) endings are also the most complete or closed.

We might hypothesize, for example, that the frequent co-occurrence of the many features in the perfect authentic cadence make it a prime candidate for schematic representation in long-

[143]Musicologists and theorists sometimes note the correspondence between expectation and probability. Barbara Barry suggests, for example, that "implications of closure are based on the probability of certain expected harmonic, textural, or dynamic gestures, either as learned procedures or remembered experiences of similar configurations"("In Search of an Ending: Reframing Mahler's Contexts of Closure," *Journal of Musicological Research* 26, no. 1 [2007]: 57).
[144]Dunsby, "Schoenberg on Cadence," 125.

term memory. Indeed, for those with sufficient exposure to classical music, the PAC category should serve as the quintessential event schema for the ends of phrases, themes, and larger sections, with the specificity of the mental representation reflected in the strength and specificity of the schematic expectations it generates. For listeners with little experience in classical music, a schema for the perfect authentic cadence may only consist of the most representative (terminal) events, resulting in relatively weak and vague expectations for exemplars heard in prospect. And yet for highly trained (or experienced) listeners, events in the initial stages of the schema, though they are less representative than the events appearing in its terminal stages, nevertheless serve as sign-posts for the impending cadential end, resulting in stronger and more specific expectations for the final event(s) of the schema.

This is especially true of perfect authentic cadences closing larger sections, such as at the end of the exposition or recapitulation in sonata-form movements. Cadences in these formal locations often feature an *expanded cadential progression* (ECP), whereby the constituent events of the schema support an entire phrase (i.e., typically at least four measures of music).[145] In such cases, the initial stages of the ECP alert the listener to the impending cadential arrival. Caplin notes, for example, that the ECP is typically composed of a four-harmony formula: I^6, ii^6 (or IV), V, and I, with the initial I^6 serving as a "conventionalized sign" or cue for the onset of the progression.[146] Additionally, the dominant stage often features a cadential six-four that both signals and significantly delays closure "from the phrase level to the highest level of structure."[147] Regarding its delay function in the classical style, Caplin notes that dominant expansion creates a "sense of heightened drama,"[148] while Meyer suggests that the durational

[145] Caplin, *Classical Form*, 254.
[146] William E. Caplin, "The "Expanded Cadential Progression": A Category for the Analysis of Classical Form," *Journal of Musicological Research* 7 (1987): 218.
[147] Meyer, *The Spheres of Music*, 245–247.
[148] Caplin, ""Expanded Cadential Progression"," 227.

emphasis indicates that "high-level closure [is] at hand."[149] Regarding its sign function, Leonard Ratner remarks in *Harmony* that "when the [six-four chord] sounds, we receive a clear and strong impression that a cadence will be made. This chord is the signal for an authentic cadence."[150] Schmalfeldt agrees, suggesting that "the six-four embellishment of the dominant transmits the unmistakable message that a cadence is underway."[151] Schoenberg relates this sign function to the conventionality of the progression and its potential to generate expectations, remarking that "the cadential six-four draws attention to itself and arouses the expectation of a certain sequel. It amounts to a certain cliché."[152]

In sum, the perfect authentic cadence is a *prospective* schema, in that the initial events serve to activate the corresponding mental representation *before* the final events take place. What is more, because the terminal events for *closing* schemata like the perfect authentic cadence are the most probabilistic, specifically envisaged, and so on, perhaps far more so than the terminal events for other schemata like the *Do-Re-Mi*, composers very often combat the tendency toward maximum certainty by introducing *designed uncertainty* at the precise moment of cadential arrival.[153] Meyer suggests, for example, that "departures from or delays in the normally expected course of musical events will be most effective where that course is most specifically and precisely envisaged, deviations will be most effective where the pattern is most complete."[154] Hence, it is precisely their position at the ends of phrases and themes that makes deviations of closing schemata so likely. For Gjerdingen, cadences from the deceptive, evaded, abandoned, and even

149 Meyer, *The Spheres of Music*, 245.
150 Leonard G. Ratner, *Harmony: Structure and Style* (New York: McGraw-Hill, 1962), 110.
151 Schmalfeldt, "Cadential Processes," 5.
152 Schoenberg, *Harmonielehre*, 141.
153 Meyer, "Meaning in Music and Information Theory," 419.
154 Meyer, *Emotion and Meaning in Music*, 50. C.P.E. Bach even relates these deviations to cadential gestures explicitly, noting that "embellishments are best applied to those places where a melody is taking shape, as it were, or where its partial, if not complete, meaning or sense has been revealed. Hence with regard to the latter case, they are found chiefly at half or full closes, caesuræ, and fermatæ"(*Essay on the True Art of Playing Keyboard Instruments*, trans. William Mitchell [New York: W. W. Norton & Co., Inc., 1949], 84).

imperfect authentic categories in Caplin's typology would therefore represent *rival closing schemata*, since they share initial events with the perfect authentic cadence but later diverge, typically at the precise moment of cadential arrival.[155] They are, in other words, primarily *retrospective* schemata because their activation very likely takes place *after* the final events have occurred.

This point does not preclude the possibility that listeners may *expect* an imperfect authentic cadence, or a deceptive cadence, or even an evaded cadence at a particular moment in the phrase-structural process—if listeners with considerable experience in the classical style have internalized the most common cadence types as a flexible *network* of interrelated mental representations, they might expect more than one potential continuation at any given moment. But the potential for events on the musical surface to activate any one of these rival closing schemata in prospect is not equally likely. Or put another way, the *strength* of the activation for each schema reflects the frequency of occurrence of the corresponding final event(s). Presumably the perfect authentic cadence serves as a schematic default for many (if not all) of the exemplars encountered in the classical style because its final events appear more frequently than those in the other categories, but listeners might also expect one (or more) of the other cadential types, albeit more weakly.[156]

But what about the other closing schemata? Could a mental representation of the imperfect authentic cadence or the half cadence serve as the default prospective schema within the network? That is, upon hearing the initial events from a given exemplar, could listeners *expect* the terminal event(s) from one of the other cadence categories, and not from the perfect authentic cadence? Recall that in Caplin's theory, the deceptive, evaded, and abandoned cadence categories initially

[155] Gjerdingen, *A Classic Turn of Phrase*, 62.

[156] Just how many continuations we might expect at any given moment is difficult to determine. As Narmour notes, "we need to know more about memory—or rather about *forgetting*, since the listener can carry only so many implicative "paths" in his head. We need to formulate some rules of memory because an analysis that enumerates every possible connection between tones can be confusing" (*Beyond Schenkerism*, 156).

promise an authentic cadence, but fundamentally deviate from the cadential formula, either at the moment of cadential arrival (DC, EV), or sometime earlier (AB). For this reason, exemplars of the DC, EV, and AB categories presumably activate a mental representation of the perfect authentic cadence in prospect and only serve as rival closing schemata in retrospect. For the other genuine categories (IAC, HC), however, it is conceivable that the initial events within Gjerdingen's representation scheme could imply the terminal event(s), resulting in a network of cadential schemata in which the genuine categories serve as schematic defaults in prospect. We might call this the *Genuine Schemas* model.

And yet because the terminal events of the perfect authentic cadence are likely the most probabilistic, many scholars seem to reject the view that all of the genuine categories may serve as default prospective schemata under the assumption that the initial events of *every* encountered exemplar—whether they form an imperfect authentic cadence, a half cadence, or any other ending formula *in toto*—most strongly activate the same essential prototype: the perfect authentic cadence. We might call this the *1-schema* model. According to this view, the PAC schema allows listeners to generate expectations concerning potential continuations, and so any deviation on the musical surface naturally results in a violation of listener expectations, and thus would be experienced as a decrease in the perceived strength of a given excerpt. Deviations from the expected final event thereby result in closing schemata of diminished strength. As a consequence, the half cadence represents the weakest cadential category; it is marked not by a deviation among the many events at cadential arrival, but rather by their absence. The half cadence is, as the term suggests, an *incomplete* cadence.

The *1-schema* model of cadential strength has received some support from music theorists. Lamenting that "a well-defined hierarchical theory of cadence-types has simply not become established," Schmalfeldt outlined the five cadential types for which the goal event closes a preceding process, which she termed *distinct closure*,[157] and of these types, she regarded the half cadence as weakest.[158] Citing Schmalfeldt's hierarchical model of cadential closure, Edward Latham also recently proposed a model that identifies and subsequently weights the criteria deemed necessary for establishing cadential closure on a 10-point scale.[159] He assigns 5 points to tonic harmony and 5 points to the preceding dominant, and he derives these scores from the scale-degrees present in the bass (1.5) and soprano (0.5), from whether the sonority is in root position (1.5), and finally from the presence of particular chord members (0.5) and a contextual feature: whether each sonority serves as a harmonic and melodic goal (1.0). Latham then scores each of Schmalfeldt's cadential types and places them along the scale. According to his criteria, the PAC category receives between 9 and 10 points (depending on whether the cadential tonic is elided), followed by IAC (8.5–9.5), DC (6.5–8.5), EV (3.5–8.5), and finally HC (3.5–5.0), positioned near the bottom of the scale. His model therefore conceptualizes a half cadence as an *incomplete* authentic cadence.

Thus, whether a half-cadential formula serves as an *incomplete* authentic cadence within the *1-schema* model on the one hand, or as a prospective schematic default in the *Genuine Schemas* model on the other, depends on whether listeners generate strong and specific expectations for the terminal events of the pattern. Caplin has suggested, for example, that the dominant, merely by virtue of its harmonic-melodic content, can represent a harmonic end: "In the half-cadential progression, the dominant itself becomes the goal harmony and so occupies the *ultimate* position. To be sure, this dominant usually resolves to tonic, one that normally initiates a new harmonic progression, but within the boundaries of the half-cadential progression itself,

[157] Schmalfeldt, "Cadential Processes," 11–12.

[158] ibid., 7. At no point, however, does Schmalfeldt explicitly compare the strength of the half cadence with modifications of the perfect authentic cadence, such as the deceptive cadence. Thus her view of the half cadence within the general hierarchy of cadential categories remains unclear.

[159] Latham, "Drei Nebensonnen," 308.

the dominant possesses enough stability to represent a harmonic end."[160] The implication here is that stability occupies a continuum, whereby the ultimate dominant in a half-cadential progression is *less* stable than the ultimate tonic in an authentic cadential progression because the half-cadential dominant—or indeed any dominant—implies further continuation to tonic harmony. And yet since this implication is presumably quite weak, that same ultimate dominant is still *more* stable than the penultimate dominant in an authentic cadential progression.

But perhaps the primary mechanism determining the stability of a given terminal event is not how strongly it implies continuation to further events, but rather how strongly it is implied by preceding events. In the *Genuine Schemas* model, the half-cadential dominant may imply further continuation, but its stability is determined at least in part by previous implications. In other words, for the IAC and HC categories, it is conceivable that the initial events within Gjerdingen's representation scheme could imply the terminal event(s), resulting in default prospective schemata for all of the genuine cadence categories.

Like the DC and EV categories, the IAC category shares initial events with the PAC category but diverges at the expected moment of cadential arrival, in this case by concluding with $\hat{3}$ in the soprano. Thus, determining if the PAC category serves as the default prospective schema for a given imperfect authentic cadence depends on whether listeners generate expectations for $\hat{3}$ at the cadential arrival. Presumably tonal melodies featuring scale degrees like $\hat{2}$ or $\hat{7}$ proceed most frequently to $\hat{1}$; indeed, corpus studies generally bear this claim out.[161] In such cases, resolutions to $\hat{3}$ are less common, and thus, less expected. On the other hand, $\hat{4}$ proceeds more often to $\hat{3}$ than to $\hat{1}$, particularly when it serves as a dissonant seventh above $\hat{5}$ in the bass, as is often the case in cadential formulæ. In other words, we might conclude that cadential melodies that descend by step to $\hat{3}$ are more likely to imply the terminal scale degree in prospect, as does

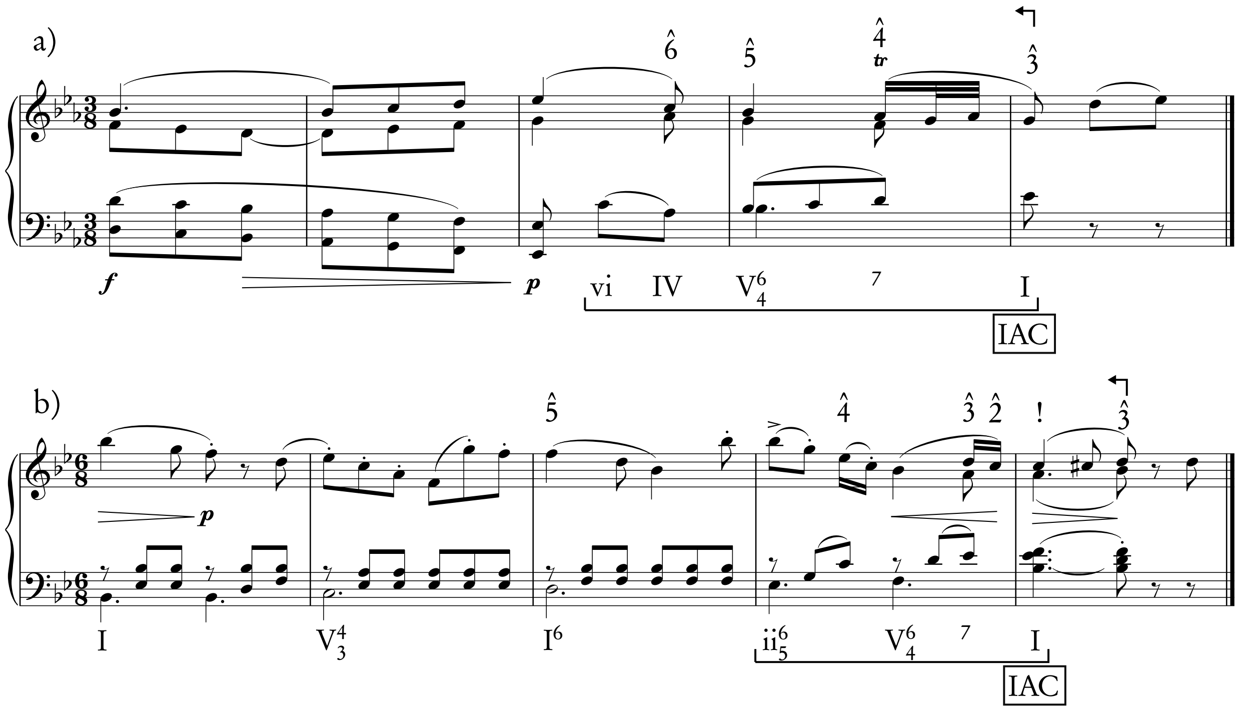
[160] Caplin, *Classical Form*, 29 (emphasis in original).
[161] For example, see Huron, *Sweet Anticipation*, 158.

Example 2.5a. Readers familiar with Gjerdingen's schemata will note the correspondence here between imperfect authentic cadences that descend to $\hat{3}$ and the *Prinner* schema, which consists of a stepwise descending tetrachord in the outer voices: from $\hat{6}$ to $\hat{3}$ in the soprano, and from $\hat{4}$ to $\hat{1}$ in the bass.[162] In this case, however, the bass presents the standard $\hat{4}$–$\hat{5}$–$\hat{1}$ cadential formula, with the stepwise descent from $\hat{4}$ to $\hat{1}$ appearing instead in the alto voice. In addition to the descending tetrachord, the soprano features a cadential trill above the penultimate dominant, another common sign-post for the impending cadential end. Thus, given the stepwise descent to $\hat{3}$, the cadential trill, and the movement's andante tempo marking, it as at least conceivable that listeners who are familiar with the IAC schema would have sufficient time to generate expectations for $\hat{3}$ before the moment of cadential arrival in m. 8.

In Example 2.5b, however, Mozart deviates from the expected PAC schema only at the moment of cadential arrival, with the melodic descent implying $\hat{1}$ and not $\hat{3}$. Here, the cadential progression appears to support an extended *Do-Si-Do* pattern in the melody, but when the cadential six-four resolves to a root-position dominant seventh, $\hat{2}$ replaces $\hat{7}$ in the soprano to effect an upward resolving suspension to $\hat{3}$ at the downbeat of m. 36, with a chromatic passing tone inserted in between.[163] Nevertheless, the sudden appearance of $\hat{2}$ at the end of m. 35 does not in itself imply the upward resolving suspension that follows; we could easily re-interpret $\hat{2}$ within a broader stepwise descent from $\hat{5}$ to $\hat{1}$ beginning in m. 34, with the *Do-Si-Do* pattern relegated to an inner voice. Thus, mm. 34–35 imply a perfect authentic cadence, with the events at the cadential arrival only activating the IAC schema in retrospect. To be sure, as Markus Neuwirth points out, analysts sometimes interpret cadences like this one as melodically

[162] Exemplars of the *Prinner* sometimes insert $\hat{5}$ between $\hat{2}$ and $\hat{1}$ in the bass, suggesting a kind of *Prinner* IAC. See, for example, Mozart's K. 282, i, mm. 2–4 (Appendix B, #12). For a discussion of the *Prinner's* cadential implications, see William E. Caplin, "Harmony and Cadence in Gjerdingen's "Prinner"," ed. Markus Neuwirth and Pieter Bergé (Leuven: Leuven University Press, 2015), 17–57.

[163] Also, see Mozart, K. 533/iii, m. 26.

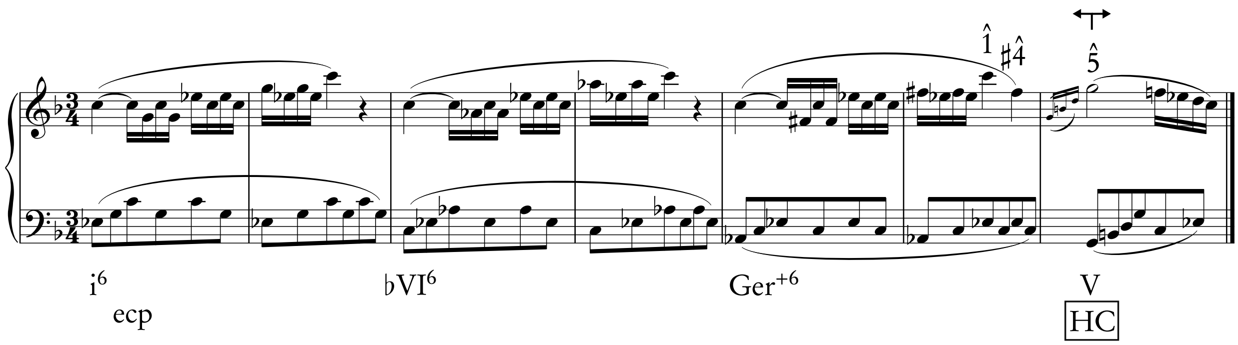


Example 2.5: a) *Prospective* IAC: Mozart, K. 281, ii, mm. 4–8. b) *Melodic Deviation* IAC: Mozart, K. 498a, iv, mm. 32–36.

deceptive deviations of the perfect authentic cadence.[164]

Whether the HC category may serve as the default prospective schema is more difficult to determine, since the terminal events of the HC schema would seem to imply a continuation that simply never materializes (see Figure 2.3). But given the prevalence of half-cadential progressions that tonicize (or strongly point to) the dominant, either by inserting $\sharp\hat{4}$ or $\flat\hat{6}$ in the bass, it seems plausible that listeners might expect the events at cadential arrival in half-cadential contexts. Shown in Example 2.6, the half cadence closing the transition in the first movement of Mozart's K. 332 features an expanded cadential progression that tonicizes the dominant with an augmented sixth chord in m. 35. As I point out later in Chapter 5, the soprano clausula $\hat{1}$–$\sharp\hat{4}$–$\hat{5}$ featured here is especially characteristic of a particular sub-type of the

[164]For contemporaneous accounts of cadential deviations featuring melodic deceptions, see Neuwirth, "Fuggir la cadenza, or the Art of Avoiding Cadential Closure," 117–130.



Example 2.6: *Prospective* HC: Mozart, K. 332, i, mm. 31–37.

half cadence that I call the *Expanding Do-Fi-Sol*, in which the (♭)$\hat{6}$–$\hat{5}$ bass clausula supports an upper register $\hat{1}$ in the soprano that leaps down to ♯$\hat{4}$ before resolving to $\hat{5}$. And since the events constituting this particular schema rarely precede the terminal events of a perfect authentic cadence, listeners with exposure to this formula might expect the events at the cadential arrival in m. 37, and not the terminal events of a perfect authentic cadence in the moments that follow.

Thus, while it is possible that listeners may *expect* a tonic resolution that simply never appears, both because half cadences often share initial events—in some cases the entire pattern—with exemplars of the PAC schema, and because half cadential formulæ are generally less common than their perfect authentic counterparts, it is also conceivable that listeners may generate expectations in prospect for exemplars from the PAC, IAC, and HC categories. What is more, recall that these categories receive special status in Caplin's theory as "the only genuine cadences... in the classical style" because they are the only categories that can achieve thematic closure.[165] In light of the preceding discussion, we might further suggest that composers reserved the PAC, IAC, and HC categories for phrases, themes, and larger sections at least in part because these categories allow listeners to generate expectations for the terminal events of the phrases, themes, or sections they close. That is, they are *genuine* because they have the potential to serve as the schematic default in prospect.

[165] Caplin, *Classical Form*, 43.

2.3.3 Terminal Events as Perceptual Boundaries

In this chapter, I have suggested that listeners who are familiar with classical music have internalized the most common cadence categories as a flexible network of rival closing schemata. During music listening, the activation of this network in prospect—and of the individual closing schemata contained within—results in the formation of expectations for the terminal events of the cadence, the fulfilment of which in retrospect serves to close off both the schema itself and the larger phrase-structural process that subsumes it.

But just how does one "close off" a *closing* schema, or indeed any schema? According to Gjerdingen, nothing links the events of a schema together except "the statistics of their common co-occurrence."[166] A schema is therefore closed or complete only when the probability between adjacent events is comparatively low. One could imagine a probability curve that might visualize what he is suggesting, where a sudden decrease in probability from one event to the next closes off the schema.

By way of example, consider the prototype for the *converging* half cadence shown in Figure 2.3. In the converging sub-type, the bass ascends by step to $\hat{5}$, and the soprano descends by step to $\hat{2}$ or $\hat{7}$.[167] For Gjerdingen, listeners internalize a schema like this one as a consequence of the frequent co-occurrence of the cluster of features presented in the first three events. If, however, a fourth event were to appear quite frequently—comprised of, say, tonic harmony in root position and $\hat{1}$ in the soprano—listeners would presumably internalize a 4-event schema,

[166] Gjerdingen, *Music in the Galant style*, 374.

[167] In Gjerdingen's theory, the converging cadence also includes an initial tonic event with $\hat{3}$ in the outer voices (ibid., 160). Nathan Martin and Julie Pedneault-Deslauriers did not include the initial tonic event in a recent corpus study of this sub-type, however (Martin and Pedneault-Deslauriers, "The Mozartean Half Cadence," 186–189). What is more, in my own corpus (presented in Part II), only 15 of the 45 half cadences from the converging sub-type feature an initial tonic event. As a consequence, I have elected to exclude the initial tonic from this prototype. This is not to say that events other than those shown in the prototype—an initial tonic with $\hat{3}$ in the bass, a cadential six-four, etc.—never appear, only that the features presented in Figure 2.3 are the most representative.

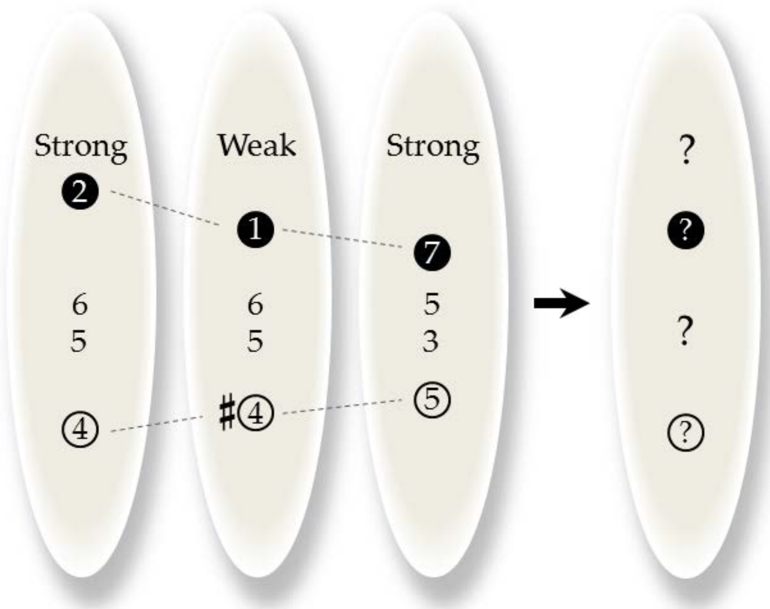


Figure 2.3: The three events of the converging cadence schema prototype followed by a fourth event, represented using Gjerdingen's notation (2007).

and not the *converging* sub-type from Figure 2.3. The distinction here thus rests on the final event of the schema.

What then, are we to make of the two cadences in Example 2.7 from the opening movement of Haydn's Op. 76, No. 2? The passage preceding the second cadence serves as a near verbatim repetition of the passage preceding the first; alterations of the original cadence appear in green. And yet, Example 2.7a culminates in a half cadence at the end of the transition, while Example 2.7b concludes with a perfect authentic cadence in the subordinate theme. How can the passages feature different schemata at their ends?

Since the cadence categories from Caplin's typology represent rival closing schemata—they share initial events but later diverge—a given exemplar of the HC schema may nest within the PAC schema.[168] Recall that I referred to such cases as incomplete authentic cadences. And yet

[168]Although schemata are generally mid-level parsings of the musical surface, Gjerdingen notes the potential for "low-level" schemata to nest within "high-level" schemata. Simple overlaps (or *elisions*) are also possible, where one

we might argue that the second chromatic event of the converging sub-type appears more often in half cadences than in the other categories because it serves to tonicize the dominant, thereby strengthening our expectations for the terminal events of the half cadence such that the HC schema would become the default mental representation during listening.

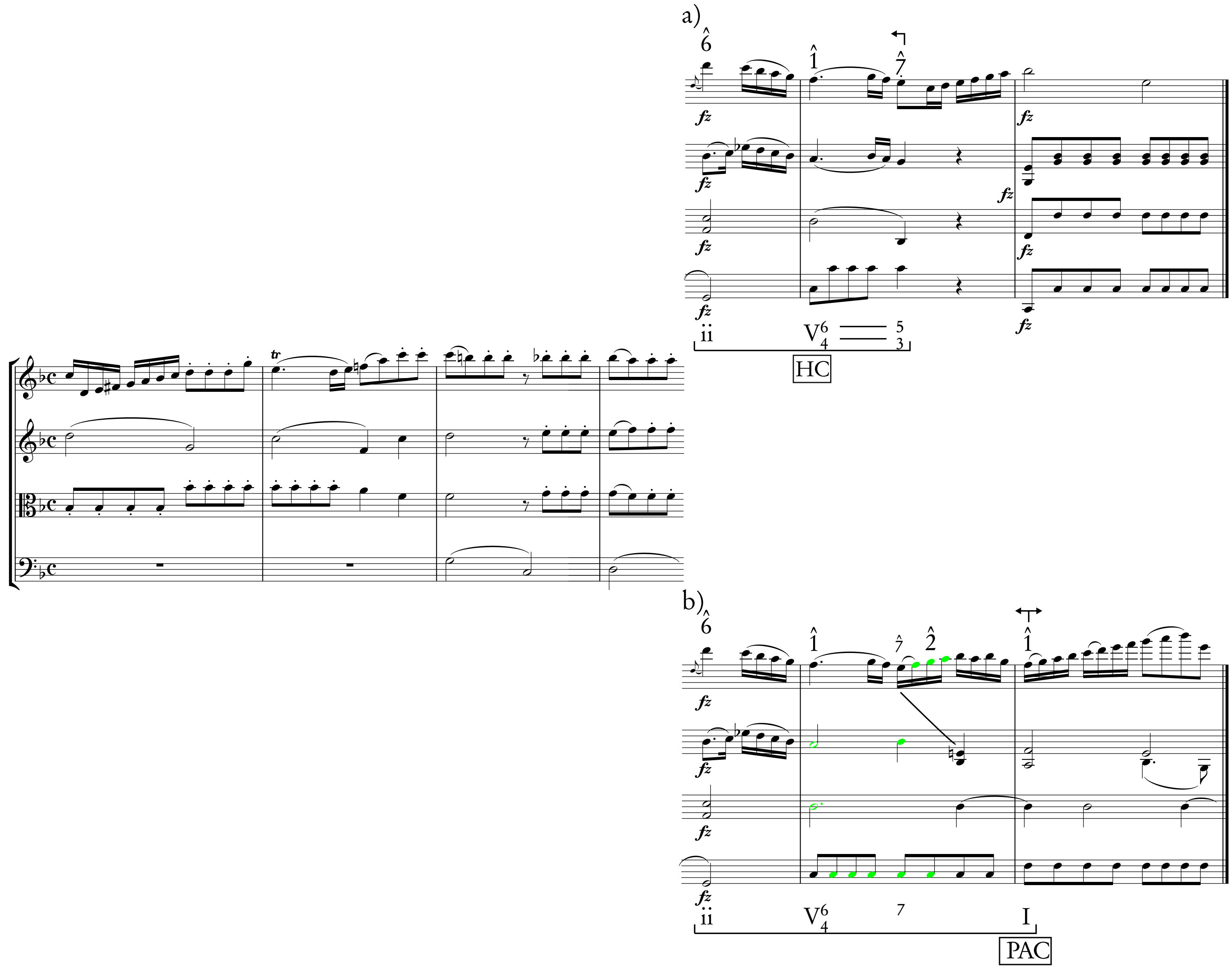
In this case, Example 2.7a is a somewhat unconventional member of the converging sub-type. The cadential bass line appears in the cello part in m. 18 at the end of the "fifths" motive (for which this quartet was named), but the superposition of the end of the fifths motive with the beginning of the cadential progression has altered the normative cadential bass line from $\hat{4}$–$\sharp\hat{4}$–$\hat{5}$ to $\hat{2}$–$\hat{5}$. Nevertheless, the cadential melody exemplifies a particularly common variant of the converging sub-type Gjerdingen has called the "High $\hat{6}$ Drop", which descends from $\hat{6}$ to $\hat{7}$ to signal the approaching close.[169] We might therefore hypothesize that the melody from Example 2.7a could activate a mental representation of the converging sub-type, thereby alerting listeners to the terminal events in m. 19; the subsequent realization of those schematic expectations then contributes to the perception of closure and reinforces the schema in memory.

When the passage returns essentially unchanged in m. 45, our schematic expectations would presumably also remain unchanged.[170] Indeed, by repeating the passage almost verbatim, Haydn likely *reinforces* our expectations for the terminal events of the HC schema. And yet rather than conclude the second passage with a *converging* half cadence in m. 49, Haydn closes the repetition with a perfect authentic cadence in m. 50. Thus, alterations of the musical materials in m. 49 are wholly unexpected; listeners expecting the converging sub-type would therefore either nest the HC schema within the PAC schema or group the events at the downbeat of m. 50 forward with the subsequent process. But if schemata like the converging sub-type amount

schema begins near the end of another schema (*Music in the Galant style*, 375–376).

[169] Ibid., 162.

[170] David Huron distinguishes schematic expectations arising from long-term memory from those *dynamic* expectations arising from short-term memories of brief—even single—exposures (*Sweet Anticipation*, 227). From this point of view, Example 2.7a engenders both types of expectation.



Example 2.7: Haydn, String Quartet in F, Op. 76/2, i, mm. 15–18. a) mm. 18–20. b) mm. 48–50. Alterations of the original cadence appear in green.

to probabilistic inferences about frequently recurring patterns, why would listeners link the event at the downbeat of m. 50 with the converging sub-type?

By considering the influence of a network of rival event schemata on the perception of passages like this one, my emphasis has been decidedly top-down. To be sure, the epigraph with which I began expresses the view that long-term memory plays an essential role in the perception of tonal cadences. To that end, I reviewed the substantial body of evidence from scholars in the learning sciences that humans internalize the many recurrent patterns they encounter in the external environment, suggesting that the perception of closure in music of the classical style depends in part on such non-verbal, implicit knowledge. Nevertheless, bottom-up sensory processes also play an important role in the perception of boundaries and the formation of event schemata.

If schematic expectations amount to probabilistic inferences, as Gjerdingen suggests, Figure 2.3 indicates that the converging schema results from the frequent co-occurrence of the first three events *and* the rare co-occurrence of any surrounding events. This schema is therefore closed or complete when the probability between events three and four is comparatively low. And yet the exemplar in Example 2.7b highlights the role played by bottom-up principles of segmental grouping on the apparent link between events three and four.

In Example 2.7a, the resolution of the cadential six-four is accompanied by a quarter-note caesura in the lower three parts, with the first violin presenting a melodic lead-in to the next measure. This decisive rhythmic break thus reinforces the perceptual boundary resulting from the realization of expectations for the terminal events of the converging sub-type. In Example 2.7b, however, the caesura is replaced by continuous surface activity and the appearance of a dissonant seventh above the cadential dominant in the second violin. What is more, the first violin regains $\hat{2}$ in preparation for the tonic resolution that follows. Thus, the events following the cadential six-four simultaneously violate schematic expectations for the terminal events of the converging sub-type while strengthening the syntactic link between the expected terminal dominant and the subsequent events at the downbeat of m. 50, thereby forcing listeners to abandon the schematic default—the HC schema—in favor of an increasingly likely alternative—the PAC schema.

The point here is that the perception of boundaries is determined not only by the fulfillment or violation of expectations for the terminal events of the previously activated schema (i.e., by top-down, prospective processes), but also by the parameters at and immediately following those events that serve to reinforce or weaken the boundary percept after the fact (i.e., by bottom-up, retrospective processes). Psychologists Christopher A. Kurby and Jeffrey M. Zacks have summarized this view as *event segmentation theory*, whereby perceivers form working memory representations of 'what is happening now,' called *event models*, and low-level discontinuities in the stimulus elicit prediction errors that force the perceptual system to update the model and segment activity into discrete events.[171] In the context of music, such discontinuities take many forms: sudden changes in texture, surface activity, rhythmic duration, dynamics, timbre, pitch, register, and so on. And when the many parameters effecting segmental grouping act together to create closure at a particular point in a work, as they do in Example 2.7a, *parametric congruence* obtains.

This view of closure matters all the more when we consider that the terminal event(s) for *closing* schemata like the converging sub-type very often determine the perceptual boundary not just for the schema itself, but also for the longer phrase-structural process it follows. In such cases, the perceived strength of the perceptual boundary—which might impact multiple levels of the structural hierarchy—depends on both top-down and bottom-up processes. Psychologists Bridgette Hard, Barbara Tversky, and David Lang note, for example, that perceivers can

[171] Chistopher A. Kurby and Jeffrey M. Zacks, "Segmentation in the Perception and Memory of Events," *Trends in Cognitive Sciences* 12, no. 2 (2008): 72–79.

hierarchically organize event sequences without prior conceptual knowledge of the commonly occurring patterns contained therein,[172] and Crystal Peebles suggests that sensory characteristics contribute to the segmentation of lower levels of hierarchically organized stimuli.[173] The perception of hierarchical structure in music is of course subject to many other phenomena, such as attention, familiarity with the stimulus,[174] and exposure to the style,[175] but it seems reasonable to conclude that a psychological approach to the perception of hierarchical event boundaries must also consider the role played by sensory processes.

Perhaps more importantly, principles of segmental grouping play an important role in the *formation* of event schemata in long-term memory. To return to my earlier point, if Example 2.7b were to occur quite frequently, listeners would presumably internalize the converging sub-type as a 4-event schema concluding with tonic harmony in root position, and not the 3-event schema shown in Figure 2.3. That is, mental representations of the parameters listeners are most likely to learn and remember depend upon the *audibility* of the many recurrent patterns contained therein. If listeners are to abstract the schema as a "discrete chunk,"[176] they must do so as a consequence of low-level sensory discontinuities that reinforce the perceptual boundaries surrounding the framing events of each encountered exemplar.[177] Gjerdingen notes, for example, that in addition to co-occurrences among the cluster of features in his representation scheme, there must also be "enough rhythmic-harmonic-melodic closure to

[172] Bridgette M. Hard, Barbara Tversky, and David S. Lang, "Making Sense of Abstract Events: Building Event Schemas," *Memory and Cognition* 34, no. 6 (2006): 1221–1235.

[173] Peebles, "The Role of Segmentation and Expectation in the Perception of Closure," 72.

[174] Elizabeth Hellmuth Margulis, "Musical Repetition Detection Across Multiple Exposures," *Music Perception* 29, no. 4 (2012): 377–385.

[175] Elizabeth Hellmuth Margulis, "Aesthetic Responses to Repetition in Unfamiliar Music," *Empirical Studies of the Arts* 31, no. 1 (2013): 45–57.

[176] Byros, "Meyer's Anvil," 290.

[177] The commitment to sensory principles in schema formation is part and parcel of modern schema theories. To use the vernacular common to schema theorists like David Rumelhart, for example, schemata are *conceptually driven* (top-down) and *data-driven* (bottom-up), which is to say that in schema-directed processing, activation goes in both directions ("Schemata," 41–42).

establish [...] schema events as perceptible points of articulation."[178] Meyer would seem to agree, suggesting that the delineation of musical patterns depends upon the emphasis they receive from other musical components, such as rhythm, dynamics, register, phrasing, and timbre.[179] This is not to say that the framing events of *every* encountered exemplar must be congruent with principles of segmental grouping for listeners to abstract the corresponding schema—Gjerdingen does not specify changes in dynamics, rhythmic duration, or tempo in his representation scheme because listeners are far less likely to remember them. Rather, my point is that principles of segmental grouping must delineate exemplars of a given schema with sufficient frequency to allow listeners to internalize the many parameters that they *will* remember: the successions of scale degrees appearing within each voice, the metric context in which they appear, the harmonies formed by their co-occurrence, and so forth. In other words, to abstract the converging sub-type, listeners should encounter Example 2.7a more frequently than they do Example 2.7b.

§2.4 Conclusions

According to Meyer, "we perceive, understand, and respond to the world, including music, in terms of the patterns and models, concepts and classifications, which have been established in our traditions—linguistic, philosophical, musical, and so on."[180] From this point of view, the replication of cadences and other closing patterns in the instrumental repertories of Haydn, Mozart, and Beethoven *define* the style to which they belong.[181] In §2.1 I argued that theories of the classical cadence described in the "New *Formenlehre*" tradition identify cadence types

[178] Gjerdingen, *A Classic Turn of Phrase*, 81. See also Gjerdingen, *Music in the Galant style*, 374–375.
[179] Meyer, *Music, the arts, and ideas*, 268; Meyer, *Explaining Music*, 83.
[180] Meyer, *Music, the arts, and ideas*, 273–274.
[181] For Meyer, a 'style' is "a replication of patterning, whether in human behavior or in the artifacts produced by human behavior, that results from a series of (tacit) choices made within some set of constraints" (*Style and Music: Theory, History, and Ideology*, 3).

according to (1) their essential surface characteristics; and (2) the temporal context in which they appear. §2.2 then described the psychological theories that explain the acquisition and mental representation of cadences, and which emphasize the role played by schematic memory in the formation of expectations during music listening. Finally, §2.3 argued that listeners with exposure to classical music have internalized the most common cadence types as a flexible network of rival closing schemata. Using Robert Gjerdingen's schema theory as a guide, I argued that the genuine cadence categories are primarily prospective schemata, whereas the cadential deviations serve as retrospective schemata.

As I hope should now be clear, the many claims made in this chapter necessitate a converging-methods approach: the first to provide a detailed study of the many recurrent closing patterns characterizing a representative corpus of instrumental works from the classical style (Part II), the second to examine the psychological relevance of existing theoretical models of the classical cadence in a series of experimental studies (Part III). Thus, in Part II I present a corpus study of the classical cadence that re-examines the cadence typology in Table 2.1 using the probabilistic approach to category formation adopted by cognitive psychologists over the last half century.

Part II

Corpus Evidence:

Eighteenth-Century Listeners

Chapter 3

Representing Closing Schemas: The Haydn Corpus

> *There is a sequence of perceptions in the mind of a listener, measured inferentially in psychology. There is a sequence of events in the air or transmission cable, measured in physics. There is an operational schema, the "score" or a "piece of music," representing certain aspects of the psychological and physical events. Each of these sequences forms an interconnected system of signs. Each sign system is closely related to the others.*
>
> Joel E. Cohen

In the previous chapter, I reviewed contemporary accounts of the classical cadence articulated in the "New *Formenlehre*" tradition and then outlined the theories that account for the acquisition and mental representation of the most common cadence types associated with the late-eighteenth-century repertories of Haydn, Mozart, and Beethoven, paying particular attention to the cadential typology presented in William E. Caplin's treatise, *Classical Form*. Citing psychologist Eleanor Rosch's work on category formation, I argued that category systems for the classical cadence are psychologically relevant if they mirror the structure of attributes encountered in classical music that listeners are likely to learn and remember.[1] Following Robert Gjerdingen's schema-theoretic approach, I then suggested that listeners who are familiar with the classical style have internalized the most common cadence types as a flexible network of rival closing schemata.

For Rosch, models of category formation depend on the statistical properties of objects and events encountered in the external environment. For our purposes, this means that the acquisition and retention of a network of cadence types depends on the frequent occurrence of these patterns on the one hand, and on a listener's repeated exposure on the other.[2] Providing evidence in support of the former claim is thus the purpose of Part II.[3]

This chapter presents the representation scheme used throughout Part II. §3.1 considers the motivations for corpus studies in music research, and §3.2 presents the corpus of expositions from Haydn's string quartets and describes how it might be digitally encoded and stored. For the purposes of pattern discovery (Chapter 4), classification (Chapter 5), and prediction (Chapter 6), §3.3 represents the most pertinent features (or *viewpoints*) from the Haydn Corpus according to the *multiple viewpoints* framework developed by Darrell Conklin and Marcus Pearce. Using Gjerdingen's schema-theoretic approach as a guide, I then represent the "core" events of the classical cadence in §3.4 according to the chromatic scale degrees and melodic contours of the outer parts (which Gjerdingen calls the "two-voice framework"), a coefficient representing the strength of the metric position, and a vertical sonority, presented as a combination of vertical interval classes or chromatic scale degrees.[4]

[1] Rosch, "Principles of Categorization," 252.

[2] Jean Mandler writes, "... repeated experiences and their internal representation lead to the phenomenon known as 'familiarity.' Because of the individual nature of experience, one person's 'familiar' organization can be another's chaos." "Categorical and Schematic Organization," 260.

[3] I leave the latter claim for Part III.

[4] Gjerdingen, *Music in the Galant style*, 142.

§3.1 Corpus Studies: Motivations

In the history of music scholarship, corpus studies are nothing new. As Gjerdingen noted in a recent issue of *Music Perception* devoted to corpus research, music historians have been collecting musical prints and manuscripts for centuries, with the word "corpus" gracing the titles of several scholarly editions.[5] But following the birth of modern computation and the proliferation of data in machine-readable formats, corpus studies have come to denote the collection and statistical analysis of large bodies of data,[6] typically using automated procedures made available by relatively recent advances in computer processing power.[7]

Computational corpus studies got their start in the early 1960s when linguists Henry Kučera and W. Nelson Francis created the first machine-readable corpus of American English at Brown University.[8] Compiled from five hundred samples of English-language text and consisting of over one million words, the *Brown Corpus* laid the groundwork for the study of natural languages using field-collected samples.[9] And as a consequence of innovations like optical character recognition (OCR), which automatically transcribes printed text into digital formats, natural language corpora using automatic transcription methods are now commonplace in language research.[10]

In the 'data-rich' environment characterizing present-day scholarship,[11] interrelated sub-

[5] Robert O. Gjerdingen, "'Historically Informed' Corpus Studies," *Music Perception* 31, no. 3 (2014): 192.

[6] David Temperley and Leigh VanHandel, "Introduction to the Special Issue on Corpus Methods," *Music Perception* 31, no. 1 (2013): 1.

[7] Cory McKay and Ichiro Fujinaga, "Style-Independent Computer-Assisted Exploratory Analysis of Large Music Collections," *Journal of Interdisciplinary Music Studies* 1, no. 1 (2007): 64.

[8] Charles F. Meyer, *English Corpus Linguistics: An Introduction* (New York: Cambridge University Press, 2002).

[9] Henry Kucera and W. Nelson Francis, *Computational Analysis of Present-Day American English* (Providence, RI: Brown University Press, 1967).

[10] Despite its relatively modest size relative to current language corpora—the Corpus of Contemporary American English is currently the largest corpus of American English at over 450 million words—the *Brown Corpus* remains a significant lexical resource in corpus linguistics. See, for example, Christopher D. Manning and Hinrich Schütze, *Foundations of Statistical Natural Language Processing* (Cambridge, MA: MIT Press, 1999).

[11] David Huron, "The New Empiricism: Systematic Musicology in a Postmodern Age," in *1999 Ernest Bloch*

fields like corpus linguistics, computational linguistics, and natural language processing provide but three examples from the plethora of emerging sub-disciplines witnessed over the past few decades. Fields like biology (computational biology, bioinformatics), psychology (cognitive science, artificial intelligence), and of course, music research (music information retrieval, empirical musicology) have all instituted encoding initiatives at one time or another, and all now borrow and share corpus-based methods quite freely.

And yet, despite the growth of corpus studies over the past few decades, both in the *Formenlehre* tradition and in the discipline at large, Markus Neuwirth has characterized the prevailing approach adopted by *Formenlehre* theorists as one based on what statistician David Fischer has called "intuitive statistics,"[12] with scholars frequently eschewing explicit statistical methods in favor of qualitative descriptions derived from empirical observation. Caplin states in the Introduction to *Classical Form*, for example, that "the account of classical form given here is a 'theory' only in an informal sense… Principles are derived from empirical observation and are largely descriptive. No attempt is made to ground the concepts in some broader system of mathematics, logic, cognition, or the like, and no proof is offered for the many assertions made."[13] Like Caplin, James Hepokoski and Warren Darcy also rely on empirical observation, characterizing *Elements of Sonata Theory* as a "research report, the product of our analyses of hundreds of individual movements by Haydn, Mozart, Beethoven, and many surrounding composers of the time (as well as later composers)."[14] But as Paul Wingfield points out, the authors fail to provide "a full account of the sample, complete descriptive statistics and an explanation of sampling methodology."[15]

Lectures (University of California, Berkeley, 1999).

[12] Neuwirth, "Recomposed Recapitulations," 34.

[13] Caplin, *Classical Form*, 5.

[14] Hepokoski and Darcy, *Elements of Sonata Theory*, v.

[15] Paul Wingfield, "Beyond 'Norms and Deformations': Towards a Theory of Sonata Form as Reception History," *Music Analysis* 27, no. 1 (2008): 141.

With the drive toward digitization now in full swing, statistical methods provide powerful analytic tools, enabling the testing of a priori hypotheses for bodies of music that often far exceed the capacities of one scholarly lifetime,[16] and allowing the analyst to uncover empirical evidence that remains open to falsification and subsequent replication.[17] To be sure, according to David Huron, the impetus for corpus studies "is not merely some obsession with things numerical, or a kleptophilic compulsion to collect, but a proper moral imperative," where the desire for truth, knowledge, "and other good things" depends on the quality and quantity of the collected evidence.[18] Leonard Meyer summarizes this point nicely:

> Since all classification and all generalization about stylistic traits are based on some estimate of relative frequency, statistics are inescapable. This being so, it seems prudent to gather, analyze, and interpret statistical data according to some coherent, even systematic, plan… it would appear desirable to define as rigorously as possible what is to count as a given trait, to gather data about such traits systematically, and to collate and analyze it consistently and scrupulously—in short, to employ the highly refined methods and theories developed in the discipline of mathematical statistics and sampling theory.[19]

But perhaps the most important motivation for corpus studies follows from the prevailing view in cognitive psychology that humans learn and comprehend complex, rule-governed structures like natural language merely by exposure during early development. If *implicit*

[16]Jonathan Wild, "A Review of the Humdrum Toolkit: UNIX Tools for Musical Research, created by David Huron," *Music Theory Online* 2.7 (1996).

[17]Gjerdingen has pointedly observed that music scholars "bandy about words like 'typical,' 'characteristic,' or 'standard' with the open confidence of embezzlers who, knowing that they alone keep the books, cannot imagine being called into account" ("Defining a Prototypical Utterance," 127). For a discussion of statistical methods in the *Formenlehre* tradition, see Neuwirth, "Recomposed Recapitulations," 25–67.

[18]David Huron, "On the Virtuous and the Vexatious in an Age of Big Data," *Music Perception* 31, no. 1 (2013): 5.

[19]Meyer, *Style and Music: Theory, History, and Ideology*, 64.

learning is indeed the primary mechanism by which we acquire knowledge about the world around us, a representative sample of works in the classical style will serve as a proxy for the musical experiences of listeners situated in that style.[20] And so for those with sufficient exposure to the eighteenth-century instrumental repertories of Haydn, Mozart, and Beethoven—whether deceased members of the Viennese courts or modern listeners who have immersed themselves in classical music—a detailed study of the recurrent patterns found therein is "at once an inventory of part of their musical knowledge."[21] Thus, my hope is that examining a large number of cadences from a relatively narrow, historically limited corpus will provide a clearer view of the cadence types characterizing the classical style, as well as offer empirical evidence for the kinds of closing patterns that listeners may learn implicitly.

§3.2 The Haydn Corpus

The Haydn Corpus consists of symbolic representations of 50 *sonata-form* expositions selected from Haydn's string quartets (1771–1803).[22] The choice to limit this corpus to one composer, genre, and form resulted from a number of considerations. First, restricting the investigation to Haydn's string quartets ensured that idiomatic differences in compositional style would not affect the reported findings. Second, this genre—perhaps more so than any other—allows the

[20] Byros, "Meyer's Anvil," 278.

[21] Gjerdingen, "Courtly Behaviors," 380–381.

[22] Music corpora exist in symbolic and audio formats. Symbolic representations include printed notes, scores, and text, and a number of encoding formats are now prevalent in the research community, including Musical Instrument Digital Interface (MIDI), Kern, and MusicXML, with software like Huron's *Humdrum Toolkit* (*The Humdrum Toolkit*, 1993, `music-cog.ohio-state.edu/Humdrum/index.html`), the *MidiToolbox* in Matlab (Tuomas Eerola and Petri Toiviainen, *MIDI Toolbox: MATLAB Tools for Music Research*, 2004), and Michael ₵hbert's *Music21* in Python providing frameworks for encoding and analysis ("music21: A Toolkit for Computer-Aided Musicology and Symbolic Music Data," in *11th International Society for Music Information Retrieval Conference*, ed. J. Stephen Downie and Remco C. Veltkamp [2010], 637–642). For reviews of the *Humdrum Toolkit* and *Music21*, see Wild, "A Review of the Humdrum Toolkit: UNIX Tools for Musical Research, created by David Huron"; Dmitri Tymoczko, Review of *Music21: a Toolkit for Computer-aided Musicology*, by Michael Cuthbert, *Music Theory Online* 19, no. 3 (2013).

analyst to model the individual instrumental parts both independently and in combination, thereby simplifying the discovery of composite events like dyads and triads in complex polyphonic textures. Finally, since most of the basic musical materials from sonata-form movements appear in the exposition, with cadences appearing less frequently in the development, and with recapitulations often repeating verbatim the cadential material from the exposition, it seemed more reasonable to restrict the corpus to expositions, thereby mitigating the potential for repeated or unnecessary material in the corpus. In sum, the Haydn Corpus is small enough to allow for analytical annotations regarding the key and mode, the position and identification of cadences of various types, and so on, yet large enough to permit the analyst to ask more general questions about the articulation of cadences in music of the classical style.

Table 3.1 presents the reference information, keys, time signatures, and tempo markings for each movement. The corpus spans much of Haydn's mature compositional style (Opp. 17–76), with the majority of the expositions selected from first movements (28) or finales (11), and with the remainder appearing in inner movements (ii: 8; iii: 3). All movements were downloaded from the KernScores database in the MIDI format and analyzed in Matlab.[23] To ensure that each instrumental part would qualify as monophonic—a pre-requisite for many of the analytical techniques that follow—all trills, extended string techniques, and other ornaments were removed.[24] Note velocities and durations were quantized exactly, and the tempo in beats-per-minute (bpm) for each movement was determined by score or convention (see the tempo markings in Table 3.1). Table 3.2 provides a few descriptors concerning the number of events in each movement for each instrumental part.

Most natural languages consist of a finite alphabet of discrete symbols (letters), combinations of which form words, phrases, and so on. As a result, the mapping between the individual letter

[23] `http://kern.ccarh.org/`.

[24] For events presenting extended string techniques (e.g., double or triple stops), I retained the note event in each part that preserved the voice leading both within and between instrumental parts.

Table 3.1: Reference information (Opus number, work, movement, measures), keys (case denotes mode), time signatures, and tempo markings for the exposition sections of each movement in the Haydn Corpus.

Excerpt	*Key*	*Time Signature*	*Tempo Marking*
Op. 17, No. 1, i, mm. 1–43	E	C	Moderato
Op. 17, No. 2, i, mm. 1–38	F	C	Moderato
Op. 17, No. 3, iv, mm. 1–26	E♭	C	Allegro molto
Op. 17, No. 4, i, mm. 1–53	c	C	Moderato
Op. 17, No. 5, i, mm. 1–33	G	C	Moderato
Op. 17, No. 6, i, mm. 1–73	D	6/8	Presto
Op. 20, No. 1, iv, mm. 1–55	E♭	2/4	Presto
Op. 20, No. 3, i, mm. 1–94	g	2/4	Allegro con spirito
Op. 20, No. 3, iii, mm. 1–43	G	3/4	Poco Adagio
Op. 20, No. 3, iv, mm. 1–42	g	C	Allegro molto
Op. 20, No. 4, i, mm. 1–112	D	3/4	Allegro di molto
Op. 20, No. 4, iv, mm. 1–49	D	C	Presto scherzando
Op. 20, No. 5, i, mm. 1–48	f	C	Allegro moderato
Op. 20, No. 6, ii, mm. 1–27	E	₵	Adagio
Op. 33, No. 1, i, mm. 1–37	b	C	Allegro moderato
Op. 33, No. 1, iii, mm. 1–40	D	6/8	Andante
Op. 33, No. 2, i, mm. 1–32	E♭	C	Allegro moderato
Op. 33, No. 3, iii, mm. 1–29	F	3/4	Adagio
Op. 33, No. 4, i, mm. 1–31	B♭	C	Allegro moderato
Op. 33, No. 5, i, mm. 1–95	G	2/4	Vivace assai
Op. 33, No. 5, ii, mm. 1–30	g	C	Largo
Op. 50, No. 1, i, mm. 1–60	B♭	₵	Allegro
Op. 50, No. 1, iv, mm. 1–75	B♭	2/4	Vivace
Op. 50, No. 2, i, mm. 1–106	C	3/4	Vivace
Op. 50, No. 2, iv, mm. 1–86	C	2/4	Vivace assai
Op. 50, No. 3, iv, mm. 1–74	E♭	2/4	Presto
Op. 50, No. 4, i, mm. 1–64	f♯	3/4	Allegro spirituoso
Op. 50, No. 5, i, mm. 1–65	F	2/4	Allegro moderato
Op. 50, No. 5, iv, mm. 1–54	F	6/8	Vivace
Op. 50, No. 6, i, mm. 1–54	D	C	Allegro
Op. 50, No. 6, ii, mm. 1–25	d	6/8	Poco Adagio
Op. 54, No. 1, i, mm. 1–47	G	C	Allegro con brio
Op. 54, No. 1, ii, mm. 1–54	C	6/8	Allegretto
Op. 54, No. 2, i, mm. 1–87	C	C	Vivace
Op. 54, No. 3, i, mm. 1–58	E	₵	Allegro
Op. 54, No. 3, iv, mm. 1–82	E	2/4	Presto
Op. 55, No. 1, ii, mm. 1–36	D	2/4	Adagio cantabile
Op. 55, No. 2, ii, mm. 1–76	f	₵	Allegro
Op. 55, No. 3, i, mm. 1–75	B♭	3/4	Vivace assai
Op. 64, No. 3, i, mm. 1–69	B♭	3/4	Vivace assai
Op. 64, No. 3, iv, mm. 1–79	B♭	2/4	Allegro con spirito
Op. 64, No. 4, i, mm. 1–38	G	C	Allegro con brio
Op. 64, No. 4, iv, mm. 1–66	G	6/8	Presto
Op. 64, No. 6, i, mm. 1–45	E♭	C	Allegretto
Op. 71, No. 1, i, mm. 1–69	B♭	C	Allegro
Op. 74, No. 1, i, mm. 1–54	C	C	Allegro moderato
Op. 74, No. 1, ii, mm. 1–57	G	3/8	Andantino grazioso
Op. 76, No. 2, i, mm. 1–56	d	C	Allegro
Op. 76, No. 4, i, mm. 1–68	B♭	C	Allegro con spirito
Op. 76, No. 5, ii, mm. 1–33	F♯	₵	Largo. Cantabile e mesto

Table 3.2: Descriptive statistics for the Haydn Corpus.

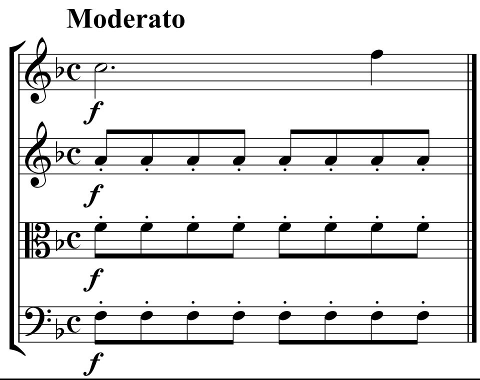
Instrumental Part	*N*	*M* (*SD*)	*Range*
Violin 1	14,506	290 (78)	133–442
Violin 2	10,653	213 (70)	69–409
Viola	9156	183 (63)	79–381
Cello	8463	169 (60)	64–326

Note. *N* refers to the number of note events, *M* denotes the mean rounded to the nearest integer, *SD* represents the standard deviation, also rounded to the nearest integer, and the *Range* indicates the lowest and highest values.

or word encountered in a printed text and its symbolic representation in a computer database is essentially one-to-one. Music encoding is considerably more complex. Notes, chords, phrases, and the like are characterized by a number of different features, and so regardless of the unit of meaning, digital encodings of individual events must concurrently represent multiple properties of the musical surface. To that end, many symbolic formats encode standard music notation as a series of discrete event sequences in an $m \times n$ matrix, where m denotes the number of events in the symbolic representation, and n refers to the number of encoded features (e.g., pitch, onset time, rhythmic duration, etc.).

By way of example, Figure 3.1 presents the note matrix provided by the *MIDI Toolbox* for the first measure from the opening movement of Haydn's String Quartet in F, Op. 17/2. The columns of the note matrix refer to the following characteristics: (1) onset time, measured in quarter-note beats; (2) duration in quarter-note beats; (3) instrumental part (or MIDI channel), with the instrumental parts ordered from 0–3 (beginning with the first violin); (4) pitch in semitones, where middle C (C_4) is 60; (5) velocity, which in MIDI nomenclature describes how quickly the key is pressed, and thus, how loudly the note is played (0–127); (6) onset time in seconds; and (7) duration in seconds.[25]

[25]Tuomas Eerola and Petri Toiviainen, "MIR in Matlab:The MIDI Toolbox," in *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR)* (2004), 22.



Onset (beats)	Duration (beats)	MIDI channel	Pitch	Velocity	Onset (seconds)	Duration (seconds)
0	3	0	72	64	0	1.8
0	0.5	1	69	64	0	0.3
0	0.5	2	65	64	0	0.3
0	0.5	3	53	64	0	0.3
0.5	0.5	1	69	64	0.3	0.3
0.5	0.5	2	65	64	0.3	0.3
0.5	0.5	3	53	64	0.3	0.3
1	0.5	1	69	64	0.6	0.3
1	0.5	2	65	64	0.6	0.3
1	0.5	3	53	64	0.6	0.3
1.5	0.5	1	69	64	0.9	0.3
1.5	0.5	2	65	64	0.9	0.3
1.5	0.5	3	53	64	0.9	0.3
2	0.5	1	69	64	1.2	0.3
2	0.5	2	65	64	1.2	0.3
2	0.5	3	53	64	1.2	0.3
2.5	0.5	1	69	64	1.5	0.3
2.5	0.5	2	65	64	1.5	0.3
2.5	0.5	3	53	64	1.5	0.3
3	1	0	77	64	1.8	0.6
3	0.5	1	69	64	1.8	0.3
3	0.5	2	65	64	1.8	0.3
3	0.5	3	53	64	1.8	0.3
3.5	0.5	1	69	64	2.1	0.3
3.5	0.5	2	65	64	2.1	0.3
3.5	0.5	3	53	64	2.1	0.3

Figure 3.1: Top: Haydn, String Quartet in F, Op. 17/2, i, m. 1. Bottom: Event representation provided by the *MIDI Toolbox*.

The note matrix is fairly self-explanatory, but two further comments are warranted here. First, when the movement begins on a metric downbeat, as does the example in Figure 3.1, the onset vector measured in quarter-note beats begins at `0` (and not 1). Second, as a consequence of the quantization step during pre-processing, the velocity vector in the note matrix does not vary, so this feature will be excluded from further discussion.

The *MIDI Toolbox* obtains the first five columns of the note matrix in Figure 3.1 from the note-event messages of the MIDI protocol, but the MIDI file also encodes meta-events that represent more general features of the music, such as the key, tempo, meter, and time signature.[26] The tempo associated with the note matrix is `100` bpm, for example, so given the onset and duration vectors measured in quarter-note beats and the tempo provided by the MIDI file, we can easily derive either of the onset and duration vectors in seconds shown in the right-most columns of the note matrix:

$$\mathtt{dur}_{\text{sec}}(e_i) = \frac{\mathtt{dur}_{\text{beats}}(e_i) \times 60}{\mathtt{bpm}(e_i)}$$

where event e_i refers to the i^{th} row of the note matrix.[27] The appeal of symbolic representations like this one is that they encode only the most important features, thereby requiring very little computer memory while remaining flexible enough to allow the researcher to derive any further features with relative ease. From the vector of pitches from the twelve-tone chromatic scale, for example, we could easily obtain a sequence of melodic contours or intervals for each instrumental part, or a vector of vertical intervals between one or more parts. If we also include information from the MIDI file's meta-events (e.g., key and time signatures, tempo changes, etc.), we can derive vectors representing metric positions, scale degrees, variations in the onset

[26]Eleanor Selfridge-Field, ed., *Beyond MIDI: The Handbook of Musical Codes* (Cambridge, MA: MIT Press, 1997), 53.

[27]To obtain the corresponding values for the onset vector measured in seconds, one need only replace `dur` with `onset` in the equation.

and duration vectors measured in seconds, and so on.

Figure 3.1 also demonstrates a few limitations of the Haydn Corpus. Because the encoded representations were not aligned with selected performances, clock-time measures like those represented in columns six and seven reflect a metronomic interpretation of musical time that necessarily departs from the kinds of everyday encounters with this repertory we might hope to study. Ideally, we could align the encoded representations with performances, either by employing tempo-alignment algorithms, or by annotating an isochronous pulse at a given metrical level in the performance and then aligning the encoded representation to the obtained tempo curve, but such is the encoding bottleneck that time-aligned symbolic representations are exceedingly rare in the research community. What is more, as I mentioned in the previous section, the MIDI format does not distinguish between enharmonic equivalents—F♯ and G♭ are represented by the same numeric value (66). As a result, the encoded representation shown in the note matrix is too reductive to capture the entire pitch alphabet, and so distributional analyses for pitch-based features like pitch class or scale degree will generally be restricted to alphabets of 12 elements.[28]

§3.3 Representing Cadences with *Multiple Viewpoints*

Representation schemes like the one presented in Figure 3.1 roughly correspond to the *multiple-viewpoint* framework first proposed by Darrell Conklin in the late 1980s, and later extended and refined by his student, Marcus Pearce.[29] Conklin's primary aim was to apply statistical

[28]John Snyder has been particularly critical of this limitation ("Entropy as a Measure of Musical Style: The Influence of A Priori Assumptions," *Music Theory Spectrum* 12, no. 1 [1990]: 121–160).

[29]Darrell Conklin, "Modelling and Generating Music Using Multiple Viewpoints," in *Proceedings of the First Workshop on Artificial Intelligence and Music* (St. Paul, MN, 1988), 125–137; Darrell Conklin, "Prediction and Entropy of Music" (MA Thesis, University of Calgary, 1990); Darrell Conklin and Ian H. Witten, "Multiple Viewpoint Systems for Music Prediction," *Journal of New Music Research* 24, no. 1 (1995): 51–73; Marcus T. Pearce and Geraint A. Wiggins, "Improved Methods for Statistical Modelling of Monophonic Music," *Journal of New*

modeling procedures from the machine learning and prediction of language to domains such as music, where events have a multidimensional structure.[30] Like the note matrix in Figure 3.1, the multiple-viewpoint framework accepts sequences of musical events that typically correspond to individual notes as notated in a score. Each event e consists of a set of *basic attributes*—what I have up until this point been calling 'features'—and each attribute is associated with a *type*, τ, which specifies the properties of that attribute. The *syntactic domain* (or alphabet) of each type, $[\tau]$, denotes the set of all unique elements associated with that type, and each element of the syntactic domain also maps to a corresponding set of elements in the *semantic domain*, $[\![\tau]\!]$. Following Conklin, attribute types appear here in typewriter font to distinguish them from ordinary text. In the twelve-tone chromatic scale, for example, the type `pitch class` would consist of the syntactic set, $\{0, 1, 2, \ldots, 11\}$, and the semantic set, {C,C♯/D♭,D,...,B}.[31]

Within this representation language, Conklin defines several distinct classes of type, but we will concern ourselves in what follows with just three: *basic*, *derived*, and *linked*.[32] Basic types are irreducible representations of the musical surface—that is, they cannot be derived from any other type. Thus, an attribute representing the sequence of pitches from the twelve-tone chromatic scale—hereafter referred to as *chromatic pitch*, or `cpitch`—would serve as a basic type in Conklin's approach because it cannot be derived from a sequence of pitch classes, scale degrees, melodic intervals, or indeed, any other attribute. What is more, basic types represent

---

Music Research 33, no. 4 (2004): 367–385; Marcus T. Pearce, "The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition" (PhD Dissertation, City University, London, 2005); Marcus T. Pearce, Darrell Conklin, and Geraint A. Wiggins, "Methods for Combining Statistical Models of Music," in *Computer Music Modelling and Retrieval*, ed. U. K. Wiil (Heidelberg, Germany: Springer Verlag, 2005), 295–312.

[30] Conklin and Witten, "Multiple Viewpoint Systems for Music Prediction," 57–58.

[31] Note that in this example the semantic domain of pitch-class names is necessarily larger than the corresponding syntactic domain as a consequence of enharmonic equivalence, hence the appearance of two labels C♯ and D♭ for the value 1.

[32] For a review of multiple-viewpoint systems, including a discussion of the viewpoint classes defined by Conklin and Pearce, see Conklin and Witten, "Multiple Viewpoint Systems for Music Prediction"; Pearce, "The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition," 49–79.

every event in the corpus. For example, a sequence of melodic contours would not constitute a basic type because either the first or last events of the melody would receive no value.[33] Indeed, an interesting property of the set of n basic types for any given corpus is that the Cartesian product of the domains of those types determines the *event space* for the corpus, denoted by ξ:

$$\xi = [\tau_1] \times [\tau_2] \times \ldots \times [\tau_n]$$

Each event consists of an n-tuple in ξ—a set of values corresponding to the set of basic types that determine the event space. ξ therefore denotes the set of all representable events in the corpus.[34] To model a corpus of Bach chorales, Conklin identified six basic types: chromatic pitch (`pitch`), key signature (`keysig`), time signature (`timesig`), fermata (`fermata`), start time (`st`), and duration (`duration`).[35] With the exception of `fermata`, the MIDI format encodes all of these types either as note or meta events, and the *MIDI Toolbox* represents three of these types explicitly in its note matrix representation (see columns 1, 2, and 4 in Figure 3.1).

As should now be clear from the examples given above, derived types like pitch class, scale degree, and melodic interval do not appear in the event space but are derived from one or more of the basic types. Thus, for every type in the encoded representation there exists a partial function, denoted by Ψ, which maps sequences of events onto elements of type τ. The term *viewpoint* therefore refers to the function associated with its type, but for convenience both authors refer to viewpoints by the types they model.[36] The function is partial because the output may be undefined for certain events in the sequence (denoted by $\perp$). Again, viewpoints

[33] In a melody of length n, a sequence of contours will necessarily be of length $n - 1$.

[34] Pearce, "The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition," 58–59.

[35] In the representation scheme outlined in his dissertation, Pearce expanded the list of basic types described by Conklin by omitting `fermata` and `timesig` and including rest duration (`deltast`), bar length (`barlength`), metric pulses (`pulses`), mode (`mode`), and phrasing (`phrase`) (Ibid., 63).

[36] Pearce explains that for basic types, Ψ_τ is simply a projection function, thereby returning as output the same values it receives as input (Ibid., 59).

for attributes like melodic contour or melodic interval demonstrate this point, since either the first or last element will receive no value (i.e., it will be undefined).

Basic and derived types attempt to model the relations within attributes, but they fail to represent the relations *between* attributes. I argued in Chapter 2 that prototypical utterances are comprised of a cluster of co-occurring features, and the relations between those features could be just as significant as their presence (or absence). In isolation, the harmonic progression V–I does not provide sufficient grounds for the identification of a perfect authentic cadence, but the co-occurrence of that progression with $\hat{1}$ in the soprano, a six-four sonority preceding the root-position dominant, or a trill above the dominant makes such an interpretation far more likely. Thus, linked viewpoints attempt to model correlations between attributes by calculating the cross-product of their constituent types.[37] We might hypothesize, for example, that the succession of scale degrees in the bass voice interacts with the chordal sonorities it supports, and a viewpoint linking these attributes measures this interaction explicitly.

§3.4 Viewpoint Selection

Taken together, basic, derived, and linked viewpoints form an elegant multiple-viewpoint system for the representation and analysis of music. But how do we select the appropriate viewpoints for the representation of cadences in Haydn's string quartets? According to Gjerdingen's schema-theoretic approach, a cadence is best understood as an instance of bass-melody co-articulation. Gjerdingen represents the "core" events of the cadence by the scale degrees and melodic contours of the outer voices (i.e., the two-voice framework), a coefficient representing the strength of the metric position (strong, weak), and a sonority, presented using figured bass notation. Given the

[37] For readers familiar with David Lewin's direct-product systems, a linked viewpoint models a *product type* (*Generalised Musical Intervals and Transformations* [New Haven, CT: Yale University Press, 1987], 1–15); Conklin and Witten, "Multiple Viewpoint Systems for Music Prediction," 12–13.

importance of melodic intervals in studies of recognition memory for melodies,[38] we might also add this attribute to Gjerdingen's list.

3.4.1 Note Events

Melodic Interval. The melodic interval of an event is represented as an integer by the attribute type `melint`. Given the chromatic pitch vector provided by the symbolic representation, which I will hereafter call `cpitch`, we derive a sequence of melodic intervals by the function:

$$\Psi_{\texttt{melint}}(e_i) = \begin{cases} \perp & \text{if } i = 1, \\ \Psi_{\texttt{cpitch}}(e_i) - \Psi_{\texttt{cpitch}}(e_{i-1}) & \text{otherwise.} \end{cases} \tag{3.1}$$

Here, we obtain the distance between adjacent pitches as an integer, where ascending intervals are positive and descending intervals are negative.

Contour. The viewpoint `contour` reduces the information present in `melint` still further. Starting from the basic type `cpitch`, we derive a sequence of melodic contours from the function:

$$\Psi_{\texttt{contour}}(e_i) = \begin{cases} -1 & \text{if } \Psi_{\texttt{cpitch}}(e_i) < \Psi_{\texttt{cpitch}}(e_{i-1}) \\ 0 & \text{if } \Psi_{\texttt{cpitch}}(e_i) = \Psi_{\texttt{cpitch}}(e_{i-1}) \\ 1 & \text{if } \Psi_{\texttt{cpitch}}(e_i) > \Psi_{\texttt{cpitch}}(e_{i-1}) \end{cases} \tag{3.2}$$

where all ascending intervals receive a value of 1, all descending intervals a value of –1, and all lateral motion a value of `0`. This viewpoint assumes all ascending motion is equivalent, whether by a semitone or an octave.

---

[38] See, for example, W. Jay Dowling, "The Importance of Interval Information in Long-Term Memory for Melodies," *Psychomusicology* 1 (1981): 30–49.

Scale Degree. I derived `melint` and `contour` from `cpitch`, but a viewpoint relating the chromatic pitches in each movement to a referential tonic pitch class cannot be obtained from the symbolic representation alone. In some studies, the referential tonic is determined from the opening key signature (e.g., four sharps), and from an assessment of the underlying mode (E major or C♯ minor), resulting in scale-degree distributions that do not control for modulations or changes in modality.[39] Key-finding algorithms have also become more common in recent decades, allowing researchers to automatically identify the key of a passage with high degrees of accuracy ($>$90%).[40] Nevertheless, the lack of available annotated corpora indicating modulations and changes of mode makes testing these algorithms quite difficult.[41] To resolve these issues, I manually annotated the key, mode, modulations, and pivot boundaries for each movement in the corpus and then included the analysis in a separate text file to accompany the MIDI representation. Thus, every note event in the corpus was associated with the viewpoints `key` and `mode`. The vector of keys assumes values in the set {0,1,2,...,11}, where `0` represents the key of C, 1 represents C♯ or D♭, and so on. Passages in the major and minor modes receive values of `0` and 1, respectively.

To derive a viewpoint relating each chromatic pitch to a referential tonic, chromatic scale degree (or `csd`) maps `cpitch` to `key` and reduces the resulting vector of chromatic scale degrees

[39] Elizabeth Hellmuth Margulis and Andrew P. Beatty, "Musical Style, Psychoaesthetics, and Prospects for Entropy as an Analytic Tool," *Computer Music Journal* 32, no. 4 (2008): 68. Note that while key signatures appear in the notated score, determining the modality necessitates an interpretation on the part of the analyst.

[40] Joshua Albrecht and Daniel Shanahan, "The Use of Large Corpora to Train a New Type of Key-Finding Algorithm: An Improved Treatment of the Minor Mode," *Music Perception* 31, no. 1 (2013): 59–67.

[41] David Temperley and Elizabeth Marvin examined distributional approaches to key finding using a corpus of classical string quartets, and they only included the opening eight measures from each movement to ensure modulations would not affect the results ("Pitch-Class Distribution and the Identification of Key," *Music Perception* 25, no. 3 [2008]: 193–212). Nevertheless, Leigh VanHandel recently noted that in 58 of the 310 movements a modulation still took place within the first eight measures ("The Role of Phrase Location in Key Identification by Pitch Class Distribution," in *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*, ed. Emilios Cambouropoulos et al. [Thessaloniki, Greece: School of Music Studies, Aristotle University of Thessaloniki, 2012], 1069–1073).

modulo 12:

$$\Psi\texttt{csd}(e_i) = (\Psi\texttt{cpitch}(e_i) - \Psi\texttt{key}(e_i)) \mod 12 \tag{3.3}$$

The domain of `csd` consists of twelve distinct symbols numbered from `0` to 11, where `0` denotes the tonic, 7 the dominant, and so on.[42] Events located within the boundaries of a pivot were encoded in both keys. In a movement that modulates to the key of the dominant (e.g., from C to G), for example, the pitch class C appearing within the pivot would receive the values {0,7}, and the pitch class E would receive the values {4,9}.

Metrical Strength. In Gjerdingen's view, a mental representation for a musical category is "likely in no particular key," and "may or may not have a particular meter."[43] For this reason, Gjerdingen generalizes across metric contexts by appealing to the *strength* of each event in the notated meter, which he characterizes as either strong or weak. Whether duple or triple, simple or compound, the internal organization of the meter is therefore largely irrelevant, as strong beats in one context are equivalent to strong beats in any other context. For our purposes, this approach greatly simplifies matters, since we would otherwise be forced to divide the corpus into its various metric conditions (e.g., $\frac{4}{4}$, $\frac{3}{4}$, $\frac{6}{8}$, etc.), examining each in isolation.

Many corpus studies determine the metric context inductively by using statistical procedures like autocorrelation, which classifies meters by finding periodicities in the note onset distribution.[44] By making minimal prior assumptions about the metric organization, this approach has the added benefit of distinguishing the perceptible meter from the notated time signature,

[42] This viewpoint is quite common in corpus studies. See, for example, Margulis and Beatty, "Musical Style, Psychoaesthetics, and Prospects for Entropy as an Analytic Tool," 68; Pearce, "The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition," 71.

[43] Gjerdingen, *Music in the Galant style*, 453.

[44] Judith C. Brown, "Determination of the Meter of Musical Scores by Autocorrelation," *Journal of the Acoustical Society of America* 94, no. 4 (1993): 1953–1957; Petri Toiviainen and Tuomas Eerola, "Autocorrelation in Meter Induction: The Role of Accent Structure," *Journal of the Acoustical Society of America* 119, no. 2 (2006): 1164–1170.

which does not always accurately reflect the real metric organization.[45] Nevertheless, corpus studies of note onset distributions using the notated measure as a rigid framework provide convincing evidence that the time signature serves as a normative template for the representation of temporal information in Western notation. And thus, while note distributions derived from large symbolic corpora sometimes disregard significant variations in metric organization within an individual composition, they also capture large-scale differences between various metric contexts with great precision. To be sure, if the metric organization of music in say, common time, is to be understood as a mental representation acquired by implicit, statistical learning, then metric strength or stability (like tonal strength or stability) should arise out of distributional statistics, and the notated measure provides a useful starting point in this regard. Thus, in what follows, I consider the metric strength of note events in the corpus by first determining their position in the notated measure.

Just as `csd` relates the sequence of chromatic scale degrees in each movement to a particular key, a viewpoint representing metric position relates the sequence of note onset times to a particular time signature. The *MIDIToolbox* provides the basic type `onset`, which refers to the onset time of each event measured in quarter-note beats, and the basic type `timesig` indicates the number of quarter-note beats in the notated measure.[46] Six meters appear in the corpus—$\frac{4}{4}$, cut, $\frac{3}{4}$, $\frac{2}{4}$, $\frac{6}{8}$, and $\frac{3}{8}$—so the domain of `timesig` consists of four values: 1.5, 2, 3, and 4 quarter-note beats. In other words, each of these values represents one or more of the meters from the Haydn Corpus: $\frac{4}{4}$ and cut consist of 4 quarter-note beats, $\frac{3}{4}$ and $\frac{6}{8}$ consist of 3 quarter-note beats, and so on. To determine the metric position of each event in the corpus, the viewpoint `metricpos` reduces `onset` modulo `timesig` and adds 1 to the resulting sequence of metric positions to ensure

---

[45] Grosvenor Cooper and Leonard Meyer, *The Rhythmic Structure of Music* (Chicago: The University of Chicago Press, 1960), 88.

[46] None of the excerpts in the Haydn Corpus feature a change of time signature during the movement.

that the downbeat in each measure = 1.

$$\Psi\texttt{metricpos}(e_i) = ((\Psi\texttt{onset}(e_i) \mod \texttt{timesig}(e_i)) + 1 \tag{3.4}$$

Thus, a note event falling on the 20th quarter note in a common-time movement would receive a value of 1 in `metricpos` because it falls on a downbeat in the notated measure $((20 \mod 4) + 1)$.

To this point I have only offered a method for determining the metric position of each event in the corpus. To determine the *strength* of each event in the notated meter, it might be useful to review a few working definitions for terms relating to metric organization that appear frequently in contemporary theory. If meter refers to nested layers of approximately equally spaced beats or pulses,[47] *metric strength* reflects the coincidence of these layers at multiple levels: the greater the number of layers that align at a given moment, the greater its metric strength.[48] The precise cocktail of features responsible for the perception of meter is not yet known, though metric attending would seem to depend on isochronous patterns of accentuation resulting from changes in harmony, dynamics, rhythmic duration (agogic accents), and register, just to name a few.

Accent is a loaded term in music research. It generally refers to how events in a musical sequence draw attention to themselves;[49] to borrow a well-known expression from Grosvenor Cooper and Leonard Meyer, an accented event is "marked for consciousness."[50] In *A Generative Theory of Tonal Music*, Fred Lerdahl and Ray Jackendoff distinguish between three types of accent: (1) *phenomenal*, which refers to stressed or emphasized events in a continuous sequence; (2) *structural*, which refers to "points of gravity" in a phrase, such as a cadence; and (3)

[47] Harald Krebs, *Fantasy Pieces: Metrical Dissonance in the Music of Robert Schumann* (New York: Oxford University Press, 1999), 22.

[48] Yonatan Malin, *Songs in Motion: Rhythm and Meter in the German Lied* (New York: Oxford University Press, 2010), 39.

[49] Ibid., 41.

[50] Cooper and Meyer, *The Rhythmic Structure of Music*, 8.

metrical, which refers to the strength of certain beats in a given metrical context. In their view, phenomenal accent "functions as a perceptual input to metrical accent," where "moments of stress in the raw signal serve as 'cues' from which the listener attempts to extrapolate a regular pattern of metrical accents."[51] Yonatan Malin would seem to agree, suggesting that phenomenal accents *generate* metrical layers when they recur at regular intervals. And like Lerdahl and Jackendoff, Malin characterizes phenomenal accents according to the usual cast of characters: dynamics, agogic accents, register, harmonic and textural change, the functional beginning of a unit, and so forth.[52]

But perhaps the clearest cue to musical meter lies in the frequency-of-occurrence of note onsets within the notated measure. Caroline Palmer and Carol Krumhansl have noted, for example, that note onset distributions readily conform to theoretical accounts of the metric hierarchy for works in both duple and triple meters.[53] Thus, we might derive a measure of metric strength empirically by examining the distribution of note onsets in the notated measure for each meter in the corpus.[54]

Note events can appear in a range of metric positions within each measure. For movements in common or cut time in the Haydn Corpus, for example, note events appeared in 51 unique

[51]Lerdahl and Jackendoff, *A Generative Theory of Tonal Music*, 17.

[52]Malin, *Songs in Motion: Rhythm and Meter in the German Lied*, 41. Although I have only cited the contributions of twentieth-century theorists, these ideas have a long history in music theory. Lerdahl and Jackendoff's accent types recall Johann Philipp Kirnberger's *Akzenttheorie*, and Cooper and Meyer's off-cited expression that accented events are marked for consciousness is reminiscent of Johann Mattheson's description of events in metrically strong positions as having an "inner content and emphasis." For a review of theories of accent, rhythm, and meter in music theory scholarship, see William E. Caplin, "Theories of Musical Rhythm in the 18th and 19th Centuries," in *The Cambridge History of Western Music Theory*, ed. Thomas Christensen (Cambridge, UK: Cambridge University Press, 2002), 657–694; Justin London, "Rhythm in Twentieth-Century Theory," in *The Cambridge History of Western Music Theory*, ed. Thomas Christensen (Cambridge, UK: Cambridge University Press, 2002), 695–725.

[53]Caroline Palmer and Carol L. Krumhansl, "Mental Representations for Musical Meter," *Journal of Experimental Psychology: Human Perception and Performance* 16, no. 4 (1990): 728–41.

[54]This approach also lends support to the claim that mental representations for various meters result from statistical learning, whereby metrically strong or stable events appear frequently at particular positions on the metric grid.

metric positions. To visualize each meter using a histogram, I have divided the notated measure into equal-sized bins, with the size of each bin corresponding to the duration of a 32nd note. Histograms typically display the number of note events in each bin, but this approach gives equal weight to each note event, regardless of duration. To resolve this issue, several recent corpus studies weighted pitch-class and scale-degree distributions by the rhythmic duration of each event, so I have also adopted that procedure here.[55]

Rather than weight the histogram for each meter by notated durations, which assumes durational equivalence across the corpus regardless of the underlying tempo,[56] I have elected to weight each histogram using Richard Parncutt's model of *durational accent*, which maps the physical inter-onset interval (IOI) between events e_i and e_{i+1} to the phenomenal accent of note event e_i.[57] Following a number of experimental studies linking IOI with metrical accentuation, Parncutt's function assumes that the perceptual or durational accent for a given note event increases with the IOI that follows it. To account for limitations of auditory processing for very short ($<$ 50 ms) and very long ($>$1–2 s) IOIs, he also includes two free parameters, represented by k and τ in the equation below.

$$accent_{dur}(e_i) = \left[1 - exp\left\{\frac{-ioi(e_i)}{\tau}\right\}\right]^k$$

The function $accent_{dur}(e_i)$ denotes the durational accent of event e_i, exp is the natural exponential function, $ioi(e_i)$ refers to the IOI following event e_i, expressed in milliseconds, τ is the saturation duration, which is proportional to the duration of echoic memory, and k is

55 The Krumhansl-Schmuckler key-finding algorithm provides one well-known example (*Cognitive Foundations of Musical Pitch*, 77-110).

56 Snyder, "Entropy as a Measure of Musical Style: The Influence of A Priori Assumptions." To normalize note distributions across movements in different meters and tempi, Snyder also suggested weighting each note as a fraction of the duration of the entire piece (141).

57 Richard Parncutt, "A Perceptual Model of Pulse Salience and Metrical Accent in Musical Rhythms," *Music Perception* 11, no. 4 (1994): 409–464.

the *accent index*, which accounts for the minimum discriminable IOI. Parncutt suggests that parameter values of $k = 2$ and $\tau = 500$ ms provide a good fit to experimental data, so I have retained those values here.[58]

Shown in Figure 3.2, the durational accent increases for small values of IOI and plateaus (or *saturates*) at around 2 s, when the IOI exceeds the duration of echoic memory. The further we move from right to left along the curve (i.e., the shorter the IOI between events), the greater the difference in accent between long and short IOIs. According to Parncutt, the shape of the curve therefore sharpens the difference between long and short events and renders metrical interpretations less ambiguous.[59]

Unfortunately, identifying physical IOIs between events in polyphonic textures is relatively complex. Do we select the IOI between adjacent events within a given instrumental part, or within the entire texture? How do we account for effects of auditory streaming? To simplify matters, I have elected to treat note duration as a rough approximation of inter-onset interval, with the hope that the results obtained here will be replicated in subsequent analyses weighted by IOI.

Using Parncutt's model, I determined the durational salience of the note durations (measured in seconds) in each movement. To demonstrate the effect of weighting the note distributions by durational salience, Figure 3.3 presents bar plots of the distribution of the proportion of note onsets within the notated measure in $\frac{4}{4}$, with the plot on the left weighted by note count, and the plot on the right weighted by summing the durational accents in each metric position. The plot below presents the arithmetic difference between these two distributions, with positive values indicating a greater proportion of note events in the duration-weighted distribution. The plot on the left demonstrates that note onsets appeared prevalently at positions throughout

[58]Ibid., 426–433.
[59]Ibid., 431–432.

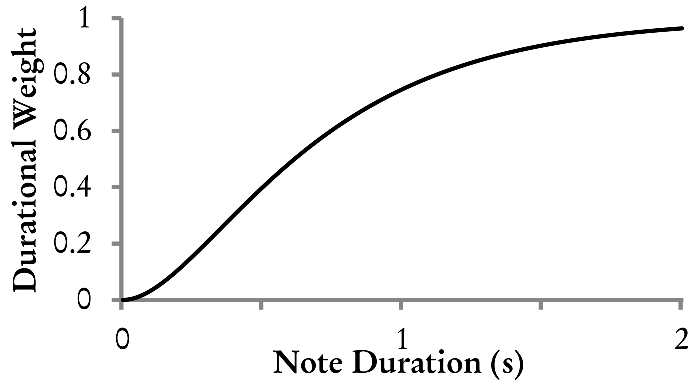


Figure 3.2: Durational weight plotted against note duration in seconds, reproduced from Figure 6 of Parncutt's "A Perceptual Model of Pulse Salience and Metrical Accent in Musical Rhythms" (1994), 431.

the notated measure, but the duration-weighted distribution shown in the plot on the right indicates that longer durations appeared far more frequently in metrically strong positions, such as beats one and three. In fact, weighting the note distribution by durational accent clarified the metrical hierarchy just as Parncutt suggested, with the downbeat receiving the greatest proportion of durations at the level of the measure, followed by beat three at the half-note level, beats two and four at the quarter-note level, and so on.

From visual inspection alone, a four-level viewpoint of metric strength would seem to provide the best fit to the underlying distribution. But to select the optimum number of levels across all of the metric conditions, quantifying the degree of fit in each case, it might be useful to examine the statistical properties of the distribution more closely. If we conceptualize metric strength along a continuous scale, with strong or stable events at one end and weak or unstable events at the other, we might reorder the metric positions within the histogram from most common to least common to reflect this scale. The black line in Figure 3.4 presents the most common sixteen metric positions from the duration-weighted note distribution shown in Figure 3.3.

Formally, the note onset distribution presented in Figure 3.4 is a discrete probability

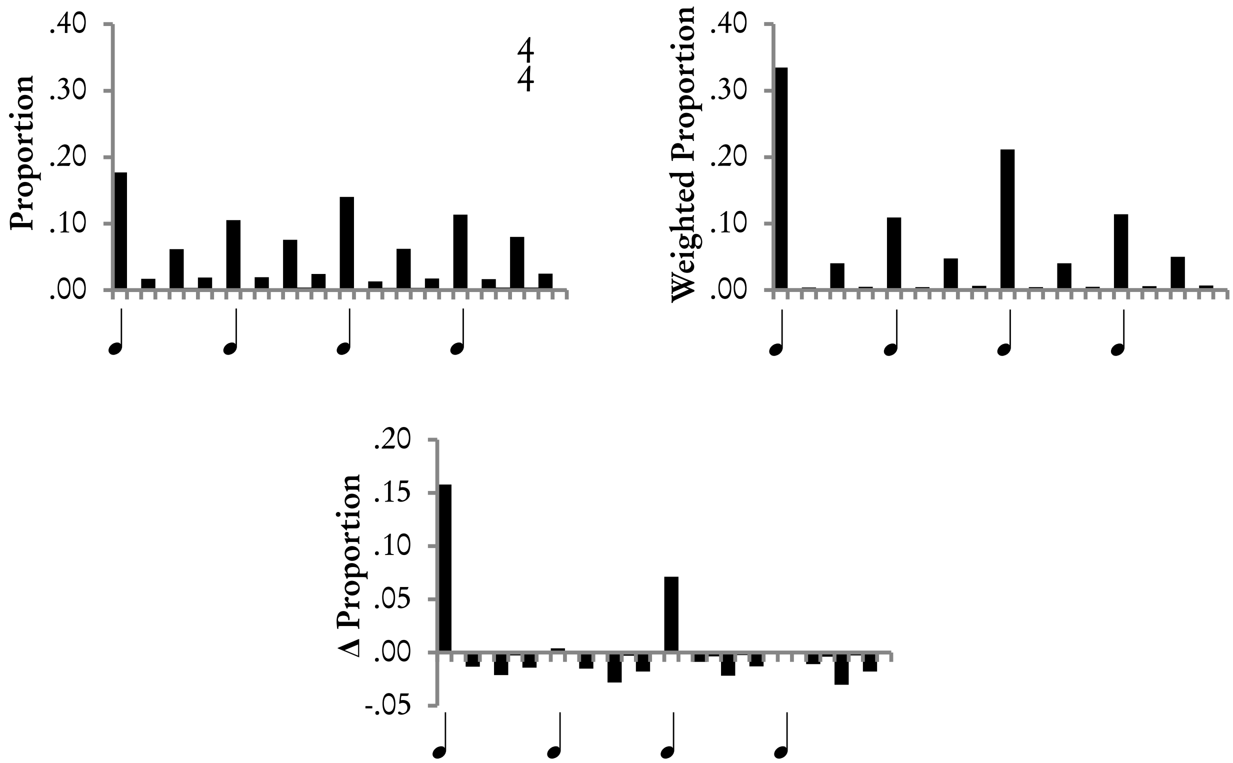


Figure 3.3: Top: Bar plots of the proportion of note onsets from the Haydn Corpus in each metric position within the notated measure in $\frac{4}{4}$, weighted by note count (left), or by summing the durational accents in each metric position (right). Bottom: Bar plot of the difference in the proportion of note onsets.

distribution, and it loosely conforms to the family of power laws used in linguistics to describe the frequency-of-occurrence of words in language corpora.[60] George Zipf noted, for example, that in many language corpora the frequency of any word is inversely proportional to its rank in the corpus.[61] According to Zipf's law, the most frequent word occurs twice as often as the second most frequent word, three times as often as the third most frequent word, and so on. In this case, however, the note onset distribution more closely resemblances an exponential function with base 2, where the most frequent metric position occurs twice as often as the

[60]Manning and Schütze, *Foundations of Statistical Natural Language Processing*, 20–29.
[61]George Kingsley Zipf, *Human Behavior and the Principle of Least Effort* (Oxford, UK: Addison-Wesley, 1949).

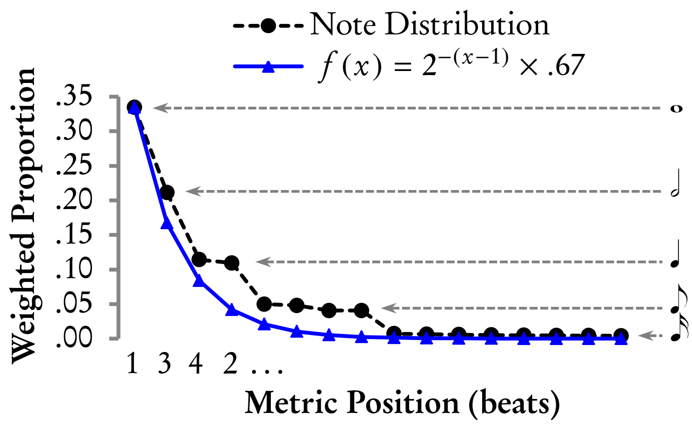


Figure 3.4: Line plot of the duration-weighted proportion of note onsets in $\frac{4}{4}$, ordered from most to least common. The right y-axis annotates the durational level reflected in the note distribution, which appears in dotted black. The exponential distribution (base 2) appears in blue.

second most frequent position and four times as often as the third most frequent position. The blue curve in Figure 3.4 visualizes the exponential function, $2^{-(x-1)} \times .67$, where x represents the rank of each metric position according to its frequency in the distribution, $2^{-(x-1)}$ halves the proportion of note onsets as we ascend in rank (i.e., from left to right); and .67 ensures that the curve passes through the top-ranked position in this particular distribution.[62]

Nevertheless, the exponential function is not a perfect fit. If metric strength is to be conceived as a continuous exponential function, as I previously suggested, we would expect note onsets in the fourth-ranked metric position to appear half as often as those in the third-ranked position, but meter distributions like this one instead demonstrate a staircase effect, where each stair conforms to a different durational level within the metric hierarchy. In this case, the fourth-ranked position (beat two) appears almost as often as the third-ranked position (beat four) because both positions reflect the quarter-note durational level in $\frac{4}{4}$. From left to right, the next four positions in the distribution represent the eighth-note level, the next eight represent

[62]To ensure this equation would fit *any* distribution, we could replace the constant .67 with $2(y_1)$, where y_1 refers to the most frequent (i.e., highest ranked) metric position in the measure.

the sixteenth-note level, and the final sixteen positions represent the thirty-second-note level (not shown).

Given this staircase effect, one possible explanation for the close fit between the note distribution and the exponential curve with base 2 might be that the durational levels of the metric hierarchy in $\frac{4}{4}$ preserve the same 2:1 ratio from top to bottom (i.e., a whole note is double the duration of a half note, a half note is double the duration of a quarter note, etc.). And if a doubling of the durational level corresponds to a doubling of the proportion of note onsets in $\frac{4}{4}$, we could further hypothesize that the ratio represented between adjacent levels of any metric hierarchy should correspond to the ratio between stairs of the corresponding note onset distribution.

Bearing this assumption in mind, I have created a hypothetical note distribution for each meter that preserves the ratio between adjacent levels of the metric hierarchy. Table 3.3 presents the durational levels for each meter, starting with the 32$^{\text{nd}}$ note value. A measure in $\frac{4}{4}$ consists of six such levels, $\frac{3}{4}$, $\frac{2}{4}$, and $\frac{6}{8}$ consist of five levels, and $\frac{3}{8}$ consists of four. To create a hypothetical note distribution, we must first determine the proportion of note onsets for each level of the metric hierarchy. The column denoted by w represents the number of 32$^{\text{nd}}$ notes contained within each durational level. To find the proportion of note onsets associated with each level L_i, we need only divide the weight w at each level by the sum of the weights for that meter:

$$prop(L_i) = \frac{w_i}{\sum w}$$

w preserves the ratios between adjacent levels, and the function $prop$ ensures the resulting values in L sum to 1. In $\frac{4}{4}$, for example, the downbeat receives a durational weight of 32, and the sum of the weights in $\frac{4}{4}$ is 63, so the estimated proportion of note onsets associated with the metric downbeat is .51.

Table 3.3: Levels of the metric hierarchy for each meter.

Metric	$\frac{4}{4}$			$\frac{3}{4}$			$\frac{2}{4}$			$\frac{6}{8}$			$\frac{3}{8}$		
Strength	Dur.	N[a]	w[b]	Dur.	N	w	Dur.	N	w	Dur.	N	w	Dur.	N	w
4	𝅝	1	32	𝅗𝅥.	1	24	𝅗𝅥	1	16	𝅗𝅥.	1	24	♩.	1	12
3	𝅗𝅥	1	16	♩	2	8	♩	1	8	♩.	1	12	♪	2	4
2	♩	2	8	♪	3	4	♪	2	4	♪	4	4	𝅘𝅥𝅯	3	2
1	♪	4	4	𝅘𝅥𝅯	6	2	𝅘𝅥𝅯	4	2	𝅘𝅥𝅯	6	2	𝅘𝅥𝅰	6	1
	𝅘𝅥𝅯	8	2	𝅘𝅥𝅰	12	1	𝅘𝅥𝅰	8	1	𝅘𝅥𝅰	12	1			
	𝅘𝅥𝅰	16	1												

[a] N refers to the number of metric positions within the notated measure for each level.
[b] w refers to the durational weight of each level, measured in 32nd notes.

Unfortunately, these values assume each level consists of just one metric position; within the notated measure, however, the lower levels of the metric hierarchy feature multiple metric positions. As a result, if we were to assign each metric position the appropriate proportion using the equation above, the sum of the resulting values in the distribution would be greater than 1. To adjust these proportions such that they accommodate the number of metric positions N associated with each level, I multiply the values of N by the corresponding proportions in L and sum them, and then divide each L_i by this sum:

$$prop_{adj}(L_i) = \frac{L_i}{\sum(L \times N)}$$

Using the adjusted proportions from the equation above, we can assign a proportion to each metric position according to its membership in the metric hierarchy. I will hereafter refer to this procedure as the *proportions model*. Figure 3.5 presents the bar plots of the duration-weighted note distributions for each movement on the left in blue, with the distributions provided by the proportions model on the right in red. The majority of the movements in the Haydn Corpus were notated in common or cut time, so the top-left distribution represents over 20,000 note events. By comparison, only one movement was notated in $\frac{3}{8}$, so the bottom-left plot represents less than 1,000 note events.

I mentioned previously that Palmer and Krumhansl have already noted the degree to which note onset distributions conform to theoretical accounts of the metric hierarchy for each meter,[63] and the distributions on the left in Figure 3.5 replicate that finding. In fact, the frequency-of-occurrence of notes in the metric positions associated with each durational level corresponds exactly with the levels of metric strength found in Table 3.3 for *every* meter—the metric position associated with the level of the measure received the highest proportion of note onsets, the metric positions associated with the next lower level received the next highest proportions, and so on.

There are a number of statistical procedures for determining the degree of fit between the predicted distributions from the proportions model and the corresponding note distributions from the Haydn Corpus. In this instance, I have elected to describe the relationship using *linear regression*, which calculates a best-fit line that minimizes the error between the predicted estimates and the actual values found in the note distribution. To understand the regression estimates that appear in Figure 3.5, R^2 refers to the fit of the model, where a value of 1 indicates that the model accounts for all of the variance in the outcome variable (i.e., a perfectly linear relationship between the predictor and the outcome), and a value of 0 indicates that the model fails to account for any of the variance.

The distributions from the proportions model provide an excellent fit for the simple meters ($\frac{4}{4}$, $\frac{3}{4}$, and $\frac{2}{4}$), suggesting that the empirical distributions reflect the ratios between adjacent levels of the metric hierarchy. Nevertheless, the fixed 2:1 or 3:1 ratios characterizing each distribution in the proportions model are somewhat variable in the empirical distributions. In $\frac{4}{4}$, the model underestimated the proportion of durations for the most frequent metric positions

[63] Palmer and Krumhansl, "Mental Representations for Musical Meter."

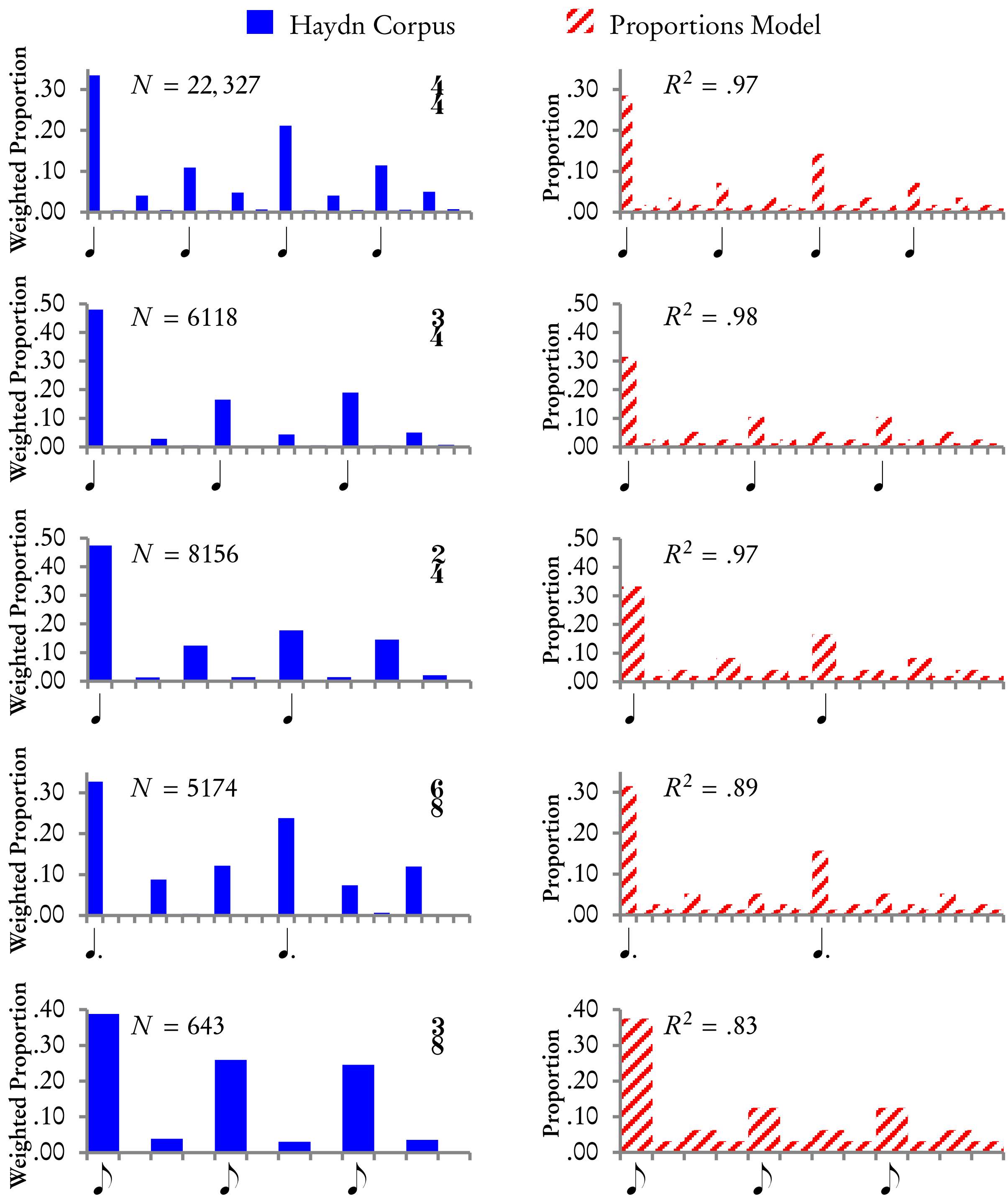


Figure 3.5: Left: Bar plots of the proportion of note onsets weighted by durational accent for movements in $\frac{4}{4}$, $\frac{3}{4}$, $\frac{2}{4}$, $\frac{6}{8}$, and $\frac{3}{8}$ (solid blue). N refers to the number of note onsets represented in each plot. Right: Bar plots of the proportion of note onsets predicted by the proportions model (dashed red). R^2 indicates model fit.

(the downbeat and beat three), but overestimated the proportion of durations for the lowest levels of the hierarchy (the metric positions associated with the levels of the 16th and 32nd note). The same trend emerged in the $\frac{3}{4}$ distributions, where the model underestimated the proportion of durations at the downbeat and overestimated the proportion of durations at the level of the 16th note and lower. Note, however, that the proportions model correctly predicted the 3:1 ratio between the dotted-half-note level and the quarter-note level in the note distribution—in both distributions, the proportion of durations at the downbeat was three times larger than the proportion of durations for metric positions at the quarter-note level.

In $\frac{2}{4}$, the proportions model predicted far smaller proportions at the eighth-note level than the empirical note distribution demonstrated. To be sure, the difference between the levels of the quarter note and eighth note are surprisingly small. If we were to derive a viewpoint of metric strength strictly from empirical observation, it would seem reasonable to assign the same value to the metric positions from these two durational levels. For the sake of consistency across all of the meters, however, I have elected to retain the theoretical metric hierarchy in $\frac{2}{4}$.

For the compound meters, the predicted models were less successful. In $\frac{6}{8}$, the second dotted quarter received a greater proportion of durations than the model predicted, just as we noted for the corresponding level from the $\frac{4}{4}$ distribution. The proportions model for $\frac{3}{8}$ suffered from the same limitation; in this case, the predicted 3:1 ratio between the levels of the dotted quarter and eighth note did not correspond with the ratio in the empirical distribution, where proportions for metric positions at the eighth-note level were much larger. In fact, the ratio between these two levels was less than 2:1 in the empirical distribution. Nevertheless, the small sample size for the $\frac{3}{8}$ distribution calls into question any similarities or differences we might observe between the model and the distribution.

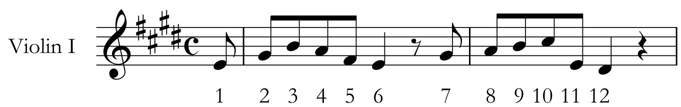
Given the degree of fit between the empirical note distributions and the corresponding proportions models, a viewpoint of metric strength reflecting the coincidence of layers of nested periodicities appears well justified. Metric strength in this context is an ordinal viewpoint, where the value associated with each metric position represents the number of periodicities (or layers) it supports within the notated meter. Since meter minimally involves two or three layers,[64] a viewpoint representing four levels of metric strength should be sensitive enough to allow us to make subtler distinctions than a two- or three-level viewpoint would permit, while still being coarse enough to generalize across all five metric contexts. Shown in Table 3.3, level 4 represents the metric downbeat; level 3 represents the duple or triple subdivision of the measure; level 2 represents the next subdivision, which typically corresponds to the level of the quarter or eighth note; and level 1 represents the lowest level of strength, which consists of all of the remaining metric positions across the measure.

3.4.2 Chord Events

To this point I have represented note events in the Haydn Corpus according to four viewpoints: melodic interval (`melint`), melodic contour (`contour`), chromatic scale degree (`csd`), and metric strength (`strength`). Figure 3.6 presents the viewpoint representation for the first violin part from the opening two measures of the first movement of Haydn's String Quartet in E, Op. 17/1. The appeal of this approach is that it represents each part in isolation, allowing us to consider the distinct roles these instrumental parts may play, both in the corpus at large, and in cadential contexts. Nevertheless, by treating the note event as the unit of analysis, this representation scheme can tell us nothing about the manifold ways in which these parts may interact. In short, it tells us nothing about the vertical sonorities that characterize this style.

Identifying events beyond the level of the note using inductive methods is a tremendous challenge. As a consequence, many analysts have elected to ignore the symbolic encoding entirely and instead annotate vertical sonorities using roman numerals, figured bass symbols,

[64] Lerdahl and Jackendoff, *A Generative Theory of Tonal Music*, 19; London, *Hearing in Time*, 47.



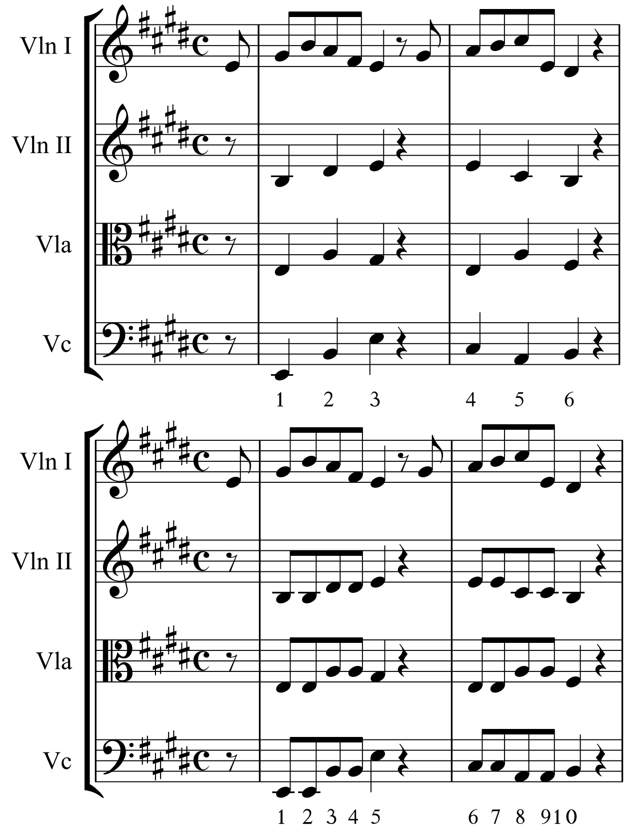
	1	2	3	4	5	6	7	8	9	10	11	12
`melint`	⊥	4	3	-2	-3	-2	4	1	2	2	-9	-1
`contour`	⊥	1	1	-1	-1	-1	1	1	1	1	-1	-1
`strength`	1	4	1	2	1	3	1	4	1	2	1	3
`csd`	0	4	7	5	2	0	4	5	7	9	0	11

Figure 3.6: Top: First violin part from Haydn's String Quartet in E, Op. 17/1, i, mm. 1–2. Bottom: Viewpoint representation.

and the like. But in recent decades, several studies have attempted to derive principles of tonal harmony from symbolic corpora by constructing composite viewpoints of chord events from simpler viewpoints derived from note events.

The simplest procedure for deriving chord events from multi-voiced textures is to partition note events into simultaneities whenever all of the instrumental parts have the same onset time. Shown at the top of Example 3.1, all four instrumental parts feature the same onset time on six occasions in the first two measures. This approach works well for homo-rhythmic textures and simple species counterpoint, but it under-partitions more complex polyphony where common onset times may not coincide with the harmonic rhythm of the passage.[65] For this reason, many current music analysis software frameworks perform a *full expansion* of the symbolic encoding, which duplicates overlapping note events at every unique onset time.[66] Shown below in the same example, expansion results in the identification of ten unique onset times for which all four instrumental parts are present. With this method, we could model the resulting sequence of note combinations directly, or sample at regular metric intervals using what Conklin calls

[65] Darrell Conklin, "Representation and Discovery of Vertical Patterns in Music," in *Music and Artifical Intelligence: Proc. ICMAI 2002*, ed. Christina Anagnostopoulou, Miguel Ferrand, and Alan Smaill, vol. 2445 (Springer-Verlag, 2002), 3–4.

[66] ibid., 4. In *Humdrum*, this technique is called *ditto*, while *Music21* calls it *chordifying*.

Example 3.1: Top: Haydn, String Quartet in E, Op. 17/1, i, mm. 1–2. Bottom: Full expansion.

threaded viewpoints.

In previous publications, Conklin represented vertical sonorities by the melodic intervals between adjacent onset times in each instrumental part. Beginning with the cello part in Example 3.1, Conklin's method derives the vertical pattern $\langle 0, 0, 0, 3\rangle$ between events 1 in 2. Ian Quinn and Panayotis Mavromatis have pointed out, however, that this approach hard-codes the ordering of the four parts when in principle, voices are permutable.[67] Swapping the alto and tenor voices in a vertical sonority would produce a completely different representation in Conklin's method, for example. As an alternative, Quinn developed a representation consisting

[67]Ian Quinn and Panayotis Mavromatis, "Voice-Leading Prototypes and Harmonic Function in Two Chorale Corpora," in *Mathematics and Computation in Music*, ed. Carlos Agon et al. (Heidelberg: Springer, 2011), 231.

of an ordered triple (S_1, S_2, I), where S_1 and S_2 are sets of intervals above the bass in semitones modulo the octave, and I is the melodic interval (again modulo the octave) from the first bass note to the second. He calls this representation a *voice-leading type*.[68]

The appeal of Quinn's representation is that the most common voice-leading types in a given corpus will have analogues in figured-bass nomenclature. Nevertheless, as a representation scheme for chord events, Quinn's voice-leading type is more promiscuous than traditional definitions of 'chord' would embrace. Whereas theorists tend to assign chordal status only to those vertical sonorities featuring stacked intervals of a third, Quinn's voice-leading types make no distinction between chord tones and non-chord tones, consonant and dissonant intervals, or diatonic and chromatic scale degrees.[69] As a result, the syntactic domain (or alphabet) of voice-leading types is enormous. Thus, to adapt Quinn's method here, we need to reduce the syntactic domain such that the resulting viewpoint corresponds more closely with the figured bass symbols in Gjerdingen's schema-theoretic framework.[70]

The viewpoint *vertical interval class combination* (`vintcc`) models the vertical intervals in semitones modulo 12 between the lowest instrumental part b and the upper parts u from the basic type `cpitch`:

$$\Psi\texttt{vintcc}(e_{i_b}) = |\Psi\texttt{cpitch}(e_{i_b}) - \Psi\texttt{cpitch}(e_{i_u})| \mod 12 \tag{3.5}$$

[68]Ian Quinn, "Are Pitch-Class Profiles Really Key for Key," *Zeitschrift der Gesellschaft der Musiktheorie* 7 (2010): 151–163; Quinn and Mavromatis, "Voice-Leading Prototypes and Harmonic Function in Two Chorale Corpora." The ELVIS team at McGill University use the same method. For more details, see `http://elvisproject.ca/`.

[69]Quinn, "Are Pitch-Class Profiles Really Key for Key," 152.

[70]Ideally, we would reduce the syntactic domain to less than, say, 30 symbols, but given the number of combinatorial possibilities for three- and four-note chords, such a feat is staggeringly difficult to achieve. At present, the creation of an alphabet-reduction algorithm that identifies such a reduced set of chord classes is beyond the scope of this dissertation, but see Christopher W. White, "Some Statistical Properties of Tonality, 1650-1900" (PhD Dissertation, Yale University, 2013); Christopher W. White, "An Alphabet-Reduction Algorithm for Chordal n-Grams," in *Proceedings of the 4th International Conference on Mathematics and Computation in Music* (Springer, 2013), 201–212.

If we only consider unique onsets that contain all four instrumental parts, the number of combinatorial possibilities is 12^3 (or 1728), but this procedure excludes combinations containing only one or two vertical interval classes.[71] By including unique onsets for combinations containing two, three, or four instrumental parts, the number of combinatorial possibilities increases to $13^3 - 1$ (or 2196), since the syntactic domain of each vertical interval class is $\{0, 1, 2, \ldots, \perp\}$.[72]

To reduce the syntactic domain of `vintcc` to a more reasonable number while retaining those combinations that approximate figured bass symbols, I have excluded note events in the upper parts that double the lowest instrumental part at the unison or octave, allowed permutations between vertical intervals, and excluded interval repetitions. Following Quinn, the assumption here is that both the precise location and repeated appearance of a given interval in the instrumental texture are inconsequential to the identity of the combination. Thus, by allowing permutations and excluding voice doublings of the lowest instrumental part, the major triads $\langle 4, 7, 0\rangle$ and $\langle 7, 4, 0\rangle$ would reduce to $\langle 4, 7, \perp\rangle$. Similarly, by eliminating repetitions, the chords $\langle 4, 4, 10\rangle$ and $\langle 4, 10, 10\rangle$ would reduce to $\langle 4, 10, \perp\rangle$. Using this procedure, the potential domain of `vintcc` reduces dramatically from 2196 to 232 unique vertical interval combinations, though the Haydn Corpus only contained 190 of the 232 possible combinations, reducing the domain yet further.

Unfortunately, `vintcc` does not represent voice-leading information, nor does it define each harmony in relation to an underlying tonic. Quinn's solution to the first limitation was to encode the melodic interval between successive events in the bass. Given the viewpoint `csd`, however, we may instead represent vertical sonorities as combinations of chromatic scale degrees.

---

[71] I did not include onsets consisting of just one instrumental part under the assumption that such instances would not constitute chord events.

[72] I excluded the combination representing just one instrumental part from the calculation, $\langle\perp, \perp, \perp\rangle$. Again, $\perp$ indicates that the vertical interval class is undefined.

The viewpoint csdc includes the chromatic scale degrees derived from csd as combinations of two, three, or four instrumental parts. Here, the number of possibilities increases exponentially to $13^4 - 13^1$ (or $28,548$), since the cello part is now encoded explicitly in combinations containing all four parts.[73] Rather than treating permutable combinations as equivalent (e.g., $\langle 0, 4, 7, \perp\rangle$ and $\langle 4, 7, 0, \perp\rangle$), as we did in vintcc, it will also be useful to retain the chromatic scale degree in the lowest instrumental part in csdc and only permit permutations in the upper parts. Excluding voice doublings and permitting permutations in the upper parts reduces the potential domain of csdc to 2784, though in the Haydn Corpus the domain reduced yet further to 688 distinct combinations.

§3.5 Conclusions

This chapter took a circuitous path through the Haydn Corpus. I began in §3.1 with a brief discussion of corpus studies in music research, and then in §3.2 presented the corpus of expositions from Haydn's string quartets and the representation scheme employed by the *MIDIToolbox*. Digital encodings of individual note or chord events typically represent multiple properties of the musical surface, so in §3.3 I adopted a multiple-viewpoint framework to encode irreducible viewpoints like chromatic pitch (cpitch), note onset in beats (onset), metric position (metricpos), key (key), and mode (mode). From these basic types I then derived a number of other viewpoints in §3.4 to represent the "core" events of the classical cadence: melodic interval (melint), contour (contour), chromatic scale degree (csd), and strength (strength) to represent note events, and vertical interval class combination (vintcc) and chromatic scale degree combination (csdc) to represent chord events.

Armed with this representation scheme, I now apply a few common statistical methods to

[73] As with vintcc, I excluded combinations representing less than two instrumental parts from the calculation (e.g., $\langle\perp, \perp, \perp, \perp\rangle$).

probe the corpus at large and consider whether cadences and other closing formulæ are indeed the most recurrent patterns in classical music. Thus, Chapter 4 attempts to reinforce the link between psychological stability and statistical frequency, providing distributional evidence in support of the view that cadences are among the most important event schemas in the tonal system.

Chapter 4

Discovering Closing Schemas: Stability and Voice

We perceive, understand, and respond to the world, including music, in terms of the patterns and models, concepts and classifications, which have been estabilished in our traditions—linguistic, philosophical, musical, and so on.

LEONARD B. MEYER

For many, the cadence concept is woven into the very fabric of the tonal system. According to Giorgio Sanguinetti, it represents "the first, most elementary of tonal structures,"[1] providing a flexible scaffold on which to build increasingly complex diminutions spanning phrases, sections, and entire pieces.[2] If the cadence is indeed the quintessential tonal schema, a "microcosm

[1]Giorgio Sanguinetti, *The Art of Partimento: History, Theory, and Practice* (Oxford, UK: Oxford University Press, 2012), 105.

[2]Few would disagree that cadences effect closure at the level of the phrase or theme (i.e., at fore-to-middle-ground levels of musical organization), but whether local cadences effect closure at higher levels of formal organization, or indeed, that the materials found within an entire work could *constitute* a high level cadence, remains very much in dispute. For a summary and critique of the role of cadential closure in the structural hierarchy of a work, see Caplin, "The Classical Cadence," 56-66, and Michael Spitzer, "The Retransition as Sign: Listener-Oriented Approaches to Tonal Closure in Haydn's Sonata-form Movements," *Journal of the Royal Musical Association* 121, no. 1 (1996): 11–45.

which summarizes the essential features... of the work it closes,"[3] we should expect to see those features reflected in global characteristics of the corpus at large.

It seems from existing corpus studies that the features on which the cadence concept depends—prominent scale degrees, sonorities, metric positions, and the like—appear with far greater frequency in existing corpora than those deemed unstable, suggesting that stable events are resistant to change, continuation, or further implication because they appear frequently in a given style. Nevertheless, the compositional framework that supports the classical cadence—and indeed, many of the cadential formulæ spanning the entire European art tradition from its inception—consists of at least two independent voices serving quite distinct roles, an observation that has yet to receive much attention in corpus research.[4] Thus, in §4.1 I provide distributional evidence for the tonal and metric stability of note and chord events appearing at the moment of cadential arrival and examine the statistical characteristics that distinguish each voice of the two-voice framework, setting aside the discovery of entire cadential patterns until §4.2.

§4.1 Stability and Voice

4.1.1 Note Events

I have represented note events in the Haydn Corpus according to four viewpoints: chromatic scale degree (`csd`), metric strength (`strength`), melodic contour (`contour`), and melodic interval (`melint`). Also, recall from Chapter 3 that to derive `csd` I annotated the key, mode, modulations, and pivot boundaries for each note event in every movement of the corpus. Figure 4.1 presents the major and minor chromatic scale degree distributions weighted by Parncutt's coefficient of durational salience for each instrumental part, and omitting note events occurring within

[3] Alfredo Casella, *The Evolution of Music throughout the History of the Perfect Cadence* (London, UK: Chester, 1924), iii.

[4] Gjerdingen, "'Historically Informed' Corpus Studies," 194.

the boundaries of a modulatory pivot.[5] As expected, both sets of distributions replicated the goodness-of-fit profiles first published in Carol Krumhansl and Edward Kessler's seminal 1982 study.[6] For the major and minor distributions from the upper parts, $\hat{1}$ received the highest proportion of durations, followed by $\hat{5}$, $\hat{3}$, the other diatonic scale degrees, and finally the remaining (chromatic) scale degrees.[7] A theory of implicit statistical learning would predict that cadences featuring scale degrees like $\hat{1}$, $\hat{5}$, and $\hat{3}$ at the expected moment of cadential arrival should be easier to learn and remember because those scale degrees appear far more frequently in the classical style. Thus, it should not be surprising that the (genuine) cadence categories in Caplin's typology—and in nearly every theory of cadence—feature precisely these scale degrees (see Table 2.1).

The distributions in Figure 4.1 also illustrate an important difference between the four instrumental parts: for the major mode distributions, $\hat{1}$ and $\hat{5}$ appeared more frequently in the cello part than in the upper parts, whereas diatonic scale degrees like $\hat{2}$, $\hat{3}$, and $\hat{6}$ appeared less frequently relative to the upper parts. In fact, the cello part presented a different tonal hierarchy altogether in both the major and minor distributions, with $\hat{1}$, $\hat{5}$, and $\hat{4}$ representing the most stable scale degrees, and not $\hat{1}$, $\hat{5}$, and $\hat{3}$, as was noted for the distributions from the upper parts. One explanation for this finding might be that the upper-part distributions reflect the statistical behavior of higher voices in a multi-voiced texture, where stepwise motion—particularly around the most stable scale degrees of the tonal system—is a hallmark of melodic organization. In the two-voice framework characterizing the "classic texture,"[8] however, the cello part exemplifies the contrapuntal organization of a lower voice, where scale degrees like $\hat{1}$, $\hat{5}$, and $\hat{4}$ (and not $\hat{3}$) harmonize particularly well with the distribution of scale degrees appearing in the upper parts.

[5]In other words, I have excluded all note events that imply more than one key.

[6]Krumhansl and Kessler, "Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys," 343.

[7]I review this and other experimental studies examining the perception of closure in §7.1.

[8]Ratner, *Classic Music*, 108.

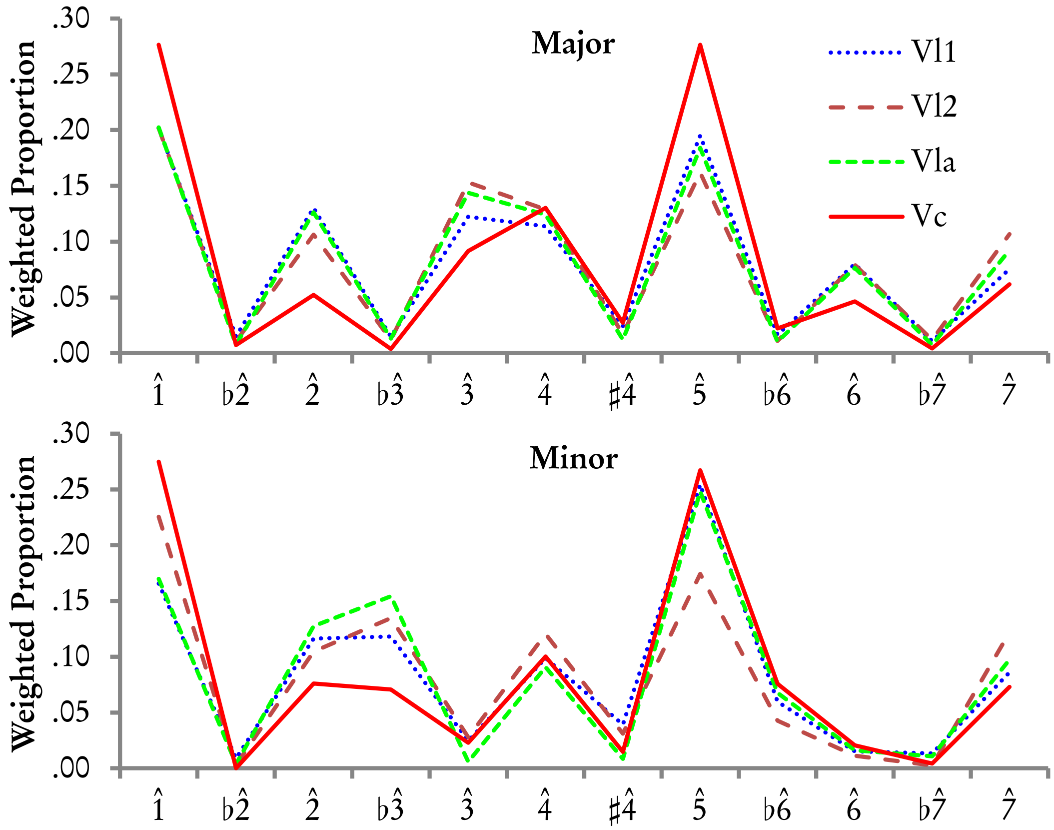


Figure 4.1: Major (top) and minor (bottom) chromatic scale degree distributions weighted by durational accent for each instrumental part.

I took care in the preceding discussion not to conflate each instrumental part with a specific voice in the musical texture (e.g., the first violin part with the soprano voice). In the classical style, for example, the cadential bass sometimes appears in an inner part when the cello part is absent. For my purposes, the generic term *voice* will exemplify what psychologist Albert Bregman has called an *auditory stream*.[9] As opposed to concepts like line, part, or voice, Bregman's "stream" distills into one term the various sensory and cognitive processes by which listeners parse the complex auditory scene into its constituent sound sources (e.g., the viola part from a Haydn string quartet), integrating those sources that display similar sequential grouping

[9]Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, 9.

cues, and segregating those sources that display dissimilar grouping cues.[10] In the classical string quartet texture, listeners may perceive many auditory streams, or as is far more likely, just one or two streams, with two or more instrumental parts presenting acoustic cues so similar as to integrate the individual sound sources into just one stream.[11]

In music-theoretical discourse, the various modifiers associated with the term *voice*—soprano, bass, discant, tenor, and so on—attempt to convey the approximate ambitus and distinct functional characteristics associated with a given line or part, often bearing in mind the presence of other voices (and their functional characteristics) in the complete musical texture. Yet at present, few corpus studies consider how the distributions characterizing these individual voices might differ. To provide a descriptive measure of the differences we might find between instrumental parts in the Haydn Corpus, I will apply a statistic from information theory that mathematician Claude Shannon called *entropy*,[12] but which is sometimes also called *Shannon entropy* to distinguish it from other usages of the term in physics and astronomy.

A number of music-analytic studies have employed Shannon entropy as a descriptive statistic for musical style,[13] and they often note the potential for information-theoretic tools to address questions in music cognition more generally.[14] Ben Duane's attempt to distinguish between

[10]Event formation by auditory fusion results from concurrent grouping cues relating to the temporal onset synchrony, spectral harmonicity, and coherent frequency and amplitude behavior of the various acoustic components characterizing the auditory scene. Auditory stream integration results from sequential grouping cues relating to the spectral, intensity, and spatial continuity of auditory events separated in time. For a review of principles of auditory scene analysis, including a description of the various cues characterizing event formation and auditory stream integration, see McAdams and Drake, "Auditory Perception and Cognition."

[11]Bregman himself was well aware that the rules governing Western contrapuntal practice correspond in many instances with principles of *auditory scene analysis* (James K. Wright and Albert S. Bregman, "Auditory Stream Segregation and the Control of Dissonance in Polyphonic Music," *Contemporary Music Review* 2 [1987]: 63–92), and David Huron conducted a large-scale corpus study to compare the two domains ("Tone and Voice").

[12]Shannon, "A Mathematical Theory of Communication."

[13]See, for example, Leon Knopoff and William Hutchinson, "Information Theory for Musical Continua," *Journal of Music Theory* 25, no. 1 (1981): 17–44.

[14]Meyer, "Meaning in Music and Information Theory"; Joseph Youngblood, "Style as Information," *Journal of Music Theory* 2, no. 1 (1958): 24–35; Leon Knopoff and William Hutchinson, "Entropy as a Measure of Style: The Influence of Sample Length," *Journal of Music Theory* 27, no. 1 (1983): 75–97; Dean Simonton, "Melodic Structure and Note Transition Probabilities: A Content Analysis of 15,618 Classical Themes," *Psychology of Music* 12, no. 3

individual lines in a complex texture using context models provides one recent example.[15] By taking an information-theoretic approach, Duane successfully differentiated melodies and countermelodies from accompanying lines in Haydn's string quartets using a measure of Shannon entropy, leading him to suggest that the "leading status" of a given line within the texture derives from a listener's natural sensitivity to the probabilistic structure of musical lines.[16]

Denoted by H, Shannon entropy represents the degree of choice or uncertainty involved in selecting a symbol from a particular message, where the message consists of a finite set of discrete symbols—a dictionary of English words, a database of nucleic acid sequences, or in our case, a corpus of Haydn string quartets.[17] The more choices available for selection, the more uncertain the eventual outcome, and thus, the higher the value of H. On the other hand, when restricted to just one choice, H is `0`, since the outcome will always be certain.

Formally, Shannon represented H as the number of *bits* (or binary digits) of information associated with a particular message, with bits expressed on a logarithmic scale with base 2. In the multiple-viewpoint framework employed here, if every distinct symbol associated with a given viewpoint is equally likely, H is maximum and equal to $\log_2(N)$, where N represents the number of symbols in the syntactic domain (or alphabet) of that viewpoint. Thus, a viewpoint whose syntactic domain consists of just two symbols represents 1 bit of information (e.g., `mode`

(1984): 3–16; Dean Simonton, "Computer Content Analysis of Melodic Structure: Classical Composers and Their Compositions," *Psychology of Music* 22, no. 1 (1994): 31–43; Margulis and Beatty, "Musical Style, Psychoaesthetics, and Prospects for Entropy as an Analytic Tool."

[15] Ben Duane, "Texture in Eighteenth- and Early Nineteenth-Century String-Quartet Expositions" (PhD Dissertation, Northwestern University, 2012); Ben Duane, "Agency and Information Content in Eighteenth- and Early Nineteenth-Century String-Quartet Expositions," *Journal of Music Theory* 56, no. 1 (2012): 87–120.

[16] Duane, "Agency and Information Content in Eighteenth- and Early Nineteenth-Century String-Quartet Expositions," 93.

[17] In the communication system proposed by Shannon, Joseph Youngblood likened the degree of choice or uncertainty to the roles played by composer and listener: "In information theory, information refers to the freedom of choice which a composer has in working with his materials or the degree of uncertainty which a listener feels in responding to the results of a composer's tonal choices" ("Style as Information," 25).

= {0,1}), a viewpoint consisting of four distinct symbols represents 2 bits (e.g., `strength` = {1,2,3,4}), and a viewpoint consisting of eight distinct symbols represents 3 bits (though no such viewpoint exists). In the case of the chromatic scale degree distributions presented in Figure 4.1, the syntactic domain of `csd` consists of twelve distinct symbols in the Haydn Corpus, so the Shannon entropy associated with `csd` is 3.59 bits (i.e., $\log_2(12) = 3.585$).

For the above examples, it can be helpful to think of each bit as a fair coin flip. One flip produces two possible outcomes: H (for heads) and T (for tails); two flips produce four possible outcomes: HH, HT, TH, TT; three flips produce eight possible outcomes: HHH, HHT, HTH, HTT, THH, THT, TTH, TTT; and so on. The more symbols in the syntactic domain of a given viewpoint, the more choice or uncertainty. But note here that to think of each bit as a coin flip also assumes that each outcome is equally likely, hence the mathematical property that H is maximum and equal to $\log_2(N)$ when the probabilities associated with each symbol in the syntactic domain are equal. I will refer to this value as H_{max}.

Returning to `csd`, the Shannon entropy of 3.59 bits assumes that the proportion of durations calculated for each chromatic scale degree is equal. In the context of the Haydn Corpus, however, the distributions in Figure 4.1 demonstrate just how much more frequently we find $\hat{1}$ than, say, $\sharp\hat{4}$. To calculate the Shannon entropy for a viewpoint whose symbols are not equally likely, the original H should represent the weighted sum of the individual values of H for each symbol in that viewpoint:

$$H = -\sum p_i \log_2 p_i, \tag{4.1}$$

where p_i represents the probability of symbol i, and the sum of the probabilities for all distinct symbols is 1. Thus, in Equation 4.1, if the probability of a given chromatic scale degree in `csd` is 1, the probabilities for all the remaining chromatic scale degrees will be 0, and $H = 0$ (i.e., maximum certainty). If all of the chromatic scale degrees are equally likely, however, the equation reduces to $\log_2(N)$.

In fact, we can infer the relative values of H for each instrumental part just by looking at the distributions in Figure 4.1. Given what we know about Shannon entropy, distributions featuring the majority of durations in just a few chromatic scale degrees (like $\hat{1}$ and $\hat{5}$) will produce lower values of H, whereas relatively flat distributions—where the durations spread evenly among the chromatic scale degrees—will produce higher values of H. In both the major and minor distributions, the durations concentrate more in $\hat{1}$ and $\hat{5}$ in the cello part than they do in the upper parts, so we can infer that the degree of choice or uncertainty—and thus, the value of H—will be lower for the cello part.

Table 4.1 presents the descriptive statistics for each instrumental part. The major mode distribution for the first violin presented in Figure 4.1 consists of 12,964 note events, which represents approximately 89% of the total number of note events in the corpus. Another 7% appear in the minor mode, with the remaining 4% occurring within the boundaries of a pivot (not shown).[18] As expected, Shannon entropy was lowest for the cello part ($H = 2.78$), followed by the viola ($H = 3.00$), the second violin ($H = 3.06$), and the first violin ($H = 3.09$). To relate the absolute entropies to the maximum entropy for that viewpoint, I have also included the *relative entropy*, denoted by H_{rel}, which represents the ratio of the Shannon entropy for a given distribution to the maximum entropy (i.e., $H_{\text{rel}} = \frac{H}{H_{\text{max}}}$). This statistic is particularly useful when comparing distributions across viewpoints whose syntactic domains differ in size.[19]

In statistical parlance, the various measures of H described here represent point estimates of unknown population parameters.[20] In other words, the first value of H reported in Table 4.1 (3.09) is a descriptive statistic drawn from a particular sample—the major mode chromatic scale-

[18] Again, these note events were excluded from the analysis.

[19] For this reason, H_{rel} is also sometimes called *normalized entropy* (Margulis and Beatty, "Musical Style, Psychoaesthetics, and Prospects for Entropy as an Analytic Tool," 69).

[20] George A. Ferguson and Yoshio Takane, *Statistical Analysis in Psychology and Education*, Sixth (Toronto, ON: McGraw-Hill, 2005), 166-167.

Table 4.1: Statistics for the major and minor chromatic scale-degree distributions weighted by durational accent for each instrumental part.

	N[a]	%[b]	H[c]	CI[d]	H_{rel}[e]	r_{KK}[f]	Vln 1	Vln 2	Vla	Vc
Major										
Vln 1	12,964	89.37	3.09	3.07–3.11	0.86	.95	–	.96	.99	.92
Vln 2	9362	87.88	3.06	3.03–3.09	0.85	.94		–	.99	.86
Vla	8125	88.74	3.00	2.98–3.03	0.84	.95			–	.89
Vc	7466	88.22	2.78	2.75–2.81	0.78	.94				–
Total	37,917	88.64	3.02	3.01–3.03	0.84	.97				
Minor										
Vln 1	1038	7.16	3.08	3.01–3.18	0.80	.74	–	.89	.98	.92
Vln 2	829	7.78	3.01	2.94–3.10	0.78	.84		–	.91	.90
Vla	672	7.34	2.93	2.86–3.02	0.76	.80			–	.88
Vc	687	8.12	2.83	2.72–2.96	0.73	.80				–
Total	3226	7.54	3.01	2.97–3.06	0.78	.82				

[a] N refers to the number of note events.
[b] % refers to the percentage of note events that did not occur within a pivot.
[c] H refers to the Shannon entropy.
[d] CI refers to the 95% bootstrap confidence interval of H using the bias-corrected and accelerated percentile method with 1000 replicates (see text and footnote 25).
[e] H_{rel} refers to the Shannon relative entropy.
[f] r_{KK} refers to the Pearson correlation coefficient calculated between the empirical distribution and the corresponding Krumhansl-Kessler goodness-of-fit profile.

degree distribution from the first violin part—that estimates the unknown (and unknowable) value of H if we could sample from the "infinite pool" of chromatic scale degrees represented in the first violin parts of Haydn's string quartets, the entire classical string quartet repertoire, or indeed, all tonal music.[21] In order to draw inferences about the potential differences between estimates of H for each instrumental part, statisticians typically calculate an interval estimate to assert with some known degree of confidence that the population statistic falls within the estimated interval, called a *confidence interval* (CI).

[21] What is meant by "population" is itself an intriguing question. Elizabeth Margulis asks, for example, "do listeners evaluate Haydn's choices in light of the choices he made in all his string quartets..., or in all his works, or in all 18th-century music, or in all tonal music?" ("Musical Style, Psychoaesthetics, and Prospects for Entropy as an

In hypothesis testing theory, confidence intervals provide an interval estimate for a specific population statistic: the arithmetic mean. But since our interest here is in Shannon entropy, the equation for obtaining the standard 95% confidence interval around a sample mean would not be appropriate to the task.[22] In such instances, statisticians often employ an alternative method called the *bootstrap*,[23] which determines the accuracy of point estimates like the arithmetic mean or Shannon entropy when the underlying theoretical distribution is unknown.[24]

To determine the confidence interval for statistics other than the mean, the bootstrap method randomly selects a data point for inclusion in the bootstrap sample and then replaces it in the original distribution, repeating this procedure n times to produce a sample of the same size as the original distribution (this procedure is called *sampling with replacement*). The algorithm then creates a large number of samples (typically 1000), calculating for each sample a replicate of the test statistic, in our case H. Given an appropriate confidence level (90%, 95%, and 99% are the most typical), the algorithm finally estimates the upper and lower boundaries of the confidence interval from the distribution of H replicates. In the case of the 95% interval,

Analytic Tool," 77). From a psychological perspective, the "infinite pool" described by Leon Knopoff and William Hutchinson should represent the long-term stylistic knowledge of listeners regarding the functional characteristics of the first violin part ("Entropy as a Measure of Style: The Influence of Sample Length," 81). Determining the relevant sample (or population) by which to examine that knowledge is an incredibly difficult task, however. See, for example, Justin London, "Building a Representative Corpus of Classical Music," *Music Perception* 31, no. 1 (2013): 68–90.

[22] In fact, Knopoff and Hutchinson provide the necessary equations to obtain the standard 95% confidence interval for Shannon entropy ("Entropy as a Measure of Style: The Influence of Sample Length," 94), but like the confidence interval calculated for the arithmetic mean, the equations assume a normal distribution. What is more, the authors note that when using standard methods to estimate confidence intervals, the sample should consist of at least 7900 characters (Ibid., 83). The bootstrap method I describe in what follows does not assume normality, and Thomas DiCiccio and Bradley Efron note that bootstrap methods are generally more robust than the standard interval estimates for small sample sizes ("Bootstrap Confidence Intervals," *Statistical Science* 11, no. 3 [1996]: 189–228).

[23] Bradley Efron and Robert J. Tibshirani, *An Introduction to the Bootstrap* (London, UK: Chapman & Hall, 1993).

[24] Standard confidence intervals depend on the assumption that the arithmetic mean will be approximately normally distributed, which statisticians call the *central limit theorem* (CLT). In short, the CLT captures the idea that if we could draw a large number of samples from the (infinite) population, calculating a mean for each sample, the hypothetical distribution of sample means would fit a normal (bell-shaped) curve. Since the bootstrap method generates this distribution empirically, it need not rely on the normality assumption.

for example, 97.5% and 2.5% represent the upper and lower boundaries, respectively.[25]

Table 4.1 presents the bootstrap confidence intervals of H for the major and minor distributions from each instrumental part. As previously mentioned, the cello part received the lowest values of H for the major and minor distributions, and given the lack of overlap between the corresponding confidence intervals and the values of H calculated for the upper parts, we can infer that the major mode distribution from the cello part differs significantly from the upper-part distributions. Shannon entropy therefore demonstrates one fundamental difference between the four instrumental parts in Haydn's compositional style: namely, that the proportion of durations in the cello part concentrate more in $\hat{1}$ and $\hat{5}$ than they do in the upper parts. And although these scale degrees also occur frequently in the upper parts, they become the focal point for melodic activity in the cello part because they serve as harmonic support for the upper parts in the complete texture.

For reference, I have also provided in the right-most columns of Table 4.1 the Pearson correlations for the major and minor distributions from each instrumental part with the corresponding Krumhansl-Kessler goodness-of-fit profiles, as well as the correlations between the distributions themselves.[26] As expected, the distributions in the major mode were all highly

[25] I have described here the *basic percentile* method for deriving bootstrap confidence intervals, but in many cases the bootstrap distribution is non-symmetric (i.e., non-normal) and the test statistic may differ systematically from the population parameter, resulting in bias in the calculated confidence interval. To resolve this issue, Bradley Efron and Robert Tibshirani recommend the bias-corrected and accelerated percentile method (or BC_a), which corrects for both bias and skewness in the bootstrap distribution. Thomas DiCiccio and Efron have demonstrated the success of BC_a to approximate the exact confidence interval for both parametric and non-parametric bootstrap distributions, so I adopt that method here. *An Introduction to the Bootstrap*, 178-201; "Bootstrap Confidence Intervals."

[26] The Pearson product-moment correlation is the most common quantitative statistic in key-finding algorithms, though other measures based on Euclidean distance have recently been considered (Albrecht and Shanahan, "The Use of Large Corpora to Train a New Type of Key-Finding Algorithm"). It represents the magnitude of the relationship between two variables X and Y, giving a value between -1 and $+1$. A negative value indicates a negative relation (e.g., X decreases as Y increases), whereas a positive value indicates a positive relation (e.g., X increases as Y increases), and 0 indicates no correlation between X and Y. For a discussion of the Pearson correlation coefficient in key-finding algorithms, see David Temperley, "What's Key for Key? The Krumhansl-Schmuckler Key-Finding Algorithm Reconsidered," *Music Perception* 17, no. 1 (1999): 67-70.

correlated with the corresponding Krumhansl-Kessler profile ($r > .93$). Thus, the degree to which each distribution conveyed a particular key did not differ between instrumental parts. The inter-correlations between the major mode distributions for each part also replicated the earlier finding obtained by the estimates of H. Here, the distributions from the upper parts were all highly correlated, whereas the cello part was generally less correlated with the other parts.[27]

Thus far I have only considered the distributions of chromatic scale degrees represented by `csd` in the Haydn Corpus. As expected, the duration-weighted distributions replicated the hierarchy of tonal stability described by previous authors, and information-theoretic measures also demonstrated an important difference between the instrumental parts in the complete texture: specifically, the preference for scale degrees like $\hat{1}$, $\hat{5}$, and $\hat{4}$ in the cello part compared to the other parts. But what about the other viewpoints examined in this corpus, viewpoints that clearly play an important role in the representation of cadences?

Table 4.2 presents the statistics for the metric strength, contour, and simple melodic interval distributions weighted by Parncutt's coefficient of durational salience for each instrumental part. Shown in Figure 4.2, the metric strength distributions demonstrated a similar trend for all four parts, with durations appearing more frequently at metrically strong positions—at the levels of the downbeat (4) or first subdivision of the measure (3)—and less frequently at metrically weak positions. This trend was especially clear in the cello part, where durations rarely appeared at lower levels of the metric hierarchy (i.e., at metric positions represented by the lowest two levels of `strength`). This trend became incrementally less clear for each adjacent

[27] That the minor mode distributions were less correlated with the corresponding Krumhansl-Kessler key profile is a well-known limitation of the correlational approach ($r < .85$) (Albrecht and Shanahan, "The Use of Large Corpora to Train a New Type of Key-Finding Algorithm"). At first glance, it also seems noteworthy that the minor mode distribution from the first violin was much less correlated with the corresponding key profile than the other parts ($r = .74$), but again, the small sample size for the minor mode distributions casts considerable doubt on any inferences we may hope to draw.

Table 4.2: Statistics for the metric strength, contour, and simple melodic interval distributions weighted by durational accent for each instrumental part.

	N	H	CI	H_{rel}	Vln 1	Vln 2	Vla	Vc
Metric Strength								
Vln 1	14,506	1.97	1.96–1.97	0.98	–	.99	.97	.98
Vln 2	10,653	1.92	1.91–1.93	0.96		–	.99	1.00
Vla	9156	1.89	1.88–1.91	0.95			–	1.00
Vc	8463	1.87	1.85–1.88	0.93				–
Contour								
Vln 1	14,456	1.49	1.48–1.50	0.94	–	1.00	1.00	.40
Vln 2	10,603	1.54	1.53–1.55	0.97		–	1.00	.36
Vla	9106	1.55	1.54–1.56	0.98			–	.33
Vc	8413	1.58	1.58–1.58	1.00				–
Simple Melodic Interval								
Vln 1	14,456	3.72	3.69–3.76	0.80	–	.97	.94	.81
Vln 2	10,603	3.62	3.59–3.66	0.78		–	.98	.90
Vla	9106	3.69	3.65–3.73	0.79			–	.94
Vc	8413	3.43	3.39–3.47	0.74				–

upper part, however. In the first violin part, for example, durations appeared less frequently at metrically strong positions and more frequently at metrically weak positions relative to the lower parts, resulting in a flatter distribution. In information-theoretic terms, we should expect the incrementally greater freedom associated with each of the adjacent upper part distributions to produce incrementally higher values of H. The values of H shown in Table 4.2 demonstrated precisely this trend, with the cello and first violin parts receiving the lowest (1.87 bits) and highest (1.97 bits) values of H, respectively.

The bottom plot in Figure 4.2 presents the contour distributions for each instrumental part. The upper parts all displayed very little lateral motion and approximately equal proportions of ascending and descending motion. For the cello part, however, the proportion of durations was spread evenly across the distribution, though it did demonstrate a slight preference for

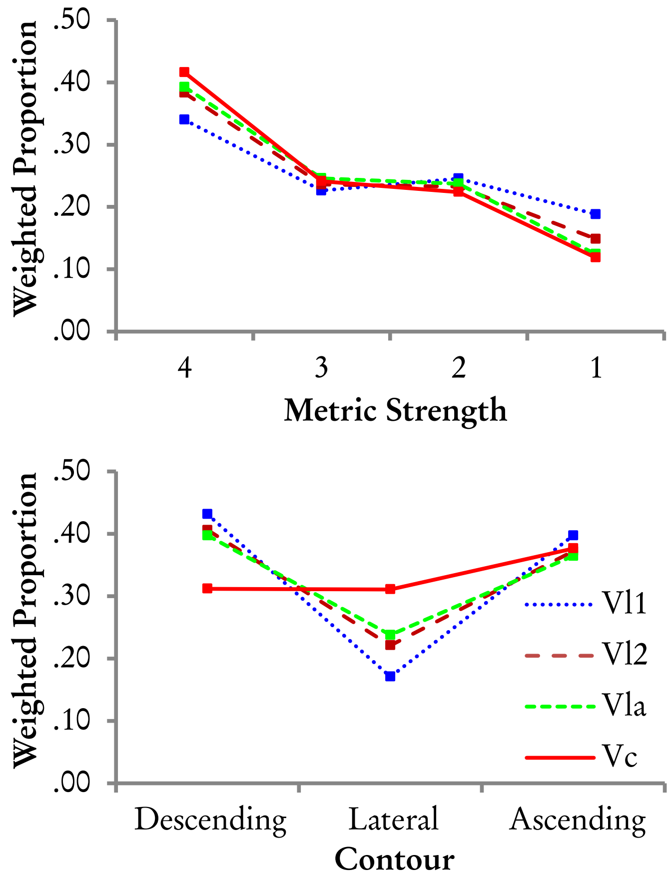


Figure 4.2: Metric strength (top) and contour (bottom) distributions weighted by durational accent for each instrumental part.

ascending motion. As a result, the values of H in Table 4.2 increased from the first violin (1.49 bits) to the cello (1.58 bits).

Unfortunately, `contour` is a coarse viewpoint because it consists of just three symbols. Figure 4.2 indicates that the cello part ascended slightly more frequently than it descended, but by which intervals? Figure 4.3 presents the distributions of simple intervals for each instrumental part (i.e., intervals up to an octave). Smaller intervals predominated in all four parts, resulting in approximately Gaussian (or normal) distributions. The Gaussian curve was particularly evident in the first violin part, presumably because smaller intervals represent the basic unit of motion

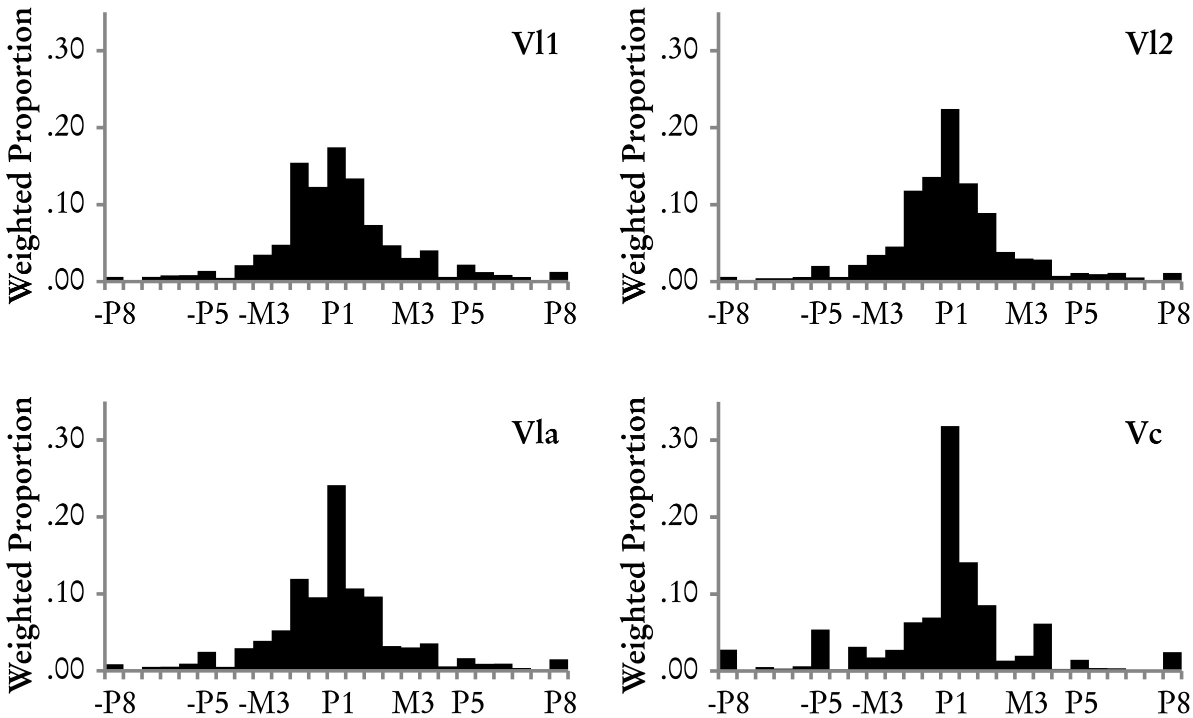


Figure 4.3: Simple melodic interval distributions weighted by durational accent for each instrumental part.

in upper-voice melodies.[28] According to psychologists Jay Dowling and Dane Harwood, this preference for smaller intervals in melodic organization is a cross-cultural universal resulting from the physiological structure of the auditory system,[29] which integrates acoustic components separated in time into one perceptual stream—what Bregman called *auditory stream integration*—partly as a consequence of pitch proximity.[30] But it is also noteworthy that the preference

[28]Existing corpus studies bear this claim out. Piet Vos and Jim Troost have noted, for example, that there is an exponential decrease in the proportional occurrence of intervals larger than five to six semitones ("Ascending and Descending Melodic Intervals: Statistical Findings and Their Perceptual Relevance," *Music Perception* 6, no. 4 [1989]: 383-384). For cross-cultural evidence of the prevalence of small intervals in melodic organization, see Dowling and Harwood, *Music Cognition*; Huron, "Tone and Voice," 25.

[29]Dowling and Harwood, *Music Cognition*, 155-156.

[30]Huron, "Tone and Voice," 22-30. Leon van Noorden's studies examining the boundaries for auditory stream integration aptly demonstrate how *pitch proximity* and *rate of change* interact during perception. Simply put, van Noorden found that for very small changes in log frequency—say, two semitones or less—listeners easily integrate the incoming sound events into one stream regardless of changes in inter-onset interval. For larger changes in

for certain large intervals increased as we descend from the first violin part to the cello part. Unisons, perfect fifths, and octaves appeared fairly infrequently in the first violin part, but in the cello part these intervals were especially common. Presumably the contrapuntal interaction between the outer voices plays some part in establishing this difference, with the bass voice moving both by step and by leap to provide harmonic support for the primarily stepwise soprano.

Shown in Table 4.2, the Shannon entropies for each part reflect these distributional differences, with the prevalence of perfect intervals in the lower parts producing significantly lower values of H. As a result of the presence (or absence) of large intervals, the distributions characterizing upper voice melodies therefore exhibit greater freedom and are most likely easier to track as individual auditory streams, whereas lower voice melodies are both less informative and more difficult to track.[31]

Finally, corpus studies have shown both for Western and non-Western melodies that small intervals tend to descend, whereas large intervals tend to ascend.[32] Shown in Figure 4.4, the interval direction distribution for the first violin part generally replicated this finding, with intervals smaller than a perfect fourth descending, and intervals of a perfect fourth or larger ascending. For the cello part, however, the results were generally reversed. With the exception of the intervals of a minor third and its inversion, intervals smaller than a perfect fifth generally

log frequency (i.e., larger than two semitones), however, changes in the temporal interval between incoming sound events play a significant role, with shorter inter-onset intervals—less than, say, 150 ms—resulting in auditory stream segregation ("Temporal Coherence in the Perception of Tone Sequences" [PhD Dissertation, Institute for Perception Research, 1975]). Based on his findings, one could argue that while the prevalence of large leaps makes bass melodies more difficult to track than upper voice melodies, the prevalence of longer durations in the bass relative to the upper voices might also make bass melodies *easier* to track. A corpus study comparing the distribution of durations from each part might bear this claim out.

[31] Again, this claim assumes that the voices do not differ with respect to rate of change (see footnote 30), but I do not consider that issue here. However, in Chapter 7 I provide experimental evidence in support of the related claim that listeners attend primarily to the soprano voice in multi-voiced textures, though musical training appears to promote flexible voice tracking (see §7.2.3).

[32] Vos and Troost, "Ascending and Descending Melodic Intervals: Statistical Findings and Their Perceptual Relevance," 388.

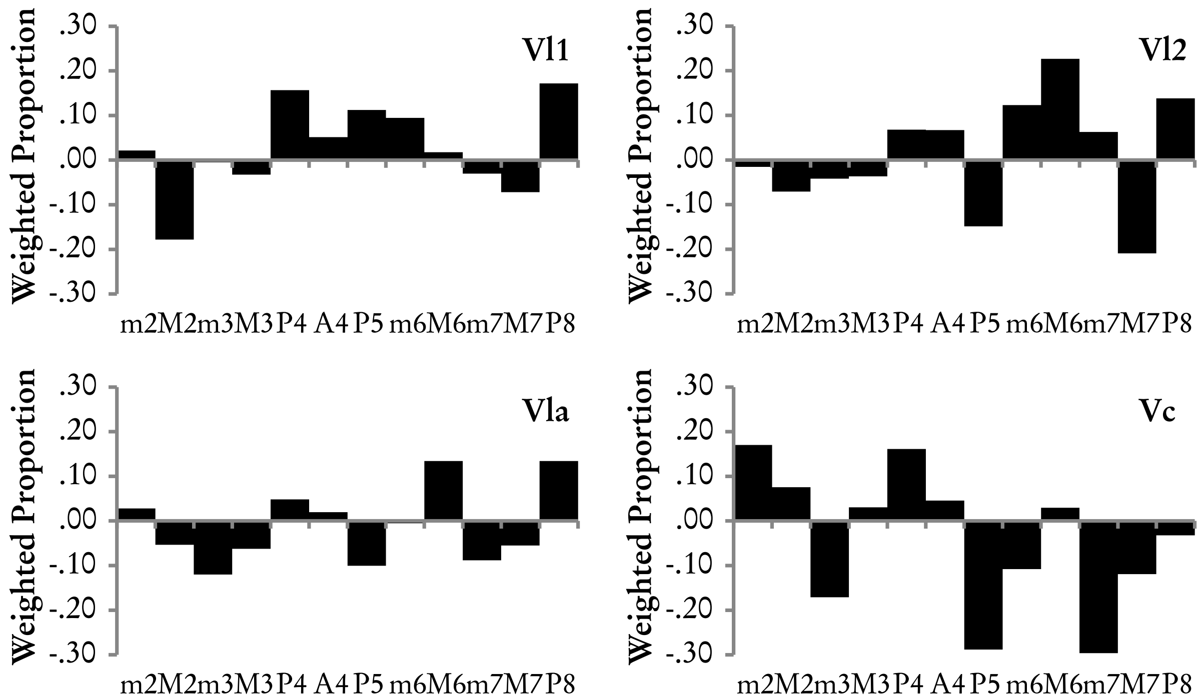


Figure 4.4: Simple interval direction distributions weighted by durational accent for each instrumental part.

ascended, whereas intervals of a perfect fifth or larger generally descended. According to psychologists Piet Vos and Jim Troost, the distribution of interval directions demonstrated in the first violin part confirms Leonard Meyer's claim that large ascending intervals function as carriers of musical tension and surprise, resulting in the gap-fill archetype that has since sparked considerable scholarly debate.[33] That we find precisely the opposite pattern of results in the lowest instrumental part suggests (at the very least) that its role as the "bearer of harmonic fundamentals" severely constrains the distribution of melodic intervals.[34] In the bass voice, the melodic intervals of the descending third, ascending fourth, and their inversions serve especially important functions in Western tonal harmony, thus differentiating bass melodies from upper

[33] Paul von Hippel, "Questioning a Melodic Archetype: Do Listeners Use Gap-Fill to Classify Melodies?," *Music Perception* 18, no. 2 (2000): 139–153.

[34] Caplin, *Classical Form*, 27.

voice melodies in multi-voiced textures.

4.1.2 Chord Events

In the preceding section I provided distributional evidence for the note events in the Haydn Corpus, noting that events deemed stable within a given tonal (`csd`) or metric (`strength`) context appeared more frequently than those deemed unstable. In this section I extend this discussion to chord events. Beginning with vertical interval class combinations, `vintcc` contained $18,007$ combinations representing 190 unique types (hereafter called `vintcc` types). Approximately 30% of the combinations presented just one distinct interval class (e.g., a perfect fifth, or $< 7, \perp, \perp >$), however, so I have omitted those combinations from the analysis in this section in order to examine the most common chord types represented by `vintcc` (major and minor triads, seventh chords, etc.). This procedure eliminated 5277 combinations and 12 distinct `vintcc` types.

Figure 4.5 presents a multi-level pie plot of the vertical interval class combinations that consisted of at least two distinct interval classes, with the proportions weighted by durational accent. Following Quinn, I collated the remaining 178 types into three categories: type, quality, and inversion.[35] The inner pie plot presents the three chord types (e.g., seventh chords); the inner concentric circle presents the chord qualities associated with each chord type (e.g., major-minor, minor, major, half-diminished, and diminished); and the outer concentric circle presents the inversions for each chord quality in clockwise order (i.e., root position, followed by first, second, and third inversion).[36] Gaps in the outer circle indicate that the associated chord quality does not contain recognizable inversions (e.g., augmented triads, diminished seventh chords, etc.). The 'other' type was divided into tertian and non-tertian `vintcc` chord qualities, where tertian events consisted of incomplete seventh chords and their inversions (e.g., $< 4, 10, \perp >$

[35]Quinn, "Are Pitch-Class Profiles Really Key for Key," 154-155.

[36]I have not included inversion labels for the minor, half-diminished, and diminished seventh chords, but the inversions appear in clockwise order.

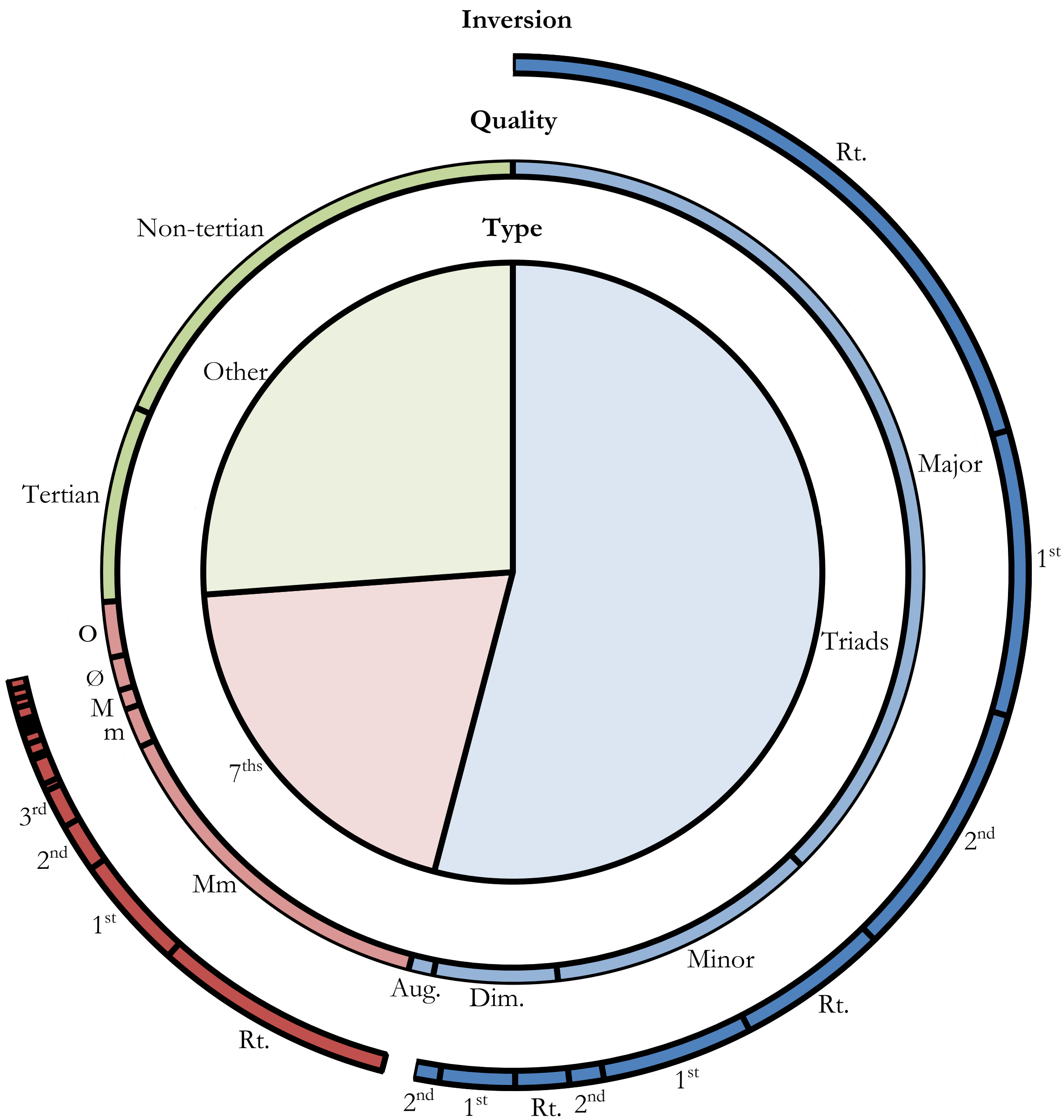


Figure 4.5: Multi-level pie plot of the vertical interval class combinations (`vintcc`) consisting of at least two distinct interval classes weighted by durational accent. The three levels represent chord type (triads in blue, seventh chords in red, and other chords in green), quality (major, minor, diminished, augmented, etc.), and inversion (root position, first, second, third). For the seventh chords, quality abbreviations represent major-minor (Mm), minor (m), major (M), half-diminished (ø), and diminished (o). Tertian combinations represent incomplete seventh chords. Inversions appear in clockwise order beginning in root position for each chord quality. $N = 12{,}730$.

represents a major-minor seventh chord with missing fifth), and non-tertian events consisted of the remaining combinations.

Major triads in root position, first inversion, and second inversion appeared most frequently, followed by major-minor seventh chords in root position, and then minor triads in root position; together these five chord types constituted 50% of the combinations in the corpus. In fact, the 27 `vintcc` types classified as triads (in blue) and seventh chords (in red) in Figure 4.5 accounted for slightly less than three quarters of the $12,730$ combinations in the Haydn Corpus. Nine tertian `vintcc` types represented another 8% of the available combinations, with the remaining 18% of combinations—presumably characterized by the presence of what we ordinarily call 'non-chord tones'— represented by the remaining 142 `vintcc` types.

Given my previous claims about the relationship between frequency-of-occurrence on the one hand and psychological stability on the other, we might conclude from Figure 4.5 that the root position major triad represents the most stable chord type. But unfortunately this rather narrow view of stability also forces us into less tenable claims—for example, that major-minor seventh chords in root position are more stable than minor triads in root position. We could argue that these findings are statistical aberrations, but they generally replicate previous results for Bach's chorales, suggesting at the very least that the distribution of chords in this corpus generalizes to other tonal repertories.[37] Alternatively, one could claim that by pooling events from both the major and minor modes, the distribution of chords in the minor mode was largely lost, since minor mode passages and movements occur far less frequently in the corpus. Assuming the minor mode distribution conforms to the findings presented thus far, minor triads would presumably appear most frequently in the minor mode distribution, followed by major-minor seventh chords and major triads. But again, if the Haydn Corpus is to serve as a

[37] Quinn, "Are Pitch-Class Profiles Really Key for Key," 154-155; Martin Rohrmeier and Ian Cross, "Statistical Properties of Tonal Harmony in Bach's Chorales," in *Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC)*, ed. Ken'ichi Miyazaki et al. (Sapporo, Japan, 2008), 619–627.

metaphor for musical experience, eighteenth-century listeners presumably heard far less music in the minor mode, and so the distribution in Figure 4.5 should better reflect the chord types Galant listeners were likely to learn and remember than would one which weights the two modes equally.

But perhaps the distribution of chords shown in Figure 4.5—and more importantly, our notion of stability—reflects more than one mechanism. As I argued in Chapter 1, psychological stability depends both on the statistical frequency of the event within a particular style *and* on the degree to which listeners organize the acoustic components characterizing that event into a coherent auditory image. For vertical sonorities like triads and seventh chords, for example, the strength of the auditory image depends (among other things) on the harmonicity of the various acoustic components characterizing the members of the chord, not to mention its approximate ambitus, the spacing of the chord members, and of course the timbre(s) voicing each member. In this instance, the tendency towards philosophical monism simply is not justified, since the sensory constraints characterizing human auditory processing play some role in determining the chord typology shown in Figure 4.5.[38]

All of this is to say that implicit learning is by no means the only mechanism underlying psychological stability, particularly for chord typologies represented by viewpoints like `vintcc` that classify sonorities without a fixed reference point. For viewpoints like `csdc`, however, chord representations deemed stable within a particular tonal context presumably appear with far greater frequency than those deemed unstable. To be sure, as we have already seen with viewpoints like `csd` and `strength`, the stability relations characterizing the Western tonal system emerge out of distributional statistics, providing evidence in support of the link between stability and implicit learning.

[38] A discussion of the various sensory mechanisms underlying event formation for vertical sonorities like triads and seventh chords merits its own study, but unfortunately such a study is beyond the scope of this dissertation.

For the chromatic scale degree combinations in the Haydn Corpus, `csdc` contained $17,277$ combinations representing 688 unique types. Approximately 30% of the combinations presented fewer than three distinct chromatic scale degrees (e.g., $< 0, 4, \perp, \perp >$), however, so I have again omitted those combinations in this section in order to examine the most common chord types represented by `csdc` (diatonic and chromatic harmonies, modal mixture, etc.). This procedure eliminated 5135 combinations and 114 distinct `csdc` types.

Shown in Figure 4.6, each multi-level pie plot presents the diatonic chromatic scale degree combinations consisting of at least three chromatic scale degrees from either the major or minor modes, with the proportions weighted by durational accent. The inner pie plot represents the diatonic harmonies, labeled with Roman numeral notation,[39] and the outer concentric circle represents each inversion (root position, first, second, and third). Just as in Figure 4.5, inversions appear in clockwise order for each harmony beginning in root position.

In the major mode, tonic harmony appeared most frequently, followed by dominant harmony, the pre-dominant harmonies IV and ii, and finally vii, vi, and iii. In the outer concentric circle, root position chords predominated for harmonies like I, IV, V, and vi, but unsurprisingly, first inversion chords appeared more frequently for ii, iii, and vii. For the minor mode distribution, the same general pattern of results emerged, with I and V appearing far more frequently than the remaining diatonic harmonies. It is noteworthy, however, that for the minor mode distribution, i, iv, and ii generally appeared much less frequently relative to the major mode distribution, resulting in comparatively larger proportions for V, VI, and vii.

Shown in the lower table in Figure 4.7, the 49 `csdc` types representing diatonic harmony (triads and seventh chords for every diatonic harmony in every inversion) for the major and minor modes accounted for approximately 62% and 68% of the chromatic scale degree combinations, respectively. The table also presents three other harmony categories: modal

[39] Following convention, upper and lower case Roman numerals denote major and minor triads, respectively.

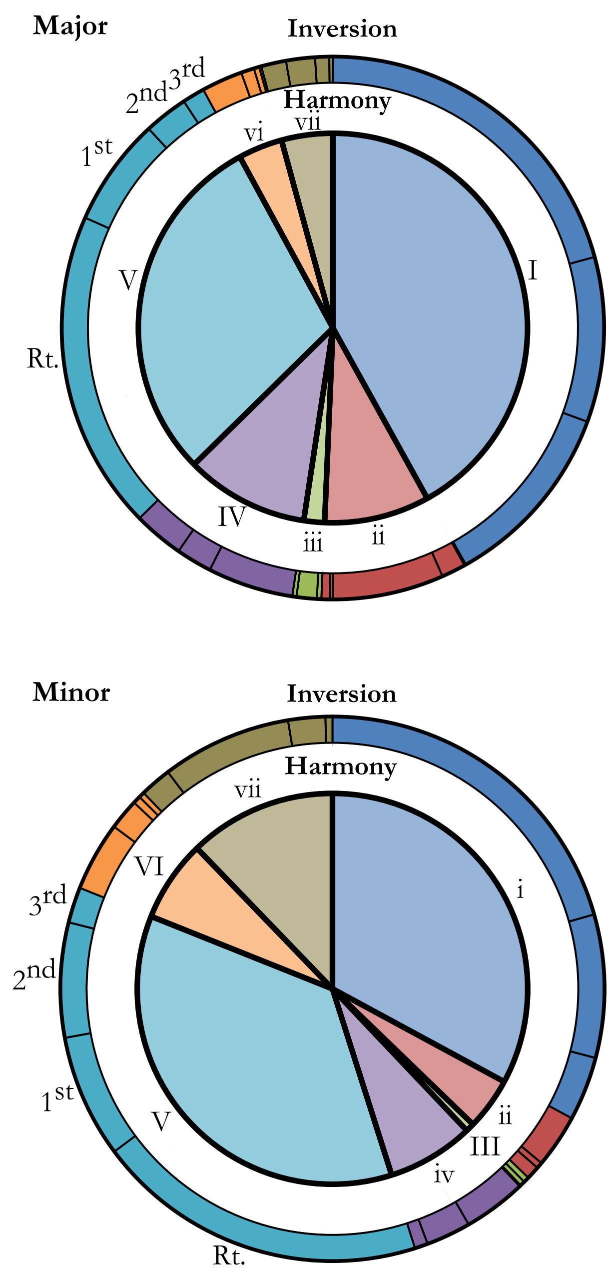
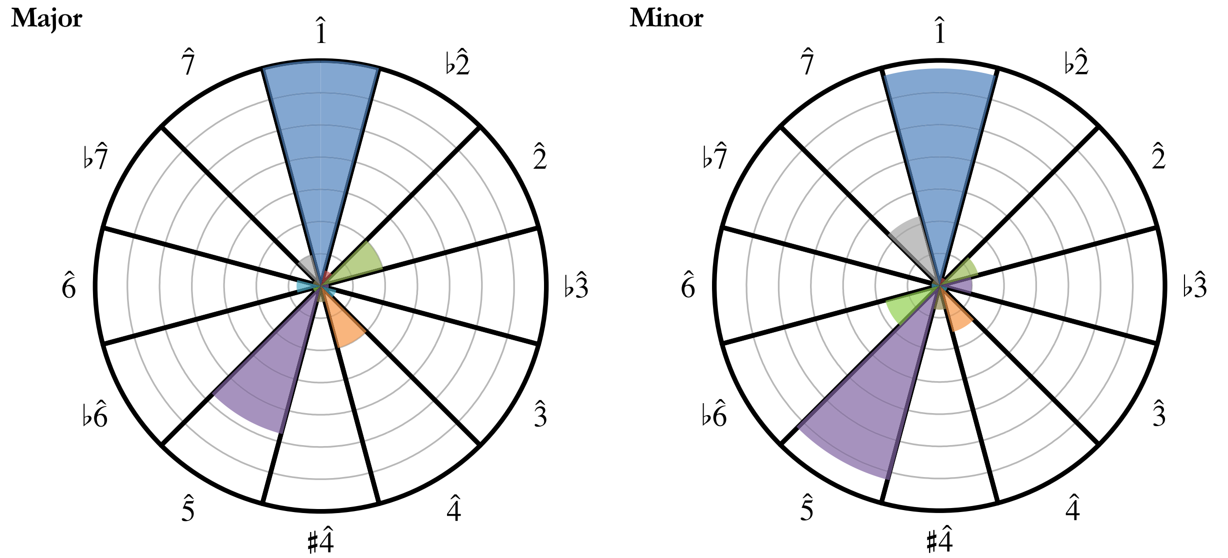


Figure 4.6: Multi-level pie plots of the diatonic chromatic scale degree combinations (`csdc`) consisting of at least three chromatic scale degrees from the major (top) and minor (bottom) modes, with the proportions weighted by durational accent. The two levels represent diatonic harmony (triads and seventh chords) and inversion (root position, first, second, third). Inversions appear in clockwise order for each harmony beginning in root position (labels only provided for dominant harmony). In both plots, iii and IV did not appear in third inversion. In the minor mode plot, i did not appear in third inversion, and ii did not appear in root position. In the major mode plot, $N = 11,253$. In the minor mode plot, $N = 885$.



	Major			*Minor*		
	N	**Duration**	**%**	*N*	**Duration**	**%**
Diatonic	6404	1229.18	61.75	540	118.06	67.55
Modal Mixture	304	65.88	3.31	19	3.35	1.91
Chromatic	707	140.87	7.08	110	26.85	15.36
Other	3838	554.75	27.87	216	24.45	15.18

Figure 4.7: Top: Radial bar plots of the chromatic scale degree combinations (`csdc`) consisting of at least three chromatic scale degrees for which each chromatic scale degree serves as root. Each grid line represents 4% of the total duration, with the proportions weighted by durational accent. Bottom: Percentage of chromatic scale degree combinations weighted by durational accent for each of four categories: diatonic harmony, modal mixture, chromatic harmony (applied dominants, augmented sixth chords, and Neapolitans), and other.

mixture, which represents the 49 `csdc` types from the opposite mode; chromatic harmony, which denotes applied triads and seventh chords (V and vii) for the non-tonic harmonies that were not also shared by the other harmony categories,[40] as well as Neapolitan and augmented

[40]In the major mode, for example, the same chromatic scale degrees represent tonic harmony and V/IV, so in such instances combinations were included in the larger of the two categories.

sixth chords; and other, which represents all other `csdc` types. Together, the diatonic, chromatic, and modal mixture categories accounted for nearly three quarters of the combinations from the major mode and nearly 85% of the combinations from the minor mode.

Finally, Figure 4.7 provides an alternative visualization of the distribution of combinations in `csdc` that includes all three harmony categories represented in the previous table. Here, radial bar plots present the chromatic scale degree combinations for which each chromatic scale degree serves as root. Each grid line represents 4% of the total duration, with the proportions weighted by durational accent. Thus, the radial bar representing $\hat{1}$ in the major mode plot accounts for 28% of the combinations in the Haydn Corpus. For the major mode, $\hat{1}$ and $\hat{5}$ served as the most frequent roots, followed by $\hat{4}$, $\hat{2}$, and $\hat{7}$. In the minor mode, $\hat{1}$ and $\hat{5}$ also occurred most frequently, followed by $\hat{7}$, $\flat\hat{6}$, $\hat{4}$, and finally $\hat{2}$. Compared to the diatonic scale degrees, chromatic scale degrees like $\sharp\hat{4}$ or $\flat\hat{7}$ rarely served as roots within the tonal system.

§4.2 Scale-Degree Schemas

4.2.1 Contiguous *N*-grams

§4.1 provided evidence in support of the view that stable events appear with far greater frequency than those deemed unstable. It would seem that the features on which the cadence concept depends are resistant to further continuation at least in part because they appear frequently in the classical style. Nevertheless, I argued in Chapter 2 that stable events do not simply appear at random. Rather, the classical style is characterized by a limited number of highly stereotyped harmonic and melodic formulæ, and the degree of finality attributed to the final events within those formulæ depends in no small part on the sequence of events they follow. In other words, to remember a perfect authentic cadence is not only to recall its final harmonic and melodic events, but rather to retain some impression of the entire temporal formula: the successions of scale degrees, melodic intervals, and contours appearing within each voice, the harmonies formed by their co-occurrence, the metric context in which they appear, and so forth. Borrowing from psychologists like David Rumelhart and Marvin Minsky,[41] Gjerdingen has called these recurrent patterns *schemas*,[42] but given the term's breadth of application in music psychology and the emphasis Gjerdingen places on scale degree patterns specifically, we might call them *scale-degree schemas*.[43]

In corpus linguistics, researchers often discover recurrent patterns by dividing the corpus into contiguous sequences of n events, called n-grams, and then determining the frequency of each distinct pattern in the corpus.[44] N-grams consisting of one, two, or three events are often called *unigrams*, *bigrams*, or *trigrams*, respectively, while longer n-grams are typically represented by the value of n.[45] Much of the discussion in §4.1, for example, represented the Haydn Corpus using unigrams, but to expand our notion of stability for viewpoints like `csdc`, we need only increase the value of n.

Each movement m consists of a contiguous sequence of combinations, so let k represent the length of the sequence for each movement. Since the Haydn Corpus consists of 50 movements, the number of n-grams in the corpus is $\sum_{m=1}^{50} k_m - n + 1$. `csdc` contains 17,277 combinations representing 688 unique types, so the Haydn Corpus contains 17,227 bigrams, 17,177 trigrams, and so on. Given what we know about the limits of sequence perception, however, we could

[41] Minsky, "A Framework for Representing Knowledge"; Rumelhart, "Notes on a Schema for Stories."

[42] Gjerdingen, *A Classic Turn of Phrase*.

[43] Temperley, Review of Music in the Galant Style, 278.

[44] For an early example of n-gram databases in music research, see J. Stephen Downie, "Evaluating a Simple Approach to Music Information Retrieval: Conceiving Melodic N-Grams as Text" (PhD Dissertation, The University of Western Ontario, 1999).

[45] Formally, n-gram *models* attempt to identify the dependencies between contiguous events using Markov chains. That is, n-gram (or *context*) models attempt to predict the next event in a sequence given a preceding context of n events. In this chapter, I examine the relevance of n-grams for pattern discovery and classification, setting aside *context* models until Chapter 6.

discard those *n*-grams that listeners are unlikely to group into smaller segments.

Psychologists have long observed that participants tend to group sequences separated in time into smaller perceptual units to overcome processing limitations.[46] According to Paul Fraisse, the perception of grouping falls within a range of 200 milliseconds to 1.8 seconds; beyond this upper limit, grouping no longer occurs.[47] This estimate tends to vary between 1.5 and 2 seconds depending on nature of the task,[48] but it generally corresponds to the upper limit of auditory sensory memory, or what Ulric Neisser termed *echoic memory*,[49] which represents the time period over which sensory experiences linger in auditory perception.[50] Thus, in the analyses that follow I have only included *n*-grams whose adjacent events occur within an inter-onset interval (IOI) of 2 seconds under the assumption that sequence perception breaks down beyond this interval.

In the preceding section I weighted the frequency for each distinct unigram using a measure of durational accent under the assumption that note or chord events with longer temporal durations are somehow more salient and/or better remembered than those with shorter temporal durations, and thus should contribute more to the final count. But how would we apply this function to temporal *patterns* (i.e., when the value of *n* is larger than 1)? We could again assume that patterns featuring longer average temporal durations should receive higher weights, but for

[46]McAdams and Drake, "Auditory Perception and Cognition," 428-429.

[47]Paul Fraisse, "Rhythm and Tempo," in *The Psychology of Music*, ed. Diana Deutsch (New York: Academy Press, 1982), 156.

[48]Studies of *subjective rhythmization*—the grouping of isochronous stimuli into groups of twos and threes—typically provide upper limits between 1.5 and 1.8 seconds. See, for example, Thaddeus L. Bolton, "Rhythm," *The American Journal of Psychology* 6, no. 2 (1894): 145–238. For a review of the boundaries of sequential grouping, see London, *Hearing in Time*, 27-47.

[49]Neisser, *Cognitive Psychology*, 190.

[50]David Huron and Richard Parncutt, "An Improved Model of Tonality Perception Incorporating Pitch Salience and Echoic Memory," *Psychomusicology* 12, no. 2 (1993): 157. The upper limit of echoic memory varies between two and four seconds in the auditory domain (Darwin and Turvey, "An Auditory Analogue of the Sperling Partial Report Procedure: Evidence for Brief Auditory Storage"), but I adopt the more conservative estimate in the analyses that follow. For a review of memory and its role in the perception of temporal sequences, see Candace Brower, "Memory and the Perception of Rhythm," *Music Theory Spectrum* 15, no. 1 (1993): 19–35.

longer *n*-grams the duration of event *a* may differ from the IOI between events *a* and *b*.[51] In this section I offer an alternative weighting scheme, one that penalizes patterns that listeners are less likely to remember. When the value of *n* is larger than 1, we might instead weight each pattern by the degree of contiguity between adjacent members of the *n*-gram. Presumably, *n*-grams that minimize the temporal distance between adjacent members are more likely to be perceived as a unitary whole, thereby minimizing the burden on echoic memory.

Hermann Ebbinghaus (1850–1909) was very likely the first psychologist to observe that memory for contiguous sequences decays exponentially over time, which he called the *forgetting curve*.[52] Over the following century, studies of sequence memory gradually refined this curve using inverse exponential functions,[53] though the optimum half-life for sensory memory—the temporal duration in which memory decays by half—varies from anywhere between 0.5 and 3 seconds in the experimental literature.[54]

Several recent studies in music cognition have modeled memory decay for note or chord sequences using inverse exponential functions. David Huron and Richard Parncutt have observed, for example, that accounting for sensory memory decay generally improves key-finding algorithms. By varying the half-life of the decay function, they also successfully derived an optimum half-life of around 1 second for two-chord progressions (i.e., for every second that passes, memory decays by half).[55] Following Huron and Parncutt, I have weighted the appearance of each distinct *n*-gram in `csdc` using an exponential decay function with a half-life of one second to account for the decay in sensory memory resulting from the potential

[51]In other words, *a* and *b* would receive the same durational accent even if a rest appeared between them.
[52]Michael Jacob Kahana, *Foundations of Human Memory* (New York, NY: Oxford University Press, 2012), 6.
[53]Ibid., 292.
[54]For a review of the published findings concerning sensory memory decay, see Huron and Parncutt, "An Improved Model of Tonality Perception Incorporating Pitch Salience and Echoic Memory," 165-166.
[55]They note that this half-life is quite long for echoic memory and suggest that the exponential decay simulated for their study may incorporate elements from short-term memory (Ibid., 166), or what Fraisse once called the *psychological present*, which generally lasts around 3 seconds ("Rhythm and Tempo").

discontiguity between contiguous events, shown in Equation 4.2.

$$d(n\text{-gram}_t) = 2^{\left[\frac{1}{n-1}\sum_{i=2}^{n}(\text{offset}_{i-1} - \text{onset}_i)\right]} \tag{4.2}$$

For each event i in n-gram$_t$, the *distance* function, denoted by d, represents the average offset-to-onset interval (OOI) between adjacent events. The values of d fall in a range between 0-1, with n-grams that do not contain rests between adjacent events receiving a value of 1 (maximal contiguity), and n-grams containing rests approaching 2 seconds in duration receiving values closer to 0 (minimal contiguity). Note here that the exponential decay for a contiguous bigram consisting of events a and b does not begin at the onset of event a (i.e., when a begins), but instead at its offset (i.e., when a ends). Although several studies have suggested that inter-onset intervals, and not offset-to-onset intervals, provide the most perceptually salient information for sequential grouping,[56] the equation above attempts to model the temporal interval over which memory begins to decay for each isolated event within the sequence, and not the approximate interval over which sequential grouping breaks down. For impulsive instrument timbres like piano and harpsichord, this distinction is probably meaningless, since sensory memory will begin to decay soon after perceptual onset. For non-impulsive, sustained timbres like bowed string instruments, however, note or chord events with longer durations are still present in perception after the moment of perceptual onset, so d will only begin to decay after those events end.[57] In other words, grouping breaks down for sequences featuring note or chord events with very long IOIs (i.e., greater than two seconds), but sensory memory for isolated events within that sequence will not decay until each event ends. Thus, d represents the degree of contiguity

[56] McAdams and Drake, "Auditory Perception and Cognition," 426-427.

[57] Ben Duane used precisely this approach to account for sensory memory decay in a context model that predicted each note in a given instrumental part given a preceding context of n events ("Agency and Information Content in Eighteenth- and Early Nineteenth-Century String-Quartet Expositions," 101-102).

(measured by OOI) between adjacent events in the *n*-gram, where an inverse exponential function models the decay in sensory memory resulting from the potential discontiguity.

Unfortunately for complex textures like string quartets, finding relevant patterns using contiguous *n*-grams is a tremendous challenge. The musical surface typically features considerable repetition, and a significant portion of the chromatic scale degree combinations in `csdc` feature fewer than three distinct chromatic scale degrees, thereby obscuring the kinds of patterns we might hope to study (e.g., trichords and tetrachords, harmonic progressions featuring more than one harmony, etc.). One solution to this problem might be to include only those *n*-grams that present a genuine harmonic change between events and consist of at least one trichord or tetrachord in the pattern. To that end, in the analyses that follow I have excluded those *n*-grams that (1) do not present at least one combination featuring more than two distinct chromatic scale degrees, (2) feature adjacent repetitions of the same combination (e.g., $< 0, 4, \perp, \perp >$ and $< 0, 4, \perp, \perp >$), (3) share the same chromatic scale degree in the bass and feature subsets or supersets of chromatic scale degrees in the upper parts (e.g., $< 0, 4, \perp, \perp >$ and $< 0, 4, 7, \perp >$), or (4) represent inversions of the same combination (e.g., $< 0, 4, 7, \perp >$ and $< 4, 0, 7, \perp >$).

Table 4.4 presents the top ten contiguous 2-grams weighted by temporal distance and determined with and without the above exclusion criteria. The combinations are represented using Roman numeral notation, and incomplete harmonies contain the subscript symbols NoRt, No3, and No5 to denote missing chord members. Parentheses indicate a pedal in the bass throughout the *n*-gram. Without exclusion, the top ten 2-grams represent approximately 11% of the 17,131 2-grams with an IOI of less than 2 seconds in the Haydn Corpus. Unsurprisingly, tonic harmony serves as the terminal event in eight out of the ten 2-grams, with the progression V^7–$\mathrm{I}_{\mathrm{No5}}$ ranked seventh overall, appearing 146 times in the corpus. Without exclusion criteria, these 2-grams demonstrate considerable redundancy, with the remaining nine 2-grams presenting repetitions of tonic or dominant harmony.

Table 4.4: Top ten contiguous 2-grams weighted by temporal distance with and without exclusion criteria.

Without Exclusion Criteria				*With Exclusion Criteria*			
d	N	csdc[a]		d	N	csdc	
446.39	450	I	I	145.37	146	V^7	I_{No5}
216	219	I_{No5}	I_{No5}	59	59	V^{6b}_{4}	V
212.78	214	I^6_4	I^6_4	57.9	59	V^6_5	I
183.12	184	I^6	I^6	53.66	54	(vii)	I_{No5}
159.98	165	I_{No5}	I	44.2	46	V^6_4	V^7
154.17	155	V	V	39	39	V^7_{No3}	V^6_4
145.37	146	V^7	I_{No5}	31	31	V^6_4	V^7_{No3}
139.78	141	V^7	V^7	30	30	V^4_2	I^6
104.34	106	I	I_{No5}	29.56	31	ii^6	V^6_4
91.22	92	i	i	28.48	29	IV	I^6

Note. The table includes contiguous 2-grams with inter-onset intervals of less than 2 seconds. Exclusion criteria: (1) any adjacent repetitions; (2) all events represent inversions of the same combination (e.g., $<0,4,7,\perp>$ and $<4,0,7,\perp>$) (3) any adjacent events sharing the same chromatic scale degree in the bass and featuring subsets or supersets of chromatic scale degrees in the upper parts (e.g., $<0,4,\perp,\perp>$ and $<0,4,7,\perp>$); (4) all events have less than three chromatic scale degrees (i.e., no trichords or tetrachords). *N*-grams in red denote partial or complete cadential progressions.

[a] Chromatic scale degree combinations are represented with Roman numerals. The subscript symbols NoRt, No3, and No5 denote missing chord members (e.g., I^6_{NoRt} represents $<4,7,\perp>$). Parentheses indicate that the chromatic scale degree in the bass does not change throughout the *n*-gram (i.e., a pedal.)

[b] All instances of the harmony, V^6_4, denote cadential six-fours (i.e., $\hat{5}$ in the bass with $\hat{1}$ and $\hat{3}$ in the upper parts).

Shown in the right-most columns of the table, 36% of the 17,131 2-grams feature a genuine harmonic change of some sort. With exclusion, progressions deemed "cadential" in the *Formenlehre* tradition rise to the top of the table, with five of the top ten progressions representing partial or complete cadential progressions (shown in red). What is more, the top ten 2-grams represent approximately 9% of the 2-grams meeting these exclusion criteria, indicating that of the progressions featuring harmonic change, cadential progressions are among the most numerous patterns in the Haydn Corpus.

For 3-grams and 4-grams, however, this procedure fails to identify patterns of any significant frequency in the corpus, instead uncovering extended prolongations of tonic and dominant harmony. For example, the top-ranked 3-gram $\mathrm{V}^7_{\mathrm{No3}}$–$\mathrm{V}^6_4$–$\mathrm{V}^5_3$ appears just 14 times in the corpus,[58] while the top-ranked 4-gram (vii)–I–(vii)–I appears just 7 times. Thus, by limiting ourselves to contiguous relationships between combinations, the power-law relationship between frequency and rank order almost entirely disappears as the value of *n* increases, resulting in increasingly flat distributions of chromatic scale degree combinations. The commitment to *contiguous* *n*-grams—the standard method in corpus research—has effectively tied our hands.

4.2.2 Non-contiguous *N*-grams

The contiguous *n*-gram approach reflects an implicit assumption about the nature of association: namely, that any event on the musical surface depends only on its immediate neighbors. For stimuli demonstrating hierarchical structure, however, non-contiguous events often serve as focal points in the syntax.[59] In the contemporary theoretical landscape, this claim should seem uncontentious: the rise of architectonic approaches in 20th century theory—Schenkerian analysis, the cognitively informed *A Generative Theory of Tonal Music*, or recent theories in the *Formenlehre* tradition, to name a few—all demonstrate an unwavering commitment to the discovery of associations that lie beneath (or beyond) the musical surface.[60]

It should be no surprise, then, that studies of sequence memory published over the past century have repeatedly demonstrated the importance of non-contiguous associations for a

[58] Here and in Tables 4.4 and 4.5, V^6_4 denotes a cadential six-four (i.e., $\hat{5}$ in the bass with $\hat{1}$ and $\hat{3}$ in the upper parts).
[59] Gjerdingen, "'Historically Informed' Corpus Studies," 195.
[60] London, "Rhythm in Twentieth-Century Theory."

variety of stimulus domains. According to psychologist Michael Kahana, the concept of *remote association* was in fact Ebbinghaus' main theoretical advance—the idea that events are not only associated in memory with their nearest neighbors in the sequence, but also with the next neighbors, and the next-next neighbors, albeit more weakly.[61] Ebbinghaus writes,

> The associative threads, which hold together a remembered series, are spun not merely between each member and its immediate successor, but beyond intervening members to every member which stands to it in any close temporal relation. The strength of the threads varies with the distance of the members, but even the weaker of them must be considered as relatively of considerable significance.[62]

Given the constraints I initially placed on contiguous *n*-grams—only including those patterns with IOIs shorter than 2 seconds, and privileging those patterns with shorter OOIs in the final count—we may also include *n*-grams featuring non-contiguous events using the same criteria.[63] Nevertheless, non-contiguous patterns may span tens of seconds, so I will add one further constraint concerning the temporal duration of the entire pattern. When the *n*-gram is contiguous and relatively short—consisting of, say, 7 ± 2 events,[64] it typically spans just a few seconds, so it would seem reasonable to assume that listeners could remember the pattern *in toto*. When n is larger than 1 and the pattern features non-contiguous events, however, it may

[61] Kahana, *Foundations of Human Memory*, 7.

[62] Hermann Ebbinghaus, *On Memory: A Contribution to Experimental Psychology* (New York: Teachers College, Columbia University, 1885/1913), 94.

[63] This approach is similar to a technique recently implemented by David Guthrie in natural language processing to identify non-contiguous *n*-grams. In his case, the temporal structure of a sequence of linguistic utterances is less clearly defined, so he identifies non-contiguous relationships by simply skipping adjacent events in the sequence, which he calls *skip* grams ("A Closer Look at Skip-gram Modelling," in *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 06)* [European Language Resources Association, 2006], 1222–1225). James Symons has also demonstrated the applicability of non-contiguous *n*-grams for the discovery of recurrent temporal patterns in a corpus of two-voice solfeggi by sampling at regular temporal intervals ("Temporal Regularity as a Key to Uncovering Statistically Significant Schemas in an Eighteenth-Century Corpus," *Paper Presented at the Society for Music Theory*, 2012, New Orleans, LA).

[64] Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information."

span tens of seconds. Clearly n-grams of this sort lie beyond the limitations of short-term or working memory, so it will be useful to specify an upper limit for the duration of the total pattern that accounts for the limitations of working memory. Justin London has suggested that the longest interval by which we can hierarchically integrate sequential events into a stable pattern is around 5 to 6 seconds,[65] so in what follows I additionally exclude those n-grams with onset intervals measured from the first to the last event of more than 6 seconds.

Without these limitations, the number of associations between events in the sequence—be it a short passage of music or an entire movement—necessarily explodes in combinatorial complexity as the value of n and the number of events in the sequence k increases, since each event may be said to depend on literally every other event in the sequence. Recall that the number of contiguous n-grams in any sequence is given by $k - n + 1$. Thus, the number of contiguous 2-grams in a 5-event sequence is 4 ($5 - 2 + 1 = 4$). Figure 4.8 depicts these contiguous associations with solid arcs. If we include non-contiguous associations, however, the number of n-grams is given by the combination equation:

$$\binom{k}{n} = \frac{k!}{n!(k-n)!} = \frac{k(k-1)(k-2)\ldots(k-n+1)}{n!} \tag{4.3}$$

The notation $\binom{k}{n}$ denotes the number of possible combinations of n events from a sequence of k events, or "k choose n."[66] By including the non-contiguous associations, depicted with dashed arcs in Figure 4.8, the number of 2-grams for a 5-event sequence increases to 10. As the values of k and n increase, the number of patterns can very quickly become unwieldy: a 20-event sequence, for example, contains 190 possible 2-grams, 1140 3-grams, 4845 4-grams, and 15,504 5-grams. Clearly not all of these patterns consist entirely of focal events. What is more,

[65] London, *Hearing in Time*, 27.

[66] The factorial symbol ! denotes the product of all the whole numbers from 1 to n. Thus, $4! = 4 \times 3 \times 2 \times 1$, or 24.

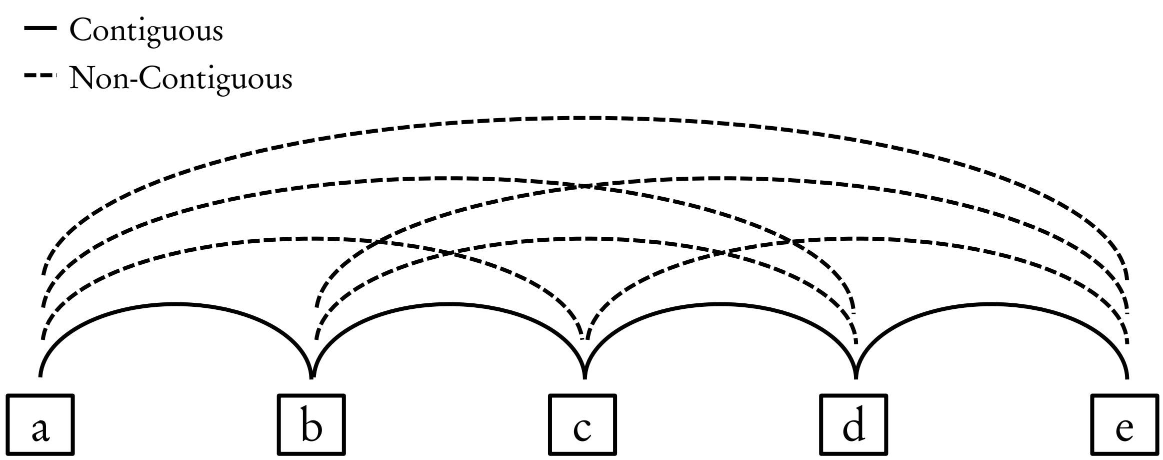


Figure 4.8: A 5-event sequence, with arcs denoting all contiguous (bold) and non-contiguous (dashed) 2-grams.

the vast majority represent redundant associations resulting from repetitions in the sequence.

Consider the following sequence of scale degrees:

$$\hat{2}\text{–}\hat{3}\text{–}\hat{2}\text{–}\hat{2}\text{–}\hat{1}$$

Each event e serves as the terminus for up to $k_e - n$ contiguous and non-contiguous n-grams. Thus, the final event ($\hat{1}$) in the sequence above forms four possible 2-grams: $\hat{2}$–$\hat{1}$ appears three times, and $\hat{3}$–$\hat{1}$ appears once. If each n-gram serves as a plausible representation of the associations listeners might form for a given terminal event in the sequence, we might only include those patterns that represent genuine alternatives (i.e., that represent distinct n-grams). In this case, the final event forms just two distinct 2-grams: $\hat{2}$–$\hat{1}$ and $\hat{3}$–$\hat{1}$. But which of the three 2-grams presenting $\hat{2}$–$\hat{1}$ do we include? I have elected to maximize the distance weight calculated for each distinct n-gram by selecting the pattern with the smallest average temporal distance. Assuming the sequence above features no rests between contiguous events (i.e., OOIs of 0 seconds), the 2-gram $\hat{2}$–$\hat{1}$ would therefore receive a distance weight of 1 because the closest instance is contiguous.

In the case of contiguous patterns, of course, each event serves as the terminus for precisely one n-gram, hence the mathematical property that the total number of n-grams is necessarily smaller than the total number of events in the sequence. But as I have already stated, the great limitation of contiguous n-grams is precisely that they offer no alternatives; every terminal event depends only on the events directly preceding it. For non-contiguous patterns, however, the number of possible distinct n-grams for every event e in the sequence ranges from 1 to $k_e - n$, so this approach often reduces the total number of redundant associations for each event in the sequence by only including the individual weights from among the genuine alternatives. Finally, just as I did with the contiguous n-grams, the analyses that follow include the exclusion criteria described earlier so as to consider only those n-grams representing a harmonic change of some sort.

Table 4.5 presents the results of these procedures for the top ten contiguous and non-contiguous 2-grams, 3-grams, and 4-grams in the Haydn Corpus. The top ten 2-grams are remarkably similar to those found in the right-most columns of Table 4.4, though the absolute counts for each n-gram have understandably increased. For example, the top-ranked 2-gram V^7–$\mathrm{I}_{\mathrm{No5}}$ appears as a contiguous sequence 146 times, but with the addition of non-contiguous sequences, this number increased to 397, with a total distance weight of 312.7. Four of the top ten progressions also feature some variation of V–I, providing further evidence of the primacy of that progression in the tonal system.

Perhaps more importantly, by including non-contiguous sequences, cadential patterns retain their position in the table even as the value of n increases. For 3-grams, the top-ranked pattern V^6_4–V^7–$\mathrm{I}_{\mathrm{No5}}$ appears almost twice as frequently as eight of the remaining nine patterns in the table, with a total distance weight of 102. What is more, the pre-dominant harmony that most often precedes this n-gram in many cadential models, ii^6, appears in the table as the seventh

Table 4.5: Top ten contiguous and non-contiguous 2-grams, 3-grams, and 4-grams weighted by temporal distance.

2-grams				3-grams					4-grams					
d	*N*	csdc[a]		*d*	*N*	csdc			*d*	*N*	csdc			
312.7	397	V^7	$\mathrm{I}_{\mathrm{No5}}$	102	136	V^{6b}_{4}	V^7	$\mathrm{I}_{\mathrm{No5}}$	57.41	82	V^7	$\mathrm{I}_{\mathrm{No5}}$	V^7	$\mathrm{I}_{\mathrm{No5}}$
155.22	210	V^6_5	I	96.19	139	$\mathrm{I}_{\mathrm{No5}}$	V^7	$\mathrm{I}_{\mathrm{No5}}$	39.51	55	ii^6	V^6_4	V^7	$\mathrm{I}_{\mathrm{No5}}$
151.92	209	V^6_4	V^7	57.07	82	V^7	$\mathrm{I}_{\mathrm{No5}}$	V^7	36.03	53	$\mathrm{I}_{\mathrm{No5}}$	V^7	$\mathrm{I}_{\mathrm{No5}}$	V^7
128.66	199	V	$\mathrm{I}_{\mathrm{No5}}$	56.63	95	I	V^7	$\mathrm{I}_{\mathrm{No5}}$	32.15	50	$\mathrm{I}_{\mathrm{No5}}$	$\mathrm{I}^6_{\mathrm{NoRt}}$	V^7	$\mathrm{I}_{\mathrm{No5}}$
109.43	163	I^6	IV	55.39	78	$\mathrm{I}_{\mathrm{No5}}$	(vii)	$\mathrm{I}_{\mathrm{No5}}$	32.08	44	(vii)	$\mathrm{I}_{\mathrm{No5}}$	(vii)	$\mathrm{I}_{\mathrm{No5}}$
108.53	146	V^6_4	V	47.94	78	I	IV^6_4	I	29.52	41	V^4_3	I	V^4_3	I
105.81	166	V^7	I	47.11	66	ii^6	V^6_4	V^7	28.34	47	$\mathrm{I}^6_{\mathrm{NoRt}}$	$\mathrm{I}_{\mathrm{No5}}$	V^7	$\mathrm{I}_{\mathrm{No5}}$
104.43	156	V^6	I	46.06	81	$\mathrm{I}_{\mathrm{No5}}$	V	$\mathrm{I}_{\mathrm{No5}}$	27.92	41	$\mathrm{ii}^6_{\mathrm{NoRt}}$ [c]	V^6_4	V^7	$\mathrm{I}_{\mathrm{No5}}$
101.03	175	V	I	44.88	71	I	V^6_5	I	27.66	44	V^7	$\mathrm{I}_{\mathrm{No5}}$	$\mathrm{I}^6_{\mathrm{NoRt}}$	$\mathrm{I}_{\mathrm{No5}}$
99.9	122	(vii)	$\mathrm{I}_{\mathrm{No5}}$	43.94	71	I	V^6	I	26.33	44	$\mathrm{I}_{\mathrm{No5}}$	$\mathrm{I}^6_{\mathrm{NoRt}}$	V	$\mathrm{I}_{\mathrm{No5}}$

Note. The table includes contiguous and non-contiguous *n*-grams with inter-onset intervals of less than 2 seconds. Exclusion criteria: (1) any adjacent repetitions; (2) all events represent inversions of the same combination (e.g., $< 0, 4, 7, \perp >$ and $< 4, 0, 7, \perp >$) (3) any adjacent events sharing the same chromatic scale degree in the bass and featuring subsets or supersets of chromatic scale degrees in the upper parts (e.g., $< 0, 4, \perp, \perp >$ and $< 0, 4, 7, \perp >$); (4) all events have less than three chromatic scale degrees (i.e., no trichords or tetrachords). *N*-grams in red denote partial or complete cadential progressions.

[a] Chromatic scale degree combinations are represented with Roman numerals. The subscript symbols NoRt, No3, and No5 denote missing chord members (e.g., $\mathrm{I}^6_{\mathrm{NoRt}}$ represents $< 4, 7, \perp >$). Parentheses indicate that the chromatic scale degree in the bass does not change throughout the *n*-gram (i.e., a pedal.)

[b] All instances of the harmony, V^6_4, denote cadential six-fours (i.e., $\hat{5}$ in the bass with $\hat{1}$ and $\hat{3}$ in the upper parts).

[c] $\mathrm{IV}_{\mathrm{No5}}$ is an alternative Roman numeral interpretation of the combination ($< 5, 9, \perp, \perp >$).

ranked 3-gram, with a total distance weight of 47.11. It should come as no surprise, then, that the pattern ii^6–V^6_4–V^7–$\mathrm{I}_{\mathrm{No5}}$ and its variant $\mathrm{ii}^6_{\mathrm{NoRt}}$–$\mathrm{V}^6_4$–$\mathrm{V}^7$–$\mathrm{I}_{\mathrm{No5}}$ appear as the second and eighth most common 4-grams, respectively. Assuming these two patterns in fact represent just one pattern, in which the pre-dominant may necessarily imply either IV or ii^6, the combined distance weight for these 4-grams of 67.43 would make the cadential progression Pre-Dom–V^6_4–V^7–$\mathrm{I}_{\mathrm{No5}}$ the most common 4-gram in the Haydn Corpus.

The appeal of non-contiguous *n*-grams for stimulus domains exhibiting hierarchical structure should now be obvious, since the patterns that appear most frequently—and thus, the patterns that listeners are most likely to learn and remember—necessarily lie beneath the surface. What is more, this approach makes minimal assumptions about the nature of the musical materials, the temporal context in which these patterns appear, or the stylistic knowledge listeners might actually possess. In short, non-contiguous *n*-grams provide evidence in support of the view that cadential patterns are among the most important schemas in the tonal system using an inductive, data-driven method.

§4.3 Conclusions

For many, the cadence is the quintessential tonal schema, a perfect distillation of the features characterizing the classical style. This chapter presented evidence in support of this view, demonstrating that the features on which the cadence concept depends occur with remarkable frequency in the Haydn Corpus. §4.1 replicated and extended the published distributional evidence for individual tonal, harmonic, and metric events, and examined the statistical evidence that distinguishes each voice of the two-voice framework using an information-theoretic tool called *Shannon entropy* as well as a statistical technique called the *bootstrap*. To discover recurrent temporal *patterns* like cadences, I then extended the canonical *n*-gram approach in §4.2—which divides the corpus into contiguous sequences of n events—by including non-contiguous sequences using a weighting function d, which awards higher weights to patterns comprised of temporally proximal members. When non-contiguous n-grams were included in the final count, the cadential progression Pre-Dom–V^6_4–V^7–I_{No5} emerged as among the most frequent patterns in the Haydn Corpus.

But what about the other cadential schemas characterizing the classical style? Surely they possess internal regularities that justify their inclusion in existing cadence typologies, despite their absence from Table 4.5. In this case, statistical analyses conducted on large corpora tend to obscure the many recurrent patterns contained therein. Gjerdingen notes, for example, that many chords can follow IV in compositional practice, but only one chord can follow IV in the Prinner schema (namely, I^6). The former is necessarily a *global* question, necessitating a statistical analysis that is limited only by the size of the corpus, whereas the latter depends on our first identifying the many instances of that pattern *before* computing statistics intended to reveal its internal organization.[67] Thus, in the next chapter I first annotate a collection of cadences from the Haydn Corpus before examining a few techniques for the classification of these cadences on the basis of the features they share.

[67]Gjerdingen, "'Historically Informed' Corpus Studies," 196.

Chapter 5

Classifying Closing Schemas: The Cadential Typology

Taxonomy (the science of classification) is often undervalued as a glorified form of filing—with each species in its folder, like a stamp in its prescribed place in an album; but taxonomy is a fundamental and dynamic science, dedicated to exploring the causes of relationships and similarities among organisms. Classifications are theories about the basis of natural order, not dull catalogues compiled to avoid chaos.

Stephen Jay Gould

The capacity to identify and categorize recurrent patterns encountered during everyday life is a tremendous perceptual and cognitive feat, one achieved by the mind's propensity to detect regularities amidst the 'blooming, buzzing confusion' of present experience and compare those clusters of co-occurring attributes against remembered exemplars.[1] Indeed, if the mind is to create order out of the 'welter of stimulation' characterizing the perceptual present,[2] it must do so by organizing objects and events into categories.

[1] William James, *The Principles of Psychology*, 2 vols. (New York: Holt, 1890), 1:488.
[2] Mandler, "Categorical and Schematic Organization," 260.

The classical cadence is a case in point. As conventional harmonic and melodic formulæ, cadences provide perhaps the clearest instances of phrase-level schematic organization in the classical style. And as the profusion of terms associated with the classical cadence demonstrates, scholars within the *Formenlehre* tradition continue to offer new cadence types, with the essential characteristics concerning the final harmonic and melodic events of each type often playing the decisive role in the final classification.

In Chapter 2 I noted that the commitment to these "criterial" or essential features in category formation is widely held in the *Formenlehre* tradition. In the perfect authentic cadence, for example, theorists often suggest that the dominant and tonic harmonies of the cadential progression must be in root position, and the tonic must support $\hat{1}$ in the soprano voice. In this chapter I consider an alternative view, one that exemplifies the probabilistic approach to category formation adopted by cognitive psychologists over the last half century, in which a category is understood as a network of overlapping attributes, and members are prototypical to the extent that they bear a *family resemblance* to—have attributes in common with—other members of the category.

To support this claim, this chapter classifies the many cadential patterns found in the Haydn Corpus using a family of techniques pioneered (or inspired) by psychologist Amos Tversky. §5.1 describes the methodology for annotating exemplars of the following five cadences in the Haydn Corpus from Caplin's cadence typology: perfect authentic (PAC), imperfect authentic (IAC), half (HC), deceptive (DC), and evaded (EV) (see Table 2.1). In §5.2, I calculate the similarity between every pair of cadences using Tversky's *ratio* model, which was later adapted by Daniel Müllensiefen and Marc Pendzich to compare melodies, and then apply (in §5.3) an additive clustering algorithm called the *neighbor-joining* method to visualize the obtained similarity estimates using phylogenetic trees.

§5.1 Collostructions: The Cadence Collection

To uncover meaningful patterns that otherwise elude global statistics, linguists Anatol Stefanowitsch and Stefan Gries abandoned traditional collocational analysis, which determines which items occur most frequently around a given item within a large corpus, in favor of what they called *collostructional analysis*, which starts with a particular construction and then investigates which items occur more or less frequently in a given slot within that construction.[3] The distribution of words following the word *day* in a corpus of English text is presumably quite large, for example, but following the construction *Don't give up your day*, the distribution very likely consists of just one word (*job*).

The benefit of this approach is that it can identify the internal regularities characterizing a particular construction without appealing to global statistics. But by *starting* with a construction, collostructional analysis is necessarily more labor-intensive, since the researchers must inspect and manually code each instance before the analysis can take place.[4] What is more, this approach assumes that we already know a great deal about the constructions we seek to study. To provide evidence that listeners actually learn and remember the cadential schemas associated with the classical style, we would ideally attempt to identify these constructions using *unsupervised* methods that might mirror the kinds of mental processes that characterize learning and memory.[5] In this case, however, I will adopt the collostructional approach, which is to manually identify the many instances of a given pattern (in this case, cadences) before

[3]Anatol Stefanowitsch and Stefan Th. Gries, "Collostructions: Investigating the Interaction of Words and Constructions," *International Journal of Corpus Linguistics* 8 (2003): 209–243.

[4]Ibid., 215.

[5]In classification tasks, *supervised* learning algorithms train on (or learn from) a set of annotated examples before attempting to classify a set of unannotated examples (i.e., they require feedback as to the correct output corresponding to any given input). *Unsupervised* learning algorithms attempt the task without such feedback (Pearce, "The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition," 2). For an introduction to machine learning algorithms in music research, see McKay and Fujinaga, "Style-Independent Computer-Assisted Exploratory Analysis of Large Music Collections."

applying techniques intended to reveal its internal organization.[6]

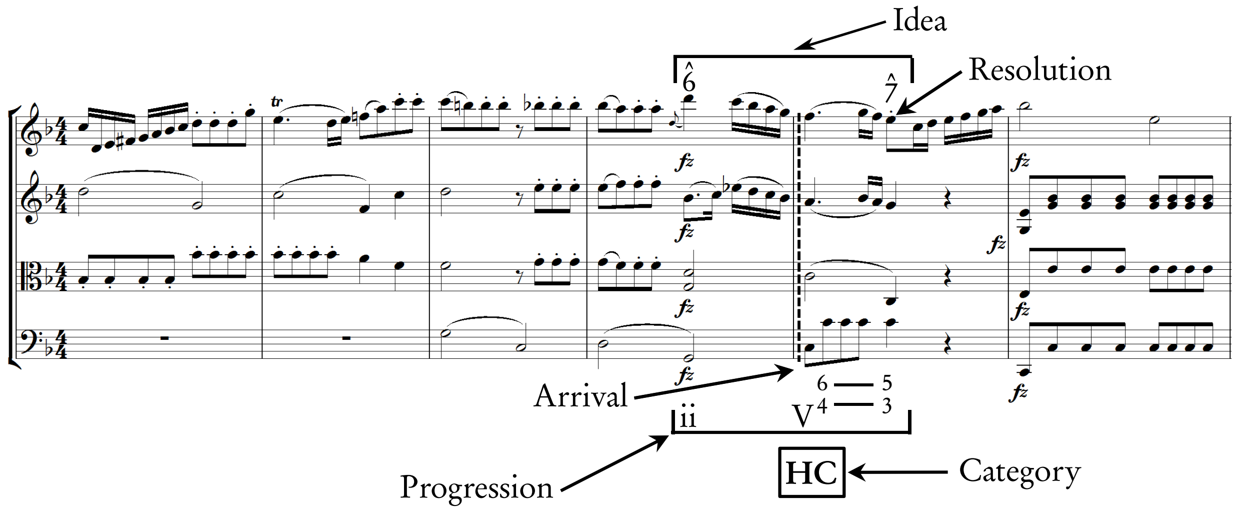
To examine the cadences found in the corpus, I classified exemplars of the five cadence categories that achieve (or at least promise) cadential arrival in Caplin's cadence typology—PAC, IAC, HC, DC, and EV (see Table 2.1). Along with the cadential classification, I have also included for each cadence: the boundaries of cadential function; the harmonies of the cadential progression; the duration of the cadential progression, expressed as a percentage of the length of the total movement; the presence of an expanded cadential progression, cadential six-four, or trill within cadential function; or a melodic surface dissonance or change in dynamics at cadential arrival. Finally, to consider the relationship between cadential articulation and formal function, I also annotated the boundaries of the thematic sections in each sonata-form exposition: Main Theme (MT), Transition (TR), and Subordinate Theme (ST). Suffice it to say that I will not explore each of these annotations in what follows, instead concentrating on those annotations that have direct bearing on the representation scheme described in Chapter 3.[7]

The Haydn Corpus contains 270 cadences (see Appendix A). Example 5.1 illustrates some of the relevant annotations for one such cadence, which appears in the opening movement of Haydn's string quartet in F, Op. 76/2. As mentioned in Chapter 2,[8] this passage features a half cadence in m. 19, with the cadential bass line appearing in the cello part in m. 18 following the end of the "fifths" motive for which this quartet was named. In this case, the cadential melody is entirely consistent with the *converging* half cadence schema shown in Figure 5.1, but the superposition of the end of the "fifth" motive with the beginning of the cadential progression

[6]Gjerdingen, "'Historically Informed' Corpus Studies," 196. As a consequence, any generalizations I might make about how we learn and remember the cadential schemas that appear in the Haydn Corpus await further experimental study using computational models, human listeners, or (preferably) both.

[7]The annotations described here are clearly subject to the idiosyncrasies or biases of the analyst. To mitigate the effects of subjective interpretation for the analyses that follow and further ensure that the exemplars from the cadence collection achieved cadential status in Caplin's typology, all of the cadential identifications were made in consultation with William E. Caplin.

[8]See Example 2.7.



Example 5.1: Haydn, String Quartet in F, Op. 76/2, i, mm. 15–20.

in the cello part has altered the normative cadential bass line of the schema from $\hat{4}$-$\sharp\hat{4}$-$\hat{5}$ to $\hat{2}$-$\hat{5}$.

The encoded cadence collection approximates the prototypes in Gjerdingen's scheme by representing each cadence using the viewpoints `csd` and `contour` for each of the outer parts, and `vintcc` and `strength` between all four parts. Each encoded cadence represents the *cadential idea* for the viewpoints from the first violin part and the *cadential progression* for the viewpoints from the cello part. In some cases, the resolution of the cadential melody does not correspond with the end of the cadential progression, as in Example 5.1.[9] In this case, the encoded viewpoints for the first violin part and the cadential progression begin at the onset of the cadential idea on beat 3 of m. 18 and end at the resolution of the dissonant suspension on beat 2 of m. 19, and the viewpoints for the cello part begin on beat 3 of m. 18 and end at the moment of cadential arrival on the downbeat of m. 19.

Recall that Gjerdingen's representation scheme depends on a two-voice framework consisting of the outer voices. I noted in Chapter 4, however, that the first violin and cello parts are by

[9]Recall from Chapter 2 that Caplin marks the onset of the *cadential arrival* at the time point where the final harmony of the cadential progression first appears, but the end of the cadential idea sometimes includes a surface dissonance that delays the moment of melodic resolution.

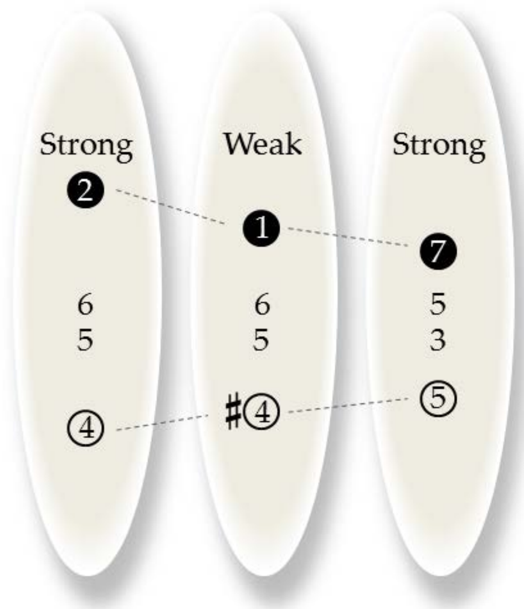


Figure 5.1: The converging cadence schema prototype, represented using Gjerdingen's notation (2007).

no means synonymous with the soprano and bass voices. In the classical string quartet, for example, the cadential bass line sometimes appears in an inner part (e.g., in the viola or second violin). Before entering a cadence into the corpus, it was therefore necessary to indicate whether the outer parts were present in the cadential function. To examine these cadences more closely using computational methods, I have elected to omit those cadences in which the cadential progression and cadential idea do not appear in the cello and first violin parts, respectively. This procedure eliminated 15 cadences. Additionally, another 10 cadences represent cadential deviations that imply more than one category (i.e., PAC-EV or DC-EV). Thus, for the analyses that follow (both in Chapters 5 and 6), the corpus consists of 245 cadences.

Figure 5.2 visualizes the cadence collection as a pie chart. As expected, the perfect authentic cadence (in blue) and the half cadence (in red) represent the most prevalent categories, followed by the cadential deviations: the deceptive and evaded categories (in magenta and yellow, respec-

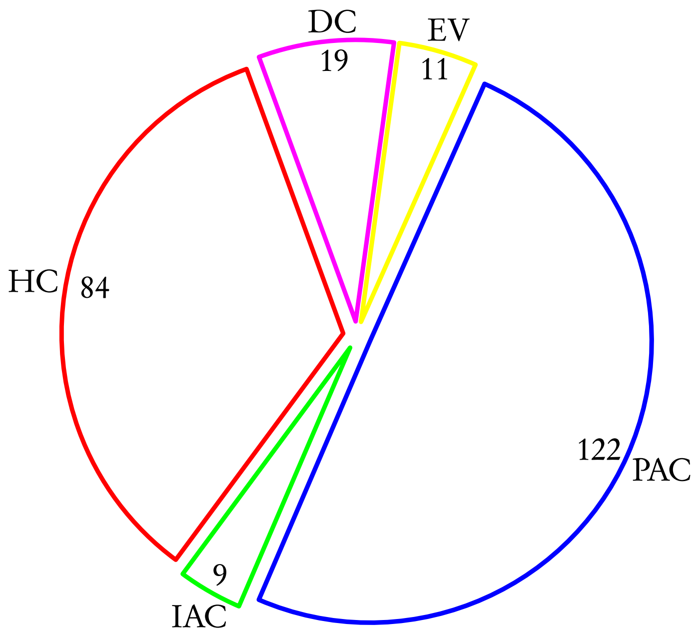


Figure 5.2: Pie chart of the cadences in the Haydn Corpus.

tively). The imperfect authentic cadence (in green) is the least common category, which perhaps reflects the late eighteenth-century stylistic preference for perfect authentic cadential closure at the ends of themes and larger sections. According to Markus Neuwirth, this distribution largely replicates previous findings for Mozart's keyboard sonatas,[10] suggesting this distribution may characterize the classical style in general.

I mentioned in Chapter 2 that this cadence typology tends to *underspecify* a number of important constituent features of closing schemata that relate principally to melodic organization, features which undoubtedly play an important role in perception and memory. To resolve this issue, I adopted Gjerdingen's representation scheme for the cadences in the Haydn Corpus. In what follows I will further classify these cadences using a similarity model first proposed by Amos Tversky and later adapted by Daniel Müllensiefen and Marc Pendzich to compare melodies.

[10]Markus Neuwirth, e-mail message to author, October 28, 2015.

§5.2 Similarity and Prototypicality

5.2.1 Tversky's *Ratio* Model

To uncover the relationships between items in a given corpus, classifiers very often depend on some notion of similarity. And because similarity is limited at one extreme (sameness) but not obviously at the other,[11] researchers often operationalize similarity as a distance metric, where items are represented as points in a multi-dimensional space and observed dissimilarities between items correspond to the metric distances between the respective points.[12] A metric distance function, denoted here by δ, is therefore a scale that assigns to every pair of items a non-negative number that represents their distance in accordance with the following three assumptions: (1) that the distance between any item and itself is 0 (self-identity); (2) that the distance between any pair of items is always the same, no matter the direction of the comparison (symmetry); and (3) that distance is transitive, such that if item a is similar to item b, and item b is similar to item c, a and c cannot be very dissimilar (triangle-inequality).[13]

Given the apparent simplicity of the approach, it should come as no surprise that the geometric (or dimensional) view described here has become the prevailing method in cognitive psychology. But in a seminal review of similarity models, psychologist Amos Tversky noted that the geometric approach does not always reflect the similarity judgments of participants in experimental conditions.[14] According to Tversky, similarity judgments are inherently directional, and thus, asymmetric. The expression, "a is like b," for example, has a subject, a, and a referent, b, and is not necessarily equivalent to the similarity statement, "b is like a." Citing

[11] Alan Marsden, "Interrogating Melodic Similarity: A Definitive Phenomenon or the Product of Interpretation?," *Journal of New Music Research* 41, no. 4 (2012): 324.

[12] Amos Tversky, "Features of Similarity," *Psychological Review* 84, no. 4 (1977): 327.

[13] Ibid., 328.

[14] Ibid.

Rosch's work on category formation, Tversky instead suggested that the choice of subject and referent depends on the prototypicality of the items, where the more prototypical item serves as the referent. And so for Tversky, similarity judgments represent an implicit evaluation of prototypicality, where the variant (subject) is always more similar to the prototype (referent) than vice versa.

To accommodate the potential asymmetry between items, Tversky pioneered a set-theoretic approach to similarity, in which each item is represented by a collection of features or attributes, and the similarity between two items is a feature-matching process that increases with the addition of common features and/or the deletion of distinctive features. In the initial formulation of this approach, Tversky offered two similarity models, but the matching function of interest to us here is the *ratio model*:

$$\delta(a,b) = \frac{f(A \cap B)}{f(A \cap B) + \alpha f(A \setminus B) + \beta f(B \setminus A)}, \alpha, \beta \geq 0, \tag{5.1}$$

where similarity lies between 0 and 1. a and b represent the items under investigation, and A and B denote the sets of features. The function $f(A \cap B)$ measures the salience of the features shared by a and b, and $f(A \setminus B)$ and $f(B \setminus A)$ measure the salience of the features that are distinct to a and b, respectively, where salience (or prominence) generally refers to the intensity, frequency, familiarity, good form, or informational content of a given feature within the larger set.[15] The terms α and β denote weights that express the degree of asymmetry calculated by the model. If $\alpha = 1$ and $\beta = 0$, for example, the salience of the features shared by a and b is only evaluated with respect to all of the features in a. This choice of weights yields an asymmetric

[15] Ibid., 332.

and directional similarity relation and $\delta(a,b)$ reduces to

$$\frac{f(A \cap B)}{f(A)},$$

where $\delta = 1$ if a shares all of its features with b, and $\delta = 0$ if a shares none of its features with b, regardless of the features that might be distinct to b. If $\alpha = \beta = 1$, however, the model is symmetric and $\delta(a,b)$ reduces to

$$\frac{f(A \cap B)}{f(A \cup B)},$$

where $f(A \cup B)$ measures all of the features in a and b. Thus, δ is only a metric model if $\alpha = \beta = 1$, (i.e., $\delta(a,b) = \delta(b,a)$).

Although geometric models continue to appear in studies of melodic similarity, Tversky's set-theoretic approach has received broad support in recent decades.[16] In a comparative study of state-of-the-art similarity algorithms, Daniel Müllensiefen and Marc Pendzich recently noted that their implementation of Tversky's *ratio* model correctly predicted court decisions on music plagiarism with high degrees of accuracy (up to 90%) relative to the other geometric models they examined.[17] In their approach, the distinct contiguous n-grams in each melody a and b from length $n=1$ to $n=4$ represent the respective feature sets A and B. To determine the salience or prominence of each n-gram (denoted by f in Equation 5.1), they applied a weighting scheme from computational linguistics called the *Inverted Document Frequency* (*IDF*), which measures

[16]See, for example, James C. Bartlett and Jay W. Dowling, "Scale Structure and Similarity of Melodies," *Music Perception* 5, no. 3 (1988): 285–314; Marsden, "Interrogating Melodic Similarity"; Naomi Ziv and Zohar Eitan, "Themes as Prototypes: Similarity Judgments and Categorization Tasks in Musical Contexts," *Musicae Scientiae* Discussion Form 4A (2007): 99–133.

[17]Daniel Müllensiefen and Marc Pendzich, "Court Decisions on Music Plagiarism and the Predictive Value of Similarity Algorithms," *Musicae Scientiae* Discussion Forum 4B (2009): 257–295.

the prevalence of a distinct n-gram τ in a collection of melodies C.

$$IDF(\tau, C) = \log\left(\frac{|C|}{|m : \tau \in m|}\right) \tag{5.2}$$

$|C|$ denotes the total number of melodies in collection C and $|m : \tau \in m|$ is the total number of melodies that contain n-gram τ at least once. High *IDF* values therefore indicate rare n-grams, whereas low values indicate common n-grams. In the context of Tversky's equation, this means that two melodies will be more similar if the n-grams they share appear infrequently in the collection.

To examine the predictive value of symmetric over asymmetric models, they applied α and β weights corresponding to a symmetric model ($\alpha = \beta = 1$), an asymmetric model (setting α or β to 1 and the other to 0), and a dynamically weighted model (representing α and β as the ratio of their shared n-grams to the total n-grams in a and b, respectively). They then combined the similarity values computed for each length n into one composite similarity estimate that represents the arithmetic average of the similarity values calculated for each length weighted by the entropy of the n-gram distribution from the larger collection. Because entropy necessarily increases as length increases, this weighting scheme awards higher weights to similarity values computed from longer n-grams under the assumption that longer exact matches play a greater role in the perception of similarity.[18]

5.2.2 Implementation

In what follows I have adapted their approach to the comparison and classification of cadences from the Haydn Corpus, but with a few noteworthy alterations to the original model. First, they restricted the feature set to contiguous n-grams, but I have also included non-contiguous n-

[18] Ibid., 274.

grams, weighting each distinct n-gram by the distance function d described in §4.2 (see Equation 4.2). For n-grams that appear more than once in the sequence, I select the highest distance weight. Second, they examined all distinct n-grams across the entire melody, weighting each n-gram equally regardless of its position within the sequence. For the classical cadence, however, events appearing at the end of cadential function play a far greater role in the final classification than those appearing at the beginning, so I have restricted the feature set to contiguous and non-contiguous n-grams whose final members terminate at the end of the sequence. As a consequence, two cadences whose sequences do not share the same final event will automatically receive a similarity value of `0` for that viewpoint, since neither cadence will share any n-gram of any length. Third, Müllensiefen and Pendzich only applied the model to the sequence of melodic intervals characterizing each melody. To approximate the representation scheme advocated by Gjerdingen, I have applied the similarity model separately to the sequences of contours and scale degrees characterizing each outer part (`contour` and `csd`), and the sequences of vertical interval class combinations and metric strengths characterizing the full texture (`vintcc` and `strength`), and then combined the resulting similarity values in a later step.

The salience function f and the weighted average depend on a collection of melodies from which to draw the requisite *IDF* and Shannon H statistics. The goal here is to select a collection that is somehow representative of the sequences under investigation. Ideally, we would derive the necessary statistics from a much larger collection of cadences, but given the restricted size of the corpus, I have instead selected a collection of excerpts that could contain the cadences as members, but which might also feature other kinds of patterns. In this corpus, all of the annotated cadences terminate on metrically strong positions at the moment of cadential arrival (`strength` $= \{3, 4\}$) and feature the outer parts throughout cadential function, and over 90% of the cadences are less than 6 seconds in duration. To ensure the collection could contain the annotated cadences as members, I randomly selected 1,000 non-overlapping excerpts that terminated on metrically strong positions, featured both of the outer parts throughout the excerpt, and lasted 6 seconds in duration. The *IDF* and H statistics were then computed from the distinct, distance-weighted contiguous and non-contiguous n-grams that terminated on the final event for each excerpt in the Haydn collection.

If we replace f, A, and B with the appropriate IDF weights, $f(A \cap B)$ can be rewritten as

$$\sum_{\tau \in A_n \cap B_n} IDF_C(\tau)$$

The numerator of Equation 5.1 is now represented as the sum of the IDF estimates for the terms shared by a and b. This is the approach favored by Müllensiefen and Pendzich,[19] but in my case, the model also includes non-contiguous terms, so simply replacing each term with its corresponding IDF estimate would put contiguous and non-contiguous terms on equal footing. To resolve this issue, we might award higher weights to terms that minimize the temporal distance between adjacent members so that uncommon terms lying closer to the surface will receive higher weights in the final model.

For terms that are distinct to either a or b, including the distance weight is straightforward. $f(A \setminus B)$ can be rewritten as

$$IDF_C(\tau_{A \setminus B}) \cdot d(\tau_{A \setminus B}),$$

which represents the dot product of the IDF estimates from the collection and the distance estimates from a that are distinct to a (i.e., the sum of the IDF estimates weighted by d for the distinct terms in a). Shared terms consist of two distance weights, however, since the term necessarily appears in both sets. In such cases, I select the higher value of d in order to maximize the similarity estimate for cadences with shared terms. Thus, if the distance of a shared term is

[19] Müllensiefen and Pendzich, "Court Decisions on Music Plagiarism and the Predictive Value of Similarity Algorithms," 272.

.8 in *A* and .5 in *B*, the term would receive a distance weight of .8.

For Tversky, an item is prototypical if it exemplifies the category to which it belongs.[20] That is, prototypical items contain features that are common to the other members of that category. Or put another way, we might say that prototypicality reflects the *absence* of distinct features, so that cadences with many (or all) shared features and few (or no) distinct features are more likely to serve as prototypes within a given category. The appeal of Tversky's ratio model is thus that the asymmetry measured between items is an explicit measure of prototypicality: the greater the difference in the similarity estimates for a pair of cadences, the more "prototypical" the referent cadence.

But perhaps the most important consequence of the set-theoretic approach described here is that asymmetries imply a "nested" relation, where the more prototypical stimulus nests within the variant. Psychologists James Bartlett and Jay Dowling have noted, for example, that asymmetries in similarity ratings for scalar and non-scalar melodies suggest that scalar melodies elicit fewer alternatives in the minds of listeners. The assumption here is that listeners implicitly categorize new stimuli into a set of possible alternatives, with the range of the imagined set determining the stability of the stimulus: the smaller the set, the more stable the stimulus.[21] And because non-scalar melodies presumably belong to a much larger set of alternative stimuli, Bartlett and Dowling expected that less stable (i.e., non-scalar) melodies would be perceived as more similar to more stable (i.e., scalar) melodies than vice versa. The results confirmed this asymmetry, suggesting that nested relations occur when the set of alternatives evoked by one stimulus (the non-scalar subject) includes the second stimulus (the scalar referent).[22]

If we assume that asymmetries reflect the presence of a nested relation, in which the more

[20] Tversky, "Features of Similarity," 347.

[21] Wendell R. Garner, "Good Patterns Have Few Alternatives: Information Theory's Concept of Redundancy Helps in Understanding the Gestalt Concept of Goodness," *American Scientist* 58, no. 1 (1970): 34–42; Wendell R. Garner, *The Processing of Information and Structure* (Potomac, MD: Lawrence Erlbaum, 1974).

[22] Bartlett and Dowling, "Scale Structure and Similarity of Melodies," 289-290.

stable referent nests within the embellished variant, we need only set β to 1 and α to 0 to apply this assumption to Tversky's ratio model. In this formulation, the ratio model reflects the features that are distinct to b, and $\delta(a,b)$ reduces to

$$\frac{f(A \cap B)}{f(A \cap B) + f(B \setminus A)} \tag{5.3}$$

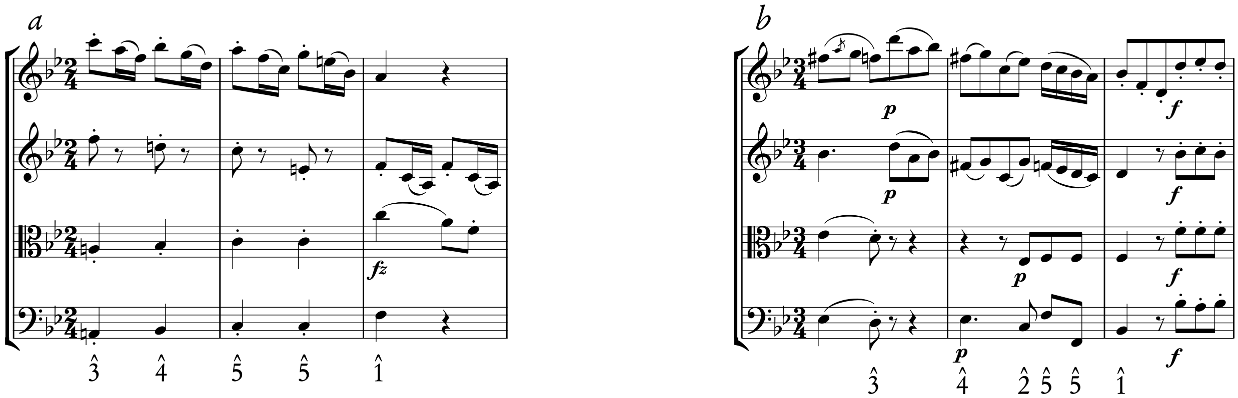
If a is the variant, b necessarily contains fewer distinct features, so $\delta(a,b) > \delta(b,a)$. If a is the referent (or prototype), however, b contains more distinct features, so $\delta(b,a) > \delta(a,b)$.

Replacing f, A, and B in Equation 5.3 with the IDF and d weights described earlier yields the final model.

$$\delta_n(a,b) = \frac{IDF_C(\tau_{A_n \cap B_n}) \cdot [d_{A_n}(\tau_{A_n \cap B_n}) \vee d_{B_n}(\tau_{A_n \cap B_n})]}{IDF_C(\tau_{A_n \cap B_n}) \cdot [d_{A_n}(\tau_{A_n \cap B_n}) \vee d_{B_n}(\tau_{A_n \cap B_n})] + IDF_C(\tau_{B_n \setminus A_n}) \cdot d_{B_n}(\tau_{B_n \setminus A_n})} \tag{5.4}$$

The model represents each term τ as the product of the IDF weight computed from the collection and the corresponding d value from either a or b. If the term is distinct, d represents the highest distance estimate calculated from the instances of the term in b ($d_{B_n}(\tau_{B_n \setminus A_n})$). If the term is shared by a and b, however, d represents the higher of the two distance estimates from a and b ($[d_{A_n}(\tau_{A_n \cap B_n}) \vee d_{B_n}(\tau_{A_n \cap B_n})]$).[23] Each similarity estimate, $\delta_n(a,b)$, is then combined into one composite similarity estimate, $\delta(a,b)$, which represents the arithmetic average of the similarity values calculated for each length weighted by the entropy of the n-gram distribution from the larger collection.

To illustrate how Tversky's equation models nested relations in cadential contexts, consider the two cadences shown in Example 5.2. Using Gjerdingen's framework, we could compare the

[23]The term $\vee$ denotes the maximum of the two distance estimates for the term shared by a and b.

a

Scale Degree	Onset (s)	Duration (s)
$\hat{3}$	0	0.5
$\hat{4}$	0.5	0.5
$\hat{5}$	1	0.5
$\hat{5}$	1.5	0.5
$\hat{1}$	2	0.5

b

Scale Degree	Onset (s)	Duration (s)
$\hat{3}$	0	0.25
$\hat{4}$	1	0.75
$\hat{2}$	1.75	0.25
$\hat{5}$	2	0.25
$\hat{5}$	2.25	0.25
$\hat{1}$	2.5	0.5

Example 5.2: a) Haydn, String Quartet in B-flat, Op. 50/1, iv, mm. 68–70. b) String Quartet in B-flat, Op. 64/3, i, mm. 3–5. Below: Event representation of the chromatic scale degrees in the cello part for *a* and *b*.

sequences of contours or chromatic scale degrees in either of the outer parts, or the sequences of metric strengths or vertical interval class combinations for the entire texture, but in this example I will confine my observations to the sequence of chromatic scale degrees appearing in each cello part. In the Neapolitan tradition, the bass line in *a*, $\hat{3}$-$\hat{4}$-$\hat{5}$-$\hat{5}$-$\hat{1}$, exemplifies the *cadenza composta*, or compound cadence, in which $\hat{5}$ in the standard cadential bass line, $\hat{3}$-$\hat{4}$-$\hat{5}$-$\hat{1}$, receives two metrical units to support a dissonant suspension.[24] The bass line in *b* is nearly identical, but also includes $\hat{2}$ between $\hat{5}$ and $\hat{1}$, resulting in the sequence $\hat{3}$-$\hat{4}$-$\hat{2}$-$\hat{5}$-$\hat{5}$-$\hat{1}$. We might therefore expect *a* to serve as the prototype and *b* to serve as the variant because the former "nests within" the latter, resulting in an asymmetry relation where $\delta(a,b) < \delta(b,a)$.

[24] Sanguinetti, *The Art of Partimento*, 105.

Table 5.1: Similarity algorithm for the cadences in Example 5.2. Top: Distance and IDF weights for the n-grams that are distinct to a ($A \setminus B$), shared by a and b ($A \cap B$), and distinct to b ($B \setminus A$). Bottom: Asymmetric similarity estimates for each length n, along with the average estimates weighted by shannon H.

$$a = \hat{3}\text{-}\hat{4}\text{-}\hat{5}\text{-}\hat{5}\text{-}\hat{1}$$

$$b = \hat{3}\text{-}\hat{4}\text{-}\hat{2}\text{-}\hat{5}\text{-}\hat{5}\text{-}\hat{1}$$

	$A \setminus B$			$A \cap B$			$B \setminus A$		
	τ	d	IDF	τ	d	IDF	τ	d	IDF
$n = 1$									
				$\hat{1}$	1	1.18			
$n = 2$									
				$\hat{3}$-$\hat{1}$	.35	3.7	$\hat{2}$-$\hat{1}$	.71	3.14
				$\hat{4}$-$\hat{1}$	.59	3.65			
				$\hat{5}$-$\hat{1}$	1	1.91			
$n = 3$									
				$\hat{3}$-$\hat{4}$-$\hat{1}$	.71	4.03	$\hat{2}$-$\hat{5}$-$\hat{1}$	.92	3.68
				$\hat{3}$-$\hat{5}$-$\hat{1}$	.71	3.46	$\hat{3}$-$\hat{2}$-$\hat{1}$	.5	3.72
				$\hat{4}$-$\hat{5}$-$\hat{1}$	.84	3.06	$\hat{4}$-$\hat{2}$-$\hat{1}$	.84	3.84
				$\hat{5}$-$\hat{5}$-$\hat{1}$	1	2.37			
$n = 4$									
				$\hat{3}$-$\hat{4}$-$\hat{5}$-$\hat{1}$	.89	3.6	$\hat{2}$-$\hat{5}$-$\hat{5}$-$\hat{1}$	1	4.3
				$\hat{3}$-$\hat{5}$-$\hat{5}$-$\hat{1}$	.89	3.85	$\hat{3}$-$\hat{2}$-$\hat{5}$-$\hat{1}$	.67	4.14
				$\hat{4}$-$\hat{5}$-$\hat{5}$-$\hat{1}$	1	3.23	$\hat{3}$-$\hat{4}$-$\hat{2}$-$\hat{1}$	.75	4.57
							$\hat{4}$-$\hat{2}$-$\hat{5}$-$\hat{1}$	.94	4.2

n	H	$A \setminus B$	$A \cap B$	$B \setminus A$	$\delta(a, b)$	$\delta(b, a)$
1	2.72	0	1.18	0	1	1
2	5.64	0	5.36	2.23	.71	1
3	8.71	0	10.26	8.47	.55	1
4	11.6	0	9.86	14.45	.41	1
				TOTAL	.56	1

Table 5.1 presents the IDF and d weights for the terms that are distinct to a ($A \setminus B$), shared by a and b ($A \cap B$), and distinct to b ($B \setminus A$). For each length n, the ratio model represents the measures of the common and distinctive terms as the dot product of the corresponding IDF and distance estimates. At $n = 1$, for example, $\hat{1}$ appears at the end of the sequence in 308 of the 1,000 sequences in the collection, so its IDF weight is small ($\log(1000/308) = 1.18$). For unigrams the distance measure makes little sense, so for $n = 1$, d corresponds to a count of the term in the sequence.[25] And because the model described here only includes n-grams whose final members terminate at the end of the sequence, the value of d will always be 1 for unigrams. Thus, at $n = 1$, the dot product of the weights for the term shared by a and b is 1.18.

At $n = 2$, the two sequences share $\hat{3}$-$\hat{1}$, $\hat{4}$-$\hat{1}$, and $\hat{5}$-$\hat{1}$. In the previous section I excluded non-contiguous n-grams with IOIs longer than 2 seconds, and so by that reasoning we might also exclude $\hat{3}$-$\hat{1}$ from b (see the event representation in Example 5.2). In this section, however, I have elected to relax that restriction under the assumption that the cadence is a perceivable whole (i.e., that the cadence serves as a stable pattern over which listeners can hierarchically integrate sequential events). To that end, I also include non-contiguous n-grams with IOIs less than the maximum interval over which hierarchical integration could presumably take place (i.e., 6 seconds). Shown in the lower table, the dot product of the estimates shared by a and b is $(.35)(3.7) + (.59)(3.65) + (1)(1.91)$, or 5.36, while the dot product of the estimates that are distinct to b is 2.23. Plugging these values into Equation 5.3, the similarity of a to b at $n = 2$ is $\frac{5.36}{5.36+2.23}$, or .71, while the similarity of b to a is $\frac{5.36}{5.36+0}$, or 1.

As n increases, the asymmetry between a and b increases. For $n > 2$, a shares all of its terms with b, so that $\delta_n(b, a) = 1$, but b contains distinct terms featuring $\hat{2}$ that lower the corresponding similarity estimates. Following Müllensiefen and Pendzich, I have combined

[25] Recall from the previous section that *distance*, denoted by d, represents the degree of contiguity (measured by OOI) between adjacent events in the n-gram. Hence, *distance* assumes $n > 1$.

these estimates into one composite similarity estimate that represents the arithmetic average of the similarity estimates weighted by the Shannon entropy of the *n*-gram distribution from the collection. Shown in the lower table, the composite similarity estimate of a to b is .56, while the composite similarity estimate of b to a is 1. Thus, for this example the asymmetry reflects a nested relation, where a serves as the referent (or prototype) because it shares all of its features with b.

§5.3 Classification Using Additive Trees

5.3.1 The *Neighbor-Joining* Method

Applying Tversky's ratio model to every pair of cadences in the Haydn Corpus, we obtain an asymmetric similarity matrix consisting of 245 rows and 245 columns for every viewpoint in Gjerdingen's scheme—`csd` and `contour` for the outer parts, and `strength` and `vintcc` for the entire texture—thereby yielding six matrices. To obtain graphic representations that classify the cadences on the basis of the features they share, I will apply a technique from cluster analysis called *additive clustering*, which represents each cadence as a node in a connected graph, called a *tree*, and which represents the dissimilarities between cadences by the lengths of the paths joining them.

Tree representations of similarity data are widespread in the scholarly community, though additive trees are less well known than their more restrictive *hierarchical clustering* counterparts, most likely because the hierarchical clustering scheme is more straightforward to calculate.[26] In agglomerative (or bottom-up) hierarchical clustering, the two items a and b with the highest

[26] Computational approaches to taxonomic organization were dominated by hierarchical clustering algorithms in the early days. See, for example, Peter H. A. Sneath and Robert R. Sokal, *Numerical Taxonomy: The Principles and Practice of Numerical Classification* (San Francisco: Freeman, 1973), 188-308. For a review of additive clustering algorithms, see Joseph Felsenstein, *Inferring Phylogenies* (Sunderland, MA: Sinauer Associates, 2004), 147–195.

similarity estimate are grouped into a single cluster, and then the similarity between the new cluster and any other item c is calculated according to a specified *linkage* method, which is often the minimum, maximum, or average of (a, c) and (b, c).[27] The algorithm then repeats this step until it obtains a single cluster that includes all of the items. Thus, hierarchical clustering produces a *metric* tree, in which the height of the lowest (internal) node connecting two items represents the distance (or dissimilarity) between them, and where items connected to lower nodes on the tree denote higher similarity relations.[28]

Although hierarchical clustering remains the most common technique in cluster analysis, it unfortunately imposes severe limitations on the resulting tree representation because it implies that (1) all intra-cluster distances are smaller than all inter-cluster distances; and (2) all inter-cluster distances are equal. Together these properties define the *ultrametric inequality* that characterizes hierarchical clustering. And because similarity judgments often violate the ultrametric inequality, Shmuel Sattath and Amos Tversky championed the more general additive clustering scheme to represent similarity data, which assumes that (1) intra-cluster distances may exceed inter-cluster distances; and (2) objects outside a cluster are not necessarily equidistant from objects inside a cluster.[29]

In graph theory nomenclature, a *tree* is a connected graph without cycles, which is to say that a tree consists of nodes, and any pair of nodes is connected by exactly one straight line (or path). External nodes representing items are called *leaves*, while the paths connecting leaves to internal nodes are called *branches*. An *additive* tree is thus a metric representation of the similarity matrix, in which the dissimilarity between items is represented by the length of the path (or *cophenetic distance*) that joins them. Extending the metaphor yet further, *ultrametric*

[27] These three functions denote the *single*, *complete*, and *average* linkage methods in hierarchical clustering algorithms.

[28] For an example of hierarchical clustering in music analysis, see Ian Quinn, "Listening to Similarity Relations," *Perspectives of New Music* 39, no. 2 (2001): 147-153.

[29] Shmuel Sattath and Amos Tversky, "Additive Similarity Trees," *Psychometrika* 42, no. 3 (1977): 319–345.

trees possess a node called the *root* that is equidistant from all external nodes. Trees derived from hierarchical clustering algorithms typically represent the root on top such that a horizontal line could connect the leaves of the tree. But because additive trees are metric but not ultrametric, the intra- and inter-cluster distances between items vary, resulting in a better-fitting (and more tree-like) representation of the similarity matrix.

The *neighbor-joining* method (NJ) is perhaps the most well-known additive clustering algorithm to date. Created by biologists Naruya Saitou and Masatoshi Nei in the late 1980s,[30] the NJ method was intended to visualize the genetic similarities between biological species,[31] and it remains the canonical method for the estimation of phylogenetic trees, though its application now extends far beyond its original mandate.[32] And since the NJ method has been shown to produce tree topologies with higher degrees of fit than the traditional hierarchical clustering scheme, it is the preferred method here.

The details of the algorithm need not concern us here,[33] but the basic procedure is much the same as the agglomerative hierarchical clustering scheme described earlier: the NJ algorithm iteratively groups the pair of items (or clusters) with the highest similarity estimate and then calculates the similarity between the resulting cluster and every other item (or cluster) in the matrix. But unlike hierarchical clustering, which places the items at equidistant points along an imaginary line, the NJ method starts with a star-like formation, with items placed at equidistant

[30] Naruya Saitou and Masatoshi Nei, "The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees," *Molecular Biology and Evolution* 4, no. 4 (1987): 406–425; James A. Studier and Karl J. Keppler, "A Note on the Neighbor-Joining Algorithm of Saitou and Nei," *Molecular Biology and Evolution* 5, no. 6 (1988): 729–731.

[31] Radu Mihaescu, Dan Levy, and Lior Pachter, "Why Neighbor-Joining Works," *Algorithmica* 54, no. 1 (2009): 1–24.

[32] For examples of the NJ method in fields like literature and music, see Albert C.-C Yang et al., "Information Categorization Approach to Literary Authorship Disputes," *Physica A: Statistical Mechanics and its Applications* 329, nos. 3-4 (2003): 473–483; Esben Paul Bugge et al., "Using Sequence Alignment and Voting to Improve Optical Music Recognition from Multiple Recognizers," in *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)* (2011), 405–410.

[33] For a worked example of the NJ algorithm, see Saitou and Nei, "The Neighbor-joining Method," 407-414.

points around a circle, and with their path lengths connected to a central node like the spokes of a wheel. At each step, the pair of items (or clusters) with the highest similarity estimate is joined to a newly created node that is connected to the central node, and then the least-squares method is used to estimate the branch lengths and to update the similarity matrix.[34]

Like most clustering algorithms, the NJ method is metric, which is to say that it requires a symmetric matrix that represents the dissimilarity between items. For the purposes of classification, it will therefore be necessary to recalculate the similarity matrix from Tversky's ratio model such that $\delta(i, j) = \delta(j, i)$ for every pair of cadences in the Haydn Corpus, but I will nevertheless retain the asymmetric matrix to determine the most prototypical cadences in each class.

The standard approach to dealing with pronounced asymmetries in a similarity matrix is to simply calculate the arithmetic mean of $\delta(i, j)$ and $\delta(j, i)$, and since the ratio model computes similarity rather than dissimilarity, I also subtract the similarity matrix from the identity relation in the ratio model so that the resulting matrix represents the dissimilarity between all pairs of cadences in the Haydn Corpus. For the cadences a and b in Table 5.1, for example, $\delta(a, b) = .56$ and $\delta(b, a) = 1$, so $\delta(a, b) = \delta(b, a) = 1 - \frac{1}{2}(.56 + 1)$, or .22, where 0 represents identity.

To demonstrate how the NJ algorithm works, Figure 5.3 presents the additive trees (or *dendrograms*) calculated from the dissimilarity matrix shown in Table 5.2 for the sequence of chromatic scale degrees from eight authentic cadential bass lines in the Haydn Corpus. Additive trees can be visualized in a number of ways. The square method shown in the top dendrogram is

[34]Saitou and Nei, "The Neighbor-joining Method," 408-409. Although Sattath and Tverksy pioneered the additive tree method (called ADDTREE), which is based on rather different underlying principles than those of the NJ method, several authors have noted that the ADDTREE and NJ methods make similar mathematical assumptions and produce identical or highly similar trees (Olivier Gascuel, "A Note on Sattath and Tversky's, Saitou and Nei's, and Studier and Keppler's Algorithms for Inferring Phylogenies from Evolutionary Distances," *Molecular Biology and Evolution* 11, no. 6 [1994]: 961; Felsenstein, *Inferring Phylogenies*, 166–170).

Table 5.2: Dissimilarity (or distance) matrix calculated by Tversky's ratio model for the sequence of chromatic scale degrees from eight authentic cadential bass lines in the Haydn Corpus.

	$\hat{3}$-$\hat{3}$-$\hat{4}$-$\hat{5}$-$\hat{1}$	$\hat{3}$-$\hat{4}$-$\hat{5}$-$\hat{1}$	$\hat{4}$-$\hat{3}$-$\hat{2}$-$\hat{5}$-$\hat{1}$	$\hat{4}$-$\hat{5}$-$\hat{1}$	$\hat{3}$-$\hat{4}$-$\hat{5}$-$\hat{5}$-$\hat{1}$	$\hat{4}$-$\hat{2}$-$\hat{5}$-$\hat{5}$-$\hat{1}$	$\hat{4}$-$\hat{2}$-$\hat{5}$-$\hat{1}$	$\hat{5}$-$\hat{1}$
$\hat{3}$-$\hat{3}$-$\hat{4}$-$\hat{5}$-$\hat{1}$	–							
$\hat{3}$-$\hat{4}$-$\hat{5}$-$\hat{1}$	.17	–						
$\hat{4}$-$\hat{3}$-$\hat{2}$-$\hat{5}$-$\hat{1}$	.59	.57	–					
$\hat{4}$-$\hat{5}$-$\hat{1}$	.22	.20	.27	–				
$\hat{3}$-$\hat{4}$-$\hat{5}$-$\hat{5}$-$\hat{1}$	.35	.16	.60	.22	–			
$\hat{4}$-$\hat{2}$-$\hat{5}$-$\hat{5}$-$\hat{1}$	.68	.68	.40	.24	.51	–		
$\hat{4}$-$\hat{2}$-$\hat{5}$-$\hat{1}$	.68	.67	.23	.24	.68	.16	–	
$\hat{5}$-$\hat{1}$	.22	.22	.22	.20	.21	.24	.25	–

the canonical representation for clustering trees because it emphasizes the hierarchical relations of the tree; in this case, the root marks the point at which the blue and red lines meet. Here, the sum of the horizontal path lengths (or *cophenetic distances*) represents the dissimilarity between two items in the tree, but bear in mind that the length of the vertical paths connecting two items or clusters is arbitrary. In this dendrogram, we can see that $\hat{4}$-$\hat{2}$-$\hat{5}$-$\hat{1}$ is very similar to $\hat{4}$-$\hat{2}$-$\hat{5}$-$\hat{5}$-$\hat{1}$, with a cophenetic distance of .16 (.08 + .08); less similar to $\hat{5}$-$\hat{1}$, with a cophenetic distance of .3 (.08 + .1 + .12); and still less similar to $\hat{3}$-$\hat{3}$-$\hat{4}$-$\hat{5}$-$\hat{1}$, with a cophenetic distance of .65 (.08 + .1 + .12 + .02 + .08 + .07 + .05 + .13). Shown in Table 5.2, these three cophenetic distances correspond closely to the dissimilarities estimated by the ratio model of .16, .25, and .68.

The appeal of dendrograms based on additive clustering algorithms is generally twofold. First, they organize items into groups (or clusters) on the basis of the features they share. In this case, $\hat{4}$-$\hat{5}$-$\hat{1}$ nests within all of the bass lines in red, whereas the bass lines in blue represent more remote variations of the $\hat{4}$-$\hat{5}$-$\hat{1}$ pattern, in which $\hat{4}$ descends to $\hat{2}$ before approaching $\hat{5}$. Second, additive trees implicitly represent prototypicality because they allow the path lengths between

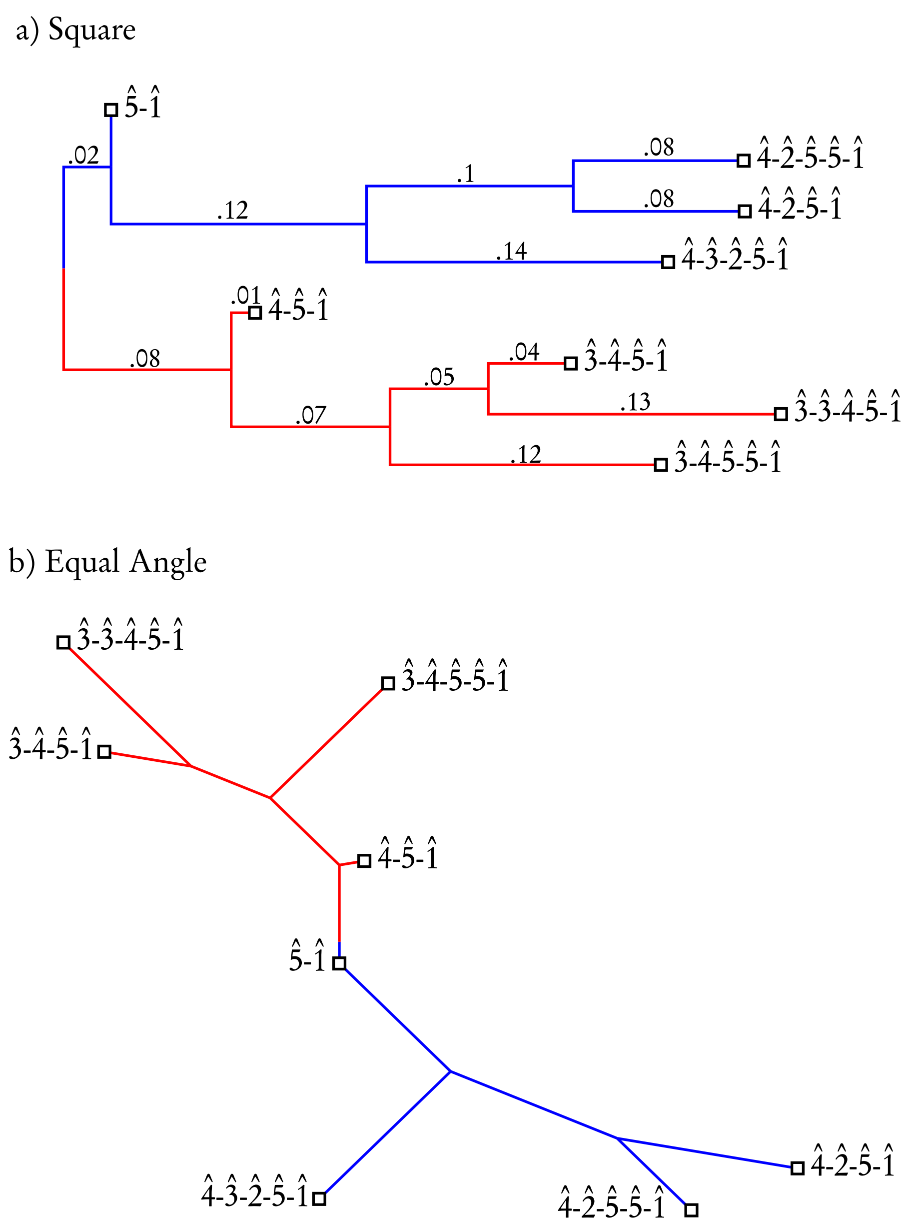


Figure 5.3: Square (a) and equal-angle (b) dendrograms calculated with the NJ algorithm for the sequence of chromatic scale degrees from eight authentic cadential bass lines in the Haydn Corpus.

items to vary such that each item's position reflects its average distance to the other items in the tree. For square dendrograms, the prototypicality of each bass line varies as a function of the horizontal position of the corresponding leaf, where bass lines appearing to the left are more prototypical than those appearing to the right. Thus, $\hat{5}$-$\hat{1}$ and $\hat{4}$-$\hat{5}$-$\hat{1}$ are the most prototypical patterns in each class, with $\hat{5}$-$\hat{1}$ serving as the central prototype for the entire tree. Given the initial assumption in the ratio model that prototypes nest within variants, this result is entirely expected, as both $\hat{5}$-$\hat{1}$ and $\hat{4}$-$\hat{5}$-$\hat{1}$ nest within the remaining bass lines in the tree, though $\hat{4}$-$\hat{5}$-$\hat{1}$ nests more directly for the bass lines in red.

One limitation of the square method is that it tends to obscure the cophenetic distances between items from different clusters. And since the length of the vertical lines connecting each pair of items (or clusters) is arbitrary, we could instead represent the tree in such a way as to preserve the cophenetic distance matrix calculated by the NJ algorithm while maximizing the angular distance between classes, thus omitting the vertical path lengths altogether. In this case, I have used the equal-angle algorithm, which starts from the root of the tree—again, at the point where the blue and red lines meet—and allocates arcs of angle to each subtree that are proportional to the number of leaves in it.[35] Shown in the lower tree in Figure 5.3, this approach clearly distinguishes the two classes without obscuring the distances between bass lines from different clusters, such as $\hat{5}$-$\hat{1}$ and $\hat{4}$-$\hat{5}$-$\hat{1}$. And remember that the same cophenetic distance matrix applies for both trees, so we could easily determine the precise values for the path lengths in the equal-angle tree by lifting the corresponding values from the square tree.

To determine how faithfully the dendrograms preserve the pairwise distances in the dissimilarity matrix, analysts sometimes compare the cophenetic distances with the original dissimilarity estimates using a Pearson correlation; in cluster analysis, this is called a *cophenetic*

[35] Felsenstein, *Inferring Phylogenies*, 578-580; Tobias H. Kloepper and Daniel H. Huson, "Drawing Explicit Phylogenetic Networks and their Integration into SplitsTree," *BMC Evolutionary Biology* 8, no. 22 (2008), doi:10.1186/1471-2148-8-22.

correlation. Because additive trees are not ultrametric—the leaves (shown with white boxes) are not equidistant from any one point on the tree—the cophenetic distance matrix generally corresponds much more closely with the original dissimilarity matrix than those derived from hierarchical clustering algorithms. Using hierarchical clustering with the complete linkage method, for example, the correlation between the cophenetic distance matrix and the original dissimilarity matrix is moderate, $r(28) = .61, p < .001$. For the cophenetic distance matrix computed by the NJ algorithm and represented by the additive trees in Figure 5.3, however, the fit is much better, $r(28) = .92, p < .001$.

5.3.2 Combining Viewpoints and Evaluation

With the similarity model and clustering algorithm now in place, we should be able to classify all of the cadences in the Haydn corpus on the basis of the features they share, and in so doing, identify the most prototypical exemplars in each class. But recall that the similarity model was applied separately to six different viewpoints—the sequences of contours and chromatic scale degrees characterizing each of the outer parts, and the sequences of metric positions and vertical interval class combinations characterizing the entire texture—under the assumption that the similarity relations characterizing a pair of cadences will reflect more than just one constituent viewpoint. To classify the cadences from the collection, the similarity matrices should therefore be combined in some way *before* applying the NJ method. But how do we determine which viewpoints to combine, and how do we combine them?

For classification problems like this one, researchers in fields like machine learning rely on *ensemble methods*, which combine individual models using a variety of techniques to (1) more faithfully represent the stimulus domain under investigation (in our case, the musical surface); and (2) improve model performance. The simplest method of combining similarity models is to compute the arithmetic mean of the similarity estimates for each pair of cadences, but doing so assumes that listeners weight each viewpoint equally when determining similarity. As an alternative, Pearce and Conklin have suggested weighting the arithmetic mean of the similarity matrices in the final model by the Shannon entropies associated with the corresponding viewpoints.[36] In their case, greater entropy (and hence uncertainty) was associated with a lower weight, but I will make the opposite assumption here.

In the ratio model, using the IDF weight assumes that two sequences will be more similar if they share very rare n-grams, so in this instance I have weighted the similarity matrices by the Shannon entropies of the corresponding viewpoints under the assumption that two cadences will be more similar if they share features from more uncertain (or unpredictable) viewpoints. Put another way, it is far more likely that a pair of cadences will share similar sequences of melodic contours than vertical interval class combinations because `contour` consists of just three distinct symbols, whereas `vintcc` consists of 190 symbols. By weighting each viewpoint by its Shannon entropy, viewpoints with larger alphabets, and thus, greater uncertainty, will receive higher weights in the final model.

Using the 3rd-order Shannon entropies derived from the Haydn collection, the viewpoints receive the following weights:

$$
\begin{aligned}
\texttt{contour}_{\text{vc}} &= 6.17\\
\texttt{csd}_{\text{vc}} &= 11.6\\
\texttt{contour}_{\text{vl1}} &= 6.04\\
\texttt{csd}_{\text{vl1}} &= 12.42\\
\texttt{strength} &= 6.95\\
\texttt{vintcc} &= 16.9
\end{aligned}
$$

With these weights, we could derive a composite similarity matrix that represents the weighted mean of *all* of the viewpoints. But given what little we know about the features

[36] Pearce, Conklin, and Wiggins, "Methods for Combining Statistical Models of Music," 302–304.

listeners actually employ in the service of similarity estimation, classification, and the like, it seems more reasonable to calculate every possible combination of models and select only that model which most closely corresponds to the labels from Caplin's typology. That is, it seems preferable to select the best model empirically, using Caplin's typology as a baseline (or *ground truth*). Of course, the labels themselves are subject to interpretation, and to employ them in this way is to assume that Caplin's typology is somehow relevant to how listeners represent, estimate, and classify cadences. But in the absence of experimental data, Caplin's typology provides us with a reasonable alternative.

Given six viewpoints, there are 63 possible combinations either of one viewpoint model, or of the weighted mean of two, three, four, five, or six viewpoint models. In what follows, I have selected the combination of viewpoint models that most closely corresponds with the labels from Caplin's typology using an external evaluation method from the fields of machine learning and information retrieval called the *weighted F-measure*.[37] The NJ algorithm was computed for all model combinations, and cluster validation was performed to find the five clusters that minimized the maximum dissimilarity between items in each cluster.[38]

Table 5.3 provides the confusion matrix and accuracy measures for the best performing model combination, which includes the weighted mean of all of the viewpoints except $\texttt{contour}_{\text{vl1}}$.[39] The x-axis of the confusion matrix represents the annotations from Caplin's typology, and the y-axis represents the model predictions. Thus, the diagonal represents correct classifications, and all other values represent confusions. Reading along the diagonal, the cluster analysis correctly classified 233 of the 245 cadences in the Haydn Corpus.

Recall captures what we often mean by *accuracy*. It represents the proportion of correctly

[37] Ricardo Baeza-Yates and Berthier Ribeiro-Neto, *Modern Information Retrieval* (New York, NY: ACM Press, 1999), 82.

[38] I used the very same cluster validation method to find the two clusters shown in red and blue in Figure 5.3.

[39] My approach here is technically *unsupervised*, so in this instance the confusion matrix is called a *matching matrix*, but I have elected to retain the more common term here.

Table 5.3: Confusion matrix and accuracy measures for the top model combination: $\texttt{contour}_{\text{vc}} \times \texttt{csd}_{\text{vc}} \times \texttt{csd}_{\text{vl1}} \times \texttt{strength} \times \texttt{vintcc}$.

		Annotation							
		PAC	IAC	HC	DC	EV	Precision	Recall	F-measure
Prediction	PAC	120	3	0	0	3	.95	.98	.97
	IAC	2	6	0	0	1	.67	.67	.67
	HC	0	0	84	0	0	1	1	1
	DC	0	0	0	17	1	.94	.89	.92
	EV	0	0	0	2	6	.75	.55	.63

classified items in each class, where the sum of the values in each column determines the number of items for that class. For the PAC category, for example, the NJ method correctly classified 120 of the 122 cadences in the Haydn Corpus. What is more, the confusion matrix tells us which other categories the model confused with the PAC category: in this case, the algorithm incorrectly classified two of the perfect authentic cadences in the IAC category.

For some categories, the model may be very accurate but not very precise, which is to say that it might classify many (or all) of the cadences from a given category into the same class, but also confuse cadences from other categories with that same class. *Precision* represents the proportion of correctly predicted items in each class, where the sum of the values in each row determines the number of model predictions for that class. For example, the NJ method correctly classified six of the nine imperfect authentic cadences in the Haydn Corpus, so its recall value is .67, but it also confused three other cadences with the IAC category, resulting in the lowest precision value in the model of .67.

To balance the recall and precision values characterizing a given model, the *F-measure* evaluates model performance by computing the weighted average of precision and recall. We could of course award higher or lower weights to precision and recall depending on the classification problem at hand, but in this case I have employed the *balanced F-measure*, which weights the two coefficients equally:

$$F\text{-}measure = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{5.5}$$

Based on the *F-measure* estimates, the NJ method classified cadences from the PAC and HC categories with the highest accuracy and precision, followed by the DC category, and finally the IAC and EV categories. For the purposes of model comparison, I have condensed these *F-measure* estimates into a single composite score that represents the arithmetic mean of the individual *F-measures* weighted by the number of cadences in each class. The idea here is to award higher weights to *F-measures* from classes with many items. As a result, the PAC and HC categories receive higher weights than the other categories because they represent the majority of the cadences in the corpus.

Table 5.4 presents the top ten model combinations based on the *balanced F-measure* weighted by the number of cadences in each class. The model described in Table 5.3 received a value of .95. The *Rand Index* is also often reported in classification tasks, so I have provided it here. It measures the similarity between two clustering solutions as the ratio of the correct classifications against all classifications.[40] From this list, it seems clear that the sequences of chromatic scale degrees in the outer parts play the most prominent role in the model's performance, as $\texttt{csd}_{vc}$ appears in every model and $\texttt{csd}_{vl1}$ appears in eight of the top ten models. Given the emphasis placed on the outer parts in Caplin's typology, this result seems quite reasonable, though

[40]The *Rand Index* represents each class by a 2×2 confusion matrix, in which the diagonal values represent the number of cadences that were correctly classified in that class (the *true positive* condition, or TP) and the number of cadences that were correctly classified in the other classes (the *true negative* condition, or TN). The remaining cells in the matrix refer to the *false positives* (or FP) and *false negatives* (or FN) for each class (i.e., the sum of the non-diagonal values for the row and column representing each class in the original matrix). For the PAC category, for example, TP is 120, TN is 132 (i.e., the sum of the remaining values along the diagonal in the original matrix), FP is 2, and FN is 6. The Rand Index sums the values from each cell of the 2×2 matrix for all of the classes in the model, and then divides TP + TN by TP + TN + FP + FN. If the original confusion matrix contains no confusions, the Rand Index is 1.

Table 5.4: Top ten model combinations based on the weighted F-measure of the cadence labels derived from Caplin's typology and the five clusters that minimized the maximum dissimilarity between cadences in each cluster in the tree.

	Model	Weighted F-measure	Rand Index	Total Correct
1.	$\texttt{contour}_{vc} \times \texttt{csd}_{vc} \times \texttt{csd}_{vl1} \times \texttt{strength} \times \texttt{vintcc}$	.95	.98	233
2.	$\texttt{contour}_{vc} \times \texttt{csd}_{vc} \times \texttt{csd}_{vl1} \times \texttt{strength}$	.92	.97	229
3.	$\texttt{contour}_{vc} \times \texttt{csd}_{vc} \times \texttt{csd}_{vl1} \times \texttt{vintcc}$	.92	.97	229
4.	$\texttt{contour}_{vc} \times \texttt{csd}_{vc} \times \texttt{contour}_{vl1} \times \texttt{csd}_{vl1}$	.91	.97	226
5.	$\texttt{csd}_{vc} \times \texttt{csd}_{vl1} \times \texttt{strength} \times \texttt{vintcc}$	.91	.97	228
6.	$\texttt{contour}_{vc} \times \texttt{csd}_{vc} \times \texttt{contour}_{vl1} \times \texttt{strength} \times \texttt{vintcc}$	.91	.97	228
7.	$\texttt{csd}_{vc} \times \texttt{vintcc}$	.91	.97	228
8.	$\texttt{csd}_{vc} \times \texttt{contour}_{vl1} \times \texttt{csd}_{vl1} \times \texttt{vintcc}$	.91	.97	227
9.	$\texttt{csd}_{vc} \times \texttt{contour}_{vl1} \times \texttt{csd}_{vl1} \times \texttt{strength} \times \texttt{vintcc}$	.90	.97	225
10.	$\texttt{csd}_{vc} \times \texttt{csd}_{vl1} \times \texttt{strength}$	.90	.97	227

it is also noteworthy that *all* of the viewpoints appear in at least four of the top ten model combinations, suggesting `csd` is not the only relevant viewpoint in Caplin's typology.

§5.4 The Cadential "Tree of Life"

Figure 5.4 presents the equal-angle dendrogram calculated with the NJ algorithm for the cadences from the Haydn Corpus. I partitioned the cadences in the left tree using the cluster validation method described earlier, but in the right tree I have clustered the cadences into the five categories from Caplin's typology. The PAC and IAC categories appear at the top of the trees in blue and green, the HC category appears at the bottom of the trees in red, and the DC and EV categories appear on the right side of the trees in magenta and yellow.

Using the weighted mean of the dissimilarity matrices from the viewpoints in the top model combination, the NJ method organized the cadences on the basis of the features they shared.

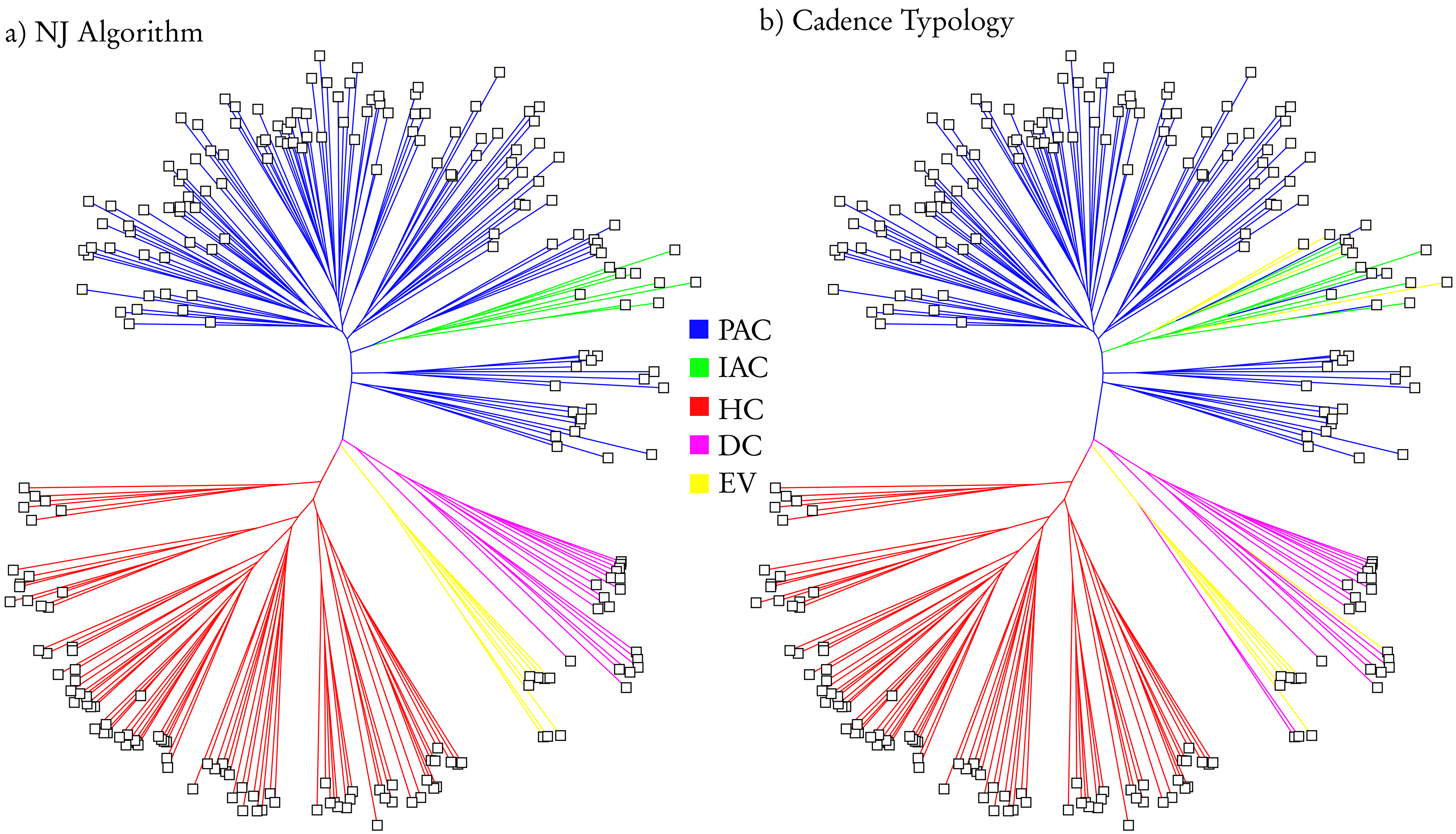


Figure 5.4: Equal-angle dendrogram calculated from the top model combination—$\texttt{contour}_{vc} \times \texttt{csd}_{vc} \times \texttt{csd}_{vl1} \times \texttt{strength} \times \texttt{vintcc}$—with the NJ algorithm for the cadences from the Haydn Corpus. a) Partitioned into the five clusters that minimized the maximum dissimilarity between cadences in each cluster. b) Partitioned into the five cadence categories from Caplin's typology.

As a result, cadences from categories displaying similar characteristics appeared closer in the tree. Cadences from the IAC category appeared as a subordinate branch of the PAC category at the top of the tree, cadences from the DC and EV categories appeared together on the right side of the tree, and cadences from the HC category, which share few characteristics with the other categories in Caplin's typology, were isolated at the bottom of the tree. To gain a more complete understanding of the similarity relations characterizing the cadences in each category, we might therefore partition the final tree into three sub-trees: the authentic cadences at the top of the tree, the cadential deviations at the right of the tree, and the half cadences at the bottom of the tree.

5.4.1 Authentic Cadences

The authentic cadence sub-tree features 135 cadences: 122 cadences from the PAC category, 9 cadences from the IAC category, and finally 4 cadences from the EV category. The NJ method clustered the cadences into six branches in the sub-tree. Counting clockwise from the left-most branch, the fourth branch consists primarily of imperfect authentic cadences, so we might expect the remaining branches to reflect other pertinent sub-types of the authentic cadence.

Previous attempts to subdivide the authentic cadence have relied upon a number of distinguishing characteristics: the final scale degree in the soprano ($\hat{1}$ or $\hat{3}$), the presence of a dissonant suspension above the cadential dominant or at the moment of cadential arrival, and the length of the cadential progression have all been suggested at one time or another. According to Giorgio Sanguinetti, Haydn's Neapolitan contemporaries typically classified the authentic cadence into three types according to the number of metrical units supporting the cadential dominant: the *cadenza semplice*, or simple cadence, which consists of just one unit (e.g., V–I, ii^6–V–I, etc.); the *cadenza composta*, or compound cadence, which consists of two units, and thus supports a cadential six-four (e.g., V^6_4–V^7–I); and finally the *cadenza doppia*, or double cadence, which consists of four units (e.g., V^5_3–V^6_4–V^5_4–V^5_3–I), though this particular bass pattern featured less prominently in the late eighteenth century.[41] But because the partimento training many composers received during this time emphasized a two-voice framework of soprano and bass,[42] Gjerdingen additionally subdivides the authentic cadence according to the behavior of the soprano voice: the descending stepwise *Mi-Re-Do* and its variant, the descending hexachord; the octave descent of the *Cudworth* cadence; and the prolonged repetition of $\hat{1}$ and $\hat{3}$ characterizing the *Pulcinella* all provide ready examples, not to mention of course the imperfect authentic (or *incomplete*) cadence, which remains a mainstay of most cadence typologies.[43]

Shown in Figure 5.5, the authentic cadence sub-tree tends to subdivide the cadences according to the final event characterizing each viewpoint in the final model combination. Reading clockwise starting on the left, the first two branches represent those cadences employing the standard cadential bass, $\hat{3}$–$\hat{4}$–$\hat{5}$–$\hat{1}$, that feature either a final ascending or final descending contour from $\hat{5}$ to $\hat{1}$. Within both branches, the NJ method then clustered the cadences according to the presence of a cadential six-four. Thus, the cadential bass descends from $\hat{5}$ to $\hat{1}$ for the cadences in the first branch, with the first nine leaves featuring semplice cadences and the interior leaves featuring composta cadences. For the second branch in the sub-tree, however, the cadential bass ascends from $\hat{5}$ to $\hat{1}$, with nearly three quarters of the leaves featuring composta cadences.

The third branch of the authentic cadence sub-tree represents cadences with bass lines either

[41] Sanguinetti, *The Art of Partimento*, 105-106.

[42] Musicologists Felix Diergarten and James Dack have both suggested that Haydn, then in his twenties, was very likely to have received partimento training by the composer Nicola Porpora, who before retiring to Vienna in 1752 at the age of 66, studied and taught at the Conservatorior dei Poveri di Gesù Cristo in Naples for nearly three decades. During his time with Porpora, Haydn learned to realize an unfigured bass and accompany a singer at the harpsichord, and since biographical evidence suggests that Haydn preferred to compose at the keyboard, it is possible that his partimento training, however brief, may have imparted the lexicon of Galant schemata from the "Italianate style" ("'The True Fundamentals of Composition': Haydn's Partimento Counterpoint," *Eighteenth Century Music* 8, no. 1 [2011]: 53–75; "Sacred Music," in *The Cambridge Companion to Haydn*, ed. Caryl Clark [Cambridge: Cambridge University Press, 2005], 141).

[43] Following Johann Gottfried Walther, Gjerdingen refers to the perfect authentic cadence as the *clausula formalis perfectissima*, or "most complete close" (*Music in the Galant style*, 139).

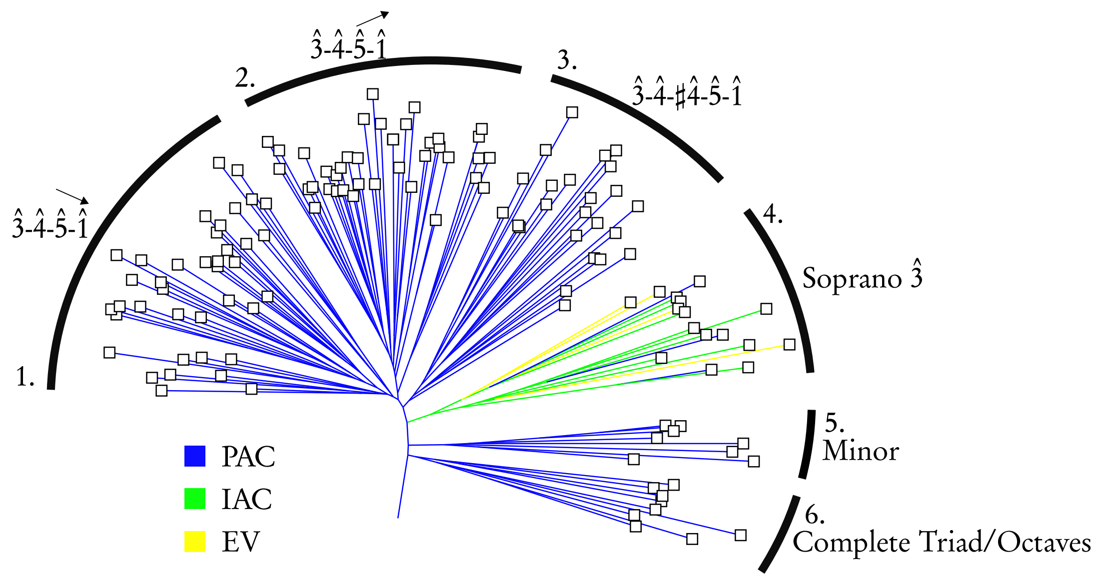


Figure 5.5: Equal-angle dendrogram calculated from the top model combination—$\texttt{contour}_{vc}$ × $\texttt{csd}_{vc}$×$\texttt{csd}_{vl1}$×`strength`×`vintcc`—for the authentic cadence sub-tree. Each cadence was clustered into the categories from Caplin's typology.

that lack the initial ascent to $\hat{5}$ (e.g., $\hat{5}$–$\hat{1}$ or $\hat{1}$–$\hat{5}$–$\hat{1}$), or that embellish the standard bass with scale degrees like $\sharp\hat{4}$ (e.g., $\hat{4}$–$\sharp\hat{4}$–$\hat{5}$–$\hat{1}$). Sixteen of the twenty four cadences in this branch also feature *expanded cadential progressions* (i.e., harmonic progressions spanning a complete phrase).[44] In fact, the first cluster of eight leaves in branch three represent exemplars of a schema Gjerdingen has called the *Grand*, in which a composta bass supports the melodic sequence, $\hat{1}$–$\hat{6}$–$\hat{5}$–$\hat{2}$–$\hat{1}$.[45] As with all of Gjerdingen's schemata, the *Grand* is often subjected to extensive elaboration—indeed, particularly so for this schema, since it typically effects closure at the ends of movements or larger sections.

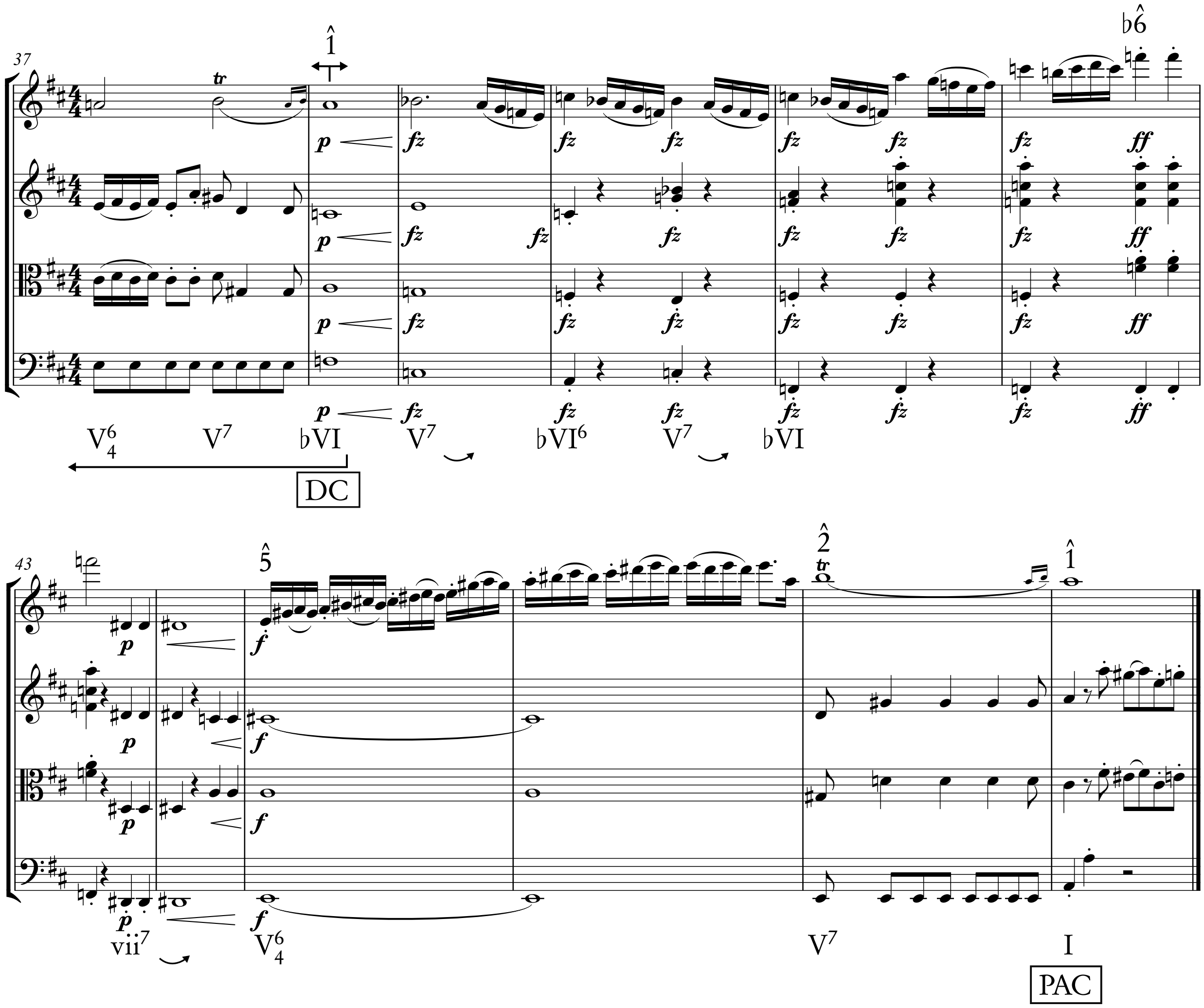
Selected from the first cluster of leaves in branch three, one particularly noteworthy instance of Gjerdingen's *Grand* cadence appears at end of the exposition in the first movement of Haydn's 'Frog' Quartet, Op. 50, No. 6 (see Example 5.3). In Gjerdingen's scheme, the *Grand* features

[44] Caplin, *Classical Form*, 254.
[45] Gjerdingen, *Music in the Galant style*, 152.

the standard composta bass, and the theme's first attempt at the final cadence features just such a bass beginning in m. 34, but at the moment of cadential arrival, the progression resolves deceptively to the flattened submediant in m. 38. Like so many other cadential deviations within sonata form, this deceptive gesture would seem to heighten the expectation for authentic cadential closure at the end of the exposition, and so in such cases, the subsequent passage often features a continuation of the previous cadential process. In this case, however, the events in m. 38 initiate a rather unexpected prolongation of ♭VI, thereby eliding the preceding cadential progression with the subsequent prolongational passage. Nevertheless, Haydn quickly rights the ship by replacing the flattened submediant with $\sharp\hat{4}$ in the bass in m. 43, and what began as an unexpected exploration of a new tonal region soon culminates in yet another expanded cadential progression, and finally, a perfect authentic cadence to close the subordinate theme in m. 48.

Despite extensive diminutions, the *Grand's* underlying two-voice scaffold remains essentially intact in this example, particularly so in the soprano voice, where each focal scale degree appears in a metrically strong position—$\hat{1}$ in m. 38, $\flat\hat{6}$ in m. 42, $\hat{5}$ in m. 45, $\hat{2}$ in m. 47, and finally $\hat{1}$ in m. 48. From this point of view, the appearance of ♭VI in m. 38 serves both as a deceptive resolution for the preceding cadential progression and as a tonic substitute for another expanded cadential progression, with the initial scale degrees, $\hat{3}$ and $\hat{4}$, of the composta bass replaced by $\flat\hat{6}$ and $\sharp\hat{4}$, resulting in a chromatic variation of Gjerdingen's schema.

Cadences like this one are indeed quite common in the corpus, though Haydn sometimes either substitutes the initial melodic events of the *Grand* for other chord members, or omits them altogether. In such cases, only the final three events of the schema remain: the $\hat{5}$–$\hat{2}$–$\hat{1}$ pattern in the soprano, and the composta bass that supports it.[46]

[46]See, for example, Op. 54/1, ii, mm. 46–52; Op. 17/3, iv, mm. 22–24; or Op. 33/4, i, mm. 25–26. The cadence in mm. 26–34 of the second movement of Op. 55/1 is particularly noteworthy in this regard. The expanded cadential progression begins with a lament bass, but upon reaching $\hat{5}$, reverts to the composta and presents the

Example 5.3: An example of Gjerdingen's *Grand* cadence: Haydn, String Quartet in D, Op. 50/6, i, mm. 37–48.

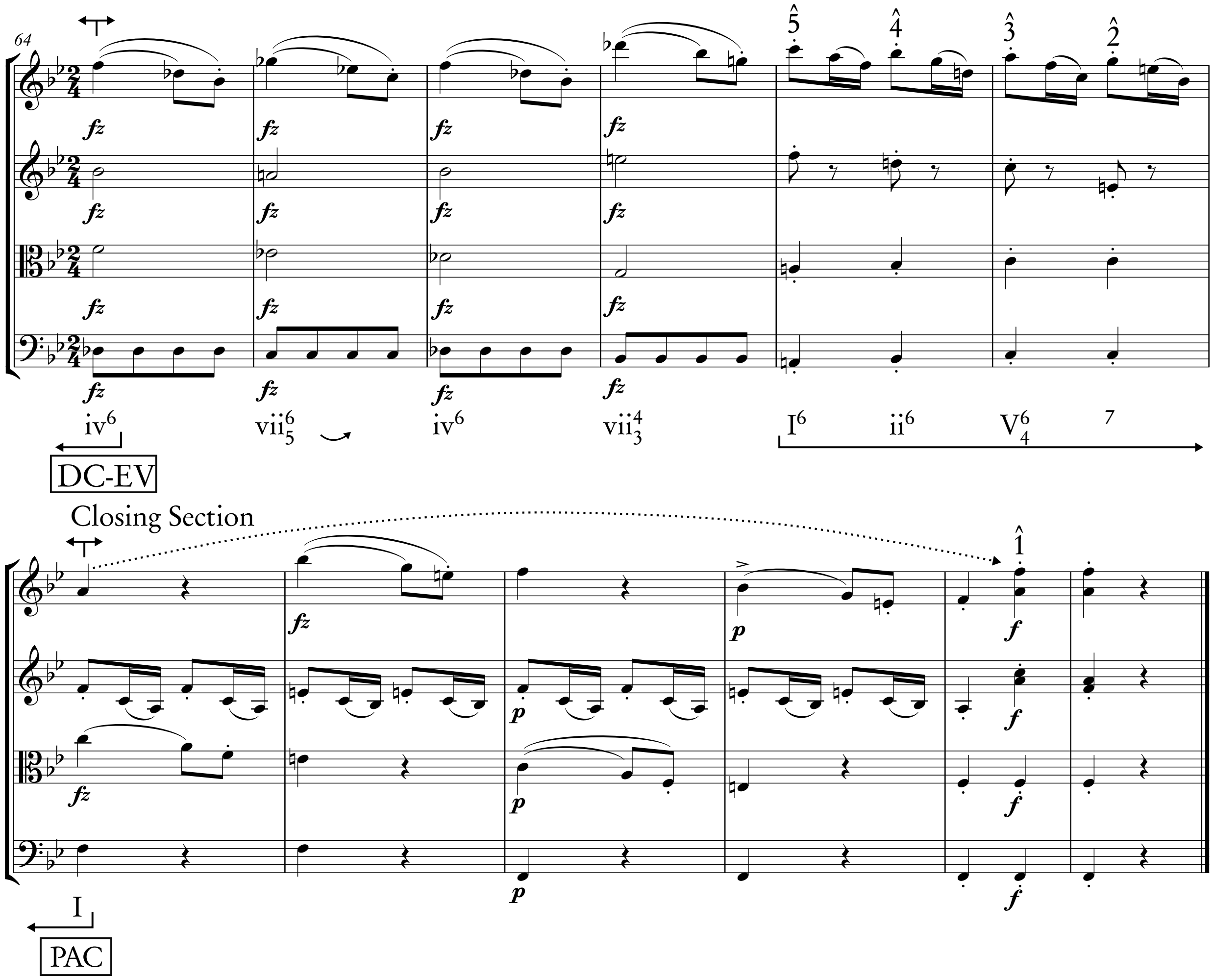
As mentioned previously, the cadences in branch four consist of authentic cadential progressions culminating in $\hat{3}$ in the first violin at the moment of cadential arrival. Thus, all nine of the cadences from the IAC category appear in branch four, but another seven cadences from the PAC and EV categories in Caplin's typology also appear in this branch. Given the restrictions placed on the similarity algorithm—including only those *n*-grams whose terminal event appears

$\hat{5}$–$\hat{2}$–$\hat{1}$ pattern in the soprano.

at the end of the sequence—these model confusions should not be surprising. Remember that the NJ method clustered the cadences on the basis of the features they shared. As a result, it successfully clustered all of the imperfect authentic cadences in the branch. But by examining such a limited number of (primarily syntactic) viewpoints, it also necessarily ignored several of the rhetorical characteristics that contribute to Caplin's typology, particularly for cadences featuring parametric noncongruence, where harmonic-melodic content alone is not always sufficient for cadential identification. In other words, for particularly problematic cases, the method sometimes struggled.

Shown in Example 5.4, the cadence closing the subordinate theme in the final movement of Haydn's Op. 50/1 appears in branch four because it contains an authentic cadential progression with a composta bass and $\hat{3}$ in the first violin at the moment of cadential arrival. Like the *Grand* cadence in the previous example, it begins with the tonicization of a tonic substitute following a cadential deviation in m. 64, in this case tonicizing the minor-mode subdominant rather than ♭VI.[47] In m. 68, iv^6 gives way to the authentic cadential progression that closes the subordinate theme, but $\hat{3}$ appears in place of $\hat{1}$ in the first violin in order to resolve the dissonant seventh from the preceding measure.

To further complicate matters, the cadential idea in the first violin presents not one but two voices in an otherwise monophonic texture: a *soprano* voice stepwise descent from $\hat{5}$ to $\hat{2}$, and

[47]In Caplin's theory, the deceptive and evaded categories differ primarily as a consequence of the grouping structure expressed by their final events at the cadential arrival. If those events group backward with the preceding cadential process, the cadence is deemed *deceptive*, but if they group forward with the subsequent passage, the cadence is said to *evade* the expected resolution. In this case, the behavior of the outer voices is consistent with a deceptive cadence, with the melody and bass resolving to $\hat{1}$ and $\flat\hat{6}$, respectively, but the change in texture and the increase in dynamics and surface activity accompanying these scale degrees indicate that the events at the downbeat of m. 64 may group forward rather than backward. Examples like this one point out the difficulties involved in making categorical distinctions about cadential deviations, since several musical features play some role in the perception of segmental grouping boundaries. Indeed, the distinction between deception and evasion in Caplin's typology has been criticized for precisely this reason (Nicholas Marston, Review of *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven*, by William E. Caplin, *Music Analysis* 20, no. 1 [2001]: 146). As I mentioned at the beginning of this section, particularly problematic cases like this one received two labels in the Haydn Corpus, and were thus omitted from the tree analysis presented here.

Example 5.4: A perfect authentic cadence clustered in the IAC branch of the authentic cadence sub-tree: Haydn, String Quartet in B-flat, Op. 50/1, iv, mm. 64–75.

an *inner voice* that initially doubles the second violin part before presenting (and resolving) the dissonant seventh. As a consequence, the model could not distinguish the immediate resolution of the inner voice from the delayed resolution of the soprano voice a few measures later. Of course, Haydn could have resolved the cadential dominant *and* realized the stepwise descending melody of the cadential idea simply by placing a double stop F-A dyad at the moment of the cadential arrival—indeed, that very dyad appears at the end of the exposition a few measures later. Instead, he elected to omit $\hat{1}$ from the first violin in order to foreground the melodic-motivic material appearing in the viola on the downbeat of m. 70, thus eliding the preceding cadence with the beginning of the closing section. And since the C_5 in the viola's upper register represents the highest note in the texture at the moment of cadential arrival, placing a double-stop in the first violin would have relegated the melodic-motivic material to the background. Haydn's choice to foreground the beginning of the closing section by omitting $\hat{1}$ altogether thereby deferred the expected melodic resolution of the perfect authentic cadence to the interior of the closing section, either at the beginning of the second codetta in m. 72, or at the end of the exposition in m. 75.

Borrowing a term from Leonard Meyer, the cadence in m. 70 is thus only *provisionally realized* at the moment of cadential arrival, with the expected melodic resolution arriving several measures later.[48] But since the encoded cadential idea concludes with $\hat{3}$ in m. 70, the NJ method classified this cadence in branch four with the other imperfect authentic cadences. Whether this cadence is indeed perfect authentic, imperfect authentic, deceptive, or evaded depends on the perceived strength and temporal function of the cadential boundary, which can serve either as a *beginning*, an *end*, or some combination of the two.[49] At the very least, the thematic elision and the omission of $\hat{1}$ at the cadential arrival have seriously weakened the cadence, thus putting the perfect authentic interpretation in doubt, but given the presence of a complete authentic cadential progression culminating in tonic harmony in root position, the caesura appearing in beat two of m. 70 in the outer parts, and the increase in surface rhythmic activity in the second violin, *deferred* or *separated* PAC seems the most convincing label in this case.[50]

[48] Meyer, *Explaining Music*, 117.

[49] In the case of abandoned cadences in Caplin's typology, the events at the expected cadential arrival can also serve as a *middle*, which is to say that no boundary exists (see Table 2.1).

[50] Mark Richards has called instances like this one *separated* cadences, where the cadential arrival may be dispersed over a span of music when the parameters responsible for articulating cadential closure are temporally mis-aligned ("Closure in Classical Themes," 35-37). Whether listeners experience a partial arrival when the bass resolves and a more complete arrival when melodic closure obtains is very difficult to judge. Clearly the perception of closure depends on whether, and to what degree, the many parameters of a musical work align at a given point in time: the longer the time interval between the various parameters effecting cadential closure, the more

Nevertheless, the NJ method classified this cadence in branch four with the other imperfect authentic cadences precisely because it contained the requisite syntactic characteristics to justify its inclusion in the IAC category.[51]

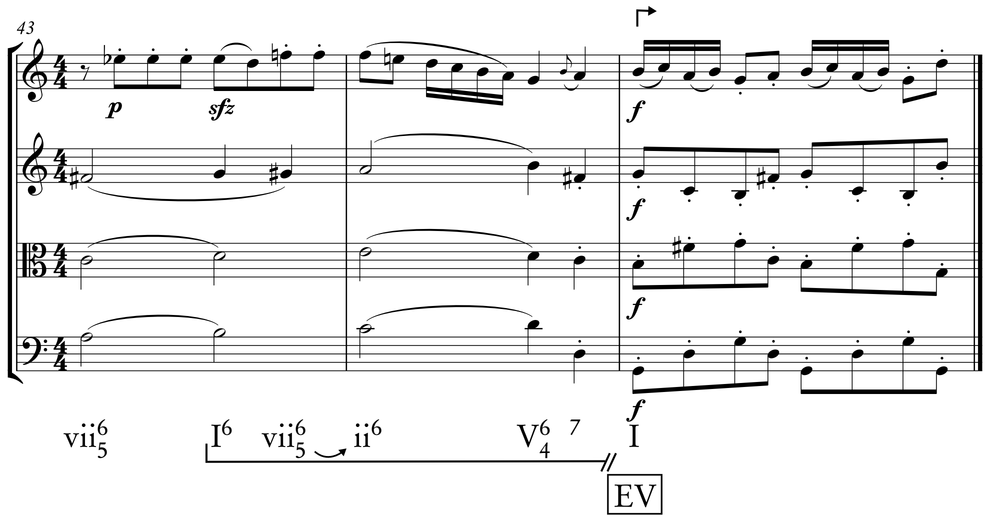
In addition to the *deferred* or *separated* perfect authentic cadences appearing in branch four,[52] the NJ method placed four evaded cadences in this branch as a consequence of the final events they shared. Unlike the other categories in Caplin's typology, evaded cadences may feature quite literally *any* scale degree in the outer voices and *any* harmony characterizing the entire texture—including $\hat{1}$ in the soprano supported by tonic harmony in root position—so long as the final events at the expected moment of cadential arrival group forward with the subsequent passage, and not backward with the preceding cadential process. For this reason, cadences from the evaded category appear in three different branches of the tree (see Figure 5.4), thereby reflecting three different categories: imperfect authentic, deceptive, and evaded.

Just as in the previous example, the evaded cadences in branch four feature an authentic cadential progression and $\hat{3}$ in the soprano at the cadential arrival. Shown in Example 5.5, the evaded cadence in m. 45 of the opening movement of Haydn's Op. 74/1 brings a composta bass and a compressed variant of the Cudworth in the soprano, but $\hat{3}$ replaces $\hat{1}$ at the expected moment of cadential arrival, a form of melodic deception that often appears—along with the

temporally remote the associations a listener must form, and thus, the greater the burden placed on attention, memory, and so forth. In Caplin's theory of form, harmony is the essential characteristic underlying (formal) closure, so for instances in which the two principal voices of a cadence achieve closure at different points in time, he tends to privilege the onset of the final harmony of the cadential progression as the clearest signal for cadential closure. There are of course very good reasons to privilege the selective attention to, or "partial hearing" of, a given syntactic parameter like harmony in a theory of closure, but experimental evidence would suggest that listeners depend on a great many parameters when determining the strength and status of a given ending. Chapter 7 attempts to tease out how these many parameters engender closure in the classical style.

[51]Since I omitted cadences if the cadential progression did not appear in the cello part, one could argue that I also could have omitted cadences if the first violin did not bring the soprano voice of the cadential idea, as is the case in Example 5.4. As Meyer points out, attempts to sharpen our concepts and categories into algorithmic form is a useful exercise "not merely because they can save enormous amounts of time but, equally important, because their use will force us to define terms and traits, classes and relationships, with precision" (Meyer, *Style and Music: Theory, History, and Ideology*, 64).

[52]See also Op. 17/4, i, mm. 45–53.



Example 5.5: An evaded cadence clustered in the IAC branch of the authentic cadence sub-tree: Haydn, String Quartet in C, Op. 74/1, i, mm. 43–45.

other cadential deviations—in the subordinate theme of sonata-form expositions. To be sure, the classification here is not altogether unreasonable, since analysts sometimes interpret cadences like this one as melodically deceptive imperfect authentic cadences.[53] Nevertheless, the sudden change in dynamics, texture, and surface activity suggest that the events at the implied cadential arrival group forward rather than backward, and thus, despite the presence of a complete authentic cadential progression and $\hat{3}$ in the soprano at the cadential arrival, the cadence was classified in the EV category in the Haydn Corpus.

Whether the cadence in Example 5.5 *is* evaded is not really the point here. Certainly this interpretation determines how we evaluate the model's performance, but our goal is to understand how the ratio model and NJ method—given a specific but limited set of viewpoints—actually classify this and every other cadence in the corpus. Clearly Caplin's cadence typology, or more specifically, the cadential annotations specific to this corpus, depend(s) on some assessment of the rhetorical features surrounding the moment of cadential arrival (i.e., on

[53]Neuwirth, "Fuggir la cadenza, or the Art of Avoiding Cadential Closure," 117–130.

features characterizing the perceived grouping structure), and so any discrepancies between the model predictions and the cadential annotations will necessarily result from the omission of these features in the present approach.

Finally, branch five consists of cadences in the minor mode, while branch six presents cadences for which the final sonority in `vintcc` features a complete triad or doubles $\hat{1}$ (i.e., $< 4, 7, \perp>$ or $< 0, \perp, \perp>$). For all of the other cadences in the authentic cadence sub-tree, the final harmony at the cadential arrival consisted only of $\hat{1}$ and $\hat{3}$ (i.e., `vintcc` $= < 4, \perp, \perp>$), and so the NJ method clustered cadences separately in branch six if they did not feature this final sonority.

Whether these cadences constitute a sub-type of the authentic cadence depends of course on the viewpoints employed in service of similarity estimation and classification. For many, the presence of certain scale degrees over others within a given sonority would seem inconsequential so long as the harmony and inversion remain essentially intact. From this point of view, the failure of the NJ method to integrate these cadences into the other branches of the tree reflects two limitations of the current approach: (1) that cadences are compared *entirely* on the basis of *n*-grams whose final event terminates at the end of the sequence, when cadences may otherwise share *n*-grams located at other positions within the sequence; and (2) that `vintcc` does not always reflect the figured bass symbols or Roman numeral annotations made by analysts, resulting in potentially unnecessary distinctions regarding the final sonorities in each cadence.

All of this is to say that the potential sub-types of the authentic cadence sub-tree reflected in Figure 5.5 do not always correspond with the sub-types offered in the *Formenlehre* tradition. Although many of the smaller clusters located at the furthest extremes of the sub-tree reflect various sub-types of the authentic cadence—the *semplici* cadences in the first cluster of branch one, the *composte* cadences in the largest cluster of branch two, or the *Grand* cadences in the first cluster of branch three—the predominant branches of the sub-tree subdivided the cadences according to the contour between the final two scale degrees in the bass (branches one and two), the presence of chromatic scale degrees in the standard bass (branch three), the presence of $\hat{3}$ in the soprano at the moment of cadential arrival (branch four), and the mode of the entire cadence or final harmony of the cadential arrival (branches five and six). To improve model performance for the authentic cadences in the Haydn Corpus, reducing the syntactic domain of `vintcc` and including *n*-grams from other positions in the sequence would be a useful starting point. Nevertheless, as we will see shortly, applying the NJ method to the best model combination may not have resulted in particularly clear sub-types for the authentic cadence, but it produced very clear sub-types for the half cadence category.

Prototypicality Reconsidered: *Nearest Prototype Analysis*

Before turning to the other sub-trees, we might first consider another issue explored by the ratio model: prototypicality. Because the ratio model and NJ method classify the cadences on the basis of the features they share, some cadences within each class will be more prototypical than others, which is to say that some cadences will share more features specific to the class than others. Presumably, finding the most prototypical exemplars in each branch—or in the sub-tree as a whole—may indicate which cadences listeners are most likely to learn and remember, or at the very least, which cadences listeners are the most likely to associate with that branch. For equal-angle additive trees, for example, prototypical cadences tend to appear closer to the middle of the tree because they are more similar to *all* of the other cadences in the tree than cadences appearing at the furthest extremes of the tree.

To classify these cadences, recall that the NJ method requires a symmetric matrix where 0 represents identity (i.e., maximal similarity). To that end, I employed a data reduction method that eliminates the asymmetry between cadences (i.e., $1 - \frac{\delta(a,b)+\delta(b,a)}{2}$). But since asymmetries in the ratio model imply nested relations where one item serves as a prototype for the other, eliminating those asymmetries in the matrix impairs prototypicality estimation in the resulting tree representation. Luckily, prototypicality estimation does not require a symmetric matrix, so we can return to the asymmetric matrices calculated by the ratio model and compute the weighted average of the same viewpoints from the best-performing model combination in the classification task ($\texttt{contour}_{\mathrm{vc}} \times \texttt{csd}_{\mathrm{vc}} \times \texttt{csd}_{\mathrm{vl1}} \times \texttt{strength} \times \texttt{vintcc}$).

Given the ratio model presented here, there are a number of ways to calculate prototypicality. More prototypical cadences will receive higher similarity values on average in the referent position than less prototypical cadences, for example, which is calculated as the arithmetic mean of the similarity values appearing in each column of the matrix.[54] More prototypical cadences will also exhibit higher asymmetries than less prototypical cadences, so we could alternatively represent prototypicality as the sum of the differences between $\delta(a, b)$ and $\delta(b, a)$ in each column.

In what follows, I have borrowed and extended a technique from Tversky and J. Wesley Hutchinson called *nearest neighbor analysis*.[55] In their formulation, item b is the nearest neighbor of item a if $\delta(a, b) > \delta(a, \Lambda)$ for all Λ in the matrix. But since b also serves as a *prototype* for a in this model if $\delta(a, b) > \delta(b, a)$, we might alternatively refer to b as the *nearest prototype* of a if $\delta(a, b) > \delta(b, a)$ and $\delta(a, b) > \delta(a, \Lambda)$. Following this procedure, we can determine the nearest prototype of each cadence in the matrix, and then simply count the number of times each nearest prototype appears. The cadence that serves as the nearest prototype most often is the most prototypical, followed by the next most common nearest prototype, and so on.

This technique works very well for the descriptive measures proposed by Tversky and Hutchinson to characterize the similarity matrix as a whole,[56] but it tends to ignore more

[54] I have just described Tversky's equal-weighted prototypicality equation ("Features of Similarity," 347-348).

[55] Amos Tversky and J. Wesley Hutchinson, "Nearest Neighbor Analysis of Psychological Spaces," *Psychological Review* 93, no. 1 (1986): 3–22.

[56] Their principal contribution is a diagnostic measure called *centrality*, which measures the degree to which the items of a given similarity matrix share a nearest neighbor. Those matrices characterized by just a few distinct

remote prototypicality relationships that could alter the final count, and thus, improve the prototypicality estimate. The chromatic scale degree sequence $\hat{4}$–$\hat{2}$–$\hat{5}$–$\hat{1}$ might serve as a more immediate prototype for $\hat{4}$–$\hat{2}$–$\hat{5}$–$\hat{5}$–$\hat{1}$ than $\hat{4}$–$\hat{5}$–$\hat{1}$, for example, but $\hat{4}$–$\hat{5}$–$\hat{1}$ is more likely to serve as a prototype for the majority of the cadential bass lines in the corpus. In this case, $\hat{4}$–$\hat{5}$–$\hat{1}$ lies beyond the nearest neighbor, but if we adjusted the final count by considering *all* of the potential prototypes for each cadence, the prototypicality estimate for $\hat{4}$–$\hat{5}$–$\hat{1}$ would necessarily increase by a much wider margin than a less prototypical bass like $\hat{4}$–$\hat{2}$–$\hat{5}$–$\hat{1}$.

Given an $n \times n$ matrix of similarity estimates, the approach employed here orders all of the potential prototypes b from most to least prototypical for each cadence a, and then each prototype receives a prototypicality score corresponding to its ordinal rank, with the nearest prototype receiving the score n, the next nearest prototype receiving the score $n-1$, the next next nearest prototype receiving the score $n-2$, and so on; any remaining cadences do not serve as prototypes and so receive a score of 0. Using the resulting $n \times n$ matrix of prototypicality scores, the final prototypicality estimate is

$$proto(a, \Lambda) = \frac{\sum scores(\Lambda, a)}{n(n-1)}, \tag{5.6}$$

where $proto(a, \Lambda)$ denotes the prototypicality of item a with respect to class Λ. In this formulation, $proto$ represents the ratio of the sum of the prototypicality scores from column a of the matrix to the maximum possible prototypicality estimate, with values falling in a range between 0–1.[57] In this approach, the prototypicality estimate for any given cadence depends on the other cadences in the class. A cadence from branch one may be very prototypical when compared to the other cadences in that branch, for example, but much less prototypical when

nearest neighbors (i.e., items with high prototypicality) will demonstrate high *centrality*.

[57] An item receiving the maximum prototypicality estimate would serve as the nearest prototype for every item in the matrix except itself, hence the denominator $n(n-1)$.

Table 5.5: Prototypicality estimates for the exemplars from the first three branches of the authentic cadence sub-tree, branches one and two and branch three of the cadential deviations sub-tree, and the three branches of the half cadence sub-tree.

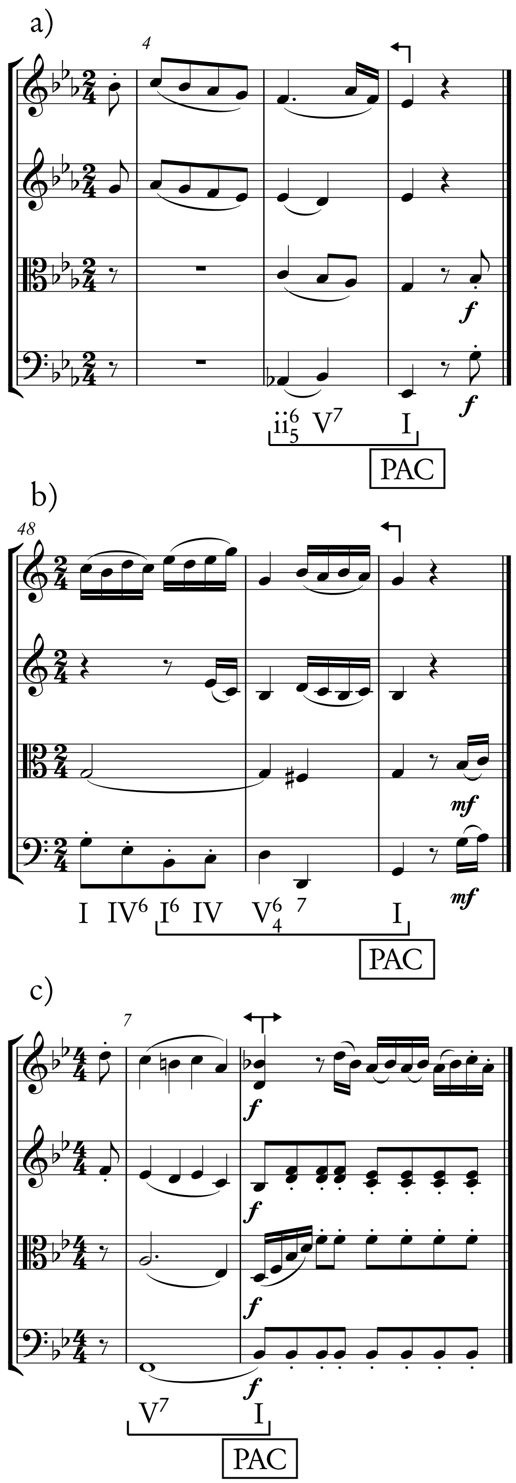
Branch Number	Excerpt	*proto* Sub-tree	*proto* Branch
Authentic Cadences			
1	Op. 20, No. 1, iv, mm. 4–6	.78	.90
2	Op. 50, No. 2, iv, mm. 48–50	.86	.86
3	Op. 71, No. 1, i, mm. 7–8	.89	.84
Cadential Deviations			
1/2	Op. 20, No. 4, i, mm. 22–24	.79	.90
3	Op. 74, No. 1, ii, mm. 30–33	.40	.89
Half Cadences			
1	Op. 17, No. 2, i, mm. 19–20	.62	.77
2	Op. 54, No. 3, i, m. 4–5	.92	.96
3	Op. 55, No. 3, i, mm. 5–8	.41	.83

compared with all of the cadences in the sub-tree, or indeed, in the tree as a whole.

Table 5.5 and Example 5.6 present the prototypicality estimates and musical examples for the most prototypical cadences from the first three branches of the authentic cadence sub-tree, which I will hereafter call *exemplars*. Shown in Example 5.6a, the perfect authentic cadence closing the main theme in the finale of Haydn's Op. 20, No. 1 received the highest prototypicality estimate from the cadences in branch one, with a *proto* of .90 for the cadences within the branch and .78 for the cadences in the authentic cadence sub-tree. It consists of a semplice bass supporting a stepwise descending hexachord in the soprano. Gjerdingen points out that such broad descents are not always in a fixed relationship with the bass, as is the case here.[58]

The exemplar from branch two appears in mm. 48–50 in the finale of Op. 50, No. 2 (see Example 5.6b). Unlike the previous example from branch one, it consists of a composta

[58]Gjerdingen, *Music in the Galant style*, 144.



Example 5.6: Exemplars from the first three branches of the authentic cadence sub-tree. a) Branch 1: String Quartet in E-flat, Op. 20/1, iv, mm. 4–6; b) Branch 2: String Quartet in C, Op. 50/2, iv, mm. 48–50; c) Branch 3: String Quartet in B-flat, Op. 71/1, i, mm. 7–8.

bass, with the second $\hat{5}$ appearing an octave lower, a particularly common signal of authentic cadential closure within the subordinate theme. In this example, the melody ascends to $\hat{6}$ in m. 48 and then "drops" to $\hat{1}$ at the downbeat of m. 49. Such sudden drops often occur within the *cadenza composta* at the onset of the cadential six-four. Just as in the previous example, the cadential content is generally very compressed and features relatively little surface activity, but remember that the ratio model awards the greatest positive asymmetries to cadences featuring the smallest possible sequences that "nest within" sequences from the other cadences in the class. As such, the cadential bass lines in Example 5.6 serve as exemplars precisely because they do not embellish the sequences characterizing each viewpoint in the final model combination.

This is particularly true of the most prototypical cadence in branch three, which closes the main theme in the opening movement of Op. 71, No. 1. Shown in Example 5.6c, the bass lacks the initial ascent to $\hat{5}$, resulting in the extraordinarily compressed cadential progression, V^7–I. In fact, this cadence received the highest prototypicality estimate of all of the cadences in the sub-tree because it shares the most characteristic features of the authentic cadence—$\hat{5}$–$\hat{1}$ in the bass and a V^7–I progression—with nearly every other cadence in the tree.[59] But note here that it also received a *lower* prototypicality estimate for the cadences in its branch, as the majority of these cadences consisted of expanded cadential progressions featuring a composta bass and $\sharp\hat{4}$ between $\hat{4}$ and $\hat{5}$. In this case, branch three is the most heterogeneous of the first three branches in the sub-tree because it features cadences that either omit the initial ascent of the standard bass, as is the case for Example 5.6c, or embellish the standard bass in some way. As a result, the cadence in Example 5.6c is less likely to serve as the nearest prototype for the cadences featuring more complex diminutions from branch three than those cadences featuring less elaborated structures from branches one and two.

[59]Only 8 of the 131 authentic cadences in the Haydn Corpus do not feature a dissonant seventh above the penultimate dominant.

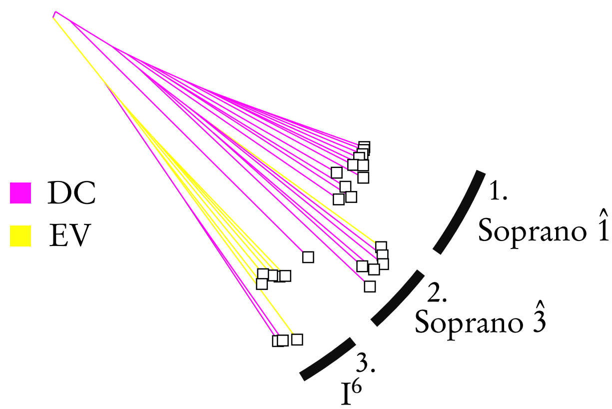


Figure 5.6: Equal-angle dendrogram calculated from the top model combination—$\texttt{contour}_{vc}$ × $\texttt{csd}_{vc}$ × $\texttt{csd}_{vl1}$ × `strength` × `vintcc`—for the sub-tree comprised of cadential deviations (i.e., cadences from the DC and EV categories). Each cadence was clustered into the categories from Caplin's typology.

5.4.2 Cadential Deviations

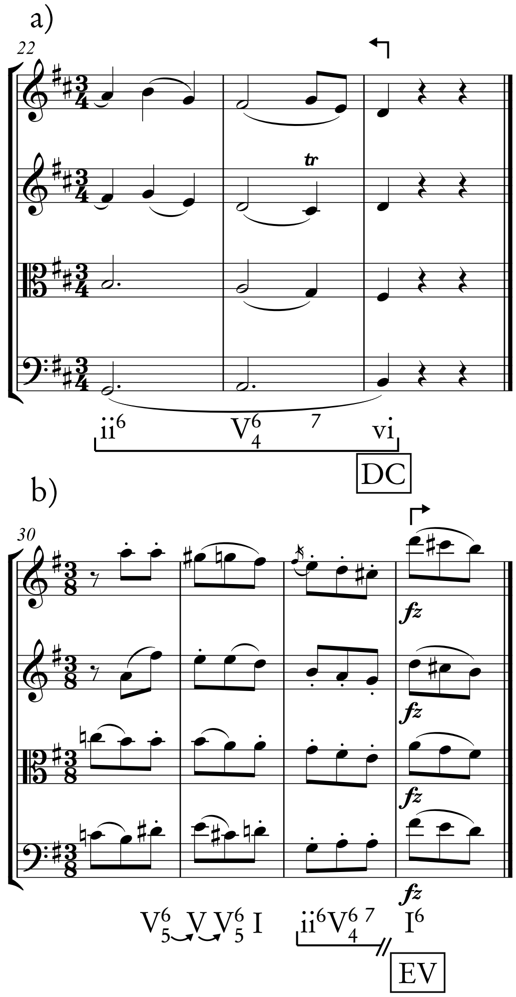
The cadential deviations sub-tree consists of 26 cadences: 19 cadences from the DC category, and 7 cadences from the EV category. Shown in Figure 5.6, the NJ method clustered the cadences into three branches in the sub-tree. Reading clockwise starting on the right, the first two branches represent those cadences featuring a standard bass that resolves deceptively to $\hat{6}$, but with the soprano resolving either to $\hat{1}$ (branch one) or $\hat{3}$ (branch two) at the moment of cadential arrival. These two branches feature just one confusion, which in this case consisted of an evaded cadence whose syntactic characteristics exemplified the DC category, but whose rhetorical features signaled cadential evasion. Branch three represents those cadences for which the bass resolves down to $\hat{3}$, in nearly every case supporting a I^6 harmony. But since I^6 could potentially group forward or backward, depending on the behavior of the many other characteristics supporting boundary function—surface activity, dynamics, texture, rhythmic duration, metric position, and so on—branch three also includes two cadences from the DC category.

Example 5.7a presents the exemplars from branches one and two of the sub-tree. Shown in

Table 5.5, this cadence received by far the highest prototypicality estimate of .90, with the next highest candidate receiving an estimate of .71. Nevertheless, it is also noteworthy that when compared to the other cadential deviations in the sub-tree, which include cadences from the EV category in branch three, Example 5.7a received a lower estimate of .79. Like the previous examples in the authentic cadence sub-tree, it features the standard, unembellished $\hat{3}$–$\hat{4}$–$\hat{5}$–$\hat{1}$ bass, in this case a composta, and the soprano resolves as expected to $\hat{1}$ following a stepwise descent from $\hat{4}$. But as is often the case with deceptive cadences, the bass resolves deceptively to $\hat{6}$, thus thwarting the expected moment of cadential arrival.

Shown in Example 5.7b, the exemplar from branch three features the same standard, composta bass. Following Johann Gottfried Walther, Gjerdingen suggests that the standard voice-leading scheme for the resolution of a dominant seventh to a root-position tonic—Walther's *clausula formalis perfectissima*, or "most complete close"—features $\hat{7}$–$\hat{1}$ in the soprano, ($\hat{5}$–)$\hat{4}$–$\hat{3}$ in the alto, $\hat{2}$–$\hat{1}$ in the tenor, and $\hat{5}$–$\hat{1}$ in the bass,[60] and it is noteworthy that this example preserves that scheme exactly. Yet, as is customary of cadences from the EV category in Caplin's theory, the cadence in Example 5.7b thwarts the expected moment of cadential arrival, in this case by leaping above the expected scale degrees in each of the four voices and replacing the expected root-position tonic with I^6. The events thus group forward as a consequence of the changes in register, harmony, dynamics, and articulation.

Although these two exemplars provide textbook examples of cadential deviations, the relatively meager sample for the DC and EV categories in the Haydn Corpus casts considerable doubt on the inferences we might hope to draw about prototypicality, particularly given the range of features cadential deviations support for the events at the moment of cadential arrival. Remember that since the DC and EV categories differ primarily as a consequence of the temporal function expressed by their final events, and not as a consequence of a specific

[60] Gjerdingen, *Music in the Galant style*, 139–140.

Example 5.7: Exemplars from the cadential deviations sub-tree. a) Branches 1/2: String Quartet in D, Op. 20/4, i, mm. 22–24; b) Branch 3: String Quartet in C, Op. 74/1, ii, mm. 30–33.

melodic, harmonic, or rhythmic goal, numerous syntactic and rhetorical events may appear at the moment of cadential arrival so long as they function either as *endings* or *beginnings*. And because the ratio model only included *n*-grams whose final events appeared at the end of the sequence, the NJ method failed to group all of the cadential deviations in the corresponding sub-tree.

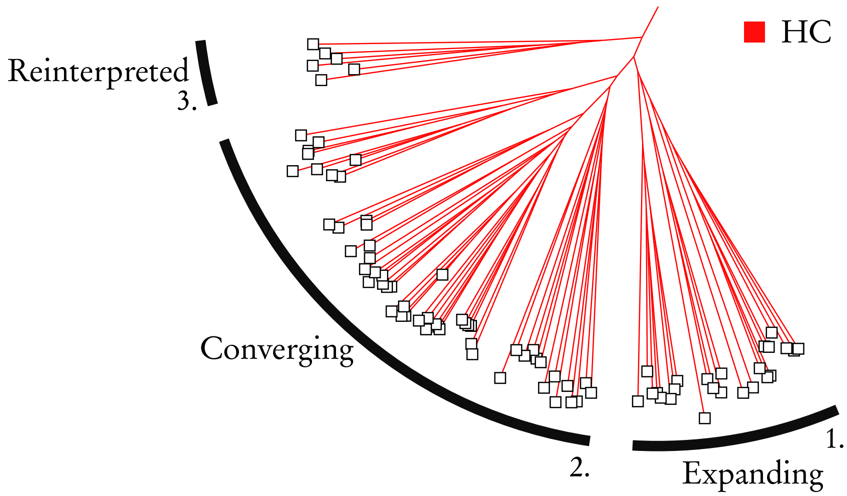


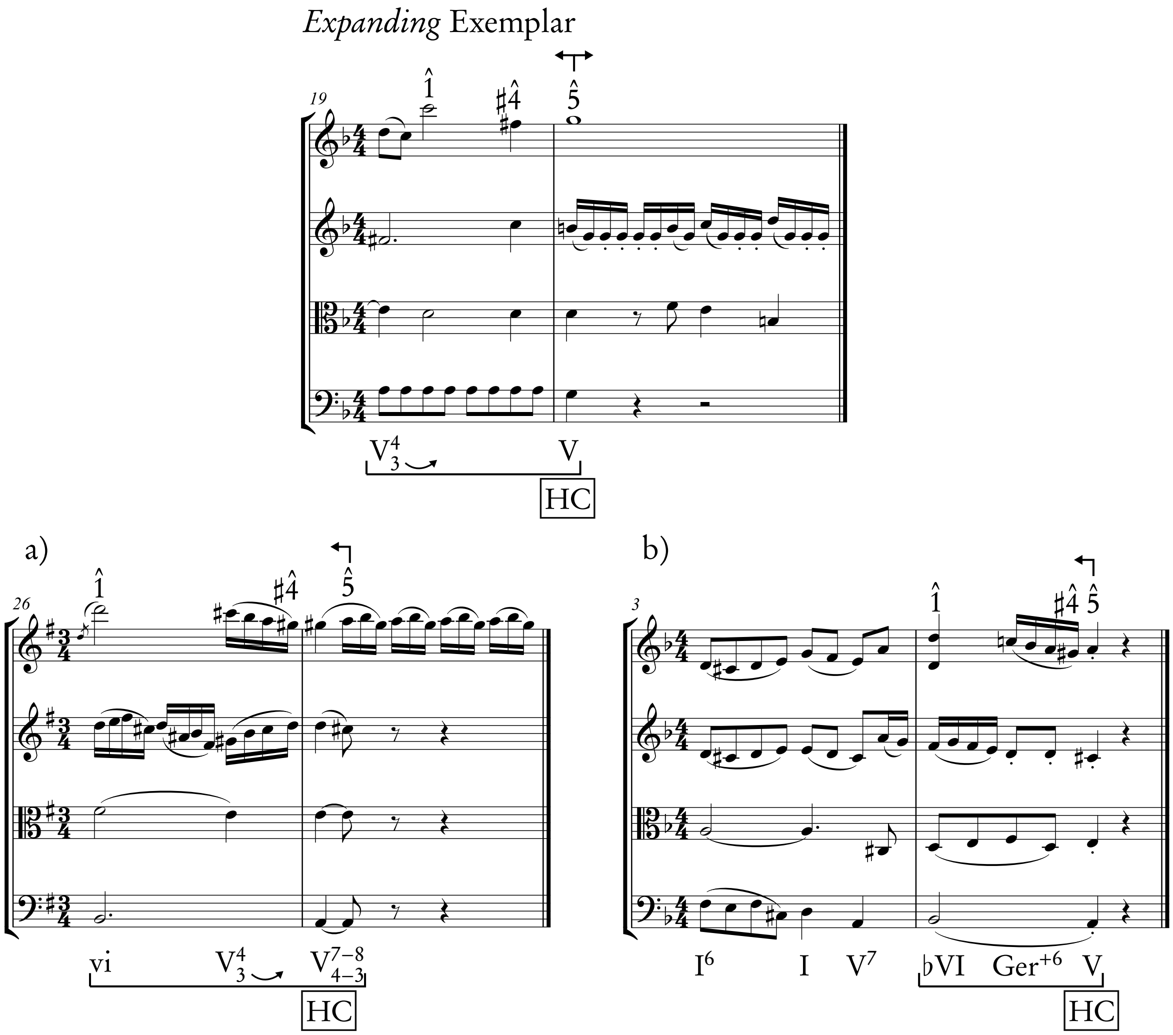
Figure 5.7: Equal-angle dendrogram calculated from the top model combination—$\texttt{contour}_{\mathrm{vc}} \times \texttt{csd}_{\mathrm{vc}} \times \texttt{csd}_{\mathrm{vl1}} \times \texttt{strength} \times \texttt{vintcc}$—for the half cadence sub-tree. Each cadence was clustered into the categories from Caplin's typology.

5.4.3 Half Cadences

The half cadence sub-tree consists of all 84 cadences from the HC category. Shown in Figure 5.7, the NJ method clustered the cadences into three branches in the sub-tree, with the second (middle) branch subdivided into three subordinate branches. Reading clockwise starting on the right, the first two branches clustered the cadences according to the contour of the bass, with the cadences in the first Expanding branch featuring a descending bass and the cadences in the second Converging branch featuring an ascending bass. The third Reinterpreted branch represents those cadences that reference the PAC formula—the resolution of a root-position dominant to a root-position tonic, with $\hat{1}$ in the soprano at the cadential arrival—in the key of the dominant. But since the home key is immediately reinstated following the cadence, Caplin refers to such cases as *reinterpreted half cadences*.[61]

For the previous sub-trees, the branches did not always reflect pertinent sub-types of the corresponding categories. For the half cadence sub-tree, however, the NJ method discovered sub-

[61]Caplin, *Classical Form*, 57.



Example 5.8: Top: *Expanding* Exemplar from the first branch of the half cadence sub-tree. String Quartet in F, Op 17/2, i, mm. 19–20. Bottom: Variants of the *Expanding Do-Fi-Sol*. a) String Quartet in G minor, Op. 20/3, iii, mm. 26–27; b) String Quartet in D minor, Op. 76/2, i, mm. 3–4.

types that have only recently been described in the *Formenlehre* tradition. Shown in Example 5.8, the cadences in branch one exemplify a sub-type Nathan Martin and Julie Pedneault-Deslauriers have called the *Expanding* 6–8 half cadence, in which the bass descends by step from $\hat{6}$ (or $\flat\hat{6}$) to $\hat{5}$, resulting in cadential progressions that tonicize the dominant, such as V^4_3/V–V, vii^6/V–V, or

+6$^{\text{th}}$–V.[62] Yet unlike the authentic cadence and its variants, whose final harmony supports a strict melodic goal (typically $\hat{1}$ or $\hat{3}$), the final harmony of the half cadence may support any chord member in the soprano. As such, it typically descends by step to $\hat{2}$ or $\hat{7}$ or ascends by step to $\hat{5}$, depending on the bass clausula supporting it. In the *Expanding* sub-type, the soprano usually ascends by half step from $\sharp\hat{4}$ to $\hat{5}$, thereby producing the schema's characteristic 6–8 intervallic progression between the outer voices.

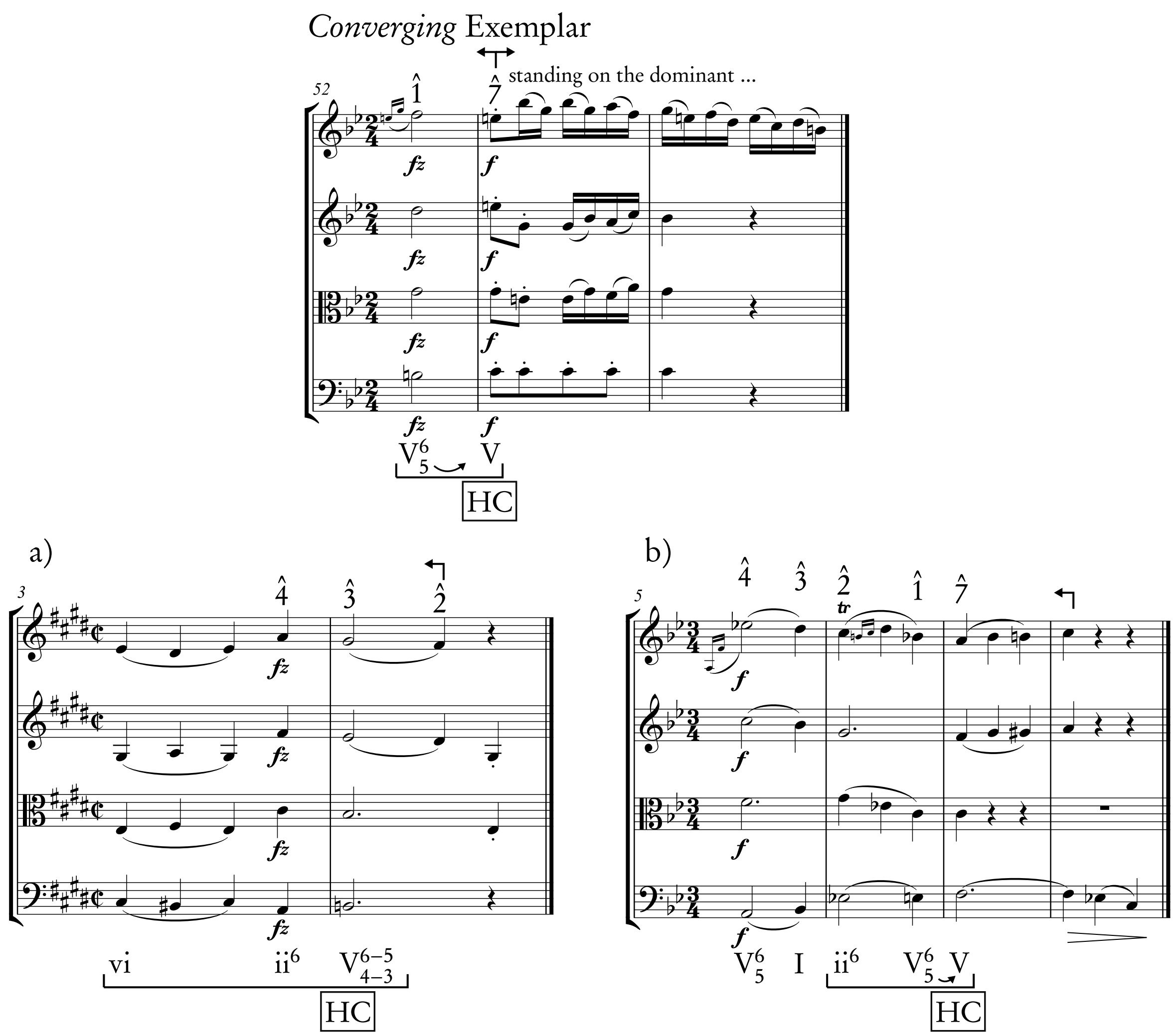
The exemplar from branch one in Example 5.8 closes the transition in the first movement of Haydn's Op. 17, No. 2. Shown in Table 5.5, it received a *proto* of .77 for the cadences within the branch and .62 for the cadences in the half cadence sub-tree. According to Martin and Pedneault-Deslauriers, the normative *Expanding* sub-type begins with $\hat{3}$ in the soprano,[63] but the most frequent pattern in Haydn's string quartets opens with $\hat{1}$ in an upper register, which then leaps down to $\sharp\hat{4}$ before resolving to $\hat{5}$. And since the $\sharp\hat{4}$–$\hat{5}$ pattern sometimes appears in an inner voice while the soprano descends to $\hat{7}$, I will call the exemplar in Example 5.8 the *Expanding Do-Fi-Sol*. Provided below are two variants of the schema that fill in the space between $\hat{1}$ and $\sharp\hat{4}$ and open with the submediant harmony supported by $\hat{6}$ (Example 5.8a) or $\flat\hat{6}$ (Example 5.8b).

As is evident in the half cadence sub-tree, the *Expanding* half cadence appears less frequently in the Haydn Corpus than the more well known *Converging* half cadence sub-type, which characterizes the majority of the cadences in branch two.[64] It consists of a stepwise ascending bass to $\hat{5}$ that often features $\sharp\hat{4}$, but unlike the *Expanding* HC, which typically concludes with $\hat{5}$ in the soprano, the *Converging* half cadence may be further subdivided by the melodic goal it supports, with the three subordinate branches comprised of cadences whose goal is $\hat{5}$, $\hat{2}$, and $\hat{7}$,

[62] Martin and Pedneault-Deslauriers, "The Mozartean Half Cadence," 190–192.

[63] Ibid., 190.

[64] Gjerdingen, *Music in the Galant style*, 160–162; Martin and Pedneault-Deslauriers, "The Mozartean Half Cadence," 186–189.



Example 5.9: Top: *Converging* Exemplar from the second branch of the half cadence sub-tree. String Quartet in B-flat, Op. 64, No. 3, iv, mm. 52–53. Bottom: Variants of the *Converging* Half Cadence. a) String Quartet in E, Op. 54, No. 3, i, mm. 4–5; b) String Quartet in B-flat, Op. 55, No. 3, i, mm. 5–8.

reading clockwise starting from the bottom.[65]

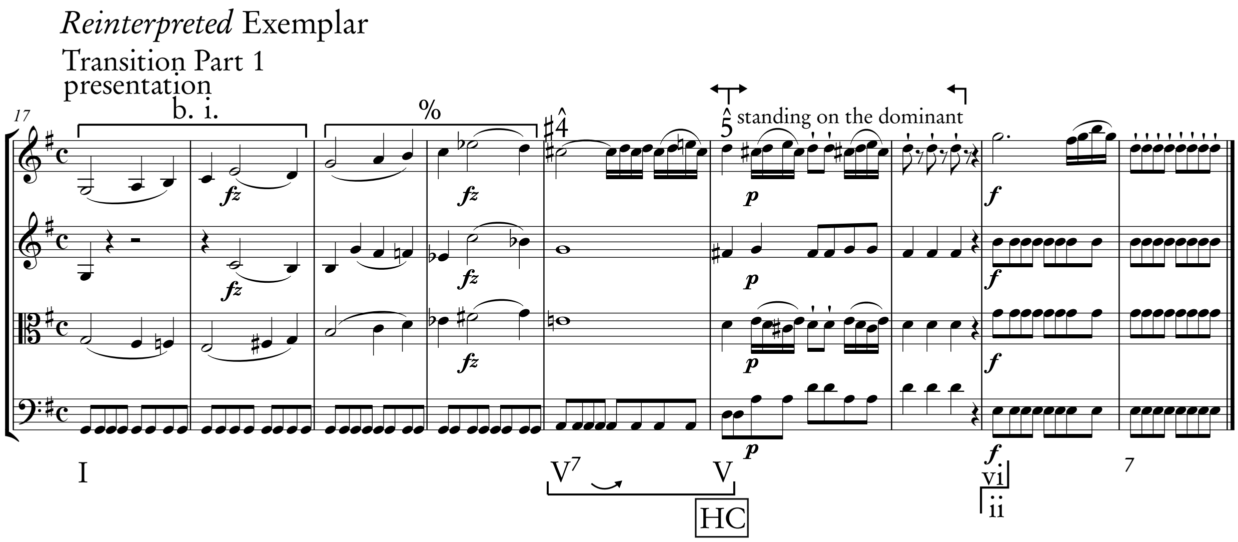
Example 5.9 presents the exemplar and two variants from the *Converging* HC branch of the

[65] $\hat{5}$ is a fairly uncommon melodic goal in the *Converging* sub-type; just six cadences from the Haydn Corpus met this criterion. Like branch three of the authentic cadence sub-tree, the first subordinate branch of the *Converging* sub-type is something of a catch-all, including half cadences with stepwise ascending bass lines and $\hat{5}$ in the soprano at the cadential arrival, as well as more elaborated structures than those found in subordinate branches two and three that feature $\hat{2}$ or $\hat{7}$ as the melodic goal.

sub-tree. Appearing in the opening measures of the first movement of Op. 54/3, the exemplar in the *Converging* HC branch received the highest prototypicality estimate of all of the cadences in the sub-tree, with a value of .92 for the cadences in the sub-tree and .96 for the cadences within the branch. It features $\sharp\hat{4}$ to $\hat{5}$ in the bass and $\hat{1}$ to $\hat{7}$ in the soprano but lacks a diatonic pre-dominant harmony like ii^6, resulting in just two harmonies in the cadential progression. The more common three-stage variant of the *Converging* sub-type is shown in Example 5.9b, where the $\hat{4}$–$\sharp\hat{4}$–$\hat{5}$ pattern in the bass supports a stepwise descent to $\hat{7}$ in the soprano. In this case, the initial melodic descent in m. 5 precedes the start of the cadential progression, and the cadential dominant is prolonged by a melodic overhang that climbs back up to $\hat{2}$ before the end of the passage. Example 5.9a presents a separate variant of the *Converging* schema in which $\sharp\hat{4}$ is no longer present in the bass and vi replaces the more typical I^6, resulting in the cadential progression vi–ii^6–V. This variant also includes a cadential six-four to support the melodic descent from $\hat{4}$ to $\hat{2}$.

Finally, Example 5.10 presents the exemplar from the third branch of the half cadence sub-tree. Following a two measure basic idea and its repetition at the beginning of the transition, the passage abruptly cadences on dominant harmony in m. 22. In the home key, the bass clausula $\hat{2}$–$\hat{5}$ supports the soprano motion from $\sharp\hat{4}$ to $\hat{5}$, which in the key of the dominant is the standard formula for a perfect authentic cadence; in fact, the voice leading exemplifies Walther’s *clausula formalis perfectissima*. Following the dominant prolongation and caesura, the beginning of the second part of the transition reasserts the home key (albeit temporarily), suggesting that the perfect authentic cadence formula signifies a half cadence in retrospect. Within the branch, this exemplar received a prototypicality estimate of .83, but given the small number of cadences in the reinterpreted HC branch, it should be unsurprising that it received a much lower value of .41 for the cadences in the sub-tree.

Two other half cadence sub-types are worthy of mention here despite being clustered in



Example 5.10: *Reinterpreted* Exemplar from the third branch of the half cadence sub-tree. String Quartet in G, Op. 54, No. 1, i, mm. 17–23.

the larger branches of the sub-tree. One instance of Vasili Byros' *Le-Sol-Fi-Sol* schema appears in the second branch of the sub-tree, for example, because, like the Converging sub-type, the bass clausula of the *Le-Sol-Fi-Sol* features ♯4̂.[66] In this case, the NJ method interpreted the *Le-Sol-Fi-Sol* as a variant of the Converging sub-type because the *ratio* model awards higher weights to *n*-grams featuring proximal members, and so ♯4̂–5̂ received a higher weight than ♭6̂–5̂. Along with the Expanding and Converging schemas, Martin and Pedneault-Deslauriers also mention a third sub-type called the *Simple* half cadence, which features a I–V progression with 2̂ in the soprano at the cadential arrival.[67] The NJ method did not cluster the simple half cadences from the Haydn Corpus in a separate branch, however, instead placing them in the third subordinate branch of the Converging sub-type because they featured an ascending (albeit leaping) bass and 2̂ at the cadential arrival. Presumably had there been more instances of this sub-type in the Haydn Corpus, the NJ method would have distinguished Simple half cadences from the other sub-types in the tree.

[66] Op. 54/2, i, mm. 50–54. Byros discusses this example in "Meyer's Anvil," 294.
[67] Martin and Pedneault-Deslauriers, "The Mozartean Half Cadence," 192–193.

§5.5 Conclusions

This chapter presented a probabilistic approach to category formation, in which a category is a network of overlapping attributes, and its members are prototypical to the extent that they bear a *family resemblance* to other members in the category. To support this view, I identified a collection of cadences from the Haydn Corpus in §5.1 and then classified the cadences using set-theoretic tools pioneered by psychologist Amos Tversky. Classifiers typically depend on some notion of similarity, so §5.2 quantified the similarity between cadences using the *ratio* model, which assumes that asymmetries between items imply a "nested" relation (i.e., where the more prototypical item nests within the variant item). Using the *neighbor-joining* method in §5.3, I then visualized and classified the cadences on the basis of the features they shared using phylogenetic trees. Finally, §5.4 presented the cadential "tree of life" for the cadences from the Haydn Corpus and identified the most prototypical exemplars from each category.

To this point I have said nothing about whether the principles motivating the development of the NJ method in evolutionary biology correspond in any way with the assumptions that led to its application here. Like any clustering algorithm, the NJ method depends on a similarity algorithm to produce a dissimilarity matrix, which in my case quantifies the dissimilarities between cadences using an asymmetric procedure, where cadences sharing most (or all) of their features with the other cadences in the corpus serve as prototypes. In the final approach, the NJ method attempted to preserve the asymmetries between cadences by placing more prototypical members near the center of the equal-angle tree. Thus, in the square tree in Figure 5.3 (see Page 210), the branch lengths represent the dissimilarity between cadential bass lines, and the horizontal position of each leaf reflects the prototypicality of the corresponding bass line.

For evolutionary biologists, distance methods typically quantify the dissimilarity between two species by the number of substitutions between their respective DNA sequences. Thus, the branch lengths of an additive tree represent the genetic changes between biological species over evolutionary time scales. To return to Figure 5.3, the branch lengths would now represent genetic change, with the horizontal position of each leaf representing the appearance of the corresponding species on the evolutionary time line. In other words, biologists interpret the horizontal position of each leaf as a measure of evolutionary time, while I have interpreted those positions as a measure of prototypicality.

One could of course assume that the distinction between time and prototypicality is meaningless and suggest that increasingly sophisticated diminutions of the closing schemas described here have evolved gradually over the history of Western music.[68] Gjerdingen notes, for example, that the more often a schema occurs, the more prototypical it will be, resulting in approximately normally-distributed instances of the schema over time.[69] But since I have applied the NJ method to a relatively homogeneous corpus of modest size and scope, it would be unreasonable to conclude that any given cadence necessarily evolved out of other, simpler exemplars from the same corpus. In my case, the path lengths do not reflect an underlying temporal continuum over which musical changes occur, but rather a continuum reflecting the depth of our schematic knowledge for a given cadential formula.

Recall from Chapter 2 that schematic knowledge inhabits a continuum from relatively deep to relatively shallow, where depth relates to availability for direct access, susceptibility to change through exposure, and scope of application.[70] As an example, I cited the perfect authentic cadence, noting that a V–I progression persisted throughout much of the history of Western music, whereas the Cudworth cadence was confined to a relatively narrow period of roughly eighty years. From this point of view, the additive trees in Figure 5.4 (see Page 218)

[68]Don Randel's article on the "emerging" V–I cadence exemplifies this kind of thinking ("Emerging Triadic Tonality in the Fifteenth Century," *The Musical Quarterly* 57, no. 1 [1971]: 73–86).
[69]Gjerdingen, *A Classic Turn of Phrase*, 99–106.
[70]Margulis, "A Model of Melodic Expectation," 666.

visualize the prototypicality (or schematic depth) of the cadences encountered in the Haydn Corpus, where cadences like the Cudworth lie near the outer extremes of the tree because they are increasingly available for access, increasingly susceptible to change through exposure, and increasingly limited in historical scope. Put another way, if one could imagine schema variation on a target, as Gjerdingen has previously suggested,[71] and then overlay this target on either of the additive trees in Figure 5.4, the more prototypical exemplars would lie closer to the bull's eye, with the less prototypical variants positioned at the furthest extremes of the target.

The point here is that the evolutionary metaphor may be appealing, but these methods do not represent changes in a given cadence category over time, but rather changes in the prototypicality of the cadences from that category. That being said, this approach to classification is not without its limitations. First, by assuming that prototypes "nest within" variants, the ratio model as I have applied it here automatically favors reductions to simpler, less embellished patterns, which can at times lead to strange results. For example, the exemplar at the top of Example 5.9 shares nearly all of its features with the other cadences from the *Converging* branch, but is it more prototypical of the *Converging* half cadence schema than, say, Example 5.9b? Perhaps a prototypical exemplar not only shares most (or all) of its features with the other members of its class, but also contains those features that cover the largest *number* of those members. This is to say that patterns like Example 5.9b—which shares features with both the *Converging* Exemplar (V^6_5/V–V) and Example 5.9a (ii^6–V)—are more likely to serve as prototypes if they maximize the number of represented members.

Second, by applying a fixed weighting scheme based on the Shannon entropy of the viewpoints in the final model combination, I have assumed that the combination of viewpoints employed in service of similarity estimation and classification would not change from one category to the next. For the half cadence-tree, this weighting scheme divided the cadences into

[71]Gjerdingen, *A Classic Turn of Phrase*, 94.

pertinent sub-types, but for the authentic cadence sub-tree, the branches rarely corresponded with many of the sub-types described in the *Formenlehre* tradition. Whether we classify cadences from each category using such a fixed weighting scheme is an open question, but it is at least possible that listeners attend to different musical features depending on the diagnostic properties of the category (i.e., the features that are most likely to distinguish one member from another). The presence of a cadential six-four or the precise sequence of scale degrees in the melody may be more relevant to the authentic cadence category than to the half cadence category, for example.

But perhaps the greatest limitation of the present approach was to exclude *n*-grams from the similarity model if their final members did not terminate at the end of the sequence. As a consequence, the model clustered the cadences into the largest branches of the tree on the basis of the final event(s) they shared. The goal here was to adopt a hybrid approach that made two assumptions: (1) that cadence categories should not be determined entirely by certain essential features, but by the *family resemblance* of their members; and (2) that the final events of the cadence should weigh more heavily than the initial events in the final classification. As an initial step, I adopted the largely *essentialist* stance that the exclusion of *n*-grams that do not terminate at the end of the sequence would improve the accuracy and precision of the classifier, despite the fact that listeners might take a more holistic approach to similarity estimation and classification. An alternative approach might be to weight each *n*-gram depending on its position in the sequence, or to compute a measure like *cue validity* (see § 2.2.2), which determines the diagnostic value of each event within the sequence empirically.

Indeed, some in the *Formenlehre* tradition classify the various categories of the classical cadence according to the strength of the cadential arrival, leading many theorists to appeal to theories of expectation as the source of the cadential percept. And since the material preceding cadential arrival elicits very definite expectations concerning the melodic scale-degree, the harmony, and the metric position of the goal event, classification would seem to depend as much on the expectations these cadences generate as it does on the features they share. Thus, in Chapter 6 I apply a *context* model developed by Marcus Pearce called the *Information Dynamics of Music* model (or IDyOM) to determine whether expectancy formation, fulfilment, and violation contribute to the hierarchy of cadence categories described in the *Formenlehre* tradition.

Chapter 6

Predicting Closing Schemas: The Cadential Hierarchy

> *... anyone speaking a language possesses, implicitly, an enormous knowledge of the statistics of the language. Familiarity with the words, idioms, clichés and grammar enables him to fill in missing or incorrect letters in proof-reading, or to complete an unfinished phrase in conversation.*
>
> CLAUDE E. SHANNON

The ability to make accurate predictions about future events is a biological imperative.[1] While descending a staircase, for example, even slightly misjudging the height or depth of each step could be fatal, so the brain activates a schematic representation of the staircase in order to predict future steps, using incoming auditory, visual, haptic, and proprioceptive cues to minimize potential prediction errors and update the representation in memory. Since our chances of survival from one moment to the next depend on the accuracy of our predictions, it

[1] Moshe Bar, "The Proactive Brain: Using Analogies and Associations to Generate Predictions," *Trends in Cognitive Sciences* 11, no. 7 (2007): 280–289; Karl Friston, "A Theory of Cortical Responses," *Philosophical Transactions of the Royal Society, London, Series B* 360 (2005): 815–836; Karl Friston, "The Free Energy Principle: A Unified Brain Theory?," *Nature Reviews Neuroscience* 11 (2010): 127–138.

is no surprise that many cognitive scientists consider predictive processing to be the "primary function" of the brain.[2] And as Psychologist Ulric Neisser points out, the brain engages these predictive mechanisms constantly in response to all kinds of sensory stimuli, including those from the auditory modality.

> The listener continuously develops more or less specific readinesses (anticipations) for what will come next, based on information he has already picked up. These anticipations—which themselves must be formulated in terms of temporal patterns, not of isolated moments—govern what he will pick up next, and in turn are modified by it. Without them, he would hear only a blooming, buzzing confusion.[3]

In Chapter 2 I characterized the most common cadence types from the classical style as relatively short time spans consisting of recurrent temporal patterns. In my view, listeners who are familiar with classical music have internalized these cadence types as a complex network of rival closing schemata. The activation of this network allows listeners to form expectations for the terminal events of the cadence, the fulfillment of which determines the perceived strength and temporal position of the cadential boundary both for the schema itself, and in some cases, for the larger phrase-structural process that subsumes it. For our purposes, this means that (1) expectancy violations for the terminal events of the cadence result in closing schemata of diminished strength (see §2.3.2); and (2) the terminal event of the cadence is the most expected event in the surrounding sequence, while the next event in the sequence is comparatively unexpected (see §2.3.3).

[2]Andreja Bubic, D. Yves von Cramon, and Ricarda I. Schubotz, "Prediction, Cognition and the Brain," *Frontiers in Human Neuroscience* 4 (2010): 1–15; Jeff Hawkins and Sandra Blakeslee, *On Intelligence* (New York, NY: Times Books, 2004), 89. For a current review of predictive processing, see Andy Clark, "Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science," *Behavioral and Brain Sciences* 36, no. 3 (2013): 1–73.

[3]Neisser, *Cognition and Reality*, 27.

From the probabilistic view of category formation just described, schematic expectations for the terminal events of the cadence amount to probabilistic inferences, with the most expected (i.e., probable) endings also being the most complete or closed. David Huron notes, for example, that "it is not simply the final note of the cadence that is predictable; the final note is often approached in a characteristic or formulaic manner. If cadences are truly stereotypic, then this fact should be reflected in measures of predictability."[4] If Huron is right, applying a probabilistic approach to the cadences from the Haydn Corpus should allow us to examine these claims empirically.

In this chapter I apply and extend a probabilistic account of expectancy formation developed by Marcus Pearce called the *Information Dynamics of Music* model (or IDyOM)—a finite-context (or *n*-gram) model that predicts the next event in a musical stimulus by acquiring knowledge through unsupervised statistical learning of sequential structure—to examine how the formation, fulfillment, and violation of schematic expectations may contribute to the perception of cadential closure during music listening.[5] IDyOM is based on a class of Markov models commonly used in statistical language modeling,[6] the goal of which is to simulate the learning mechanisms underlying human cognition. Pearce explains,

> It should be possible to design a statistical learning algorithm possessing no prior knowledge of sequential dependencies between melodic events but which, given exposure to a reasonable corpus of music, would exhibit similar patterns of melodic expectation to those observed in experiments with human participants.[7]

Unlike language models, which typically deal with unidimensional inputs, IDyOM generates

[4]Huron, *Sweet Anticipation*, 154.

[5]Pearce, "The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition."

[6]Manning and Schütze, *Foundations of Statistical Natural Language Processing*.

[7]Marcus T. Pearce and Geraint A. Wiggins, "Expectation in Melody: The Influence of Context and Learning," *Music Perception* 23, no. 5 (2006): 386–387.

predictions for multidimensional melodic sequences using the *multiple-viewpoint* framework developed by Darrell Conklin and described in Chapter 3,[8] which is to say that Pearce's model generates predictions for basic viewpoints like `cpitch` by combining viewpoint predictions using a set of simple heuristics to minimize model uncertainty.[9] In the past decade, several studies have demonstrated the degree to which IDyOM can simulate the responses of listeners in melodic segmentation tasks,[10] subjective ratings of predictive uncertainty,[11] subjective and psychophysiological emotional responses to expectancy violations,[12] and behavioral,[13] electrophysiological,[14] and neural measures of melodic pitch expectations.[15]

Pearce provides the technical details of IDyOM elsewhere,[16] but a summary and discussion of the mathematical formalism will be useful to us here.[17] To that end, §6.1 describes the methods for estimating the conditional probability function for individual viewpoints like `cpitch` or `csd` using *maximum likelihood* estimation and the *prediction by partial match* algorithm. In §6.2, I describe Pearce's procedure for improving model performance by combining these

[8] Conklin, "Modelling and Generating Music Using Multiple Viewpoints"; Conklin, "Prediction and Entropy of Music"; Conklin and Witten, "Multiple Viewpoint Systems for Music Prediction."

[9] Pearce, Conklin, and Wiggins, "Methods for Combining Statistical Models of Music." Recall from §3.3 that `cpitch` denotes the vector of pitches in each instrumental part from the twelve-tone chromatic scale.

[10] Marcus T. Pearce, Daniel Müllensiefen, and Geraint A. Wiggins, "The Role of Expectation and Probabilistic Learning in Auditory Boundary Perception: A Model Comparison," *Perception* 39 (2010): 1367–1391.

[11] Niels Chr. Hansen and Marcus T. Pearce, "Predictive Uncertainty in Auditory Sequence Processing," *Frontiers in Psychology* 5 (2014): 1–17, doi:`10.3389/fpsyg.2014.1052`.

[12] Hauke Egermann et al., "Probabilistic Models of Expectation Violation Predict Psychophysiological Emotional Responses to Live Concert Music," *Cognitive, Affective, and Behavioral Neuroscience* 13, no. 2 (2013).

[13] Diana Omigie, Marcus T. Pearce, and Lauren Stewart, "Tracking of Pitch Probabilities in Congenital Amusia," *Neuropsychologia* 50 (2012): 1483–1493; Pearce and Wiggins, "Expectation in Melody"; Marcus T. Pearce et al., "Unsupervised Statistical Learning Underpins Computational, Behavioural, and Neural Manifestations of Musical Expectation," *NeuroImage* 50 (2010): 302–313.

[14] Diana Omigie et al., "Electrophysiological Correlates of Melodic Processing in Congenital Amusia," *Neuropsychologia* 51 (2013): 1749–1762.

[15] Pearce et al., "Unsupervised Statistical Learning Underpins Computational, Behavioural, and Neural Manifestations of Musical Expectation."

[16] I draw much of the following discussion from Pearce and Wiggins, "Improved Methods for Statistical Modelling of Monophonic Music."

[17] Documentation and downloads for IDyOM are available here: `https://code.soundsoftware.ac.uk/projects/idyom-project`.

viewpoint models into a single composite prediction for each event in the sequence. Finally, §6.3 applies IDyOM to the events surrounding each cadence in the Haydn Corpus to consider the following three claims about the link between expectancy and cadential closure: (1) terminal events from cadential contexts are more predictable than those from non-cadential contexts (§6.3.1); (2) the probabilistic account exemplified by IDyOM corresponds with models of cadential strength advanced in the *Formenlehre* tradition (§6.3.2); and (3) a significant decrease in predictability follows the terminal event of the cadential sequence (§6.3.3).

§6.1 IDyOM: A Cognitive Model of Musical Expectation

6.1.1 Maximum Likelihood

The goal of context models like IDyOM is to derive from a corpus of example sequences a model which estimates the probability of event e_i given a preceding sequence of events e_1 to e_{i-1}, notated here as e_1^{i-1}. Thus, the function $p(e_i|e_1^{i-1})$ assumes that the identity of each event in the sequence depends only on the events that precede it. In principle, the length of the context is limited only by the length of the sequence e_1^{i-1}, but context models typically stipulate a global order bound such that the probability of the next event depends only on the previous $n-1$ events, or $p(e_i|e_{(i-n)+1}^{i-1})$. Following the Markov assumption (see § 4.2), the model described here is an $(n-1)^{\text{th}}$ order Markov model, but researchers also sometimes call it an n-gram model because the sequence $e_{(i-n)+1}^{i}$ is an n-gram consisting of a *context* $e_{(i-n)+1}^{i-1}$ and a single-event *prediction* e_i.

To estimate the conditional probability function $p(e_i|e_{(i-n)+1}^{i-1})$ for each event in the *test* sequence, IDyOM must first acquire the frequency counts for a collection of such sequences from a *training* set. When the trained model is exposed to the test sequence, it then uses the frequency counts to estimate the probability distribution governing the identity of the next event in the sequence given the $n-1$ preceding events.[18] In this case, IDyOM relies on *maximum likelihood* (ML) estimation.

$$p(e_i|e_{(i-n)+1}^{i-1}) = \frac{c(e_i|e_{(i-n)+1}^{i-1})}{\sum_{e \in A} c(e|e_{(i-n)+1}^{i-1})} \tag{6.1}$$

The numerator terms represent the frequency count c for the n-gram $e_i|e_{(i-n)+1}^{i-1}$, and the denominator terms represent the sum of the frequency counts c associated with all of the possible events e in the alphabet A following the context $e_{(i-n)+1}^{i-1}$. Thus, Equation 6.1 represents the ratio of the number of times e_i follows the context $e_{(i-n)+1}^{i-1}$ to the number of times *any* event e from A follows the context $e_{(i-n)+1}^{i-1}$.

Armed with the Haydn Corpus and the equation above, we could generate probability estimates for nearly any encountered sequence.[19] To illustrate, Table 6.1 provides the counts and probabilities for the chromatic scale degrees from the first violin in the Haydn Corpus for the sequence $\hat{4}$-$\hat{3}$-$\hat{2}$-e_4. If we ignore the preceding context (i.e., $n = 1$), $\hat{5}$ receives the highest probability estimate of .187 ($\frac{2619}{13{,}994}$). As the value of n increases, however, the model predictions increasingly favor $\hat{1}$ as the most likely continuation, with a probability estimate of .322 in the first-order model, .546 in the second-order model, and finally .566 in the third-order model. Thus, the higher-order models predict what we might expect: namely, that $\hat{1}$ or perhaps $\hat{3}$ will follow the sequence $\hat{4}$-$\hat{3}$-$\hat{2}$. On this basis, we might assume that higher-order models better simulate listener expectations than lower-order models,[20] but how do we justify this claim in

[18] Pearce and Wiggins, "Expectation in Melody," 389.

[19] When n is larger than i, such as at the beginning of the sequence, context models introduce padding symbols to provide the necessary context.

[20] Alistair Moffat, "Implementing the PPM Data Compression Scheme," *IEEE Transactions on Communications* COM-38, no. 11 (1990): 1917.

Table 6.1: Counts and probabilities for the chromatic scale degrees from the first violin in the Haydn Corpus following the contexts $\hat{4}$-$\hat{3}$-$\hat{2}$, $\hat{3}$-$\hat{2}$, $\hat{2}$, and no context (n = 1).

$n = 4$					$n = 3$					$n = 2$					$n = 1$			
			c	p				c	p				c	p			c	p
$\hat{4}$-$\hat{3}$-$\hat{2}$	→	$\hat{1}$	176	.566	$\hat{3}$-$\hat{2}$	→	$\hat{1}$	280	.546	$\hat{2}$	→	$\hat{1}$	598	.322	→	$\hat{1}$	2390	.171
	→	$\flat\hat{2}$	11	.035		→	$\flat\hat{2}$	22	.043		→	$\flat\hat{2}$	57	.031	→	$\flat\hat{2}$	171	.012
	→	$\hat{2}$	27	.087		→	$\hat{2}$	52	.101		→	$\hat{2}$	203	.109	→	$\hat{2}$	1859	.133
	→	$\flat\hat{3}$	4	.013		→	$\flat\hat{3}$	5	.01		→	$\flat\hat{3}$	93	.05	→	$\flat\hat{3}$	307	.022
	→	$\hat{3}$	45	.145		→	$\hat{3}$	61	.119		→	$\hat{3}$	336	.181	→	$\hat{3}$	1823	.13
	→	$\hat{4}$	21	.068		→	$\hat{4}$	49	.096		→	$\hat{4}$	186	.1	→	$\hat{4}$	1647	.118
	→	$\sharp\hat{4}$	1	.003		→	$\sharp\hat{4}$	1	.002		→	$\sharp\hat{4}$	9	.005	→	$\sharp\hat{4}$	406	.029
	→	$\hat{5}$	14	.045		→	$\hat{5}$	19	.037		→	$\hat{5}$	151	.081	→	$\hat{5}$	2619	.187
	→	$\hat{6}$	8	.026		→	$\hat{6}$	12	.023		→	$\flat\hat{6}$	3	.002	→	$\flat\hat{6}$	247	.018
	→	$\hat{7}$	4	.013		→	$\hat{7}$	12	.023		→	$\hat{6}$	40	.022	→	$\hat{6}$	1126	.08
											→	$\flat\hat{7}$	2	.001	→	$\flat\hat{7}$	135	.01
											→	$\hat{7}$	181	.097	→	$\hat{7}$	1264	.09

Note. n refers to the length of the context, c denotes the count for each chromatic scale degree continuation, and p represents the probability of each continuation without exclusion.

the absence of experimental evidence?

6.1.2 Performance Metrics

The most common performance metrics for context models derive from information-theoretic measures introduced by Claude E. Shannon in the late 1940s, one of which I summarized in §4.2. Context models for communication systems like natural language essentially began in Shannon's laboratory.[21] However, his main contribution was to formalize the relationship between probability theory and the quantification and compression of information, and in so doing, to derive a method for the transduction of messages from communication systems

[21] Claude E. Shannon, "Prediction and Entropy of Printed English," *Bell System Technical Journal* 30 (1951): 50–64.

like English into compressed numerical codes.[22] The principles outlined in his 1948 article "A Mathematical Theory of Communication" form the basis for the digital encoding, storage, and retrieval of information in modern data compression theory.[23]

At first glance, context modeling and compression theory might seem unrelated, but Shannon demonstrated that for seemingly stochastic communication systems (i.e., where the system produces sequences of discrete events according to certain probabilities), "parameters of engineering importance such as time, bandwidth, and number of relays, etc., tend to vary linearly with the logarithm of the number of possibilities," such that events associated with fewer possible outcomes in the transduced message also require less time (or bandwidth) or fewer relays during transmission.[24] Thus, the mathematical formalism on which his transduction scheme depends is typically a base 2 logarithmic function. To demonstrate how this scheme might work, Shannon encoded sequences of information according to the number of *bits* (or binary digits, represented by 0 or 1) associated with their selection: the fewer the number of choices (or higher the probability) for a particular event in the message, the fewer the number of bits required to encode it, and thus, the less informative the event.

Returning to our earlier equation, if the probability of e_i is given by the conditional probability function $p(e_i|e_{(i-n)+1}^{i-1})$, *information content* (IC) represents the minimum number of bits required to encode e_i in context.[25]

$$\mathrm{IC}(e_i|e_{(i-n)+1}^{i-1}) = \log_2 \frac{1}{p(e_i|e_{(i-n)+1}^{i-1})} \tag{6.2}$$

IC is inversely proportional to p and thus represents the degree of contextual *unexpectedness*

[22] Conklin and Witten, "Multiple Viewpoint Systems for Music Prediction," 53.

[23] Timothy C. Bell, John G. Cleary, and Ian H. Witten, *Text Compression* (Englewood Cliffs, NJ: Prentice Hall, 1990).

[24] Shannon, "A Mathematical Theory of Communication," 379–380.

[25] David J. C. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge, UK: Cambridge University Press, 2003), 32.

or surprise associated with e_i. In Table 6.1, for example, $p(\hat{1}|\hat{4}\text{-}\hat{3}\text{-}\hat{2})$ is .566, so $\mathrm{IC}(\hat{1}|\hat{4}\text{-}\hat{3}\text{-}\hat{2})$ is $\log_2(\frac{1}{.566})$, or .821 bits. Researchers often prefer to report IC over p because it has a more convenient scale (p can become vanishingly small), and since it also has a well-defined interpretation in data compression theory,[26] I will prefer it in the analyses that follow.

Whereas IC represents the degree of unexpectedness associated with a particular event e_i in the sequence, *Shannon entropy* (H) represents the degree of contextual uncertainty associated with the probability distribution governing that outcome, where the probability estimates are independent and sum to one. I have already described the formula for H in Equation 4.1, but I present an alternative formula here to demonstrate the relationship between IC and H.

$$H(e_{(i-n)+1}^{i-1}) = \sum_{e \in A} p(e_i|e_{(i-n)+1}^{i-1})\mathrm{IC}(e_i|e_{(i-n)+1}^{i-1}) \tag{6.3}$$

H is computed by averaging the information content over all e in A following the context $e_{(i-n)+1}^{i-1}$. According to Shannon's equation, if the probability of a given outcome is 1, the probabilities for all of the remaining outcomes will be 0, and $H = 0$ (i.e., maximum certainty). If all of the outcomes are equally likely, however, H is maximum (i.e., maximum uncertainty). Thus, we can assume here that the best performing models will minimize model uncertainty. In principle, this means that each of the values of H computed from the probability distributions in Table 6.1 estimates the model's uncertainty for the possible continuations following each context. In this case, model uncertainty decreases from the zeroth-order to the first-order model ($n = 1, H = 3.14$; $n = 2, H = 2.83$), indicating that the model predictions improve as the context grows.[27]

[26]Pearce et al., "Unsupervised Statistical Learning Underpins Computational, Behavioural, and Neural Manifestations of Musical Expectation," 304.

[27]Note that I did not compute the values of H for the higher-order models in Table 6.1 because their distributions contain chromatic scale degrees for which the probability is 0 (i.e., $\flat\hat{6}$ and $\flat\hat{7}$ never follow the contexts $\hat{4}\text{-}\hat{3}\text{-}\hat{2}$ and $\hat{3}\text{-}\hat{2}$ in the Haydn Corpus). I mention in the next section that since $\log_2(0)$ is undefined, information-theoretic

In practice, we rarely know the true probability distribution of the stochastic process,[28] so it is often necessary to evaluate model performance using an alternative measure called *cross entropy*, denoted by H_m.

$$H_m(p_m, e_1^j) = -\frac{1}{j}\sum_{i=1}^{j} \log_2 p_m(e_i|e_1^{i-1}) \tag{6.4}$$

Whereas H represents the average information content over all e in the alphabet A, H_m represents the average information content for the model probabilities estimated by p_m over all e in the sequence e_1^j. That is, cross entropy provides an estimate of how uncertain a model is, on average, when predicting a given *sequence* of events.[29] Thus, while H is a good way of discriminating between models when the outcome is not yet known for a given event in the sequence, H_m allows us to discriminate between models based on the uncertainty of their predictions for every event in the sequence. As a consequence, H_m is frequently used to evaluate the performance of different context models on some corpus of data for tasks like speech recognition, machine translation, and spelling correction because, as Peter Brown and his co-authors note, "models for which the cross entropy is lower lead directly to better performance."[30]

6.1.3 Prediction by Partial Match

In Chapter 2 I suggested that the strength and specificity of our expectations for musical materials from the classical style depend on our schematic knowledge of the many recurrent

measures like IC and H depend on non-zero probability estimates for every possible prediction.

[28] Pearce and Wiggins, "Improved Methods for Statistical Modelling of Monophonic Music," 3.

[29] Pearce and Wiggins, "Improved Methods for Statistical Modelling of Monophonic Music," 3; Manning and Schütze, *Foundations of Statistical Natural Language Processing*, 74–76.

[30] Peter F. Brown et al., "An Estimate of an Upper Bound for the Entropy of English," *Computational Linguistics* 18, no. 1 (1992): 39.

formulæ found therein. If Pearce is correct that IDyOM simulates the "actual cognitive processes" involved in perceiving these phenomena,[31] it therefore follows that context models—like listeners—should improve as the value of n increases, just as we found for the probability distributions in Table 6.1. And yet because listeners are unlikely to remember temporal patterns beyond a certain value of n, and because such high-order patterns do not recur with any significant frequency in classical music, it also seems reasonable to impose an upper bound on the relationship between model uncertainty and model order. To be sure, if low-order models fail to account for the structural influence of the context on expectations, high-order models prevent IDyOM from capturing many of the statistical regularities present in the corpus.[32] The frequency counts in Table 6.1 demonstrate this fact: At $n = 1$, the model consists of 13,994 chromatic scale degree patterns, but by $n = 4$, the model consists of just 311 patterns.

What is more, because the number of potential patterns decreases dramatically as the value of n increases, high-order models often suffer from the *zero-frequency problem*,[33] in which n-grams encountered in the test set do not appear in the training set. The pattern $\hat{4}$-$\hat{3}$-$\hat{2}$-$\flat\hat{7}$ never appears in the Haydn Corpus, for example, so a third-order context model would return an estimated probability of 0. And since $\log_2(0)$ is undefined, it is therefore necessary for information-theoretic measures like IC, H, and H_m to provide non-zero probability estimates for every possible event in the alphabet. Simply put, context models like IDyOM assume that "no sequence of events is impossible, however unlikely it may be."[34]

To resolve this issue, IDyOM applies a data compression scheme called *Prediction by Partial Match* (PPM),[35] which adjusts the ML estimate for each event in the sequence by combining (or

[31] Marcus T. Pearce and Geraint A. Wiggins, "Auditory Expectation: The Information Dynamics of Music Perception and Cognition," *Topics in Cognitive Science* 4 (2012): 641.

[32] Pearce and Wiggins, "Expectation in Melody," 388–389.

[33] Ian H. Witten and Timothy C. Bell, "The Zero-Frequency Problem: Estimating the Probabilities of Novel Events in Adaptive Text Compression," *IEEE Transactions on Information Theory* 37, no. 4 (1991): 1085–1094.

[34] Conklin and Witten, "Multiple Viewpoint Systems for Music Prediction," 54.

[35] John G. Cleary and Ian H. Witten, "Data Compression Using Adaptive Coding and Partial String Matching,"

smoothing) the predictions generated at higher orders with less sparsely estimated predictions from lower orders. Context models estimated with the PPM scheme typically use a procedure called *backoff smoothing* (or *blending*), which assigns some portion of the probability mass from each distribution to an *escape probability* using an escape method to accommodate predictions that do not appear in the training set. When an event like ♭$\hat{7}$ does not appear in the $n-1$ order distribution, for example, PPM stores the escape probability and then iteratively backs off to lower-order distributions until it predicts the event or reaches the zeroth-order distribution, at which point it transmits the probability estimate for a uniform distribution over A (i.e., where every event in the alphabet is equally likely). PPM then multiplies these probability estimates together to obtain the final (smoothed) estimate.

Unfortunately there is no sound theoretical basis for choosing the appropriate escape method,[36] but Pearce and Wiggins recently demonstrated the potential of *method C* to minimize model uncertainty in melodic prediction tasks,[37] so I employ that method here.[38]

$$\gamma(e_{(i-n)+1}^{i-1}) = \frac{t(e_{(i-n)+1}^{i-1})}{\sum_{e \in A} c(e|e_{(i-n)+1}^{i-1}) + t(e_{(i-n)+1}^{i-1})} \tag{6.5}$$

Escape method C represents the escape count t as the number of distinct symbols that follow the context $e_{(i-n)+1}^{i-1}$. To calculate the escape probability for events that do not appear in the

IEEE Transactions on Communications 32, no. 4 (1984): 396–402.

[36] Witten and Bell, "The Zero-Frequency Problem: Estimating the Probabilities of Novel Events in Adaptive Text Compression."

[37] Pearce and Wiggins, "Improved Methods for Statistical Modelling of Monophonic Music," 4.

[38] Moffat, "Implementing the PPM Data Compression Scheme." In the original PPM scheme, John G. Cleary and Ian H. Witten offered two escape methods, named A and B, respectively ("Data Compression Using Adaptive Coding and Partial String Matching"). Researchers attempting to improve upon these methods have generally preserved the convention of using letter names for method designations. IDyOM implements a number of these escape methods, but again, Pearce points out that method C generally outperforms the other methods, so I will restrict my commentary to that method here. For a discussion of the escape methods implemented by IDyOM, see Pearce and Wiggins, "Improved Methods for Statistical Modelling of Monophonic Music," 4.

training set, γ represents the ratio of the escape count t to the sum of the frequency counts c and t for the context $e_{(i-n)+1}^{i-1}$. The appeal of this escape method is that it assigns greater weighting to higher-order predictions (which are more specific to the context) over lower-order predictions (which are more general) in the final probability estimate.[39]

By way of example, consider again the probability estimates for the sequence $\hat{4}$-$\hat{3}$-$\hat{2}$-♭$\hat{7}$ in Table 6.1. For a context model featuring a third-order global bound (i.e., $n = 4$), ♭$\hat{7}$ fails to follow the context beyond $n = 2$, so PPM stores the escape probability and then backs off to the lower-order distribution in both cases. For the $n = 4$ distribution, ten chromatic scale degrees follow the context, so $t = 10$, resulting in the escape probability $\frac{10}{311+10}$. For the $n = 3$ distribution, t remains the same, but the total number of patterns increases to 513, so $\frac{10}{513+10}$. For the $n = 2$ distribution, however, ♭$\hat{7}$ appears twice, so PPM terminates with the probability estimate $\frac{2}{1859+12}$. Note here that the ML estimate for $\hat{2}$-♭$\hat{7}$ *also* includes the escape count t (of 12), demonstrating a basic feature of the PPM scheme: the probability mass includes the escape count in *every* distribution whether or not the to-be-predicted event follows the context.

Bearing that assumption in mind, we need to revise the ML method from Equation 6.1 in the following way:

$$\alpha(e_i|e_{(i-n)+1}^{i-1}) = \frac{c(e_i|e_{(i-n)+1}^{i-1})}{\sum_{e \in A} c(e|e_{(i-n)+1}^{i-1}) + t(e_{(i-n)+1}^{i-1})} \tag{6.6}$$

Thus, the first two probability estimates $\frac{10}{321}$ and $\frac{10}{523}$ represent the escape probabilities for the third-order and second-order models, respectively, while $\frac{2}{1871}$ represents the revised ML estimate for the first-order model, denoted by α. Using backoff smoothing, the final probability estimate for ♭$\hat{7}$ in the sequence $\hat{4}$-$\hat{3}$-$\hat{2}$-♭$\hat{7}$ is therefore $\frac{10}{321} \times \frac{10}{523} \times \frac{2}{1871}$, or .0000006. As mentioned

[39] Suzanne Bunton, "On-Line Stochastic Processes in Data Compression" (PhD Dissertation, University of Washington, 1996), 83; Pearce and Wiggins, "Expectation in Melody," 389.

previously, probability estimates can become vanishingly small in situations like this one, so the IC estimate is often preferred, which in this case is 20.58 bits.

Context models like the one just described also often use a technique called *exclusion*, which improves the final probability estimate by reclaiming a portion of the probability mass in lower-order models that is otherwise wasted on redundant predictions. In other words, the counts for events that were predicted in the higher-order distributions do not need to be included in the calculation of the lower-order distributions. In the previous example, the $n = 4$ distribution already predicted ten of the twelve chromatic scale degrees, so the counts from these predictions can be excluded from the lower-order distributions.[40] The $n = 4$ escape probability remains the same ($\frac{10}{311+10}$), but the $n = 3$ escape probability excludes the counts from all ten chromatic scale degrees since they were already predicted in the higher-order distribution, resulting in an escape probability of $\frac{10}{0+10}$. For the $n = 2$ distribution, two novel chromatic scale degrees follow the context: $\flat\hat{6}$, which appears thrice; and $\flat\hat{7}$, which appears twice. Thus, the terminal probability estimate is $\frac{2}{5+12}$. Using backoff smoothing with exclusion, the final probability estimate is now $\frac{10}{321} \times \frac{10}{10} \times \frac{2}{17}$, or .004 (8.09 bits).

Given how infrequently $\flat\hat{7}$ follows *any* context in the Haydn Corpus, it should not be surprising that it receives such a low probability estimate in the sequence $\hat{4}$-$\hat{3}$-$\hat{2}$-$\flat\hat{7}$. For sequences that do not require escape probabilities, backoff smoothing is far simpler to compute. In the sequence $\hat{4}$-$\hat{3}$-$\hat{2}$-$\hat{1}$, for example, $\hat{1}$ follows the context in the highest-order distribution, so the PPM scheme terminates with the first probability estimate. Note, however, that since we now include an escape count in the distribution, the probability estimate for $\hat{1}$ deflates slightly from $\frac{176}{311}$, or .566 (.821 bits) in Table 6.1, to $\alpha = \frac{176}{311+10}$, or .548 (.867 bits).

The PPM scheme just described remains the canonical method in many context models,[41]

[40]Note, however, that exclusion only applies to the counts; the total number of events t following the context remains the same in every distribution.

[41]John G. Cleary and W. J. Teahan, "Unbounded Length Contexts for PPM," *The Computer Journal* 40, nos.

but Suzanne Bunton has since provided a variant smoothing technique called *mixtures* that generally improves model performance,[42] but which, following Stanley F. Chen and Joshua Goodman, I will call *interpolated smoothing*.[43] The central idea behind interpolated smoothing is to compute a weighted combination of higher-order and lower-order models for *every* event in the sequence—regardless of whether that event features n-grams with non-zero counts—under the assumption that the addition of lower-order models might generate more accurate probability estimates.[44]

Formally, interpolated smoothing estimates the probability function $p(e_i|e_{(i-n)+1}^{i-1})$ by recursively computing a weighted combination of the $(n-1)^{\text{th}}$ order distribution with the $(n-2)^{\text{th}}$ order distribution.[45]

$$p(e_i|e_{(i-n)+1}^{i-1}) = \begin{cases} \alpha(e_i|e_{(i-n)+1}^{i-1}) + \gamma(e_{(i-n)+1}^{i-1})p(e_{(i-n)+2}^{i-1}) & \text{if } e_{(i-n)+2}^{i-1} \neq \varepsilon \\ \frac{1}{|A|+1-t(\varepsilon)} & \text{otherwise} \end{cases} \tag{6.7}$$

I have already presented the equations for α and γ above, but in the context of interpolated smoothing it can be helpful to think of γ as a weighting function, with α serving as the weighted ML estimate. Unlike the backoff smoothing procedure, which terminates at the first non-zero prediction, interpolated smoothing recursively adjusts the probability estimate for each order—regardless of whether the corresponding n-gram features a non-zero count—and then terminates with the probability estimate for ε, which represents a uniform distribution

2/3 (1997): 67.

42 Suzanne Bunton, "Semantically Motivated Improvements for PPM Variants," *The Computer Journal* 40, nos. 2/3 (1997): 76–93.

43 Stanley F. Chen and Joshua Goodman, "An Empirical Study of Smoothing Techniques for Language Modeling," *Computer Speech & Language* 13 (1999): 363; Pearce and Wiggins, "Improved Methods for Statistical Modelling of Monophonic Music," 2.

44 Marcus Pearce, e-mail message to author, February 26, 2016.

45 Pearce and Wiggins, "Improved Methods for Statistical Modelling of Monophonic Music," 2; Pearce and Wiggins, "Expectation in Melody," 389.

over $|A| + 1 - t(\varepsilon)$ events (i.e., where every event in the alphabet is equally likely).[46]

Returning to the sequence $\hat{4}$-$\hat{3}$-$\hat{2}$-$\flat\hat{7}$, we can replace the formulæ for α and γ with the appropriate values from the third-order distribution in Table 6.1. Recall that $\flat\hat{7}$ never follows the context $\hat{4}$-$\hat{3}$-$\hat{2}$, so the weighted ML estimate α is $\frac{0}{311+10}$, and the escape probability γ is $\frac{10}{311+10}$.

$$p_{\text{smooth}}(\flat\hat{7}) = \overbrace{\frac{0}{321} + \frac{10}{321}}^{n\,=\,4} \cdot \; p(e_i|e^{i-1}_{(i-n)+2})$$

The second-order model only consists of veteran chromatic scale degrees (i.e., chromatic scale degrees that were already predicted in the higher-order model) and $\flat\hat{7}$ again fails to appear in the distribution, so the PPM scheme excludes all of the state counts from γ but preserves the frequency counts in α, resulting in the estimates $\frac{0}{513+10}$ for α and $\frac{10}{0+10}$ for γ.

$$p_{\text{smooth}}(\flat\hat{7}) = \overbrace{\frac{0}{321} + \frac{10}{321}}^{n\,=\,4} \cdot \left(\overbrace{\frac{0}{523} + \frac{10}{10}}^{n\,=\,3} \cdot \; p(e_i|e^{i-1}_{(i-n)+3})\right)$$

Interpolated smoothing continues to include weighted combinations of lower-order models until it reaches the zeroth-order model, at which point it terminates with the uniform distribution over A. By this point, the model has already encountered every possible chromatic scale degree, so the uniform distribution is $\frac{1}{0+1-0}$.

$$p_{\text{smooth}}(\flat\hat{7}) = \overbrace{\frac{0}{321} + \frac{10}{321}}^{\substack{n\,=\,4\\ \alpha+\gamma}} \cdot \left(\overbrace{\frac{0}{523} + \frac{10}{10}}^{\substack{n\,=\,3\\ \alpha+\gamma}} \cdot \left(\overbrace{\frac{2}{1871} + \frac{12}{17}}^{\substack{n\,=\,2\\ \alpha+\gamma}} \cdot \left(\overbrace{\frac{135}{14006} + \frac{12}{12}}^{\substack{n\,=\,1\\ \alpha+\gamma}} \cdot \overbrace{\frac{1}{1}}^{\substack{n\,=\,0\\ \frac{1}{|A|+1-t(\varepsilon)}}} \right)\right)\right)$$

[46]In the PPM scheme, the alphabet A increases by one event to accommodate the escape count t but decreases by the number of events in A that never appear in the corpus. If $\flat\hat{7}$ never appeared in the Haydn Corpus, for example, the probability estimate for ε would be $\frac{1}{12+1-1}$.

The full model for $\flat\hat{7}$ from the sequence $\hat{4}$-$\hat{3}$-$\hat{2}$-$\flat\hat{7}$ returns a final smoothed probability estimate of .022, or 5.491 bits. For comparison, I have also provided the full model for $\hat{1}$ from the sequence $\hat{4}$-$\hat{3}$-$\hat{2}$-$\hat{1}$, shown below.

$$p_{\text{smooth}}(\hat{1}) = \overbrace{\frac{176}{321}+\frac{10}{321}}^{\substack{n=4\\ \\ \alpha+\gamma}} \cdot \left(\overbrace{\frac{280}{523}+\frac{10}{10}}^{\substack{n=3\\ \\ \alpha+\gamma}} \cdot \left(\overbrace{\frac{598}{1871}+\frac{12}{17}}^{\substack{n=2\\ \\ \alpha+\gamma}} \cdot \left(\overbrace{\frac{2390}{14006}+\frac{12}{12}}^{\substack{n=1\\ \\ \alpha+\gamma}} \cdot \overbrace{\frac{1}{1}}^{\substack{n=0\\ \\ \frac{1}{|A|+1-t(\varepsilon)}}} \right)\right)\right)$$

Here, the final smoothed probability estimate is .601, or .735 bits.[47]

6.1.4 Variable Orders

In the previous examples, I arbitrarily selected a global order bound of $n = 4$, but as Pearce points out, the optimal order for context models depends on the nature of the corpus, which in the absence of a priori knowledge can only be determined empirically.[48] To resolve this issue, IDyOM employs an extension to PPM called PPM*,[49] which includes contexts of variable length and thus "eliminates the need to impose an arbitrary order bound."[50] In the PPM* scheme, the context length is allowed to vary for each event in the sequence, with the maximum context length selected using simple heuristics to minimize model uncertainty. Specifically, PPM* exploits the fact that the observed frequency of novel events is much lower than expected for contexts that feature exactly one prediction, called *deterministic* contexts. As a result, the entropy of the distributions estimated at or below deterministic contexts tends to be lower than in non-deterministic contexts. Thus, PPM* selects the shortest deterministic context to serve as the global order bound for each event in the sequence. If such a context does not exist, PPM*

[47] Pearce and Wiggins point out that with interpolated smoothing, the conditional probability distribution rarely sums to one, so IDyOM computes the entire distribution and then renormalizes the component probabilities such that they do sum to one ("Improved Methods for Statistical Modelling of Monophonic Music," 5).

[48] Ibid., 2.

[49] Cleary and Teahan, "Unbounded Length Contexts for PPM."

[50] Pearce and Wiggins, "Improved Methods for Statistical Modelling of Monophonic Music," 6.

then selects the longest matching context. In §5.1, for example, I noted that the distribution of words following the expression *Don't give up your day* very likely consists of just one word (*job*). In the context of PPM*, this means that *Don't give up your day* would serve as a deterministic context. Assuming no shorter deterministic contexts exist (e.g., *give up your day*, *up your day*, and so on), PPM* would then select $n = 6$ as the global order bound.

§6.2 Combining Models

6.2.1 Long-term vs. Short-term

To improve model performance, IDyOM separately estimates and then combines two subordinate models trained on different subsets of the corpus: a *long-term* model (LTM), which trains on the entire corpus to simulate the long-term, *schematic* knowledge of listeners; and a *short-term* model (STM), which is initially empty for each individual composition and then trains incrementally to simulate the short-term, *dynamic* knowledge of listeners.[51] Thus, the long-term model reflects inter-opus statistics from a large corpus of compositions, whereas the short-term model only reflects intra-opus statistics, some of which may be specific to that composition.[52] Like the STM, we might also slightly improve the LTM by incrementally training on the composition being predicted, called LTM+. Thus, LTM+ represents both the statistics from the training set *and* the statistics from that portion of the test set that has already been predicted. Both models generate a conditional probability distribution for each event in the corpus, so Pearce typically combines LTM+ and STM using a weighted geometric mean,[53]

[51] Pearce and Wiggins, "Auditory Expectation: The Information Dynamics of Music Perception and Cognition," 632.

[52] Pearce and Wiggins, "Improved Methods for Statistical Modelling of Monophonic Music," 6; Pearce and Wiggins, "Expectation in Melody," 389. According to Conklin, the short-term model is *transitory* since it is discarded after a particular sequence is predicted, and *dynamic* because it adapts to a particular sequence ("Multiple Viewpoint Systems for Music Prediction," 57).

[53] Pearce, Conklin, and Wiggins, "Methods for Combining Statistical Models of Music."

yielding a single conditional distribution, called BOTH+.

The LTM (or LTM+) is taken to represent long-term stylistic knowledge, or what Keith Potter, Geraint A. Wiggins and Marcus Pearce refer to as "typical Western musical experience,"[54] so it seems reasonable to conclude that probability estimates from the LTM simulate schematic expectations arising from long-term memory. Pearce and Wiggins concede, however, that the LTM exhibits total recall and its memory never fails, which may explain why it outperforms humans in some implicit learning tasks.[55] It remains far less clear which memory system the STM might simulate, however. Pearce and Wiggins claim, for example, that only the LTM and BOTH are "serious candidates as models of human cognition," while the STM alone is included in IDyOM for "completeness."[56] Potter, Wiggins, and Pearce compared the estimates from both models for two works by Philip Glass, *Gradus* and *Two Pages*, noting that the LTM exhibited a tendency towards tonal-melodic structure while the STM emphasized local melodic structure, leading them to suggest that the STM is a surrogate for short-term memory.[57] But since the STM only discards statistics when it reaches the end of the composition, it far surpasses the supposed upper limits for short-term and working memory, sometimes by several minutes.[58]

Nevertheless, the STM corresponds quite closely with David Huron's *dynamic* expectancy type, which arises from short-term memories of brief—even single—exposures.[59] Although Huron concedes that the upper limit of short-term auditory memory is likely no longer than 10–12 seconds, he notes that repetitions of patterns throughout the composition could increase

---

[54] Keith Potter, Geraint A. Wiggins, and Marcus T. Pearce, "Towards Greater Objectivity in Music Theory: Information-Dynamic Analysis of Minimalist Music," *Musicae Scientiae* 11, no. 2 (2007): 300.

[55] Rohrmeier, Rebuschat, and Cross, "Incidental and Online Learning of Melodic Structure."

[56] Pearce and Wiggins, "Auditory Expectation: The Information Dynamics of Music Perception and Cognition," 632.

[57] Potter, Wiggins, and Pearce, "Towards greater objectivity in music theory: Information-dynamic analysis of minimalist music," 299, 310.

[58] Bob Snyder suggests an upper limit of 10–12 seconds for short-term or working memory (*Music and Memory*, 50).

[59] Huron, *Sweet Anticipation*, 227.

the likelihood that the corresponding mental representation will be passed to "intermediate-term memory, and then potentially into long-term memory."[60] From this point of view, it is perhaps possible that the probability estimates from the STM reflect dynamic expectations arising from some combination of short- and intermediate-term memory, though the experimental evidence for a separate intermediate-term memory system is not well supported.

But recall from Chapter 2 that in my view, cadences exemplify exactly the kinds of inter-opus patterns that listeners are likely to store in long-term memory. If the cadence is indeed the quintessential tonal *schema*, the STM should be irrelevant for the present purposes, since it reflects the kinds of intra-opus patterns that listeners are far less likely to remember. To that end, I have elected to omit the STM in the analyses that follow and only present the probability estimates from LTM+, but with the hope that future studies examining the relationship between expectancy and cadential closure might compare the results obtained here with probability estimates from the STM and BOTH+.

6.2.2 Performance Evaluation

Context models like IDyOM depend on a training set and a test set, but in this case the Haydn Corpus will need to serve as both. To accommodate small corpora like this one, IDyOM employs a resampling approach called *k-fold cross-validation*,[61] using cross entropy as a measure of performance.[62] The corpus is divided into k disjoint subsets of approximately equal size, and the LTM+ is trained k times on $k-1$ subsets, each time leaving out a different subset for testing. IDyOM then computes an average of the k cross entropy values as a measure of the model's performance. Following Pearce and Wiggins, I use 10-fold cross validation for the models that

[60] Huron, *Sweet Anticipation*, 228.

[61] Thomas G. Dietterich, "Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms," *Neural Computation* 10, no. 7 (1998): 1895–1923.

[62] Conklin and Witten, "Multiple Viewpoint Systems for Music Prediction."

follow.[63]

6.2.3 Viewpoint Selection

With the PPM* scheme now in place, we should be able to estimate the conditional probability of every event in the Haydn Corpus, but which viewpoints do we predict? Recall from Chapter 3 that the multiple-viewpoint framework represents sequences of note events from a notated score according to attribute types like chromatic pitch (`cpitch`), melodic interval (`melint`) and chromatic scale degree (`csd`). IDyOM is capable of individually predicting any one of these viewpoints using the PPM* scheme, but it can also *combine* viewpoint predictions using a weighted multiplicative combination scheme that assigns greater weights to viewpoint models whose predictions are associated with lower entropy at that point in the sequence.[64] To determine the combined probability distribution for each event in the test sequence, IDyOM then computes the product of the weighted probability estimates from each viewpoint model for each possible value of the predicted viewpoint.

What is more, IDyOM can automate the viewpoint selection process using a hill-climbing procedure called *forward stepwise selection*, which picks the combination of viewpoints that yields the richest structural representations of the musical surface and minimizes model uncertainty. Given an empty set of viewpoints, the stepwise selection algorithm iteratively selects the viewpoint model additions or deletions that yield the most improvement in cross entropy, terminating when no addition or deletion yields an improvement.[65] The assumption here is that, like IDyOM, listeners seek to minimize their uncertainty about future events by selecting reduced representations of the musical surface. By way of example, Pearce and Wiggins note

[63] Pearce and Wiggins, "Improved Methods for Statistical Modelling of Monophonic Music," 10.

[64] Pearce, Conklin, and Wiggins, "Methods for Combining Statistical Models of Music," 304.

[65] Pearce, "The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition," 121–122; Potter, Wiggins, and Pearce, "Towards greater objectivity in music theory: Information-dynamic analysis of minimalist music," 302.

that infants demonstrate absolute pitch, but as exposure to music increases, the vast majority quickly learn relative pitch.[66] Viewpoint representations of the former (like `cpitch`) require far more symbols than those of the latter (like `melint`), so it is conceivable that humans compress multidimensional stimulus domains like music by discarding representations like `cpitch` in favor of those like `melint`.

IDyOM was originally intended to simulate melodic pitch expectations, so the majority of the published corpus studies using IDyOM predict pitch-based viewpoints like `cpitch` or `melint`,[67] or linked viewpoints like `csd` ⊗ `melint`.[68] What is more, only a few studies employ the viewpoint selection scheme just described.[69] Viewpoint models predicting rhythmic or metric attributes are much less common, though Pearce, Müllensiefen, and Wiggins recently demonstrated the degree to which IDyOM can simulate the behavior of listeners in segmentation tasks using viewpoint predictions of inter-onset interval and offset-to-onset interval.[70]

In Chapter 3, I represented the "core" events of the classical cadence according to pitch-based viewpoints from the outer parts (`csd` and `contour`), a coefficient representing the strength of the metric position (`strength`), and a vertical sonority, presented as a combination of vertical interval classes (`vintcc`) or chromatic scale degrees (`csdc`).[71] For the majority of the encoded cadences from the cadence collection (see §5.1), the terminal events at the moment of cadential

[66] I should point out, however, that some studies have shown that infants already demonstrate relative pitch abilities at an early stage of development. See, for example, Judy Plantinga and Laurel J. Trainor, "Memory for Melody: Infants Use a Relative Pitch Code," *Cognition* 98 (2005): 1–11.

[67] Pearce and Wiggins, "Improved Methods for Statistical Modelling of Monophonic Music"; Ian H. Witten, Leonard C. Manzara, and Darrell Conklin, "Comparing Human and Computational Models of Music Prediction," *Computer Music Journal* 18, no. 1 (1992): 70–80.

[68] Pearce et al., "Unsupervised Statistical Learning Underpins Computational, Behavioural, and Neural Manifestations of Musical Expectation"; Conklin and Witten, "Multiple Viewpoint Systems for Music Prediction."

[69] Pearce and Wiggins, "Expectation in Melody"; Potter, Wiggins, and Pearce, "Towards greater objectivity in music theory: Information-dynamic analysis of minimalist music." Conklin also uses a variant of this technique ("Multiple Viewpoint Systems for Music Prediction").

[70] Pearce, Müllensiefen, and Wiggins, "The Role of Expectation and Probabilistic Learning in Auditory Boundary Perception."

[71] Gjerdingen, *Music in the Galant style*, 142.

arrival appear in strong metric positions (i.e., `strength`={3,4}), and few of the cadences feature unexpected durations or inter-onset intervals at the cadential arrival, so I have elected to omit viewpoint models for rhythmic or metric attributes from the present investigation and concentrate instead on those viewpoints representing either melodic or harmonic expectations.

Table 6.2 presents the reference information, domain, and cardinality for each of the predicted viewpoints in the analyses that follow. For `cpitch`, for example, the MIDI protocol specifies 128 possible chromatic pitches, but only 64 appear in the first violin and cello parts in the Haydn Corpus. To examine melodic pitch expectations for the resolving note event at the moment of cadential arrival in the first violin and cello, I have included viewpoint predictions for `cpitch`, `melint`, and `csd`. Following Gjerdingen's schema-theoretic approach, the analyses in Chapters 4 and 5 also included a viewpoint representing the melodic contour between note events, but in this case the `contour` model received a much higher cross entropy estimate than the other viewpoint models, so it was excluded here. To derive the optimal viewpoint system for the representation of melodic expectations, I also employed the stepwise selection procedure mentioned above for the following viewpoints: `cpitch`, `melint`, `csd`, and `contour`. In this case, IDyOM begins with the above set of viewpoint models, but also includes the linked viewpoints derived from that set (i.e., `cpitch` ⊗ `melint`, `cpitch` ⊗ `csd`, `cpitch` ⊗ `contour`, `melint` ⊗ `csd`, `melint` ⊗ `contour`, `csd` ⊗ `contour`), resulting in a pool of ten individual viewpoint models from which to derive the optimal combination of viewpoints.

Shown in Table 6.3, viewpoint selection derived the same combination of viewpoint models for the first violin and the cello. Recall from §6.1.2 that cross entropy, denoted by H_m, estimates how uncertain a model is, on average, when predicting a given sequence of events. Thus, for our purposes, the model with the lowest cross entropy will serve as the best performing model because it minimizes uncertainty. For the Haydn Corpus, `melint` was the best performing viewpoint model in the first step, receiving a cross entropy estimate of 3.006 in the first violin

Table 6.2: Reference information, domain, and cardinality for the predicted viewpoints used in the present research.

Viewpoint	*Description*	*Domain*	*Cardinality in Haydn Corpus*
Note Events			
`cpitch`	chromatic pitch	$\{0,\ldots,127\}$	64
`melint`	melodic pitch interval	$\{-126,\ldots,126\}$	63
`csd`	chromatic scale degree	$\{0,\ldots,11\}$	12
Chord Events			
`vintcc`	vertical interval class combination	$\{<0,\perp,\perp>,\ldots,<9,10,11>\}$	190
`csdc`	chromatic scale degree combination	$\{<0,\perp,\perp,\perp>,\ldots,<11,8,9,10>\}$	688

and 2.798 in the cello. In the second step, the combination of `melint` with the linked viewpoint `csd` ⊗ `cpitch` decreased the cross entropy estimate to 2.765 in the first violin and 2.556 in the cello. Including any of the remaining viewpoints did not improve model performance, however, so the stepwise selection procedure terminated with this combination of viewpoints. In §6.3 I refer to this viewpoint model as `selection`.

Currently, IDyOM is not designed to accommodate viewpoints for chord events, but in Chapter 3 I extended the multiple-viewpoint framework to include the viewpoints `vintcc` and `csdc`. Recall from §3.4.2 that the chord types from `vintcc` loosely approximate symbols from figured bass nomenclature, whereas `csdc` defines its chord types in relation to an underlying tonic. As such, the probabilities estimated by IDyOM for `vintcc` and `csdc` simulate the formation of harmonic expectations during listening.

Finally, I have also created a composite viewpoint that might represent those viewpoint models characterizing pitch-based (i.e., melodic *and* harmonic) expectations more generally. To simulate the cognitive mechanisms underlying melodic segmentation, Pearce, Müllensiefen, and Wiggins found it beneficial to combine viewpoint predictions for basic attributes like chromatic pitch, inter-onset interval, and offset-to-onset interval by multiplying the component

Table 6.3: The results of viewpoint selection for reduced cross entropy in the Haydn Corpus.

Step	Viewpoint Added	H_m	
		Violin 1	Cello
1	`melint`	3.006	2.798
2	`csd` ⊗ `cpitch`	2.765	2.556

probabilities to reach an overall probability for each note in the sequence as the joint probability of the individual basic attributes being predicted.[72] Following their approach, the viewpoint model `composite` represents the product of the `selection` viewpoint model from the first violin (to represent melodic expectations) and the `csdc` viewpoint model (to represent harmonic expectations) for each unique onset time for which a note *and* chord event are present in the Haydn Corpus. In this case I have preferred `csdc` over `vintcc` in the `composite` model because the former viewpoint explicitly encodes the chromatic scale-degree successions in the lowest instrumental part along with the relevant scale-degrees from the upper parts. Thus, `csdc` better reflects the underlying tonal context, so it should better simulate harmonic expectations.

§6.3 Results

6.3.1 Cadences vs. Non-Cadences

I noted in Chapter 2 that Caplin classifies every possible cadence category according to two fundamental types: those for which the goal of the cadential progression is tonic (the perfect authentic cadence and its variants), and those for which the goal is dominant (the half cadence and its variants). These two categories account for 206 of the 245 cadences from the collection. Thus, it seems reasonable that listeners with sufficient exposure to music of the classical style

[72] Pearce, Müllensiefen, and Wiggins, "The Role of Expectation and Probabilistic Learning in Auditory Boundary Perception," 1376.

will form schematic expectations for the terminal events of exemplars from the PAC and HC categories. What is more, if cadences are the most predictable formulæ in all of tonal music, we should expect to find lower IC estimates for the terminal events from the aforementioned cadence categories compared to those from non-cadential closing contexts even if they both share similar or even identical terminal events. That is, cadences should be more predictable than their non-cadential counterparts.

To compare the PAC and HC categories against non-cadential contexts exhibiting varying degrees of closure or stability, each of the viewpoints estimated by IDyOM was analyzed for the terminal note events from the first violin and cello—represented by the viewpoints `cpitch`, `melint`, `csd`, and `selection`—and the terminal chord events from the entire texture—represented by the viewpoints `vintcc`, `csdc`, and `composite`—using a one-way analysis of variance (ANOVA) with a three-level between-groups factor called *closure*. To examine the IC estimates for Caplin's first (tonic) type, *tonic closure* consists of three levels: *PAC*, which consists of the IC estimates for the terminal events from the 122 exemplars of the PAC category; *tonic*, which consists of an equal-sized sample of events selected randomly from the Haydn Corpus that appear in strong metric positions ($\texttt{strength} > 1$) and feature tonic harmony in root position and $\hat{1}$ in the soprano; and *non-tonic*, which again consists of an equal-sized sample of events selected randomly from the Haydn Corpus that appear in strong metric positions, but that feature any harmony and any scale-degree in the soprano.

To examine the IC estimates for Caplin's second (dominant) type, *dominant closure* was designed in much the same way. *HC* consists of the IC estimates for the terminal events from the 84 exemplars of the HC category, while *dominant* and *non-dominant* consist of equal-sized samples of dominant and non-dominant events selected at random. Dominant events appear in strong metric positions and feature dominant harmony in root position but (like their cadential counterparts) permit any chord member in the soprano, while non-dominant events appear in strong metric positions but feature any harmony and any scale-degree.

In hypothesis testing theory, an ANOVA computes Fisher's F ratio,[73] which represents the probability of obtaining differences between the group means arising from the experimental sample (i.e., the between-group variability) that are equal to or greater than the differences between means arising from the population (i.e., the within-group variability).[74] If the probability is small—less than, say, .05 (called the *significance level*, and represented by α)—we may claim that the observed differences are unlikely to have occurred due to sampling error alone and thus reject the *null hypothesis* that no differences exist between the group means.[75]

Given the popularity of the F ratio in experimental research,[76] the specifics of the technique need not concern us here, but the statistical assumptions made prior to its estimation deserve further mention. Inferential statistical procedures are said to be *robust* when the probability statements resulting from their application are insensitive to violations of the assumptions made in their development.[77] In this case, ANOVAs assume that the distributions characterizing each group are normal (or bell-shaped), that the group variances are more or less equal (or homoscedastic), and that the observations in each group are statistically independent. ANOVAs are generally quite robust to violations of normality, but much less so to violations of homo-

[73] Ronald A. Fisher, *Statistical Methods, Experimental Design, and Scientific Inference* (Oxford: Oxford University Press, 1990).

[74] Ferguson and Takane, *Statistical Analysis in Psychology and Education*, 180.

[75] The significance level is generally a matter of convention. Ronald A. Fisher is perhaps the first statistician to suggest a significance level of .05 in his famous "lady tasting tea" experiment, which tested the claim that a lady could discern just by tasting a cup of tea with milk whether the milk or the tea infusion was first added to the cup. He noted that for eight cups of tea presented in random order—four prepared first by adding the milk, and four by first adding the tea—the chances of correctly discerning the order is one in seventy. Fisher would only believe her claim if she could correctly classify all eight cups of tea; his standard of evidence, or *significance level*, was thus $\frac{1}{70}$, or $p = .014$. "It is open to the experimenter to be more or less exacting in respect to the smallness of the probability he would require before he would be willing to admit that his observations have demonstrated a positive result... [but] it is usual and convenient for experimenters to take 5 per cent. as a standard level of significance, in the sense that they are prepared to ignore all results which fail to reach this standard" (*The Design of Experiments*, 8th [New York: Hafner Publishing Company, 1971], 13) (my addition).

[76] There are many fine introductory texts on hypothesis testing and inferential statistics. See, for example, Ferguson and Takane, *Statistical Analysis in Psychology and Education*.

[77] Ibid., 191.

geneity of variance. If the ratio of the variances between any two groups is greater than 2:1 (or smaller than 1:2), for example, such heteroscedastic groups tend to inflate statistics like the F ratio, leading researchers to incorrectly reject the null hypothesis (called a *Type I* error). This problem is particularly acute when the groups are unbalanced (i.e., not of equal size), since variance generally increases as sample size decreases. The largest and smallest cadence categories from the cadence collection consist of 122 and 9 cadences, for example, which is very likely to lead to heterogeneous variances among the groups. As a result, implementing traditional inferential techniques like the F ratio for the datasets from this chapter will almost invariably increase the risk of producing a Type I error.

For every between-groups factor examined in this section, Levene's equality of variances, which tests the null hypothesis that the population variances are equal, revealed significant differences between groups for nearly every viewpoint model. Thus, in what follows I employ an alternative to Fisher's F ratio that is generally robust to heteroscedastic data, called the Welch F ratio.[78] Again, this technique is fairly common in inferential statistics, so I will forgo the formula here. Along with some measure of the F ratio, it is also common practice in hypothesis testing to measure the magnitude of the differences between groups (called the *effect size*).[79] To determine the effect size both for the Welch F ratio and for the planned comparisons described shortly, I use Barry Cohen's recent notation of a common effect size measure called estimated ω^2.

$$\text{est. } \omega^2 = \frac{\text{df}_{\text{bet}}(F - 1)}{\text{df}_{\text{bet}}(F - 1) + N_T}$$

The term df_{bet} denotes the between-groups degrees of freedom, F refers to the Welch F-ratio,

[78] Bernard L. Welch, "On the Comparison of Several Mean Values: An Alternative Approach," *Biometrika* 38, nos. 3/4 (1951): 330–336.

[79] Remember that the F ratio indicates whether the differences among group means are significant; it does not measure the size (or magnitude) of those differences.

and N_T represents the total number of IC estimates in the model.[80] Thus, est. ω^2 falls in a range between 0 and 1. Unlike the more common η^2, which estimates the effect size for measures calculated from the sample, Cohen's formula tends to produce smaller effect sizes because it represents an unbiased estimate of the proportion of variance accounted for in the population. For a three-group factor like *closure*, for example, an est. ω^2 of .01 accounts for one percent of the variance in the outcome variable and represents a small but interpretable effect size,[81] whereas an est. ω^2 of .14 or greater generally represents a large effect.

The F ratio is an *omnibus* test, which is to say that it indicates whether the differences among group means are statistically significant, not *where* these differences lie. To address more specific hypotheses about the potential differences between groups, each model also includes planned comparisons that do not assume equal variances. Thus, in this section each model included two planned comparisons to examine differences in the IC estimates for the terminal events from cadential and non-cadential contexts: the first to determine whether the IC estimates from the corresponding cadence category differ significantly from the two non-cadential levels (*Cadences* vs. *Non-Cadences*), and the second to determine whether the IC estimates from the corresponding cadence category differ significantly from the second (*tonic* or *dominant*) level of *closure* (*PAC* vs *Tonic* or *HC* vs. *Dominant*). Unfortunately, these additional tests increase the risk of committing a Type I error, so I apply a Bonferroni correction to the planned comparisons, which multiplies the p value by the number of tests. As such, an initial p value of .02 would still be significant at $\alpha = .05$ ($.02 \times 2 = .04$), whereas a p value of .04 would no longer be significant ($.04 \times 2 = .08$). This approach is admittedly quite conservative, but it further ensures that the reported findings could not have occurred by chance alone.

The top bar plots in Figure 6.1 display the mean IC estimates for the terminal note event in

[80]Barry H. Cohen, *Explaining Psychological Statistics* (Hoboken, NJ: John Wiley & Sons, Inc., 2008), 374–375.
[81]Ibid., 375.

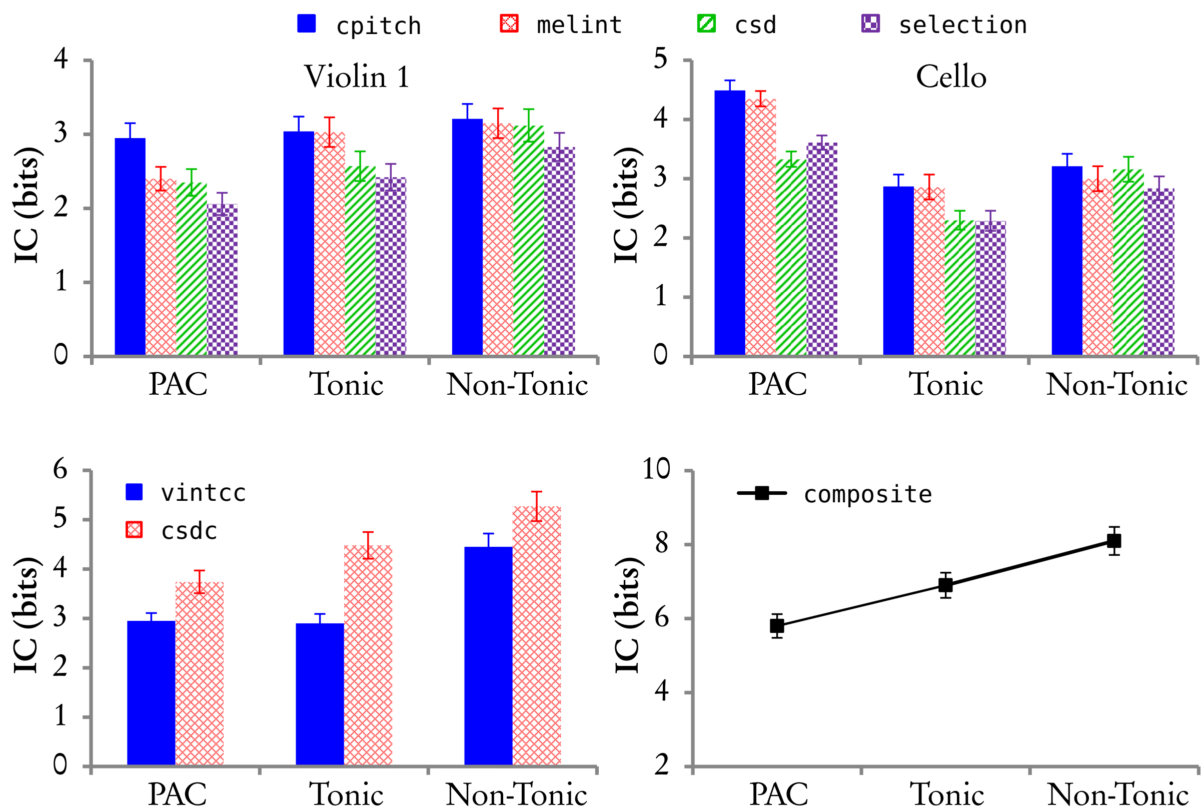


Figure 6.1: Top: Bar plots of the mean information content (IC) estimated for the terminal note event in the first violin (left) and cello (right) for each level of *tonic closure*. Viewpoints include `cpitch`, `melint`, `csd`, and an optimized combination called `selection`, which represents `melint` and the linked viewpoint `csd` ⊗ `cpitch`. Bottom left: Bar plot of the mean information content (IC) estimated for the terminal `vintcc` and `csdc` for each level of *tonic closure*. Bottom right: Line plot of the mean information content (IC) estimated for the combination called `composite`, which represents the dot product of $\texttt{selection}_{\text{vl1}}$ and `csdc`. Whiskers represent ±1 standard error.

the first violin (left) and cello (right) for each level of *tonic closure*. Beginning with the first violin, one-way ANOVAs of the IC estimates revealed a main effect for the viewpoints `melint`, $F(2, 238.95) = 5.18$, $p < .01$, est. $\omega^2 = .02$, `csd`, $F(2, 240.42) = 3.84$, $p < .05$, est. $\omega^2 = .02$, and the optimized combination `selection`, $F(2, 239.01) = 4.98$, $p < .01$, est. $\omega^2 = .02$. The ANOVA for `cpitch` was not significant, $F(2, 241.94) = .45$, $p > .05$. Mean IC estimates increased significantly from *PAC* to the non-cadential levels of *closure* for `melint`, $t(294.75) = -3.19$,

$p < .01$, $r = .18$, and `selection`, $t(293.73) = -2.80$, $p < .01$, $r = .16$, but the comparison for `csd` was marginal, $t(278.36) = -2.12$, $p = .07$, $r = .13$.[82] Although this trend also emerged for the planned comparison between *PAC* and *tonic*, only the `melint` model revealed a significant effect, $t(230.09) = -2.42$, $p < .05$, $r = .16$. Thus, the viewpoint models for the first violin demonstrated that terminal note events from cadential contexts are more predictable than those from non-cadential contexts.

For the cello, one-way ANOVAs of the IC estimates revealed a main effect of *tonic closure* for every viewpoint (`cpitch`, $F(2, 240.13) = 21.83$, $p < .001$, est. $\omega^2 = .10$; `melint`, $F(2, 227.14) = 26.96$, $p < .001$, est. $\omega^2 = .12$; `csd`, $F(2, 234.38) = 12.67$, $p < .001$, est. $\omega^2 = .06$; `selection`, $F(2, 230.64) = 21.78$, $p < .001$, est. $\omega^2 = .10$), but the direction of the effect was reversed. Mean IC estimates decreased in every model from *PAC* to the non-cadential levels of *tonic closure* (`cpitch`, $t(282.33) = 6.46$, $p < .01$, $r = .36$; `melint`, $t(349.56) = 7.33$, $p < .01$, $r = ..37$; `csd`, $t(314.87) = 3.15$, $p < .01$, $r = .17$; `selection`, $t(336.39) = 5.95$, $p < .01$, $r = .31$), as well as from *PAC* to *tonic* (`cpitch`, $t(237.07) = 6.18$, $p < .01$, $r = .37$; `melint`, $t(201.16) = 6.18$, $p < .01$, $r = .40$; `csd`, $t(234.27) = 4.92$, $p < .01$, $r = .31$; `selection`, $t(217.90) = 6.44$, $p < .01$, $r = .40$). Thus, contrary to our predictions, the terminal events in the cello from cadential contexts were actually *less* predictable than those from non-cadential contexts. Why might this be so?

In §4.1 I presented distributional evidence that distinguished each instrumental part in Haydn's string quartets, noting that the most stable scale degrees and metric positions appear more frequently in the cello than they do in the upper parts. What is more, the simple interval distributions presented in Figures 4.3 and 4.4 indicate that perfect intervals like unisons and

[82] Recall from §4.1 that the Pearson correlation, denoted by r, is a common effect size measure that represents the magnitude of the relationship between two variables X and Y, giving a value between -1 and $+1$. A negative value indicates a negative relation (e.g., X decreases as Y increases), whereas a positive value indicates a positive relation (e.g., X increases as Y increases), and 0 indicates no correlation between X and Y (see Chapter 4, footnote 26).

fifths tend to appear more frequently in the cello than they do in the upper parts, and that larger intervals tend to descend in the cello and ascend in the upper parts. To explain these differences, I suggested that the contrapuntal interaction between the outer voices plays an especially important role, with the bass voice moving by step and by leap to provide harmonic support for the primarily stepwise soprano. And yet the kinds of melodic intervals found in the bass voice in cadential contexts—such as the leap by ascending fourth or descending fifth from $\hat{5}$ to $\hat{1}$—still appear with far less frequency than motions by smaller intervals like unisons and seconds. In other words, even though larger intervals appear more often in the cello than they do in the upper parts, these intervals are still quite rare in the cello part, resulting in Markov models that expect motion by unison, repetitions of stable scale degrees like $\hat{1}$ and $\hat{5}$, and so forth. As a consequence, IDyOM produces considerably higher IC estimates for successions like $\hat{5}$-$\hat{1}$ than for successions like $\hat{1}$-$\hat{1}$, $\hat{2}$-$\hat{1}$, or $\hat{7}$-$\hat{1}$.

My claim, then, is that statistical models conducted on the bass voice tend to obscure the many recurrent patterns contained therein because these patterns depend on the interaction among numerous parameters across the entire texture. To determine whether listeners expect the terminal events of cadences and other frequently-occurring patterns in the lower voice of the two-voice framework, we might instead examine the harmonies formed between the cello part and the upper parts.

The bottom-left bar plot in Figure 6.1 displays the mean IC estimates for the terminal chord event—represented by `vintcc` and `csdc`—for each level of the between-groups factor. One-way ANOVAs of the IC estimates revealed a main effect of *tonic closure* for `vintcc`, $F(2, 233.40) = 13.41$, $p < .001$, est. $\omega^2 = .06$, and for `csdc`, $F(2, 240.79) = 8.12$, $p < .001$, est. $\omega^2 = .04$. As expected, both viewpoint models increased from *PAC* to the non-cadential levels of *tonic closure* (`vintcc`, $t(315.61) = -3.19$, $p < .01$, $r = .18$; `csdc`, $t(287.99) = -3.68$, $p < .01$, $r = .21$). A marginal trend also emerged for `csdc`, with mean IC estimates increasing from *PAC* to *tonic*, $t(237.61) = -2.09$, $p = .08$, $r = .13$, but this trend was not present in `vintcc`, $t(235.46) = .19$, $p > .05$. Thus, for both viewpoint models the terminal chord events from cadential contexts were more predictable than those from non-cadential contexts. Moreover, `csdc` indicated an ascending staircase, with the mean IC estimates increasing from *PAC* to *tonic* to *non-tonic*.

I noted in the previous section that `csdc` is more representative of the lower voice in the two-voice framework because it explicitly encodes the chromatic scale-degree successions in the lowest instrumental part along with the relevant scale-degrees from the upper parts. To represent the predictability of the harmony and melody in a single IC estimate for each note/chord event, I created a `composite` viewpoint that reflects the dot product of the estimates from `csdc` and $\texttt{selection}_{\mathrm{vl1}}$. The bottom-right line plot in Figure 6.1 displays the mean IC estimates for the terminal `composite` event for each level of *tonic closure*. A one-way ANOVA of the IC estimates revealed a significant main effect, $F(2, 240.79) = 10.62$, $p < .001$, est. $\omega^2 = .05$. Mean IC estimates increased from *PAC* to the non-cadential levels of *closure*, $t(270.22) = -4.13$, $p < .001$, $r = .24$, as well as from *PAC* to *tonic*, $t(240.83) = -2.34$, $p < .01$, $r = .15$. Thus, `composite` again demonstrated an ascending staircase for the levels of *tonic closure*, with *PAC* receiving the lowest IC estimates and *non-tonic* receiving the highest IC estimates.

The top bar plots in Figure 6.2 display the mean IC estimates for the terminal note event in the first violin (left) and cello (right) for each level of *dominant closure*. For the first violin, one-way ANOVAs of the mean IC estimates revealed a main effect for the viewpoints `cpitch`, $F(2, 163.24) = 3.50, p < .05$, est. $\omega^2 = .01$, `melint`, $F(2, 158.07) = 5.95, p < .01$, est. $\omega^2 = .03$, and `selection`, $F(2, 161.43) = 4.62, p < .05$, est. $\omega^2 = .02$, but not for `csd`, $F(2, 163.40) = 2.47, p = .09$. Excepting the `csd` model, the mean IC estimates increased significantly from *HC* to the non-cadential levels of *dominant closure* for every viewpoint model (`cpitch`, $t(205.51) = -2.37, p < .05, r = .16$; `melint`, $t(223.14) = -3.05, p < .01, r = .20$; `selection`, $t(217.95) =$

$-2.95, p < .01, r = .20$). The size of these effects increased in the second planned comparison when the *non-dominant* level was omitted (`cpitch`, $t(151.89) = -2.64, p < .05, r = .21$; `melint`, $t(128.28) = -3.45, p < .01, r = .29$; `selection`, $t(146.49) = -2.87, p < .01, r = .23$), indicating that terminal events from half-cadential contexts are more predictable than those from non-cadential contexts featuring dominant harmony in root-position and a chord member in the soprano voice (*dominant*), but not necessarily more predictable than non-cadential contexts featuring any harmony and any scale-degree (*non-dominant*).

For the cello, the mean IC estimates demonstrated a significant effect of *dominant closure* for `cpitch`, $F(2, 165.13) = 12.32, p < .001$, est. $\omega^2 = .06$, and `melint`, $F(2, 158.07) = 5.95, p < .01$, est. $\omega^2 = .03$, but not for `csd`, $F(2, 163.40) = .46, p > .05$, and `selection`, $F(2, 164.82) = 1.14, p > .05$. The direction of the effect was again reversed, with *HC* receiving higher IC estimates than the non-cadential contexts (`cpitch`, $t(186.99) = 3.67, p < .01, r = .26$; `melint`, $t(200.46) = 2.64, p < .05, r = .18$). What is more, although a similar trend emerged for the mean IC estimates for `vintcc` and `csdc` compared to the estimates for the first violin, there was no main effect of *dominant closure* for `vintcc`, $F(2, 157.58) = .51, p > .05$, or `csdc`, $F(2, 165.15) = .80, p > .05$, suggesting that the terminal note and chord events represented in the cello and the entire multi-voiced texture were not more predictable in half-cadential contexts than in non-cadential contexts. The `composite` viewpoint did demonstrate a significant main effect of *dominant closure*, however, $F(2, 163.27) = 3.17, p < .05$, est. $\omega^2 = .01$. Mean IC estimates marginally increased from *HC* to the non-cadential levels of *dominant closure*, $t(206.73) = -2.21$, $p = .06$, $r = .15$, and also significantly increased from *HC* to *dominant*, $t(158.72) = -2.51$, $p < .05$, $r = .20$.

In sum, the effects were generally smaller for *dominant closure* than for *tonic closure*, but both between-groups factors demonstrated significantly lower mean IC estimates for the terminal events from cadential contexts compared to those from non-cadential contexts. The factor

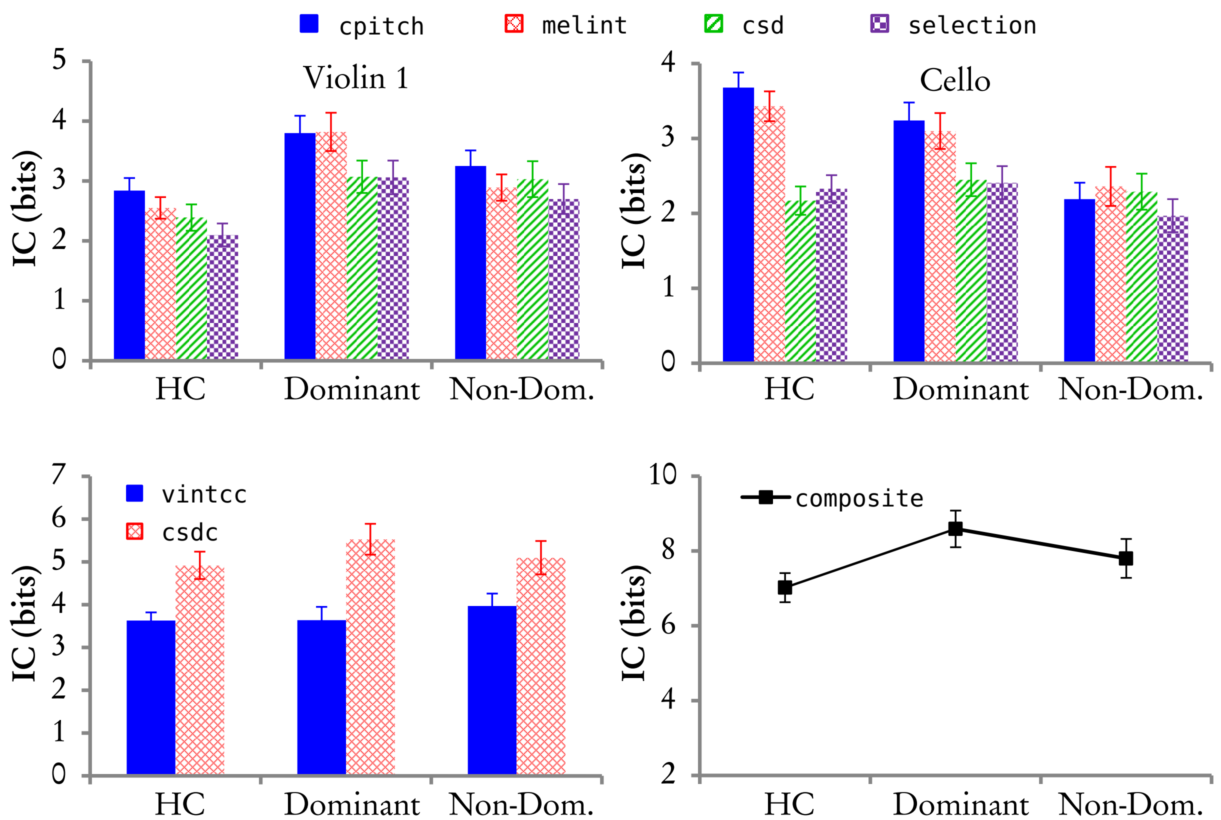


Figure 6.2: Top: Bar plots of the mean IC estimated for the terminal note event in the first violin (left) and cello (right) for each level of *dominant closure*. Viewpoints include `cpitch`, `melint`, `csd`, and an optimized combination called `selection`, which represents `melint` and the linked viewpoint `csd` ⊗ `cpitch`. Bottom left: Bar plot of the mean IC estimated for the terminal `vintcc` and `csdc` for each level of *dominant closure*. Bottom right: Line plot of the mean IC estimated for the combination called `composite`, which represents the dot product of $\texttt{selection}_{\text{vl1}}$ and `csdc`. Whiskers represent ±1 standard error.

tonic closure elicited significant effects for viewpoints representing both voices of the two-voice framework, with greater effect sizes appearing for viewpoint models characterizing harmonic progressions (`vintcc`, `csdc`, and `composite`). For *dominant closure*, significant effects were generally limited to viewpoints representing the first violin, and half-cadential contexts only elicited lower mean IC estimates in reference to non-cadential root-position dominants. To be sure, events from the *HC* condition were no more (or less) predictable than non-dominant events selected at random from the Haydn Corpus.

Given our earlier assumptions about schematic expectations for dominant events, these results should not be surprising. According to IDyOM, the terminal events from perfect authentic cadences are generally more predictable than those from non-cadential contexts, but half cadences appear to be no more predictable on average than any of the other events in the classical style. Nevertheless, it remains unclear whether terminal events from half cadences receive higher IC estimates on average because the preceding context fails to stimulate strong expectations for any particular continuation, or because the actual continuation is unexpected.[83] And yet, by only considering the potential differences between cadential and non-cadential contexts, the previous analysis failed to directly compare the cadence categories from Caplin's typology. We might imagine, for example, that the strength and specificity of our schematic expectations formed in prospect and their subsequent realization in retrospect contributes to the perception of cadential *strength*, where the most expected (i.e., probable) endings are also the most complete or closed. From this point of view, the probabilities estimated by IDyOM might correspond with models of cadential strength advanced in the *Formenlehre* tradition.

6.3.2 Cadential Strength

To compare the mean IC estimates for the terminal events from each cadence category, each of the viewpoints was again analyzed for the terminal note events from the first violin and cello and the terminal chord events from the entire texture using a one-way ANOVA with a five-level between-groups factor called *cadence category* (PAC, IAC, HC, DC, and EV). For many of the viewpoint models, Levene's test revealed heteroscedastic groups for the unbalanced levels of *cadence category*, so I again report the Welch F ratio and estimate effect size using est. ω^2.

[83] Pearce, Müllensiefen, and Wiggins, "The Role of Expectation and Probabilistic Learning in Auditory Boundary Perception," 1374–1375.

To examine the potential differences in the IC estimates for the terminal events from each cadence category, each model includes three planned comparisons that do not assume equal variances. To decrease the risk of a Type I error, I again apply the Bonferroni correction (i.e., $\frac{\alpha}{3}$). In the first comparison I coded each level of *cadence category* to represent two models of cadential strength: *Genuine Schemas* (*PAC*→*IAC*→*HC*→*DC*→*EV*) and *1-Schema* (*PAC*→*IAC*→*DC*→*EV*→*HC*). Polynomial contrasts with linear and quadratic terms were then included to estimate the goodness-of-fit for each model. In what follows I report the contrast whose linear or quadratic trend accounts for the greatest proportion of variance in the outcome variable, represented here by R^2. The second comparison examines my earlier claim that the genuine cadence categories elicit lower IC estimates on average than the cadential deviations (*Genuine* vs. *Deviations*). Finally, the third comparison determines whether the authentic cadence categories elicit lower IC estimates on average than the HC category (*AC* vs. *HC*).

Figure 6.3 displays line plots of the mean IC estimates for the terminal note event in the first violin (left) and cello (right) for each level of *cadence category*. For the first violin, one-way ANOVAs of the mean IC estimates revealed a main effect for the viewpoints `cpitch`, $F(4, 32.42) = 3.19, p < .05, \text{est. } \omega^2 = .03$, `csd`, $F(4, 29.95) = 3.02, p < .05, \text{est. } \omega^2 = .03$, and `selection`, $F(4, 29.80) = 3.77, p < .05, \text{est. } \omega^2 = .04$, but not for `melint`, $F(4, 31.57) = 2.34, p = .08$. Moreover, the best-fitting polynomial contrast revealed a positive (increasing) linear trend in the *Genuine Schemas* model (i.e., from the PAC to the EV categories) for every viewpoint model (`cpitch`, $t(13.47) = 3.40, p < .05, r = .68, R^2 = .46$; `melint`, $t(12.13) = 3.00, p < .05, r = .65, R^2 = .43$; `csd`, $t(18.99) = 3.43, p < .01, r = .62, R^2 = .38$; `selection`, $t(16.38) = 3.85, p < .01, r = .69, R^2 = .47$). The genuine cadence categories also received lower mean IC estimates than the cadential deviations in every viewpoint model (`cpitch`, $t(17.19) = -3.64, p < .01, r = .66$; `melint`, $t(14.14) = -2.94, p < .05, r = .62$; `csd`, $t(30.51) = -3.17, p < .01, r = .50$; `selection`, $t(26.03) = -3.66, p < .01, r = .58$). The authentic cadence

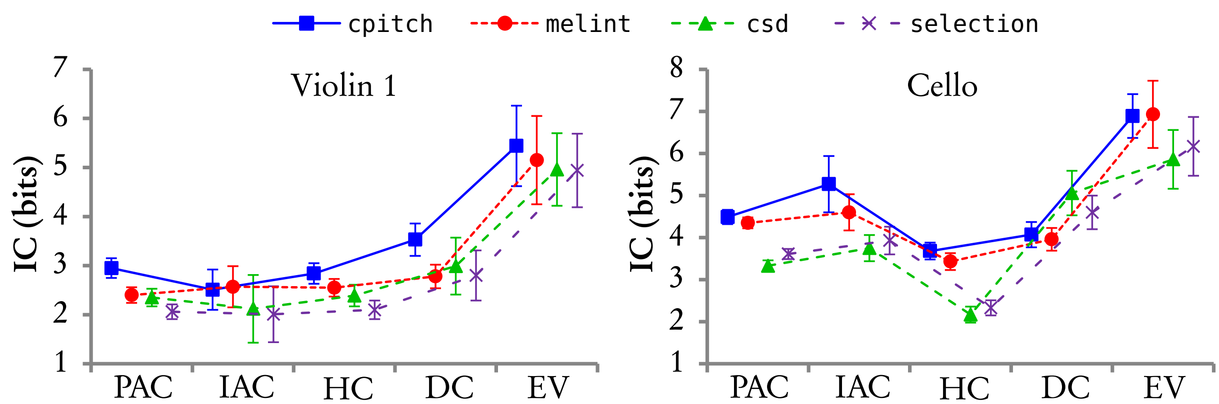


Figure 6.3: Line plots of the mean IC estimated for the terminal note event in the first violin (left) and cello (right) for each cadence category. Viewpoints include `cpitch`, `melint`, `csd`, and an optimized combination called `selection`, which represents `melint` and the linked viewpoint `csd` ⊗ `cpitch`. Whiskers represent ±1 standard error.

categories did not differ significantly from the HC category, however.

For the cello, one-way ANOVAs of the IC estimates revealed a main effect of *cadence category* for every viewpoint (`cpitch`, $F(4, 31.18) = 8.86, p < .001$, est. $\omega^2 = .11$; `melint`, $F(4, 30.72) = 6.81, p < .001$, est. $\omega^2 = .09$; `csd`, $F(4, 31.25) = 13.99, p < .001$, est. $\omega^2 = .18$; `selection`, $F(4, 30.66) = 14.99, p < .001$, est. $\omega^2 = .19$). The *Genuine Schemas* model again produced the best fit for every viewpoint model, with polynomial contrasts revealing positive (U-shaped) quadratic trends for `cpitch`, $t(25.39) = 4.38, p < .001, r = .66, R^2 = .43$, and `melint`, $t(14.04) = 4.11, p < .01, r = .74, R^2 = .55$, but positive (increasing) linear trends for `csd`, $t(14.83) = 4.09, p < .01, r = .73, R^2 = .53$, and `selection`, $t(13.41) = 3.83, p < .01, r = .72, R^2 = .52$. The U shape exhibited in the `cpitch` and `melint` models for the cello probably reflects the statistical preference for smaller melodic intervals in the Haydn Corpus, resulting in lower mean IC estimates for categories that feature stepwise motion in the bass (HC and DC), and higher estimates for categories featuring large leaps (PAC, IAC, and EV). This U shape was not demonstrated in the `csd` and `selection` viewpoint models, however, as the DC category received higher IC estimates relative to the other categories in these viewpoint models, thereby resulting in positive linear trends. Presumably, the HC category received the lowest IC estimates on average because scale-degree successions like $\hat{4}$–$\hat{5}$ are more common than successions like $\hat{5}$–$\hat{1}$. And yet successions like $\hat{5}$–$\hat{6}$ are also evidently *less* common than $\hat{5}$–$\hat{1}$, hence the higher IC estimates for the DC category and the increasing linear trend, *PAC→IAC→DC→EV*.

As expected, for the second planned comparison the genuine cadence categories received lower mean IC estimates than the cadential deviations in every viewpoint model (`cpitch`, $t(27.62) = -2.59, p < .05, r = .44$; `melint`, $t(15.93) = -2.93, p < .05, r = .59$; `csd`, $t(24.39) = -5.18, p < .01, r = .72$; `selection`, $t(19.91) = -4.95, p < .01, r = .74$). Given the statistical preference for stepwise motion in the cello, it should also not be surprising that the terminal events for exemplars from the HC category received lower IC estimates on average than those from the authentic cadence categories for every viewpoint model (`cpitch`, $t(16.38) = 2.99, p < .05, r = .59$; `melint`, $t(29.02) = 3.47, p < .01, r = .54$; `csd`, $t(47.17) = 5.36, p < .01, r = .62$; `selection`, $t(39.40) = 5.65, p < .01, r = .67$).

The left line plot in Figure 6.4 displays the mean IC estimates for the terminal chord event—represented by `vintcc` and `csdc`—for *cadence category*. One-way ANOVAs of the IC estimates revealed a main effect for `vintcc`, $F(4, 29.94) = 6.68, p < .001$, est. $\omega^2 = .08$, and for `csdc`, $F(4, 31.90) = 7.02$, $p < .001$, est. $\omega^2 = .09$. The best-fitting polynomial contrast revealed a positive (increasing) linear trend in the *Genuine Schemas* model for `vintcc`, $t(16.96) = 3.56, p < .01, r = .65, R^2 = .43$, and `csdc`, $t(27.04) = 4.14, p < .001, r = .62, R^2 = .39$. The genuine cadence categories also received lower mean IC estimates than the cadential deviations for `vintcc`, $t(28.01) = -3.92, p < .01, r = .60$, and `csdc`, $t(39.39) = -3.84, p < .01, r = .52$. Finally, the terminal chord events from the authentic cadence categories received lower mean IC estimates than those from the HC category, but this trend was not significant for `vintcc`, $t(19.95) = -2.25, p = .11, r = .45$, or `csdc`, $t(35.67) = -1.60, p > .05, r = .26$.

It is also noteworthy that the terminal events from the EV category generally received lower

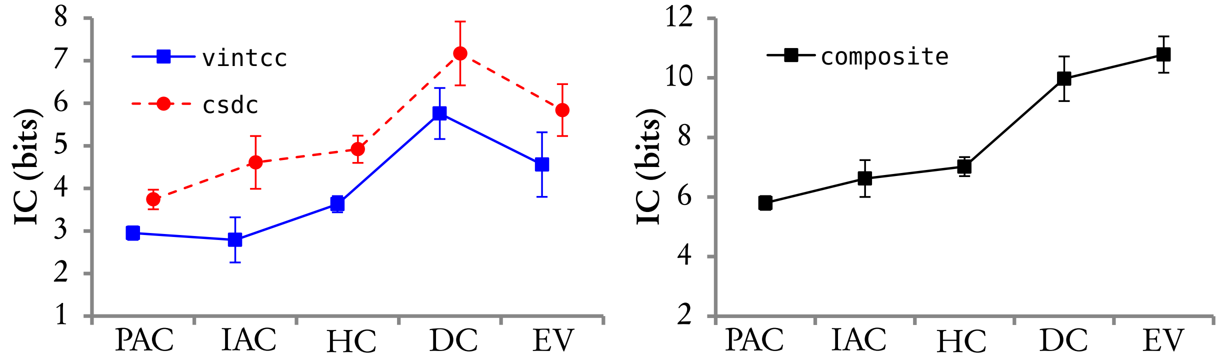


Figure 6.4: Left: Line plot of the mean IC estimated for the resolving chord event for each cadence category. Right: Line plot of the mean IC estimated for the combination called `composite`, which represents the product of $\texttt{selection}_{\text{vl1}}$ and `csdc`. Whiskers represent ±1 standard error.

IC estimates than those from the DC category. Recall from Chapter 2 that evaded cadences are typically characterized not by a deviation in the harmonic progression (though such a deviation may take place), but rather by a sudden interruption in the projected resolution of the melody. To be sure, 10 of the 11 evaded cadences in the collection feature tonic harmony either in root position or in first inversion at the moment of cadential arrival. Given how often this harmony appears in the Haydn Corpus (see Figure 4.6), it is therefore not too surprising that the mean IC estimates decreased from the DC to the EV category.

The right line plot in Figure 6.4 displays the mean IC estimates for the terminal `composite` event for *cadence category*. A one-way ANOVA of the IC estimates revealed a main effect for `composite`, $F(4, 30.91) = 7.21, p < .001$, est. $\omega^2 = .09$. The best-fitting polynomial contrast revealed a positive (increasing) linear trend for the *Genuine Schemas* model, $t(19.23) = 4.81, p < .001, r = .74, R^2 = .55$. Thus, the model *PAC→IAC→HC→DC→EV* accounted for roughly 55% of the variance in the mean IC estimates for `composite`, which represents the largest effect demonstrated across all of the polynomial contrasts from every viewpoint model. The genuine cadence categories also received lower mean IC estimates than the cadential deviations, $t(31.87) = -4.58, p < .01, r = .63$, but the authentic cadence categories did not differ significantly from the HC category, $t(27.98) = -1.32, p > .05$.

In short, the mean IC estimates from IDyOM provide strong evidence in support of the *Genuine Schemas* model of cadential strength. Polynomial contrasts revealed significant positive (increasing) linear trends for the viewpoints `vintcc`, `csdc`, `composite`, `csd`$_{\text{vc}}$, and `selection`$_{\text{vc}}$, as well as significant positive (U-shaped) quadratic trends for `cpitch`$_{\text{vc}}$, `melint`$_{\text{vc}}$, and all of the viewpoints for the first violin. Furthermore, my earlier claim that the genuine cadence categories elicit the strongest and most specific schematic expectations appears to be well supported by the experimental results from the second planned comparison, which revealed that the terminal events from the genuine cadence categories produced the lowest IC estimates on average for the viewpoint models from the first violin and across the entire texture, whereas the cadential deviations generally received the highest IC estimates on average. Finally, the third planned comparison examining the potential differences between the authentic cadence categories and the HC category received less support. Although the predicted trend emerged for `composite`, `csdc`, and `vintcc`, with the terminal events from the AC categories receiving lower IC estimates on average compared to those from the HC category, only `vintcc` demonstrated a significant difference.

Taken together, the reported findings support the role for expectancy in models of cadential strength, with the most complete or closed cadences also serving as the most expected or probabilistic. From this point of view, the perceived strength of the phrase-structural boundary corresponds with the strength of the schematic expectations it generates in prospect. But recall from Chapter 2 that the perception of closure also depends on the cessation of expectations *following* the terminal events of the cadence. That is, the strength of the potential boundary between two sequential events results in part from the *increase* in information content (or decrease in probability) from the first to the second event (i.e., the last event of one group to the first event of the following group). The preceding analyses examined terminal events from cadential and non-cadential contexts in isolation, so the next section considers the role played by schematic expectations in boundary perception and event segmentation by examining the time course of IC estimates surrounding the terminal events of the cadence.

6.3.3 Terminal Events as Perceptual Boundaries

In this section I examine two claims about the relationship between expectancy and boundary perception: (1) that the terminal event of a group is the most expected (i.e., predictable) event in the surrounding sequence; and (2) that the next event in the sequence is comparatively unexpected (i.e., unpredictable). The assumption here is that unexpected events engender prediction errors that lead the perceptual system to segment the event stream into discrete chunks.[84] If the terminal events from genuine cadential contexts are highly predictable, then prediction errors for the comparatively unpredictable events that follow should force listeners to segment the preceding cadential process. Thus, in addition to the between-groups factor of *cadence category*, which allows us to consider how the IC estimates differ from one category to the next, this section also considers how these IC estimates change over time. In this case, the between-groups factor of *time* consists of three events: e_t, which represents the terminal event of the group, and e_{t-1} and e_{t+1}, which represent the immediately surrounding events.

With more complex designs like this one, the number of significance tests can become prohibitively large, so I will restrict the investigation to just four dependent variables under

[84] Kurby and Zacks, "Segmentation in the Perception and Memory of Events." I noted in Chapter 1 that the brain generates expectations about future events to increase our chances of survival from one moment to the next (Clark, "Whatever next? Predictive brains, situated agents, and the future of cognitive science"). Psychologists Christopher A. Kurby and Jeffrey M. Zacks have argued that this biological imperative also explains why listeners segment the musical surface into motives, phrases, and themes. In their view, perceptual systems segment activity into memorable chunks whenever errors in prediction increase transiently. The end result of this process, which they call *event segmentation theory*, is a hierarchically encoded mental representation of our temporal experience that saves on processing resources and improves comprehension.

the assumption that their corresponding viewpoints serve as reasonable approximations of the two-voice framework characterizing the classical cadence: `selection`$_{\text{vl1}}$ and `selection`$_{\text{vc}}$ will represent the soprano and bass, respectively, and `vintcc`, and `csdc` will each represent the entire texture. Thus, this section analyzes these viewpoints for the note events from the first violin and cello and the chord events from the entire texture using a 5×3 two-way ANOVA with between-groups factors of *cadence category* (PAC, IAC, HC, DC, and EV), and *time* (e_{t-1}, e_t, e_{t+1}).

By moving from one to two between-groups factors, the number of omnibus statistics and planned comparisons necessarily increases, and since Levene's test also revealed heteroscedastic groups for all four of the 5×3 viewpoint ANOVAs, our risk of committing a Type I error is considerably greater here than in any of the previous analyses. In this case, the two assumptions mentioned above concern the interaction between *cadence category* and *time*: namely, whether the IC estimates for each cadence category increase or decrease significantly from one event to the next. Thus, we can essentially ignore the main effects and concentrate only on the interaction term of the two-way ANOVA. If the interaction is significant, I will instead report *simple* main effects, which represent one-way ANOVAs with *time* as a factor for each level of cadence category. Because the levels of *cadence category* are generally heteroscedastic, I again report the Welch F ratio and estimate effect size using est. ω^2. To decrease the risk of committing a Type I error resulting from the additional tests of the simple main effects, I apply a Bonferroni correction, which in this case multiplies the p value by the number of levels of the between-groups factor. The factor *cadence category* consists of five levels, so an initial p value of .008 would still be significant at $\alpha = .05$ ($.008 \times 5 = .04$). Finally, to examine the potential differences in the IC estimates for the levels of *time* for each cadence category, each simple main effect included two simple planned comparisons that do not assume equal variances: (1) whether the IC estimate for e_t is lower on average than the surrounding events, e_{t-1} and e_{t+1};

and (2) whether the IC estimate for e_{t+1} is higher on average than the estimate for e_t. As before, I apply a Bonferroni correction to these two comparisons to minimize the risk of a Type I error.

Figure 6.5 displays line plots of the mean IC estimates for the note events over time in the first violin (top) and cello (bottom) for each level of *cadence category*. To gain a more global picture of the IC time course, I have provided the mean IC estimates for the seven-event sequence surrounding the terminal event of each cadence category, but the ANOVA models only consider the interior three events. For the first violin, a two-way ANOVA of the mean IC estimates revealed a significant interaction between *cadence category* and *time* for the viewpoint $\texttt{selection}_{\text{vl1}}$, $F(8, 718) = 3.88, p < .001, \text{est. } \omega^2 = .03$. One-way ANOVAs of the mean IC estimates for each level of *cadence category* revealed simple main effects for *PAC*, $F(2, 236.76) = 67.63, p < .01, \text{est. } \omega^2 = .35$, and *HC*, $F(2, 162.48) = 32.34, p < .01, \text{est. } \omega^2 = .20$, but the remaining categories were no longer significant using Bonferroni correction (*IAC*, $F(2, 14.95) = 3.89, p > .05, \text{est. } \omega^2 = .02$; *DC*, $F(2, 35.86) = 2.84, p > .05, \text{est. } \omega^2 = .01$; *EV*, $F(2, 18.67) = 5.22, p > .05, \text{est. } \omega^2 = .03$).

Despite the non-significant simple main effects for the IAC, DC, and EV categories, simple planned comparisons revealed significant trends over time for every cadence category.[85] As expected, the terminal event in the first violin received lower IC estimates on average than the immediately surrounding events for the genuine cadence categories (*PAC*, $t(298.45) = -9.88, p < .001, r = .50$; *IAC*, $t(18.48) = -2.79, p < .05, r = .54$; *HC*, $t(209.41) = -6.92, p < .001, r = .43$), as well as for the DC category, though this trend was marginal, $t(36.92) = -2.10, p = .08, r = .33$. Thus, for cadences featuring melodies that resolve to stable scale-degrees, IDyOM indicates that the terminal event of the group is also the most predictable event in the

[85] That the planned comparisons are significant when the simple main effects are non-significant may seem surprising, but remember that the comparisons simply measure the difference between specific levels of *time* (e.g., e_t vs. e_{t+1}), whereas the analysis of variance compares the differences between the group means against the grand mean. In this case, the comparisons between the levels of *time* might be significant for a given level of *cadence category* (e.g., *IAC*), but the difference between the group means (relative to the grand mean) is not.

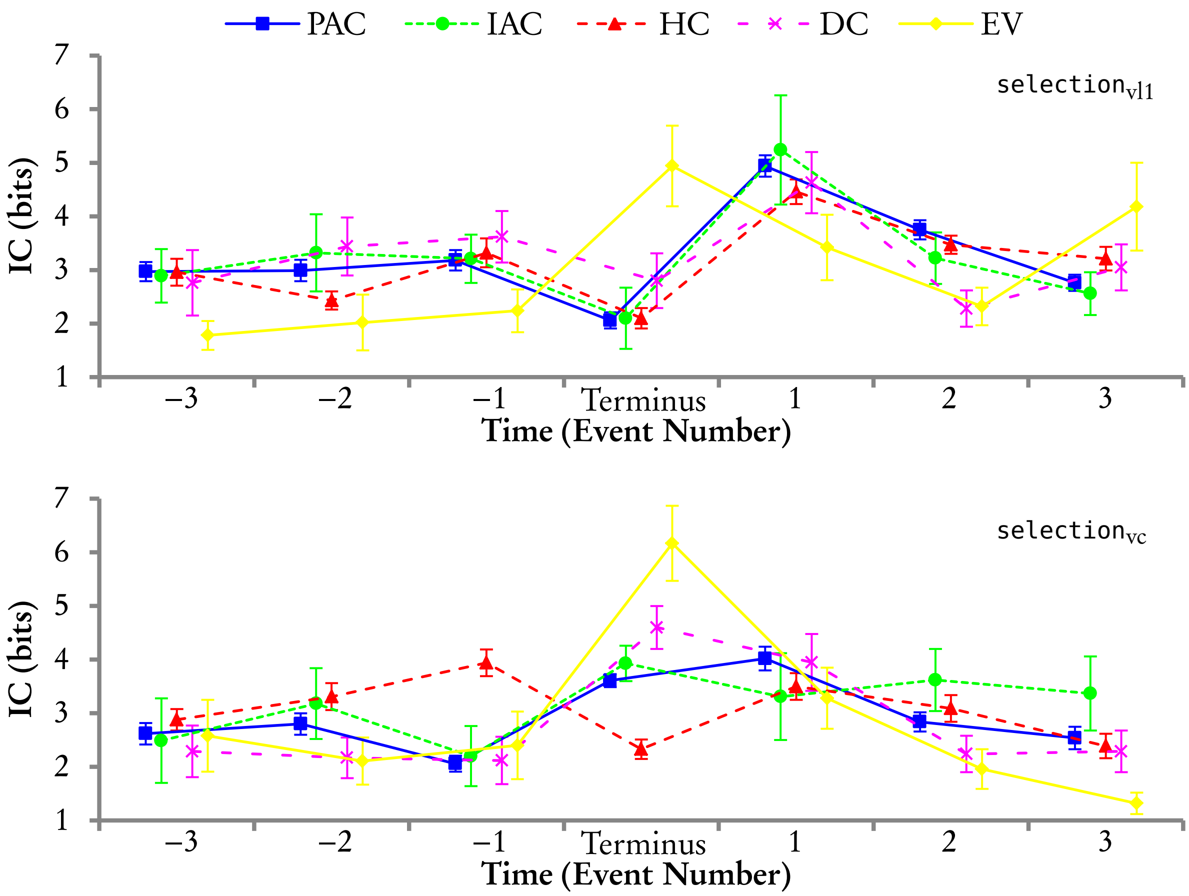


Figure 6.5: Time course of the mean IC estimated for the events surrounding the terminal note event in the first violin (top) and cello (bottom) for each cadence category using the viewpoint `selection`, which represents `melint` and the linked viewpoint `csd` ⊗ `cpitch`. The statistical analysis pertains to times −1, 0 (or Terminus), and 1. Whiskers represent ±1 standard error.

sequence.

For the PAC, IAC, HC, and DC categories, the mean IC estimates increased significantly from e_t to e_{t+1} (*PAC*, $t(223.50) = 11.64, p < .001, r = .61$; *IAC*, $t(12.53) = 2.78, p < .05, r = .54$; *HC*, $t(160.95) = 8.02, p < .001, r = .53$; *DC*, $t(35.62) = 2.41, p < .05, r = .55$), thereby supporting the view that the strength of the perceptual boundary depends on the increase in information content following the terminal event of the cadence. And yet since the EV category replaces the expected terminal event in the melody with material that clearly initiates the subsequent process—often by leaping up to an unexpected scale-degree like $\hat{5}$—we might therefore predict that a significant increase in information content should occur *at* (and not following) the expected terminal event of the group. This is exactly what we observe, with the expected terminal events from the EV category receiving the highest mean IC estimate in the sequence, $t(14.87) = 2.54, p < .05, r = .55$. Thus, the pattern of results from $\texttt{selection}_{vl1}$ is entirely consistent with our previous assumptions: (1) that the terminal event of a group is the most predictable event in the sequence, and (2) that the next event is comparatively unpredictable. Here, the mean IC estimates for the first violin increased significantly following the predicted boundary for every cadence category in the collection.

For the cello, a two-way ANOVA of the mean IC estimates revealed a significant interaction between *cadence category* and *time* for the viewpoint $\texttt{selection}_{vc}$, $F(8, 717) = 13.02, p < .0001$, est. $\omega^2 = .12$. Excepting *IAC*, one-way ANOVAs of the mean IC estimates also revealed simple main effects for every level of *cadence category* (*PAC*, $F(2, 227.67) = 42.14, p < .001$, est. $\omega^2 = .25$; *HC*, $F(2, 161.48) = 15.57, p < .001$, est. $\omega^2 = .11$; *DC*, $F(2, 35.50) = 8.93, p < .01$, est. $\omega^2 = .06$; *EV*, $F(2, 19.87) = 8.29, p < .01$, est. $\omega^2 = .06$). As expected, the terminal event in the cello received lower IC estimates on average than the immediately surrounding events for the HC category, $t(213.66) = -5.44, p < .001, r = .35$, but the trend was reversed for the PAC, DC, and EV categories (*PAC*, $t(322.14) = 3.18, p < .01, r = .17$; *DC*, $t(43.02) = 2.98, p < .01, r = .41$; *EV*, $t(17.49) = 4.06, p < .01, r = .70$. The trend for the IAC category was not significant, $t(21.98) = 1.98, p > .05, r = .39$.

Thus, for the HC category at least, the terminal event was also the most predictable event in the sequence. What is more, the significant increase in information content in the cello at the expected terminal event in the DC and EV categories is consistent with what we know about cadential deviations. For the former category the bass typically resolves deceptively to scale-degrees like $\hat{6}$, thereby violating our expectations for $\hat{1}$, whereas the latter category evades the expected resolution by leaping to other scale-degrees to support harmonies like I^6. The significant increase in information content for the terminal event of the PAC category is somewhat more surprising, however. Recall from § 6.3.2 that the mean IC estimates for the terminal events from each cadence category in the cello demonstrated a U shape, with the HC category receiving the lowest IC estimates (see Figure 6.3). In that case I suggested that small melodic intervals appear more abundantly in the Haydn Corpus than large intervals, resulting in higher IC estimates for categories featuring large leaps (PAC, IAC, and EV). From this point of view, it therefore seems reasonable that the mean IC estimates for the cello would increase at e_t for categories featuring large leaps or unexpected scale-degree continuations, as is the case with the PAC, IAC, DC, and EV categories.

Given this pattern of results for the cello, it should also come as little surprise that *HC* was the only category to demonstrate a significant increase in information content following the terminal event of the cadence, $t(151.11) = 3.76, p < .001, r = .29$. To be sure, the IC estimates for the cello did not significantly increase at e_{t+1} for the PAC and IAC categories, thereby undermining my claim that for the genuine cadence categories at least, the perceived boundary follows the terminal events of the cadence. When we consider the results from the first violin and the cello together, *HC* was also the only category for which the IC estimates from $\texttt{selection}_{\mathrm{vl1}}$ and $\texttt{selection}_{\mathrm{vc}}$ decreased at e_t and then increased at e_{t+1}. If the PAC and IAC categories also generate strong and specific melodic *and* harmonic expectations for the terminal events of the cadence, the viewpoint models representing both voices of the two-voice framework should demonstrate congruent behavior. I suggested in § 6.3.1 that $\texttt{selection}_{\mathrm{vc}}$ is not the most suitable representation of the lower voice in the two-voice framework because our conception of the bass voice depends on the interaction among numerous parameters across the entire texture. To determine the role played by the bass voice in the perception of cadential boundaries, we might therefore examine the harmonies formed between the cello part and the upper parts.

Figure 6.6 displays line plots of the mean IC estimates for `vintcc` (top) and `csdc` (bottom) over time for each level of *cadence category*. Two-way ANOVAs of the mean IC estimates revealed a significant interaction between *cadence category* and *time* for the viewpoints `vintcc`, $F(8, 719) = 3.13, p < .01, \text{est. } \omega^2 = .03$ and `csdc`, $F(8, 704) = 2.99, p < .01, \text{est. } \omega^2 = .03$. Simple main effects and planned comparisons were not significant for `csdc`, however, so it will not be reported here. For `vintcc`, one-way ANOVAs of the mean IC estimates revealed simple main effects for the genuine cadence categories (*PAC*, $F(2, 235.55) = 9.74, p < .001, \text{est. } \omega^2 = .07$; *IAC*, $F(2, 15.21) = 6.48, p < .05, \text{est. } \omega^2 = .04$; *HC*, $F(2, 161.39) = 8.26, p < .01, \text{est. } \omega^2 = .06$), but not for the cadential deviations. As expected, the terminal chord event received lower IC estimates on average compared to the surrounding events for the genuine cadence categories (*PAC*, $t(313.01) = -3.63, p < .001, r = .20$; *IAC*, $t(22.58) = -3.49, p < .01, r = .59$; *HC*, $t(216.80) = -4.07, p < .001, r = .27$). Although the trend was reversed for the cadential deviations, with the mean IC estimates increasing from e_{t-1} to e_t, the difference was not significant for either category (*DC*, $t(38.02) = 1.14, p > .05, r = .18$; *EV*, $t(18.50) = .25, p > .05, r = .06$). Finally, the mean IC estimates increased significantly from e_t to e_{t+1} for *PAC*, $t(220.21) = 4.42, p < .001, r = .29$, and *HC*, $t(157.62) = 3.25, p < .001, r = .25$, but this trend was marginal for *IAC*, $t(13.71) = 2.09, p > .05, r = .49$.

In sum, `vintcc` demonstrated a similar trend to that found in $\texttt{selection}_{\text{vl1}}$ for the genuine cadence categories, with the mean IC estimates decreasing from e_{t-1} to e_t, and then increasing from e_t to e_{t+1}. These two viewpoint models also displayed congruent behavior for the EV category, with both models increasing from e_{t-1} to e_t, suggesting that the perceptual boundary precedes (rather than follows) the expected terminal event in evaded cadences. For the DC category, parametric noncongruence obtained, with the mean IC estimates at e_t decreasing

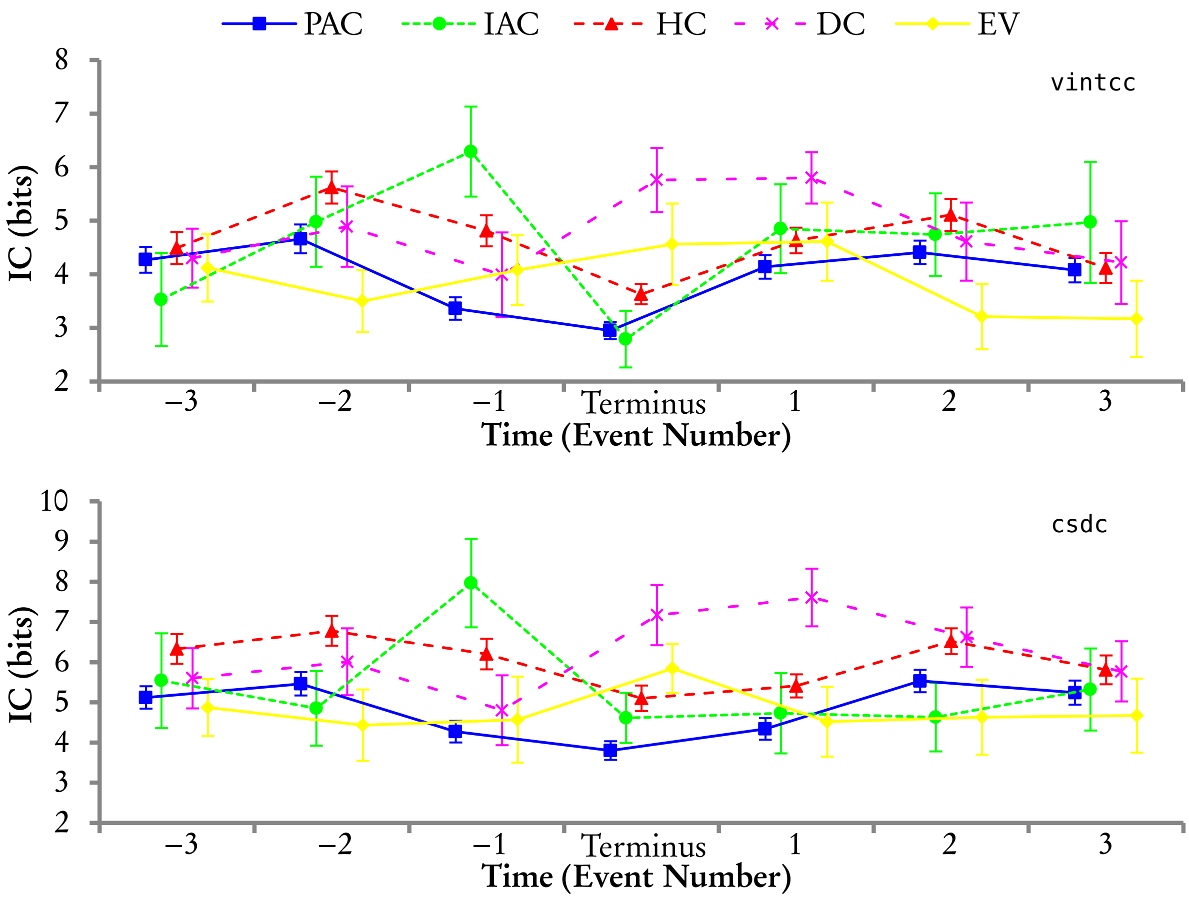


Figure 6.6: Time course of the mean IC estimated for the events surrounding the terminal chord event for `vintcc` (top) and `csdc` (bottom) for each cadence category. The statistical analysis pertains to times −1, `0` (or Terminus), and 1. Whiskers represent ±1 standard error.

in $\texttt{selection}_{\mathrm{vl1}}$ but increasing in `vintcc`. Thus, across the levels of *cadence category* and *time*, the $\texttt{selection}_{\mathrm{vl1}}$ and `vintcc` viewpoint models support my initial assumptions about the role played by schematic expectations in boundary perception: (1) that the terminal event of a group is the most expected (i.e., predictable) event in the sequence; and (2) that the next event is comparatively unexpected (i.e., unpredictable).

It is also noteworthy—if not statistically significant—that the IC estimates for the DC category demonstrated a greater increase from e_{t-1} to e_t compared to the estimates for the EV category in both the `vintcc` and `csdc` models. This result may seem counterintuitive, but remember that 10 of the 11 evaded cadences in the collection feature tonic harmony either in root position or first inversion. Since deceptive cadences typically feature vi at the moment of cadential arrival, it seems reasonable that cadences from the EV category were less harmonically unexpected than the DC category, and thus would receive lower IC estimates on average. Nevertheless, the increase in information content from e_{t-1} to e_t was not statistically significant for either category. It is also somewhat surprising that `csdc` revealed no significant effects when the trends over time for each cadence category were nearly identical to those found for `vintcc`. The alphabet of *csdc* consists of 688 distinct chord types compared to just 190 types for `vintcc`, so it is possible that the greater variety of possible continuations influenced the model estimates in such a way as to obscure the differences between the cadence categories over time. Given the rather meager sample size for three of the five cadence categories in the collection, however, it is difficult to speculate about the IC estimates from `csdc` without first increasing the sample size for the imperfect authentic, deceptive, and evaded cadence categories.

§6.4 Conclusions

This chapter examined three claims about the relationship between expectancy and cadential closure: (1) terminal events from cadential contexts are more predictable than those from non-cadential contexts; (2) models of cadential strength advanced in the *Formenlehre* tradition reflect the formation, violation, and fulfillment of schematic expectations during music listening; and (3) a significant decrease in predictability follows the terminal note and chord events of the cadential process. In §6.1 I reviewed the mathematical formalism behind IDyOM, a context (or *n*-gram) model developed by Marcus Pearce that simulates melodic and harmonic expectations by acquiring knowledge through unsupervised statistical learning of sequential and simultaneous structure, and §6.2 described the methods for combining viewpoint models estimated by IDyOM. Finally, §6.3 provided evidence in support of the link between expectancy and cadential closure.

In §6.3.1 I found that the terminal note and chord events from perfect authentic cadences are more predictable than (1) non-cadential events featuring tonic harmony in root position and $\hat{1}$ in the soprano, and (2) non-cadential events featuring any harmony and any scale-degree in the soprano. For half cadences, significant effects were limited to viewpoints representing the first violin, but the terminal events from half-cadential contexts were still more predictable than those from non-cadential root-position dominants. In §6.3.2 I provided strong evidence in support of the *Genuine Schemas* model of cadential strength (*PAC→IAC→HC→DC→EV*), with the genuine cadence categories (PAC, IAC, HC) and cadential deviations (DC, EV) eliciting the lowest and highest IC estimates on average, respectively. Finally, §6.3.3 indicated that unexpected events—like those directly following the terminal note and chord events from genuine cadences—engender prediction errors that presumably lead the perceptual system to segment the event stream immediately following the cadential process.

Taken together, the reported findings support the role of expectancy in models of cadential closure, with the most complete or closed cadences also serving as the most expected or probable. Nevertheless, future studies will need to address a number of limitations in the current investigation. First, the rather meager sample size for three of the five cadence categories in the collection—as well as the Haydn Corpus more generally—casts some doubt upon the generalizability of the reported findings. That the estimates from IDyOM corresponded so well with theoretical predictions suggests that they may be robust to issues of sample size, but future studies should look to expand the collection considerably, as well as to consider how the relationship between expectancy and cadential closure varies for other genres and style periods. Second, Chapters 4 and 5 demonstrated the utility of non-contiguous *n*-grams for the discovery and classification of cadential patterns, but IDyOM is not presently capable of including those sorts of patterns.[86] Attempts to include non-contiguous (or *distance*) *n*-grams now exist in the language modeling community,[87] however, so revisions to IDyOM's architecture could be feasible in the near future.

But perhaps the greatest limitation of the present research lies in my continued dependence on simulation. Indeed, Part II provided a detailed study of the many cadential patterns characterizing a representative corpus of Haydn's string quartets, but to examine the psychological relevance of the many claims made therein requires an entirely different approach, one in which the listener, rather than the music, represents the object of study. Thus, Part III extends the findings from Part II to a collection of cadences from Mozart's keyboard sonatas, examining the perception and cognition of cadential closure in a series of experimental studies. In Chapter 7 participants provide completion ratings for cadences heard both in and out of context to examine the roles played by syntactic and rhetorical parameters in models of cadential strength, while Chapter 8 examines the formation of expectations during music listening using retrospective ratings, continuous ratings, and an implicit reaction-time task based on the priming paradigm.

[86]However, it is possible to construct *threaded* viewpoints in IDyOM that sample events from a *base* viewpoint like `cpitch` according to some *test* viewpoint that represents positions in the sequence, such as metric downbeats or phrase boundaries (Pearce, "The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition," 60–62).

[87]Guthrie et al., "A Closer Look at Skip-gram Modelling"; Milind Huang, Doug Beeferman, and X. D. Huang, "Improved Topic-Dependent Language Modeling Using Information Retrieval Techniques," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing* (Washington, DC: IEEE Computer Society Press, 1999); M. Simons, H. Ney, and S. C. Martin, "Distant Bigram Language Modeling Using Maximum Entropy," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing* (Washington, DC: IEEE Computer Society Press, 1997), 787–790.

Part III

Experimental Evidence:

Twenty-First-Century Listeners

Chapter 7

Perceiving Closing Schemas: Completion Ratings

The ear is so accustomed to the perfect cadence or the plagal cadence as the termination of all polyphonic compositions, that most hearers do not consider a piece finished which has not the chord of the tonic as its final chord.

René Lenormand

In Part I I suggested that listeners with exposure to music of the classical style possess schematic knowledge of the many recurrent patterns found therein. To support this view, Part II examined a corpus of Haydn's string quartets and presented a few analytical techniques for the discovery, classification, and prediction of cadences that might simulate the learning mechanisms underlying human cognition. And yet despite all this interest in the effects of learning and memory on the perception and cognition of cadences, listeners have yet to represent the object of study. To be sure, recall from Chapter 2 that the cadence concept embraces genres and styles beyond those encapsulated within the high classical period. Surely in today's vast musical landscape, such anachronistic patterns as the Cudworth or the Expanding Do-Fi-Sol delimit too narrow a repertory to characterize the schematic knowledge of modern listeners. Robert Gjerdingen seems to agree:

> "Native listeners" of eighteenth-century court music—whether we mean by that term the deceased members of those courts or modern listeners who have immersed themselves in the galant style to the point of acquiring it as a second language—hear in its compositions discrete chunks that match memories of meaningful gestures and phrases. Other, more casual listeners will perceive a pleasant flow of tones, gross changes in texture and dynamics, and those elements of musical syntax that may transcend the period in question.[1]

And since the musical knowledge of eighteenth-century listeners presumably "stemmed from relatively homogeneous compositions of similar ethnic, geographic, social, and chronological derivation,"[2] we might therefore assume that the evidence accumulated in Part II characterizes the musical knowledge of Haydn and his contemporaries, a form of "culturally situated cognition."[3] How do the findings from Part II relate to the schematic knowledge of modern listeners? And do the sorts of patterns found in Haydn's string quartets generalize to the works of other classical composers and genres? It is thus the goal of Part III to examine more critically the influence of musical expertise on the perception and cognition of the most common cadence types from the classical style—as well as to extend the findings from Part II to the works of another classical composer and genre—using the many methods of inference developed in the experimental sciences.

This chapter examines the perception of cadential closure using an explicit task, in which participants provided *completion* ratings on a 7-point continuous scale for excerpts drawn from

[1] Gjerdingen, "Courtly Behaviors," 380.

[2] Ibid., 380–381.

[3] Vasili Byros, "Foundations of Tonality as Situated Cognition, 1730–1830" (PhD Dissertation, Yale University, 2009).

Mozart's keyboard works. In §7.1, I review the treatment of closure in music psychology, attending specifically to the available experimental evidence for the perception and cognition of tonal cadences. §7.2 presents Experiment I, which examines the perception of cadences presented *out of context*, such that participants remained unaware of the larger context surrounding each excerpt. To examine how the material following cadential arrival might also contribute to the perception of closure in §7.3, Experiment II presented the same excerpts from the previous experiment *in context*, such that participants could also consider the material following the moment of cadential arrival when determining the perceived strength or finality of the cadence. I conclude in §7.4 by discussing the relevant findings from Experiments I and II and their impact on the many issues surrounding hierarchical models of cadential strength.

§7.1 Cadences: Experimental Evidence

Although the term 'cadence' appears frequently in the music psychology literature as a perceptually-relevant concept, little experimental research explicitly investigates the perception of cadential closure. Instead, a vast number of studies employ cadences and other ending formulæ as stimuli under the assumption that the experience of closure during music listening is simply a by-product of more general cognitive processes. Questions as to how listeners store cognitive representations of harmonic, tonal, and rhythmic structure in long-term memory, as well as to how these mental representations affect various aspects of music perception (e.g., the formation of expectations during music listening, the perception of dynamic variations in tension, etc.), continue to resonate with music psychologists, resulting in considerably fewer studies devoted to the perception of closure itself.

Nonetheless, closure often plays a prominent role in studies otherwise concerned with other aspects of music perception and cognition. Krumhansl's seminal studies examining the tonal hierarchy provide one notable example. In her initial experiment with Edward Kessler, Krumhansl presented a short tonal context consisting of a major triad or one of three harmonic progressions (IV–V–I, II–V–I, and VI–V–I), and then asked participants to rate how well each of the twelve members of the chromatic scale fit with the preceding context. The average probe-tone ratings for each of the 12 tones of the scale revealed a hierarchy of tonal stability, with the tonic and dominant scale degrees receiving the highest fit ratings, followed next by the other diatonic members of the scale, and with the non-diatonic members receiving the lowest ratings.[4] To explain the effect of tonal context on the obtained probe-tone ratings, Krumhansl and Kessler proposed that listeners possess a cognitive representation of the tonal hierarchy. What is more, a comparison of the goodness-of-fit ratings with the frequency-of-occurrence of these scale-degrees in various corpora from Western music revealed a significant correlation, with scale-degrees that occur more frequently receiving higher "fit" ratings,[5] leading Krumhansl to suggest that listeners form an internal representation of the tonal hierarchy by internalizing the distribution properties of Western tonal music.[6]

Publications over the past two decades of numerous key-finding algorithms incorporating Krumhansl and Kessler's major and minor key profiles provide convincing evidence for the psychological reality of the tonal hierarchy.[7] Yet scholars like David Butler and Bret Aarden have since raised several objections both to Krumhansl's probe-tone method and to her interpretation

[4]Krumhansl and Kessler, "Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys." Krumhansl and Kessler also examined minor key tonal contexts using a minor triad or one of the same three harmonic progressions.

[5]Krumhansl, *Cognitive Foundations of Musical Pitch*, 68–69.

[6]Ibid., 286.

[7]The first model incorporating the Krumhansl-Kessler key profiles was proposed by Krumhansl herself, in collaboration with Mark Schmuckler. For a discussion of the model, see *Cognitive Foundations of Musical Pitch*, 77–110. Figure 4.1 also replicates those profiles. Huron and Parncutt have since suggested methods for incorporating effects of echoic memory and pitch salience on tonality perception ("An Improved Model of Tonality Perception Incorporating Pitch Salience and Echoic Memory"), and Temperley has provided numerous revisions to the Krumhansl-Schmuckler model that improve the mathematical efficiency of the algorithm and address issues related to modulation ("What's Key for Key?").

of the results.[8] Noting a disparity between the distribution of the goodness-of-fit ratings with the scale-degree distributions from tonal music, Aarden examined whether Krumhansl's major and minor key profiles may actually reflect the distribution properties of tones located specifically at the ends of phrases.[9] To examine the effect of phrase position on mental representations of the tonal hierarchy, Aarden conducted a reaction-time study in which participants responded to the contour of each event in a short tonal melody. His first experiment was designed to test the assumption that scale-degrees receiving a higher "fit" rating in Krumhansl's major and minor key profiles would lead to faster reaction times, but the results did not support this hypothesis. Yet when he asked participants in a second experiment to respond only to the contour of the final event of each melody, he observed a close correspondence between participant reaction times and the "fit" ratings of the final events,[10] suggesting that the probe-tone method employed by Krumhansl and Kessler actually encouraged listeners to treat the probe tone as a phrase-final event. Thus, Aarden claimed, Krumhansl and Kessler's major and minor key profiles reflect a cognitive representation of the tonal hierarchy that pertains specifically to *endings*.[11]

Krumhansl, Bharucha and Kessler noted a similar closure effect for two-chord harmonic progressions selected from the chords of the diatonic scale. Following the presentation of an ascending major scale, participants were given a two-chord progression and asked to indicate how well the second chord followed the first on a 7-point scale. The relatedness judgments the authors obtained for these progressions revealed a hierarchy of stability (I > V > IV > VI > II > III > VII), with more stable chords serving as better continuations in the

[8]For a succinct summary and critique of Krumhansl's methodology and the interpretation of her results, see Bret Aarden, "Dynamic Melodic Expectancy" (PhD Dissertation, The Ohio State University, 2003), 11–26.

[9]Although the correlations between the scale-degree distributions of various corpora with Krumhansl's goodness-of-fit ratings are quite high ($r > .80$), several discrepancies remain unexplained, the most noteworthy example being that $\hat{5}$ normally appears more frequently than $\hat{1}$ in the various corpora, yet in Krumhansl and Kessler's key profiles, the tonic receives the highest "fit" rating.

[10]Aarden, "Dynamic Melodic Expectancy," 75.

[11]Ibid.

two-chord context.[12] Noting that stable harmonies like I and V typically received higher continuation judgments when they followed, rather than preceded, the remaining diatonic harmonies, Bharucha and Krumhansl further proposed that stable tones and chords appear at the ends of phrases because they serve as *cognitive reference points*,[13] an expression coined by Eleanor Rosch for elements that are characterized by their asymmetric temporal relations with less stable elements.[14]

The effect of harmonic and melodic closure also has significant effects on short-term memory during music listening, with participants demonstrating a decrease in performance in short-term memory tasks involving melodic notes that straddle a phrase boundary. In an experimental study examining the effect of phrase position on memory for melodies, Norma Tan and her co-authors suggested that listeners use stable harmonies like I and V to segment melodic phrases even in the absence of an explicit harmonic context, resulting in improved memory performance for melodic notes that appear within phrases, and decreased performance for notes lying on either side of a phrase boundary.[15] Lola Cuddy, Annabel Cohen, and D. J. Mewhort also reported a similar finding using a memory recognition task for entire melodies, noting that short melodies were better remembered if they contained a strong tonal ending.[16]

The evidence is thus overwhelming that an internal representation of the tonal hierarchy affects the perception of melodic closure, though it remains far less clear whether such a

[12] Carol L. Krumhansl, Jamshed J. Bharucha, and Edward J. Kessler, "Perceived Harmonic Structure of Chords in Three Related Musical Keys," *Journal of Experimental Psychology: Human Perception and Performance* 8, no. 1 (1982): 32.

[13] Bharucha and Krumhansl, "The Representation of Harmonic Structure in Music."

[14] Rosch, "Cognitive Reference Points." In the color domain, for example, Rosch demonstrated through a series of experiments that natural categories like blue or red have reference point stimuli in relation to which variants of those categories are judged. This relationship between variant and referent is therefore inherently asymmetrical, a characteristic Bharucha and Krumhansl also identified in the relatedness judgments for two-chord progressions.

[15] Norma Tan, Rita Aiello, and Thomas G. Bever, "Harmonic Structure as a Determinant of Melodic Organization," *Memory and Cognition* 9, no. 5 (1981): 533–539.

[16] Lola L. Cuddy, Annabel J. Cohen, and D. J. K. Mewhort, "Perception of Structure in Short Melodic Sequences," *Journal of Experimental Psychology: Human Perception and Performance* 7, no. 4 (1981): 869–883.

representation pertains only to the final event of a phrase, or instead, whether it may pertain to a *series* of events, thereby suggesting that listeners possess schematic representations of various melodic closing formulæ. Marilyn Boltz and David Butler have reported effects of serial order on both the perception of melodic closure and the identification of tonal center respectively,[17] but the issue as to whether listeners store melodic closing patterns in long-term memory remains open. For bass line motion, however, the serial order of the events preceding a phrase ending is fundamental to current definitions of cadential closure in the *Formenlehre* tradition; the cadential status of the final tonic in an authentic cadence, for example, is crucially determined by the harmony of the preceding event. Thus, music scholars frequently treat harmonic closure as a temporal process, an idea that has gained acceptance in the experimental literature.

To consider the effect of a number of musical parameters on the perception of harmonic closure, Burton Rosner and Eugene Narmour asked participants to judge which of a pair of two-chord progressions seemed more closed; they then quantified variables relating to the position of the soprano and bass voices with respect to the root of each chord, the number of common tones shared between the two chords, and the motion of the soprano voice.[18] In addition to these parameters, they also considered style-specific variables that corresponded to music-theoretic notions of cadential closure, such as the root progression of each stimulus and the position of the final melodic event within the tonal hierarchy. To their surprise, parametric variables such as the soprano position, bass inversion, and the number of shared common

[17]In Boltz's study, melodies ending with the tonic-to-leading tone progression received the lowest ratings of melodic completeness, whereas the leading tone-to-tonic progression received the highest ratings, leading Boltz to claim that the precise temporal order of these scale degrees results in the highest perceived resolution ("Perceiving the End: Effects of Tonal Relationships on Melodic Completion," *Journal of Experimental Psychology: Human Perception and Performance* 15, no. 4 [1989]: 754). Butler showed how the temporal order of certain rare intervals (such as the tritone) presented as harmonic dyads can alter the identification of the tonal center. For a summary of the previous studies he conducted with Helen Brown, see David Butler, "Describing the Perception of Tonality in Music: A Critique of the Tonal Hierarchy Theory and a Proposal for a Theory of Intervallic Rivalry," *Music Perception* 6, no. 3 (1989): 234–236.

[18]Rosner and Narmour, "Harmonic Closure."

tones did not affect the closure preference ratings. Instead, schematic representations of root progressions common to known cadences appeared to play the most prominent role, leading the authors to claim that the various harmonic formulæ located at phrase endings result in the formation of schematic representations of harmonic closure. They explain,

> ...when evaluating closure, listeners presumably invoke learned harmonic structures as stylistic schemata. Such schemata come into play when the stimulus displays a sufficient number of featured properties to activate them. This process relies on previously learned stylistic patterns and should be central to closural evaluation.[19]

That listeners may possess both melodic and harmonic closing schemata might also explain why cadences play such a prominent role in the perception of tension, a topic that has received a great deal of attention over the past two decades. In a study initially investigating Lerdahl and Jackendoff's model of tonal tension, Emmanuel Bigand and Richard Parncutt asked listeners to rate their perception of musical tension for each pair of successive chords in Chopin's Prelude in E major.[20] Although they expected Lerdahl and Jackendoff's model to perform best, they were surprised to find that the simple encoding of authentic and half cadences best explained listener ratings of tension, leading them to conclude that cadences provide important reference points for the perception of tension during music listening.[21]

From an examination of the experimental literature, it appears that cadences play a vital

[19] Rosner and Narmour, "Harmonic Closure," 397–398.

[20] Emmanuel Bigand and Richard Parncutt, "Perceiving Musical Tension in Long Chord Sequences," *Psychological Research* 62 (1999): 237–254.

[21] Ibid., 254. In a related finding that is pertinent to this research, the authors also suggest that listeners perceive tension from within a short temporal window. As a result, they claim tension ratings for a given harmonic event remain more or less independent of non-adjacent events. The effect of such hierarchical interpretations of musical structure on the perception of both closure and tension is, however, very much in dispute. Lerdahl has since proposed that the results obtained by Bigand and Parncutt reflect a conflation of stability/instability—terms Lerdahl associates with tonal tension—with closure/non-closure. He rightly points out that a highly stable event, such as the tonic initiating a phrase, may nonetheless imply continuation, and thus, non-closure (Fred Lerdahl and Carol L. Krumhansl, "Modeling Tonal Tension," *Music Perception* 24, no. 4 [2007]: 357).

role in the perception of tonal music. And it seems intuitive that listeners possess cognitive representations for various ending patterns that appear frequently in tonal music. But what remains absolutely essential to such a claim is that the strength of the schematic representation depends on a listener's exposure to the musical style. A growing body of evidence reveals that listeners develop implicit knowledge of tonal and harmonic structure simply as a result of passive exposure to Western music. The psychological reality of Krumhansl's tonal and harmonic hierarchies therefore reflects the general purpose implicit learning mechanism I described in detail in Chapters 1 and 2, in which knowledge about the external environment is acquired without conscious awareness.[22] Researchers have since extended this claim to explain how listeners process harmony and melody, proposing a connectionist framework to account for Western harmonic syntax,[23] and employing artificial grammars to examine how listeners respond to novel harmonic and melodic contexts.[24] Moreover, children appear to develop a sensitivity to harmonic structure at around 5 to 7 years of age, even in the absence of explicit formal training.[25] Researchers have also reported harmonic priming effects in children between 5-11 years of age using both behavioral and neural measures,[26] and Eugenia Costa-Giomi has even extended this claim explicitly to the perception of cadences, suggesting that by 6 years of

[22]Arthur S. Reber, "Implicit Learning and Tacit Knowledge," *Journal of Experimental Psychology: General* 118, no. 3 (1989): 219–235.

[23]Bharucha, "Music Cognition and Perceptual Facilitation"; Jamshed J. Bharucha and Keiko Stoeckig, "Reaction Time and Musical Expectancy: Priming of Chords," *Journal of Experimental Psychology: Human Perception and Performance* 12, no. 4 (1986): 403–410; Barbara Tillmann, Jamshed J. Bharucha, and Emmanuel Bigand, "Implicit Learning of Tonality: A Self-organizing Approach," *Psychological Review* 107, no. 4 (2000): 885–913.

[24]Jonaitis and Saffran, "Learning Harmony"; Rohrmeier, Rebuschat, and Cross, "Incidental and Online Learning of Melodic Structure"; Barbara Tillmann and Bénédicte Poulin-Charronnat, "Auditory Expectations for Newly Acquired Structures," *The Quarterly Journal of Experimental Psychology* 63, no. 8 (2010): 1646–1664.

[25]Kathleen A. Corrigall and Laurel J. Trainor, "Musical Enculturation in Preschool Children: Acquisition of Key and Harmonic Knowledge," *Music Perception* 28, no. 2 (2010): 195–200; Laurel J. Trainor and Sandra E. Trehub, "Key Membership and Implied Harmony in Western Tonal Music: Developmental Perspectives," *Perception and Psychophysics* 56, no. 2 (1994): 125–132.

[26]Stefan Koelsch et al., "Children Processing Music: Electric Brain Responses Reveal Musical Competence and Gender Differences," *Journal of Cognitive Neuroscience* 15, no. 5 (2003): 683–693; E. Glenn Schellenberg et al., "Children's Implicit Knowledge of Western Music," *Developmental Science* 8, no. 6 (2005): 551–566.

age, children notice the lack of a conclusive cadence if it is missing from a progression, and by 8 they can discriminate between conclusive and inconclusive cadences.[27]

Unfortunately, the effect either of explicit musical training or passive exposure on the perception of closure remains unclear, with many studies reporting contradictory findings. Boltz asked both musicians and nonmusicians to provide melodic completion ratings on a 10-point scale for several ending patterns, but she failed to observe a difference between the two groups, leading her to claim that implicit exposure, rather than explicit training, accounts for the perception of melodic completion.[28] Barbara Tillmann and her co-authors also reported a similar finding using cadential patterns, in which musicians and nonmusicians provided completion ratings for both half cadences and authentic cadences (in the context of 16-m. minuets).[29] Tillmann therefore proposed that participants apply the same perceptual principles when assessing musical closure, regardless of expertise, though nonmusicians may be less efficient than musicians.

Other scholars, however, have noted significant effects of musical expertise on the perception of closure. Michel Vallières et al. asked participants to categorize a series of short excerpts from Mozart's keyboard sonatas as beginnings, middles, or ends; and for the ending excerpts, Vallières selected only perfect authentic cadences. The results revealed a significant difference between musicians and nonmusicians, as the musician group correctly identified these cadences as "ends" with nearly perfect accuracy, while nonmusicians were considerably less accurate, correctly identifying ends roughly 80% of the time.[30] Margaret Weiser also reported an effect of

---

[27]Eugenia Costa-Giomi, "Young Children's Harmonic Perception," *Annals of the New York Academy of Sciences* 999, no. 1 (2003): 477–484.

[28]Boltz, "Perceiving the End," 753. Like Aarden, Boltz also considered how cognitive representations of tonal structure affect music perception, but she employed an explicit rating task rather than a reaction-time task.

[29]Barbara Tillmann, Emmanuel Bigand, and Francois Madurell, "Local versus Global Processing of Harmonic Cadences in the Solution of Musical Puzzles," *Psychological Research* 61 (1998): 168.

[30]Michel Vallières et al., "Perception of Intrinsic Formal Functionality: An Empirical Investigation of Mozart's Materials," *Journal of Interdisciplinary Music Studies* 3, nos. 1-2 (2009): 23. The correct identification of 80%, of course, is still significantly better than chance, but as Vallières's analysis later revealed, this effect of expertise could

expertise for two-chord cadences (authentic, half, plagal, and deceptive), in which participants were asked to rate the stability of the final chord on a 5-point scale. The results led her to suggest that musical training facilitates flexible voice-tracking, while the absence of such training results in an attentional bias toward the soprano voice.[31] Finally, the findings obtained over a series of experiments investigating the perception of harmonic and melodic cadential patterns led Roland Eberlein and Jobst Fricke to theorize that experienced listeners of tonal music form schematic representations for frequently occurring cadential formulæ. They conclude that differences of expertise during the perception of closure result from differences in familiarity with the tonal idiom.[32]

Such contradictory reports as to the role of explicit formal training or implicit exposure on the perception of closure may reflect differences either in the choice of experimental task or in the use of stimuli, as researchers often prefer to use homorhythmic, four-part chorale representations of cadential progressions rather than attempt to find examples of cadences from tonal repertories. And there are certainly very good reasons for doing so; by eliminating variations in dynamics, tempo, and rhythm, as well as disregarding a number of features that appear frequently in compositional practice (e.g., a trill at the cadential dominant, the cadential six-four, the suspension dissonance at cadential arrival), such abstract paradigms provide greater experimental control and are much easier to alter to satisfy specific experimental needs. But

not be attributed simply to greater variability between subjects in the nonmusician group, but rather to explicit differences in the way the two groups perceived cadential patterns.

[31]Margaret Weiser, "Rating Cadence Stability: The Effects of Chord Structure, Tonal Context, and Musical Training" (PhD Dissertation, McMaster University, 1992), 40–46. There has been some empirical support for the claim that nonmusicians appear to privilege parameters related to melodic motion, such as pitch proximity and contour, while musicians attend principally to harmonic factors, such as the size of the interval between two events (Piet Vos and Dennis Pasveer, "Goodness Ratings of Melodic Openings and Closures," *Perception and Psychophysics* 64, no. 4 [2002]: 631–639), a claim that will be pertinent to the results presented here.

[32]Roland Eberlein and Jobst Fricke, *Kadenzwahrnehmung und Kadenzgeschichte: ein Beitrag zu einer Grammatik der Musik* (Frankfurt/M.: P. Lang, 1992), 258. Eberlein has also succinctly summarized his theory and proposed a rough model for the effect of familiarity on the perception of closure. "A Method of Analysing Harmony, Based on Interval Patterns or "Gestalten?," in *Music, Gestalt, and Computing*, ed. Marc Leman, vol. 1317, Lecture Notes in Computer Science (Springer Berlin / Heidelberg, 1997), 232–233.

these paradigms also misrepresent the ways in which composers often articulate phrase endings in tonal music (and consequently the ways in which listeners might actually perceive these endings), as they disregard many of the features of cadences that might contribute to the perception of closure. Perhaps worse, such an approach often leads researchers to generalize the behavioral responses elicited by these simple melodic and harmonic formulæ to all tonal music, though the characteristics of closure present in Prokofiev's piano sonatas might differ markedly from those found in Mozart's symphonies.[33] To be sure, the goal of many of the experimental studies employing cadential stimuli is to determine how listeners represent tonal structure in long-term memory. As a result, the examples that they employ serve to probe the various cognitive representations of tonal patterns listeners have abstracted from previous experience. Whether or not a given musical example could actually appear in the repertoire might therefore seem largely irrelevant. But the precision with which we may examine these various representations ultimately depends on a careful understanding of the music to which listeners are consistently exposed. In comparing a listener competent in Mozart's keyboard style with a diverse group of participants, for example, we might find very similar ratings for rhythmically isochronous, harmonic formulæ (and indeed, as a few of the previous studies I have just mentioned can attest, we sometimes do), yet when presented with an excerpt written in that keyboard style, our listener may possess distinctions of a much finer grain than those possessed by the wider group.

Indeed, whereas theorists attend principally to the syntactic parameters of tonal music, in compositional practice each cadence may be realized in nearly countless ways, entailing parameters of rhythm, meter, texture, and instrumentation. Thus, cadences also differ as a

[33]Indeed, Courtenay Harter has outlined some of the characteristic differences of cadential articulation found between Prokofiev and composers of the common practice. "Bridging Common Practice and the Twentieth Century: Cadences in Prokofiev's Piano Sonatas," *Journal of Music Theory Pedagogy* 23 (2009): 57–77.

result of their unique nonsyntactic or rhetorical content,[34] an issue that has not been considered in an experimental setting. In Mozart's keyboard style, for example, cadences may also be characterized according to the formal context from which they were drawn (e.g., main theme, transition, subordinate theme, etc.) or by the presence of a melodic dissonance at the cadential arrival.

The recent revival of interest in the *Formenlehre* tradition has also largely gone unnoticed in the music psychology community, as those studies explicitly examining the perception of closure rarely employ the wider variety of cadential categories found in the "common practice" period. Techniques for cadential deviation, in particular, serve an important formal and expressive function in the classical style, but they have yet to be considered in an experimental setting. Indeed, the experimental study of cadential failure could serve to explore rich areas of inquiry in music psychology—the perception of closure, the processing of harmonic syntax, and the generation and violation of expectations—using musical examples selected from the extant literature.

The experimental studies I will summarize in this chapter attempted to address these issues directly. While an exploration of the underlying sensory and cognitive mechanisms responsible for the perception of closure in tonal music is the ultimate aim of this research, my initial approach was more limited in scope, concentrating as it does on a closing schema that appears frequently in tonal music: the classical cadence. Limiting the initial investigation to cadential closure also afforded the opportunity to consider issues germane to music theory. In the analysis of musical form, the capacity to discern amongst various cadential categories is paramount to the identification of the formal function of a specific musical passage, and this study provides evidence as to whether expert and non-expert listeners can make such distinctions in real time, without the aid of the score. Furthermore, as discussed in Chapter 4, analysts frequently

[34] Caplin, "The Classical Cadence," 106–107.

appeal to a hierarchy of cadential closure (e.g., Janet Schmalfeldt, William Caplin, and, more recently, Edward Latham),[35] and Rosner and Narmour have suggested that style-structural closing schemata may be stored hierarchically in long-term memory,[36] but it remains unclear how various cadential categories—perfect authentic, imperfect authentic, half, etc.—may be positioned within the hierarchy, or how the various musical parameters—melody, harmony, rhythm, etc.—contribute to the perception of closure.

The participants were presented with 50 short excerpts from Mozart's keyboard sonatas that contained an equal number of perfect authentic, imperfect authentic, half, deceptive, and evaded cadences. These categories were chosen both on the basis of their frequency in the classical style and on their assumed relevance to scholarship in music theory and music perception. Each excerpt contained at least the entire cadential progression as defined by Caplin,[37] with some excerpts including music preceding the onset of that progression.[38] Thus, each cadential category differs at the moment of the cadential arrival, which represents the crucial variable distinguishing each excerpt.[39]

To consider syntactic and rhetorical features that occur frequently in Mozart's cadences but are not embraced by cadence category membership, I further subdivided each category into two subtypes to consider issues of formal context (in the case of the PAC and HC), the presence of a melodic dissonance at cadential arrival (for the IAC and HC), as well as the melodic scale-degree and harmony at cadential arrival (for the DC and EV, respectively).[40] Table 7.1 displays the

---

[35]Schmalfeldt, "Cadential Processes"; Caplin, *Classical Form*, 101–111; Latham, "Drei Nebensonnen," 308–309.

[36]Rosner and Narmour, "Harmonic Closure," 406.

[37]Caplin, *Classical Form*, 24–29.

[38]I included additional material preceding the onset of the cadential progression in instances in which I felt the duration of the excerpt was too short to provide a sufficient tonal context.

[39]Of course, a number of other parameters within the cadential progression itself might necessarily imply a given cadential category. For example, metrical placement and duration serve to distinguish a dominant harmony in a half cadence from a dominant in an authentic cadence. But for the purposes of the experimental design it was useful to differentiate each cadential category according to a specific temporal event, in this instance the moment of cadential arrival.

[40]I also considered a number of other parameters for inclusion as subtypes in the study, such as the presence of

Table 7.1: Cadence categories, syntactic characteristics, subtypes, and reference information (Köchel index, movement, measures) for each excerpt.

Cadence Categories	*Subtypes*	*Excerpts*
Perfect Authentic	Main Theme	K. 281, i, mm. 5–8
		K. 281, iii, mm. 5–8
		K. 283, i, mm. 5–10
		K. 311, i, mm. 19–24
		K. 333, ii, mm. 5–8
	Subordinate Theme	K. 284, i, mm. 44–50
	(Expanded Cadential Progression)	K. 309, i, mm. 48–54
		K. 333, i, mm. 54–59
		K. 333, iii, mm. 31–36
		K. 545, i, mm. 20–26
Imperfect Authentic	Melodic Dissonance	K. 311, ii, mm. 27–32
		K. 330, i, mm. 4–8
		K. 330, iii, mm. 39–43
		K. 498a, iv, mm. 32–36
		K. 533, iii, mm. 23–26
	No Melodic Dissonance	K. 281, ii, mm. 4–8
		K. 282, i, mm. 2–4
		K. 284, ii, mm. 21–25
		K. 309, ii, mm. 1–4
		K. 333, iii, mm. 28–32
Half	Main Theme	*K. 284, iii, mm. 1–4
		*K. 311, ii, mm. 1–4
		*K. 331, i, mm. 1–4
		K. 332, ii, mm. 3–4
	Transition	K. 279, iii, mm. 11–18
		*K. 280, i, mm. 21–26
		K. 281, i, mm. 12–16
		*K. 281, ii, mm. 22–26
		K. 310, i, mm. 11–16
		K. 332, i, mm. 31–37
Deceptive	Failed PAC	K. 280, ii, mm. 16–19
		K. 281, ii, mm. 32–35
		K. 282, i, mm. 11–13
		K. 282, iii, mm. 25–31
		K. 309, iii, mm. 58–65
		K. 457, i, mm. 42–48
		K. 533, i, mm. 16–22
	Failed IAC	K. 279, i, mm. 7–10
		K. 330, i, mm. 27–31
		K. 457, ii, mm. 9–11
Evaded	Tonic Harmony	K. 281, i, mm. 30–34
		K. 281, iii, mm. 30–35
		K. 309, i, mm. 13–18
		K. 309, i, mm. 43–46
		K. 309, iii, mm. 11–16
	Non-Tonic Harmony	K. 279, ii, mm. 1–4
		K. 280, i, mm. 3–10
		K. 281, ii, mm. 96–99
		K. 332, ii, mm. 14–16
		K. 333, iii, mm. 84–89

Note. Excerpts from the HC category marked with an asterisk contain a surface dissonance at cadential arrival, whereas the remaining excerpts from that category do not.

cadential categories, the essential characteristics, the subtypes, and the reference information for each excerpt, and Appendix B provides musical examples of the complete stimulus set.

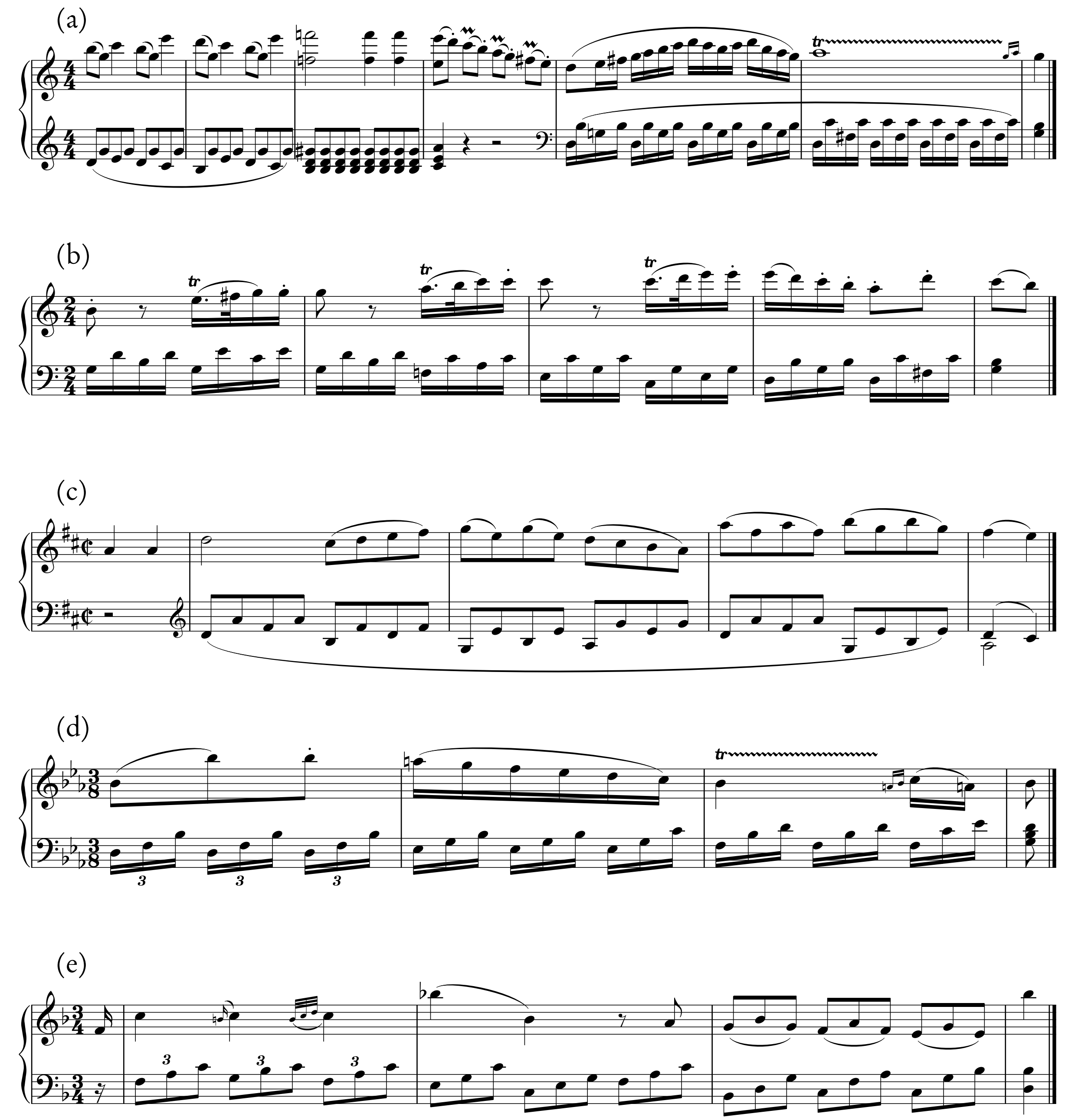
The category of perfect authentic cadences was subdivided according to formal location, selected either from the main theme or the subordinate theme. The excerpts chosen from subordinate themes feature an expanded cadential progression (ECP) (see Example 7.1a), which, in addition to its longer duration (compared to those cadences selected from main themes), is characterized by a dramatic increase in surface activity, usually resulting from an Alberti bass in the left hand and the appearance of a cadential trill above the penultimate dominant.[41] Indeed, that surface activity may affect the perception of closure has been suggested by Michel Vallières, as he found that higher average event density as well as the sudden decrease in event density at cadential arrival significantly affected the categorization of endings by nonmusicians.[42] Figure 7.1 displays the average event density, calculated as the total number of notes per second, for each of the last five seconds of the two subtypes of the PAC category, the EV category, and finally the other categories aggregated together. Both the PAC subordinate-theme subtype and the EV category feature a significant increase in surface activity in the last moments before cadential arrival, at which point the activity ceases almost entirely. For the other categories, however, surface activity does not vary much within the cadential progression. It is thus quite possible that PAC excerpts selected from subordinate themes will yield significantly higher completion ratings than excerpts from the other categories.

The IAC category was subdivided according to the presence or absence of a melodic

---

a surface dissonance at cadential arrival in the perfect authentic cadence category, but the time constraints imposed by the experimental session precluded a design examining more than two subtypes for each category. Moreover, my intent was to select subtypes that reflect the most prevalent features of Mozart's compositional style. In doing so, however, it should be acknowledged that whereas the features reflected in each subtype may play a prominent role in phrase endings from a number of different style periods, they may also be idiomatic to Mozart.

[41] For a discussion of the ECP, see Caplin, “"Expanded Cadential Progression".”

[42] Michel Vallières, “Beginnings, Middles, and Ends: Perception of Intrinsic Formal Functionality in the Piano Sonatas of W. A. Mozart” (PhD Dissertation, McGill University, 2011), 106.

Example 7.1: Five excerpts representing the five cadential categories. (a) PAC category, Subordinate Theme subtype: K. 309/i, mm. 48–54. (b) IAC category, Melodic Dissonance subtype: K. 330/iii, mm. 39–43. (c) HC category, Main Theme subtype: K. 284/iii, mm. 1–4. (d) DC category, Failed PAC subtype: K. 281/ii, mm. 32–35. (d) EV category, Non-Tonic subtype: K. 279/ii, mm. 1–4

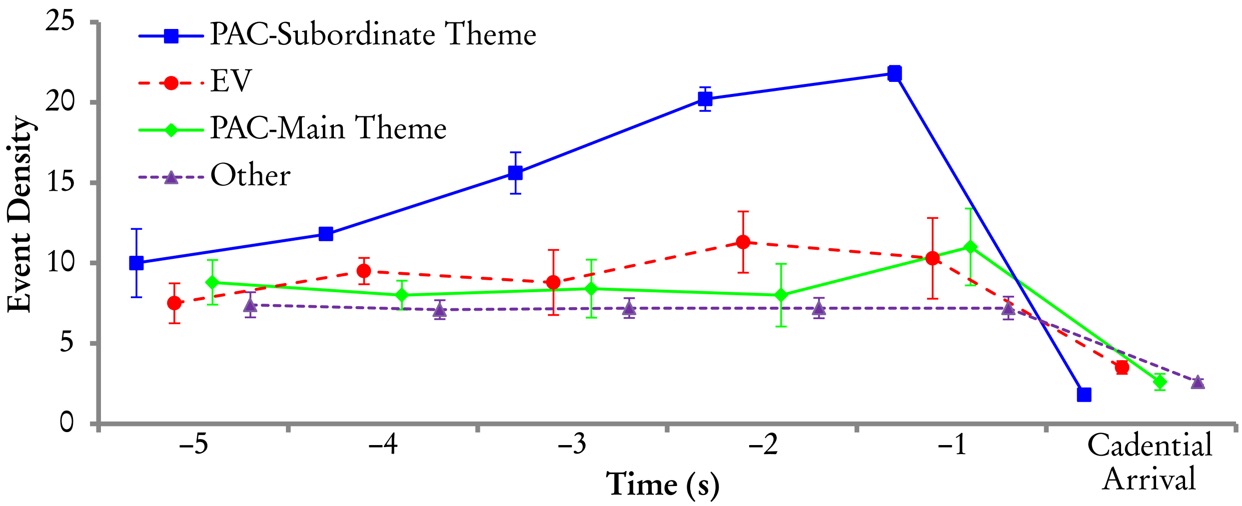


Figure 7.1: Time series plot of the mean event density calculated in a window of 1 second for the two subtypes of the PAC category, the EV category, and the other categories aggregated together. Whiskers represent ±1 standard error of the mean.

dissonance at cadential arrival (Example 7.1b displays the former case, an accented passing tone embellishing the melodic goal). Although a number of other features might serve to differentiate imperfect authentic cadences, such as the metric placement of cadential arrival or the contour of the melody preceding cadential arrival (ascending vs. descending), the presence of a surface dissonance in the melody at the moment of cadential arrival (defined by the appearance of the final tonic harmony) is a prominent attribute of Mozart's imperfect authentic cadences.[43]

As with the PAC category, half cadences were subdivided according to their formal location, selected either from the main theme (as in Example 7.1c, which forms the antecedent phrase of an 8-m. period) or from the end of the transition. As the material within the transition in sonata form typically modulates to the subordinate key, the resulting tonal instability is often accompanied by passages preceding cadential arrival that are frequently characterized by

[43]Imperfect authentic cadences featuring a surface dissonance typically include $\hat{2}$ at cadential arrival (see K. 330/i, m. 8; K. 311/ii, m. 32). To prolong the dissonance at cadential arrival, $\hat{2}$ also sometimes appears in the soprano as an upward resolving suspension to $\hat{3}$, with a chromatic passing tone inserted in between (see K. 498a/iv, m. 36; K. 533/iii, m. 26).

increased energy (relative to the main theme).[44] These features serve to dramatically differentiate transition half cadences from those appearing in the main theme. In addition, the ten excerpts from the HC category were also separately classified according to the presence or absence of a melodic dissonance at the moment of cadential arrival.[45]

Depending upon the degree of melodic closure, the deceptive cadence category has been further subdivided according to whether the melody arrives on $\hat{1}$, which I will refer to as a failed perfect authentic cadence (as in Example 7.1d) or on $\hat{3}$, which I will refer to as a failed imperfect authentic cadence.[46]

Finally, as I mentioned in Chapter 2, the evaded cadence is characterized by a sudden interruption in the projected resolution of the melodic line; instead of resolving to $\hat{1}$, the melody leaps up, often to $\hat{5}$, thereby replacing the expected ending with material that clearly initiates the subsequent phrase. Thus, the evaded cadence projects no sense of ending whatsoever, as the event located at the point of expected cadential arrival, which should group backward by ending the preceding thematic process, instead groups forward by initiating the subsequent process. In order to consider issues of harmonic context associated with the evaded cadence, the category has been subdivided according to which harmony appears at the moment of expected cadential arrival—tonic harmony (which is typically inverted, but may sometimes be in root position) or non-tonic harmony (as in Example 7.1e).

Although the cadential subtypes permit an extended examination of the role played by parameters within each cadence category, these subtypes still fail to consider the relative contribution of a variety of additional parameters that appear frequently in cadential contexts

[44] Caplin, *Classical Form*, 125; Hepokoski and Darcy, *Elements of Sonata Theory*, 93.

[45] Unfortunately, the formal location subtypes for the HC category do not contain an equal number of excerpts: the main theme subtype contains four, whereas the transition subtype contains six. However, the surface dissonance subtypes for the HC category contain an equal number of excerpts.

[46] Because deceptive cadences occur less frequently in Mozart's keyboard sonatas than the other cadence categories selected for this study, the two subtypes do not contain an equal number of excerpts: the failed perfect authentic cadence subtype contains seven while the failed imperfect authentic cadence subtype contains three.

(e.g., the presence of a cadential trill, the use of a dissonant six-four chord suspended above the cadential dominant, the temporal duration of the cadential progression, etc.). Thus, in addition to the subtypes, I also characterized each excerpt according to a number of rhetorical parameters in order to examine their relative contribution in a regression model (see §7.2.3).

§7.2 Experiment I

7.2.1 Method

Participants

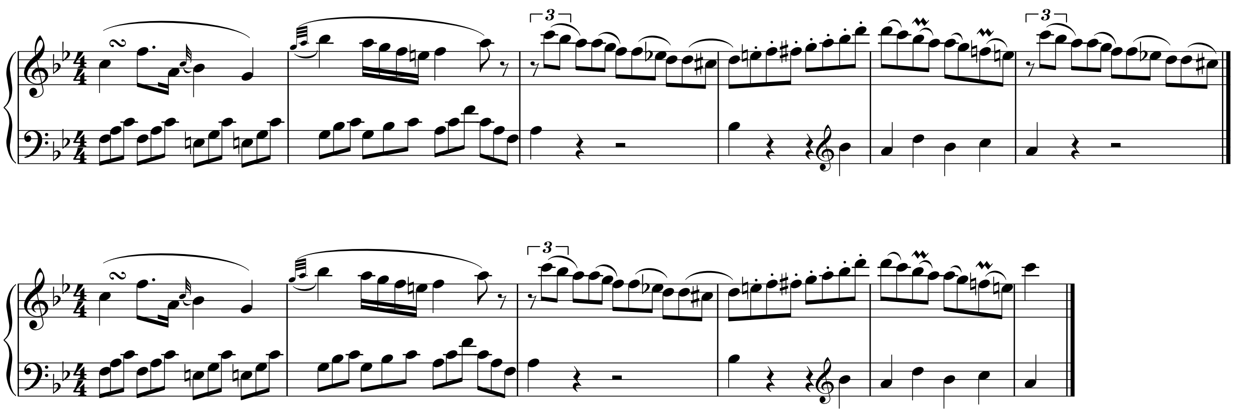
Participants were 40 members (19 male) of the Montreal community recruited through the Schulich School of Music and the McGill University classified ads. Ages ranged from 18 to 48 ($M = 24$, $SD = 6$). Twenty participants with music training equivalent or superior to second-year-university level formed the musician group, and twenty participants with less than one year of music training comprised the nonmusician group. To limit any effects caused by familiarity with the stimuli, no participant with more than one year of formal study on the piano was permitted to take part.

A questionnaire was administered to assess musical preferences and training. On average, musicians had 11.4 years of study on a musical instrument (other than piano), 3.5 years of ear training, 3.0 years of instruction in harmony, and 2.9 years of instruction in music analysis. At the time of their participation, they additionally reported spending an average of 15.2 hours each week engaged in instrumental practice. Participants also listened to an average of 11.3 hours of music each week. All of the participants reported normal hearing, which was confirmed with a pure-tone audiometric test using a MAICO MA 39 audiometer in which participants were required to demonstrate minimum hearing thresholds at or below 20 dB HL for octave-spaced frequencies from 125 Hz to 8 kHz.[47] None of the participants indicated they had absolute pitch.

Materials

The stimuli consisted of 50 short excerpts (average 8.6 s) selected from Mozart's keyboard sonatas. To limit the number of variables under consideration, performance features (such as dynamics and rubato) were neutralized and the tempo of each excerpt was determined by convention. Each stimulus was first created with the notation software Finale and then realized as a .wav sound file using a piano physical model created by PianoTeq. Finally, a 4-s fade-in was inserted at the beginning of each excerpt to encourage participants to attend specifically to the excerpt's end.

Unfortunately, the extraction of each excerpt from its surrounding material introduced a number of factors at the moment of cadential arrival that might confound the experimental outcome. To eliminate these unwanted effects while preserving the stylistic integrity of each excerpt, it was necessary to impose a few constraints on the materials appearing at the cadential arrival. First, any chord tones appearing after cadential arrival (e.g., an Alberti bass pattern) were verticalized to the moment of cadential arrival and all subsequent material was removed. This alteration was necessary in order to eliminate differences in surface activity among excerpts, in particular for instances in which the absence of the third of the triad at the point of arrival would have resulted in an unstylistic open octave. Second, I recomposed the duration of the cadential arrival to one full tactus to ensure that differences in duration would not affect the perception of closure. This change still resulted in small variations in the duration of the final

[47] ISO 389-8, *Acoustics: Reference Zero for the Calibration of Audiometric Equipment—Part 8: Reference Equivalent Threshold Sound Pressure Levels for Pure Tones and Circumaural Earphones*, Technical Report (International Organization for Standardization, 2004); Frederick N. Martin and Craig A. Champlin, "Reconsidering the Limits of Normal Hearing," *Journal of the American Academy of Audiology* 11 (2000): 64–66.

Example 7.2: EV category, Tonic Harmony subtype: K. 281/iii, mm. 30–35. Top: From score. Bottom: Recomposed.

event for each excerpt, but these differences were assumed to be too small to significantly affect the completion ratings. Third, because I did not wish to consider the effect of cadential absence—such as when a rest replaces the expected tonic at cadential arrival—in two instances the events following the rest were shifted back to cadential arrival (see Example 7.2). Finally, any melodic dissonances appearing at the cadential arrival were retained so as not to fundamentally alter the excerpt (for example, in evaded cadences the melodic line frequently features an appoggiatura at the point of the expected cadential arrival).[48]

Design and Procedure

The participants were seated in a double-walled IAC Model 1203 sound-isolation chamber. The stimuli were reproduced on a Macintosh G5 PowerPC, output as S/PDIF using an M-Audio Audiophile 192 sound card, converted to analog using a Grace Design m904 monitor system, and presented over a pair of Dynaudio BM6A monitors. The stimuli were presented at 55 dB SPL, which was kept constant for all participants throughout the experimental session. The experimental program, subject interface, and data collection were programmed using the

[48]See, for example, Appendix B, Example 45.

Max/MSP environment from Cycling 74' controlled by the PsiExp software environment.[49]

Participants were presented with a randomized set of all 50 excerpts in two blocks. After listening to each excerpt up to three times, participants were instructed to rate the degree of *completion* of each excerpt on a 7-point continuous analogical-categorical scale,[50] which consists of an analog scale subdivided into seven discrete categories labeled from 1 to 7. Completion was defined as: "the expectation that the music will not continue. A value of 1 indicates that the excerpt would certainly continue, whereas a value of 7 indicates that the excerpt could end at that moment without the need for anything further." Participants were encouraged to use the full range of the scale over the course of the experiment. In addition, at no point was the term 'cadence' ever mentioned during the session under the assumption that its usage might unintentionally bias musicians toward consciously categorizing the excerpts' endings.

In addition to a completion judgment, participants rated on 7-point scales the *confidence* of their completion rating and their *familiarity* with the excerpt. To distinguish between those excerpts potentially rated in the center of the completion scale, the participants also responded to the following two statements on a 4-point Likert scale labeled from *strongly agree* to *strongly disagree*: "this excerpt could complete an entire work or movement," and "this excerpt could complete a phrase or short passage of music." The aim of the additional 4-point Likert response scales was to ask participants to differentiate between endings located at the conclusion of a longer work from those endings they may deem sufficient to complete a phrase or short passage within that work, a distinction assumed too subtle to be captured by the completion scale, particularly for those excerpts that participants placed in the center of the scale (i.e., excerpts rated as neither entirely complete nor entirely incomplete). The participants were

[49] Bennett K. Smith, "PsiExp: An Environment for Psychoacoustic Experimentation Using the IRCAM Musical Workstation," Paper presented at the Society for Music Perception and Cognition, 1995, Berkeley, CA.

[50] Reinhard Weber, "The Continuous Loudness Judgement of Temporally Variable Sounds with an 'Analog' Category Procedure," in *Fifth Oldenburg Symposium on Psychological Acoustics*, ed. A. Schick, J. Hellbrück, and R. Weber (Oldenburg: BIS, 1991), 267–294.

also reminded that the two scales should not necessarily co-vary. By strongly agreeing that an excerpt could complete an entire work or movement, a participant might also assume it could complete a phrase or short passage of music. The reverse is not necessarily true, however, as an excerpt might provide an entirely implausible ending for an entire work, yet sound satisfactory at the end of a short passage within the work.

To familiarize the participants both with the range of stimuli as well as with the experimental task, the experimental session began with an exposure phase consisting of ten additional excerpts (two excerpts exemplifying each cadence category), and a practice phase in which the participants rated each of the ten excerpts. After completing the experiment, participants filled out a short questionnaire addressing their music background.

Analysis

Each of the dependent variables was analyzed using a 5 × 2 mixed-design analysis of variance (ANOVA) with a within-participant factor of cadential category (PAC, IAC, HC, DC, and EV) and a between-participant factor of musical training (musicians, nonmusicians). To further consider differences of formal context (PAC and HC), melodic dissonance (IAC), melodic scale-degree (DC), and harmony at the cadential arrival (EV), a separate 2 ×2 ANOVA was calculated for each cadential category. For example, for the perfect authentic cadence category, a 2 ×2 model was calculated with a within-participant factor of formal context (main theme, subordinate theme) and a between-participant factor of musical training (musicians, nonmusicians). Because the behavioral scales for the completion, confidence, and familiarity ratings are bounded on both sides (by 1 and 7), in a few cases the aggregated data for the perfect authentic and evaded cadential categories—the categories expected to elicit very high and very low completion ratings, respectively—violated assumptions of normality and homogeneity of variance due to their skewed distributions. To eliminate these issues before calculating the model, the completion, confidence, and familiarity ratings were normalized to the range [0-1] and an arcsin transformation was applied. In the figures that follow, however, the raw data were retained for the purposes of visualization. To counteract violations of sphericity, degrees of freedom are reported using the Greenhouse-Geisser correction where appropriate. Finally, all post hoc statistics were conducted using t-tests with Bonferroni adjustment.

7.2.2 Results

I will first describe the results as they relate to the cadential categories and then discuss differences arising between the various subtypes. In §7.2.3, I present methods for modeling the completion ratings using multiple linear regression.

Cadence Categories

Figure 7.2 displays bar plots of the completion, confidence, and familiarity ratings for each of the five cadential categories. A mixed 5×2 ANOVA of the completion ratings revealed main effects of cadential category, $F(3.14, 119.27) = 264.56$, $\varepsilon = 0.79$, $p < .001$, $\eta^2 = .85$, and music training, $F(1, 38) = 8.39$, $p < .01$, $\eta^2 = .18$, as well as a significant interaction, $F(3.14, 119.27) = 8.52$, $p < .001$, $\eta^2 = .03$. For the musician group, post hoc analyses revealed significant differences between each pair of cadential categories descending from PAC to EV ($p < .01$), with the exception of a marginal difference appearing for the HC-DC pair ($p = .06$). The membership of each excerpt to a cadential category therefore appeared to significantly affect the completion ratings for the musician group. Musicians and nonmusicians did not differ in their ratings for any of the categories of genuine cadences (PAC, IAC, HC), but the results did reveal an effect of music training for the failed cadences (DC, EV), as nonmusicians provided higher completion ratings for both deceptive and evaded cadences than did musicians ($p < .001$). Indeed, nonmusicians did not rate half cadences as any more complete than evaded

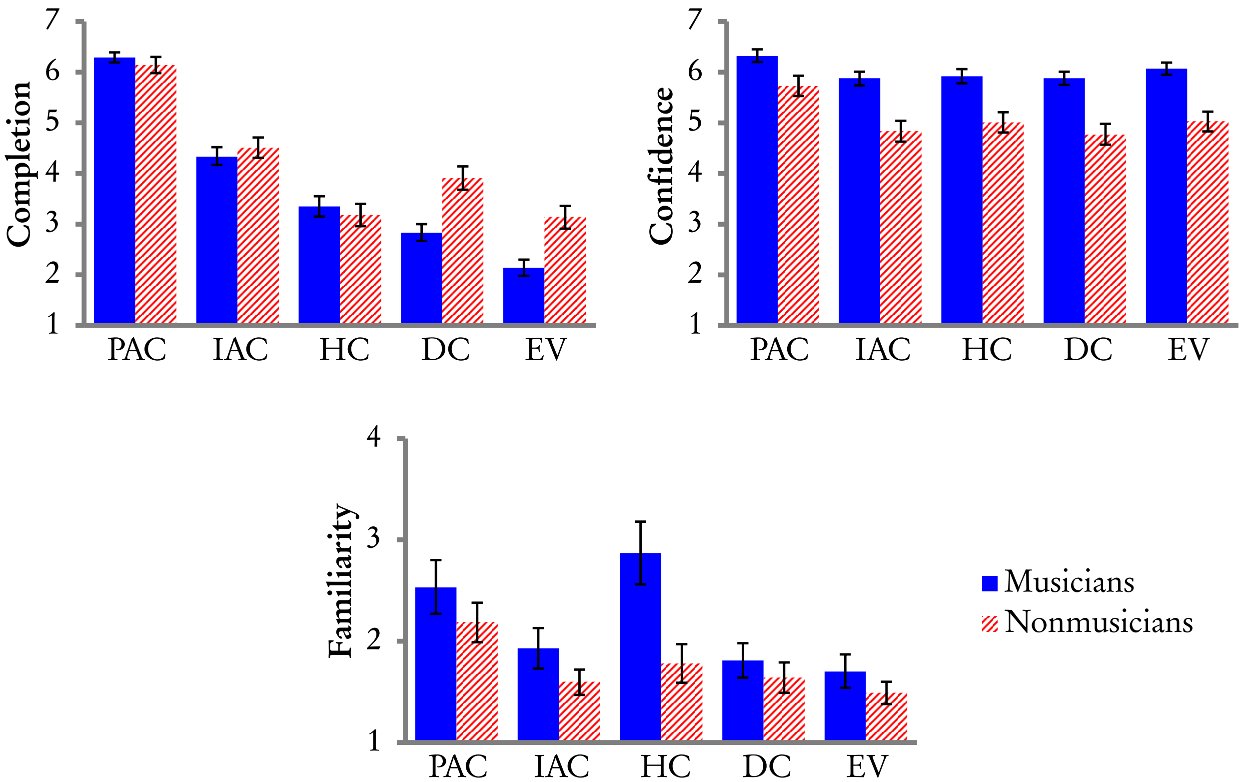


Figure 7.2: Bar plots of mean completion, confidence, and familiarity ratings of musicians (solid blue) and nonmusicians (diagonal red) for each cadential category. Whiskers represent the 95% confidence interval.

cadences ($p > .05$).

As expected, confidence ratings averaged across participants were weakly correlated with the completion ratings, Spearman's rank correlation coefficient $r_s(48) = .36$, $p = .01$. A mixed 5×2 ANOVA revealed a significant effect of cadence category, $F(2.52, 95.75) = 25.44$, $\varepsilon = 0.63$, $p < .001$, $\eta^2 = .39$, as both groups provided higher confidence ratings for excerpts from the PAC category than for any of the other cadence categories ($p < .01$). Confidence ratings were also higher for musicians than for nonmusicians, $F(1, 38) = 18.93$, $p < .001$, $\eta^2 = .33$. Although both groups provided low familiarity ratings on average, completion ratings were nonetheless weakly correlated with familiarity, $r_s(48) = .50$, $p < .001$. Familiarity judgments also revealed main effects of cadential category, $F(2.95, 112.01) = 34.23$, $\varepsilon = 0.74$, $p < .001$, $\eta^2 = .41$, and training, $F(1, 38) = 4.47$, $p = .04$, ${\eta_p}^2 = .11$, as well as an unexpected interaction resulting from differences in the way musicians and nonmusicians specifically rated half cadences, $F(2.95, 112.01) = 12.26$, $p < .001$, $\eta^2 = .14$. As shown in Figure 3, whereas both groups provided higher familiarity ratings for the perfect authentic cadence category than for the other categories ($p < .05$), musicians also rated excerpts from the half cadence category as more familiar than those from the remaining categories ($p < .05$). This effect was not observed in nonmusicians, however, as they provided very low familiarity ratings for excerpts ending in half cadences, nor did these ratings differ from those of the other categories. The intention behind providing a scale for familiarity was to determine if explicit prior exposure to a particular excerpt would affect completion ratings, assuming that knowledge of the material following the end of the excerpt might alter the interpretation of that excerpt's ending, thus confounding the experimental outcome. However, the observed difference in familiarity ratings of excerpts ending with half cadences between musicians and nonmusicians instead suggests a difference in their exposure to, and subsequent knowledge of, half cadences in general, a particularly compelling finding that appears to contradict the completion data, in which no training effect was observed for half cadences.

Cadence Subtypes

Figure 7.3 presents bar plots of the completion ratings for the cadential subtypes of each cadence category. Beginning with the perfect authentic cadence category, both groups rated PACs selected from the subordinate theme as more complete than PACs from the main theme, $F(1, 38) = 23.43$, $p < .001$, $\eta^2 = .38$, and there was no effect of training, $F(1, 38) < 1$. For the imperfect authentic cadence category, the presence of a melodic dissonance at the cadential arrival only affected completion ratings for nonmusicians ($p < .01$). In addition, nonmusicians

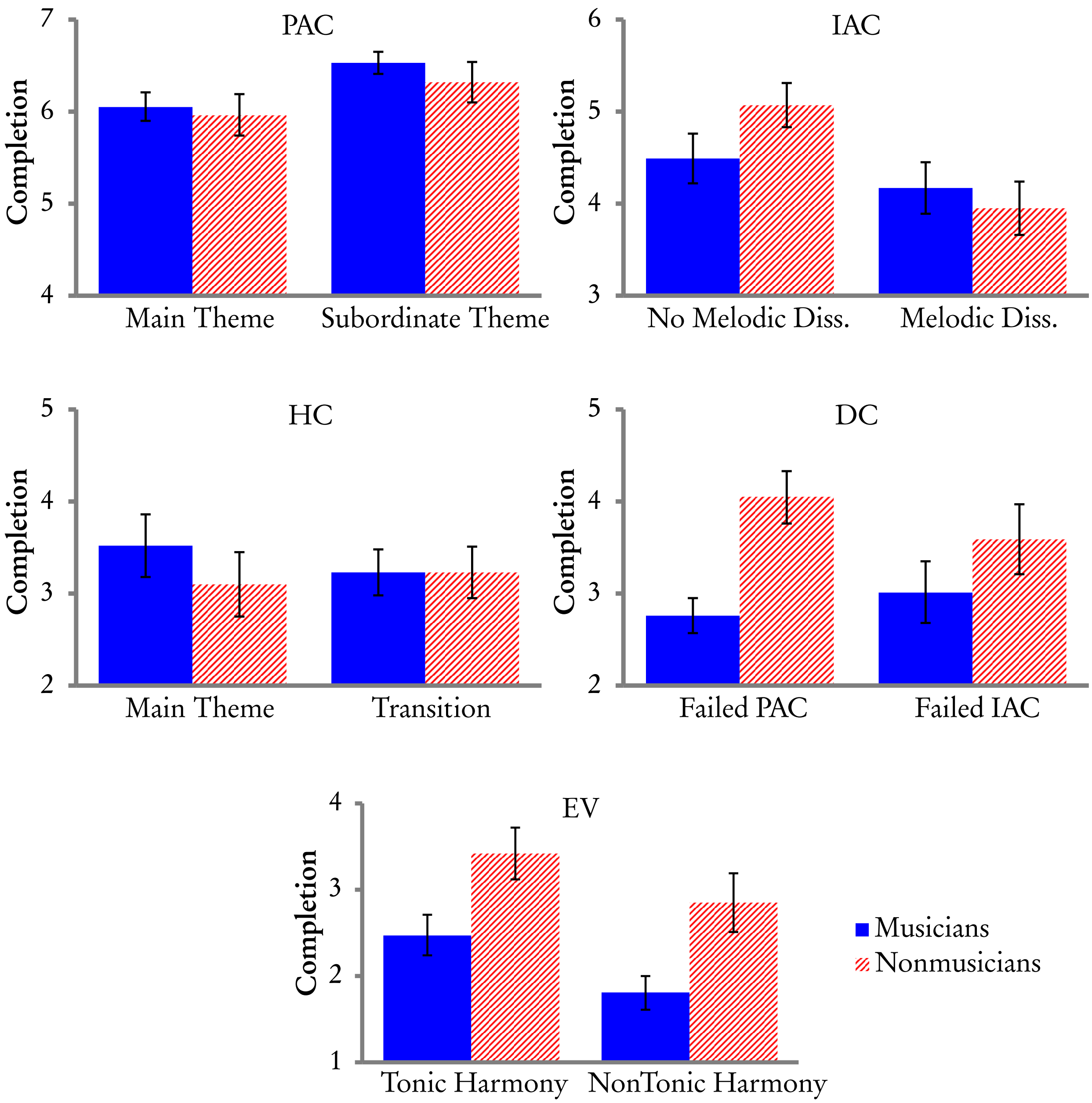


Figure 7.3: Bar plots of mean completion ratings of musicians (solid blue) and nonmusicians (diagonal red) for each subtype. Whiskers represent the 95% confidence interval.

provided higher completion ratings for imperfect authentic cadences without a dissonance at the cadential arrival than did musicians ($p < .05$). For the half cadence category, musicians tended to rate half cadences selected from transitions as less complete than those selected from the main theme, but this interaction between formal context and music training did not reach significance, $F(1, 38) = 3.10$, $p = .08$, nor did the completion ratings differ with respect to training, $F(1, 38) < 1$. However, excerpts from the HC category were also classified according to the presence or absence of a melodic dissonance at cadential arrival, and for both groups excerpts that contained a melodic dissonance received significantly lower completion ratings than excerpts that did not, $F(1, 38) = 4.5$, $p < .05$, $\eta^2 = .15$, and there was no effect of training, $F(1, 38) < 1$.

The completion ratings for the deceptive cadence category revealed a significant effect of expertise, $F(1, 38) = 24.0$, $p < .001$, $\eta^2 = .39$, and the scale degree at the cadential arrival significantly affected the completion ratings of nonmusicians ($p < .05$), as they provided higher ratings for deceptive cadences featuring melodic $\hat{1}$ than for those featuring melodic $\hat{3}$. However, the scale degree at the cadential arrival did not affect the ratings of the musician group. Finally, the evaded cadence category featured a significant effect of expertise, $F(1, 38) = 28.16$, $p < .001$, $\eta^2 = .43$, with musicians providing much lower completion ratings for evaded cadences than nonmusicians. Additionally, the harmony at the cadential arrival significantly affected the completion ratings of both participant groups, as evaded cadences with tonic harmony at the cadential arrival received higher completion ratings than cadences with non-tonic harmony ($p < .05$).

Movement Completion and Phrase Completion Ratings

The movement ratings provided very few notable results, as both groups tended to *strongly disagree* with the statement that the excerpts could complete a work or entire movement.

However, the phrase ratings revealed a significant effect of music training for the half cadence category. Figure 7.4 provides a bar plot of the distribution of the percentage of responses for each cadential category for the statement, “this excerpt could complete a phrase or short passage of music,” with musician ratings above and nonmusician ratings below the x-axis. This representation therefore visually estimates the similarity between the two groups by evaluating the symmetry about the x-axis.

A chi-square test was performed to determine the minimum number of trials necessary to reach significant agreement for each category. Out of 200 trials within each category (20 participants × 10 excerpts), a minimum of 68 identical responses (or 34%) was necessary to achieve significance, $\chi^2(1) = 3.89$, $p < .05$. The horizontal dotted lines above and below the x-axis indicate this minimum agreement threshold. The very first category in the musician group, for example, indicates that in 85% of all trials musicians *strongly agreed* that excerpts from the PAC category could complete a phrase or short passage of music. For the IAC and HC categories musicians generally *agreed* with this statement, although they generally *disagreed* that excerpts from the EV category could complete a short passage. Concerning the DC category, both groups wavered between *agree* (40%) and *disagree* (37.5%), and although musicians minimally preferred to disagree with the statement whereas nonmusicians preferred to agree, this difference was not significant, Mann-Whitney $U = 18,017$, $p = .067$, $r = -.09$. Indeed, whereas both groups differed with respect to the absolute percentage of agreement of their responses, the shape of the distribution for each cadential category remained fairly similar between the two groups. In the case of the HC category, however, the difference in the responses of musicians and nonmusicians was significant, $U = 13,923$, $p < .001$, $r = -.28$; in over 66% of their responses musicians *agreed* or *strongly agreed* that a half cadence could complete a phrase or short passage, whereas nonmusicians *disagreed* or *strongly disagreed* in 63% of their responses.

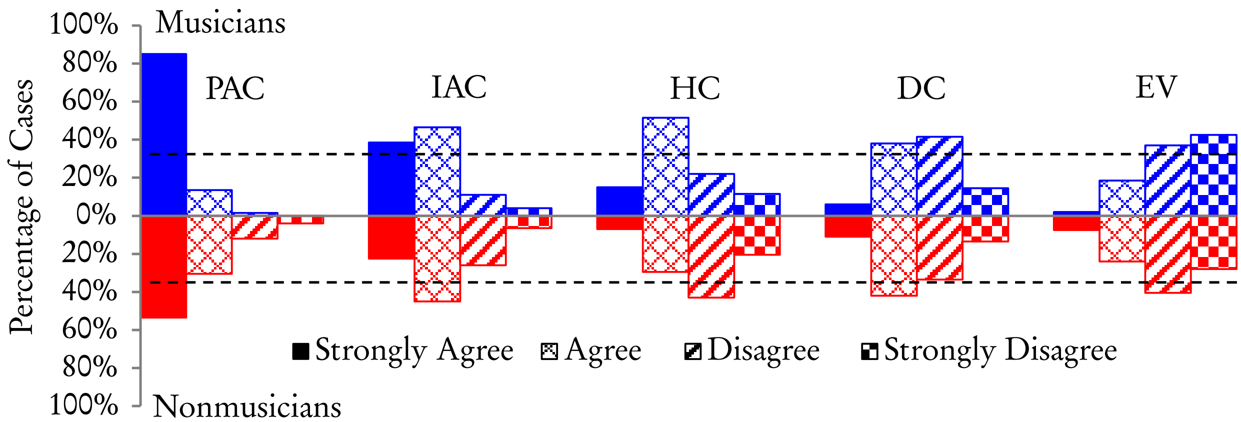


Figure 7.4: Bar plot of the distribution of the percentage of responses for each cadential category for the statement, “this excerpt could complete a phrase or short passage of music,” with musicians’ ratings above the x-axis in blue and nonmusicians’ ratings below in red. Pattern fills denote response types. The horizontal dotted lines indicate the minimum threshold necessary to reach significant agreement.

7.2.3 Modeling the Completion Ratings

Rhetorical Features

The purpose of the preceding analysis was to determine the role various cadential categories might play in the perception of closure. To that end, five distinct cadential categories were selected on the basis of their harmonic-melodic content. But in selecting examples from the extant musical literature, the stimuli fail to control for a number of rhetorical parameters that may affect participants’ ratings of completion (e.g., tempo, event density, a cadential trill, etc.). To be sure, the syntactic characteristics that distinguish the cadence categories employed in this study are frequently accompanied by a number of other features that may also facilitate the identification of cadences. A trill, for example, may serve as a contextual cue that alerts the listener to an impending cadential ending. The following analysis therefore considers the degree to which these rhetorical parameters might predict the participant completion ratings. Unfortunately, the small number of stimuli employed for this study (50) prohibits establishing a reliable multiple linear regression (MLR) model embracing the vast number of musical parameters implicated in the articulation of cadences. Nonetheless, this correlational approach may lead to the identification of a small number of parameters ($k < 6$) to be examined in future studies.[51]

By considering rhetorical parameters, my assumption was that the cadence categories alone would not account for all of the variance in the completion ratings. Before encoding the rhetorical parameters, it was therefore necessary to determine the effect of cadence category membership in a regression model. Each excerpt's cadential category was encoded along an ordinal scale from PAC to EV, following the ranking displayed in the musician completion ratings (see the completion ratings in Figure 7.2). This ranking accounts for 84% of the variance in the mean completion ratings from the musician group (standardized coefficient $\beta = .92$). However, the ranking only explains roughly 53% of the variance in the nonmusician completion ratings ($\beta = .73$). This difference in the estimates of the MLR models may result from differences in selective attention during the perception of closure, or it may suggest that rhetorical parameters have a more significant impact on nonmusicians.

To consider the role these rhetorical parameters might play, 12 features were selected that characterize (1) the entire stimulus, (2) the cadential progression, and (3) the moment of cadential arrival (see Table 7.2):

1. **Entire Stimulus**. Four features characterize the entire stimulus: the tempo in beats per minute (*Tempo*), the total number of notes per second (*Event Density*), the median pitch height in MIDI note values (*Median Pitch Height*), and the duration of the stimulus in seconds (*Stimulus Duration*).

2. **Cadential Progression**. Three dichotomous features and one continuous feature char-

[51] Jeremy Miles and Mark Shevlin, *Applying Regression and Correlation* (London: Sage Publications, 2001).

Table 7.2: Descriptive statistics for the 12 rhetorical features.

Rhetorical Features	*M* (*SD*)	*Range*	*Mode* (*Frequency*)
Entire Stimulus			
(1) *Tempo* (bpm)	104 (31)	40–152	
(2) *Event Density*[a]	6.3 (2.6)	2.7–13.3	
(3) *Median Pitch Height* (MIDI note number)	68 (5)	58–81	
(4) *Stimulus Duration* (s)	9.6 (2.3)	5.6–16.8	
Cadential Progression			
(5) *Complete*[b]			Present (30)
(6) *Cadential Trill*			Absent (33)
(7) *Cadential* $^6_4$			Present (38)
(8) *Cadential Progression Duration* (s)	3.9 (2.3)	0.9–12.2	
Cadential Arrival			
(9) *Melodic Dissonance*[c]			Absent (39)
(10) *Metric Downbeat*			Present (42)
(11) Δ *Event Density*[d]	6.7 (5.9)	2–23	
(12) *Tactus Duration* (s)	0.8 (0.5)	0.2–2.3	

[a] *Event Density* refers to the number of notes per second.
[b] *Complete* refers to an authentic cadential progression that includes an initial tonic, a pre-dominant, a dominant, and a final tonic, or to a half cadential progression that includes an initial tonic, a pre-dominant, and a dominant.
[c] *Melodic Dissonance* may refer to an appoggiatura, an accented passing tone, or a dissonant suspension.
[d] Δ *Event Density* was calculated as the difference between the sum of the events in a 1-s window preceding cadential arrival to the sum of the events in a 1-s window beginning at cadential arrival.

acterize the cadential progression: the presence of every harmonic function within the boundaries of the cadential progression (*Complete*), the presence of a cadential six-four (*Cadential* $^6_4$), a trill within the cadential dominant (*Cadential Trill*), and the duration of the cadential progression in seconds (*Cadential Progression Duration*).

3. **Cadential Arrival**. Two dichotomous features and two continuous features characterize the cadential arrival: the presence of a surface dissonance in the melody (*Melodic Dissonance*), the metric location of the final harmony, which can occur either on or off the downbeat (*Metric Downbeat*), the change in event density at the cadential arrival (Δ *Event*

Density), and the duration of the events at the cadential arrival (*Tactus Duration*).

To limit the number of predictors input into the final MLR models, correlations were calculated for each of the rhetorical features with the completion ratings from both groups. Shown in Table 7.3, intercorrelations between the rhetorical features displayed very few noteworthy results, with only three correlation coefficients registering above .50. Two features in particular, *Complete* and *Cadential Trill*, were not correlated with any of the other rhetorical features. Moreover, the majority of the moderate-to-strong correlations shown in Table 7.3 arose as a result of features that characterize temporal aspects of the stimulus. For example, stimulus tempo was correlated with the duration of the tactus at cadential arrival, $r(48) = -.46$, as well as of the stimulus as a whole, $r(48) = -.36$. However, the correlations between *Melodic Dissonance* and two other features—*Metric Downbeat* and *Tactus Duration*—were noteworthy, as excerpts that contained a melodic dissonance at cadential arrival also lengthened the duration of cadential arrival and placed it on a metric downbeat, a result that suggests a compositional strategy to accentuate the effect of the dissonance.

The musician completion ratings were significantly correlated with four rhetorical features—*Median Pitch Height*, *Complete*, *Cadential Trill*, and Δ *Event Density*—and the non-musician ratings were correlated with two features—*Median Pitch Height* and *Cadential Trill*. However, these correlations could simply result from their collinearity with a third variable, the rank order of cadential categories. It was therefore necessary to control for cadence category membership first by calculating the correlation between the completion ratings and the rank order of cadential categories, and then correlating each of the rhetorical features with the residuals. Controlling for the effect of cadential category resulted in significant correlations for four features with the musician ratings—*Melodic Dissonance*, semi-partial correlation coefficient $sr(48) = .43$, Δ *Event Density*, $sr(48) = .40$, *Cadential Trill*, $sr(48) = .35$, and *Cadential*

Table 7.3: Intercorrelations between the rhetorical features and the mean completion ratings of musicians and nonmusicians.

Features & Ratings	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Entire Stimulus														
(1) *Tempo*		−.06	.05	−.36**	−.05	.2	−.02	−.2	−.05	−.05	.55***	−.46***	.15	.04
(2) *Event Density*			−.02	.03	.04	.07	.2	.08	−.03	−.07	.2	−.01	−.1	−.05
(3) *Median Pitch Height*				.1	.03	−.11	.06	−.04	−.11	.08	.09	−.28*	.29*	.33*
(4) *Stimulus Duration*					.07	−.23	.31*	.37**	.05	.34*	−.35*	.44***	.13	.24
Cadential Progression														
(5) *Complete*						−.28	−.17	−.19	.16	−.13	.09	.18	.31*	.16
(6) *Cadential Trill*							.21	−.2	−.03	.08	−.02	−.15	−.42**	−.41**
(7) *Cadential* $^6_4$								−.08	−.7	.27	−.16	.00	−.1	−.11
(8) *Cadential Progression Duration*									.01	.05	.09	.22	.17	.25
Cadential Arrival														
(9) *Melodic Dissonance*										.43**	−.14	.57***	−.05	−.28
(10) *Metric Downbeat*											−.19	.22	−.08	−.15
(11) Δ *Event Density*												−.57***	.33*	.21
(12) *Tactus Duration*													−.08	−.12
Completion Ratings														
(13) *Musicians*														.87***
(14) *Nonmusicians*														

Note. $N = 50$.
* $p < .05$ ** $p < .01$ *** $p < .001$, two-tailed.

Progression Duration, $sr(48) = .26$—and four features with the nonmusician ratings—*Melodic Dissonance*, $sr(48) = .55$, *Stimulus Duration*, $sr(48) = .30$, *Cadential Progression Duration*, $sr(48) = .28$, and *Cadential Trill*, $sr(48) = .28$.

Given the small number of remaining features, a stepwise MLR model with forward selection was fitted for both the musician and nonmusician completion ratings using the cadential ranking $PAC > IAC > HC > DC > EV$ and the rhetorical features provided above. To examine effects of multicollinearity within the two models, correlations were calculated for all of the predictors. Of the five predictors input in the two models, only two—the cadential ranking and *Cadential Trill*—were weakly correlated, $r(48) = .30$, $p < .05$. Moreover, the mean Variance Inflation Factors calculated for the predictors in the musician ($meanVIF = 1.10$) and nonmusician ($meanVIF = 1.02$) models demonstrated very weak multicollinearity.[52]

Shown in Table 7.4, the musician model selected the cadential rankings, *Melodic Dissonance*, *Cadential Trill*, and Δ *Event Density*. These four predictors accounted for 91% of the variance in the completion ratings. The combined size of the effect for the selected rhetorical features was small ($\Delta R^2 = .07$), however, which suggests that the cadential categories play the most substantial role in the musicians' ratings. The nonmusician model selected the cadential rankings, *Melodic Dissonance*, and *Stimulus Duration*, and these three predictors accounted for 72% of the variance in the completion ratings. Moreover, the addition of *Melodic Dissonance* and *Stimulus Duration* significantly improved the fit of the model ($\Delta R^2 = .19$), indicating that

[52]Multicollinearity refers to the presence of significant correlations between model predictors. In regression models, multicollinearity violates a basic assumption of linear regression (and parametric statistics more generally), because in such instances it would be virtually impossible to determine which of the correlated predictors was actually responsible for variations in the dependent variable. The variance inflation factor is a common collinearity diagnostic that quantifies the degree of collinearity for each predictor included in the original model. A variance inflation factor of 1 indicates that the predictor is orthogonal to the other predictors, whereas a value much greater than 1 (say, for example, > 10) indicates that the predictor is highly correlated with one or more of the remaining predictors. Bruce Bowerman and Richard O'Connell suggest a mean variance inflation factor substantially greater than 1 indicates that the model may be biased (*Linear Statistical Models: An Applied Approach* [Pacific Grove, CA: Duxbury, 1990]).

Table 7.4: Summary of stepwise regression analysis predicting the completion ratings of musicians and nonmusicians with the cadential rankings and the correlated rhetorical features from Table 7.2.

		B	$SE\ B$	β
Musicians				
	Step 1			
	Constant	0.85	0.21	
	PAC > IAC > HC > DC > EV	0.98	0.06	.92**
	Step 2			
	Constant	0.27	0.26	
	PAC > IAC > HC > DC > EV	1.01	0.06	.94**
	Melodic Dissonance	0.64	0.19	.18*
	Step 3			
	Constant	0.75	0.28	
	PAC > IAC > HC > DC > EV	0.96	0.06	.90**
	Melodic Dissonance	0.64	0.18	.17*
	Cadential Trill	0.49	0.16	.15*
	Step 4			
	Constant	0.68	0.26	
	PAC > IAC > HC > DC > EV	0.92	0.05	.86**
	Melodic Dissonance	0.54	0.17	.15*
	Cadential Trill	0.51	0.15	.16*
	Δ Event Density	0.04	0.01	.15*
Nonmusicians				
	Step 1			
	Constant	2.2	0.3	
	PAC > IAC > HC > DC > EV	0.67	0.09	.73**
	Step 2			
	Constant	1.12	0.35	
	PAC > IAC > HC > DC > EV	0.71	0.08	.78**
	Melodic Dissonance	1.19	0.26	.38**
	Step 3			
	Constant	−0.1	0.53	
	PAC > IAC > HC > DC > EV	0.7	0.07	.77**
	Melodic Dissonance	1.22	0.25	.39*
	Stimulus Duration	0.12	0.04	.23*

Note. Musicians $R^2 = .84$ for Step 1; $\Delta R^2 = .03$ for Step 2 ($ps < .01$); $\Delta R^2 = .02$ for Step 3 ($ps < .01$); $\Delta R^2 = .02$ for Step 4 ($ps < .01$). Nonmusicians $R^2 = .53$ for Step 1; $\Delta R^2 = .14$ for Step 2 ($ps < .01$); $\Delta R^2 = .05$ for Step 3 ($ps < .01$).
* $p < .01$ ** $p < .001$.

the two rhetorical features played a more substantial role in the nonmusician model than those selected in the musician model.

Melody and Harmony

By retaining the cadential categories, the above models unfortunately fail to consider how features of the soprano and bass voice may independently contribute to the perception of closure, particularly at the moment of cadential arrival. The claim that strategies of attention may differ as a result of music training therefore necessitates a statistical approach in which the harmonic and melodic information of each excerpt is encoded separately, thereby permitting us to abandon the cadential categories proper.

To consider how harmony and melody may independently contribute to participant ratings of completion, we must quantify each predictor according to a set of criteria. First, a simple and fairly intuitive method might be to evaluate the melodic and harmonic content of each excerpt according to concepts of closure derived from music theory. For the purposes of this experiment, the harmony of each excerpt was assigned a value of 2 for a tonic triad in root position at the cadential arrival, 1 for a dominant triad in root position, and 0 for any other harmony in any inversion.[53] The melody of each excerpt was assigned a value of 2 for $\hat{1}$ at cadential arrival, 1 for $\hat{3}$, and 0 for any other scale-degree; henceforth I will refer to the estimates obtained from these variables as the *syntax* model.

While this approach is certainly intuitive, it is also glaringly imprecise, as it fails to consider the effect of each of the possible scale-degrees that might appear at the end of each excerpt. In a second approach, I assigned the mean goodness-of-fit ratings obtained from Krumhansl and Kessler's major and minor key profiles to the scale degrees appearing in the soprano and

[53] Bigand and Parncutt employed precisely this rating system to assess the effect of cadential patterns on tension ratings. See "Perceiving Musical Tension in Long Chord Sequences," 250.

bass line of each excerpt at cadential arrival under the assumption that their profiles signify a cognitive representation of the tonal hierarchy pertaining specifically to endings; I will refer to the estimates obtained from these variables as the *KK* model.[54]

Before entering the tonal stability values of the soprano and bass voices as predictors in a regression model, the two variables were correlated to determine if they violated the assumption of multicollinearity. In this instance, the variables were not correlated, $r(48) = -.07$, $p = .33$. Table 7.5 displays the estimates of the *syntax* and *KK* models for musicians. The *syntax* model selected harmony, with a β of .61, in the first step, accounting for about 42% of the variance in their ratings. The selection of melody in the second step, with a β of .47, significantly improved the fit of the model, which produced a final R^2 of .64. Stepwise selection therefore indicated that the harmony predictor played the most substantial role in accounting for musicians' ratings of completion. Applying the *KK* predictors improved the fit of the regression model, with the two predictors accounting for 72% of the variance in musicians' ratings. The *KK* model also produced similar standardized β weights, with the bass-line scale-degree again playing the more prominent role. Thus, as predicted, musicians placed greater emphasis on the bass voice at cadential arrival.

Table 7.6 displays the estimates of the *syntax* and *KK* models for nonmusicians. The estimates of the *syntax* model for the nonmusician ratings were a near mirror image of those found for the musicians, with the parameters of melody and harmony accounting for 65% of the variance, but with melody, with a standardized β of .64, playing a more prominent role than harmony. With the addition of the goodness-of-fit ratings in the *KK* model, however, the

[54]Presumably the metric position and rhythmic duration of the final events also affect participant ratings of completion, but in only 8 of the 50 stimuli does the cadential arrival appear in a metric position other than the downbeat (see Table 7.2). What is more, the rhetorical feature measuring the duration of the events at cadential arrival, called *Tactus Duration*, was not significantly correlated with the participant completion ratings. In noncadential contexts, these variables might play an important role in the perception of closure, but in these stimuli, the metric position and rhythmic duration of the events concluding each stimulus could not be examined in greater detail.

Table 7.5: Summary of stepwise regression analysis predicting the completion ratings of musicians with the *Syntax* and *KK* models.

Model		*B*	*SE B*	β
Syntax				
	Step 1			
	Constant	2.42	0.28	
	Harmony	1.14	0.19	.66*
	Step 2			
	Constant	1.68	0.26	
	Harmony	1.05	0.15	.61*
	Melody	0.84	0.15	.47*
KK				
	Step 1			
	Constant	−1.10	0.73	
	Bass	0.93	0.14	.70*
	Step 2			
	Constant	−4.16	0.71	
	Bass	0.97	0.10	.73*
	Soprano	0.58	0.09	.49*

Note. *Syntax* R^2 = .42 for Step 1; ΔR^2 = .22 for Step 2 (p < .001); *KK* R^2 = .49 for Step 1; ΔR^2 = .24 for Step 2 (p < .001).
* p < .001.

lopsided influence of the soprano voice diminished somewhat, with the bass-line scale-degree playing a more significant role. Thus, it appears that harmonic and melodic content can predict the completion ratings of both groups, and both models indicate unequivocally that musicians privilege the bass voice. With nonmusicians, however, the relative contribution of the two parameters is less clear-cut. In the *syntax* model, melody accounted for a greater proportion of the variance, while in the *KK* model, regression estimates for the two parameters were nearly identical.

These models support the claim that, in the perception of cadential closure, musicians appear to privilege the bass voice while nonmusicians are more sensitive to subtle differences in

Table 7.6: Summary of stepwise regression analysis predicting the completion ratings of nonmusicians with the *Syntax* and *KK* models.

Model		*B*	*SE B*	β
Syntax				
	Step 1			
	Constant	3.13	0.21	
	Melody	1.05	0.16	.69*
	Step 2			
	Constant	2.42	0.22	
	Melody	0.98	0.13	.64*
	Harmony	0.65	0.13	.44*
KK				
	Step 1			
	Constant	1.12	0.57	
	Soprano	0.63	0.11	.62*
	Step 2			
	Constant	−2.63	0.60	
	Soprano	0.67	0.08	.66*
	Bass	0.68	0.08	.60*

Note. *Syntax* R^2 = .47 for Step 1; ΔR^2 = .18 for Step 2 (p < .001); *KK* R^2 = .39 for Step 1; ΔR^2 = .36 for Step 2 (p < .001).
* p < .001.

the soprano voice. That musical training may indeed influence attention in the perception of closure supports Weiser's claim that training facilitates flexible voice-tracking.[55] Furthermore, a recent study conducted by Psyche Loui and David Wessel showed that, even when presented with a task that explicitly directed participants to attend to the contour of the melody, violations in harmonic expectancy still influenced the behavioral responses of musicians.[56] And because this effect was not observed for nonmusicians, the authors claimed repeated exposure to Western music results in the formation of automatic expectations to harmonic progressions that musicians simply cannot ignore, even when asked to attend to other features of the stimulus. It remains unclear, however, whether attention to bass-line motion in cadential contexts reflects a flexible voice-tracking strategy promoted during explicit formal training (i.e., in a pedagogical setting), or an attentional bias formed simply through implicit exposure to Western music.

Cadential Strength

The previous analysis sought to explain effects of expertise by appealing to differences in the attentional strategies employed during music listening. But perhaps differences in the completion ratings of musicians and nonmusicians might also be explained by taking another approach altogether, one in which we retain the cadential categories and propose a general model of hierarchical cadential closure.

In a model of cadential strength, the perfect authentic cadence represents a good place to begin. From even a cursory glance at the literature, it occupies a central position in music theory, as it clearly represents the *locus classicus* for establishing thematic closure in the high classical period. In Chapter 2 I also suggested that listeners versed in tonal music not only possess a cognitive representation of the tonal hierarchy, but perhaps for those listeners especially familiar with the classical style, a schematic representation explicitly for authentic cadential closure.[57] During music listening, a number of parameters located within the cadential progression may activate our schematic representation of closure in real time, allowing listeners to generate harmonic and melodic expectations concerning the moment of cadential arrival. Accordingly, I noted that any deviation on the musical surface would naturally result in a violation of listener

[55] Weiser, "Rating Cadence Stability," 40–46.

[56] Psyche Loui and David Wessel, "Harmonic Expectation and Affect in Western Music: Effects of Attention and Training," *Perception and Psychophysics* 69, no. 7 (2007): 1084–1092. In a selective attention task, the authors asked participants to respond to the contour of a melody as they were presented with harmonic progressions that were either highly expected, slightly unexpected, or extremely unexpected. They found that the expectancy condition affected the speed and accuracy of the contour judgment for musicians, but had no effect on nonmusicians.

[57] By this I mean not only a representation for harmonic closure, though such a claim has already been made by Rosner and Narmour ("Harmonic Closure," 397–398), but rather for a number of potential characteristics—both syntactic and rhetorical—that appear within the cadential progression. However, the question as to whether listeners actually possess such a representation remains open.

expectations, and thus would be experienced as a decrease in the cadential strength of a given excerpt. Deviations in melodic scale-degree and harmony at the cadential arrival thereby result in cadential categories of diminished strength.[58] In this view, the half cadence represents the weakest cadential category; it is marked not by a deviation in the melodic and harmonic content at cadential arrival, but rather by the absence of that content.

So according to this view, every cadential context is compared to one essential prototype: the perfect authentic cadence. In Chapter 2, I referred to this model of cadential closure as the *1-Schema* model and noted its correspondence to Edward Latham's model, which identifies and subsequently weighs the criteria deemed necessary for establishing cadential closure on a 10-point scale.[59] Recall that he assigns 5 points to tonic harmony and 5 points to the preceding dominant, and he derives these scores from the scale-degrees present in the bass (1.5) and soprano (0.5), from whether the sonority is in root position (1.5), and finally from the presence of particular chord members (0.5) and a contextual feature: whether each sonority serves as a harmonic and melodic goal (1.0). According to his criteria, the PAC category receives between 9 and 10 points (depending on whether the cadential tonic is elided), followed by IAC (8.5–9.5), DC (6.5–8.5), EV (3.5–8.5), and finally HC (3.5–5.0), positioned near the bottom of the scale. His model therefore conceptualizes a half cadence as an *incomplete* authentic cadence.

At the heart of the half-cadence issue is an inherent contradiction: that a dominant, which is the penultimate harmony in an authentic cadential progression, can serve as a satisfactory goal. Indeed, many scholars besides Latham envision the half cadence as an incomplete authentic cadence, one in which the expected resolution to tonic simply never appears.[60] And perhaps

[58]But such deviations need not only pertain to harmonic and melodic expectations. James Hepokoski and Warren Darcy's *attenuated PAC*, in which the moment of cadential arrival is marked by a sudden drop in dynamics or an unexpected shift to the minor mode, provides one such example. See *Elements of Sonata Theory*, 170.

[59]Latham, "Drei Nebensonnen," 308.

[60]See, for example, Hepokoski and Darcy, *Elements of Sonata Theory*. They describe the half cadential dominant as an active dominant (24), one that necessarily implies resolution to an existing or implied tonic (xxv).

the results of our experiment reflect this contradiction. Recall that while the two groups did not differ in their completion ratings for the HC category, musicians provided much higher familiarity ratings than nonmusicians. In contrast to musicians, nonmusicians also generally disagreed that half cadences could complete a phrase or short passage of music. Thus, the effect of expertise on the perception of half cadences remains patently unclear.

Caplin has posited another view of half cadence, in which a dominant, merely by virtue of its harmonic-melodic content, can represent a harmonic end: “In the half-cadential progression, the dominant itself becomes the goal harmony and so occupies the *ultimate* position. To be sure, this dominant usually resolves to tonic, one that normally initiates a new harmonic progression, but within the boundaries of the half-cadential progression itself, the dominant possesses enough stability to represent a harmonic end.”[61] Moreover, recall that Caplin distinguishes the half cadence—one of the genuine cadence categories—from the deceptive and evaded cadences, which represent failed attempts to achieve authentic cadential closure. In Chapter 2, I postulated an alternative to the *1-Schema* model, in which listeners may possess schematic representations for each of the genuine cadences, called the *Genuine Schemas* model. Accordingly, we may rank the strength of cadential closure beginning with the PAC category, followed by IAC and HC, followed by the syntactically weaker cadential categories: the DC and EV categories.

By ranking each cadential category, we may compare the two models with the completion ratings. Given that Latham has provided a method for quantifying closure, I have also calculated the strength of closure for each excerpt using Latham’s criteria, which I will refer to as the *Latham* model. Table 7.7 provides the estimates for each model. For the musicians, the *Genuine Schemas* model accounts for 84% of the variance in their ratings, while the *1-Schema* and *Latham* models were less successful, accounting for between 55–60% of the ratings. For the nonmusicians, however, the *1-Schema* model provided the best fit, accounting for approximately

[61] Caplin, *Classical Form*, 29 (emphasis in original).

Table 7.7: Summary of regression analysis predicting the mean completion ratings for each excerpt using the *Genuine Schemas*, *1-Schema*, and *Latham* models as predictors.

Model	Cadential Strength	*B*	*SE B*	β	R^2
Musicians					
Genuine Schemas	PAC > IAC > HC > DC > EV	0.98	0.06	0.92*	.84
1-Schema	PAC > IAC > DC > EV > HC	0.81	0.10	0.76*	.56
Latham		0.54	0.06	0.77*	.59
Nonmusicians					
Genuine Schemas	PAC > IAC > HC > DC > EV	0.66	0.09	0.74*	.53
1-Schema	PAC > IAC > DC > EV > HC	0.73	0.08	0.8*	.63
Latham		0.47	0.05	0.79*	.61

* $p < .001$.

63% of the variance in their ratings.

What are we to make of this result? The *1-Schema* model assumes that, when presented with a cadential excerpt, listeners have no knowledge of the future, and thus, of the material that may follow cadential arrival. Yet for a listener familiar with the classical style, the material that follows instances of cadential failure often differs considerably from the material following genuine cadential closure. By thwarting the expected moment of cadential arrival, theorists typically conceptualize cadential deception and evasion as a kind of derailment. And in order to attain the cadential closure initially promised, the subsequent passage typically features a continuation of an earlier process, sometimes even a direct repetition of the previous cadential progression itself, a compositional procedure Schmalfeldt refers to as the "one more time" technique.[62] Thus, Caplin refers to the PAC, IAC, and HC categories as *genuine* specifically because they are the only categories that can achieve thematic closure.

What these results may suggest, however, is that the PAC, IAC, and HC categories also achieve *genuine* status by virtue of the material following cadential arrival. A genuine cadence

[62]Schmalfeldt, "Cadential Processes."

therefore not only provides sufficient closure to permit the introduction of new initiating material, but the perceived strength of such a cadence is also crucially influenced by the function of the material following cadential arrival in the retrospective stage; or more generally put, the surrounding context may be crucial to determining the strength of a given cadence.

So perhaps during music listening (and particularly during a first listening), the *1-Schema* model is the default for determining the strength of closure of a given ending, but the material following each cadence subsequently compels listeners to retrospectively re-evaluate their earlier impression, and thus, to adopt a model that embraces a theory of *genuine* cadential closure. Given enough exposure to the style, however, listeners may apply this model even to excerpts presented out of context, which would explain why musicians and nonmusicians disagreed as to whether a half cadence could complete a phrase or short passage of music. But to consider this claim in an experimental setting would nevertheless require stimuli that include the material following cadential arrival, an approach that our experimental design did not permit. To address this limitation, the stimuli in Experiment II include the material following cadential arrival in order to examine how retrospective understanding influences the perception of cadential strength.

§7.3 Experiment II

Experiment II considers two factors that might influence the perception of cadential closure in the retrospective stage: (1) principles of segmental grouping, in which syntactic and rhetorical features either reinforce, weaken, or violate schematic expectations for the terminal event(s) of the cadence; and (2) the formal function of the subsequent passage (i.e., whether it serves as a beginning, middle, or end for a new phrase or theme). The former entails the entire range of parameters responsible for the perception of segment boundaries at various levels of the grouping hierarchy, such as the harmony, melodic scale-degree, metric position, rhythmic duration, and surface rhythmic activity characterizing the events at cadential arrival, the presence of an elision or caesura, a dramatic change in texture or dynamics, and so on. The latter consists of the essential characteristics that distinguish the formal functions residing at the phrase and theme levels in Caplin's form-functional hierarchy, such as a post-cadential closing section or standing on the dominant, a resumption of continuation and/or cadential function within a thematic region, or the beginning of a new theme altogether.

To determine how the completion ratings might change following the inclusion of material after the cadence, it was necessary to design the experiment with two context conditions, such that participants first provide ratings for each excerpt presented *out of context*, as in Experiment I, and then provide ratings again for the excerpts presented *in context*. This within-participant design permits us to consider whether the ratings for excerpts presented *out of context* replicate the ratings obtained in Experiment I, as well as provides a basis for comparing the ratings obtained from both context conditions of Experiment II. Moreover, because the objective in Experiment II was to examine the effect of the subsequent context on the perception of cadential closure, it was also necessary to create a revised stimulus set, with the cadences further categorized according to systematic differences relating to segmental grouping and formal function.

In compositional practice, a number of formal options might follow the genuine cadence categories, but for the deceptive and evaded categories the options are considerably more limited. Thus, for the genuine cadence categories I selected two formal functions to serve as contextual subtypes and one formal function to follow the cadential deviations. To examine the effect of the formal function of the subsequent passage on the change in completion ratings, it was also important to select excerpts in each subtype that maintained a consistent formal function whenever possible, and to conclude each excerpt before reaching another cadence, double barline, or caesura, as these subsequent arrivals could bias participants towards rating the later ending. I also attempted to select formal options following each of the genuine cadences that appear with enough regularity in Mozart's keyboard sonatas to warrant further experimental investigation. Finally, the results obtained in Experiment I also indicated that surface dissonances influence the perception of cadential closure for nonmusicians. For Experiments II-V, however, it was desirable to maintain consistent stimuli to provide a basis of comparison for the results obtained across all of the experimental tasks. Experiment V specifically required participants to indicate as quickly as possible whether the final harmony of the cadential progression was in or out of tune (see Chapter 8), so it was essential to synchronize the events at cadential arrival and remove any surface dissonance that could affect participant accuracies and reaction times. Table 7.8 presents the cadence categories, a description of the post-cadential contextual subtypes, and the reference information for each excerpt. The excerpts in bold were not selected in Experiment I, and excerpts that included a surface dissonance at cadential arrival are marked with an asterisk in the table.

Excerpts from the PAC category were classified according to whether the post-cadential context featured either the beginning of the transition following the end of the main theme, or a closing section appearing at the end of the subordinate theme. Example 7.3a presents a perfect authentic cadence from the second movement of Mozart's keyboard sonata in B-flat, K. 333, with the resolution of the cadential idea appearing *out of context* to the right and above the initial cadential progression, and with Mozart's realization of the cadence presented *in context* below. The arrow symbol in the caption indicates that the subtype "follows" the cadence, as in Example 7.3a, in which the transition follows the PAC. The transition begins in the penultimate measure in the example, one measure after the cadence (i.e., PAC→Transition), but the appearance of a new accompanimental bass pattern in the left hand coincides with the moment of cadential arrival, potentially weakening the finality of the cadence. Since the structural beginning of the

Table 7.8: Cadence categories, a description of the post-cadential context, and reference information (Köchel index, movement, measures) for each excerpt. These excerpts comprised the stimulus set for Experiments II-V.

Cadence Categories	*Post-Cadential Context*	*Excerpts*
Perfect Authentic	→ *Beginning*	K. 281, i, mm. 5–11
	• Transition	**K. 283, i, mm. 10–20**
		K. 310, ii, mm. 5–10
		K. 333, ii, mm. 5–10
	→ *After-the-End*	**K. 279, i, mm. 26–33**
	• Closing Section	K. 309, i, mm. 48–57
		K. 333, i, mm. 54–62
		K. 545, i, mm. 20–27
Imperfect Authentic	→ *Beginning*	**K. 282, i, mm. 3–6**
	• Consequent	K. 309, ii, mm. 1–6
	• Transition	**K. 311, i, mm. 1–6**
		***K. 333, ii, mm. 22–28**
	→ *Middle/End*	*K. 311, ii, mm. 27–35
	• Varied Repetition	*K. 330, i, mm. 4–11
		*K. 330, iii, mm. 39–47
		*K. 533, iii, mm. 23–29
Half	→ *Beginning*	*K. 284, iii, mm. 1–7
	• Consequent	*K. 311, ii, mm. 1–6
		K. 332, ii, mm. 3–5
		***K. 547a, i, mm. 3–11**
	→ *Beginning*	*K. 280, i, mm. 21–30
	• MC → *ST*	***K. 330, i, mm. 14–21**
		K. 333, iii, mm. 60–67
		K. 533, i, m. 36–44
Deceptive	→ *Middle/End*	K. 279, i, mm. 7–12
	• Varied Repetition	K. 280, ii, mm. 16–20
		K. 281, ii, mm. 32–38
		*K. 282, i, mm. 11–14
		*K. 282, iii, mm. 25–34
		***K. 332, iii, mm. 22–31**
		K. 457, i, mm. 42–52
		*K. 457, ii, mm. 9–12
Evaded	→ *Middle/End*	K. 279, ii, mm. 1–5
	• Varied Repetition	**K. 280, i, mm. 3–12
		K. 281, ii, mm. 96–102
		K. 281, iii, mm. 30–38
		K. 309, i, mm. 13–20
		K. 309, i, mm. 43–49
		K. 309, iii, mm. 11–18
		K. 333, iii, mm. 84–90

Note. Excerpts in bold were not selected in Experiment I.
* denotes the removal of a surface dissonance at cadential arrival.
** denotes the removal of an arpeggiated triad at cadential arrival.

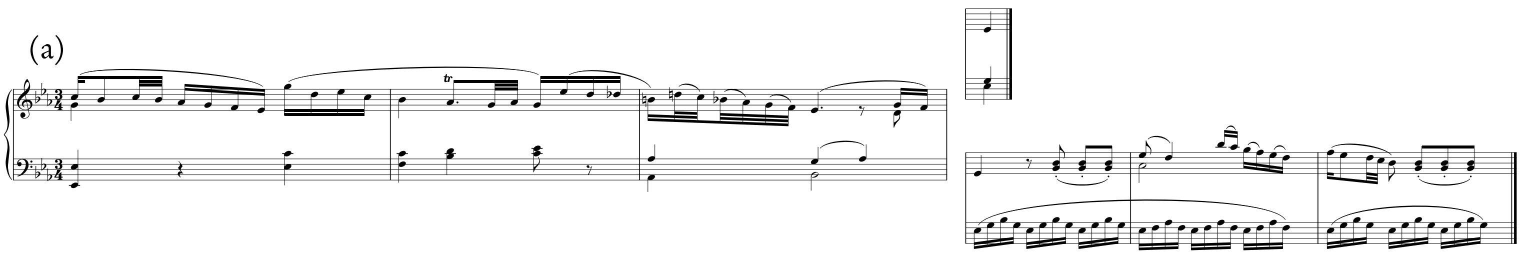
transition appears one measure *after* the moment of cadential arrival, Caplin distinguishes this procedure, which he calls *accompanimental overlap*, from a genuine thematic *elision*, in which the end of the main theme coincides with the beginning of the transition. Accompanimental overlaps like this one are quite common at the juncture between main theme and transition within sonata form, and they feature in three of the four examples selected for this subtype.[63]

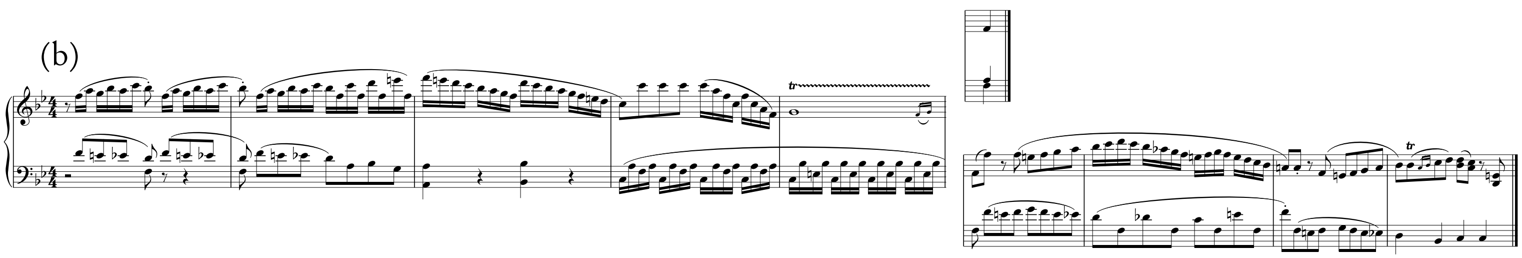
In Example 7.3b, a closing section comprised of a two-measure codetta and its repetition follows the cadence. As stated previously, for examples like this one that feature a cadence within the closing section, I concluded the excerpt before the resolution of the second cadence to discourage listeners from rating the more recent cadence (see the cadential progression in m. 62).[64] All four excerpts from this subtype were selected from Allegro movements and feature an expanded cadential progression and a cadential trill above the penultimate dominant.

For the IAC category, excerpts were classified according to whether a new initiating phrase/theme or a varied repetition of the preceding continuation/cadential material followed the cadence. Thus, cadences in the former subtype appeared most frequently in the main theme in a period theme type, whereas cadences in the latter subtype appeared within the subordinate theme as cadential deviations of the perfect authentic cadence. The IAC in Example 7.4a ends the main theme, and like the previous example, features an accompanimental overlap between the cadential arrival in m. 4 and the transition in m. 5. Two of the other examples from this subtype end an antecedent phrase in a period (K. 311/i, mm. 1–10; K. 333/ii, mm. 22–28). The remaining example from the IAC→Beginning subtype was difficult to classify, however,

[63]In K. 289/i, mm. 10–20, a caesura appears between the moment of cadential arrival and the initiation of the transition.

[64]Caplin reserves cadential status for a PAC formula that closes a *theme*, so he refers to a PAC ending a post-cadential phrase as a cadence of "limited scope" ("The Classical Cadence," 86–89). For the PAC ending the subordinate theme in the opening movement of K. 545, however, I was unable to conclude the excerpt before the appearance of such a cadence in the closing section, as both codettas feature an abbreviated PAC of limited scope (not shown). Visual inspection of the completion ratings did not reveal any significant differences between K. 545 and the remaining perfect authentic cadences, however.

(a)

(b)



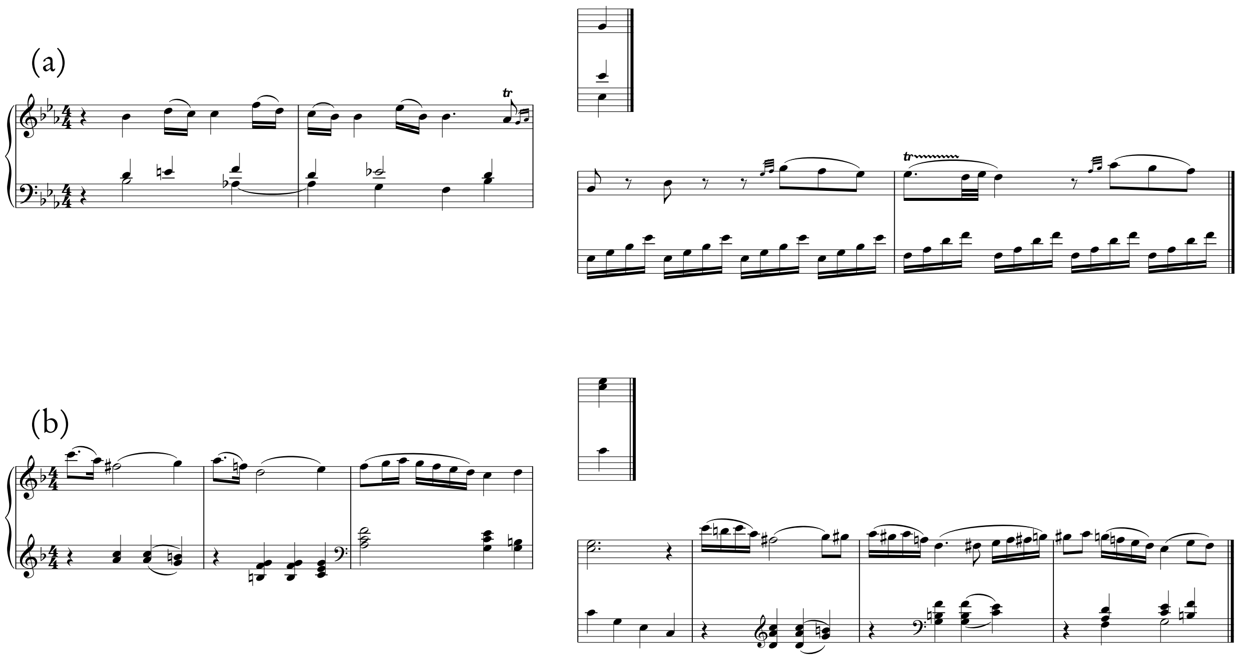
Example 7.3: PAC subtypes, with the moment of cadential arrival presented *out of context* above and the post-cadential passage presented *in context* below. (a) PAC→Transition subtype: K. 333, ii, mm. 5–10; (b) PAC→Closing Section subtype: K. 333, i, mm. 54–62.

as the cadence was not followed by an unambiguous consequent phrase featuring the return of the basic idea, but by a continuation phrase. Shown in Example 7.5, the cadence in m. 4 consists of a Prinner IAC ending the opening four-measure antecedent of a sixteen-measure compound period.[65] The continuation phrase follows the IAC in mm. 5–8, which suggests the excerpt should be categorized in the Middle/End contextual subtype. But unlike the excerpts selected for the Middle/End subtype, the continuation phrase in Example 7.5 does not feature a varied repetition of prior continuation or cadential material following the moment of cadential arrival (cf. Example 7.4b). What is more, the cadential idea consists of a stepwise descending melody that resolves to $\hat{3}$, a characteristic found frequently in *genuine* IACs ending a phrase or theme, whereas the cadential melody for the excerpts from the Middle/End subtype often features stepwise ascending motion from $\hat{2}$ to $\hat{3}$ at the cadential arrival, suggesting a kind of melodic deception that is more consistent with the cadential deviations found in subordinate themes.[66] Finally, a caesura and a melodic lead-in follow the cadential arrival in m. 4, and Michel Vallières has noted the degree to which listeners interpret the unaccompanied melodic lead-in as a prominent signal of a new beginning.[67] For these reasons, I elected to categorize this excerpt in the IAC→Beginning subtype in order to differentiate imperfect authentic cadences found in compound periods within main themes from those found in the continuation and cadential material located in subordinate themes.

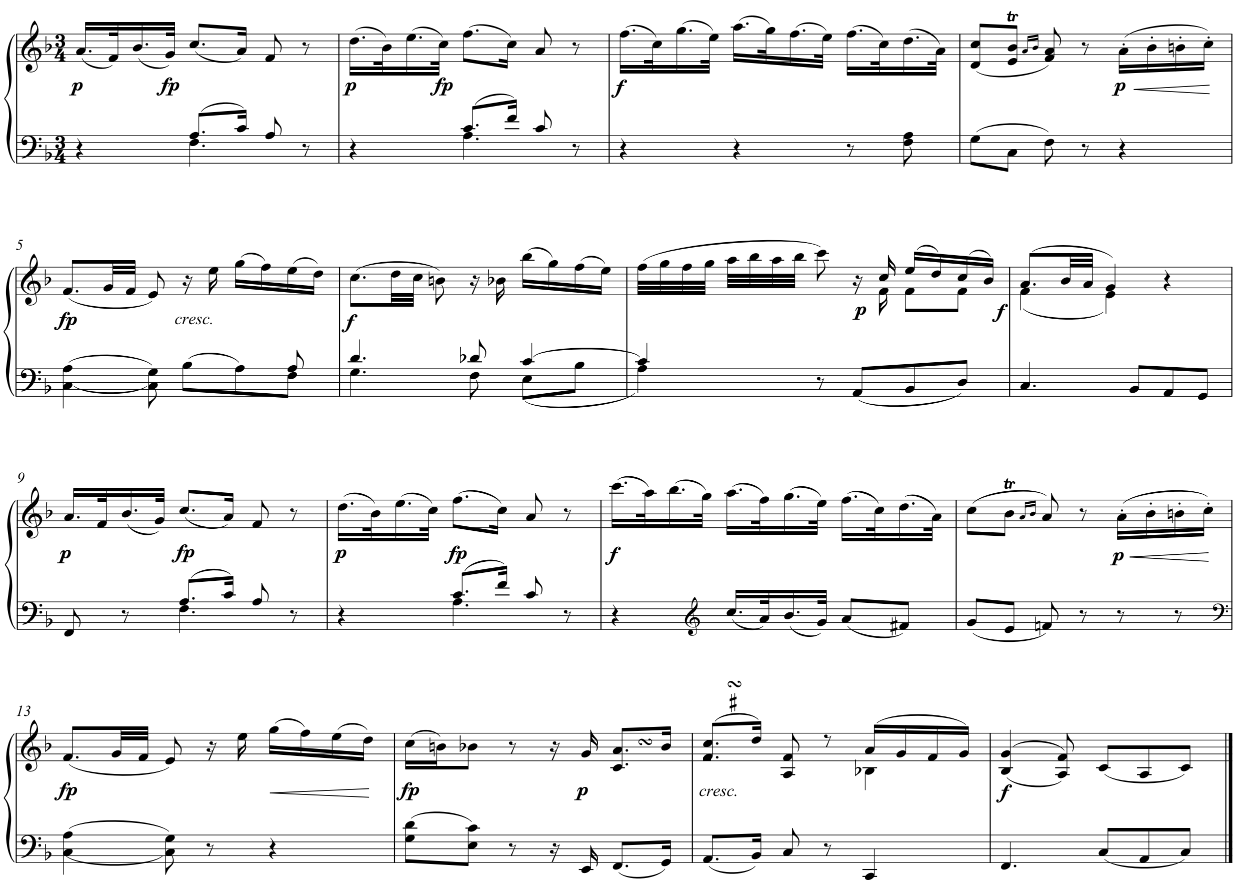
The half cadence category was subdivided according to whether the moment of cadential arrival is followed by a consequent phrase or by a prominent caesura and new thematic region.

[65] Caplin has recently examined a particular variant of Robert Gjerdingen's Prinner schema that also functions in certain contexts as an imperfect authentic cadence. In Gjerdingen's theory, the Prinner consists of a two-voice framework, with each voice presenting a descending scalar tetrachord: from $\hat{6}$ to $\hat{3}$ in the soprano, and from $\hat{4}$ to $\hat{1}$ in the bass. Caplin points out, however, that by inserting $\hat{5}$ between the final two notes of the bass tetrachord, the Prinner takes on a distinctly cadential function, as in Examples 7.4a and 7.5. For a discussion of the "Prinner cadence," see Caplin, "Harmony and Cadence in Gjerdingen's "Prinner"."

[66] Neuwirth, "Fuggir la cadenza, or the Art of Avoiding Cadential Closure," 127-130.

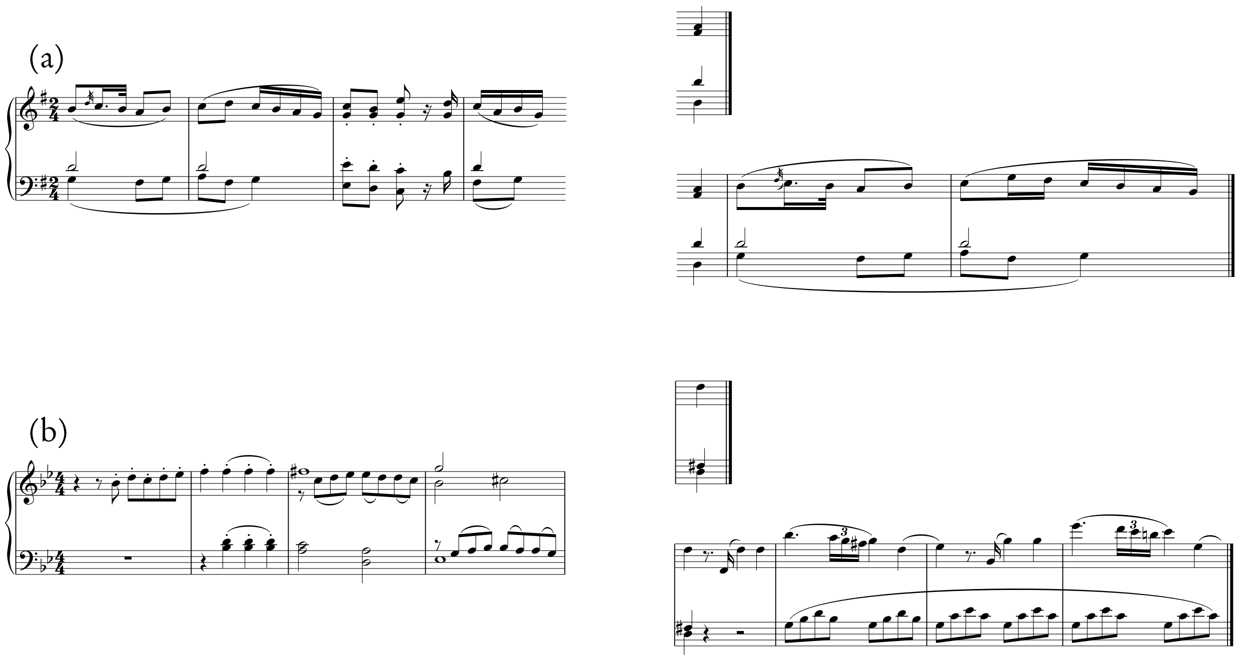
[67] Vallières et al., "Perception of Intrinsic Formal Functionality," 28.

Example 7.4: IAC subtypes, with the moment of cadential arrival presented *out of context* above and the post-cadential passage presented *in context* below. (a) IAC→Beginning subtype: K. 282, i, mm. 2–5; (b) IAC→Middle/End subtype: K. 533, iii, mm. 23–29.

Although a post-cadential prolongation of dominant harmony often follows half cadences in the classical style, I elected not to include those cases but rather to examine two subtypes that feature new beginnings following the cadential arrival in order to examine the degree to which parameters effecting segmental grouping might reinforce the perception of closure.[68] Compared to half cadences ending an antecedent phrase, half cadences preceding a prominent caesura and the beginning of a new thematic region tend to close material residing at a higher level of the formal hierarchy. As a consequence, we might expect these half cadences to receive higher completion ratings relative to those ending an antecedent. Example 7.6a features a *Simple* half cadence ending an antecedent phrase, with the consequent phrase beginning in the subsequent

[68]I also considered including another contextual subtype for the half cadence category that would examine half cadences for which parameters related to segmental grouping do not reinforce the harmonic and melodic arrival, but the duration of the experimental session at 90 minutes precluded the possibility of including further subtypes.

Example 7.5: Mozart, Piano Sonata No. 7, K. 309, ii, mm. 1–16. The example presented *in context* consists of mm. 1–6.

measure, while Example 7.6b presents an *Expanding* half cadence with *Do-Fi-Sol* melody (see §5.4). The presence of a root-position V^7 two measures before the cadential arrival is also reminiscent of Nathan Martin and Julie Pedneault-Deslauriers' *Doppia* half cadence type,[69] though the bass moves up in the subsequent measure to E♭, thus replacing the expected vii^7/V sonority with a German augmented sixth chord.

As I mentioned in Experiment I, these contextual subtypes still fail to consider the contri-

[69] Martin and Pedneault-Deslauriers, "The Mozartean Half Cadence," 193–196. In their formulation, the *Doppia* half cadence consists of the harmonic progression, V^7–vii^7/V–V, and the stepwise melodic descent, $\hat{4}$–$\flat\hat{3}$–$\hat{2}$. By retaining the *Do-Fi-Sol* melody in Example 7.6b, Haydn was forced to abandon the *Doppia* progression so as to prevent doubling the applied leading tone in the soprano and bass.

Example 7.6: HC subtypes, with the moment of cadential arrival presented *out of context* above and the post-cadential passage presented *in context* below. (a) HC→Consequent subtype: K. 311, ii, mm. 1–6; (b) HC→MC/ST subtype: K. 333, iii, mm. 60–67.

bution of a variety of other retrospective parameters that may affect the perception of closure. Thus, in addition to these contextual subtypes, I have also characterized each excerpt in §7.3.3 according to parameters related to segmental grouping and formal function to examine the role they might play in a correlational model.

7.3.1 Method

Participants

Participants were 30 members (17 female) of the Montreal community recruited through the Schulich School of Music and the McGill University classified ads. Ages ranged from 18 to 33 ($M = 22$, $SD = 4$). Fifteen participants with music training equivalent or superior to second-year-university level formed the musician group, and fifteen participants with less than one year of music training comprised the nonmusician group. To limit any effects caused by familiarity with the stimuli, no participant with more than two years of formal study on the piano was permitted to take part.

A questionnaire was administered to assess musical preferences and training. On average, musicians had 10.6 years of study on a musical instrument (other than piano), 3.5 years of ear training, 2.9 years of instruction in harmony, and 2.7 years of instruction in music analysis. At the time of their participation, they additionally reported spending an average of 22.9 hours each week engaged in instrumental practice. Participants also listened to an average of 19 hours of music each week. All of the participants reported normal hearing, which was confirmed with a standard audiogram administered before the experiment, and three participants indicated they had absolute pitch.

Materials

The stimuli consisted of 40 excerpts selected from Mozart's keyboard sonatas that contained an equal number of perfect authentic, imperfect authentic, half, deceptive, and evaded cadences (8 each). For the purpose of the experiment, two versions of each of the 40 excerpts were created: an *out of context* version that did not include any material following cadential arrival (M = 8.5 s, SD = 2.6 s), and an *in context* version that included on average an additional 7.2 s of post-cadential material (M = 15.7 s, SD = 4.1 s).

Just as in Experiment I, performance features (such as dynamics and rubato) were neutralized and the tempo of each excerpt was determined by convention. To ensure that unwanted differences at cadential arrival would not affect completion ratings, the duration of the event(s) at cadential arrival were recomposed to 900 ms (onset to offset) and any melodic dissonances at cadential arrival were removed. Each stimulus was first created with the notation software Sibelius and then realized as a .wav sound file at a sampling rate of 44.1 kHz and 16-bit resolution

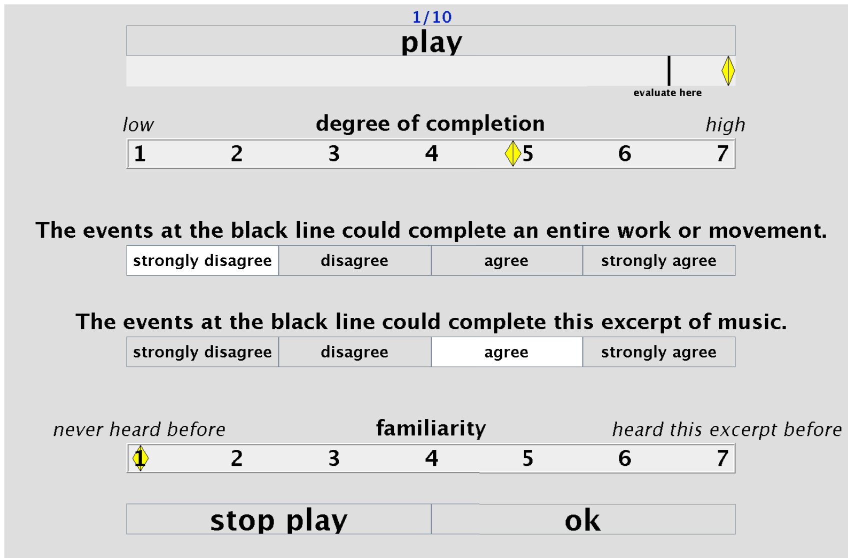


Figure 7.5: Screen shot of the interface used in Experiment II.

using a piano physical model created by PianoTeq (Modartt S.A.S., Ramonville Saint Agne). Finally, to encourage participants to attend specifically to the moment of cadential arrival, a 1-s fade-in was inserted at the beginning of each excerpt, and a 2-s fade-out was inserted at the end of each excerpt presented *in context*.

Design and Procedure

The design for Experiment II was nearly identical to that of Experiment I, except that in Experiment II the stimuli were presented at a constant level of 60 db SPL across the experimental session, which was measured with a Brüel Kjær Type 2205 sound-level meter (A-weighting) placed at the level of the listener's ears.

The experimental task was also nearly identical to Experiment I, with the exception that for excerpts presented *in context* the moment of cadential arrival did not represent the final harmonic and melodic events of each excerpt. As a result, it was necessary to provide a visual cue to alert participants to the moment of cadential arrival (see Figure 7.5). In each trial a playback cursor was provided at the top of the screen that followed along with the excerpt, and a black vertical line was placed along the playback bar to mark the onset of cadential arrival. After listening to each excerpt at least two times, participants were instructed to provide *completion*, *confidence*, and *familiarity* ratings for the event(s) marked by the black vertical line, as well as respond to two separate statements regarding whether the event(s) at the black line could complete (1) an entire work or movement, or (2) a phrase or short passage within that movement.

The 40 excerpts were presented *in context* and *out of context* in two blocks, with participants rating the excerpts presented *out of context* first. In the first block, participants were given identical instructions to those found in Experiment I, but in the second block, the participants were additionally instructed to consider the material following the black line when making their ratings. To familiarize the participants both with the range of stimuli as well as with the experimental task, the session began with an exposure phase and a practice phase consisting of 10 additional excerpts preceding the first block, and another practice phase consisting of 5 additional excerpts preceding the second block. After completing the experiment, participants filled out a short questionnaire addressing their music background.

7.3.2 Results

Cadence Categories

Figure 7.6 presents bar plots of the completion ratings for each of the five cadence categories. When the excerpts were presented *out of context* (top left), the ratings for both groups replicated the results from Experiment 1 (see Figure 7.2). A mixed 5×2 ANOVA revealed a main effect of cadence category, $F(3.13, 87.51) = 166.72$, $\varepsilon = 0.78$, $p < .001$, $\eta^2 = .84$, and a significant interaction, $F(3.13, 87.51) = 4.12$, $p < .01$, $\eta^2 = .02$, but the main effect of training was not significant. A polynomial contrast of the cadence categories also identified the same descending linear trend from the PAC to EV categories that was observed in Experiment I, $F(1,28) = 367.31$, $p < .001$, $\eta^2 = .91$. For the musician group, post hoc analysis revealed significant differences between each adjacent pair of cadence categories ($p < .01$), with the exception of the DC-EV pair. Musicians and nonmusicians also did not differ in their ratings of the genuine cadence categories, nor did post hoc tests reveal significant differences between the HC, DC, and EV categories for the ratings of the nonmusician group. In Experiment 1, excerpts from the DC category also elicited significantly higher ratings from nonmusicians relative to those from the HC category. In the context of Experiment II, the same trend also emerged, but this difference was not significant ($p > .05$).

When the excerpts were presented *in context* (top right), however, nonmusicians' ratings demonstrated the same descending linear trend that was observed in the ratings of the musician group from Experiment I and in both of the context conditions from Experiment II. To consider how the completion ratings changed when the excerpts were heard *in context*, the ratings made for excerpts heard *out of context* were subtracted from the corresponding ratings made for excerpts heard *in context*. Shown in the bottom plot of Figure 7.6, the solid blue bar in the PAC category indicates that musicians on average rated perfect authentic cadences heard *in context* nearly one unit lower on the completion scale.[70] A 5×2 ANOVA of the change in completion ratings revealed a main effect of cadence category, $F(2.88, 80.66) = 22.45$, $\varepsilon = 0.72$, $p < .001$, $\eta^2 = .44$, but there was no effect of training, and no interaction. A polynomial contrast of the cadence categories also revealed a significant quadratic trend from the PAC to EV categories, $F(1,28) = 51.79$, $p < .001$, $\eta^2 = .64$, with the observed means in both groups exhibiting an upside down U-shape from the outer cadence categories (PAC and EV) to the inner category

[70] As a general rule of thumb, if the error bar for a given bar does not cross the x-axis, the difference between the completion ratings for excerpts presented *in context* compared to *out of context* is significant.

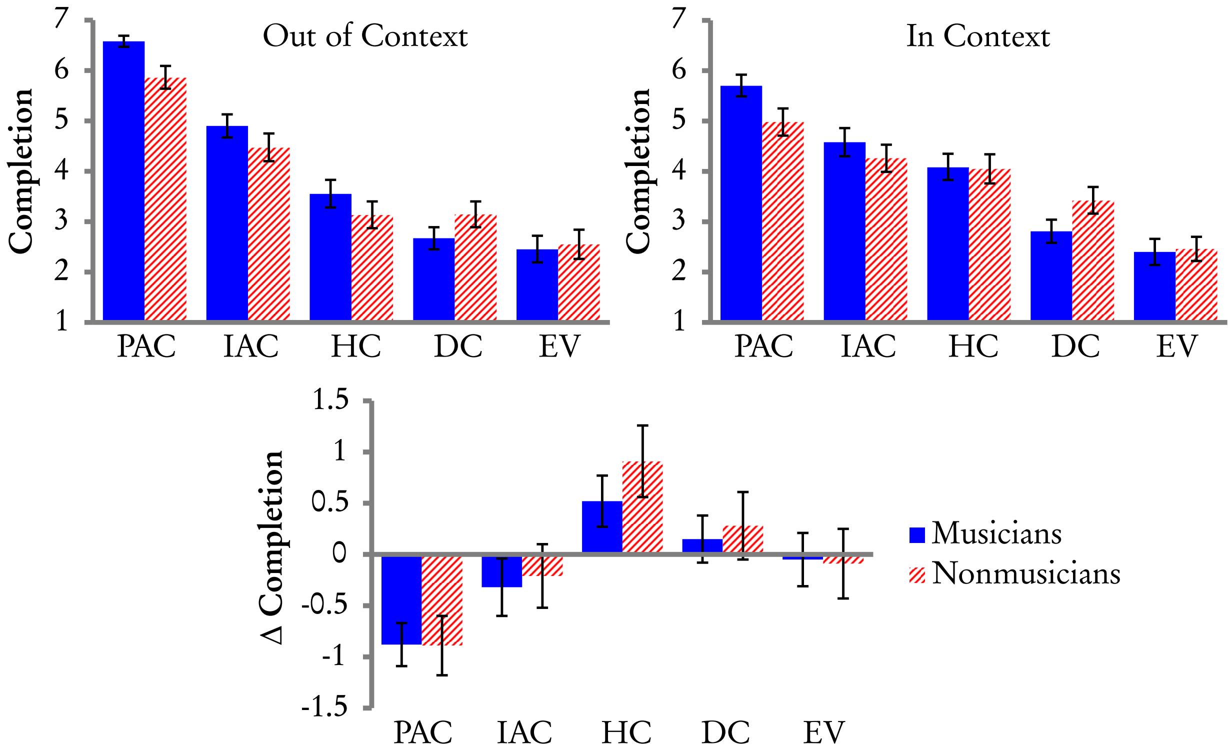


Figure 7.6: Top: Bar plots of the mean completion ratings for the *out-of-context* (left) and *in-context* (right) conditions for each cadential category, with musicians in solid blue and nonmusicians in diagonal red. Bottom: Bar plot of the mean difference in the completion ratings between the *out-of-context* and *in-context* conditions. Whiskers represent the 95% confidence interval.

(HC). A 2×2 ANOVA comparing the completion ratings for the two context conditions of the PAC category indicated that both groups provided lower ratings for excerpts heard *in context*, $F(1, 28) = 27.23$, $p < .001$, $\eta^2 = .49$, and musicians also provided generally higher ratings than nonmusicians on average, $F(1, 28) = 7.68$, $p = .01$, $\eta^2 = .22$. The participants also tended to provide lower ratings for excerpts from the IAC category when they were presented *in context*, regardless of training, but this tendency did not reach significance, $F(1, 28) = 2.55$, $p > .05$. For the HC category, both groups rated excerpts heard *in context* as more complete than those heard *out of context*, $F(1, 28) = 19.16$, $p < .001$, $\eta^2 = .39$. The context condition did not significantly influence completion ratings for the remaining categories.

In the previous section, I suggested that musicians may apply a model of genuine cadential closure even to excerpts presented out of context as a result either of extensive exposure or explicit training, whereas nonmusicians must rely to a greater extent on the material following cadential arrival to determine the perceived strength of the cadential ending. The crucial point here is that the cadential status of any potential ending is at least partially dependent on the material *following* the moment of cadential arrival, where parameters related to segmental grouping will likely play a greater role. For the half cadence, the results presented in Figure 7.6 provide evidence in support of that hypothesis, since half cadences are characterized by the appearance of various metrical, textural, and rhythmic devices after the onset of cadential arrival that serve to reinforce an otherwise active and unstable sonority. And indeed, when presented *out of context*, as in Experiment I and in the corresponding condition of Experiment II, the nonmusician group implicitly positioned the half cadence near the bottom of the cadential hierarchy. Yet when these excerpts were presented *in context*, the nonmusician group positioned the half cadence somewhere in the middle, thereby replicating the cadential hierarchy demonstrated in the musician completion ratings from Experiment I.

For the PAC category, however, both groups provided significantly lower ratings when they heard the excerpts *in context*. The apparent contradiction embedded within the task—determining whether the events at cadential arrival *could* end the excerpt "without the need for anything further" when more material *did* always follow—forced participants to choose between prospective and retrospective vantage points. Faced with such overwhelming retrospective evidence that the cadential arrival failed to serve as the excerpt's literal end, both groups elected to lower the completion ratings for excerpts from the PAC category and reserve the very top of the scale for passages which feature silence following the events at cadential arrival.

Still, if the context surrounding the cadential arrival plays a significant role in determining the strength of the ending, one would expect that excerpts from the EV category presented *in context* would receive significantly lower ratings than those presented *out of context*, since the latter excerpts impose an artificial boundary at the cadential arrival that might elicit higher ratings from participants. As mentioned previously, the EV category is characterized by the appearance of events at the precise moment of cadential arrival that interrupt the perceived grouping structure and initiate the subsequent phrase. What is more, the excerpts selected for the EV category in Experiment II "back up" and re-initiate the preceding continuation/cadential material so as to attempt the cadence "one more time," thus providing the listener with yet clearer retrospective evidence that the expected cadence was not realized. Presumably in such instances listeners with access to the surrounding context would provide significantly lower completion ratings relative to those without such access, but these findings do not justify that assumption. We may only conjecture from the results presented here that the position of the half cadence within the general cadential hierarchy is dependent on parameters that both precede *and* follow cadential arrival, whereas for the remaining categories the role played by parameters that appear after the events at cadential arrival is less clearly defined.

Contextual Subtypes

Figure 7.7 presents bar plots of the change in completion ratings observed for the contextual subtypes of the genuine cadence categories. Beginning with the PAC category, a $2 \times 2 \times 2$ ANOVA of the completion ratings with within-participant factors of context (in context, out of context), subtype (transition, closing section), and training (musicians, nonmusicians) revealed a significant effect of subtype, $F(1, 28) = 8.84$, $p < .01$, $\eta^2 = .24$. As in Experiment I, both groups provided higher ratings for excerpts selected from the subordinate theme compared to those selected from the main theme (cf. Figure 7.3). Shown in the top left plot in Figure 7.7, however, the particular subtype that followed the category did not elicit significantly different

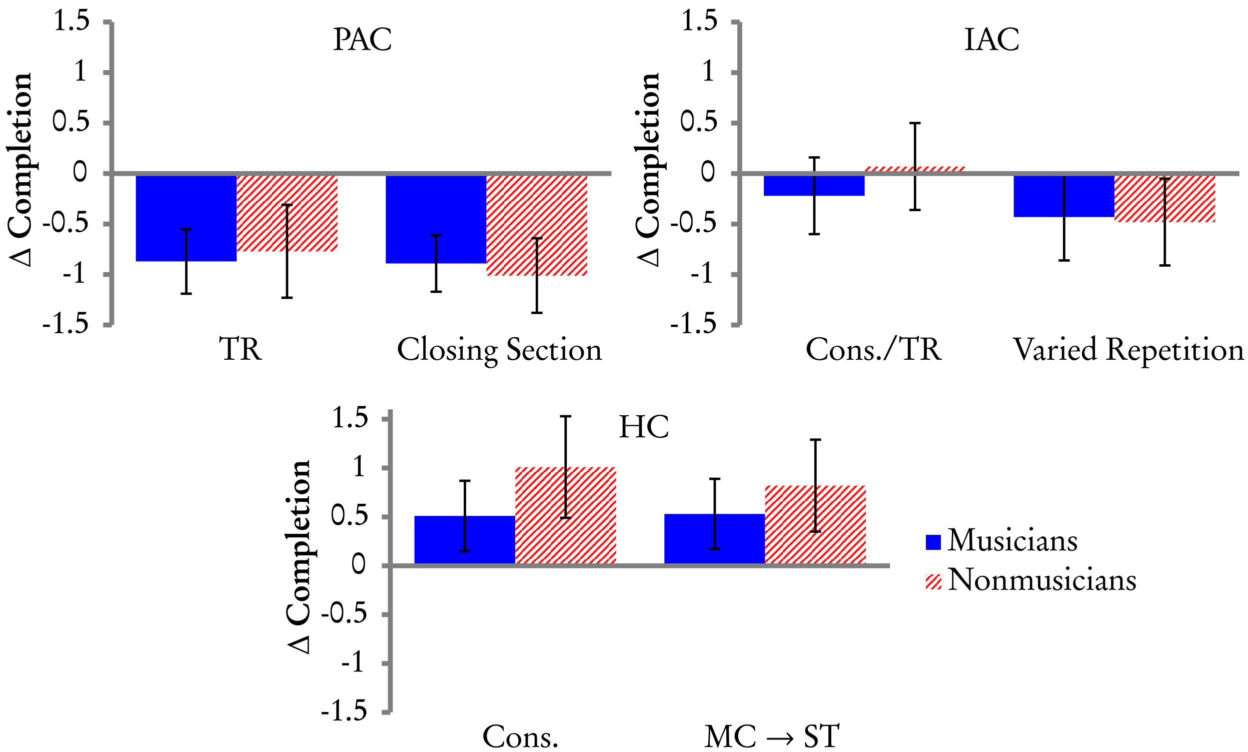


Figure 7.7: Top: Bar plots of the mean difference in the completion ratings between the *out-of-context* and *in-context* conditions for the genuine cadence categories, with musicians in solid blue and nonmusicians in diagonal red. Whiskers represent the 95% confidence interval.

completion ratings for excerpts heard *in context* compared to *out of context*. Instead, both groups provided significantly lower completion ratings for excerpts heard *in context*, regardless of subtype.

For the IAC category, musicians and nonmusicians provided higher completion ratings for excerpts from the main theme subtype that were followed either by a consequent phrase (i.e., in a period or period hybrid) or by the beginning of the transition, $F(1, 28) = 45.70$, $p < .001$, $\eta^2 = .61$. Moreover, excerpts from the subordinate theme that were followed by a varied repetition of previous continuation or cadential material received significantly lower ratings when they were heard *in context* compared to *out of context*, $F(1, 28) = 5.53$, $p < .05$, $\eta^2 = .16$, providing further evidence that the perceived strength of a cadence—in this case an imperfect authentic cadence—may be influenced by the surrounding context.

Finally, both groups rated half cadences selected from the main theme that were followed by a medial caesura and the beginning of the transition as more complete than half cadences followed by a consequent phrase within the main theme, $F(1, 28) = 6.08$, $p < .05$, $\eta^2 = .16$. This effect was also more pronounced in nonmusicians, $F(1, 28) = 4.77$, $p < .05$, $\eta^2 = .12$, which again suggests segmental grouping may play a more prominent role for nonmusicians in judgments of completion.

Movement Completion and Phrase Completion Ratings

Just as in Experiment I, the movement ratings provided few notable results, so the following analysis attends specifically to the phrase completion ratings for excerpts from the HC category. Figure 7.8 provides a bar plot of the distribution of the percentage of responses for excerpts from the half cadence category, with ratings for excerpts heard *out of context* above and *in context* below the x-axis. When the excerpts were presented *out of context*, both groups wavered between general agreement (47%) and general disagreement (53%), but when the excerpts were presented *in context*, in nearly 68% of their responses the participants generally *agreed* with the statement that the half cadence could complete a phrase or short passage of music, and this difference was significant, $U = 20,850$, $p < .001$, $r = -.25$.

To visualize the difference between the two training groups more clearly, Figure 7.9 presents the degree of change in the Likert scale ratings when the excerpts were heard *in context*, with musicians' ratings above the x-axis in solid blue and nonmusicians' below in diagonal red. The x-axis indicates whether the ratings for excerpts from the HC category remained the same (0), increased (>0), or decreased (<0) when they were heard *in context*. The plot visualizes the earlier finding that both groups very rarely provided comparatively lower ratings when the excerpts were presented *in context*, instead exhibiting a tendency to provide higher ratings.

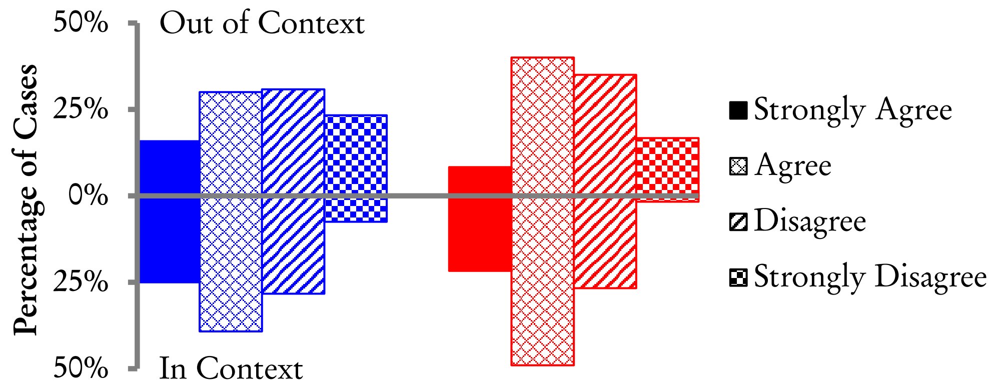


Figure 7.8: Bar plot of the distribution of the percentage of responses for the half cadence category for the statement, "this excerpt could complete a phrase or short passage of music," with excerpts rated *out of context* appearing above the x-axis and excerpts rated *in context* appearing below. The ratings of musicians and nonmusicians appear in blue (on the left) and red (on the right), respectively. Pattern fills denote response types.

7.3.3 Modeling the Change in Completion Ratings

Taken together, these results support an increasingly temporal view of cadential perception, whereby the events at and following cadential arrival may realize, revise, or indeed, entirely contradict schematic expectations for cadential closure. In this retrospective stage, segmental grouping and form-functional identification inform previous interpretations of cadential closure, challenging initial impressions in some instances, reinforcing them in others. The half cadence category is exemplary in this regard. In the context of Experiment I, nonmusicians rated half cadences near the bottom of the cadential hierarchy, but in Experiment II, both groups provided significantly higher completion ratings for half cadence excerpts presented *in context*. What is more, nonmusicians provided lower ratings for imperfect authentic cadences from the subordinate theme that were followed by a varied repetition of the continuation and cadential material, suggesting the temporal function of the passage following cadential arrival may also influence the perception of closure. But whereas parameters in the retrospective stage apparently play an important role in the perception of cadential closure for nonmusicians, it is noteworthy

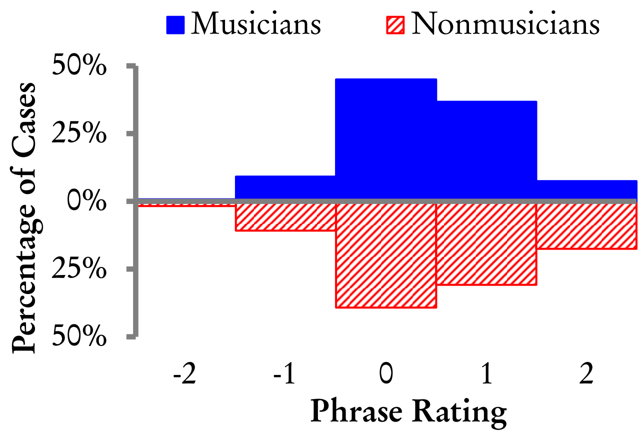


Figure 7.9: Bar plot of the distribution of the change in responses for the half cadence category for the statement, "this excerpt could complete a phrase or short passage of music," with musicians' ratings above the x-axis in solid blue and nonmusicians' ratings below in diagonal red.

that few such effects emerged in the musician ratings. To be sure, if prediction is indeed the "primary thing" the brain does, as Jeff Hawkins and Sandra Blakeslee have recently suggested,[71] perhaps long-term stylistic knowledge acquired by passive exposure or explicit training biases musicians toward increasingly prospective processing during music listening.

Nevertheless, the preceding analysis takes a rather coarse-grained approach to the examination of retrospective parameters on the perception of cadential closure, in which I attempted to distill a complex of co-varying features characterizing the events at and following cadential arrival—including texture, dynamics, rhythmic duration, metric position, surface rhythmic activity, register, and so on—into a few general subtypes to examine the post-cadential formal functions that appear frequently following cadences. The selection of these cadential categories and contextual subtypes therefore presents a rather incomplete picture of retrospective processing. Thus, as in Experiment I, we may benefit from a correlational approach to the change in completion ratings, one that seeks to identify those parameters that explain changes in the

[71] Hawkins and Blakeslee, *On Intelligence*, 89.

completion ratings without regard for the contextual subtypes I initially selected.

To consider the role these retrospective parameters might play, I selected 11 features following cadential arrival that characterize (1) the presence of a segment boundary, and (2) the post-cadential passage (see Table 7.9).

1. **Segment Boundary**. Four continuous features, one dichotomous feature, and one ordinal feature characterize segment boundaries: the next note onset following the onset of cadential arrival in seconds (*Next Note Onset*), the next note onset that appears in the bass voice in seconds (*Next Bass Note Onset*), the next note onset that appears in the soprano voice in seconds (*Next Soprano Note Onset*), the number of notes appearing in a 1.5 s window beginning at the onset of cadential arrival (*Event Density*), the presence of a rest in all four instrumental parts (*Caesura*), and the superposition of a new intrathematic region, an accompanimental overlap in the bass, or a melodic lead-in at the onset of cadential arrival (*Elision*).[72]

2. **Post-Cadential Passage**. Two categorical, one dichotomous, and two continuous features characterize the passage following cadential arrival: the formal function of the subsequent passage at the theme level, identified as *before-the-beginning*, *beginning*, *middle*, *end*, or *after-the-end* (*Interthematic Function*), the formal function of the subsequent passage at the phrase level, identified as *beginning*, *middle*, or *end* (*Intrathematic Function*), the presence of a varied repetition of material from the continuation and cadential progression preceding the onset of cadential arrival (*Repetition*), the duration of the stimulus in seconds (*Stimulus Duration*), and the duration of the stimulus from the onset of cadential arrival to the end of the excerpt in seconds (*Stimulus Duration from CA*).

[72]The excerpts selected for Experiment II did not feature interthematic elision (e.g., an elision of the main theme with the transition).

Table 7.9: Descriptive statistics for the 11 retrospective features.

Retrospective Features	*M* (*SD*)	*Range*	*Mode* (*Frequency*)
Segment Boundary			
(1) *Next Note Onset* (s)	.57 (.44)	.1–1.8	
(2) *Next Bass Note Onset (s)*	.97 (.77)	.1–2.8	
(3) *Next Soprano Note Onset* (s)	.99 (.66)	.1–2.6	
(4) *Event Density*[a]	9.98 (5.25)	3–19	
(5) *Caesura*[b]			Absent (28)
(6) *Elision*[c]			None (26)
Post-Cadential Passage			
(7) *Interthematic Function*[d]			End (20)
(8) *Intrathematic Function*[e]			Beginning (19)
(9) *Repetition*			Present (25)
(10) *Stimulus Duration* (s)	15.68 (4.14)	9.2–27.6	
(11) *Stimulus Duration from CA* (s)	7.20 (2.15)	3.6–14.1	

[a] *Event Density* (1.5) refers to the number of notes identified in a 1.5 s window starting from the onset of cadential arrival.
[b] *Caesura* refers to the presence of a rest across all four instrumental parts.
[c] *Elision* refers to the superposition of a new intrathematic phrase at the moment of cadential arrival, an accompanimental overlap in the bass, or a melodic lead-in.
[d] *Interthematic Function* refers to one of the following temporal functions to characterize the passage at the theme level following cadential arrival: Before-the-Beginning, Beginning, Middle, End, After-the-End.
[e] *Intrathematic Function* refers to either the Beginning, Middle, or End functions that characterize the passage at the phrase level following cadential arrival.

Of the features presented above, *Elision* requires further explanation. This feature characterizes the superposition of initiating material at cadential arrival on an ordinal scale, where an intrathematic elision (e.g., a cadential phrase with a post-cadential standing on the dominant) constitutes the clearest case of form-functional elision in this study, thereby receiving the highest ranking on the *Elision* scale. In many instances, however, an accompanimental bass pattern in the left hand coincides with the moment of cadential arrival, but the moment of structural beginning appears instead in the following measure. For this reason, the sense of elision is weakened by the absence of initiating melodic-motivic material at the moment of cadential arrival. Finally, a few of the excerpts selected for this study feature an unaccompanied melodic lead-in beginning at the cadential arrival, but the absence of material in the lower voices results in a covered caesura, thereby weakening the impression of elision still further. Stimuli that did not include any of these features received the lowest score.

Correlations were calculated for each of the retrospective features with the change in completion ratings from both groups. Shown on the left in Table 7.10, the musician ratings were significantly correlated with four retrospective features—*Next Note Onset*, *Next Bass Note Onset*, *Caesura*, and *Elision*—and the nonmusician ratings were correlated with seven retrospective features—*Next Note Onset*, *Next Bass Note Onset*, *Event Density*, *Caesura*, *Elision*, *Interthematic Function*, and *Stimulus Duration*. After controlling for cadence category membership, three further correlated features emerged for the musician ratings—*Next Soprano Note Onset*, *Interthematic Function*, and *Intrathematic Function*—and *Stimulus Duration* was replaced by *Intrathematic Function* for the nonmusician ratings.

Significant correlations for *Next Note Onset*, *Next Bass Note Onset*, and *Next Soprano Note Onset*—both before and after controlling for cadence category membership—suggest changes in the completion ratings resulted in part from surface activity at cadential arrival, where excerpts characterized by longer durations between the onset of cadential arrival and the next note onset received positive changes in completion ratings when they were presented *in context*. After controlling for cadence category, features characterizing the temporal function of the passage following cadential arrival also displayed significant correlations with the change in completion ratings, where excerpts that were followed either by interthematic or intrathematic beginnings received positive changes in completion ratings.

Despite the apparent significance of these correlations, however, many of the retrospective features identified for this study were themselves also highly correlated. In regression models, collinearity amongst model predictors violates a basic assumption of linear regression, since it

Table 7.10: Left: Correlations of retrospective features with the change in completion ratings of musicians and nonmusicians. Right: Semi-partial correlations controlling for cadential category.

	Retrospective Features	r	*Retrospective Features*	sr^{a}
Musicians				
	Next Note Onset	.50**	Next Note Onset	.70***
	Next Bass Note Onset	.46**	Next Bass Note Onset	.47**
	Caesura	−.35*	Next Soprano Note Onset	.35*
	Elision	.47**	Caesura	−.39*
			Elision	.38*
			Interthematic Function	−.48**
			Intrathematic Function	−.41*
Nonmusicians				
	Next Note Onset	.58***	Next Note Onset	.72***
	Next Bass Note Onset	.53***	Next Bass Note Onset	.53**
	Event Density	−.37*	Event Density	−.42**
	Caesura	−.48**	Caesura	−.51**
	Elision	.43**	Elision	.35*
	Interthematic Function	−.40*	Interthematic Function	−.57***
	Stimulus Duration	−.33*	Intrathematic Function	−.59***

Note. $N = 40$.
[a] sr = semi-partial (or part) correlation.
* $p < .05$ ** $p < .01$ *** $p < .001$.

would be impossible to determine which of two highly correlated predictors in a regression model—say *Next Note Onset* and *Next Bass Note Onset*—is actually responsible for variations in the dependent variable. Given the multicollinearity demonstrated in the feature set, as well as the small sample size of the stimulus set, a regression model embracing multiple predictors would thus not be advisable in the context of Experiment II.

But then again, this finding is not particularly surprising given that the very idea of *parametric congruence* is essentially synonymous with characterizing closure as a point of arrival exhibiting a high degree of statistical covariation. To be sure, scholars often articulate closure as a convergence of features interacting at various levels of the grouping hierarchy to achieve moments of rest, relaxation, or repose. In this sense, segment boundaries at relatively local levels *should* be marked by clusters of covariation between a number of parameters, and the intercorrelations shown in Table 7.9 support this view. To examine these issues yet further, however, isolating the contribution of each parameter and determining its degree of covariation with the other parameters that effect closure, would entail a careful recomposition of the selected stimuli to examine surface rhythmic activity, elision, and formal function, issues that extend far beyond the results obtained in this experiment.

§7.4 Conclusions

The goal of this chapter was to explore the underlying mechanisms responsible for the perception of cadential closure in Mozart's keyboard sonatas using an explicit rating task. The findings from Experiment I indicate that, regardless of training, listeners appear to differentiate among the categories of genuine cadences (PAC, IAC, HC). The harmonic and melodic content at the cadential arrival (i.e., the syntactic parameters of tonal music) therefore play a pivotal role in the perception of closure. Moreover, both groups provided higher completion, confidence, and familiarity ratings for the PAC category than for the other categories, providing converging evidence in support of the claim that listeners familiar with Western music possess a schematic representation for authentic cadential closure.[73]

From the inclusion of subtypes within each category, a number of conclusions may be drawn. First, the formal context of the PAC category significantly affected the perception of completion, regardless of music training, with excerpts drawn from subordinate themes receiving higher completion ratings. As I noted in Figure 7.1, perfect authentic cadences ending

[73] Eberlein, "A Method of Analysing Harmony, based on Interval Patterns or "Gestalten?"; Eberlein and Fricke, *Kadenzwahrnehmung und Kadenzgeschichte: ein Beitrag zu einer Grammatik der Musik*; Gjerdingen, *Music in the Galant style*; Rosner and Narmour, "Harmonic Closure"; John A. Sloboda, *The Musical Mind: The Cognitive Psychology of Music* (Oxford: Oxford University Press, 1985).

subordinate themes exhibit a number of unique characteristics that might explain this result: the expanded duration of the cadential progression, the increased surface activity (usually in the form of an accompanimental Alberti bass), and the appearance of a melodic trill just prior to the cadential arrival. These results also support Vallières' claim that the sudden decrease in surface activity at the cadential arrival may affect the perception of closure.[74]

Second, several results from the completion ratings in Experiment I for subtypes of the IAC and DC categories suggest that nonmusicians attend predominantly to the melody when assessing the completion of a given excerpt. First, they provided much higher completion ratings than did musicians for deceptive cadences. Second, they were more sensitive to the presence of a surface dissonance in the melody at the cadential arrival, as evidenced by their lower ratings for that subtype of the imperfect authentic cadence category. Finally, differences in the melodic scale degree in the deceptive cadence category significantly affected the ratings of nonmusicians, with melodies featuring $\hat{1}$ receiving higher ratings than those featuring $\hat{3}$, a result that was not replicated in the ratings of the musician group.

Contrary to the nonmusician group, musicians appeared to be much more sensitive to variations in harmony at the cadential arrival, as they provided much lower completion ratings for deceptive cadences than did nonmusicians. What is more, the harmony at the cadential arrival in the evaded cadence category also significantly affected musicians' ratings. Thus, the observed results might suggest a difference in attending strategies, with melody playing a more prominent role for nonmusicians, harmony a more prominent role for musicians.

Indeed, the regression estimates for the rhetorical features and tonal stability values strengthen this claim, as both the *Melodic Dissonance* feature and the tonal stability values for the soprano voice played a more substantial role in the nonmusician models. This finding supports Weiser's

[74] Vallières, "Beginnings, Middles, and Ends: Perception of Intrinsic Formal Functionality in the Piano Sonatas of W. A. Mozart," 106.

claim that music training may modulate attention, whereby nonmusicians demonstrate an attentional bias toward the soprano voice and musicians appear to flexibly track between the soprano and bass voices.[75] There has been some empirical support for the claim that nonmusicians privilege parameters related to melodic motion, such as pitch proximity and contour, whereas musicians attend principally to harmonic factors, such as the interval size between two events.[76] Furthermore, as mentioned previously, Loui and Wessel have shown that these differences of attention might not be conscious.[77]

Given the emphasis placed on the bass voice in identifying and categorizing cadences in music theory, the musician estimates for the tonal stability values in the bass and soprano voices are therefore not unexpected. It remains unclear, however, whether attention to bass-line motion in cadential contexts reflects a flexible voice-tracking strategy promoted during explicit formal training (i.e., in a pedagogical context) or an attentional bias formed simply through implicit exposure to Western music, a distinction that requires further attention in the experimental literature.[78] What these results do suggest is that, when faced with an explicit completion task, musicians appear to privilege the bass voice, whereas nonmusicians appear to be more sensitive to subtle differences in the soprano voice.

However, the hypothesis that differences in the completion ratings for the failed cadence categories may result from differences in attending strategy does not explain the significantly lower familiarity ratings from the nonmusician group for half cadences, nor does it explain why nonmusicians generally disagreed with the statement that half cadences could complete a phrase or short passage of music. Indeed, the completion ratings also suggest a different ordinal ranking of the cadential strength of each category for the two groups. The musician group

[75] Weiser, "Rating Cadence Stability," 40–46.

[76] Vos and Pasveer, "Goodness Ratings of Melodic Openings and Closures."

[77] Loui and Wessel, "Harmonic Expectation and Affect in Western Music."

[78] Emmanuel Bigand, "More about the Musical Expertise of Musically Untrained Listeners," *Annals of the New York Academy of Sciences* 999 (2003): 304–312.

provided the highest completion ratings for the PAC category, followed by the IAC, HC, DC, and EV categories, and this ranking provides empirical support for what I have called, following Caplin,[79] the *Genuine Schemas* model (see §7.2.3, *Cadential Strength*). For the nonmusicians, however, the HC category did not receive significantly higher ratings than either of the DC or EV categories, and the *1-Schema* model provided the best fit to their ratings.

The present findings may therefore suggest that listeners aware of the classical style have learned to expect the material that typically follows these cadence categories. Differences in the completion ratings for the HC, DC, and EV categories may thus result from a limit in the experimental design. By imposing an artificial boundary at the end of the cadential arrival, the findings from Experiment I do not consider the degree to which the perception of closure may be affected by the material following the cadential arrival. And the results from Experiment II largely confirmed this hypothesis, as both groups provided significantly higher completion ratings for half cadences heard *in context*. What is more, the nonmusician completion ratings for the excerpts heard *in context* demonstrated the same descending linear trend found in the musician ratings in both experiments, which suggests that musical training improves prospective processing, with musicians' ratings demonstrating a genuine model of cadential closure even without hearing the selected cadences *in context*.

A number of issues remain underexplored in these studies, however. First, any attempt to model the perception of closure in tonal music must also account for the effect of rhetorical parameters, an approach that traditional definitions of cadence generally do not embrace. The appearance of a dissonance in the melody at cadential arrival, the presence of a cadential trill, the sudden decrease in event density at the cadential arrival, and the duration of the stimulus all significantly affected participants' ratings of completion and may serve as important cues to the impending cadence. Second, the prevailing view of cadential closure held in the

[79] Caplin, *Classical Form*, 43.

Formenlehre tradition may unnecessarily privilege harmony and melody at the expense of rhetorical features related to segmental grouping that reinforce or contradict the syntactic formula at the moment of cadential arrival. Indeed, the findings from Experiment II suggest that the perception of closure is at least partly influenced by the surrounding context, where the presence of continued surface activity at the cadential arrival, an accompanimental overlap, or an intrathematic elision could weaken the perceived strength of the preceding cadence. Given the exploratory nature of the statistical approach (i.e., a correlational design), however, any future investigations considering the relative importance of either rhetorical or retrospective features must necessarily adopt a more controlled experimental design. Additionally, by selecting excerpts from a stylistically narrow repertoire—Mozart's piano sonatas—the characteristics that define closure in these excerpts (and thus, the characteristics that may lead to the development of learned schemata), may be idiomatic to this composer and genre. Unfortunately, relatively little is currently known regarding the degree to which listeners internalize conventional closing patterns that appear in other style periods or genres (e.g., romantic, rock, jazz).

Perhaps most importantly, these experiments do not provide direct evidence in support of the claim that the perception of cadential closure results from the formation of expectations during music listening. To be sure, one essential goal in selecting the stimuli was to explore the effect of cadential failure on the perception of closure, an issue which has yet to be considered in the experimental literature. Techniques for cadential deviation are a prevalent feature of the classical style and serve an important formal and expressive function. Instances of cadential failure could therefore provide ideal stimuli for future studies adopting a priming paradigm, as cadential deviations represent a violation of expectation when the listener's expectations are highest. The next chapter therefore takes precisely this approach, examining the interaction between expectations and cadential closure in three experiments.

Chapter 8

Expecting Closing Schemas: Converging Methods

Contemplating the future consequences of present actions has a proud lineage among us primates, and is one of the secrets of what is still, by and large, the stunningly successful story of humans on Earth.

Carl Sagan

Over the past three decades, the resurgence of associationist conceptions of mental processing in experimental psychology—demonstrated by the emergence of theories like implicit learning, connectionism, and predictive coding—has placed the study of expectations front and center.[1] In the intellectual climate now prevalent, many scholars view the brain as a "prediction machine" that generates expectations about future events by forming associations between co-occurring attributes within the external environment.[2] Philosopher Andy Clark suggests, for example, that brains are "statistical sponges structured by individual learning and evolutionary inheritance so as to reflect and register relevant aspects of the causal structure of the world itself."[3] For

[1]Bar, "The proactive brain: Using analogies and associations to generate predictions."
[2]Clark, "Whatever next? Predictive brains, situated agents, and the future of cognitive science."
[3]Ibid., 19.

Clark, it is precisely the brain's capacity to detect and remember such statistical regularities that explains how humans can navigate complex sensory environments from one moment to the next.

In Chapter 2 I suggested that music exploits this characteristic of the sensory-cognitive apparatus by organizing events on the musical surface to reflect the kinds of statistical regularities that listeners learn and remember. The classical cadence is one such example. According to the top-down, schema-theoretic view of memory described in Part I, I argued that listeners who are familiar with classical music have internalized the most common cadence categories as a flexible network of rival closing schemata. During music listening, the activation of this network in prospect results in the formation of expectations for the terminal events of the cadence. Thus, a theory of expectation should appeal to the study of cadences because it provides a direct, causal link between events on the musical surface and the schematic knowledge of listeners.

There is a good deal of support for the role played by expectancy in the perception of closure,[4] and numerous scholars besides myself have suggested that listeners might possess schematic representations for cadences and other recurrent closing patterns.[5] Yet there is currently very little experimental evidence justifying the link between expectancy and the variety of cadences in music of the classical style, or indeed, in tonal music more generally. This point is somewhat surprising given that the classical cadence is the quintessential compositional device for suppressing expectations for further continuation.[6] To be sure, the harmonic progression and melodic contrapuntal motion preceding the moment of cadential arrival elicit very definite expectations concerning the harmony, the melodic scale-degree, and the metric position of the

[4]Huron, *Sweet Anticipation*; Margulis, "Melodic Expectation"; Meyer, *Emotion and Meaning in Music*; Narmour, *The analysis and cognition of basic melodic structures*.

[5]Eberlein, "A Method of Analysing Harmony, based on Interval Patterns or "Gestalten?"; Eberlein and Fricke, *Kadenzwahrnehmung und Kadenzgeschichte: ein Beitrag zu einer Grammatik der Musik*; Gjerdingen, *A Classic Turn of Phrase*; Meyer, *Music, the arts, and ideas*; Rosner and Narmour, "Harmonic Closure"; David Temperley, *The Cognition of Basic Musical Structures* (Cambridge, MA: The MIT Press, 2004).

[6]Margulis, "Melodic Expectation," 263.

goal event.

Chapter 8 offers three experimental studies to examine whether the formation, fulfillment, and violation of expectations contribute to the perception of cadential closure during music listening. In §8.1 I review the evidence for melodic and harmonic expectations in experimental psychology and consider the roles played by long-term and short-term memory in the formation of expectations. §8.2 presents Experiment III, in which participants provided explicit retrospective expectancy ratings both before and after hearing the moment of cadential arrival for the cadential excerpts from Experiment II (see Table 7.8). To examine how these ratings might vary over time for each cadence category, §8.3 presents Experiment IV, in which participants provided continuous expectancy ratings on a slider while listening. To provide an implicit measure of expectancy in §8.4, Experiment V employs a reaction time task under the assumption that expected melodic and harmonic events at the moment of cadential arrival are primed, thus facilitating processing speed. In this case, participants indicated as quickly as possible whether the final chord event from each cadence was in or out of tune relative to the preceding context. Finally, §8.5 simulates the findings for the stimuli from Experiments III-V using three sensory-cognitive models of musical processing: the model of auditory sensory (or echoic) memory developed by Marc Leman,[7] which represents a sensory model of auditory expectations; Petr Janata's *Tonal Space* model,[8] which projects the output of Leman's model to the surface of a torus using a self-organizing map (SOM) algorithm that has been trained on the zeroth-order pitch distribution information from a tonal corpus, and thus combines sensory and cognitive approaches to expectancy formation; and finally IDyOM,[9] the cognitive approach

[7] Marc Leman, "An Auditory Model of the Role of Short-Term Memory in Probe-Tone Ratings," *Music Perception* 17, no. 4 (2000): 481–509.

[8] Petr Janata et al., "The Cortical Topography of Tonal Structures Underlying Western Music," *Science* 298 (2002): 2167–2170.

[9] Pearce, "The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition."

presented in Chapter 6, which predicts the next event in a musical stimulus by acquiring simulated long-term knowledge from a tonal corpus through unsupervised statistical learning of sequential structure.

§8.1 Expectations: Experimental Evidence

8.1.1 Explicit and Implicit Methods

According to psychologists Neal J. Roese and Jeffrey W. Sherman, expectancies are "beliefs about a future state of affairs, subjective estimates of the likelihood of future events ranging from merely possible to virtually certain."[10] These beliefs may be consciously held, reflecting declarative knowledge accumulated through explicit training, or beneath conscious awareness, reflecting tacit, procedural knowledge acquired through implicit exposure. As a consequence, both explicit and implicit measures of expectancy appear in the experimental literature. David Huron notes, however, that expectancy is a theoretical construct that is rarely directly observable. As a result, researchers must operationally define a measurable quantity that is somehow related to, if not synonymous with, the construct of interest.[11] In this case, researchers often employ a converging-methods approach in the hopes that demonstrable evidence from multiple methodologies, both explicit and implicit, might better reflect the formation of expectations.[12]

Explicit measures typically consist of subjective ratings paradigms, in which participants monitor and self-report their expectations on a scale or slider. In retrospective ratings tasks, participants hear a musical context and then indicate the strength and specificity of their

[10] Neal J. Roese and Jeffrey W. Sherman, "Expectancy," in *Social Psychology: Handbook of Basic Principles*, 2nd ed., ed. Arie W. Kruglanski and E. Tory Higgins (New York: The Guilford Press, 2013), 91.

[11] Huron, *Sweet Anticipation*, 42.

[12] The converging-methods approach is particularly common in emotion studies. See, for example, Lorraine Chuen, David Sears, and Stephen McAdams, "Psychophysiological Responses to Auditory Change," *Psychophysiology*, 2016, doi:10.1111/psyp.12633.

expectations for further continuation,[13] or provide a measure of uncertainty for the range of possible future outcomes.[14] Experimenters also sometimes follow the context with a target (or probe) event and ask participants to indicate how well the target "fit" with (or completed) the preceding context.[15] In cases where the target event appears in the middle of the stimulus, experimenters alert participants to the impending target using a visual cue,[16] or they dispense with the retrospective ratings paradigm entirely and instead employ a continuous ratings task, in which participants use a slider to indicate how their expectations vary over time. Recent examples of such tasks include continuous predictability judgments made during melodies,[17] and continuous judgments of how well the musical context fit with a continuously sounding probe tone.[18] Finally, production tasks represent an alternative explicit measure in which participants produce the expected continuation following a musical context, either by singing,[19]

[13]Mark A. Schmuckler, "Expectation in Music: Investigation of Melodic and Harmonic Processes," *Music Perception* 7, no. 2 (1989): 109–150.

[14]Hansen and Pearce, "Predictive Uncertainty"; Huron, *Sweet Anticipation*, 46.

[15]Emmanuel Bigand and Marion Pineau, "Global Context Effects on Musical Expectancy," *Perception and Psychophysics* 59, no. 7 (1997): 1098–1107; Lola L. Cuddy and Carole A. Lunney, "Expectancies Generated by Melodic Intervals: Perceptual Judgments of Melodic Continuity," *Perception and Psychophysics* 57, no. 4 (1995): 451–462; Krumhansl and Kessler, "Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys"; E. Glenn Schellenberg, "Expectancy in Melody: Tests of the Implication-Realization Model," *Cognition* 58 (1996): 75–125; E. Glenn Schellenberg, "Simplifying the Implication-Realization Model of Melodic Expectancy," *Music Perception* 14, no. 3 (1997): 295–318; Mark A. Schmuckler, "Expectancy Effects in Memory for Melodies," *Canadian Journal of Experimental Psychology* 51, no. 4 (1997): 292–306; Barbara Tillmann et al., "Harmonic Priming in an Amusic Patient: The Power of Implicit Tasks," *Cognitive Neuropsychology* 24, no. 6 (2007): 603–622.

[16]Pearce et al., "Unsupervised Statistical Learning Underpins Computational, Behavioural, and Neural Manifestations of Musical Expectation"; Barbara Tillmann and Frédéric Marmel, "Musical Expectations Within Chord Sequences: Facilitation Due to Tonal Stability Without Closure Effects," *Psychomusicology* 23, no. 1 (2013): 1–5.

[17]Tuomas Eerola and Carol L. Krumhansl, "Real-Time Prediction of Melodies: Continuous Predictability Judgments and Dynamic Models," in *Proceedings of the Seventh International Conference on Music Perception and Cognition*, ed. Catherine Stevens et al. (Adelaide: Causal Productions, 2002), 473–476.

[18]Petri Toiviainen and Carol L. Krumhansl, "Measuring and Modeling Real-Time Responses to Music: The Dynamics of Tonality Induction," *Perception* 32 (2003): 741–766.

[19]James C. Carlsen, Pierre I. Divenyi, and Jack A. Taylor, "A Prelimary Study of Perceptual Expectancy in Melodic Configurations," *Bulletin of the Council for Research in Music Education* 22 (1970): 4–12; James C. Carlsen, "Some Factors Which Influence Melodic Expectancy," *Psychomusicology* 1, no. 1 (1981): 12–29; Sean Hutchins and Caroline Palmer, "Repetition Priming in Music," *Psychology of Popular Media Culture* 1(S) (2011): 69–88.

playing,[20] or notating the expected continuation.[21]

Explicit measures of expectancy have been criticized for confounding expectations derived from explicit training with those resulting from implicit exposure. Emmanuel Bigand notes, for example, that for tasks requiring the judgment of an aspect of musical structure for which musicians have been explicitly trained, effects of expertise are sometimes confused with familiarity with the experimental task, the musical stimuli, or both.[22] Production tasks are particularly culpable in this regard, since singing, playing, or composing expected continuations requires advanced musical training.[23] What is more, Huron points out that explicit measures require "conscious, contrived, reflective responses … when under normal listening conditions expectations may be largely unconscious and effortless."[24] For example, Barbara Tillmann and her co-authors have found that individuals with amusia—a congenital disorder characterized by severe deficiencies in the processing and production of pitch variations (commonly known as tone-deafness)—often demonstrate severe impairments in expectancy-related tasks requiring explicit processing, but behave similarly to healthy controls when responding to implicit tasks.[25] These dissociations between implicit and explicit task performance have been found in various neurological disorders (e.g., aphasia), suggesting that in most cases the formation, fulfillment, and violation of expectations lie beneath conscious awareness, involving neurological substrates that are more resistant to neurological attacks than those expectancies derived from explicit

[20] Mark A. Schmuckler, "Expectation in Music: Additivity of Melodic and Harmonic Processes" (PhD Dissertation, Cornell University, 1988); Schmuckler, "Expectation in Music."

[21] Steve Larson, "Continuations as Completions: Studying Melodic Expectations in the Creative Microdomain Seek Well," in *Music, Gestalt, and Computing: Studies in Cognitive and Systematic Musicology*, ed. Marc Leman (Berlin: Springer-Verlag, 1997), 321–334.

[22] Bigand, "More about the Musical Expertise of Musically Untrained Listeners," 304–305.

[23] Larson, "Continuations as Completions: Studying Melodic Expectations in the Creative Microdomain Seek Well."

[24] Huron, *Sweet Anticipation*, 45.

[25] Tillmann et al., "Harmonic Priming in an Amusic Patient." See also Omigie, Pearce, and Stewart, "Tracking of pitch probabilities in congenital amusia"; Omigie et al., "Electrophysiological correlates of melodic processing in congenital amusia."

knowledge.[26]

For these reasons, implicit tasks are far more common in the experimental literature. To measure expectations implicitly, experimental studies depend on the priming paradigm, which assumes that the processing of incoming events is affected by the context in which they appear; related or repeated events are primed, thus facilitating processing. In behavioral priming studies, researchers employ a secondary task that draws the participants' attention to features of the stimulus other than those being assessed, but that may still be affected by the relatedness (i.e., expectedness) of the target event.[27] Accuracies and response times (RTs) are then collected to determine whether the preceding context *primed* participants to expect related or repeated target events.

In psycholinguistics, the priming paradigm is used to examine the semantic relatedness between word pairs using a lexical decision task. Participants are presented with related word pairs like *lion-tiger* and unrelated word pairs like *table-tiger* and asked to indicate as quickly as possible whether the target is an existing word from the English language or a nonword foil (e.g., *tiger* vs. *tigel*). RT studies employing the lexical decision task have repeatedly demonstrated that the semantic relatedness of the word pair influences the accuracy and speed of the response, providing evidence in support of the view that semantically related targets are primed.[28]

In their landmark 1986 study, "Reaction Time and Musical Expectancy: Priming of Chords," Jamshed J. Bharucha and Keiko Stoeckig adapted the lexical decision task to tonal harmony by examining the tonal relatedness between chord pairs.[29] Participants were presented with pairs

---

[26] Arthur S. Reber, "The Cognitive Unconscious: An Evolutionary Perspective," *Consciousness and Cognition* 1, no. 2 (1992): 93–133.

[27] Tillmann et al., "Harmonic Priming in an Amusic Patient," 604.

[28] See, for example, James H. Neely, "Semantic Priming and Retrieval from Lexical Memory: Evidence for Facilitatory and Inhibitory Processes," *Memory & Cognition* 4, no. 5 (1976): 648–654; James H. Neely, "Semantic Priming and Retrieval from Lexical Memory: Roles of Inhibitionless Spreading Activation and Limited-Capacity Attention," *Journal of Experimental Psychology: General* 106, no. 3 (1977): 226–254.

[29] Bharucha and Stoeckig, "Reaction Time and Musical Expectancy."

of major or minor triads that were either closely related (e.g., C major and G major) or very distantly related (e.g., C major and F♯ major) on the circle of fifths, and asked to determine whether the target triad was in or out of tune (in the out-of-tune condition, the fifth of the chord was tuned flat by an eighth tone, or 25 cents). The authors found that participants were faster and more accurate for the closely related targets, indicating that the preceding context triad primed the participants to expect a target triad from a related tonal context.

Following their study, the priming paradigm became the canonical implicit method for the study of expectations during music listening. Over the past three decades, researchers have examined priming effects for melody,[30] harmony,[31] and timbre,[32] as well as complex interactions between multiple parameters, such as melody and harmony,[33] and voice leading and harmony.[34] These studies have also employed a number of secondary discrimination tasks, such as mode (major vs. minor),[35] timbre (same vs. different),[36] intonation (in tune vs. out

[30]Aarden, "Dynamic Melodic Expectancy"; Frédéric Marmel and Barbara Tillmann, "Tonal Priming Beyond Tonics," *Music Perception* 26, no. 3 (2009): 211–221.

[31]Bharucha and Stoeckig, "Reaction Time and Musical Expectancy"; Bigand and Pineau, "Global Context Effects on Musical Expectancy"; Emmanuel Bigand et al., "Effect of Global Structure and Temporal Organization on Chord Processing," *Journal of Experimental Psychology: Human Perception and Performance* 25, no. 1 (1999): 184–197.

[32]Barbara Tillmann and Emmanuel Bigand, "Musical Priming: Schematic Expectations Resist Repetition Priming," in *Proceedings of the 8th International Conference on Music Perception & Cognition*, ed. Scott D. Lipscomb et al. (Adelaide, Australia: Causal Productions, 2004), 674–676.

[33]Loui and Wessel, "Harmonic Expectation and Affect in Western Music."

[34]Benedicte Poulin-Charronnat and Emmanuel Bigand, "The Influence of Voice Leading on Harmonic Priming," *Music Perception* 22, no. 4 (2005): 613–627.

[35]Bharucha and Stoeckig, "Reaction Time and Musical Expectancy."

[36]Marmel and Tillmann, "Tonal Priming Beyond Tonics"; Elizabeth Hellmuth Margulis and William H. Levine, "Timbre Priming Effects and Expectation in Melody," *Journal of New Music Research* 35, no. 2 (2006): 175–182; Tillmann and Bigand, "Musical Priming: Schematic Expectations Resist Repetition Priming."

of tune),[37] dissonance (consonant vs. dissonant),[38] contour (up vs. down),[39] and articulation (legato vs. staccato).[40] For every task, participants were slower and less accurate for target events from a distantly related tonal context.

In addition to behavioral methods, researchers have examined priming effects using neural and psychophysiological measures. Studies measuring central nervous system (CNS) activity using electroencephalographic recordings (EEG),[41] functional magnetic resonance imaging (fMRI),[42] and magnetoencephalography (MEG)[43] have all demonstrated significant priming effects for tonal materials. In EEG studies, for example, expectancy violations produce a notable evoked response potential (or ERP) in the electrical currents produced by neuronal activity measured on the surface of the scalp. Harmonic expectancy violations produce two characteristic (i.e., replicable) components of the ERP: an early (right) anterior negativity E(R)AN occurring between 150–280 ms after the onset of the target chord, which is believed to reflect the cognitive processing of tonal harmony (what Stefan Koelsch has called the *music-*

[37] Bharucha, "Music Cognition and Perceptual Facilitation"; Hasan Gürkan Tekman and Jamshed J. Bharucha, "Implicit Knowledge Versus Psychoacoustic Similarity in Priming of Chords," *Journal of Experimental Psychology: Human Perception and Performance* 24, no. 1 (1998): 252–260; Timothy Justus and Jamshed Bharucha, "Modularity in Musical Processing: The Automaticity of Harmonic Priming," *Journal of Experimental Psychology: Human Perception and Performance* 27, no. 4 (2001): 1000–1011.

[38] Bigand and Pineau, "Global Context Effects on Musical Expectancy"; Bigand et al., "Effect of Global Structure and Temporal Organization on Chord Processing"; Poulin-Charronnat and Bigand, "The Influence of Voice Leading on Harmonic Priming."

[39] Loui and Wessel, "Harmonic Expectation and Affect in Western Music."

[40] Seung-Goo Kim, June Sic Kim, and Chun Kee Chung, "The Effect of Conditional Probability of Chord Progression on Brain Response: An MEG Study," *PLoS ONE* 6, no. 2 (2011): 1–9, doi:10.1371/journal.pone.0017337.

[41] Petr Janata, "ERP Measures Assay the Difference of Expectancy Violation of Harmonic Contexts in Music," *Journal of Cognitive Neuroscience* 7, no. 2 (1995): 153–164; Aniruddh D. Patel et al., "Processing Syntactic Relations in Language and Music: An Event-Related Potential Study," *Journal of Cognitive Neuroscience* 10, no. 6 (1998): 717–733.

[42] Stefan Koelsch et al., "Adults and Children Processing Music: An fMRI Study," *NeuroImage* 25 (2005): 1068–1076; Barbara Tillmann, Petr Janata, and Jamshed J. Bharucha, "Activation of the Inferior Frontal Cortex in Musical Priming," *Cognitive Brain Research* 16 (2003): 145–161.

[43] Burkhard Maess et al., "Musical Syntax Is Processed in Broca's Area: An MEG Study," *Nature Neuroscience* 4, no. 5 (2001): 540–545; Asuka Otsuka et al., "Neuromagnetic Responses to Chords are Modified by Preceding Musical Scale," *Neuroscience Research* 60 (2008): 50–55.

syntactic mismatch negativity, or MMN);[44] and a later bilaterial or right-lateralized negativity (N5) occurring approximately 500 ms after the onset of the target chord, which is assumed to reflect the integration of the target into the preceding harmonic context.[45] Along with CNS measures, priming effects have also been reported for peripheral psychophysiological measures of autonomic nervous system (ANS) activity such as the skin conductance response,[46] suggesting that both the CNS and ANS respond to expectancy violations in music.

The appeal of neural and psychophysiological methods in priming studies is that they measure implicit processes without requiring the experimenter to employ a competing secondary task. For this reason, researchers sometimes also consider the influence of attention on expectancy. In an early ERP study by Stefan Koelsch and his co-authors, for example, participants were presented with harmonic progressions featuring a syntactic violation, and then asked either to attend explicitly to the stimulus, or to respond to a timbre discrimination task. The authors

---

[44]Stefan Koelsch, Schmidt Björn-Helmer Schmidt, and Julia Kansok, "Effects of Musical Expertise on the Early Right Anterior Negativity: An Event-Related Brain Potential Study," *Psychophysiology* 39 (2002): 657–663.

[45]Stefan Koelsch et al., "Brain Indices of Music Processing: "Nonmusicians" Are Musical," *Journal of Cognitive Neuroscience* 12, no. 3 (2000): 520–541; Koelsch, Schmidt, and Kansok, "Effects of Musical Expertise on the Early Right Anterior Negativity"; Stefan Koelsch et al., "Effects of Unexpected Chords and of Performer's Expression on Brain Responses and Electrodermal Activity," *PLoS ONE* 3, no. 7 (2008): 1–10; Sakari Leino et al., "Representation of Harmony Rules in the Human Brain: Further Evidence From Event-Related Potentials," *Brain Research* 1142 (2007): 169–177; Psyche Loui et al., "Effects of Attention on the Neural Processing of Harmonic Syntax in Western Music," *Cognitive Brain Research* 25 (2005): 678–687; Maess et al., "Musical Syntax is Processed in Broca's Area: An MEG Study"; Mira Müller et al., "Aesthetic Judgments of Music in Experts and Laypersons—An ERP Study," *International Journal of Psychophysiology* 76 (2010): 40–51; Patel et al., "Processing Syntactic Relations in Language and Music"; Nikolaus Steinbeis, Stefan Koelsch, and John A. Sloboda, "The Role of Harmonic Expectancy Violations in Musical Emotions: Evidence from Subjective, Physiological, and Neural Responses," *Journal of Cognitive Neuroscience* 18, no. 8 (2006): 1380–1393; Eduardo G.A. Villarreal et al., "Distinct Neural Responses to Chord Violations: A Multiple Source Analysis Study," *Brain Research* 1389 (2011): 103–114. ERP studies of melodic expectations have produced fewer characteristic (i.e., replicable) MMNs in response to expectancy violations, but see Mireille Besson and Frédérique Faıta, "An Event-Related Potential (ERP) Study of Musical Expectancy: Comparison of Musicians with Nonmusicians," *Journal of Experimental Psychology: Human Perception and Performance* 21, no. 6 (1995): 1278–1296; Robbin A. Miranda and Michael T. Ullman, "Double Dissociation Between Rules and Memory in Music: An Event-Related Potential Study," *Neuroimage* 38, no. 2 (2007): 331–345; Pearce et al., "Unsupervised Statistical Learning Underpins Computational, Behavioural, and Neural Manifestations of Musical Expectation."

[46]Koelsch et al., "Effects of Unexpected Chords and of Performer's Expression on Brain Responses and Electrodermal Activity"; Steinbeis, Koelsch, and Sloboda, "The Role of Harmonic Expectancy Violations in Musical Emotions."

found that the presence of a secondary task had no effect on the ERP components, suggesting that attentional processes play very little role in the processing of harmonic syntax.[47] In a later study Psyche Loui and her co-authors reported similar null effects even when participants were told to ignore the stimuli and study reading comprehension passages, though the authors noted that the ERAN had a smaller amplitude and later onset in the reading condition.[48] Thus, violations of harmonic expectations elicit measurable ERPs even when participants explicitly attend to other stimuli.

In sum, the evidence is overwhelming that the formation of expectations during music listening plays a significant role in the processing of harmony and melody, parameters that remain essential to the perception of (cadential) closure. Nevertheless, explicit and implicit methods suffer from a variety of limitations that might endanger the generalizability of the reported findings, so many studies take a converging-methods approach by employing both explicit and implicit methods across several experiments.[49] Thus, the first goal of the experiments reported in this chapter was to investigate the link between expectancy and cadential closure using both explicit and implicit behavioral methods. In Experiment III, participants were presented with a *truncated* cadential excerpt, which omitted the final, target harmonic and melodic events at the moment of cadential arrival, and a *non-truncated* excerpt that included the final events. Following the truncated excerpt, participants indicated the strength and specificity of their expectations for a musical continuation. Following the non-truncated excerpt, they then indicated how well the final target events fit with the expectations they had formed during the preceding context. In Experiment IV, participants provided continuous ratings of the

[47] Koelsch et al., "Brain Indices of Music Processing: "Nonmusicians" Are Musical."

[48] Loui et al., "Effects of Attention on the Neural Processing of Harmonic Syntax in Western Music."

[49] Bigand and Pineau, "Global Context Effects on Musical Expectancy"; Koelsch et al., "Effects of Unexpected Chords and of Performer's Expression on Brain Responses and Electrodermal Activity"; Steinbeis, Koelsch, and Sloboda, "The Role of Harmonic Expectancy Violations in Musical Emotions"; Tillmann et al., "Harmonic Priming in an Amusic Patient."

strength of their expectation that the end of the excerpt was imminent on a one-dimensional analog scale. Finally, in Experiment V, participants indicated as quickly as possible whether the final target harmonic and melodic events at the moment of cadential arrival were in or out of tune, where out-of-tune targets were raised by 40 cents relative to the preceding context.

8.1.2 Stimuli: Staircasing Expectancy

Priming effects have generally been limited to stimuli featuring two or three expectancy levels: an expected condition, which is usually represented by a root-position tonic triad in harmonic contexts or $\hat{1}$ in melodic contexts; and an unexpected condition, which may feature any number of harmonies or scale degrees from distantly related tonal contexts. Two stimulus sets feature prominently in harmonic priming studies. The first appeared in a 1997 study by Emmanuel Bigand and Marion Pineau, and has since been used in a number of other studies published in the laboratories of Bigand and Tillmann. It consists of a set of four-voiced, homorhythmic eight-chord sequences that culminate in a V–I authentic cadential progression in the expected condition. The authors then preserved the two target chords (e.g., G major to C major) in the unexpected condition but altered the preceding six-chord context to reflect the dominant key, resulting in a I–IV target progression.[50] Across both behavioral and neural measures, all of the studies using this stimulus set replicated the findings from the initial study: participants were faster and more accurate in response to the V–I progression relative to the I–IV progression,

[50] Bigand and Pineau, "Global Context Effects on Musical Expectancy"; Bigand et al., "Effect of Global Structure and Temporal Organization on Chord Processing"; Poulin-Charronnat and Bigand, "The Influence of Voice Leading on Harmonic Priming"; Barbara Tillmann and Emmanuel Bigand, "Global Context Effect in Normal and Scrambled Musical Sequences," *Journal of Experimental Psychology: Human Perception and Performance* 27, no. 5 (2001): 1185–1196; Tillmann, Janata, and Bharucha, "Activation of the Inferior Frontal Cortex in Musical Priming"; Barbara Tillmann et al., "The Influence of Musical Relatedness on Timbre Discrimination," *European Journal of Cognitive Psychology* 18, no. 3 (2006): 343–358; Tillmann et al., "Harmonic Priming in an Amusic Patient"; Barbara Tillmann et al., "Tonal Centers and Expectancy: Facilitation or Inhibition of Chords at the Top of the Harmonic Hierarchy," *Journal of Experimental Psychology: Human Perception and Performance* 34, no. 4 (2008): 1031–1043.

suggesting that the preceding tonal context primed listeners to expect the more common terminal harmonic progression.

The second most common stimulus set first appeared in a study by Stefan Koelsch, Tomas Gunter, and Angela D. Friederici. It also consists of four-voiced, homorhythmic chord sequences that observe the rules of common-practice voice leading, and it also employs the same authentic cadential progression in the expected condition. In the unexpected condition, however, Koelsch replaced the expected root-position tonic with $\flat\mathrm{II}^6$, a choice which, while syntactically incongruous, also rarely occurs in tonal music. Again, Koelsch and his colleagues have repeatedly demonstrated harmonic expectancy violations for $\flat\mathrm{II}^6$, using both behavioral and neural measures.[51]

Along with these stimulus sets, priming studies have also demonstrated expectancy violations for diatonic harmonies and scale degrees from the same tonal context, such as I vs. V,[52] and I vs. vi in harmonic contexts,[53] and $\hat{1}$ vs. $\hat{4}$ in melodic contexts.[54] These increasingly subtle priming effects are perhaps best exemplified in a recent study by Barbara Tillmann and her co-authors that compared RTs for the three chords at the top of the harmonic hierarchy: I, V, and IV. Their findings revealed a graded hierarchy of tonal stability, with participants responding most quickly

[51] Koelsch et al., "Brain Indices of Music Processing: "Nonmusicians" Are Musical"; Koelsch, Schmidt, and Kansok, "Effects of Musical Expertise on the Early Right Anterior Negativity"; Koelsch et al., "Children Processing Music"; Koelsch et al., "Adults and Children Processing Music: An fMRI Study"; Leino et al., "Representation of Harmony Rules in the Human Brain"; Loui et al., "Effects of Attention on the Neural Processing of Harmonic Syntax in Western Music"; Loui and Wessel, "Harmonic Expectation and Affect in Western Music"; Maess et al., "Musical Syntax is Processed in Broca's Area: An MEG Study."

[52] Barbara Tillmann et al., "The Costs and Benefits of Tonal Centers for Chord Processing," *Journal of Experimental Psychology: Human Perception and Performance* 29, no. 2 (2003): 470–482; Tillmann et al., "Tonal Centers and Expectancy: Facilitation or Inhibition of Chords at the Top of the Harmonic Hierarchy"; Tillmann and Marmel, "Musical Expectations Within Chord Sequences: Facilitation Due to Tonal Stability Without Closure Effects."

[53] Stefan Koelsch et al., "Untangling Syntactic and Sensory Processing: An ERP Study of Music Perception," *Psychophysiology* 44, no. 3 (2007): 476–490; Kim, Kim, and Chung, "The Effect of Conditional Probability of Chord Progression on Brain Response: An MEG Study."

[54] Frédéric Marmel, Barbara Tillmann, and Charles Delbé, "Priming in Melody Perception: Tracking Down the Strength of Cognitive Expectations," *Journal of Experimental Psychology: Human Perception and Performance* 36, no. 4 (2010): 1016–1028.

for tonic targets, followed by dominant targets, and then subdominant targets. Frédéric Marmel and Tillmann also extended this non-tonic priming effect to melodic contexts, observing that participants were faster and more accurate for $\hat{3}$ relative to $\hat{7}$.[55]

These stimulus sets have since been criticized for conflating general harmonic expectations formed during music listening with those expectations pertaining to the "closing" progressions of tonal music. Tillmann and Marmel recently noted, for example, that the terminal position of the target could prime participants to expect a tonic target, regardless of the context preceding it. Using the same stimulus set, the authors selected tonic and dominant targets located at various positions within each stimulus and employed a visual cue to alert the participants to the impending target.[56] They found that participants were faster for tonic over dominant targets regardless of the temporal position of the target, though participants were also significantly faster for targets appearing at later positions. Thus, participants were primed by the length and content of the preceding context, and not by an implicit awareness of the temporal position of the target.[57] Nevertheless, the authors could not discount the possibility that mid-stimulus tonic targets might elicit closing effects. Moreover, they did not consider the potential reciprocal influence between expectancy and closure: A cadential progression at the end of a phrase or theme might affect the strength and specificity of our expectations, but presumably those expectations also contribute to the perception of closure.

To be sure, much of the debate surrounding the link between expectancy and closure in the priming literature concerns the manner of the cognitive representation(s) that a tonal context presumably primes.[58] Given the degree to which zeroth-order frequency distributions of note

[55]Marmel and Tillmann, "Tonal Priming Beyond Tonics."

[56]A number of priming studies have used this approach. See, for example, Aarden, "Dynamic Melodic Expectancy"; Pearce et al., "Unsupervised Statistical Learning Underpins Computational, Behavioural, and Neural Manifestations of Musical Expectation."

[57]Tillmann and Marmel, "Musical Expectations Within Chord Sequences: Facilitation Due to Tonal Stability Without Closure Effects."

[58]Researchers even disagree as to whether priming effects result from mental representations of tonal structure

and chord events have dominated the field in recent decades,[59] priming studies often assume that the variable-order relations between events in the preceding context contribute very little to tonal priming effects. Nevertheless, several studies select target melodic and harmonic events by appealing to the syntactic relations between the context and the target using either fixed, first-order context models (i.e., transition probabilities),[60] or more sophisticated variable-order context models.[61] The assumption in these studies is that listeners with exposure to tonal music possess schematic representations for recurrent temporal patterns such that a stimulus context activating one or more of those representations will prime certain targets over others. That is, the specific *order* of events in the context determines whether, and to what degree, the target will be primed. Following Helen Brown, we might call priming effects induced by schematic knowledge of zeroth-order frequency distributions *structural priming*, whereas priming effects resulting from knowledge of harmonic and melodic syntax might be called *functional priming*.[62]

To compare structural and functional tonal priming effects, Tillmann and Bigand used Bigand and Pineau's original stimulus set but introduced a scrambled condition, in which the six context triads were swapped 2 by 2 (e.g., 2-1-4-3-6-5), or 4 by 4 (4-1-5-2-6-3). The results demonstrated that accuracies and reaction times were relatively unaffected by the

stored in long-term memory, or from psychacoustic similarities between the context and the target stored in auditory sensory memory (see 8.1.3).

[59] See, for example, Krumhansl and Kessler, "Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys."

[60] Janata, "ERP Measures Assay the Difference of Expectancy Violation of Harmonic Contexts in Music"; Schmuckler, "Expectation in Music"; Kim, Kim, and Chung, "The Effect of Conditional Probability of Chord Progression on Brain Response: An MEG Study."

[61] Egermann et al., "Probabilistic Models of Expectation Violation Predict Psychophysiological Emotional Responses to Live Concert Music"; Pearce et al., "Unsupervised Statistical Learning Underpins Computational, Behavioural, and Neural Manifestations of Musical Expectation."

[62] According to Brown, a structural account of tonality defines a tonal context by examining the pitch content of a set of tones, so a "zeroth-order frequency distribution" is simply a count of each distinct pitch or pitch class in a passage of music. Brown contrasts this synoptic and atemporal account with a functional account, which assumes that listeners track tonal music by constantly assessing the positions of the tones as they are heard ("The Interplay of Set Content and Temporal Context in a Functional Theory of Tonality Perception," *Music Perception* 5, no. 3 [1988]: 222).

scrambling condition, suggesting that the temporal order of the triads in the context only weakly contributes to harmonic priming effects.[63] Nevertheless, simulations of tonal priming effects using computational approaches have demonstrated the degree to which k^{th}-order context models can better predict RTs than those using zeroth-order models. In a meta-study comparing several computational approaches, Tom Collins and his co-authors found that a closure variable representing the first-order transition probability between the context and the target significantly improved the model, with high probabilities of closure leading to faster RTs.[64] Thus, the role played by functional priming effects remains patently unclear, as few studies have attempted to replicate and extend these findings, either by using more controlled stimuli, or by using computational models that might simulate the behavior of listeners.

Perhaps worse, few of the secondary tasks mentioned in the previous section permit the behavioral study of musical events experienced in "real-time" and at tempi similar to those found in real music. In nearly all of the previously mentioned studies, experimenters selected or composed rhythmically isochronous passages presented at relatively long IOIs, and then recomposed the target or the preceding context in the unexpected condition. Marcus Pearce and his co-authors have criticized priming studies for precisely this reason, noting that ecological validity tends to be low in the harmonic priming literature because participants are less likely to encounter such artificially constructed stimuli in the natural environment, thus endangering the generalizability of the collected evidence.[65] To resolve this issue, a few recent studies have demonstrated melodic priming effects for genuine musical materials using behavioral and

[63] Tillmann and Bigand, "Global Context Effect in Normal and Scrambled Musical Sequences."

[64] Tom Collins et al., "A Combined Model of Sensory and Cognitive Representations Underlying Tonal Expectations in Music: From Audio Signals To Behavior," *Psychological Review* 121, no. 1 (2014): 33–65. It is also worth noting that Petri Toiviainen and Carol Krumhansl compared models based on zeroth-order and first-order transition probabilities to account for goodness-of-fit ratings and found that the models performed equally well, so the efficacy of functional over structural models is still very much in dispute ("Measuring and Modeling Real-Time Responses to Music").

[65] Pearce et al., "Unsupervised Statistical Learning Underpins Computational, Behavioural, and Neural Manifestations of Musical Expectation," 303.

psychophysiological measures.[66] Nevertheless, none of these studies extended the reported findings to stimuli featuring multi-voiced textures, where the syntactic relations between harmonies play a greater role in the formation of tonal expectations. Indeed, I noted in Chapter 7 that instances of cadential failure could provide ideal stimuli for studies adopting a priming paradigm, as cadential deviations represent a violation of expectation when the listener's expectations are highest.[67] In other words, instances of cadential deception and evasion provide ready examples of expectancy violation derived from real music.

Thus, the second goal of the present study was to examine the formation, fulfillment, and violation of melodic and harmonic expectations for cadences from the classical repertoire. To that end, the stimulus set for Experiments III, IV, and V consists of the 40 out-of-context excerpts from Experiment II (see Table 7.8). The hypothesis here is that genuine cadences will elicit the highest fit ratings and strongest priming effects because they fulfill listener expectations for the target melodic and harmonic events at the moment of cadential arrival. Conversely, cadential deviations will elicit the lowest fit ratings and weakest priming effects since they violate expectations for the melodic and harmonic events at the cadential arrival. What is more, by including excerpts from the HC category, the present study also explicitly examines how expectancies in authentic cadential contexts might differ from those in half-cadential contexts.

8.1.3 Sensory and Cognitive Accounts

To explain how listeners form expectations during music listening, researchers typically offer either sensory or cognitive accounts, or in some cases some combination of the two. According to sensory accounts, facilitation effects arise when the preceding context shares sensory features

[66] Aarden, "Dynamic Melodic Expectancy"; Egermann et al., "Probabilistic Models of Expectation Violation Predict Psychophysiological Emotional Responses to Live Concert Music"; Pearce et al., "Unsupervised Statistical Learning Underpins Computational, Behavioural, and Neural Manifestations of Musical Expectation."

[67] Meyer, *Emotion and Meaning in Music*.

with the target. Mark Schmuckler suggests, for example, that "a chord sharing component tones, or overtones, with a preceding chord will be more highly anticipated than a continuation containing no overlapping frequencies with its predecessor."[68] When the preceding context shares sensory or psychoacoustic information with the target, the resulting facilitation effect is called *sensory priming*; researchers have also demonstrated facilitation effects resulting from the repetition of note or chord events from the target in the preceding context, called *repetition priming*, but for our purposes repetition priming is simply a special case of sensory priming in which the context shares entire note or chord events with the target. Thus, sensory accounts appeal to the *psychoacoustic similarities* between note or chord events accumulated over just a few seconds in echoic memory and assume that bottom-up representations of the auditory environment alone can account for the processing of tonal syntax during music listening.

According to cognitive accounts, listeners generate expectations by detecting and remembering the statistical relationships between co-occurring events in the auditory environment. If events in the stimulus co-occur with considerable frequency, the preceding context will activate schematic representations of the stimulus and then prime listeners to expect certain targets over others. Thus, both structural and functional priming effects appeal to the *statistical learning* of (co-)occurrences between note or chord events over the course of one's life and assume that top-down schematic representations of the auditory environment govern what the sensory apparatus will pick up next. Whether the target shares sensory features with the preceding context is thus inconsequential to cognitive accounts unless those features co-occur with enough frequency to justify their prior representation in long-term memory. Psychologist Aniruddh Patel summarizes this view thusly: "one strong piece of evidence for a cognitivist view of tonal syntax is that certain psychological properties of musical elements derive from their context

[68] Schmuckler, "Expectation in Music," 134.

and structural relations rather than from their intrinsic physical features."[69]

Unfortunately, disentangling low-level sensory influences from cognitive accounts of tonal expectancy remains a tremendous challenge, and appealing to the musical materials themselves tends to complicate rather than clarify matters. Sensory or psychoacoustic accounts of tonal harmony are deeply rooted in the history of Western music theory,[70] and continue to find favor in contemporary scholarship.[71] Emmanuel Bigand and his co-authors note, for example, that tonal syntax reflects the constraints of acoustic structure, such as octave equivalence and harmonic overtones, as well as general auditory mechanisms related to the perception of acoustic dissonance,[72] virtual pitch perception,[73] and principles of auditory stream analysis.[74] They also point out that the note and chord events associated with the tonic and dominant feature strong overlaps in harmonic spectra, which suggests that the acoustic properties of sounds provide an "acoustic foundation" for tonal syntax.[75] Max Matthews, John R. Pierce, and Linda A. Roberts have summarized this view as the *acoustic nucleus hypothesis*. They write, "with new materials, it is necessary to have an acoustic nucleus on which to grow powerful musical connotations via long-term learning. The acoustic nucleus consists of sound qualities that are perceivable at a low peripheral level."[76] According to this view, the contextual relations between note and chord

---

[69] Aniruddh D. Patel, *Music, Language, and the Brain* (Oxford: Oxford University Press, 2008), 260.

[70] See, for example, Jean-Phillippe Rameau, *Treatise on Harmony*, trans. Phillip Gossett (New York: Dover, 1971).

[71] Edward W. Large et al., "A Neurodynamic Account of Musical Tonality," *Music Perception* 33, no. 3 (2016): 319–331; Leman, "An Auditory Model of the Role of Short-Term Memory in Probe-Tone Ratings"; Parncutt, *Harmony: A Psychoacoustical Approach*.

[72] Plomp and Levelt, "Tonal Consonance and Critical Bandwidth"; William A. Sethares, "Consonace-Based Spectral Mappings," *Computer Music Journal* 22, no. 1 (1998): 56–72.

[73] Terhardt, "Pitch, Consonance, and Harmony"; Ernst Terhardt, Gerhard Stoll, and Manfred Seewann, "Pitch of Complex Signals According to Virtual-Pitch Theory: Tests, Examples, and Predictions," *Journal of the Acoustical Society of America* 71 (1982): 671–678.

[74] Wright and Bregman, "Auditory Stream Segregation and the Control of Dissonance in Polyphonic Music."

[75] Emmanuel Bigand et al., "Empirical Evidence for Musical Syntax Processing? Computer Simulations Reveal the Contribution of Auditory Short-Term Memory," *Frontiers in Systems Neuroscience* 8 (2014): 3.

[76] Max V. Matthews, John R. Pierce, and Linda A. Roberts, "Harmony and New Scales," in *Harmony and Tonality*, ed. Johan Sundberg (Stockholm: Royal Swedish Academy of Music, publ. no. 54, 1987), 83.

events in the tonal system could simply be "an emergent property" of echoic memory.[77]

To study the cognitive components of musical expectancies, researchers typically control for the sensory or psychoacoustic similarities between the preceding context and the target by manipulating the experimental stimuli so that (1) the context and the target share no component note or chord events (repetition priming); and/or (2) the selected timbre consists of a simple periodic waveform like a sine tone to minimize the potential for shared overtones (sensory priming). In the last three decades, priming studies have demonstrated that after short contexts, targets from related tonal contexts were processed faster than unrelated targets even when the context and target did not share sensory information,[78] and when they were separated by a silent interval or a white noise burst.[79] In longer contexts, participants also demonstrated facilitated processing for related targets compared to less related targets,[80] and this priming effect persisted even when the less-related targets shared more tones with the context than related targets.[81] What is more, these schematic priming effects remained unaffected by veridical knowledge about how each stimulus might proceed. In two recent studies, participants were permitted to preview the entire stimulus before completing the secondary task, but they were still slower to respond to unrelated targets, indicating that schematic expectations resist veridical knowledge about the stimulus.[82]

Despite the wealth of counter evidence from experimental studies for the limited role played by sensory priming in the formation of tonal expectations, advocates of both sensory and

[77]Bigand et al., "Empirical Evidence for Musical Syntax Processing? Computer Simulations Reveal the Contribution of Auditory Short-Term Memory," 19.

[78]Bharucha, "Music Cognition and Perceptual Facilitation."

[79]Hasan Gürkan Tekman and Jamshed J. Bharucha, "Time Course of Chord Priming," *Perception and Psychophysics* 51 (1992): 33–39.

[80]Koelsch et al., "Untangling Syntactic and Sensory Processing: An ERP Study of Music Perception"; Marmel, Tillmann, and Delbé, "Priming in Melody Perception: Tracking Down the Strength of Cognitive Expectations."

[81]Emmanuel Bigand et al., "Sensory Versus Cognitive Components in Harmonic Priming," *Journal of Experimental Psychology: Human Perception and Performance* 29, no. 1 (2003): 159–171.

[82]Justus and Bharucha, "Modularity in Musical Processing: The Automaticity of Harmonic Priming"; Tillmann and Poulin-Charronnat, "Auditory Expectations for Newly Acquired Structures."

cognitive priming accounts have pointed to the success of computational models to simulate the observed priming effects. Richard Parncutt's psychoacoustic model of tonal harmony provides one early example.[83] Following Ernst Terhardt's studies on virtual pitch and pitch salience,[84] Parncutt quantified the harmonic relatedness between two chords according to the number of pitches they shared, taking into account the relative perceptual salience of each pair of pitches, which he called *pitch commonality*. In the 1997 priming study by Bigand and Pineau mentioned in the previous section, the authors simulated the observed priming effects by calculating the pitch commonality values between the prime and target triads, weighted according to recency. They reported larger pitch commonality values for tonic targets over subdominant targets, and the values also correlated significantly with the correct response times, thereby demonstrating the degree to which sensory and cognitive accounts make parallel predictions.[85]

Given the source of the stimuli selected for the experiments in this chapter, controlling for sensory effects by eliminating repetitions of component tones and overtones from the target in the preceding context is not feasible. Thus, the third and final goal of the present study is to consider how the most well-known sensory and cognitive computational models of tonal expectancy simulate the results from Experiments III, IV, and V. §8.5 evaluates the simulated priming effects from three models: (1) the *Auditory Echoic Memory* model developed by Marc Leman,[86] which computes the correlation of an auditory image of the target with the corresponding auditory image of the preceding context, and which has been shown to simulate priming effects for stimulus sets from 18 separate studies;[87] (2) the *Tonal Space* model,[88] a

[83] Parncutt, *Harmony: A Psychoacoustical Approach*.

[84] Terhardt, Stoll, and Seewann, "Algorithm for Extraction of Pitch and Pitch Salience from Complex Tonal Signals."

[85] Bigand and Pineau, "Global Context Effects on Musical Expectancy."

[86] Leman, "An Auditory Model of the Role of Short-Term Memory in Probe-Tone Ratings."

[87] Bigand et al., "Empirical Evidence for Musical Syntax Processing? Computer Simulations Reveal the Contribution of Auditory Short-Term Memory."

[88] Janata et al., "The Cortical Topography of Tonal Structures Underlying Western Music."

sensory-structural priming model that projects the output of Leman's model to the surface of a torus, and which has been shown to simulate the priming effects for the RT data from seven separate studies;[89] and (3) IDyOM, the context model presented in Chapter 6, which simulates functional priming effects by acquiring long-term knowledge about the sequential dependencies between contiguous note and chord events on the musical surface.

§8.2 Experiment III

8.2.1 Method

Participants

Participants were 40 members (20 female) of the Montreal community recruited through the Schulich School of Music and the McGill University classified ads. Ages ranged from 18 to 46 ($M = 24, SD = 6$). Twenty participants with musical training equivalent or superior to second-year-university level formed the musician group, and twenty participants with less than one year of musical training comprised the nonmusician group. To limit any effects caused by familiarity with the stimuli, no participant with more than two years of formal study on the piano was permitted to take part.

A questionnaire was administered to assess musical preferences and training. Participants reported listening to an average of 19 hours of music each week, and all but two participants self-identified as music lovers. The musicians practiced their primary instruments for an average of 20 hours each week, and had been playing their primary instruments for an average of 6 years. Musicians also averaged 4.9 years of ear training, 3.2 years of instruction in harmony, and 3.4 years of instruction in music analysis. All of the participants reported normal hearing,

[89] Collins et al., "A Combined Model of Sensory and Cognitive Representations Underlying Tonal Expectations in Music."

which was confirmed with a standard audiogram administered before the experiment,[90] and five musicians reported the ability to identify pitches absolutely.

Materials

The stimuli consisted of the 40 excerpts presented *out of context* in Experiment II. To examine harmonic and melodic expectations both before and after the events at cadential arrival, two versions of each excerpt were created: a *truncated* version that omits the final harmonic and melodic events at the cadential arrival, and a *non-truncated* version that includes the final events. Unfortunately, because these excerpts only present phrase endings, it was assumed that the selected stimuli would not represent the full range of the expectancy strength and specificity scales, since excerpts terminating at the beginning or middle of a musical phrase could potentially generate weaker expectations.[91] So as not to bias ratings of expectation toward one end of the expectancy scales, eight foil stimuli that terminate in the middle of a musical phrase were also selected from Mozart's keyboard sonatas. The foils were inserted such that each group of six consecutive trials contained one foil that was neither the first nor the last member of the group.

Following the experimental design employed in Experiments I and II, performance features (such as dynamics and rubato) were neutralized and the tempo of each excerpt was determined by convention. To ensure that unwanted differences at cadential arrival would not affect expectancy ratings, the duration of cadential arrival was recomposed to 900 ms and any melodic dissonances at cadential arrival were removed. Each stimulus was first created with the notation

[90] 389-8, *Acoustics: Reference Zero for the Calibration of Audiometric Equipment—Part 8*; Martin and Champlin, "Reconsidering the Limits of Normal Hearing."

[91] Nicolas Escoffier and Barbara Tillmann, "The Tonal Function of a Task-Irrelevant Chord Modulates Speed of Visual Processing," *Cognition* 107, nos. 1070–1083 (2008); Pearce et al., "Unsupervised Statistical Learning Underpins Computational, Behavioural, and Neural Manifestations of Musical Expectation"; Tillmann and Marmel, "Musical Expectations Within Chord Sequences: Facilitation Due to Tonal Stability Without Closure Effects."

software Sibelius and then realized as a .wav sound file at a sampling rate of 44.1 kHz and 16-bit resolution using a piano physical model created by PianoTeq (Modartt S.A.S., Ramonville Saint Agne).

Design and Procedure

Participants were presented with a randomized set of 40 excerpts and 8 interspersed foils. After listening to the *truncated* version of each excerpt, participants rated on 7-point continuous analogical-categorical scales the strength of their expectation that the music would continue (*Expectancy Strength*), and the specificity of their expectation for a musical continuation (*Expectancy Specificity*). For the expectancy strength scale, participants were instructed that a value of 1 indicates that they had no expectations that the music would continue, whereas a value of 7 indicates that they had very strong expectations that the music would continue. On the expectancy specificity scale, a value of 1 indicates that they had no specific idea how the music would continue, whereas 7 indicates that they had a very specific idea how the music would continue. In addition to their expectancy ratings, participants also responded to the statement: "Following this excerpt, the end of the passage is imminent," on a 4-point Likert scale labeled from *strongly agree* to *strongly disagree*. Next, participants listened to the *non-truncated* version of the same excerpt and rated on a 7-point scale how well the final chord fit with the expectations they had formed when they heard the truncated version, with a rating of 1 indicating that the musical continuation fit very poorly with the expectations they had formed while listening to the previous truncated excerpt, and a rating of 7 indicating that the musical continuation fit very well with their expectations.

To familiarize the participants both with the range of stimuli as well as with the experimental task, the session began with an exposure phase and a practice phase consisting of 12 additional excerpts. After completing the experiment, participants filled out a short questionnaire addressing their music background.

Analysis

Data were analyzed with a linear mixed effects model (LMM) approach,[92] an alternative to conventional regression models (MLR, ANOVA, etc.) that allows the researcher to control for random sources of variance without the loss of statistical power resulting from data aggregation across subjects or stimuli, which is a frequent preliminary step in repeated-measures designs (e.g., F1 and F2 ANOVAs, RM-ANOVA). Mixed effects models have become increasingly common in the analysis of linguistic priming data because they can accommodate both continuous and binary response data,[93] which regularly violate assumptions of normality and homogeneity of variance in repeated-measures designs,[94] and often lead to unbalanced datasets as a result of the omission of incorrect responses from the analysis (as was the case in Experiments IV and V).

As suggested by Harald Baayen and his co-authors,[95] I included crossed random effects for participants and items (musical excerpts). Equation 8.1 presents the general model formulation they provided:

$$y_{ij} = X_{ij}\beta + S_i s_i + W_j w_j + \epsilon_{ij} \tag{8.1}$$

[92]Brady T. West, Kathleen B. Welch, and Andrzej T. Galecki, *Linear Mixed Models. A Practical Guide using Statistical Software* (Boca Raton: Chapman & Hall/CRC, 2007).

[93]The *Journal of Memory and Language* recently devoted a special issue to emerging data-analytic methods, with mixed effects modeling procedures receiving substantial attention (Kenneth I. Forster and Michael E. J. Masson, "Special Issue: Emerging Data Analysis," *Journal of Memory and Language* 59, no. 4 [2008]: 387–556). For a discussion of the issues surrounding the analysis of interval data using LMMs, see R. Harald Baayen, Doug J. Davidson, and Douglas M. Bates, "Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items," *Journal of Memory and Language* 59 (2008): 390–412; R. Harold Baayen and Petar Milin, "Analyzing Reaction Times," *International Journal of Psychological Research* 3, no. 2 (2010): 12–28. For a discussion of how LMMs generalize to ordinal and categorical data, see Peter Dixon, "Models of Accuracy in Repeated-Measures Designs," *Journal of Memory and Language* 59 (2008): 447–456; T. Florian Jaeger, "Categorical Data Analysis: Away from ANOVAs (Transformation or Not) and Towards Logit Mixed Models," *Journal of Memory and Language* 59 (2008): 434–446.

[94]Dixon, "Models of Accuracy in Repeated-Measures Designs"; Jaeger, "Categorical Data Analysis."

[95]Baayen, Davidson, and Bates, "Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items."

The vector y_{ij} represents the responses of subject i to item j. X_{ij} is the design matrix, consisting of an initial column of ones (representing the intercept) and followed by columns representing factor contrasts and covariates. This matrix is multiplied by the vector of population coefficients β. The $S_i s_i$ and $W_j w_j$ terms constitute the model's random effects structure, which serve to make the model's predictions more precise with respect to the subjects and items actually examined in the experiment. S_i and W_j are full copies of the X_{ij} matrix, and both are multiplied with the vectors s_i and w_j, respectively, which specify for subject i (in the case of the S_i matrix) and item j (in the case of the W_j matrix) the adjustments required. The last term is a vector of the residual errors, which includes one error for each combination of subject and item.

Previous studies have treated the specification of the random effects structure as an empirical problem, only including random intercepts and by-unit random slopes if they improve model fit (i.e., if they explain a significant proportion of the variance in the outcome variable).[96] Dale Barr and his co-authors have recently argued, however, that the standard practice in mixed-model ANOVAs has always been to specify "maximal" random-effects structures because underspecified random-effects structures (e.g., random-intercepts only LMMs) can lead to significantly higher Type I error rates.[97] They therefore advocate for the selection of a maximal random effects structure a priori, which produces nominal Type I error rates while substantially improving statistical power relative to traditional parametric models (e.g., RM-ANOVA). All mixed-effects analyses were conducted with the software R (2.15) using the packages lme4[98] and languageR.[99] Ratings of expectation strength, specificity, and fit were analyzed with linear

[96]Baayen, Davidson, and Bates, "Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items."

[97]Dale J. Barr et al., "Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal," *Journal of Memory and Language* 68 (2013): 255–278.

[98]D.M. Bates, M. Maechler, and B. Bolker, *lme4: Linear Mixed-Effects Models Using S4 Classes, (R package version 0.98.501) [computer software].*, 2011.

[99]R. Harold Baayen, *languageR: Data Sets and Functions with "Analyzing Linguistic Data: A Practical Introduction to Statistics". R package version 1.4.*, 2012.

mixed effects models using the lmer function. Following Barr et. al,[100] all models included a full random effects structure as specified by the design of the experiment, with intercepts for each participant and by-participant slopes for the within-subject fixed factor of cadence category (PAC, IAC, HC, DC, EV), and with intercepts for each musical stimulus and by-stimulus slopes for the between-subjects factor of musical training (musicians, nonmusicians). To calculate omnibus tests and parameter estimates, models were fit using sum coding for the predictor variables so that levels of the fixed effects would represent deviations from the grand mean, as is the approach in traditional ANOVA pedagogy.[101] Tests of main effects and interactions were calculated using the Anova function from the car package,[102] and pairwise comparisons and polynomial contrasts of the levels of the fixed factors were calculated using the lsmeans package.[103] Finally, to visualize the effects of the included fixed factors on participant ratings after controlling for the random variance in the dataset, the figures present the estimated means and standard errors determined by the model.

8.2.2 Results

Figure 8.1 displays line plots of the estimated mean expectation strength, specificity, and fit ratings of musicians and nonmusicians for each cadential category. Excerpts from the foil condition have also been plotted for comparison. As expected, musicians and nonmusicians provided lower strength and specificity ratings for the foil condition than for any of the other cadence conditions, suggesting phrase endings generate stronger and more specific expectations than excerpts terminating in the middle of a phrase. The intention behind including the foils,

[100]Barr et al., "Random Effects Structure for Confirmatory Hypothesis Testing."
[101]Ibid.
[102]John Fox and Sanford Weisberg, *An R Companion to Applied Regression*, 2nd ed. (Thousand Oaks, CA: Sage, 2011).
[103]Russell V. Lenth, "lsmeans: Least-Squares Means. R package version 2.00-4," 2013, http://CRAN.R-project.org/package=lsmeans.

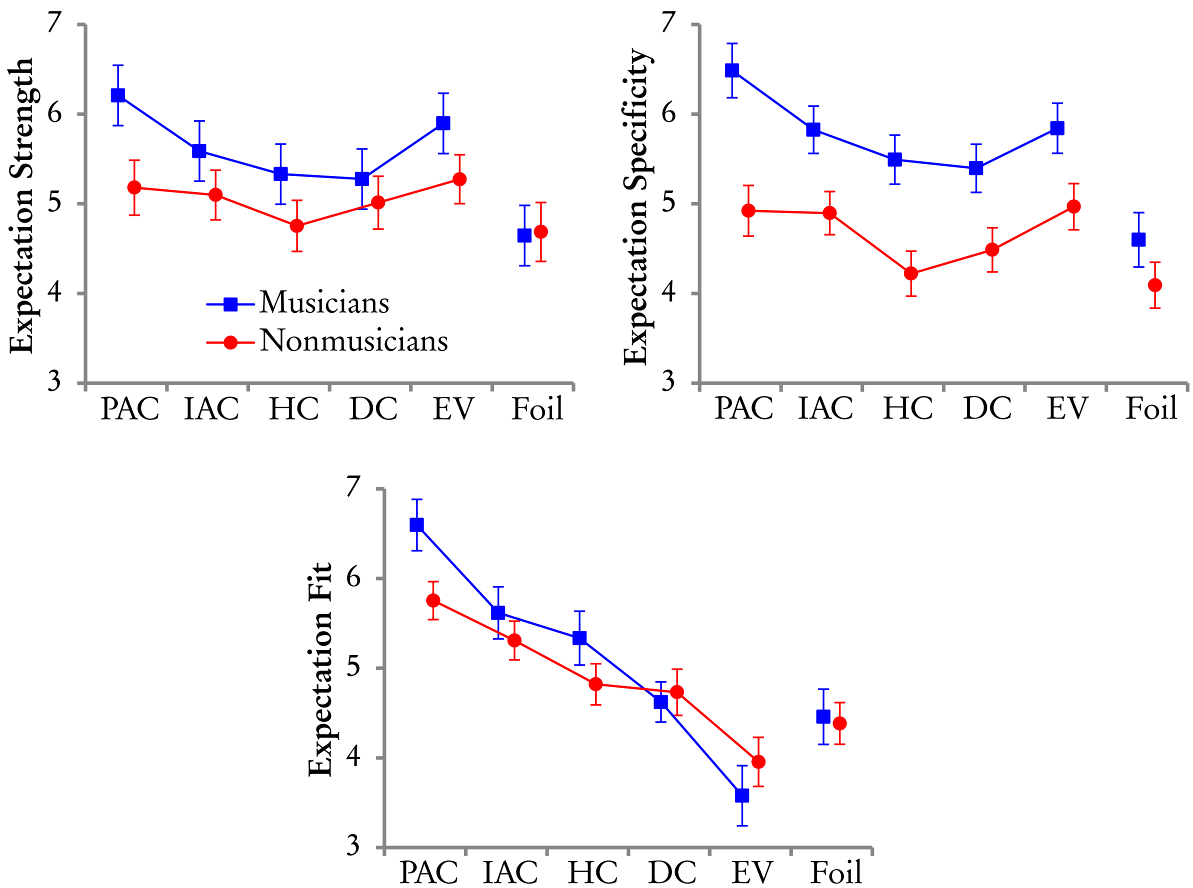


Figure 8.1: Line plots of the estimated mean expectation strength, specificity, and fit ratings of musicians and nonmusicians for each cadence category and the foil category. Whiskers represent ±1 standard error.

however, was simply to encourage participants to use the full range of the scale, so the foil trials were employed as fillers, and only responses to the cadence categories are presented in the analyses that follow, resulting in a preliminary dataset of 1600 trials (40 stimuli × 40 participants).

Shown in Table 8.1, Type III Wald *F* tests of a mixed 5×2 LMM of the expectation strength ratings with Kenward-Roger approximation for the denominator degrees of freedom revealed a significant effect of cadence category.[104] Because the genuine and deviation conditions of

[104]Ulrich Halekoh and Søren Højsgaard, “A Kenward-Roger Approximation and Parametric Bootstrap Methods

the authentic cadence only differ at the moment of cadential arrival, it was predicted that the truncated excerpts from the HC category would receive the lowest expectancy strength ratings. As expected, pairwise comparisons of the cadence categories using the Tukey HSD test revealed significant differences between the PAC–HC and HC–EV categories ($p < .05$), with the HC category receiving the lowest strength ratings overall. Moreover, a polynomial contrast of the cadence categories revealed a significant quadratic trend from the PAC to EV categories, $B = 1.99, t = 3.37, p < .01$, with the estimated means exhibiting a U shape from the outer cadence categories (PAC and EV) to the inner category (HC). The model estimates also suggested a main effect of training on expectation strength, with nonmusicians providing lower ratings than musicians, but this effect was not significant, $F(1, 37.98) = 2.77, p = .10$, and there was no interaction.

Expectation specificity ratings were moderately correlated with the strength ratings, $r(1600) = .56, p < .001$, and visual inspection of the model estimates indicated a similar pattern of responses to those found for the strength ratings (see Figure 8.1). Type III Wald F tests revealed main effects of cadence category and training, as well as a significant interaction (see Table 8.1). The PAC and HC categories received the highest and lowest specificity ratings, respectively, and polynomial contrasts revealed the same U-shaped quadratic trend in the ratings for both groups (musicians, $B = 2.45, t = 2.65, p < .05$; nonmusicians, $B = 2.02, t = 2.43, p < .05$). Musicians also provided significantly higher specificity ratings than nonmusicians, suggesting that increased exposure to the music of a given style may influence the specificity of expectations formed during music listening.

The 4-point Likert-scale ratings for the statement, "following this excerpt the end of the passage is imminent," provided similar results to those observed for the analogical-categorical scales of expectation strength and specificity. The upper plot in Figure 8.2 presents a bar plot

for Tests in Linear Mixed Models – The R Package pbkrtest," *Journal of Statistical Software* 59, no. 9 (2014).

Table 8.1: Analysis of deviance table for maximal linear mixed effects models predicting ratings of expectation strength, specificity, and fit.

	df^{a}	Wald F
Expectation Strength		
Cadence Category	39.41	3.00*
Training	37.98	2.77
Cadence Category × Training	23.81	1.54
Expectation Specificity		
Cadence Category	39.91	3.57*
Training	39.14	17.38***
Cadence Category × Training	27.90	2.98*
Expectation Fit		
Cadence Category	43.74	14.26***
Training	43.77	1.27
Cadence Category × Training	37.03	2.14

Note. $N = 1600$. [a]Denominator degrees of freedom for Type III Wald F tests reported with Kenward-Roger approximation. Independent variables are factor variables with sum coding (e.g., *musicians* = 1, *nonmusicians* = −1). Results of Wald F test: * $p < .05$ ** $p < .01$ *** $p < .001$. A maximum random effects structure was included, with a random intercept for participants and by-participant slopes for cadence category, and a random intercept for musical stimuli and by-stimulus slopes for musical training.

of the distribution of the proportion of responses for each cadence category, with musicians' ratings above and nonmusicians' ratings below the x-axis. Like the previous bar graphs from Experiments I and II (see Figures 7.4 and 7.8), this representation estimates the similarity between the ratings of musicians and nonmusicians by visualizing the symmetry of their responses about the x-axis. The very first category in the musician group, for example, indicates that for truncated excerpts from the PAC category musicians *strongly agreed* in 65% of all trials that the end was imminent. As expected, both groups generally *agreed* or *strongly agreed* with the statement for the genuine and deviation conditions of the authentic cadence. But whereas musicians generally *disagreed* or *strongly disagreed* that the end was imminent for excerpts from the HC category (58%), nonmusicians demonstrated a preference to *strongly agree* or *agree* with the statement throughout the experimental session, even for the excerpts from the foil condition

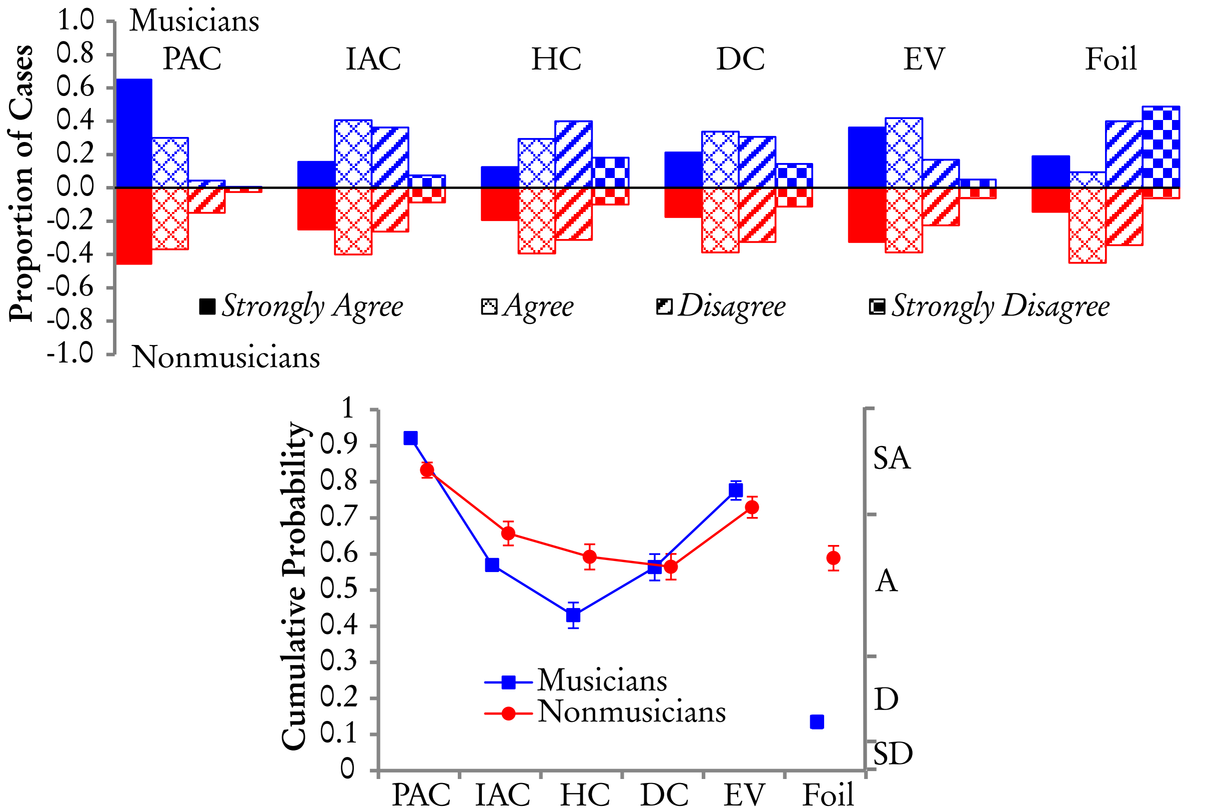


Figure 8.2: Top: Bar plot of the distribution of the proportion of responses for each cadential category and the foil category for the statement, "the ending is imminent," with musicians' ratings above the x-axis in blue and nonmusicians' ratings below in red. Pattern fills denote response types. Bottom: Line plot of the estimated cumulative probabilities calculated from the Likert scale ratings. Right y-axis presents thresholds of the response scale: *Strongly Agree* (SA), *Agree* (A), *Disagree* (D), and *Strongly Disagree* (SD). Whiskers represent ±1 standard error.

that terminated in the middle of a musical phrase. The design of the experiment—presenting a *truncated* and *non-truncated* version of each excerpt in direct succession in each trial—may therefore have biased nonmusicians to always expect a phrase ending following the end of each truncated excerpt.

Given the ordinal nature of the dependent variable, the Likert-scale ratings were entered into a proportional-odds logistic regression model using the polr function from the MASS library.[105] In order to estimate a linear regression model on ordinal data, proportional-odds models transform the cumulative probabilities calculated across all values of the response variable (in this case, from *strongly agree* to *strongly disagree*) onto an unbounded log-odds scale. For transparency, however, in the bottom plot in Figure 8.2 the estimates were back-transformed onto a cumulative probability scale [0-1] on the left y-axis, with the estimated threshold values between adjacent levels of the likert scale shown on the right y-axis. Higher cumulative probability estimates indicate a tendency to *strongly agree* or *agree*, whereas lower cumulative probability estimates indicate a tendency to *disagree* or *strongly disagree*.

Type III Wald chi-square tests revealed a main effect of cadence category, $\chi^2(4) = 223.38, p < .001$, and the effect of training was not significant, $\chi^2(1) = .02, p > .05$, but there was a significant interaction, $\chi^2(4) = 30.13, p < .001$. The PAC and HC categories again received the highest and lowest ratings, respectively, and a quadratic trend was observed in the ratings of both the musicians, $B = 7.42, z = 12.75, p < .001$, and nonmusicians, $B = 3.53, z = 6.4, p < .001$, thereby replicating the U-shaped curves found in the strength and specificity estimates. Musicians were also more likely to *disagree* or *strongly disagree* that the end was imminent following truncated excerpts from the HC category, $B = -.65, z = -3.19, p < .01$, with nonmusicians demonstrating a bias to *agree* or *strongly agree* throughout the experimental session.

Shown in Table 8.1, a mixed 5×2 LMM of the expectation fit ratings for the non-truncated excerpts revealed a significant effect of cadence category, but musical training and the interaction were not significant. The PAC category received the highest fit ratings from both groups, and

[105] William N. Venables and Brian D. Ripley, *Modern Applied Statistics with S* (New York: Springer, 2002). Unfortunately, the proportional-odds model applied here does not account for random variance associated with the participants and items used in the experiment. Attempts to apply mixed effects procedures to ordinal models are currently still in development. At present the mixed effects logistic regression models available in R (e.g., clmm or MCMCglmm) do not support the calculation of omnibus tests or pairwise comparisons. For this reason, I elected to omit the random effects structure entirely (Barr et al., "Random Effects Structure for Confirmatory Hypothesis Testing").

pairwise comparisons with the Tukey HSD test revealed significant differences between the PAC category and the remaining cadence categories excepting the IAC category. Moreover, polynomial contrasts revealed a significant decreasing linear trend from the PAC to the EV categories for the fit ratings of both musicians, $B = -7.03, t = -6.84, p < .001$, and nonmusicians, $B = -4.16, t = -5.03, p < .001$, a finding that replicates the cadential hierarchy observed in the IC estimates from IDyOM in Chapter 6 (see Figures 6.3 and 6.4), as well as the completion ratings from Experiments I and II in Chapter 7 (see Figures 7.2 and 7.6).

8.2.3 Discussion

By excluding the events at cadential arrival, Experiment III sought to determine whether the half cadence is unique among the selected cadence categories. Indeed, if the tonic is the most stable sonority within the tonal system, and accordingly, if listeners have internalized the distributional properties that characterize that system, presumably any cadence category for which the goal harmony is dominant will elicit weaker and less specific expectations relative to those for which the goal harmony is tonic. The above findings lend support to this claim, as ratings of expectation strength, specificity, and phrase completion for excerpts from the truncated condition exhibited a U-shaped pattern, with excerpts from the PAC and HC categories receiving the highest and lowest ratings, respectively. In fact, *all* of the categories with tonic harmony as the expected goal elicited higher ratings of expectancy strength and specificity for the moment of cadential arrival than did excerpts from the half cadence category, suggesting that the half cadence is best understood as an *incomplete authentic cadence* prior to the moment of cadential arrival.

When these excerpts included the events at cadential arrival, however, ratings of expectation fit demonstrated the same descending linear trend observed in the previous two experiments, with the half cadence category positioned not at the bottom of the cadential hierarchy, but somewhere in the middle. Taken together, these two findings are consistent with the Janus-faced view of closure espoused by Eugene Narmour,[106] whereby the half cadence serves as the weakest category in prospect as a result of the relatively weak and unspecific expectations it affords, yet finds itself near the middle of the cadential hierarchy in retrospect by virtue of the fulfilment of those expectations, however weakly formed. Thus, the cadential hierarchy itself reflects a bipolar continuum of schematic expectancy characterized by two opposing end points, with the fulfilment or confirmation of listener expectations on one end, and the negation, denial, or violation of those expectations on the other. A central property of this scale is precisely that it represents both the direction *and* intensity of the realization, where direction on the schematic scale corresponds to the location of the realized event(s) near one end or the other (i.e., either as a fulfilment or violation), and where intensity (or strength) corresponds to the distance of those event(s) from the center of the scale.

According to this scale, the authentic cadence categories receive very high completion and expectancy fit ratings because the events at cadential arrival largely confirm or fulfil harmonic and melodic expectations formed in prospect, doing so with such intensity as to elicit ratings to one extreme of the scale. And by violating expectations when the listener's expectations are highest, deviations of the authentic cadence categories necessarily lie on the other extreme of the schematic spectrum. But whereas the cadence categories supporting the tonic as the goal harmony stipulate both direction and intensity, expectations for the events at cadential arrival for the half cadence category remain so weakly formed that their fulfillment positions the category very near the center of the scale, resulting in the cadential hierarchy PAC→IAC→HC→DC→EV.

Nevertheless, to demonstrate that half cadences elicit weaker and less specific expectations

[106] Narmour, *Beyond Schenkerism*; Steven Haflich, Review of *Beyond Schenkerism: The Need for Alternatives in Music Analysis*, by Eugene Narmour, *Journal of Music Theory* 23, no. 2 (1979): 290.

in prospect compared to the other cadence categories necessitates an alternative method, one that measures how expectations for the impending end might vary over time. To that end, Experiment IV required participants to continuously rate their expectations for the end of the excerpt on a slider. The hypothesis here is that expectancies will gradually increase over the course of the excerpt until the moment of cadential arrival for each cadence category, but that half cadences will elicit weaker expectations—and thus, lower ratings—relative to the other cadence categories.

§8.3 Experiment IV

8.3.1 Method

Participants

The participants were the same as those who participated in Experiment III.

Materials

The excerpts were the same as those employed in Experiment III, but all foil excerpts and excerpts from the truncated condition were omitted, resulting in a stimulus set consisting of 40 excerpts. Pilot testing also revealed that excerpts shorter than 8 seconds in duration were potentially too short to rate accurately in a continuous response task, so 13 of the stimuli were extended by the necessary number of measures before their original start to increase their duration beyond 8 seconds ($M = 11.5$ s, $SD = 2$ s).

Design and Procedure

Participants were presented with a randomized set of the 40 revised, non-truncated excerpts and asked to continuously rate the strength of their expectation that the end of the excerpt was imminent on a one-dimensional analog scale. The term “imminent” was defined as “within the next one to two seconds,” and the left and right limits of the scale were labeled with “very weak” and “very strong,” respectively. Following the onset of the final chord, participants were told to move the slider back to the left limit of the scale as quickly as possible to indicate that the excerpt had ended.[107]

The slider was connected to an Arduino-based USB interface (Arduino, Torino, Italy) that recorded the slider values on a continuous scale from 1 to 7 at a sampling rate of 100 Hz, and the computer interface provided instructions on the screen and allowed the participant to advance through the trials by clicking the mouse on a button on-screen. Before each trial began, stimulus playback would not occur until the slider was positioned to the left limit of the scale, and participants were encouraged not to begin moving the slider until they started to expect that the end of the passage was imminent. To familiarize the participants with the experimental task, the session began with a practice phase consisting of five additional excerpts. Because the same group of participants completed Experiments III and IV in the same session, the order of presentation for the two experiments was counterbalanced across participants.

Analysis

The continuous slider data were processed in MATLAB (The Mathworks, Inc., Natick, MA). To remove extraneous information and ensure a smooth time series in each trial, the data were

[107] I elected to use a unipolar scale as opposed to the bipolar scale described in §8.2.3 because this experiment examines whether expectations formed in prospect might differ between the half cadence category and the other cadence categories. Thus, the end of the expectancy scale corresponding to negation, denial, or violation was unnecessary for the present experiment.

low-pass filtered with a cutoff frequency of 4 Hz using a linear phase filter, which was based on the convolution of a 1st-order Butterworth filter impulse response that was also convolved with itself in time reverse to avoid phase shifting. To obtain a measure of the rate of change in the slider ratings (also referred to as rating *velocity*), each time series was first downsampled to 2 Hz using cubic spline interpolation, and first-order derivatives were then calculated from the resulting time series.

Following Stephen McAdams and his co-authors,[108] trials were excluded from further analysis if the slider ratings did not meet the following criteria: the slider was not positioned to the left of the expectancy scale when the trial began (2 trials); the participant failed to move the slider throughout the trial (44 trials); the participant failed to move the slider until after the onset of the final chord of the excerpt (132 trials). These criteria resulted in the exclusion of 178 of 1600 trials (11%).

8.3.2 Results

To visualize the slider ratings for each cadence category, the grand mean time course for the untransformed and velocity-transformed slider ratings was calculated for musicians and nonmusicians using a time window from five seconds preceding to three seconds following the onset of the cadential arrival. Shown in Figures 8.3 and 8.4, the dotted lines indicate the 95% confidence bounds around the grand mean time course for both training groups, with the bounds around the musician ratings shaded in blue. To examine how the slider ratings varied over time using a mixed effects modeling procedure, means were calculated for 1 s epochs centered from 4 s before the onset of the cadential arrival to 0 s. LMMs of the untransformed and velocity-transformed ratings therefore included fixed factors of cadence category (5 levels),

[108] Stephen McAdams et al., "Influences of Large-Scale Form on Continuous Ratings in Response to a Contemporary Piece in a Live Concert Setting," *Music Perception* 22, no. 2 (2004): 297–350.

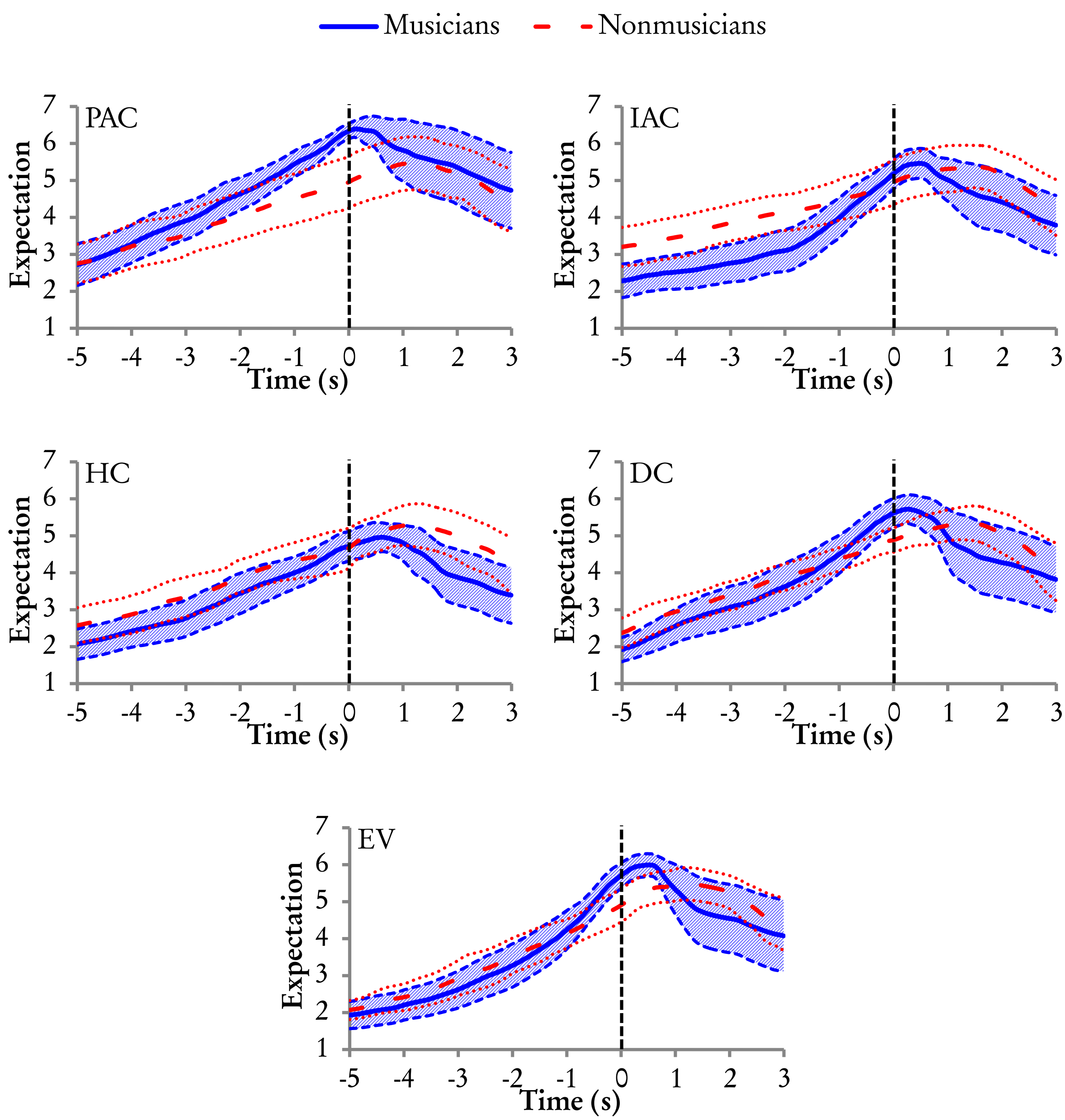


Figure 8.3: Grand mean time course for the slider ratings of musicians in blue and nonmusicians in dotted red for each cadential category. Equidistant dotted lines indicate 95% confidence bounds around the mean ratings, with the confidence bounds around the musician ratings shaded in blue. The vertical dotted line indicates the onset of the final chord.

musical training (2 levels), and time (5 levels), with random intercepts for each participant and by-participant slopes for the within-subject fixed factors of cadence category and time, and with intercepts for each musical stimulus and by-stimulus slopes for the between-subjects factor of musical training. Table 8.2 presents the omnibus tests calculated from the LMMs of the untransformed and velocity-transformed slider ratings.

Beginning with the untransformed slider ratings in Figure 8.3, Type III Wald F tests reported with Kenward-Roger approximation revealed a significant effect of time, $F(33.7) = 93.73, p < .001$, with the mean time course increasing until the moment of cadential arrival for every cadence category and for both training groups, at which point it tended to decrease. The model also revealed significant interactions between cadence category and training, $F(51.8) = 4.05, p < .01$, cadence category and time, $F(6679.5) = 5.06, p < .001$ and training and time, $F(33.7) = 3.48, p < .05$. For the PAC category, polynomial contrasts demonstrated significant linear increasing trends across time for the ratings of both musicians, $B = 7.37, t = 16.78, p < .0001$, and nonmusicians, $B = 4.78, t = 10.00, p < .0001$. Although the average time courses for both groups appeared at approximately the same position on the scale at 5 s preceding cadential arrival, musicians' ratings also exhibited a significantly steeper slope compared to nonmusicians, $B = .26, t = 3.53, p < .05$, which explains the higher peak at cadential arrival. For the remaining categories for which the tonic was the goal harmony, musicians also demonstrated an exponential increasing trend across time (IAC, $B = 2.17, t = 3.97, p < .0001$; DC, $B = 1.25, t = 2.35, p < .05$; EV, $B = 2.21, t = 4.22, p < .0001$), with a relatively slower and more gradual rate of increase between a period of roughly 5 s and 2 s preceding a sudden and more steep increase in ratings within the final 2 s. This exponential increase suggests that musicians were less aware of the impending cadential arrival for the IAC, DC, and EV categories compared to the PAC category until approximately 2 s before the onset of the cadential arrival, which is perhaps when the cadential dominant first appeared.

Table 8.2: Analysis of deviance table for maximal linear mixed effects models predicting slider ratings and first-order derivatives of the slider ratings.

	df^{a}	Wald F
Slider Ratings		
Cadence Category	38.4	1.35
Training	41.5	.02
Time	33.7	93.73***
Cadence Category × Training	51.8	4.05**
Cadence Category × Time	6679.5	5.06***
Training × Time	33.7	3.48*
Cadence Category × Training × Time	6679.5	.75
Slider Velocity Ratings		
Cadence Category	32.8	4.72**
Training	36.3	11.53**
Time	33.7	4.85**
Cadence Category × Training	18.8	.90
Cadence Category × Time	6683.9	4.98***
Training × Time	33.7	3.84*
Cadence Category × Training × Time	6683.9	1.46

Note. N = 8000. [a]Denominator degrees of freedom for Type III Wald F tests reported with Kenward-Roger approximation to one decimal place. Independent variables are factor variables with sum coding (e.g., *musicians* = 1, *nonmusicians* = −1). Results of Wald F test: * $p < .05$ ** $p < .01$ *** $p < .001$. A maximum random effects structure was included, with a random intercept for participants and by-participant slopes for cadence category and time, and a random intercept for musical stimuli and by-stimulus slopes for musical training.

Shown in Figure 8.4, the velocity-transformed slider ratings capture this exponential rate of increase quite well for the IAC, DC, and EV categories, and particularly for the IAC category, in which a sharp increase in average velocity appears within the final 2 s before cadential arrival. The ratings of the nonmusician group did not demonstrate this exponential rate of increase, however, nor did their ratings differ significantly for *any* of the cadence categories in general; both the starting slider position and the rate of increase over time were nearly identical for every category.

For the half cadence category, visual inspection of the musician time course in Figure 8.3

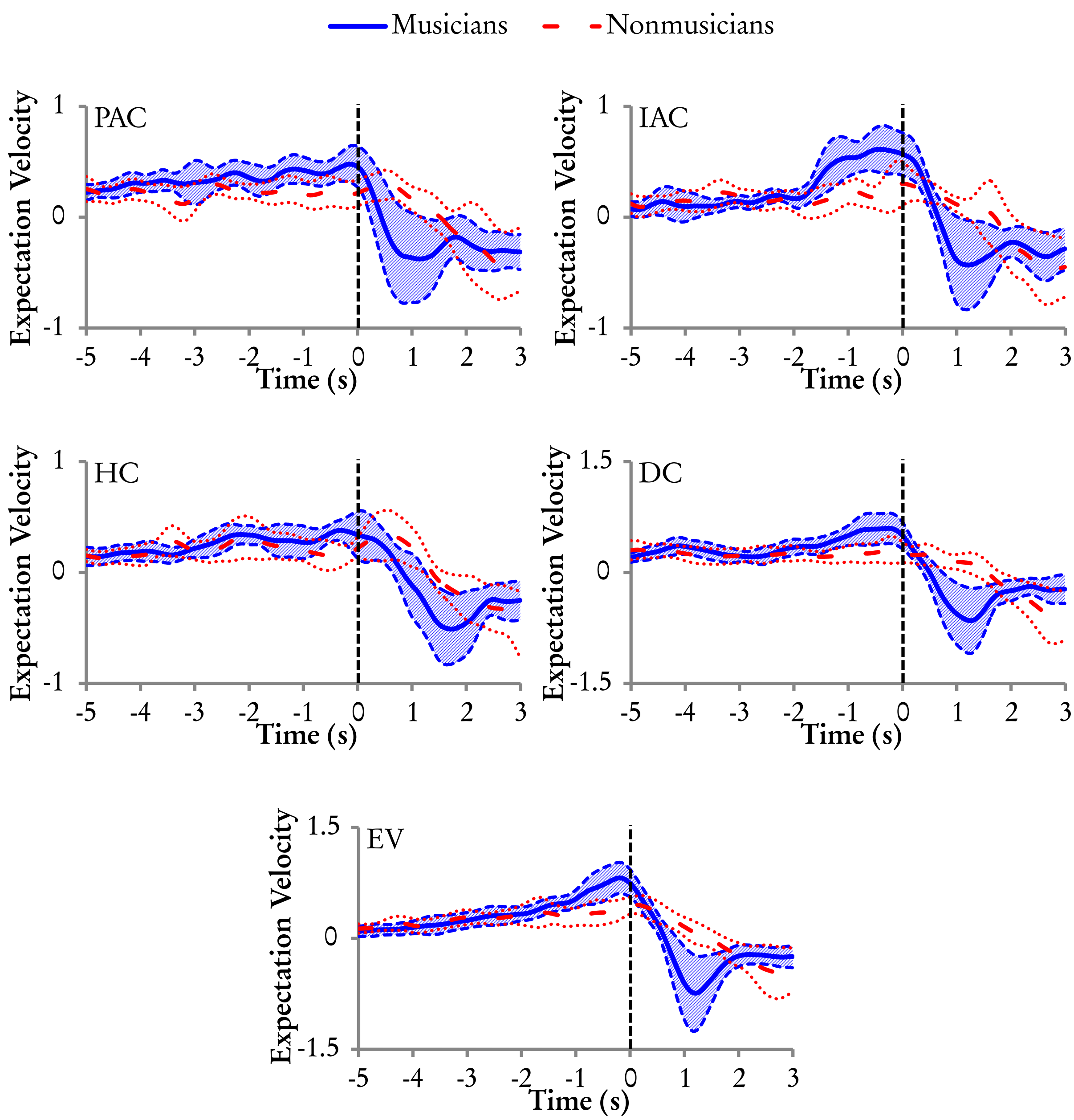


Figure 8.4: Grand mean time course for the first derivative of the slider ratings of musicians in blue and nonmusicians in dotted red for each cadential category. Equidistant dotted lines indicate 95% confidence bounds around the mean ratings, with the confidence bounds around the musician ratings shaded in blue. The vertical dotted line indicates the onset of the final chord.

would suggest that half cadences elicited the lowest peak ratings relative to the other cadence categories. To be sure, since participants were tasked with moving the slider to the bottom of the scale as quickly as possible following the moment of cadential arrival, the value and time index of the maximum rating represent the crucial moment in which the participants' expectations are highest, so they should be of interest to us here. Thus, if half cadences elicit significantly weaker expectations for the cadential arrival compared to the other categories, participants should reach a *lower* peak on the expectancy scale at a *later* point in time.

In some trials, the peak rating appeared *before* the moment of cadential arrival, suggesting that participants anticipated the end of the excerpt and consequently reached a plateau in their ratings. In other trials, however, the peak rating appeared *after* the cadential arrival, indicating that participants did not anticipate the impending end. Unfortunately, by calculating means in 1 s epochs, determining the average position and time index of the maximum slider rating would be impossible. Thus, in the analysis that follows, the average maximum rating was calculated for a 4 s window surrounding the cadential arrival, and trials were excluded if the slider ratings did not reach a maximum during this window, resulting in a dataset of 1114 trials. The left line plot in Figure 8.5 presents the estimated mean rating of the slider maxima, and the right plot presents the estimated time indices for those maxima. The horizontal dotted line indicates the onset of cadential arrival. Thus, for the PAC category, musicians reached the slider maximum 300 ms before the cadential arrival on average.

Type III Wald F tests of the fixed effects from the 5×2 LMM of the slider maxima revealed a significant effect of cadence category, $F(4, 37.27) = 2.63, p < .05$, and a significant interaction, $F(4, 26.00) = 4.10, p < .05$, but there was no main effect of training. As expected, the half cadence category received the lowest maximum rating on average, and polynomial contrasts revealed a quadratic trend in the ratings of the musicians, $B = 3.26, t = 4.27, p < .0001$, thereby replicating the U-shaped curves found in the strength, specificity, and Likert-scale ratings from

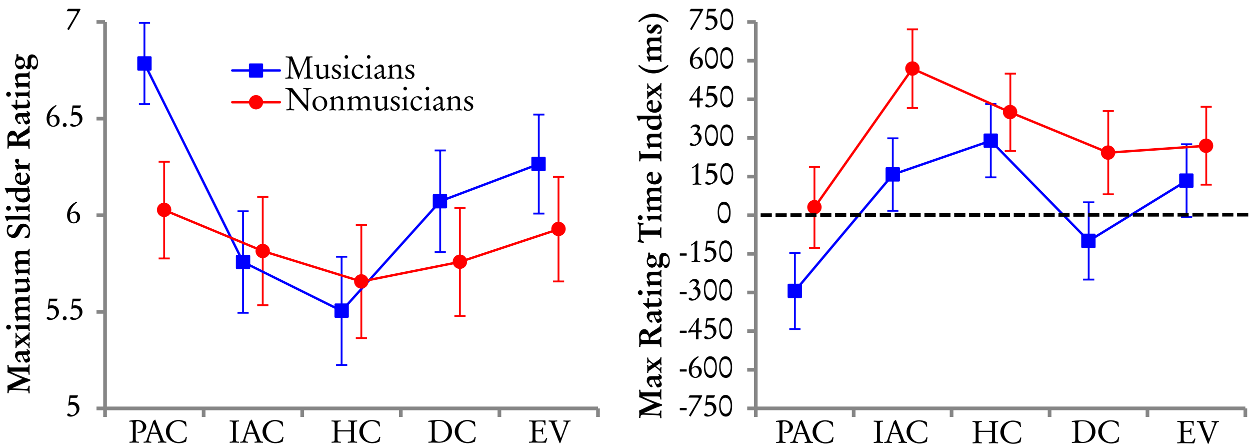


Figure 8.5: Left: Line plot of the estimated means of the slider rating maxima that occurred within a 4 s window surrounding the onset of the cadential arrival for musicians and nonmusicians for each cadence category $N = 1114$. Right: Line plot of the estimated means of the time indices for the maximum slider ratings. The horizontal dotted line indicates the onset of the cadential arrival. Whiskers represent ±1 standard error.

Experiment III. Although the same U-shaped trend emerged in the ratings of the nonmusician group, the polynomial contrast was not significant.

Type III Wald F tests from the LMM of the time indices of the slider maxima revealed a significant effect of cadence category, $F(4, 36.17) = 2.67, p < .05$, with both groups reaching the slider maximum the most quickly for the PAC category, $M = -130$ ms, $SE = 130$ ms. Musicians also reached the slider maximum more quickly than nonmusicians on average, $F(1, 41.42) = 7.00, p < .05$, and the PAC category provided the only estimated time index in which musicians anticipated the cadential arrival, $M = -290$ ms, $SE = 150$ ms, which indicates that perfect authentic cadences elicit the strongest and most specific expectations for the moment of cadential arrival. As predicted, the participants also reached the slider maximum for the HC category latest on average, $M = 340$ ms, $SE = 130$ ms, and polynomial contrasts revealed a weaker but still significant quadratic trend across the cadence categories, $B = -.96, t = -2.12, p < .05$. Taken together, the estimated ratings and time indices for the slider maxima therefore suggest that the PAC and HC categories generated the strongest and weakest expectations, respectively, with the remaining cadence categories falling somewhere in the middle.

8.3.3 Discussion

As predicted, expectations for closure increased over the course of each excerpt and then peaked at or near the moment of cadential arrival. For musicians, slider ratings demonstrated an increasing linear trend over time up to the cadential arrival for the PAC category. For the remaining cadence categories for which tonic harmony was the expected goal (IAC, DC, EV), however, expectations for closure increased exponentially over time, suggesting that musicians were less aware of the impending cadential arrival for these categories compared to the PAC category. Any number of features may have contributed to this difference (e.g., the length of the cadential progression, the presence of a cadential trill, a cadential six-four, etc.), but the point here is that excerpts from the PAC category feature syntactic and rhetorical parameters within cadential function that must have alerted the musician group far sooner to the impending cadential arrival relative to the other categories, resulting in a higher starting position and a generally linear (as opposed to exponential) increasing trend for the musician time course.

What is more, closer inspection of the average ratings and time indices for the slider maxima at or near the moment of cadential arrival revealed the same U-shaped pattern observed in the strength, specificity, and phrase completion ratings from Experiment III, with excerpts from the PAC and HC categories receiving the highest/earliest and lowest/latest maximum ratings, respectively. This finding provides converging evidence in support of the view that half cadences elicit weaker expectations in prospect than the cadence categories for which tonic harmony serves as the expected cadential goal.

For nonmusicians, the slider ratings did not differ for any of the cadence categories; the starting slider position and the rate of increase over time were nearly identical for every category, suggesting either that nonmusicians were simply unaffected by differences in the selected cadence categories—a hypothesis that seems unlikely given the pronounced differences observed in the nonmusician ratings for every other task employed across the five experiments—or that the task itself—asking participants to continuously monitor and rate their own subjective experiences while *simultaneously* listening to the musical excerpt—was simply too demanding for the nonmusician group. Indeed, the attentional and vigilance demands placed on participants in continuous ratings tasks may interfere with explicit processes related to the formation of expectations during music listening.[109] Perhaps worse, these tasks may fail to tap into the largely unconscious, automatic expectancies resulting from implicit processes during auditory perception. To measure these sorts of expectancies for the events at cadential arrival, Experiment V therefore adopts a priming paradigm and uses a competing secondary task to orient the participants' attention to other features of the stimulus. In this case, participants indicated as quickly as possible whether the note and chord events at the moment of cadential arrival were in or out of tune, where out-of-tune foil trials were tuned 40 cents sharp relative to the preceding context.

[109] McAdams et al., "Influences of Large-Scale Form on Continuous Ratings in Response to a Contemporary Piece in a Live Concert Setting"; Emery Schubert, "Measurement and Time Series Analysis of Emotion in Music" (PhD Dissertation, University of New South Wales, 1999).

§8.4 Experiment V

8.4.1 Method

Participants

Participants were 30 members (13 female) of the Montreal community recruited through the Schulich School of Music and the McGill University classified ads. Ages ranged from 18 to 35 ($M = 22, SD = 4$). Fifteen participants with musical training equivalent or superior to second-year-university level formed the musician group, and fifteen participants with less than one year of musical training comprised the nonmusician group. To limit any effects caused by familiarity with the stimuli, no participant with more than two years of formal study on the piano was permitted to take part.

A questionnaire was administered to assess musical preferences and training. Participants reported listening to an average of 16 hours of music each week, and all but four participants self-identified as music lovers. The musicians practiced their primary instruments for an average of 20 hours each week, and had been playing their primary instruments for an average of 12 years. Musicians also averaged 2.9 years of ear training, 3.3 years of instruction in harmony, and 3.4 years of instruction in music analysis. All of the participants reported normal hearing, which was confirmed with a standard audiogram administered before the experiment, and one musician reported the ability to identify pitches absolutely.

Materials

The excerpts were the same as those employed in Experiment III, but all foil excerpts and excerpts from the truncated condition were omitted, resulting in a stimulus set consisting of 40 excerpts. To create the intonation task for Experiment V, the final harmonic and melodic events at cadential arrival in each excerpt were presented both in tune and out of tune (i.e., tuned 40 cents sharp relative to the preceding musical context), resulting in 40 in-tune and 40 out-of-tune (foil) trials for the session.

Design and Procedure

The experimental session was divided into two phases. In the first training phase, participants were presented with a randomized set of 20 trials. In each trial, a playback cursor was provided at the top of the screen that followed along with the stimulus, and a black vertical line was placed on the playback bar along with a black circle placed directly above the line to mark the onset of the cadential arrival (see Figure 8.6). When the playback cursor reached the black line, the circle turned green, at which point participants were instructed to judge as quickly and accurately as possible whether the chord marked by the black line was in or out of tune by pressing one of two buttons on the keyboard, labeled “in” and “out,” respectively. Following the completion of each trial in the training phase, visual feedback was provided on the screen to indicate whether the response was correct or incorrect. In the second experimental phase, participants performed the in-tune/out-of-tune judgment without feedback, and the 80 trials were randomized such that the target and foil conditions of each excerpt were not presented within five experimental trials of each other. After completing the experiment, participants filled out a short questionnaire addressing their music background.

Analysis

As suggested in two recent articles by Florian Jaeger and by Hugo Quené and Huub van den Bergh,[110] response accuracies were analyzed with mixed-effects logistic regression models

[110] Jaeger, “Categorical Data Analysis”; Hugo Quené and Huub van den Bergh, “Examples of Mixed-Effects Modeling with Crossed Random Effects and with Binomial Data,” *Journal of Memory and Language* 59 (2008): 413–425.

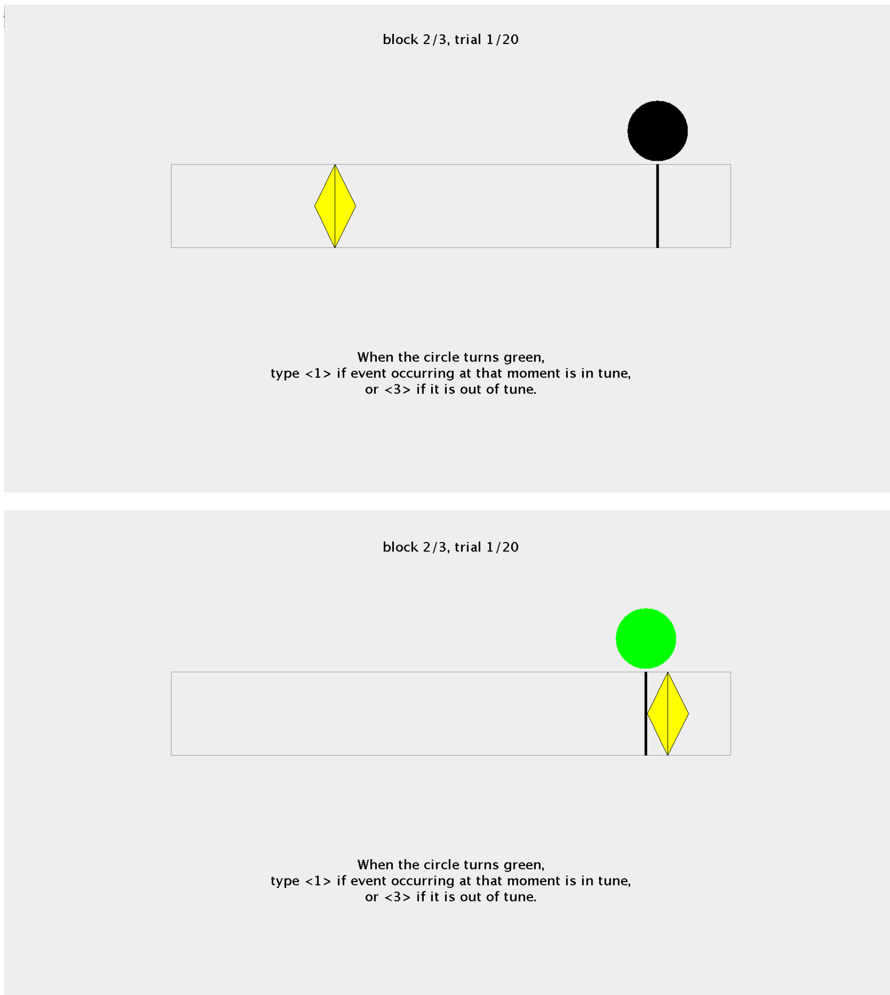


Figure 8.6: Screen shots of the interface used in Experiment V.

(GLMMs) using the glmer function, with the accuracy of the response as a binomial dependent variable. In order to estimate a linear regression model on proportion data, GLMMs transform proportions onto an unbounded log-odds scale, and all estimates in the following analyses are reported on the log-odds scale. For transparency, however, the plotted GLMM estimates were back-transformed onto a probability scale [0-1]. Correct RTs were analyzed with linear mixed effects models (LMMs) using the lmer function. Due to violations of normality observed in the residuals of the estimated RT models, analyses were conducted using log-transformed RTs.

8.4.2 Results

Following Emmanuel Bigand and his co-authors,[111] out-of-tune foils were excluded from the analysis under the assumption that they do not constitute lawful musical events.[112] Foil trials were instead employed as fillers, and only the data from in-tune target trials are presented here, resulting in a preliminary dataset of 1200 trials. Before entering accuracy into a GLMM, any responses occurring within 150 ms of the onset of the cadential arrival or later than 5 s following the onset the cadential arrival were deemed too early and too late, respectively, resulting in the omission of 14 trials (1%).

Before examining the effect of the individual cadence categories themselves, a binomial factor variable called *ending* was created to compare those cadences that achieve genuine thematic closure—PAC, IAC, and HC—against those cadences that represent deviations of the authentic cadence—DC and EV—under the assumption that the latter cadence categories would elicit the strongest violations of harmonic and melodic expectancy at cadential arrival. Type III Wald chi-square tests of a mixed 2×2 GLMM of the participant accuracies revealed significant effects of ending, $\chi^2(1) = 13.49, p < .001$, and training, $\chi^2(1) = 22.38, p < .001$, as well as a significant interaction, $\chi^2(1) = 15.57, p < .001$. For the musician group, the pairwise comparison between the genuine and deviation conditions did not reveal a significant difference, $B = .15, z = .6, p > .05$, but for nonmusicians this difference was significant, $B = 1.32, z = 4.93, p < .001$. Shown in Figure 8.7, the model predictions demonstrate that musicians were extremely adept at the intonation task, regardless of the syntactic ending appearing at cadential arrival. This effect was not demonstrated for the nonmusician group, however, as they performed no better than chance in the deviation condition. Indeed, compared

[111] Bigand et al., "Sensory Versus Cognitive Components in Harmonic Priming."

[112] See also Tillmann and Marmel, "Musical Expectations Within Chord Sequences: Facilitation Due to Tonal Stability Without Closure Effects."

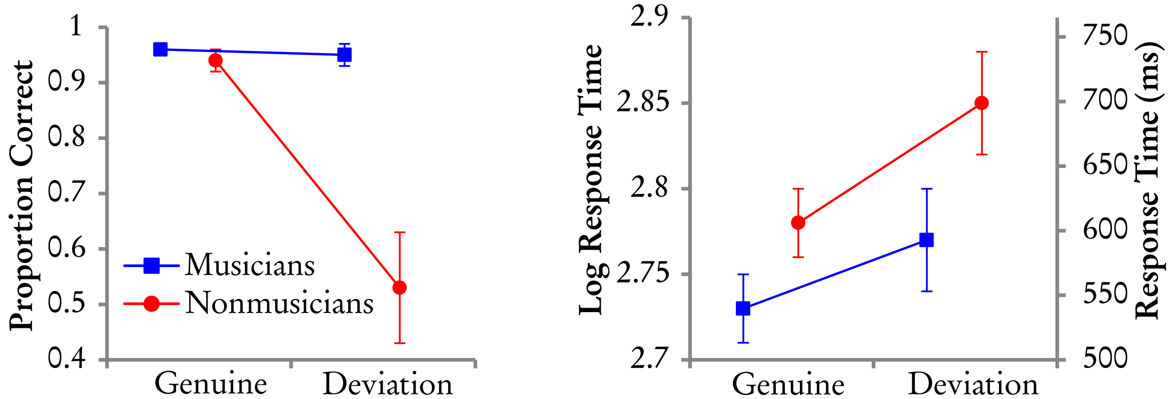


Figure 8.7: Line plots of the estimated mean proportion correct and response times of musicians and nonmusicians for each ending. Whiskers represent ±1 standard error.

to the grand mean, nonmusicians were roughly half as likely to accurately identify the intonation of the final chord, $B = -.58$, odds ratio $= .56$, Wald $Z = -3.68, p < .001$.

Type III Wald F tests of the logRTs revealed a significant effect of ending, $F(1, 47.71) = 10.9, p < .01$, and pairwise comparisons indicated logRTs were significantly slower in the deviation condition for both musicians, $B = -.04, t = -2.34, p < .05$, and nonmusicians, $B = -.07, t = -3.23, p < .01$. Shown in Figure 8.7, the model estimates also suggested a main effect of training, with nonmusicians appearing to be slower in their responses than musicians, but this effect was marginal, $F(1, 28.4) = 3.63, p = .067$, and there was no interaction.

To examine the effect of the cadence categories individually, I specified 5×2 mixed-effects models that included a within-subject factor of cadence category and a between-subjects factor of musical training. Type III Wald chi-square tests revealed significant main effects of cadence category, $\chi^2(4) = 15.13, p < .01$, and training, $\chi^2(1) = 16.76, p < .001$, and a significant interaction, $\chi^2(4) = 18.17, p < .01$. For the musician group, none of the cadence estimates differed significantly from the intercept, indicating that the cadential categories had no impact on the accuracy of the intonation judgment. For the nonmusician group, however, the estimated odds of correctly identifying the intonation of the target chord for excerpts from the PAC category

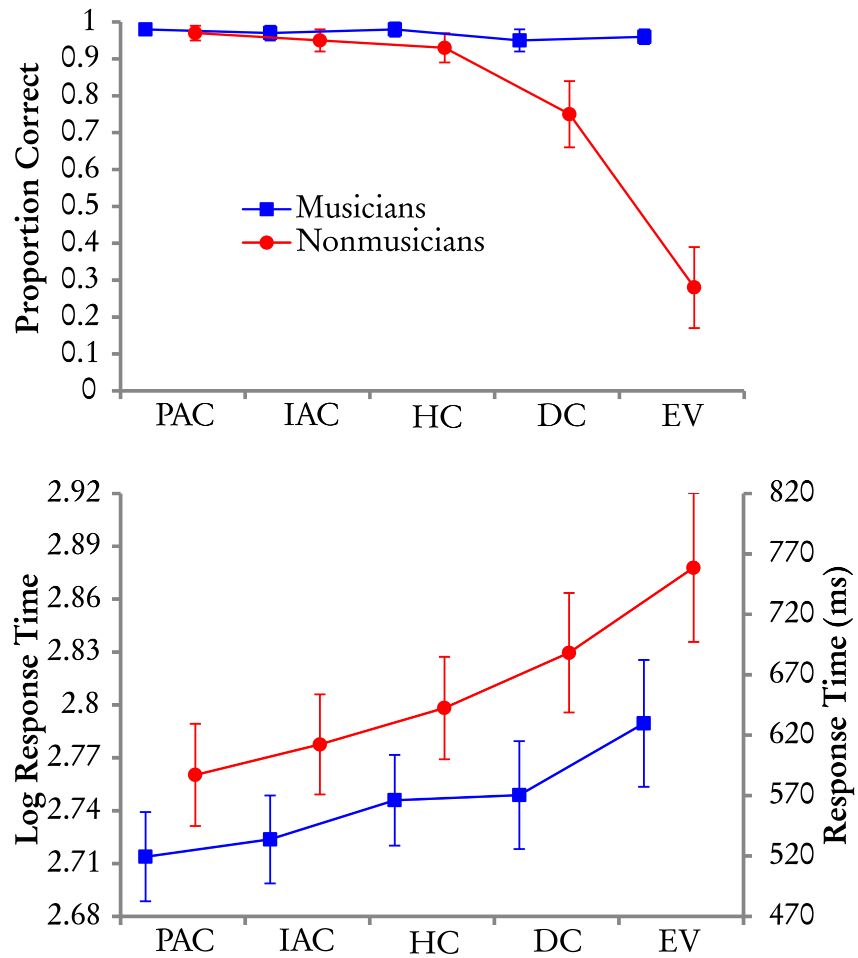


Figure 8.8: Line plots of the estimated mean proportion correct and response times of musicians and nonmusicians for each cadence category. Whiskers represent ±1 standard error.

was over five times higher than the estimated odds calculated across the entire experiment, $B = 1.73$, odds ratio $= 5.64$, $Z = 2.76$, $p < .05$. Pairwise comparisons of the nonmusician responses using the Tukey HSD test revealed a significant difference between the PAC–DC pair, and the EV category also differed significantly from the other cadence categories (see Figure 8.8). Polynomial contrasts of the nonmusician responses also revealed a significant decreasing linear trend from the PAC to EV categories, $B = -10.90$, $z = -5.25$, $p < .0001$.

Type III Wald F tests of the fixed effects from the 5×2 LMM of the correct logRTs revealed a significant effect of cadence category, $F(4, 37.54) = 3.43, p < .05$, and a marginal effect of training, $F(1, 28.48) = 3.59, p = .07$, but the interaction was not significant. Polynomial contrasts revealed a significant increasing linear trend in logRTs for both musicians, $B = .18, df = 35.47, t = 2.73, p < .01$, and nonmusicians, $B = .29, df = 50.78, t = 3.37, p < .01$, which again corresponds to the cadential hierarchy observed in the IC estimates in Chapter 6 and in Experiments I, II, and III in Chapters 7 and 8.

8.4.3 Discussion

Participants were faster and more accurate for the events at the cadential arrival from the genuine cadence categories compared to the cadential deviations. As expected, logRTs from both groups also demonstrated a significant increasing linear trend from the PAC to the EV categories, thereby replicating similar findings from Experiments I (see Figure 7.2), II (see Figure 7.6), and III (see Figure 8.1). In this experiment, target events in the authentic cadence categories (PAC, IAC) were primed, thus facilitating processing, while target events in the cadential deviations (DC, EV) presumably elicited inhibition effects resulting from unexpected harmonies and/or scale degrees in the target.

For the half cadence category, participant responses again appeared somewhere in the middle, indicating a processing benefit relative to the cadential deviations but a processing cost relative to the authentic cadence categories. Thus, whether participant logRTs for the half cadence category indicate somewhat weak facilitation or inhibition effects relative to those from the other cadence categories remains unclear. In psycholinguistic studies employing the priming paradigm, researchers typically determine whether the preceding context facilitates or inhibits target events by creating a baseline condition for which no priming effect occurs (i.e., where the preceding context is neither related nor unrelated to the target).[113] By comparing

[113]Tillmann et al., "Tonal Centers and Expectancy: Facilitation or Inhibition of Chords at the Top of the

the logRTs for the baseline condition against those from the selected cadence categories, we could then determine the processing benefit (or cost) for each category. In the absence of such a condition, however, we can only assume that the position of the half cadence category in the participant logRTs results from the presence of note and chord events associated with the dominant at the moment of cadential arrival, events which, according to scholars like Carol Krumhansl and Jamshed J. Bharucha, are generally less stable in the tonal system.[114]

In Experiments III and IV, the cadence categories elicited larger differences in the ratings of musicians compared to nonmusicians. In Experiment V, however, this trend was reversed, with the speed and accuracy of the responses from the nonmusician group demonstrating larger differences in the selected cadence categories relative to the musician group. Given the size of the effects reported here, the relatively weaker effects found in Experiments III and IV for the nonmusician group may have been due to the musician group's increased familiarity either with the experimental tasks, with technical terms relating to the experiments, or with the purpose(s) of the experiments themselves. To be sure, Bigand champions implicit behavioral tasks like the priming paradigm for precisely these reasons, noting that they can "determine the structures 'naturally' treated by the musical ear … without a conscious effort underpinned by explicit response strategies."[115] Thus, the implicit method employed in Experiment V demonstrates that priming effects occur in cadential contexts regardless of explicit musical training.

But how do we account for the formation of expectations during musical listening? Do the priming effects observed here result from implicit exposure over the course of many years, or from sensory processes accumulated over echoic memory? Or perhaps the sensory-cognitive apparatus combines the two processes in some way. The next section simulates the observed

Harmonic Hierarchy."

[114] Bharucha and Krumhansl, "The Representation of Harmonic Structure in Music"; Krumhansl, Bharucha, and Kessler, "Perceived Harmonic Structure of Chords in Three Related Musical Keys."

[115] Bigand, "More about the Musical Expertise of Musically Untrained Listeners," 305.

priming effects from Experiment V using the most current computational approaches in the priming literature.

§8.5 Simulations

Echoic Memory

Of the available sensory computational models of tonal expectancy,[116] Marc Leman's model of echoic memory (EM) is perhaps the most well known.[117] Leman's aim was to demonstrate that a model comparing the immediate pitch percept with the integrated pitch image computed over the window of echoic memory could explain tonal probe-tone judgments,[118] thus challenging cognitive accounts of tonal expectation.[119] Emmanuel Bigand and his co-authors have since simulated tonal priming effects for stimulus sets from 19 separate priming studies using the EM model, suggesting that processes related to echoic memory play a substantial role in the formation of tonal expectations.[120]

Shown in Figure 8.9, the model consists of four stages: (1) peripheral auditory system, (2) pitch periodicity analysis, (3) echoic memory, and (4) tonal contextuality. In the first stage, the EM model produces *auditory nerve images* (ANIs) that simulate the mechanisms by which the auditory peripheral nervous system transduces the acoustic signal into patterns of neural firing rate-codes in the auditory nerves. It first low-pass filters the acoustic signal to match

[116]For a review and discussion of these models, see Collins et al., "A Combined Model of Sensory and Cognitive Representations Underlying Tonal Expectations in Music," 39–43.

[117]Leman, "An Auditory Model of the Role of Short-Term Memory in Probe-Tone Ratings." It has also been called the *Tonal Contextuality* (TC) model, the *periodicity pitch* (PP) model, and the auditory short-term memory (ASTM) model in recent publications. To avoid confusion regarding the names of the models implemented here, and since I am only applying the model to simulate echoic memory, I will prefer "EM" model.

[118]Krumhansl and Kessler, "Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys."

[119]Krumhansl, *Cognitive Foundations of Musical Pitch*.

[120]Bigand et al., "Empirical Evidence for Musical Syntax Processing? Computer Simulations Reveal the Contribution of Auditory Short-Term Memory."

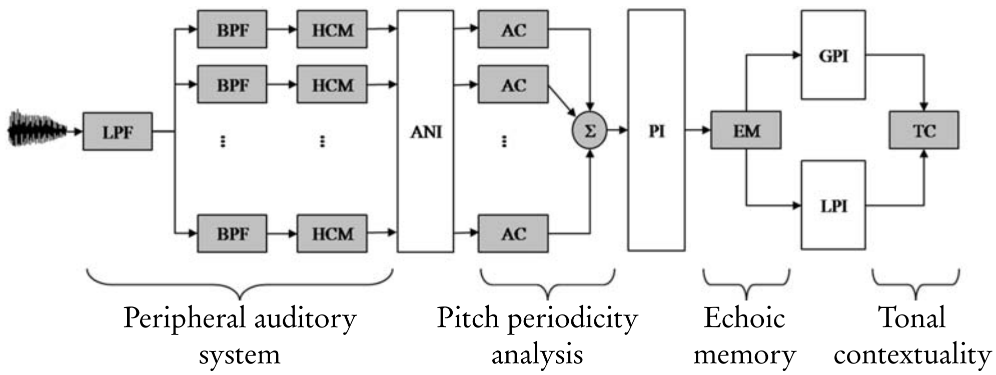


Figure 8.9: Schematic diagram of Leman's model of echoic memory (EM), reproduced from Bigand et. al's "Empirical Evidence for Musical Syntax Processing?" (2014), 6. LPF = lowpass filter; BPF = band-pass filters; HCM = hair cell models; ANI = auditory nerve image; AC = auto-correlation functions; PI = pitch image; EM = echoic memory; GPI = Global Pitch Image; LPI = Local Pitch Image; TC = Tonal Contextuality.

the filtering processes of the outer and middle ear, and then applies a bank of 40 band-pass filters (or *channels*) whose bandwidths correspond to critical bands along the basilar membrane. To complete the first stage, the EM model finally converts these band-pass-filtered channels into neural rate-code patterns using hair-cell models that amplify the signals using half-wave rectification and dynamic range compression.

In the second stage, the EM model produces a *pitch image* (PI) that represents the estimated periodicities in the firing patterns of the ANI channels in a range defined between 80 Hz and 1250 Hz. It first applies a windowed autocorrelation function that detects periods in each channel at shifted windows of 60 ms. A coincidence mechanism then sums these periodicities and stores the results in a summary pitch image. Thus, for a stimulus with a frequency component at 600 Hz, the ANI channel corresponding to 600 Hz will produce periods at multiples of $\frac{1}{600}$ in the resulting PI (e.g., 1.66, 3.33, 5, 6.66, 8.33 ms).[121]

In the third stage, the EM model incorporates effects of echoic memory. Using the PI as

[121] Ibid., 4.

input, Leman produces *echoic images* at both local and global time scales (LPI and GPI) using a leaky integrator, which updates the image at each time step t by adding a certain amount of the previous image to the new incoming pitch image, using an exponential half-decay function to simulate the decay in echoic memory over the duration of the selected time scale.[122] In the most recent studies employing the EM model,[123] the LPI was obtained using a short integration time of 0.1 s to represent the immediate pitch percept, while the GPI was obtained with a longer integration time of anywhere between 1.5 and 4 s to represent the global pitch percept stored in echoic memory.

Finally, the fourth stage estimates the similarity (or *tonal contextuality*) between the LPI and the GPI at each t using the Pearson correlation coefficient r. According to Bigand and his co-authors, the tonal contextuality index produced by the EM model represents the "tension" of the LPI with respect to the GPI, with high values indicating high correlations, and thus, low levels of tension.[124] Thus, the EM model produces a continuous time series of tonal contextuality values that reflect the relatedness between the immediate pitch percept and the preceding pitch context.

Tonal Space

Unlike Leman's EM model, the *Tonal Space* (TS) model assumes that tonal contexts are maintained in regions of the brain mediating interactions between sensory, cognitive, and affective information.[125] Thus, Petr Janata and his co-authors designed the TS model to account both for sensory *and* cognitive priming effects by projecting the output pitch images from Leman's

[122]Leman, "An Auditory Model of the Role of Short-Term Memory in Probe-Tone Ratings," 489.
[123]Bigand et al., "Empirical Evidence for Musical Syntax Processing? Computer Simulations Reveal the Contribution of Auditory Short-Term Memory"; Collins et al., "A Combined Model of Sensory and Cognitive Representations Underlying Tonal Expectations in Music."
[124]Bigand et al., "Empirical Evidence for Musical Syntax Processing? Computer Simulations Reveal the Contribution of Auditory Short-Term Memory," 6.
[125]Janata et al., "The Cortical Topography of Tonal Structures Underlying Western Music," 2169.

EM model to the surface of a torus using a self-organizing map (SOM) algorithm,[126] the central purpose of which is to learn and remember the statistical regularities governing complex stimulus domains like tonal music.[127] To be sure, Gjerdingen has suggested that connectionist models like the SOM algorithm "present an appealing analog of the ordinary listener, someone who without formal training has nevertheless developed a strong sense of how music 'works.' "[128] For this reason, the SOM algorithm has become the canonical method for simulating cognitive representations of tonal materials over the past three decades, with Bharucha's *MUSACT* model,[129] Gjerdingen's *L'ART pour l'art*,[130] and Tillmann's SOM model of harmonic perception providing ready examples.[131]

The mathematical details of SOMs have been described elsewhere,[132] but a brief summary of the approach will be useful to us here. A SOM is a hierarchical artificial neural network that learns the distributional information associated with a particular corpus without explicit teaching inputs. Conventional SOMs typically consist of an input layer, which represents the incoming stimuli, and a topological layer, which represents the neuronal activity of the cerebral cortex.[133] To simulate the tonotopic organization of sensory information in the cortex,[134] the topological layer is composed of n neurons (or *units*) positioned on a two-dimensional grid (or *map*) that are connected by synapses (or *connection strengths*) to the input layer such that

[126] Teuvo Kohonen, *Self-Organizing Maps* (Berlin, Germany: Springer, 1995).
[127] Tillmann, Bharucha, and Bigand, "Implicit Learning of Tonality," 891.
[128] Robert O. Gjerdingen, "Using Connectionist Models to Explore Complex Musical Patterns," *Computer Music Journal* 13, no. 3 (1989): 67.
[129] Jamshed J. Bharucha, "MUSACT: A Connectionist Model of Musical Harmony," in *Proceedings of the Cognitive Science Society* (Hillsdale, NJ: Erlbaum Press, 1987).
[130] Gjerdingen, "Using Connectionist Models to Explore Complex Musical Patterns"; Robert O. Gjerdingen, "Categorization of Musical Patterns by Self-Organizing Neuronlike Networks," *Music Perception* 7, no. 4 (1990): 339–370.
[131] Tillmann, Bharucha, and Bigand, "Implicit Learning of Tonality."
[132] For a thorough review of SOMs in the priming literature, see ibid., 887–891.
[133] Charles Delbé, "Recurrent Self-Organization of Sensory Signals in the Auditory Domain," in *From Associations to Rules: Connectionist Models of Behavior and Cognition*, ed. Robert M. French and Elizabeth Thomas (Singapore: World Scientific, 2008), 184.
[134] Tillmann, Bharucha, and Bigand, "Implicit Learning of Tonality," 891.

similar stimulus inputs will elicit maximal responses from *nearby* units (i.e., units located near each other on the map). If the sides of this map share the same unit, the global shape of the map becomes a cylinder or a toroid, as is the case with the SOM developed by Janata and his co-authors.[135]

The original algorithm is based on competitive learning, with each unit corresponding to a prototype or reference vector, which could represent a note, a triad, a harmonic progression, or nearly any other feature deemed noteworthy by the experimenter. In the case of the TS model, the prototype vectors represent the 24 major and minor keys of the tonal system. For each stimulus input, the algorithm measures the Euclidean distance between the input vector and the prototype vectors representing all of the units on the map. The prototype vector with the smallest distance value then wins the competition. During training, the connection strength between the input stimulus and the winning neuron is then updated such that future activations of that unit will be stronger for similar input stimuli and weaker for dissimilar input stimuli. Perhaps more importantly, the input stimulus also activates the other units on the map, albeit more weakly, thereby reorganizing the order of the units to reflect their degree of activation. When training is complete, each unit is specialized to represent a particular feature of the input stimuli (e.g., a particular tonal context), and the map is topographically organized such that input stimuli with similar features will activate nearby units on the map (e.g., excerpts in C major will be closer to those in G major than to those in F♯ major).

When presented with a harmonic progression in C major, the output pitch image from the TS model triggers a cascade of spreading activation across the map at each time t, with the highest activation appearing at the unit representing the key of C major, followed by weaker activations for neighboring units like G major and A minor, and yet weaker activations for less proximal units like D major. For the TS model, the authors trained the SOM using the

[135] Delbé, "Recurrent Self-Organization of Sensory Signals in the Auditory Domain," 185.

pitch images from the EM model integrated with a 2 s time constant that were extracted from a melody that was explicitly composed to modulate through all 24 major and minor keys over the course of approximately 8 minutes.[136] Like Leman's EM model, Janata and his co-authors also employ an exponential decay function that integrates the activation patterns over time to incorporate effects of echoic memory. Thus, at any point during the stimulus input, the relative activations across the map that have been accumulated over echoic memory represent the effects of long-term schematic knowledge on the tonal expectancies of listeners.

Information Dynamics of Music

Although the TS model simulates the effects of long-term schematic knowledge on tonal expectancies, it fails to consider whether a SOM can account for tonal priming effects in isolation. What is more, SOMs have been criticized in recent years because they generally do not represent the contextual relations between contiguous events on the musical surface.[137] To be sure, although the gradual memory decay included in the TS model allows the analyst to consider long-range temporal dependencies,[138] SOMs including such a decay essentially model how a structural account of tonal expectations varies over time.[139] In other words, unlike Markov models, conventional SOMs like those used in tonal priming studies do not explicitly model the sequential dependencies between events in tonal music.[140] Tom Collins and his co-authors argue, for example, that priming experiments to date have mainly tested musical expectations based on zeroth-order probabilities when higher-order models are more

[136] Collins et al., "A Combined Model of Sensory and Cognitive Representations Underlying Tonal Expectations in Music," 42.

[137] Ibid.

[138] John Ashley Burgoyne, "Stochastic Processes and Database-Driven Musicology" (PhD Dissertation, McGill University, 2012), 80.

[139] See Footnote 62.

[140] Researchers are beginning to consider the potential for *recursive* SOMs to model tonal expectancies over time, the purpose of which is to represent temporal *sequences* as opposed to zeroth-order distributions. See, for example, Delbé, "Recurrent Self-Organization of Sensory Signals in the Auditory Domain."

appropriate. They write,

> Though the importance of harmonic syntax for describing the tonal trajectory of a piece of music and associated expectations is undisputed by music theorists . . . , explicit consideration of the perception of chord transitions and the locations of those transitions within phrasal structure has yet to gain momentum in the field of music psychology.[141]

In this regard, functional interpretations of cognitive priming effects—such as the Information Dynamics of Music (IDyOM) model described in Chapter 6—might offer a suitable alternative to artificial neural networks like the SOM algorithm. Thus, in what follows I have also simulated the priming effects for the stimulus set employed in Experiment V using IDyOM.

8.5.1 Method

For the EM and TS models, the local pitch image was set to .1 s and the global pitch image was set to 4 s. Following Collins and his co-authors, Pearson's r was computed over each stimulus from Experiment V at a sampling rate of 26 Hz for both models.[142] To obtain a single *tonal congruency* estimate for each stimulus comparing the target chord event with the preceding context, the correlation time series for the EM and TS models were averaged over the time window corresponding to the target chord at the expected moment of cadential arrival.

For IDyOM, providing a suitable training corpus of Mozart keyboard sonatas was beyond the scope of the present study, so for the sake of simplicity I have trained the LTM+ model from IDyOM on the note events from the first violin in the Haydn Corpus under the assumption that the sequential dependencies between melodic events in Haydn's compositional style will roughly

[141] Collins et al., "A Combined Model of Sensory and Cognitive Representations Underlying Tonal Expectations in Music," 54.
[142] Ibid.

correspond to those found in the soprano-voice melodies in Mozart's keyboard sonatas. Given the pronounced differences in texture between the string quartet and piano genres, however, the present investigation does not consider bass-line or harmonic expectancies. Thus, the EM and TS models were computed over the entire audio stimulus, but IDyOM was only computed over the soprano-voice melody from each stimulus. To examine melodic pitch expectations for the resolving note event in the melody at the moment of cadential arrival, I derived the viewpoint combination from the stepwise selection procedure described in §6.2.3. As before, viewpoint selection derived the following combination of viewpoint models: `melint`, followed by the linked viewpoint `csd` ⊗ `cpitch`, which I will again refer to as `selection`.

8.5.2 Results

To examine the effect of cadences that achieve genuine thematic closure—PAC, IAC, and HC—against those cadences that represent deviations of the authentic cadence—DC and EV—the model simulations were first analyzed for the factor variable *ending* with a two-tailed independent samples t-test. Levene's test revealed heteroscedastic groups for one of the models, however, so I report t-tests that do not assume equal variances. Shown in Figure 8.10, the EM and TS models did not produce higher correlation estimates for the genuine cadences compared to the cadential deviations (EM, $t(25.83) = .44, p > .05$; TS, $t(37.03) = .12, p > .05$). Note that the y-axis is upside down so that decreasing estimates correspond to increasing logRTs. For IDyOM, however, the cadential deviations received significantly higher IC estimates relative to the genuine cadences, $t(16.95) = -3.02, p < .01$.

Figure 8.11 presents line plots of the simulation estimates for the EM, TS, and IDyOM models for each cadence category. To examine the effect of the cadence categories individually, one-way ANOVAs were specified for each model simulation. Again, Levene's test revealed heteroscedastic groups for two of the three models, so I report Welch's F test here and estimate

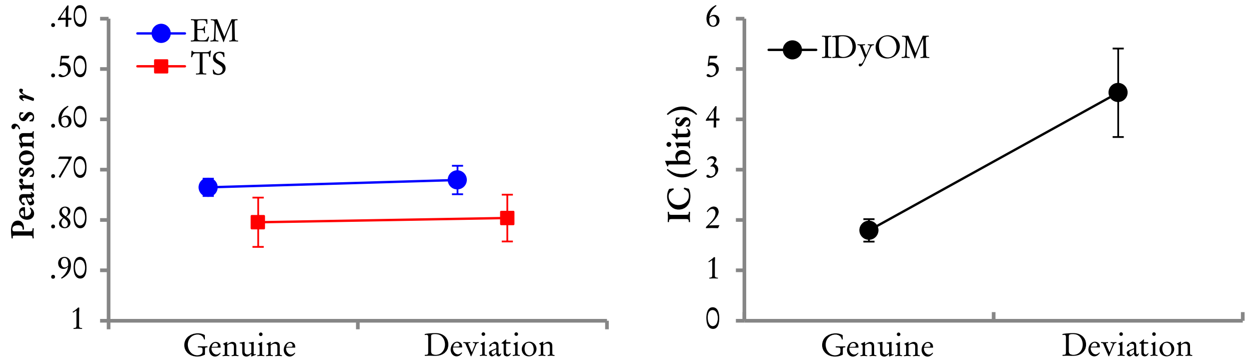


Figure 8.10: Left: Line plots of the correlation estimates from the EM and TS models for each ending. The y-axis is upside down so that decreasing estimates correspond to increasing logRTs. Right: Line plot of the information content estimates from IDyOM for each ending. Whiskers represent ±1 standard error.

effect size using est. ω^2 (see §6.3.1). To address whether the model simulations corresponded with the linear increase demonstrated in the logRTs in Experiment V, a polynomial contrast was also included that estimates the goodness-of-fit of the predicted trend.

A one-way ANOVA for the correlation estimates did not reveal a significant effect of cadence category for the EM model, $F(4, 17.24) = 1.98, p > .05$, est. $\omega^2 = .09$, and the polynomial contrast did not exhibit the increasing linear trend reflected in the logRTs, $B = -.10, df = 15.33, t = -.90, p > .05$. For the TS model, however, the mean correlation estimates demonstrated a significant effect of cadence category, $F(4, 15.61) = 3.02, p < .05$, est. $\omega^2 = .17$, but with excerpts from the HC category receiving the lowest estimates on average, $M = .61, SE = .12$. As a consequence, the polynomial contrast was not significant, $B = -.35, df = 8.43, t = -1.90, p = .09$. Finally, the IC estimates from IDyOM revealed a significant effect of cadence category, $F(4, 14.87) = 14.05, p > .001$, est. $\omega^2 = .57$, and the polynomial contrast exhibited a significant increasing linear trend from the PAC to the EV categories, $B = 12.39, df = 7.53, t = 6.85, p > .001$.

To this point I have only considered whether the model simulations correspond with the

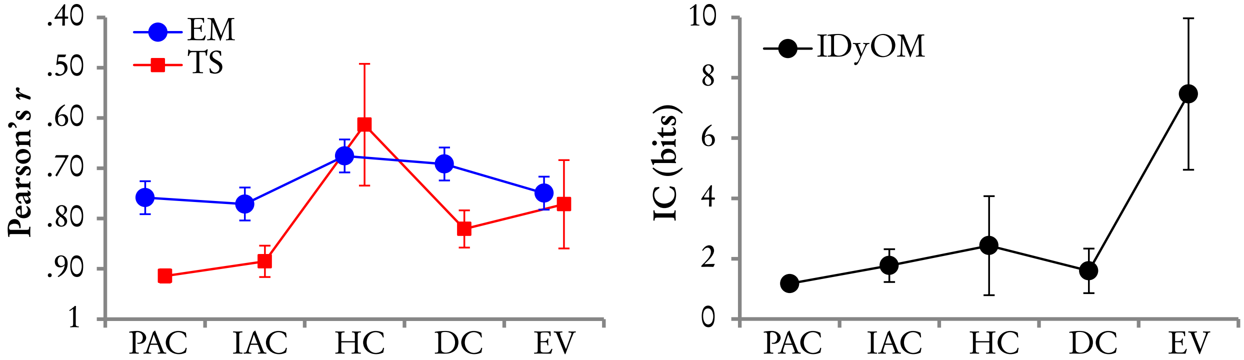


Figure 8.11: Left: Line plots of the correlation estimates from the EM and TS models for each cadence category. The y-axis is upside down so that decreasing estimates correspond to increasing logRTs. Right: Line plot of the information content estimates from IDyOM for each cadence category. Whiskers represent ±1 standard error.

effects of cadence category observed in the participant logRTs. To examine the relationship between the model simulations and the participant responses specifically, Table 8.3 presents the intercorrelations between the model simulations and the mean fit ratings and logRTs collected in Experiments III and V, respectively. As expected, the pooled fit ratings and logRTs were significantly correlated, $r(38) = -.51, p < .01$, which indicates that participants were faster for target chords at the cadential arrival if they received the highest fit ratings.

For the model simulations, the EM and TS models demonstrated just one significant correlation with the participant responses: musicians were faster for excerpts that received high correlation estimates from the TS model, $r(38) = -.33, p < .05$. This correlation was weak, however, explaining only 11% of the variance in the musician ratings, and no other significant correlations emerged. Conversely, IDyOM featured moderate-to-strong correlations with every response variable, r^2 range: .17-.50: in each case, participants were faster and provided higher fit ratings for target events that received lower IC estimates by IDyOM.

Table 8.3: Intercorrelations between the model simulations and the mean logRTs (Experiment V) and fit ratings (Experiment III) of musicians and nonmusicians.

	LogRTs (Experiment V)			**Fit Ratings (Experiment III)**			**Simulations**		
	Mus.	Nonmus.	Pooled	Mus.	Nonmus.	Pooled	EM	TS	IDyOM
LogRTs									
Musicians		$.60^{***}$	$.92^{***}$	$-.59^{***}$	$-.39^{*}$	$-.54^{***}$	$-.28$	$-.33^{*}$	$.71^{***}$
Nonmusicians			$.88^{***}$	$-.34^{*}$	$-.30$	$-.34^{*}$	$.18$	$.13$	$.41^{*}$
Pooled				$-.55^{***}$	$-.38^{*}$	$-.51^{**}$	$-.18$	$-.24$	$.63^{***}$
Fit Ratings									
Musicians					$.86^{***}$	$.98^{***}$	$.15$	$.16$	$-.64^{***}$
Nonmusicians						$.95^{***}$	$.14$	$.14$	$-.60^{***}$
Pooled							$.15$	$.16$	$-.64^{***}$
Simulations									
EM								$.66^{***}$	$-.10$
TS									$-.30$
IDyOM									

Note. $N = 40$.
$^{*}\ p < .05$ $^{**}\ p < .01$ $^{***}\ p < .001$, two-tailed.

8.5.3 Discussion

Of the model simulations examined here, IDyOM provided the best fit with the experimental data from Experiments III and V. What is more, the IC estimates shown in Figure 8.11 correspond quite well with the IC estimates calculated for the cadences in the Haydn Corpus (see Figure 6.3). In both cases, deceptive cadences received much lower IC estimates on average than evaded categories, presumably because the former category features a stable scale degree in the melody at the cadential arrival (e.g., $\hat{1}$ or $\hat{3}$), whereas the latter category typically features an unexpected leap. To be sure, when considered in isolation, melodic expectations for the moment of cadential arrival in deceptive cadential contexts should not differ from those in authentic cadential contexts, as is the case here. Conversely, the unexpected leap characterizing the evaded cadences employed in Experiments II–V may explain the higher IC estimates observed for excerpts from that category.

To obtain the increasing linear trend found for the cadence collection in Chapter 6 using IDyOM, recall that I created a `composite` viewpoint that combined the IC estimates from $\texttt{selection}_{\mathrm{vl1}}$ and a harmonic viewpoint called `csdc`. In that case, a linear combination of melodic and harmonic expectations reflected the increasing linear trend that has since been observed in the completion ratings from Experiments I and II, the ratings of expectancy fit from Experiment III, and the logRTs from Experiment V. Thus, the model simulations provided by IDyOM might better reflect the corresponding trend in the logRTs if I also include a viewpoint representing the formation of harmonic expectancies during listening.

Nevertheless, the effects reported here provide evidence in support of a functional interpretation of tonal processing, in which listeners with exposure to tonal music retain long-term, schematic knowledge about the statistical dependencies between contiguous events. This knowledge allows listeners to generate expectations during music listening, with the syntactic relationships between tonal events activating schematic representations that either facilitate or inhibit the processing of continuations heard later. This is not to say that sensory or psychoacoustic explanations of the priming effects observed here—or reported elsewhere—play no role in expectancy formation; only that the sensory account considered in isolation generally fails to account for the priming effects demonstrated in these experiments. What is more, given the paucity of experimental studies examining tonal expectancies for genuine musical materials, these findings might motivate researchers to consider stimulus sets that will better reflect the musical styles and genres to which listeners might be consistently exposed.

§8.6 Conclusions

This chapter examined the link between expectancy and cadential closure in the keyboard sonatas of Wolfgang Amadeus Mozart. In §8.1 I reviewed the evidence for tonal expectancies during music listening, noting that genuine musical materials rarely appear in the experimental literature. In Experiment III (§8.2), participants provided the lowest strength and specificity ratings for excerpts from the HC category. When the terminal events at the moment of cadential arrival followed the preceding context (i.e., the non-truncated condition), however, half cadences were placed in the center of the expectancy fit scale, suggesting that listener expectations reflect a bipolar continuum, with fulfillment on one end and violation on the other, and where distance from the center of the scale corresponds to the strength of the experienced expectation. In Experiment IV (§8.3), continuous ratings of expectations for closure reinforced the interpretation of results from Experiment III, with excerpts from the PAC and HC categories receiving the highest/earliest and lowest/latest maximum ratings, respectively. Finally, Experiment V (see §8.4) demonstrated facilitation effects in the correct RTs of both musicians and nonmusicians for excerpts from the PAC category, suggesting that authentic cadential contexts prime listeners to expect a root-position tonic harmony with $\hat{1}$ in the soprano. What is more, the mean logRTs across all five cadence categories demonstrated the same ascending linear trend that was previously observed in the model simulations in Chapter 6, as well as the completion and expectancy fit ratings in Experiments I, II, and III. Finally, in §8.5, the model simulations provided by sensory or psychoacoustic accounts of tonal priming generally failed to explain the pattern of results observed here. Instead, the functional account represented by IDyOM provided the best fit to the data, suggesting that listeners generate expectations for potential continuations as a consequence of the frequent co-occurrence of events on the musical surface.

Conclusions

> *The constraints of a style are learned by composers and performers, critics and listeners. Usually such learning is largely the result of experience in performing and listening rather than of explicit formal instruction in music theory, history, or composition. In other words, knowledge of a style is usually "tacit": that is, a matter of habits properly acquired (internalized) and appropriately brought into play... It is the goal of music theorists and style analysts to explain what the composer, performer, and listener know in this tacit way... This can be done only by making inferences from observable data—the replicated patternings present in works of art—to general principles.*
>
> LEONARD B. MEYER

I have argued in this thesis that listeners who are familiar with classical music have internalized the most common cadence categories as a flexible network of rival closing schemata. Once learned, the activation of this network during music listening—and of the individual closing schemata contained within—results in the formation of expectations for the terminal events of the cadence, the fulfillment of which serves to close off both the schema itself and, in many cases, the larger phrase-structural process that subsumes it. To support this view, the corpus-analytic studies in Part II presented a few analytical techniques for the discovery, classification, and prediction of cadences from Caplin's typology that might simulate the learning mechanisms underlying human cognition. The experimental studies in Part III extended these findings to human listeners by examining the psychological relevance of existing theoretical models of the classical cadence.

In some ways, this particular study took a deductive path, beginning with the categories identified in Caplin's theory of cadential closure and then working down to the parameters at the musical surface that characterize those categories. But this is not to say one could not have taken an inductive approach, whereby the musical parameters dictate the number and position of the categories within the schematic network. In Chapter 4, for example, I extended the canonical *n*-gram approach to include non-contiguous sequences and found that cadential progressions like ii^6–Cad^6_4–V^7–I are among the most frequent patterns in the Haydn Corpus. Nevertheless, in large part my choice was to examine whether exemplars of the cadence categories advanced in the "New *Formenlehre*" tradition influence listeners in some meaningful way. As a consequence, this study did not consider those musical passages that defy ready categorization in traditional theories of cadential closure, yet still elicit an ending percept. In Chapter 5, for example, I excluded ten cadences from the cadence collection because they implied more than one category (i.e., PAC-EV or DC-EV). By abandoning the cadential categories entirely, we might therefore consider the entire range of musical parameters responsible for the perception of closure without recourse to theories of cadence, which attempt to reveal the procedures by which composers articulated closing patterns in the classical style, but which do not always directly correspond with the schematic knowledge of listeners. Indeed, what is essential for a theory of cadence may not always be tenable for a psychological theory of closure.[1] Certainly, empiricism provides a method for applying constraints to our theoretical models, weeding out the impossible from the possible, but the concept of closure advanced in theories of cadential closure need not dispense with an examination of compositional procedures in favor of exclusively explaining how listeners perceive and process closing patterns.

However, the desire to explain or clarify how we experience musical endings can provide

[1]Nicholas Cook makes the same point in reference to the perception of tonal closure ("The Perception of Large-Scale Tonal Closure," *Music Perception* 5, no. 2 [1987]: 205).

common ground for further cross-disciplinary work. In my view, the application of experimental methods in future studies offers researchers the opportunity to gain a more complete understanding of the underlying sensory and cognitive mechanisms responsible for the perception of closure in music. Indeed, these methods seem especially worthwhile if we accept the view advocated by scholars in the learning sciences that the first-time listening is never "completely innocent of analysis."[2] As Carol Krumhansl points out,

> ... musical knowledge is highly constrained, stable, and permanent. It is presumed to be the consequence of years of exposure to a system of music, which is itself shaped at a very basic level by acoustic properties and the way they are encoded by the sensory system. Once acquired, this knowledge is thought to impose its organization on all subsequent musical experiences.[3]

This is not to say that schematic knowledge is fixed across a group of listeners; the depth (or specificity) of that knowledge will vary from one person to another. When confronted with a Cudworth Cadence, for example, listeners with relatively little exposure to this galant mannerism may hear a V–I progression, while those with a great deal of experience in the instrumental repertories of Pergolesi, Scarlatti, and Haydn may possess a schematic representation that is nearly isomorphic with the encountered exemplar. Finally, those listeners with no familiarity with Western tonal music might have to rely entirely on the biological constraints of the auditory periphery, and so may hear only a 'blooming, buzzing confusion.'[4]

But no matter the depth or specificity of the representation, the important point here is that much of this knowledge lies beneath the conscious surface, reflects implicit rather than explicit learning strategies, and goes far beyond verbal description. Jay Dowling summarizes

[2]Edward T. Cone, "Three Ways of Reading a Detective Story or a Brahms Intermezzo," *The Georgia Review* 31, no. 3 (1977): 80.

[3]Krumhansl, *Cognitive Foundations of Musical Pitch*, 284.

[4]James, *The Principles of Psychology*, 1:488.

this view thusly:

> We tend to think of what we "really" know as what we can talk about, and to disparage knowledge that we can't verbalize. When we possess two representations of a musical structure, one declarative and the other procedural, we tend to prefer the declarative one because of its accessibility to theorizing and formal manipulation. We must come to realize that most of our brain representations of musical structure are first developed through years of perceptual learning in listening to and performing music, and that the corresponding declarative representations are typically in the form of rationalizations at the conscious level of subtler and richer implicit representations at the subconscious level. At the conscious level we inevitably discard information in the interests of clarity of formalization—information that the brain "knows" procedurally to be important and does not forget.[5]

If we share Leonard Meyer's view that the goal of music researchers is "to explain what the composer, performer, and listener know in this tacit way,"[6] then it seems reasonable to conclude that the cadence concept is merely a declarative shorthand for those features of music that so exquisitely exploit the sensory-cognitive apparatus.

[5]Dowling and Harwood, *Music Cognition*, 252. Eric Clarke has similarly argued that "we live in a tremendously logocentric culture, in which our capacity to express ourselves in language seems sometimes to be regarded as virtually synonymous with knowledge" ("Issues in Language and Music," *Contemporary Music Review* 4 [1989]: 10).

[6]Meyer, *Style and Music: Theory, History, and Ideology*, 10.

Appendix A

The Haydn Corpus Cadence Collection

The Haydn Corpus consists of midi representations of 50 sonata-form expositions from Haydn's string quartets (see Table 3.1), as well as accompanying text files that include the position of each prominent key area, the boundaries of thematic functions—Main Theme (MT), Transition (TR), and Subordinate Theme (ST)—and a number of annotations about the cadences appearing in each movement.

The cadence collection consists of exemplars of the five cadence categories in the Haydn Corpus that achieve, or promise, cadential arrival in Caplin's cadence typology—PAC, IAC, HC, DC, EV (see Table 2.1). In the table below, I have included the cadential classification and bass clausula for each cadence (see Chapter 5), and I have also indicated whether each cadence includes a cadential six-four, an expanded cadential progression (ECP), a surface dissonance at cadential arrival (e.g., a 4-3 suspension), and a trill above the penultimate dominant. The electronic corpus also includes a number of other annotations for each cadence, such as the harmonies of the cadential progression, the duration of the cadential progression expressed as a percentage of the length of the total movement, whether the events at the moment of cadential arrival *elide* with the following passage, and so on.

Number	Cadence	Bass Clausula	Cad $^6_4$	ECP	Diss.	Trill	Excerpt
1	PAC	Semplice	No	Yes	No	No	Op. 17, No. 1, i, mm. 4-6
2	PAC	Semplice	No	No	No	No	Op. 17, No. 3, iv, m. 4
3	PAC	Semplice	No	No	No	No	Op. 17, No. 4, i, mm. 6-8
4	PAC	Semplice	No	Yes	No	No	Op. 17, No. 5, i, mm. 10-12
5	PAC	Semplice	No	No	No	No	Op. 17, No. 5, i, mm. 24-25
6	PAC	Semplice	No	No	No	No	Op. 17, No. 6, i, mm. 22-24
7	PAC	Semplice	No	No	No	No	Op. 20, No. 1, iv, mm. 5-6
8	PAC	Semplice	No	No	No	No	Op. 20, No. 3, i, mm. 5-7
9	PAC	Semplice	No	Yes	No	No	Op. 20, No. 3, iii, mm. 35-41
10	PAC	Semplice	No	No	No	No	Op. 20, No. 4, i, mm. 66-67
11	PAC	Semplice	No	No	No	No	Op. 20, No. 4, iv, mm. 5-6
12	PAC	Semplice	No	No	No	No	Op. 20, No. 4, iv, mm. 39-40
13	PAC	Semplice	No	No	No	No	Op. 20, No. 5, i, mm. 4-5
14	PAC	Semplice	No	No	No	No	Op. 20, No. 5, i, mm. 30-31
15	PAC	Semplice	No	No	No	No	Op. 20, No. 5, i, mm. 45-46
16	PAC	Semplice	No	No	Yes	No	Op. 20, No. 6, ii, mm. 6-8
17	PAC	Semplice	No	No	No	No	Op. 33, No. 2, i, mm. 3-4
18	PAC	Semplice	No	No	No	No	Op. 33, No. 2, i, mm. 11-12
19	PAC	Semplice	No	No	No	No	Op. 33, No. 2, i, mm. 20-21
20	PAC	Semplice	No	Yes	No	No	Op. 33, No. 3, iii, mm. 14-17
21	PAC	Semplice	No	No	No	No	Op. 33, No. 3, iii, mm. 25-27
22	PAC	Semplice	No	No	No	No	Op. 33, No. 5, i, mm. 7-10
23	PAC	Semplice	No	No	No	No	Op. 33, No. 5, i, mm. 29-32
24	PAC	Semplice	No	No	No	No	Op. 50, No. 1, i, mm. 11-12
25	PAC	Semplice	No	Yes	No	No	Op. 50, No. 1, i, mm. 18-21
26	PAC	Semplice	No	Yes	No	Yes	Op. 50, No. 2, i, mm. 93-100
27	PAC	Semplice	No	No	No	No	Op. 50, No. 3, iv, mm. 46-47
28	PAC	Semplice	No	Yes	No	No	Op. 50, No. 5, i, mm. 5-8
29	PAC	Semplice	No	Yes	No	No	Op. 50, No. 5, i, mm. 47-50
30	PAC	Semplice	No	Yes	No	No	Op. 50, No. 5, iv, mm. 9-12
31	PAC	Semplice	No	Yes	No	No	Op. 50, No. 6, i, mm. 11-16
32	PAC	Semplice	No	No	No	No	Op. 50, No. 6, ii, m. 4
33	PAC	Semplice	No	No	No	Yes	Op. 54, No. 1, i, mm. 43-44
34	PAC	Semplice	No	Yes	Yes	No	Op. 55, No. 3, i, mm. 64-67
35	PAC	Semplice	No	No	No	No	Op. 64, No. 3, i, mm. 3-5
36	PAC	Semplice	No	No	Yes	No	Op. 64, No. 4, iv, mm. 15-16
37	PAC	Semplice	No	No	No	No	Op. 64, No. 4, iv, mm. 58-60
38	PAC	Semplice	No	No	No	No	Op. 64, No. 6, i, mm. 7-8
39	PAC	Semplice	No	Yes	No	Yes	Op. 71, No. 1, i, mm. 62-66
40	PAC	Composta	Yes	No	No	Yes	Op. 17, No. 1, i, mm. 39-41
41	PAC	Composta	Yes	No	No	Yes	Op. 17, No. 2, i, mm. 34-36
42	PAC	Composta	Yes	No	No	No	Op. 17, No. 3, iv, mm. 16-18
43	PAC	Composta	Yes	No	No	No	Op. 17, No. 4, i, mm. 49-51
44	PAC	Composta	Yes	Yes	No	Yes	Op. 17, No. 5, i, mm. 30-32
45	PAC	Composta	Yes	No	No	Yes	Op. 17, No. 6, i, mm. 60-62
46	PAC	Composta	Yes	Yes	No	No	Op. 20, No. 1, iv, mm. 38-48

Number	Cadence	Bass Clausula	Cad $^6_4$	ECP	Diss.	Trill	Excerpt
47	PAC	Composta	Yes	Yes	No	Yes	Op. 20, No. 3, iv, mm. 35-38
48	PAC	Composta	Yes	Yes	No	No	Op. 20, No. 4, i, mm. 78-87
49	PAC	Composta	Yes	Yes	No	No	Op. 20, No. 4, i, mm. 95-99
50	PAC	Composta	Yes	Yes	No	Yes	Op. 20, No. 6, ii, mm. 21-25
51	PAC	Composta	Yes	No	No	No	Op. 33, No. 1, i, mm. 10-11
52	PAC	Composta	Yes	No	No	Yes	Op. 33, No. 1, i, mm. 31-33
53	PAC	Composta	Yes	No	No	No	Op. 33, No. 1, iii, mm. 29-30
54	PAC	Composta	Yes	No	No	No	Op. 33, No. 1, iii, mm. 37-38
55	PAC	Composta	Yes	No	No	Yes	Op. 33, No. 2, i, mm. 26-28
56	PAC	Composta	Yes	No	Yes	No	Op. 33, No. 3, iii, mm. 6-8
57	PAC	Composta	Yes	No	No	Yes	Op. 33, No. 4, i, mm. 25-26
58	PAC	Composta	Yes	Yes	No	Yes	Op. 33, No. 5, i, mm. 83-89
59	PAC	Composta	Yes	No	No	Yes	Op. 33, No. 5, ii, mm. 7-8
60	PAC	Composta	Yes	No	No	No	Op. 50, No. 1, i, mm. 49-50
61	PAC	Composta	Yes	Yes	No	Yes	Op. 50, No. 1, i, mm. 51-56
62	PAC	Composta	Yes	No	No	No	Op. 50, No. 1, iv, mm. 68-70
63	PAC	Composta	Yes	No	No	No	Op. 50, No. 2, i, mm. 7-9
64	PAC	Composta	Yes	No	No	No	Op. 50, No. 2, i, mm. 27-29
65	PAC	Composta	Yes	No	No	No	Op. 50, No. 2, iv, mm. 48-50
66	PAC	Composta	Yes	No	No	Yes	Op. 50, No. 2, iv, mm. 74-77
67	PAC	Composta	Yes	No	No	No	Op. 50, No. 3, iv, mm. 22-24
68	PAC	Composta	Yes	Yes	No	Yes	Op. 50, No. 3, iv, mm. 54-65
69	PAC	Composta	Yes	Yes	No	Yes	Op. 50, No. 4, i, mm. 51-56
70	PAC	Composta	Yes	Yes	No	No	Op. 50, No. 5, i, mm. 17-20
71	PAC	Composta	Yes	Yes	No	No	Op. 50, No. 5, i, mm. 51-54
72	PAC	Composta	Yes	No	No	Yes	Op. 50, No. 5, i, mm. 60-63
73	PAC	Composta	Yes	No	No	Yes	Op. 50, No. 5, iv, mm. 43-45
74	PAC	Composta	Yes	Yes	No	Yes	Op. 50, No. 6, i, mm. 43-48
75	PAC	Composta	Yes	Yes	No	Yes	Op. 50, No. 6, ii, mm. 16-22
76	PAC	Composta	Yes	No	No	No	Op. 54, No. 1, i, mm. 10-13
77	PAC	Composta	Yes	Yes	No	Yes	Op. 54, No. 1, i, mm. 37-40
78	PAC	Composta	Yes	Yes	No	No	Op. 54, No. 1, ii, mm. 17-20
79	PAC	Composta	Yes	Yes	No	Yes	Op. 54, No. 1, ii, mm. 31-34
80	PAC	Composta	Yes	Yes	No	Yes	Op. 54, No. 1, ii, mm. 46-52
81	PAC	Composta	Yes	No	No	No	Op. 54, No. 2, i, mm. 23-25
82	PAC	Composta	Yes	No	No	No	Op. 54, No. 2, i, mm. 61-63
83	PAC	Composta	Yes	No	No	No	Op. 54, No. 2, i, mm. 71-73
84	PAC	Composta	Yes	No	No	No	Op. 54, No. 2, i, mm. 78-81
85	PAC	Composta	Yes	No	No	No	Op. 54, No. 3, i, mm. 7-8
86	PAC	Composta	Yes	No	No	No	Op. 54, No. 3, iv, mm. 14-16
87	PAC	Composta	Yes	Yes	No	Yes	Op. 54, No. 3, iv, mm. 63-72
88	PAC	Composta	Yes	No	No	No	Op. 55, No. 2, ii, mm. 15-16
89	PAC	Composta	Yes	No	No	No	Op. 55, No. 2, ii, mm. 46-48
90	PAC	Composta	Yes	Yes	No	No	Op. 55, No. 2, ii, mm. 56-60
91	PAC	Composta	Yes	Yes	No	No	Op. 55, No. 2, ii, mm. 65-71
92	PAC	Composta	Yes	No	No	No	Op. 55, No. 3, i, mm. 14-16
93	PAC	Composta	Yes	Yes	No	No	Op. 64, No. 3, iv, mm. 7-10

Number	Cadence	Bass Clausula	Cad $^6_4$	ECP	Diss.	Trill	Excerpt
94	PAC	Composta	Yes	No	No	No	Op. 64, No. 3, iv, mm. 65-67
95	PAC	Composta	Yes	No	No	No	Op. 64, No. 4, i, mm. 7-8
96	PAC	Composta	Yes	No	No	No	Op. 64, No. 4, i, mm. 30-32
97	PAC	Composta	Yes	No	No	No	Op. 64, No. 4, i, mm. 34-37
98	PAC	Composta	Yes	Yes	No	No	Op. 71, No. 1, i, mm. 27-30
99	PAC	Composta	Yes	No	No	No	Op. 71, No. 1, i, mm. 49-51
100	PAC	Composta	Yes	Yes	No	Yes	Op. 74, No. 1, i, mm. 13-18
101	PAC	Composta	Yes	Yes	No	Yes	Op. 74, No. 1, i, mm. 48-52
102	PAC	Composta	Yes	No	No	No	Op. 74, No. 1, ii, mm. 49-51
103	PAC	Composta	Yes	No	No	No	Op. 76, No. 2, i, mm. 11-12
104	PAC	Composta	Yes	No	No	No	Op. 76, No. 4, i, mm. 20-22
105	PAC	Composta	Yes	Yes	No	Yes	Op. 76, No. 4, i, mm. 55-60
106	PAC	Composta	Yes	No	No	No	Op. 76, No. 5, ii, mm. 7-9
107	PAC	Composta	Yes	Yes	No	No	Op. 76, No. 5, ii, mm. 28-33
108	PAC*	Composta	Yes	No	No	No	Op. 64, No. 6, i, mm. 35-36
109	PAC*	Composta	Yes	No	No	Yes	Op. 74, No. 1, i, mm. 39-41
110	PAC	None	Yes	Yes	No	Yes	Op. 17, No. 3, iv, mm. 22-24
111	PAC	None	No	No	No	No	Op. 20, No. 3, iv, mm. 9-10
112	PAC	None	No	No	No	No	Op. 20, No. 4, i, mm. 29-30
113	PAC	None	Yes	No	No	Yes	Op. 33, No. 5, ii, mm. 24-26
114	PAC	None	No	No	No	No	Op. 50, No. 1, iv, mm. 7-8
115	PAC	None	No	No	No	No	Op. 50, No. 1, iv, mm. 14-16
116	PAC	None	No	No	No	No	Op. 50, No. 4, i, mm. 7-8
117	PAC	None	Yes	No	No	No	Op. 50, No. 4, i, mm. 42-44
118	PAC	None	Yes	No	No	No	Op. 55, No. 1, ii, mm. 15-16
119	PAC	None	Yes	Yes	No	Yes	Op. 55, No. 1, ii, mm. 26-34
120	PAC	None	No	No	No	No	Op. 64, No. 3, i, mm. 16-17
121	PAC	None	No	No	No	Yes	Op. 64, No. 3, i, mm. 64-65
122	PAC	None	Yes	Yes	No	No	Op. 64, No. 6, i, mm. 39-45
123	PAC	None	No	No	No	No	Op. 71, No. 1, i, mm. 7-8
124	PAC	None	Yes	No	No	No	Op. 76, No. 2, i, mm. 48-50
125	PAC*	None	Yes	No	No	No	Op. 55, No. 1, ii, mm. 7-8
126	PAC*	None	Yes	Yes	No	No	Op. 71, No. 1, i, mm. 53-58
127	PAC-EV**	Evaded	Yes	No	No	Yes	Op. 50, No. 2, iv, mm. 65-68
128	PAC-EV**	Semplice	No	No	No	No	Op. 76, No. 2, i, mm. 30-32
129	PAC-EV**	Composta	Yes	No	No	No	Op. 20, No. 3, iv, mm. 27-29
130	PAC-EV**	Composta	Yes	Yes	No	No	Op. 54, No. 3, i, mm. 39-42
131	PAC-EV**	Composta	Yes	Yes	Yes	No	Op. 74, No. 1, ii, mm. 9-14
132	PAC-EV**	Composta	Yes	No	No	No	Op. 74, No. 1, ii, mm. 36-38
133	IAC	Semplice	No	No	No	No	Op. 20, No. 1, iv, mm. 2-3
134	IAC	Semplice	No	Yes	Yes	No	Op. 50, No. 5, i, mm. 13-16
135	IAC	Semplice	No	Yes	No	Yes	Op. 74, No. 1, i, mm. 7-10
136	IAC	Composta	Yes	No	Yes	No	Op. 20, No. 3, iii, mm. 6-8
137	IAC	Composta	Yes	Yes	No	Yes	Op. 64, No. 3, i, mm. 58-62
138	IAC	Composta	Yes	Yes	No	No	Op. 64, No. 4, iv, mm. 48-52
139	IAC*	Composta	Yes	No	Yes	No	Op. 64, No. 6, i, mm. 29-30
140	IAC	None	No	No	No	No	Op. 50, No. 1, iv, mm. 3-4

Number	Cadence	Bass Clausula	Cad $^6_4$	ECP	Diss.	Trill	Excerpt
141	IAC	None	No	No	No	No	Op. 50, No. 1, iv, mm. 10-12
142	IAC	None	No	No	No	Yes	Op. 50, No. 2, iv, mm. 6-8
143	HC	Converging	No	No	Yes	No	Op. 17, No. 1, i, mm. 17-18
144	HC	Converging	No	No	No	No	Op. 17, No. 3, iv, m. 8
145	HC	Converging	No	No	No	No	Op. 17, No. 4, i, mm. 32-33
146	HC	Converging	Yes	No	Yes	No	Op. 17, No. 5, i, m. 8
147	HC	Converging	No	No	No	No	Op. 20, No. 1, iv, mm. 17-18
148	HC	Converging	Yes	No	Yes	No	Op. 20, No. 3, i, mm. 12-14
149	HC	Converging	No	No	No	No	Op. 20, No. 3, i, mm. 39-41
150	HC	Converging	Yes	No	Yes	No	Op. 20, No. 4, iv, mm. 17-18
151	HC	Converging	No	No	No	No	Op. 20, No. 5, i, mm. 24-25
152	HC	Converging	No	Yes	No	No	Op. 20, No. 5, i, mm. 32-35
153	HC	Converging	No	No	No	No	Op. 20, No. 6, ii, mm. 15-17
154	HC	Converging	Yes	No	Yes	No	Op. 33, No. 1, i, m. 6
155	HC	Converging	No	No	Yes	No	Op. 33, No. 1, iii, mm. 15-16
156	HC	Converging	Yes	No	Yes	No	Op. 33, No. 2, i, m. 8
157	HC	Converging	Yes	No	Yes	No	Op. 33, No. 3, iii, mm. 3-4
158	HC	Converging	No	No	Yes	No	Op. 33, No. 4, i, mm. 16-17
159	HC	Converging	No	No	No	No	Op. 33, No. 4, i, mm. 20-21
160	HC	Converging	No	No	No	No	Op. 33, No. 5, i, mm. 43-45
161	HC	Converging	No	Yes	No	No	Op. 33, No. 5, ii, mm. 14-17
162	HC	Converging	No	No	No	No	Op. 50, No. 1, i, mm. 32-33
163	HC	Converging	No	No	No	No	Op. 50, No. 2, iv, mm. 59-60
164	HC	Converging	Yes	No	Yes	No	Op. 50, No. 3, iv, mm. 10-11
165	HC	Converging	No	No	Yes	No	Op. 50, No. 3, iv, mm. 39-40
166	HC	Converging	No	No	No	No	Op. 50, No. 5, i, mm. 27-28
167	HC	Converging	Yes	Yes	Yes	No	Op. 50, No. 5, iv, mm. 21-24
168	HC	Converging	No	No	No	No	Op. 50, No. 6, i, mm. 22-23
169	HC	Converging	No	No	No	No	Op. 50, No. 6, ii, mm. 7-8
170	HC	Converging	No	No	No	No	Op. 54, No. 1, ii, mm. 11-12
171	HC	Converging	No	No	No	No	Op. 54, No. 1, ii, mm. 25-26
172	HC	Converging	Yes	No	Yes	No	Op. 54, No. 3, i, mm. 3-4
173	HC	Converging	Yes	No	Yes	No	Op. 54, No. 3, i, mm. 21-22
174	HC	Converging	Yes	No	Yes	No	Op. 54, No. 3, i, mm. 29-30
175	HC	Converging	No	Yes	No	No	Op. 54, No. 3, iv, mm. 31-34
176	HC	Converging	No	No	No	No	Op. 54, No. 3, iv, mm. 42-43
177	HC	Converging	No	No	No	No	Op. 55, No. 1, ii, mm. 20-22
178	HC	Converging	No	No	No	No	Op. 55, No. 2, ii, mm. 25-27
179	HC	Converging	No	Yes	No	No	Op. 55, No. 2, ii, mm. 31-34
180	HC	Converging	No	No	No	No	Op. 55, No. 3, i, mm. 6-7
181	HC	Converging	No	No	No	No	Op. 64, No. 3, iv, mm. 52-53
182	HC	Converging	No	No	No	No	Op. 64, No. 4, i, mm. 13-14
183	HC	Converging	No	No	No	No	Op. 64, No. 4, iv, mm. 23-24
184	HC	Converging	No	No	No	No	Op. 64, No. 4, iv, mm. 33-34
185	HC	Converging	No	No	No	No	Op. 64, No. 6, i, mm. 15-16
186	HC	Converging	No	No	No	No	Op. 71, No. 1, i, mm. 14-16
187	HC	Converging	No	No	No	No	Op. 76, No. 4, i, mm. 31-32

Number	Cadence	Bass Clausula	Cad $^6_4$	ECP	Diss.	Trill	Excerpt
188	HC*	Converging	No	No	Yes	No	Op. 20, No. 3, iii, mm. 17-18
189	HC*	Converging	No	No	No	No	Op. 50, No. 2, iv, mm. 30-32
190	HC*	Converging	No	No	No	No	Op. 50, No. 3, iv, mm. 29-31
191	HC*	Converging	No	No	No	No	Op. 55, No. 3, i, mm. 24-25
192	HC	Expanding	No	No	No	No	Op. 17, No. 1, i, m. 2
193	HC	Expanding	No	No	No	No	Op. 17, No. 2, i, mm. 19-20
194	HC	Expanding	No	No	No	No	Op. 17, No. 2, i, mm. 23-25
195	HC	Expanding	No	No	No	No	Op. 17, No. 4, i, mm. 18-19
196	HC	Expanding	No	No	No	No	Op. 17, No. 5, i, m. 18
197	HC	Expanding	No	No	No	No	Op. 17, No. 6, i, mm. 55-56
198	HC	Expanding	No	No	No	No	Op. 20, No. 3, i, mm. 3-4
199	HC	Expanding	No	No	Yes	No	Op. 20, No. 3, iii, mm. 26-27
200	HC	Expanding	No	No	No	No	Op. 20, No. 3, iv, mm. 16-17
201	HC	Expanding	No	Yes	No	No	Op. 20, No. 3, iv, mm. 21-25
202	HC	Expanding	No	Yes	No	No	Op. 20, No. 4, iv, mm. 23-28
203	HC	Expanding	No	No	No	No	Op. 33, No. 1, iii, mm. 7-8
204	HC	Expanding	No	No	No	No	Op. 33, No. 1, iii, mm. 23-24
205	HC	Expanding	Yes	No	Yes	No	Op. 33, No. 3, iii, mm. 12-13
206	HC	Expanding	No	No	No	No	Op. 50, No. 1, iv, mm. 50-52
207	HC	Expanding	No	No	No	No	Op. 50, No. 2, i, mm. 34-35
208	HC	Expanding	No	No	No	No	Op. 50, No. 2, iv, mm. 37-38
209	HC	Expanding	No	No	No	No	Op. 50, No. 4, i, mm. 20-21
210	HC	Expanding	No	Yes	No	No	Op. 54, No. 2, i, mm. 50-54
211	HC	Expanding	No	No	No	No	Op. 54, No. 3, i, mm. 52-54
212	HC	Expanding	No	No	No	No	Op. 55, No. 2, ii, mm. 7-8
213	HC	Expanding	No	No	No	No	Op. 64, No. 3, i, mm. 40-42
214	HC	Expanding	No	No	No	No	Op. 64, No. 3, iv, mm. 33-35
215	HC	Expanding	No	No	No	No	Op. 64, No. 4, i, mm. 20-21
216	HC	Expanding	No	No	No	No	Op. 71, No. 1, i, mm. 34-36
217	HC	Expanding	No	No	No	No	Op. 76, No. 2, i, m. 4
218	HC	Expanding	No	No	No	No	Op. 76, No. 2, i, mm. 40-41
219	HC	Expanding	No	No	No	No	Op. 76, No. 4, i, mm. 49-50
220	HC*	Expanding	No	No	No	No	Op. 20, No. 5, i, mm. 8-10
221	HC*	Expanding	No	No	No	No	Op. 50, No. 1, iv, mm. 23-24
222	HC*	Expanding	No	No	No	No	Op. 74, No. 1, i, mm. 26-27
223	HC	Leaping	Yes	No	Yes	No	Op. 20, No. 6, ii, mm. 13-14
224	HC	Leaping	No	No	No	No	Op. 64, No. 3, i, mm. 12-13
225	HC	Leaping	No	No	No	No	Op. 64, No. 3, iv, mm. 3-4
226	HC	Leaping	No	No	No	No	Op. 64, No. 6, i, m. 4
227	HC	Leaping	Yes	No	Yes	No	Op. 76, No. 2, i, mm. 18-19
228	HC*	Leaping	Yes	No	Yes	No	Op. 33, No. 1, i, mm. 15-16
229	HC*	Leaping	No	No	No	No	Op. 54, No. 3, iv, mm. 7-8
230	HC*	Leaping	Yes	No	Yes	No	Op. 74, No. 1, ii, mm. 26-27
231	HC	Reinterpreted	No	No	No	No	Op. 17, No. 6, i, mm. 11-12
232	HC	Reinterpreted	No	No	No	No	Op. 33, No. 5, i, mm. 17-20
233	HC	Reinterpreted	No	No	No	No	Op. 50, No. 2, i, mm. 12-13
234	HC	Reinterpreted	No	No	No	No	Op. 50, No. 6, ii, m. 2

Number	Cadence	Bass Clausula	Cad $^6_4$	ECP	Diss.	Trill	Excerpt
235	HC	Reinterpreted	No	No	No	No	Op. 54, No. 1, i, mm. 21-22
236	HC	Reinterpreted	No	No	No	No	Op. 64, No. 4, i, mm. 3-4
237	DC	Deceptive	Yes	Yes	No	No	Op. 17, No. 1, i, mm. 31-37
238	DC	Deceptive	Yes	Yes	No	Yes	Op. 17, No. 2, i, mm. 26-29
239	DC	Deceptive	No	No	Yes	No	Op. 17, No. 4, i, mm. 4-6
240	DC	Deceptive	Yes	No	No	Yes	Op. 17, No. 4, i, mm. 39-41
241	DC	Deceptive	No	No	No	No	Op. 17, No. 5, i, mm. 23-24
242	DC	Deceptive	Yes	No	No	No	Op. 20, No. 3, i, mm. 56-60
243	DC	Deceptive	Yes	Yes	No	No	Op. 20, No. 3, i, mm. 79-85
244	DC	Deceptive	Yes	No	No	Yes	Op. 20, No. 4, i, mm. 22-24
245	DC	Deceptive	Yes	Yes	No	Yes	Op. 20, No. 5, i, mm. 40-43
246	DC	Deceptive	No	No	No	No	Op. 33, No. 1, i, mm. 28-29
247	DC	Deceptive	Yes	No	Yes	No	Op. 33, No. 1, iii, mm. 35-36
248	DC	Deceptive	No	No	No	No	Op. 33, No. 4, i, mm. 8-9
249	DC	Deceptive	No	No	No	No	Op. 33, No. 4, i, mm. 9-11
250	DC	Deceptive	No	No	Yes	No	Op. 50, No. 1, i, mm. 9-10
251	DC	Deceptive	Yes	No	No	No	Op. 55, No. 3, i, mm. 42-44
252	DC	Deceptive	No	Yes	Yes	No	Op. 64, No. 4, iv, mm. 39-42
253	DC	Deceptive	Yes	No	No	No	Op. 64, No. 6, i, mm. 38-39
254	DC	Deceptive	Yes	Yes	No	No	Op. 76, No. 4, i, mm. 13-18
255	DC	Deceptive	No	No	Yes	No	Op. 76, No. 5, ii, mm. 25-27
256	DC-EV**	Deceptive	Yes	Yes	No	Yes	Op. 50, No. 1, iv, mm. 59-64
257	DC-EV**	Deceptive	Yes	Yes	No	Yes	Op. 50, No. 6, i, mm. 34-38
258	DC-EV**	Evaded	Yes	No	No	No	Op. 50, No. 2, i, mm. 90-93
259	DC-EV**	Composta	Yes	No	No	Yes	Op. 20, No. 3, iii, mm. 25-26
260	EV	Evaded	Yes	Yes	No	Yes	Op. 17, No. 1, i, mm. 25-31
261	EV	Evaded	Yes	No	Yes	No	Op. 17, No. 4, i, mm. 45-47
262	EV	Evaded	Yes	Yes	No	No	Op. 33, No. 3, iii, mm. 19-22
263	EV	Evaded	Yes	No	No	No	Op. 33, No. 5, i, mm. 81-83
264	EV	Evaded	Yes	No	No	No	Op. 50, No. 2, i, mm. 56-58
265	EV	Evaded	Yes	Yes	No	No	Op. 50, No. 2, i, mm. 58-63
266	EV	Evaded	Yes	No	No	No	Op. 74, No. 1, ii, mm. 32-33
267	EV	Evaded	Yes	No	No	No	Op. 76, No. 2, i, mm. 26-28
268	EV	Semplice	No	No	No	No	Op. 74, No. 1, i, mm. 46-47
269	EV	Composta	Yes	No	No	No	Op. 33, No. 5, i, mm. 63-65
270	EV	Composta	Yes	No	No	No	Op. 74, No. 1, i, mm. 43-45

* Excluded from the analyses in Chapters 5 and 6 because one or both of the outer parts was not present at the moment of cadential arrival.
** Excluded from the analyses in Chapters 5 and 6 because the cadence implied more than one category.

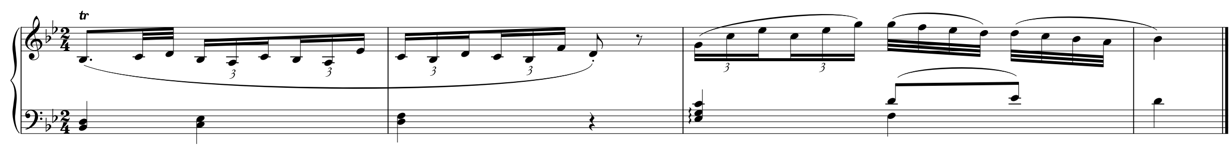
Appendix B

Experiment I: Stimuli

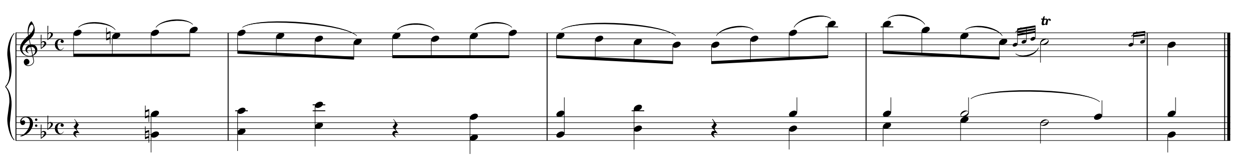
Each of the five cadence categories examined in the stimulus set was subdivided into two subtypes. The PAC category was subdivided according to formal location, selected either from the main theme or the subordinate theme (MT vs. ST). The IAC category was subdivided according to the presence or absence of a melodic dissonance at cadential arrival (Diss. vs. No Diss.). The HC category was subdivided according to formal location, selected either from the antecedent of a period theme type or from the end of the transition of a sonata-form movement (Theme vs. TR). The excerpts from the HC category were also separately classified according to the presence or absence of a melodic dissonance at the moment of cadential arrival (Diss vs. No Diss.). The DC category was subdivided according to whether the melody arrives on $\hat{1}$ or on $\hat{3}$ (Failed PAC vs. Failed IAC). Finally, the EV category were subdivided according to the harmony appearing at the moment of expected cadential arrival—tonic harmony, which is typically inverted, but may sometimes be in root position, or non-tonic harmony (Tonic vs. Non-tonic).

The extraction of each excerpt from its surrounding material introduced a number of factors at the cadential arrival that could confound the results, so it was necessary to apply the following constraints: (1) any chord tones appearing after cadential arrival (e.g., an Alberti bass pattern) were verticalized to the moment of cadential arrival and all subsequent material was removed; (2) the duration of the cadential arrival was recomposed to one full tactus; (3) for two excerpts a rest appeared at the expected cadential arrival, so the events following the rest were shifted back to cadential arrival. For further details, see §7.2.

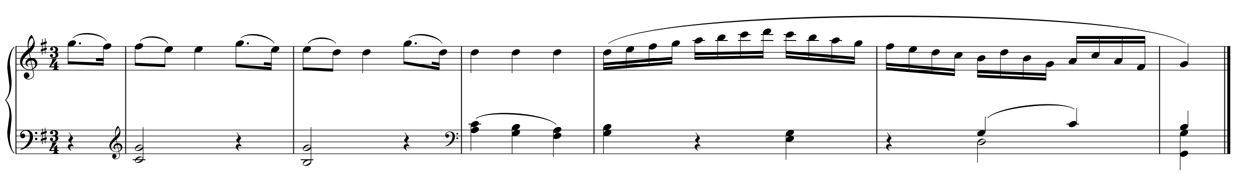
1. K. 281, i, mm. 5–8 (Allegro, eighth = 132) PAC – MT



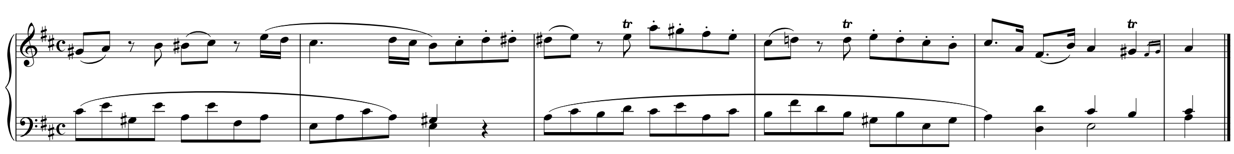
2. K. 281, iii, mm. 5–8 (Allegro, half = 85) PAC – MT



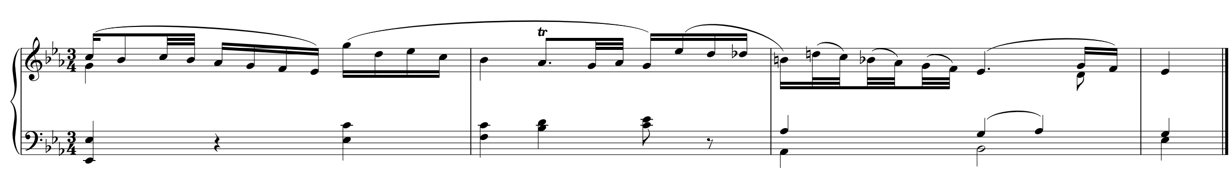
3. K. 283, i, mm. 5–10 (Allegro, quarter = 138) PAC – MT



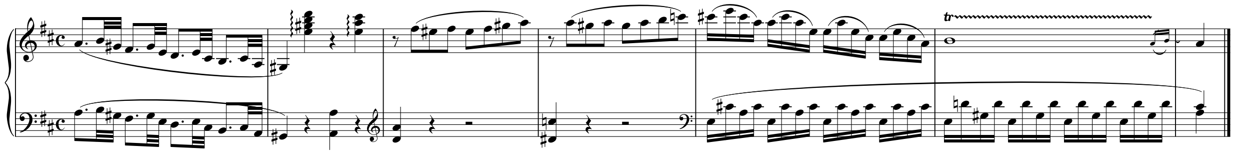
4. K. 311, i, mm. 19–24 (Allegro con spirito, quarter = 132) PAC – MT



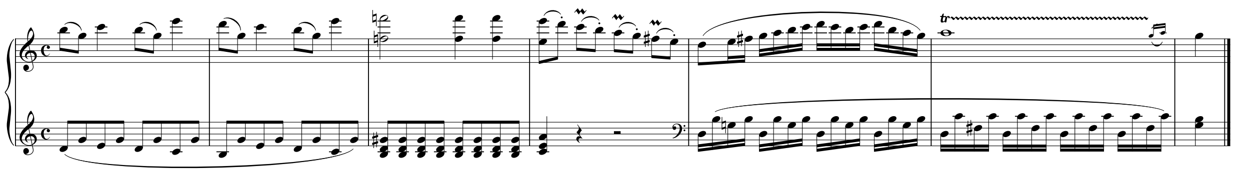
5. K. 333, ii, mm. 5–8 (Andante cantabile, quarter = 56) PAC – MT



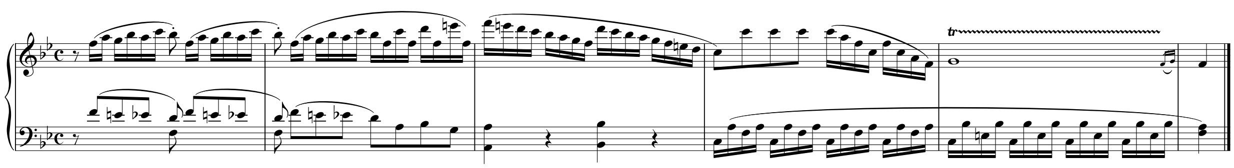
6. K. 284, i, mm. 44–50 (Allegro, quarter = 136) PAC – ST



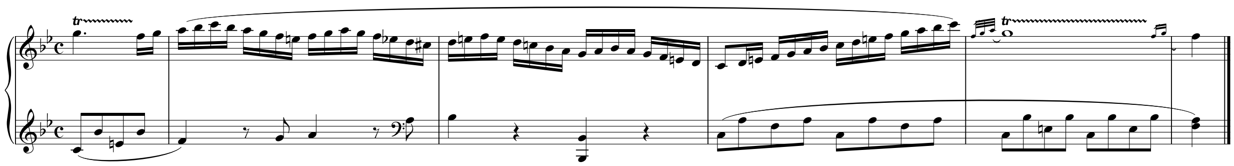
7. K. 309, i, mm. 48–54 (Allegro con spirito, quarter = 144) PAC – ST



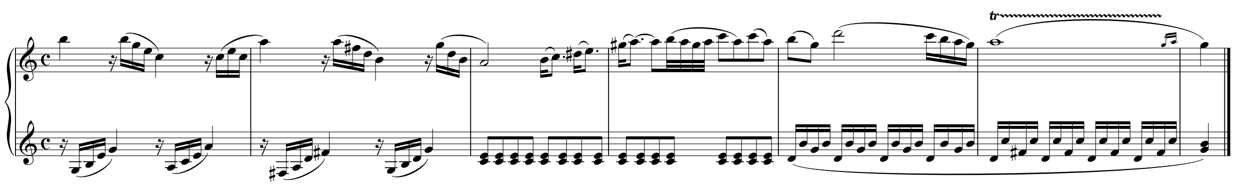
8. K. 333, i, mm. 54–59 (Allegro, quarter = 126) PAC – ST



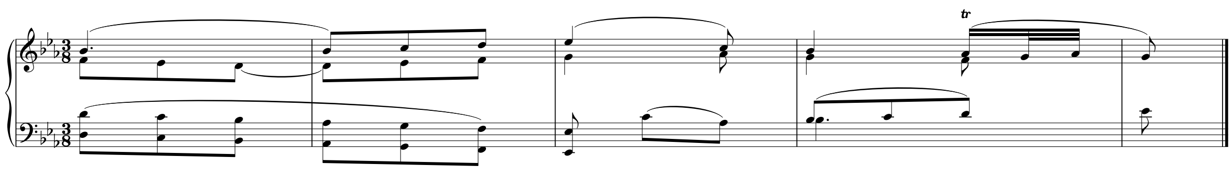
9. K. 333, iii, mm. 31–36 (Allegretto grazioso, quarter = 138) PAC – ST

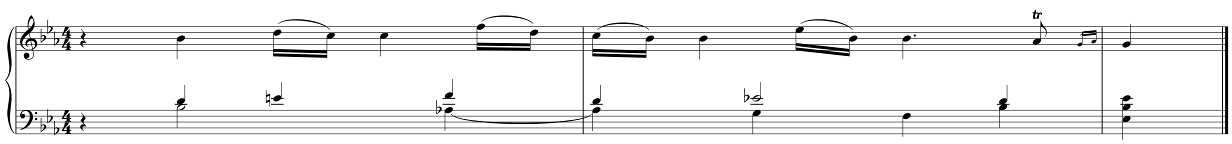


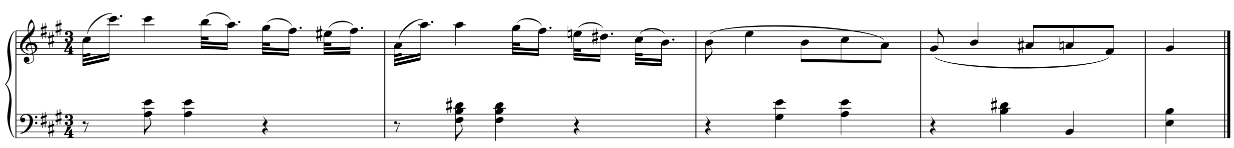
10. K. 545, i, mm. 20–26 (Allegro, quarter = 132) PAC – ST

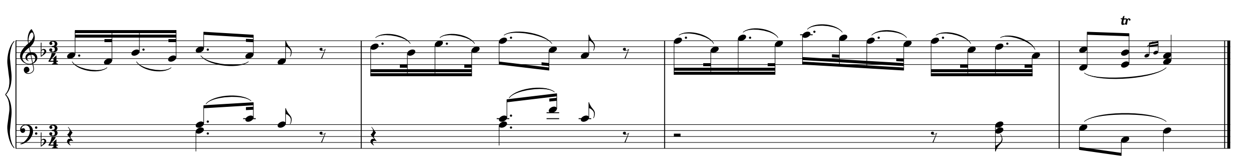


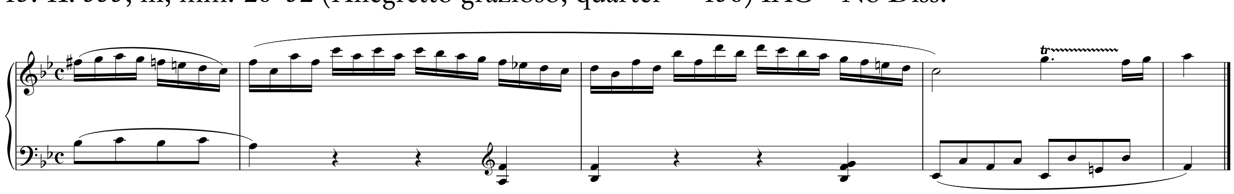
11. K. 281, ii, mm. 4–8 (Andante, eighth = 80) IAC – No Diss.



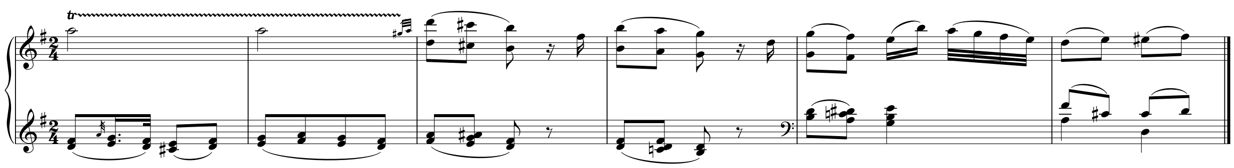
12. K. 282, i, mm. 2–4 (Adagio, quarter = 40) IAC – No Diss.

13. K. 284, ii, mm. 21–25 (Andante, quarter = 72) IAC – No Diss.



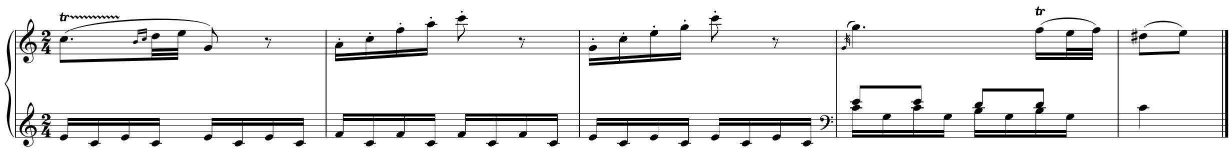
14. K. 309, ii, mm. 1–4 (Andante un poco adagio, quarter = 50) IAC – No Diss.

15. K. 333, iii, mm. 28–32 (Allegretto grazioso, quarter = 138) IAC – No Diss.



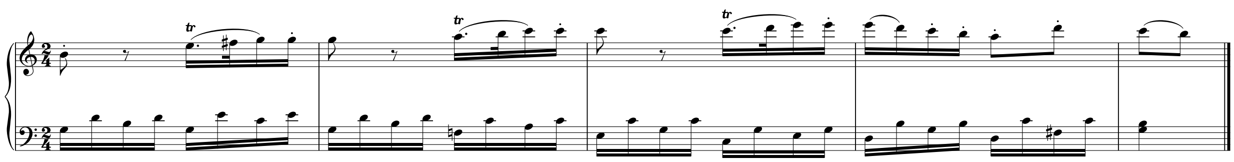
16. K. 311, ii, mm. 27–32 (Andante con espressione, eighth = 96) IAC – Diss.



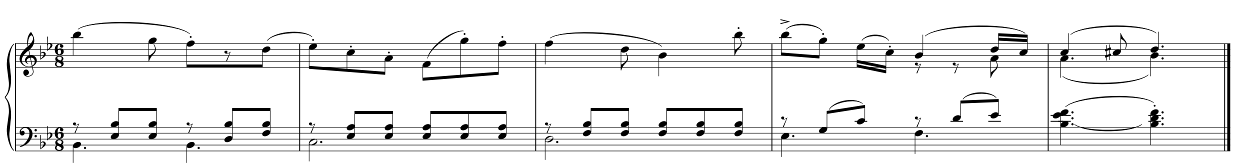
17. K. 330, i, mm. 4–8 (Allegro moderato, eighth = 132) IAC – Diss.



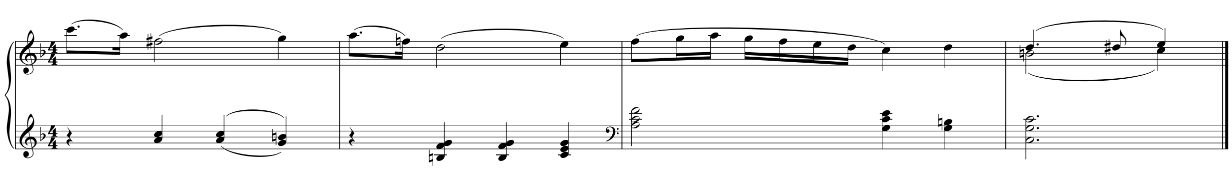
18. K. 330, iii, mm. 39–43 (Allegretto, quarter = 88) IAC – Diss.



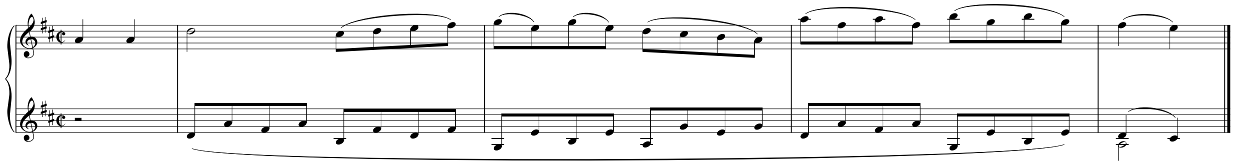
19. K. 498a, iv, mm. 32–36 (Allegro, dotted quarter = 100) IAC – Diss.



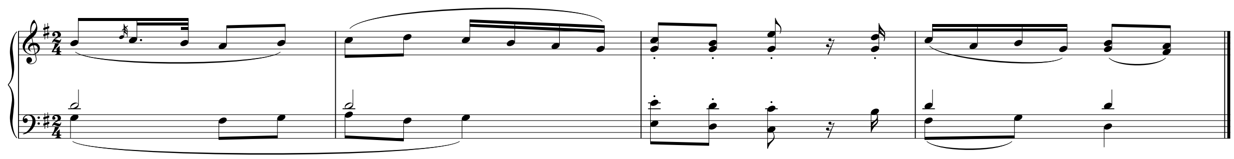
20. K. 533, iii, mm. 23–26 (Allegretto, half = 63) IAC – Diss.

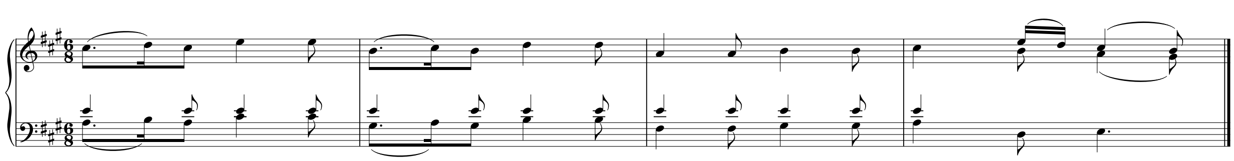


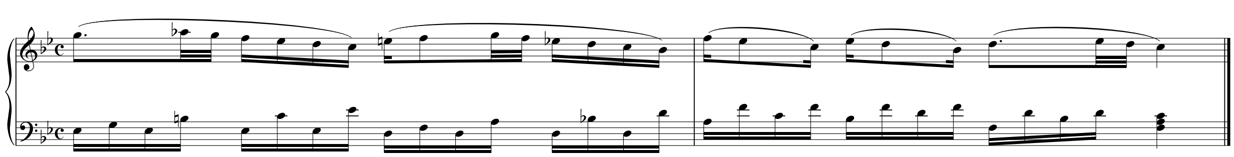
21. K. 284, iii, mm. 1–4 (Andante, quarter = 120) HC – Theme/Diss.



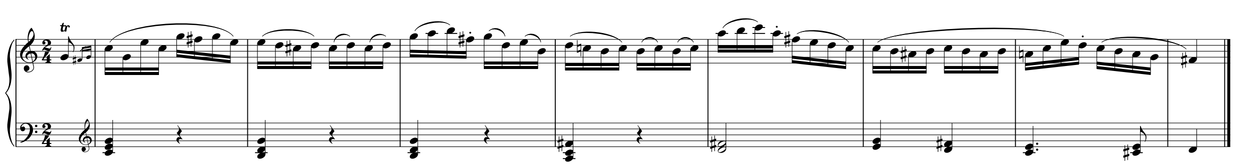
22. K. 311, ii, mm. 1–4 (Andante, eighth = 96) HC – Theme/Diss.



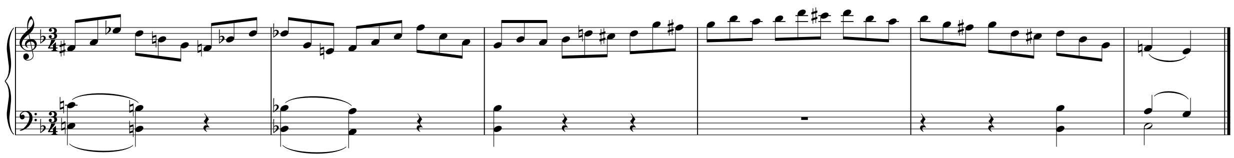
23. K. 331, i, mm. 1–4 (Andante, eighth = 120) HC – Theme/Diss.

24. K. 332, ii, mm. 3–4 (Adagio, eighth = 84) HC – Theme/No Diss.



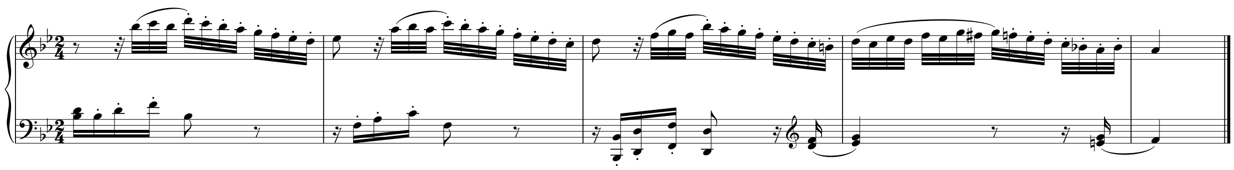
25. K. 279, iii, mm. 11–18 (Allegro, quarter = 120) HC – Transition/No Diss.



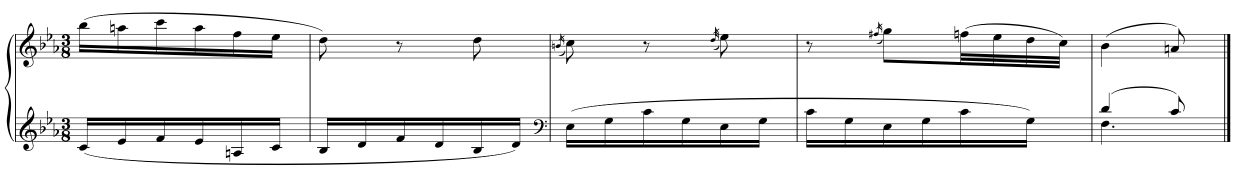
26. K. 280, i, mm. 21–26 (Allegro assai, quarter = 138) HC – Transition/Diss.



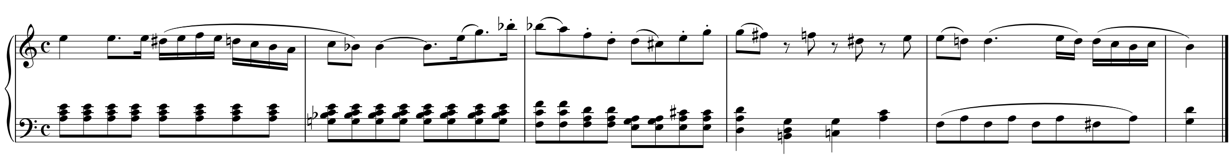
27. K. 281, i, mm. 12–16 (Allegro, eighth = 132) HC – Transition/No Diss.

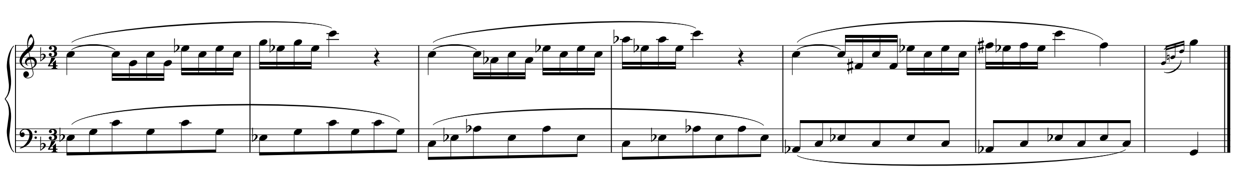


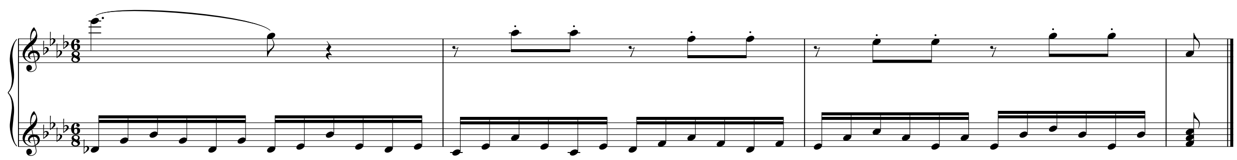
28. K. 281, ii, mm. 22–26 (Andante, eighth = 96) HC – Transition/Diss.



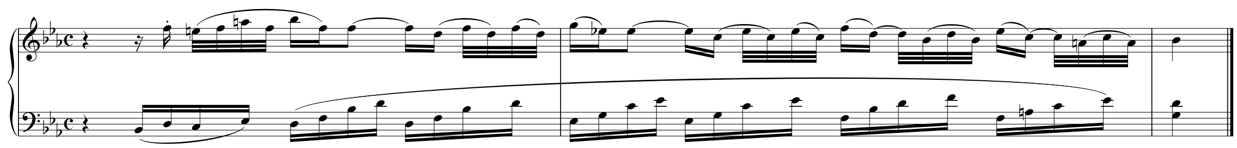
29. K. 310, i, mm. 11–16 (Allegro maestoso, quarter = 116) HC – Transition/No Diss.

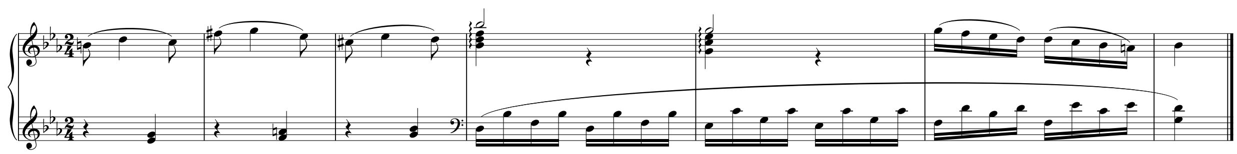


30. K. 332, i, mm. 31–37 (Allegro, quarter = 152) HC – Transition/No Diss.

31. K. 280, ii, mm. 16–19 (Adagio, eighth = 76) DC – Failed PAC

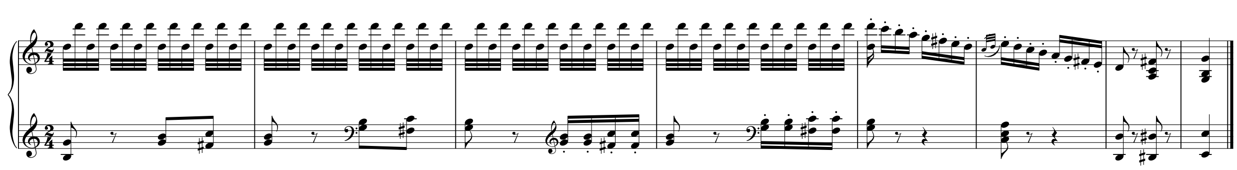


32. K. 281, ii, mm. 32–35 (Andante, eighth = 96) DC – Failed PAC

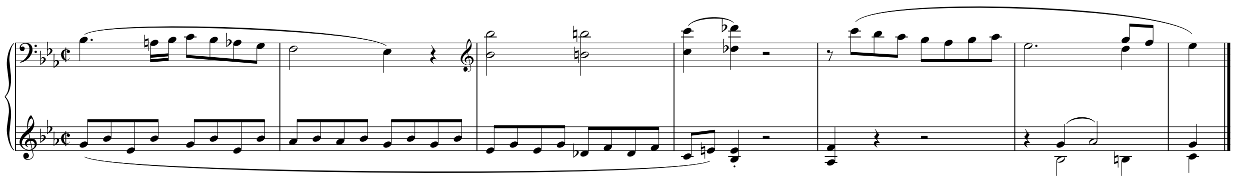
33. K. 282, i, mm. 11–13 (Adagio, quarter = 45) DC – Failed PAC

34. K. 282, iii, mm. 25–31 (Allegro, quarter = 120) DC – Failed PAC



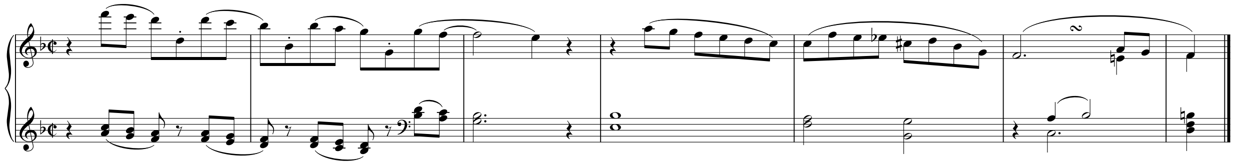
35. K. 309, iii, mm. 58–65 (Allegretto grazioso, quarter = 88) DC – Failed PAC



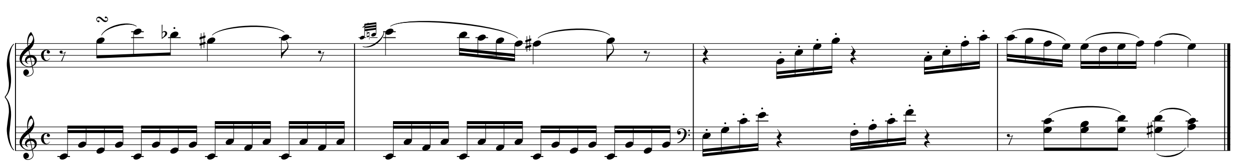
36. K. 457, i, mm. 42–48 (Molto Allegro, half = 84) DC – Failed PAC



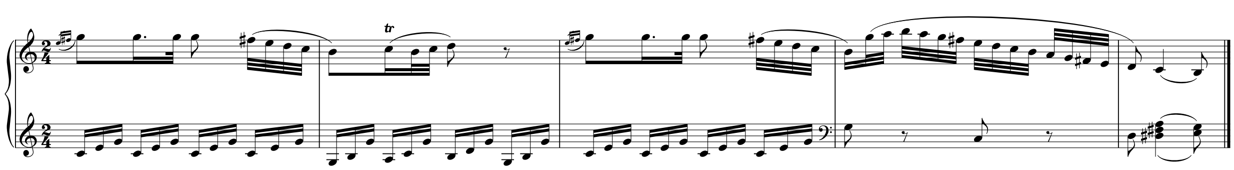
37. K. 533, i, mm. 16–22 (Allegro, half = 72) DC – Failed PAC (not VI)



38. K. 279, i, mm. 7–10 (Allegro, quarter = 112) DC – Failed IAC

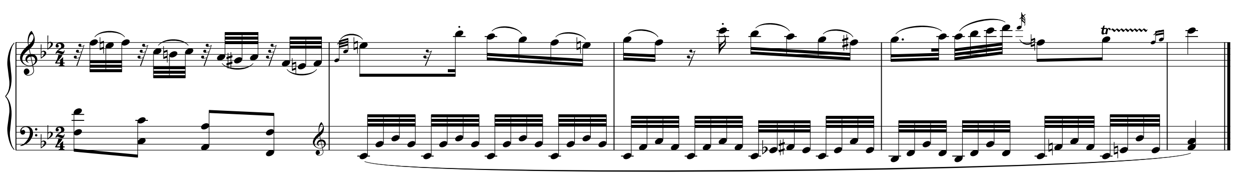


39. K. 330, i, mm. 27–31 (Allegro moderato, eighth = 126) DC – Failed IAC

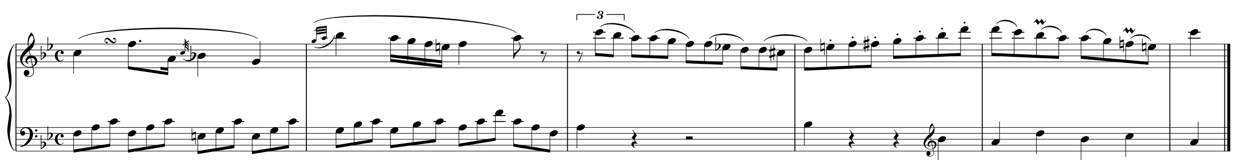


40. K. 457, ii, mm. 9–11 (Adagio, eighth = 69) DC – Failed IAC

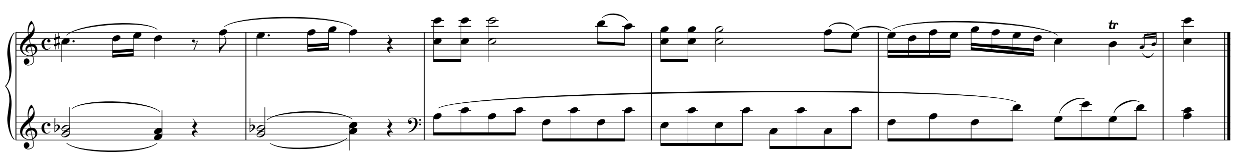
41. K. 281, i, mm. 30–34 (Allegro, quarter = 132) EV – Tonic



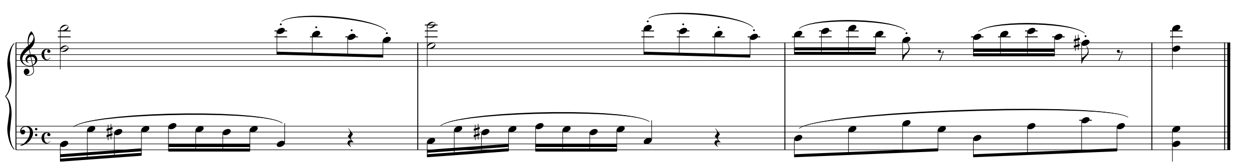
42. K. 281, iii, mm. 30–35 (Allegro, half = 75) EV – Tonic



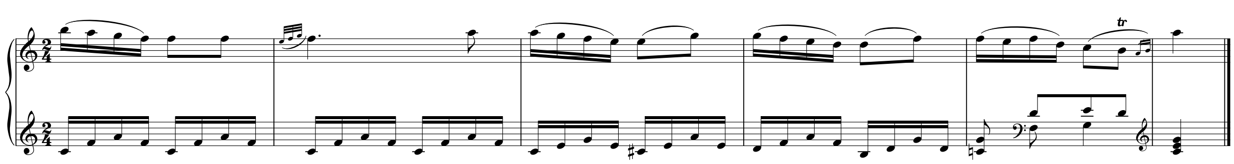
43. K. 309, i, mm. 13–18 (Allegro con spirito, quarter = 144) EV – Tonic (Sub.)



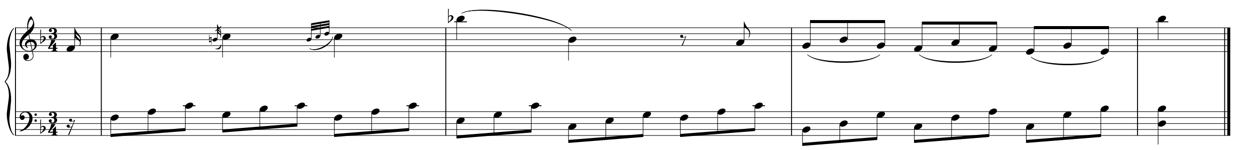
44. K. 309, i, mm. 43–46 (Allegro con spirito, quarter = 144) EV – Tonic



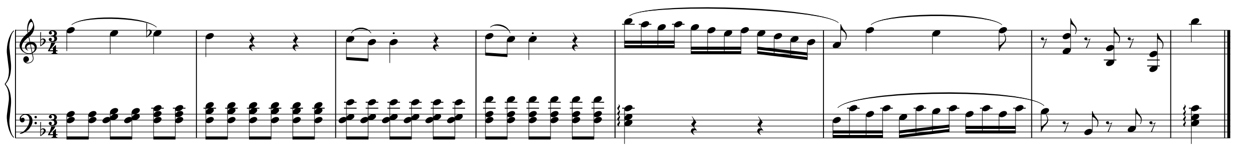
45. K. 309, iii, mm. 11–16 (Allegretto grazioso, quarter = 75) EV – Tonic

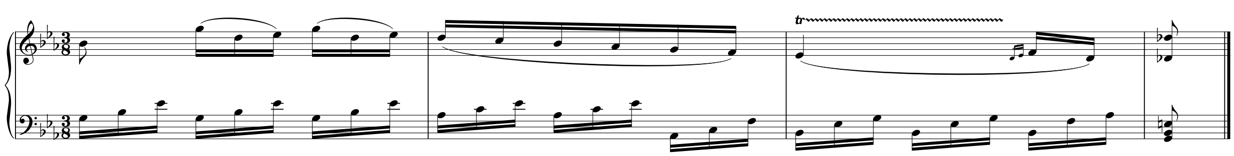


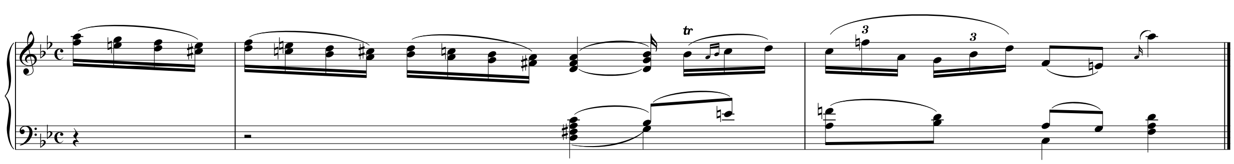
46. K. 279, ii, mm. 1–4 (Andante, quarter = 60) EV – Non-Tonic

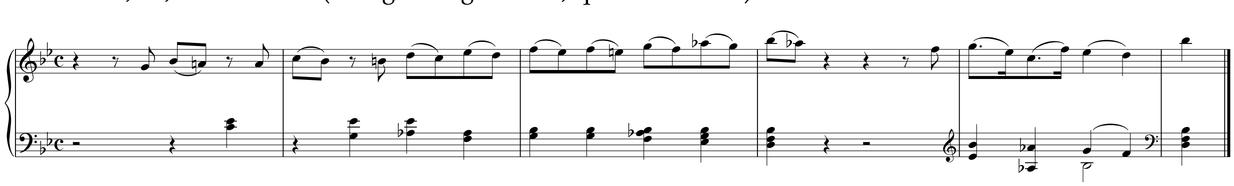


47. K. 280, i, mm. 3–10 (Allegro assai, quarter = 138) EV – Non-Tonic



48. K. 281, ii, mm. 96–99 (Andante, eighth = 96) EV – Non-Tonic

49. K. 332, ii, mm. 14–16 (Adagio, eighth = 84) EV – Non-Tonic



50. K. 333, iii, mm. 84–89 (Allegretto grazioso, quarter = 138) EV – Non-Tonic

Bibliography

389-8, ISO. *Acoustics: Reference Zero for the Calibration of Audiometric Equipment—Part 8: Reference Equivalent Threshold Sound Pressure Levels for Pure Tones and Circumaural Earphones*. Technical Report. International Organization for Standardization, 2004.

Aarden, Bret. "Dynamic Melodic Expectancy." PhD Dissertation, The Ohio State University, 2003.

———. "Expectancy vs. Retrospective Perception: Reconsidering the Effects of Schema and Continuation Judgments on Measures of Melodic Expectancy." In *Proceedings of the 7th International Conference on Music Perception and Cognition*, edited by Catherine Stevens, Denis Burnham, Gary McPherson, Emory Schubert, and James Renwick. Adelaide: Causal Productions, 2002.

Agawu, Kofi. "Concepts of Closure and Chopin's Opus 28." *Music Theory Spectrum* 9 (1987): 1–17.

Agmon, Eytan. "Functional Harmony Revisited: A Prototype-Theoretic Approach." *Music Theory Spectrum* 17, no. 2 (1995): 196–214.

Albrecht, Joshua, and Daniel Shanahan. "The Use of Large Corpora to Train a New Type of Key-Finding Algorithm: An Improved Treatment of the Minor Mode." *Music Perception* 31, no. 1 (2013): 59–67.

ANSI. *Psychoacoustical Terminology*. New York: American National Standards Institute, 1973.

Anson-Cartwright, Mark. "Concepts of Closure in Tonal Music: A Critical Study." *Theory and Practice* 32 (2007): 1–18.

Arnheim, Rudolf. *Art and Visual Perception: A Psychology of the Creative Eye*. Berkeley: University of California Press, 1954.

Atkinson, Richard C., and Richard M. Shiffrin. "Human Memory: A Proposed System and Its Control Processes." In *The Psychology of Learning and Motivation*, edited by Kenneth Wartenbee Spence and Janet T. Spence, 2:89–195. New York: Academic Press, 1968.

Attneave, Fred. "Transfer of Experience with a Class-Schema to Identification-Learning of Patterns and Shapes." *Journal of Experimental Psychology* 54, no. 2 (1957): 81–88.

Baayen, R. Harald, Doug J. Davidson, and Douglas M. Bates. "Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items." *Journal of Memory and Language* 59 (2008): 390–412.

Baayen, R. Harold. *languageR: Data Sets and Functions with "Analyzing Linguistic Data: A Practical Introduction to Statistics". R package version 1.4.*, 2012.

Baayen, R. Harold, and Petar Milin. "Analyzing Reaction Times." *International Journal of Psychological Research* 3, no. 2 (2010): 12–28.

Bach, C. P. E. *Essay on the True Art of Playing Keyboard Instruments*. Translated by William Mitchell. New York: W. W. Norton & Co., Inc., 1949.

Baddeley, Alan. *Human Memory: Theory and Practice*. Hillsdale, NJ: Erlbaum, 1990.

Baddeley, Alan, and Graham Hitch. "Working Memory." In *The Psychology of Learning and Memory*, edited by Gordon H. Bower, 8:47–89. New York: Academic Press, 1974.

Baeza-Yates, Ricardo, and Berthier Ribeiro-Neto. *Modern Information Retrieval*. New York, NY: ACM Press, 1999.

Bain, Jennifer. "Theorizing the Cadence in the Music of Machaut." *Journal of Music Theory* 47, no. 2 (2003): 325–362.

Bar, Moshe. "The Proactive Brain: Using Analogies and Associations to Generate Predictions." *Trends in Cognitive Sciences* 11, no. 7 (2007): 280–289.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. "Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal." *Journal of Memory and Language* 68 (2013): 255–278.

Barry, Barbara. "In Search of an Ending: Reframing Mahler's Contexts of Closure." *Journal of Musicological Research* 26, no. 1 (2007): 55–68.

Bartlett, Frederic C. *Remembering: A Study in Experimental and Social Psychology*. London, United Kingdom: Cambridge University Press, 1932.

Bartlett, James C., and Jay W. Dowling. "Scale Structure and Similarity of Melodies." *Music Perception* 5, no. 3 (1988): 285–314.

Bates, D.M., M. Maechler, and B. Bolker. *lme4: Linear Mixed-Effects Models Using S4 Classes, (R package version 0.98.501) [computer software].*, 2011.

Bell, Timothy C., John G. Cleary, and Ian H. Witten. *Text Compression*. Englewood Cliffs, NJ: Prentice Hall, 1990.

Besson, Mireille, and Frédérique Faıta. "An Event-Related Potential (ERP) Study of Musical Expectancy: Comparison of Musicians with Nonmusicians." *Journal of Experimental Psychology: Human Perception and Performance* 21, no. 6 (1995): 1278–1296.

Bharucha, Jamshed J. "Anchoring Effects in Music: The Resolution of Dissonance." *Cognitive Psychology* 16, no. 4 (1984): 485–518.

———. "Event Hierarchies, Tonal Hierarchies, and Assimilation: A Reply to Deutsch and Dowling." *Journal of Experimental Psychology: General* 113, no. 3 (1984): 421–425.

———. "MUSACT: A Connectionist Model of Musical Harmony." In *Proceedings of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum Press, 1987.

———. "Music Cognition and Perceptual Facilitation: A Connectionist Framework." *Music Perception* 5, no. 1 (1987): 1–30.

Bharucha, Jamshed J., and Carol L. Krumhansl. "The Representation of Harmonic Structure in Music: Hierarchies of Stability as a Function of Context." *Cognition* 13 (1983): 63–102.

Bharucha, Jamshed J., and Keiko Stoeckig. "Reaction Time and Musical Expectancy: Priming of Chords." *Journal of Experimental Psychology: Human Perception and Performance* 12, no. 4 (1986): 403–410.

Bharucha, Jamshed, and Keiko Stoeckig. "Priming of Chords: Spreading Activation or Overlapping Frequency Spectra." *Perception and Psychophysics* 41, no. 6 (1987): 519–524.

Bigand, Emmanuel. "More about the Musical Expertise of Musically Untrained Listeners." *Annals of the New York Academy of Sciences* 999 (2003): 304–312.

Bigand, Emmanuel, Charles Delbé, Bénédicte Poulin-Charronnat, Marc Leman, and Barbara Tillmann. "Empirical Evidence for Musical Syntax Processing? Computer Simulations Reveal the Contribution of Auditory Short-Term Memory." *Frontiers in Systems Neuroscience* 8 (2014): 1–27.

Bigand, Emmanuel, François Madurell, Barbara Tillman, and Marion Pineau. "Effect of Global Structure and Temporal Organization on Chord Processing." *Journal of Experimental Psychology: Human Perception and Performance* 25, no. 1 (1999): 184–197.

Bigand, Emmanuel, and Richard Parncutt. "Perceiving Musical Tension in Long Chord Sequences." *Psychological Research* 62 (1999): 237–254.

Bigand, Emmanuel, Richard Parncutt, and Fred Lerdahl. "Perception of Musical Tension in Short Chord Sequences: The Influence of Harmonic Function, Sensory Dissonance, Horizontal Motion, and Musical Training." *Perception and Psychophysics* 59 (1996): 1098–1107.

Bigand, Emmanuel, and Marion Pineau. "Global Context Effects on Musical Expectancy." *Perception and Psychophysics* 59, no. 7 (1997): 1098–1107.

Bigand, Emmanuel, Benedicte Poulin, Barbara Tillmann, Francois Madurell, and D. A. D'Adamo. "Sensory Versus Cognitive Components in Harmonic Priming." *Journal of Experimental Psychology: Human Perception and Performance* 29, no. 1 (2003): 159–171.

Bigand, Emmanuel, Barbara Tillmann, Benedicte Poulin-Charronnat, and D. Manderlier. "Repetition Priming: Is Music Special?" *The Quarterly Journal of Experimental Psychology* 58A, no. 8 (2005): 1347–1375.

Bingham, W. Van Dyke. *Studies in Melody*. Baltimore: Review Publishing Company, 1910.

Blombach, Ann. "Phrase and Cadence: A Study of Terminology and Definition." *Journal of Music Theory Pedagogy* 1 (1987): 225–251.

Bolton, Thaddeus L. "Rhythm." *The American Journal of Psychology* 6, no. 2 (1894): 145–238.

Boltz, Marilyn. "Perceiving the End: Effects of Tonal Relationships on Melodic Completion." *Journal of Experimental Psychology: Human Perception and Performance* 15, no. 4 (1989): 749–761.

Bowerman, Bruce L., and Richard T. O'Connell. *Linear Statistical Models: An Applied Approach*. Pacific Grove, CA: Duxbury, 1990.

Bregman, Albert S. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press, 1990.

———. "Progress in Understanding Auditory Scene Analysis." *Music Perception* 33, no. 1 (2015): 12–19.

Brower, Candace. "Memory and the Perception of Rhythm." *Music Theory Spectrum* 15, no. 1 (1993): 19–35.

Brown, Helen. "The Interplay of Set Content and Temporal Context in a Functional Theory of Tonality Perception." *Music Perception* 5, no. 3 (1988): 219–249.

Brown, Judith C. "Determination of the Meter of Musical Scores by Autocorrelation." *Journal of the Acoustical Society of America* 94, no. 4 (1993): 1953–1957.

Brown, Peter F., Vincent J. Della Pietra, Stephen A. Della Pietra, Jennifer C. Lai, and Robert L. Mercer. "An Estimate of an Upper Bound for the Entropy of English." *Computational Linguistics* 18, no. 1 (1992): 31–40.

Bubic, Andreja, D. Yves von Cramon, and Ricarda I. Schubotz. "Prediction, Cognition and the Brain." *Frontiers in Human Neuroscience* 4 (2010): 1–15.

Bugge, Esben Paul, Kim Lundsteen Juncher, Brian Søborg Mathiasen, and Jakob Grue Simonsen. "Using Sequence Alignment and Voting to Improve Optical Music Recognition from Multiple Recognizers." In *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, 405–410. 2011.

Bunton, Suzanne. "On-Line Stochastic Processes in Data Compression." PhD Dissertation, University of Washington, 1996.

———. "Semantically Motivated Improvements for PPM Variants." *The Computer Journal* 40, nos. 2/3 (1997): 76–93.

Burgoyne, John Ashley. "Stochastic Processes and Database-Driven Musicology." PhD Dissertation, McGill University, 2012.

Burnham, Scott. "Form." In *The Cambridge History of Western Music Theory*, edited by Thomas Christensen, 880–906. Cambridge: Cambridge University Press, 2002.

Burns, Edward M. "Intervals, Scales, and Tuning." In *The Psychology of Music*, 2nd, edited by Diana Deutsch, 215–264. New York: Academic Press, 1998.

Burns, Edward M., and W. Dixon Ward. "Intervals, Scales, and Tuning." In *The Psychology of Music*, edited by Diana Deutsch, 241–269. New York: Academic Press, 1982.

Burstein, L. Poundie. "Mid-section Cadences in Haydn's Sonata-Form Movements." *Studia Musicologica* 51, nos. 1-2 (2010): 91–107.

———. "The Half Cadence and Other Such Slippery Events." *Music Theory Spectrum* 36, no. 2 (2014): 203–227.

Butler, David. "Describing the Perception of Tonality in Music: A Critique of the Tonal Hierarchy Theory and a Proposal for a Theory of Intervallic Rivalry." *Music Perception* 6, no. 3 (1989): 219–42.

Byros, Vasili. "Foundations of Tonality as Situated Cognition, 1730–1830." PhD Dissertation, Yale University, 2009.

———. "Meyer's Anvil: Revisiting the Schema Concept." *Music Analysis* 31, no. 3 (2012): 273–346.

Caplin, William E. *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven*. New York: Oxford University Press, 1998.

———. "Harmonic Variants of the Expanded Cadential Progression." In *A Composition as Problem II: Proceedings of the Second International Estonian Music Theory Conference*, edited by M. Humal, 49–71. Tallinn: Estonian Academy of Music, 2000.

———. "Harmony and Cadence in Gjerdingen's "Prinner"," edited by Markus Neuwirth and Pieter Bergé, 17–57. Leuven: Leuven University Press, 2015.

———. "Schoenberg's 'Second Melody', or, 'Meyer-ed' in the Bass." In *Communication in Eighteenth-Century Music*, edited by Danuta Mirka and Kofi Agawu, 160–186. Cambridge, UK: Cambridge University Press, 2008.

———. "Teaching Classical Form: Strict Categories versus Flexible Analyses." *Tijdschrift voor Muziektheorie (Dutch Journal of Music Theory)* 18, no. 3 (2013): 119–135.

———. "The Classical Cadence: Conceptions and Misconceptions." *Journal of the American Musicological Society* 57, no. 1 (2004): 51–118.

———. "The Classical Sonata Exposition: Cadential Goals and Form-Functional Plans." *Tijdschrift voor muziektheorie (Dutch Journal of Music Theory)* 6, no. 3 (2001): 195–209.

———. "The "Expanded Cadential Progression": A Category for the Analysis of Classical Form." *Journal of Musicological Research* 7 (1987): 215–257.

———. "Theories of Musical Rhythm in the 18th and 19th Centuries." In *The Cambridge History of Western Music Theory*, edited by Thomas Christensen, 657–694. Cambridge, UK: Cambridge University Press, 2002.

———. "William Caplin Responds." *Intersections: Canadian Journal of Music* 31, no. 1 (2010): 67–76.

Caplin, William E., James Hepokoski, and James Webster. *Musical Form, Forms, & Formenlehre: Three Methodological Reflections*. Edited by Pieter Bergé. Leuven: Leuven University Press, 2009.

Carlsen, James C. "Some Factors Which Influence Melodic Expectancy." *Psychomusicology* 1, no. 1 (1981): 12–29.

Carlsen, James C., Pierre I. Divenyi, and Jack A. Taylor. "A Prelimary Study of Perceptual Expectancy in Melodic Configurations." *Bulletin of the Council for Research in Music Education* 22 (1970): 4–12.

Casella, Alfredo. *The Evolution of Music throughout the History of the Perfect Cadence*. London, UK: Chester, 1924.

Casey, Michael A., Remco C. Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, and Malcolm Slaney. “Content-Based Music Information Retrieval: Current Directions and Future Challenges.” *Proceedings of the IEEE* 96, no. 4 (2008): 668–696.

Chen, Stanley F., and Joshua Goodman. “An Empirical Study of Smoothing Techniques for Language Modeling.” *Computer Speech & Language* 13 (1999): 359–394.

Chuen, Lorraine, David Sears, and Stephen McAdams. “Psychophysiological Responses to Auditory Change.” *Psychophysiology*, 2016. doi:10.1111/psyp.12633.

Clark, Andy. “Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science.” *Behavioral and Brain Sciences* 36, no. 3 (2013): 1–73.

Clarke, Eric. “Issues in Language and Music.” *Contemporary Music Review* 4 (1989): 9–22.

Cleary, John G., and W. J. Teahan. “Unbounded Length Contexts for PPM.” *The Computer Journal* 40, nos. 2/3 (1997): 67–75.

Cleary, John G., and Ian H. Witten. “Data Compression Using Adaptive Coding and Partial String Matching.” *IEEE Transactions on Communications* 32, no. 4 (1984): 396–402.

Clifford, Robert. “Perennial Questions.” *Tempo* 59, no. 234 (2005): 29–33.

Cohen, Barry H. *Explaining Psychological Statistics*. Hoboken, NJ: John Wiley & Sons, Inc., 2008.

Collins, Tom, Barbara Tillmann, Frederick S. Barrett, Charles Delbé, and Petr Janata. “A Combined Model of Sensory and Cognitive Representations Underlying Tonal Expectations in Music: From Audio Signals To Behavior.” *Psychological Review* 121, no. 1 (2014): 33–65.

Cone, Edward T. “Conversations with Roger Sessions.” *Perspectives of New Music* 4, no. 2 (1966): 29–46.

———. “Three Ways of Reading a Detective Story or a Brahms Intermezzo.” *The Georgia Review* 31, no. 3 (1977): 554–574.

Conklin, Darrell. “Modelling and Generating Music Using Multiple Viewpoints.” In *Proceedings of the First Workshop on Artificial Intelligence and Music*, 125–137. St. Paul, MN, 1988.

———. “Prediction and Entropy of Music.” MA Thesis, University of Calgary, 1990.

———. "Representation and Discovery of Vertical Patterns in Music." In *Music and Artifical Intelligence: Proc. ICMAI 2002*, edited by Christina Anagnostopoulou, Miguel Ferrand, and Alan Smaill, 2445:32–42. Springer-Verlag, 2002.

Conklin, Darrell, and Ian H. Witten. "Multiple Viewpoint Systems for Music Prediction." *Journal of New Music Research* 24, no. 1 (1995): 51–73.

Cook, Nicholas. "Methods for Analysing Recordings." In *The Cambridge Companion to Recorded Music*, edited by Nicholas Cook, Eric Clarke, Daniel Leech-Wilkinson, and John Rink, 221–246. Cambridge: Cambridge University Press, 2009.

———. "The Perception of Large-Scale Tonal Closure." *Music Perception* 5, no. 2 (1987): 197–206.

Cook, Norman D., and Takashi X. Fujisawa. "The Psychophysics of Harmony Perception: Harmony is a Three-Tone Phenomenon." *Empirical Music Review* 1 (2006): 106–126.

Cooper, Grosvenor, and Leonard Meyer. *The Rhythmic Structure of Music*. Chicago: The University of Chicago Press, 1960.

Corrigall, Kathleen A., and Laurel J. Trainor. "Musical Enculturation in Preschool Children: Acquisition of Key and Harmonic Knowledge." *Music Perception* 28, no. 2 (2010): 195–200.

Costa-Giomi, Eugenia. "Young Children's Harmonic Perception." *Annals of the New York Academy of Sciences* 999, no. 1 (2003): 477–484.

Craik, Kenneth. *The Nature of Explanation*. Cambridge: Cambridge University Press, 1943.

Cuddy, Lola L., Annabel J. Cohen, and D. J. K. Mewhort. "Perception of Structure in Short Melodic Sequences." *Journal of Experimental Psychology: Human Perception and Performance* 7, no. 4 (1981): 869–883.

Cuddy, Lola L., and Carole A. Lunney. "Expectancies Generated by Melodic Intervals: Perceptual Judgments of Melodic Continuity." *Perception and Psychophysics* 57, no. 4 (1995): 451–462.

Cudworth, Charles L. "Cadence galante: The Story of a Cliché." *The Monthly Musical Record* 79 (1949): 176–178.

Curtis, Meagan E., and Jamshed J. Bharucha. “Memory and Musical Expectation for Tones in Cultural Context.” *Music Perception* 26, no. 4 (2009): 365–375.

Cuthbert, Michael Scott, and Christopher Ariza. “music21: A Toolkit for Computer-Aided Musicology and Symbolic Music Data.” In *11th International Society for Music Information Retrieval Conference*, edited by J. Stephen Downie and Remco C. Veltkamp, 637–642. 2010.

Dack, James. “Sacred Music.” In *The Cambridge Companion to Haydn*, edited by Caryl Clark. Cambridge: Cambridge University Press, 2005.

Darwin, Christopher J., and Michael T. Turvey. “An Auditory Analogue of the Sperling Partial Report Procedure: Evidence for Brief Auditory Storage.” *Cognitive Psychology* 3 (1972): 255–267.

Delbé, Charles. “Recurrent Self-Organization of Sensory Signals in the Auditory Domain.” In *From Associations to Rules: Connectionist Models of Behavior and Cognition*, edited by Robert M. French and Elizabeth Thomas, 178–190. Singapore: World Scientific, 2008.

Deutsch, Diana. “Grouping Mechanisms in Music.” In *The Psychology of Music*, 3rd ed., edited by Diana Deutsch, 183–248. New York: Academic Press, 2013.

DiCiccio, Thomas J., and Bradley Efron. “Bootstrap Confidence Intervals.” *Statistical Science* 11, no. 3 (1996): 189–228.

Dienes, Zoltán, and Christopher Longuet-Higgins. “Can Musical Transformations Be Implicitly Learned?” *Cognitive Science* 28 (2004): 531–558.

Diergarten, Felix. “‘The True Fundamentals of Composition’: Haydn’s Partimento Counterpoint.” *Eighteenth Century Music* 8, no. 1 (2011): 53–75.

Dietterich, Thomas G. “Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms.” *Neural Computation* 10, no. 7 (1998): 1895–1923.

Dixon, Peter. “Models of Accuracy in Repeated-Measures Designs.” *Journal of Memory and Language* 59 (2008): 447–456.

Dowling, W. Jay. “Scale and Contour: Two Components of a Theory of Memory for Melodies.” *Psychological Review* 85, no. 4 (1978): 341–354.

———. "The Importance of Interval Information in Long-Term Memory for Melodies." *Psychomusicology* 1 (1981): 30–49.

Dowling, W. Jay, and Dane L. Harwood. *Music Cognition*. New York: Academic, 1986.

Downie, J. Stephen. "Evaluating a Simple Approach to Music Information Retrieval: Conceiving Melodic N-Grams as Text." PhD Dissertation, The University of Western Ontario, 1999.

———. "Music Information Retrieval." Chap. 7 in *Annual Review of Information Science and Technology*, edited by Blaise Cronin, 295–340. Medford, NJ: Information Today, 2003.

Duane, Ben. "Agency and Information Content in Eighteenth- and Early Nineteenth-Century String-Quartet Expositions." *Journal of Music Theory* 56, no. 1 (2012): 87–120.

———. "Texture in Eighteenth- and Early Nineteenth-Century String-Quartet Expositions." PhD Dissertation, Northwestern University, 2012.

Dunsby, Jonathan. "Schoenberg on Cadence." *Journal of the Arnold Schoenberg Institute* 4, no. 1 (1980): 41–49.

Ebbinghaus, Hermann. *On Memory: A Contribution to Experimental Psychology*. New York: Teachers College, Columbia University, 1885/1913.

Eberlein, Roland. "A Method of Analysing Harmony, Based on Interval Patterns or "Gestalten?" In *Music, Gestalt, and Computing*, edited by Marc Leman, 1317:225–236. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 1997.

Eberlein, Roland, and Jobst Fricke. *Kadenzwahrnehmung und Kadenzgeschichte: ein Beitrag zu einer Grammatik der Musik*. Frankfurt/M.: P. Lang, 1992.

Eerola, Tuomas. "The Dynamics of Musical Expectancy: Cross-Cultural and Statistical Approaches to Melodic Expectations." PhD Dissertation, University of Jyväskylä, 2004.

Eerola, Tuomas, and Carol L. Krumhansl. "Real-Time Prediction of Melodies: Continuous Predictability Judgments and Dynamic Models." In *Proceedings of the Seventh International Conference on Music Perception and Cognition*, edited by Catherine Stevens, Denis Burnham, Gary McPherson, Emery Schubert, and James Renwick, 473–476. Adelaide: Causal Productions, 2002.

Eerola, Tuomas, and Petri Toiviainen. *MIDI Toolbox: MATLAB Tools for Music Research*, 2004.

Eerola, Tuomas, and Petri Toiviainen. "MIR in Matlab:The MIDI Toolbox." In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR)*, 22–27. 2004.

Efron, Bradley, and Robert J. Tibshirani. *An Introduction to the Bootstrap*. London, UK: Chapman & Hall, 1993.

Egermann, Hauke, Marcus T. Pearce, Geraint A. Wiggins, and Stephen McAdams. "Probabilistic Models of Expectation Violation Predict Psychophysiological Emotional Responses to Live Concert Music." *Cognitive, Affective, and Behavioral Neuroscience* 13, no. 2 (2013).

Escoffier, Nicolas, and Barbara Tillmann. "The Tonal Function of a Task-Irrelevant Chord Modulates Speed of Visual Processing." *Cognition* 107, nos. 1070–1083 (2008).

Farnsworth, Paul R. "A Modification of the Lipps-Meyer Law." *Journal of Experimental Psychology: General* 9, no. 3 (1926): 253–258.

———. "Ending Preferences in Two Musical Situations." *The American Journal of Psychology* 37, no. 2 (1926): 237–240.

———. "The Effect of Repetition on Ending Preferences in Melodies." *The American Journal of Psychology* 37, no. 1 (1926): 116–122.

Fay, Richard R. "Auditory Stream Segregation in Goldfish (*Carassius auratus*)." *Hearing Research* 1230, nos. 1–2 (1998): 69–76.

Felsenstein, Joseph. *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates, 2004.

Ferguson, George A., and Yoshio Takane. *Statistical Analysis in Psychology and Education*. Sixth. Toronto, ON: McGraw-Hill, 2005.

Fiser, József, and Richard N. Aslin. "Statistical Learning of Higher-Order Temporal Structure from Visual Shape Sequences." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28, no. 3 (2002): 458–467.

Fisher, Ronald A. *Statistical Methods, Experimental Design, and Scientific Inference*. Oxford: Oxford University Press, 1990.

———. *The Design of Experiments*. 8th. New York: Hafner Publishing Company, 1971.

Fishman, Yonatan I., David H. Reser, Joseph C. Arezzo, and Mitchell Steinschneider. “Neural Correlates of Auditory Stream Segregation in Primary Auditory Cortex of the Awake Monkey.” *Hearing Research* 151, nos. 1–2 (2001): 167–187.

Forster, Kenneth I., and Michael E. J. Masson. “Special Issue: Emerging Data Analysis.” *Journal of Memory and Language* 59, no. 4 (2008): 387–556.

Fox, John, and Sanford Weisberg. *An R Companion to Applied Regression*. 2nd ed. Thousand Oaks, CA: Sage, 2011.

Fraisse, Paul. “Rhythm and Tempo.” In *The Psychology of Music*, edited by Diana Deutsch, 149–180. New York: Academy Press, 1982.

Friston, Karl. “A Theory of Cortical Responses.” *Philosophical Transactions of the Royal Society, London, Series B* 360 (2005): 815–836.

———. “The Free Energy Principle: A Unified Brain Theory?” *Nature Reviews Neuroscience* 11 (2010): 127–138.

Garner, Wendell R. “Good Patterns Have Few Alternatives: Information Theory’s Concept of Redundancy Helps in Understanding the Gestalt Concept of Goodness.” *American Scientist* 58, no. 1 (1970): 34–42.

———. *The Processing of Information and Structure*. Potomac, MD: Lawrence Erlbaum, 1974.

Gascuel, Olivier. “A Note on Sattath and Tversky’s, Saitou and Nei’s, and Studier and Keppler’s Algorithms for Inferring Phylogenies from Evolutionary Distances.” *Molecular Biology and Evolution* 11, no. 6 (1994): 961–963.

Gati, Itamar, and Amos Tversky. “Weighting Common and Distinctive Features in Perceptual and Conceptual Judgments.” *Cognitive Psychology* 16, no. 3 (1984): 341–370.

Gjerdingen, Robert O. *A Classic Turn of Phrase: Music and the Psychology of Convention*. Philadelphia: University of Pennsylvania Press, 1988.

———. “Categorization of Musical Patterns by Self-Organizing Neuronlike Networks.” *Music Perception* 7, no. 4 (1990): 339–370.

———. “Courtly Behaviors.” *Music Perception* 13, no. 3 (1996): 365–382.

Gjerdingen, Robert O. "Defining a Prototypical Utterance." *Psychomusicology* 10, no. 2 (1991): 127–139.

———. "'Historically Informed' Corpus Studies." *Music Perception* 31, no. 3 (2014): 192–204.

———. "Learning Syntactically Significant Temporal Patterns of Chords: A Masking Field Embedded in an ART 3 Architecture." *Neural Networks* 5 (1992): 551–564.

———. *Music in the Galant Style: Being an Essay on Various Schemata Characteristic of Eighteenth-Century Music*. New York: Oxford University Press, 2007.

———. "The Psychology of Music." In *The Cambridge History of Western Music Theory*, edited by Thomas Christensen, 956–981. Cambridge: Cambridge University Press, 2002.

———. "Using Connectionist Models to Explore Complex Musical Patterns." *Computer Music Journal* 13, no. 3 (1989): 67–75.

Gould, Stephen Jay. *Wonderful Life*. New York, NY: W. W. Norton & Company, 1989.

Green, Burdette, and David Butler. "From Acoustics to Tonpsychologie." In *The Cambridge History of Western Music Theory*, edited by Thomas Christensen, 246–271. Cambridge: Cambridge University Press, 2002.

Green, Douglass M. *Form in Tonal Music: An Introduction to Analysis*. 2nd ed. New York: Holt, Rinehart and Winston, 1979.

Griffiths, Thomas L. "Manifesto for a New (Computational) Cognitive Revolution." *Cognition* 135 (2015): 21–23.

Guthrie, David, Ben Allison, Wei Liu, Louise Guthrie, and Yorick Wilks. "A Closer Look at Skip-gram Modelling." In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 06)*, 1222–1225. European Language Resources Association, 2006.

Haflich, Steven. Review of *Beyond Schenkerism: The Need for Alternatives in Music Analysis*, by Eugene Narmour. *Journal of Music Theory* 23, no. 2 (1979): 287–304.

Halekoh, Ulrich, and Søren Højsgaard. "A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models – The R Package pbkrtest." *Journal of Statistical Software* 59, no. 9 (2014).

Hankinson, Andrew. “Optical Music Recognition Infrastructure for Large-Scale Music Document Analysis.” PhD Thesis, McGill University, 2014.

Hansen, Niels Chr., and Marcus T. Pearce. “Predictive Uncertainty in Auditory Sequence Processing.” *Frontiers in Psychology* 5 (2014): 1–17. doi:10.3389/fpsyg.2014.1052.

Hard, Bridgette M., Barbara Tversky, and David S. Lang. “Making Sense of Abstract Events: Building Event Schemas.” *Memory and Cognition* 34, no. 6 (2006): 1221–1235.

Harter, Courtenay L. “Bridging Common Practice and the Twentieth Century: Cadences in Prokofiev’s Piano Sonatas.” *Journal of Music Theory Pedagogy* 23 (2009): 57–77.

Hatten, Robert. “Interpreting Deception in Music.” *In Theory Only* 12, no. 5 (1992): 31–50.

Hawkins, Jeff, and Sandra Blakeslee. *On Intelligence*. New York, NY: Times Books, 2004.

Heider, Eleanor R. “Universals in Color Naming and Memory.” *Journal of Experimental Psychology* 93, no. 1 (1972): 10–20.

Helmholtz, Hermann L. F. *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. 2nd ed. Translated by Alexander J. Ellis. London: Longmans, Green, 1877/1885.

Hepokoski, James, and Warren Darcy. *Elements of Sonata Theory: Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata*. New York: Oxford University Press, 2006.

Herlinger, Jan. “Medieval Canonics.” In *The Cambridge History of Western Music Theory*, edited by Thomas Christensen, 168–192. Cambridge: Cambridge University Press, 2002.

Herskovits, Melville J. “Patterns of Negro Music.” *Illinois State Academy of Science* 34 (1941): 19–23.

Hiller, Lejaren, and Ramon Fuller. “Structure and Information in Webern’s Symphonie, Op. 21.” *Journal of Music Theory* 11, no. 1 (1967): 60–115.

Hippel, Paul von. “Questioning a Melodic Archetype: Do Listeners Use Gap-Fill to Classify Melodies?” *Music Perception* 18, no. 2 (2000): 139–153.

Huang, Milind, Doug Beeferman, and X. D. Huang. “Improved Topic-Dependent Language Modeling Using Information Retrieval Techniques.” In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*. Washington, DC: IEEE Computer Society Press, 1999.

Hubert, Lawrence. “Min and Max Hierarchical Clustering Using Asymmetric Similarity Measures.” *Psychometrika* 38, no. 1 (1973): 63–72.

Hulse, Stewart H., Scott A. MacDougall-Shackleton, and Amy B. Wisniewski. “Auditory Scene Analysis By Songbirds: Stream Segregation of Birdsong by European Starlings (*Sturnus vulgaris*).” *Journal of Comparative Psychology* 111, no. 1 (1997): 3–13.

Huron, David. “On the Virtuous and the Vexatious in an Age of Big Data.” *Music Perception* 31, no. 1 (2013): 4–9.

———. Review of *Harmony: A Psychoacoustical Approach*, by Richard Parncutt. *Psychology of Music* 19, no. 2 (1991): 219–222.

———. *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press, 2006.

———. *The Humdrum Toolkit*, 1993. `music-cog.ohio-state.edu/Humdrum/index.html`.

———. “The New Empiricism: Systematic Musicology in a Postmodern Age.” In *1999 Ernest Bloch Lectures*. University of California, Berkeley, 1999.

———. “Tone and Voice: A Derivation of the Rules of Voice-Leading from Perceptual Principles.” *Music Perception* 19, no. 1 (2001): 1–64.

Huron, David, and Elizabeth Hellmuth Margulis. “Musical Expectancy and Thrills.” In *Handbook of Music and Emotion: Theory, Research, Applications*, edited by Patrick N. Juslin and John Sloboda, 575–604. Oxford, UK: Oxford University Press, 2010.

Huron, David, and Richard Parncutt. “An Improved Model of Tonality Perception Incorporating Pitch Salience and Echoic Memory.” *Psychomusicology* 12, no. 2 (1993): 154–71.

Hutchins, Sean, and Caroline Palmer. “Repetition Priming in Music.” *Psychology of Popular Media Culture* 1(S) (2011): 69–88.

Hutchinson, William, and Leon Knopoff. "The Acoustic Component of Western Consonance." *Interface* 7, no. 1 (1978): 1–29.

———. "The Significance of the Acoustic Component of Consonance in Western Triads." *Journal of Musicological Research* 3, nos. 1-2 (1979): 5–22.

Hyer, Brian. "Tonality." In *Cambridge History of Western Music Theory*, edited by Thomas Christensen, 726–752. Cambridge: Cambridge University Press, 2002.

Hyland, Anne. "Rhetorical Closure in the First Movement of Schubert's Quartet in C Major: A Dialogue with Deformation." *Music Analysis* 28, no. 1 (2009): 111–42.

Jackson, Roland. "Gesualdo's Cadences: Innovation Set Against Convention." In *Musicologia humana: Studies in Honor of Warren and Ursula Kirkendale*, 275–289. Firenze, Italy: Leo S. Olschki, 1994.

Jaeger, T. Florian. "Categorical Data Analysis: Away from ANOVAs (Transformation or Not) and Towards Logit Mixed Models." *Journal of Memory and Language* 59 (2008): 434–446.

James, William. *The Principles of Psychology*. 2 vols. New York: Holt, 1890.

Janata, Petr. "ERP Measures Assay the Difference of Expectancy Violation of Harmonic Contexts in Music." *Journal of Cognitive Neuroscience* 7, no. 2 (1995): 153–164.

Janata, Petr, Jeffrey L. Birk, John D. Van Horn, Marc Leman, Barbara Tillmann, and Jamshed J. Bharucha. "The Cortical Topography of Tonal Structures Underlying Western Music." *Science* 298 (2002): 2167–2170.

Janata, Petr, and D. Reisberg. "Response-Time Measures as a Means of Exploring Tonal Hierarchies." *Music Perception* 6, no. 2 (1988): 161–172.

Johns, Donald. "On the Nature of Cadence." In *Music in Performance and Society: Essays in Honor of Roland Jackson*, edited by John Koegel and Malcolm S. Cole, 393–410. Warren, MI: Harmonie Park, 1997.

Johnson, Mark. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: The University of Chicago Press, 1987.

Johnson-Laird, Phil N., Olivia E. Kang, and Yuan Chang Leong. "On Musical Dissonance." *Music Perception* 30, no. 1 (2012): 19–35.

Joichi, Janet. "Closure, Context, and Hierarchical Grouping in Music: A Theoretical and Empirical Investigation." PhD Dissertation, Northwestern University, 2006.

Jonaitis, Erin, and Jenny Saffran. "Learning Harmony: The Role of Serial Statistics." *Cognitive Science* 33 (2009): 951–968.

Justus, Timothy, and Jamshed Bharucha. "Modularity in Musical Processing: The Automaticity of Harmonic Priming." *Journal of Experimental Psychology: Human Perception and Performance* 27, no. 4 (2001): 1000–1011.

Kahana, Michael Jacob. *Foundations of Human Memory*. New York, NY: Oxford University Press, 2012.

Kermode, Frank. *The Sense of an Ending: Studies in the Theory of Fiction*. London: Oxford University Press, 1968.

Kim, Seung-Goo, June Sic Kim, and Chun Kee Chung. "The Effect of Conditional Probability of Chord Progression on Brain Response: An MEG Study." *PLoS ONE* 6, no. 2 (2011): 1–9. doi:10.1371/journal.pone.0017337.

Kloepper, Tobias H., and Daniel H. Huson. "Drawing Explicit Phylogenetic Networks and their Integration into SplitsTree." *BMC Evolutionary Biology* 8, no. 22 (2008). doi:10.1186/1471-2148-8-22.

Knopoff, Leon, and William Hutchinson. "Entropy as a Measure of Style: The Influence of Sample Length." *Journal of Music Theory* 27, no. 1 (1983): 75–97.

———. "Information Theory for Musical Continua." *Journal of Music Theory* 25, no. 1 (1981): 17–44.

Koch, Heinrich Christoph. *Introductory Essay on Musical Composition*. Translated by Nancy Kovaleff Baker. New Haven, CT: Yale University press, 1983.

Koelsch, Stefan, Thomas Fritz, Katrin Schulze, David Alsop, and Gottfried Schlaug. "Adults and Children Processing Music: An fMRI Study." *NeuroImage* 25 (2005): 1068–1076.

Koelsch, Stefan, Tobias Grossmann, Thomas C. Gunter, Anja Hahne, Erich Schröger, and Angela D. Friederici. "Children Processing Music: Electric Brain Responses Reveal Musical Competence and Gender Differences." *Journal of Cognitive Neuroscience* 15, no. 5 (2003): 683–693.

Koelsch, Stefan, Tomas Gunter, Angela D. Friederici, and Erich Schröger. "Brain Indices of Music Processing: "Nonmusicians" Are Musical." *Journal of Cognitive Neuroscience* 12, no. 3 (2000): 520–541.

Koelsch, Stefan, S. Jentschke, D. Sammler, and D. Mietchen. "Untangling Syntactic and Sensory Processing: An ERP Study of Music Perception." *Psychophysiology* 44, no. 3 (2007): 476–490.

Koelsch, Stefan, Simone Kilches, Nikolaus Steinbeis, and Stefanie Schelinski. "Effects of Unexpected Chords and of Performer's Expression on Brain Responses and Electrodermal Activity." *PLoS ONE* 3, no. 7 (2008): 1–10.

Koelsch, Stefan, and Juul Mulder. "Electric Brain Responses to Inappropriate Harmonies During Listening to Expressive Music." *Clinical Neurophysiology* 113 (2002): 862–869.

Koelsch, Stefan, Schmidt Björn-Helmer Schmidt, and Julia Kansok. "Effects of Musical Expertise on the Early Right Anterior Negativity: An Event-Related Brain Potential Study." *Psychophysiology* 39 (2002): 657–663.

Koffka, Kurt. *Principles of Gestalt Psychology*. New York: Harcourt, Brace & Co., 1935.

Kohonen, Teuvo. *Self-Organizing Maps*. Berlin, Germany: Springer, 1995.

Krebs, Harald. *Fantasy Pieces: Metrical Dissonance in the Music of Robert Schumann*. New York: Oxford University Press, 1999.

Krumhansl, Carol L. *Cognitive Foundations of Musical Pitch*. New York, NY: Oxford University Press, 1990.

Krumhansl, Carol L., Jamshed J. Bharucha, and Edward J. Kessler. "Perceived Harmonic Structure of Chords in Three Related Musical Keys." *Journal of Experimental Psychology: Human Perception and Performance* 8, no. 1 (1982): 24–36.

Krumhansl, Carol L., and Edward Kessler. "Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys." *Psychological Review* 89, no. 4 (1982): 334–368.

Kucera, Henry, and W. Nelson Francis. *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press, 1967.

Kurby, Chistopher A., and Jeffrey M. Zacks. "Segmentation in the Perception and Memory of Events." *Trends in Cognitive Sciences* 12, no. 2 (2008): 72–79.

Ladinig, Olivia, Henkjan Honing, Gábor Háden, and István Winkler. "Probing Attentive and Preattentive Emergent Meter in Adult Listeners Without Extensive Music Training." *Music Perception* 26, no. 4 (2009): 377–386.

Lamont, Alexandra, and Nicola Dibben. "Motivic Structure and the Perception of Similarity." *Music Perception* 18, no. 3 (2001): 245–274.

Large, Edward W., Ji Chui Kim, Nichole Kristine Flaig, Jamshed J. Bharucha, and Carol L. Krumhansl. "A Neurodynamic Account of Musical Tonality." *Music Perception* 33, no. 3 (2016): 319–331.

Large, Edward W., and Caroline Palmer. "Perceiving Temporal Regularity in Music." *Cognitive Science* 26, no. 1 (2002): 1–37.

Larson, Steve. "Continuations as Completions: Studying Melodic Expectations in the Creative Microdomain Seek Well." In *Music, Gestalt, and Computing: Studies in Cognitive and Systematic Musicology*, edited by Marc Leman, 321–334. Berlin: Springer-Verlag, 1997.

———. *Musical Forces: Motion, Metaphor, and Meaning in Music*. Bloomington and Indianapolis: Indiana University Press, 2012.

———. "On Rudolf Arnheim's Contribution to Music Theory." *The Journal of Aesthetic Education* 27, no. 4 (1993): 97–104.

Latham, Edward. "Drei Nebensonnen: Forte's Linear-Motivic Analysis, Korngold's Die Tote Stadt, and Schubert's Winterreise as Visions of Closure." *Gamut* 2, no. 1 (2009): 299–345.

Leino, Sakari, Elvira Brattico, Mari Tervaniemi, and Peter Vuust. "Representation of Harmony Rules in the Human Brain: Further Evidence From Event-Related Potentials." *Brain Research* 1142 (2007): 169–177.

Leman, Marc. "An Auditory Model of the Role of Short-Term Memory in Probe-Tone Ratings." *Music Perception* 17, no. 4 (2000): 481–509.

Leman, Marc, Micheline Lesaffre, and Koen Tanghe. *Toolbox for Perception-based Music Analysis*. Ghent: IPEM-Department of Musicology, Ghent University, 2001. http://www.ipem.ugent.be/Toolbox/IT_PaperMeeting.pdf.

Lenth, Russell V. "lsmeans: Least-Squares Means. R package version 2.00-4," 2013. http://CRAN.R-project.org/package=lsmeans.

Lerdahl, Fred, and Ray Jackendoff. *A Generative Theory of Tonal Music*. Cambridge, MA: The MIT Press, 1983.

Lerdahl, Fred, and Carol L. Krumhansl. "Modeling Tonal Tension." *Music Perception* 24, no. 4 (2007): 329–66.

Lewin, David. *Generalised Musical Intervals and Transformations*. New Haven, CT: Yale University Press, 1987.

London, Justin. "Building a Representative Corpus of Classical Music." *Music Perception* 31, no. 1 (2013): 68–90.

———. *Hearing in Time: Psychological Aspects of Musical Meter*. New York: Oxford University Press, 2004.

———. "Rhythm in Twentieth-Century Theory." In *The Cambridge History of Western Music Theory*, edited by Thomas Christensen, 695–725. Cambridge, UK: Cambridge University Press, 2002.

Loui, Psyche. "Acquiring a New Musical System." PhD Dissertation, University of California, Berkeley, 2007.

———. "Humans Rapidly Learn Grammatical Structure in a New Musical Scale." *Music Perception* 27, no. 5 (2010): 377–388.

Loui, Psyche, Tineke Grent-'t-Jong, Dana Torpey, and Marty Woldorff. "Effects of Attention on the Neural Processing of Harmonic Syntax in Western Music." *Cognitive Brain Research* 25 (2005): 678–687.

Loui, Psyche, and David Wessel. "Harmonic Expectation and Affect in Western Music: Effects of Attention and Training." *Perception and Psychophysics* 69, no. 7 (2007): 1084–1092.

MacKay, David J. C. *Information Theory, Inference, and Learning Algorithms*. Cambridge, UK: Cambridge University Press, 2003.

Maess, Burkhard, Stefan Koelsch, Thomas C. Gunter, and Angela D. Friederici. "Musical Syntax Is Processed in Broca's Area: An MEG Study." *Nature Neuroscience* 4, no. 5 (2001): 540–545.

Malin, Yonatan. *Songs in Motion: Rhythm and Meter in the German Lied*. New York: Oxford University Press, 2010.

Mandler, George. "Organization, Memory, and Mental Structures." In *Memory Organization and Structure*, edited by C. Richard Puff, 303–319. New York, NY: Academic Press, 1979.

Mandler, Jean M. "Categorical and Schematic Organization in Memory." In *Memory Organization and Structure*, edited by C. Richard Puff, 259–299. New York, NY: Academic Press, 1979.

Manning, Christopher D., and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press, 1999.

Margulis, Elizabeth Hellmuth. "A Model of Melodic Expectation." *Music Perception* 22, no. 4 (2005): 663–714.

———. "Aesthetic Responses to Repetition in Unfamiliar Music." *Empirical Studies of the Arts* 31, no. 1 (2013): 45–57.

———. "Melodic Expectation: A Discussion and Model." PhD Dissertation, Columbia University, 2003.

———. "Musical Repetition Detection Across Multiple Exposures." *Music Perception* 29, no. 4 (2012): 377–385.

———. "Surprise and Listening Ahead: Analytic Engagements with Musical Tendencies." *Music Theory Spectrum* 29, no. 2 (2007): 197–217.

Margulis, Elizabeth Hellmuth, and Andrew P. Beatty. "Musical Style, Psychoaesthetics, and Prospects for Entropy as an Analytic Tool." *Computer Music Journal* 32, no. 4 (2008): 64–78.

Margulis, Elizabeth Hellmuth, and William H. Levine. "Timbre Priming Effects and Expectation in Melody." *Journal of New Music Research* 35, no. 2 (2006): 175–182.

Marmel, Frédéric, and Barbara Tillmann. "Tonal Priming Beyond Tonics." *Music Perception* 26, no. 3 (2009): 211–221.

Marmel, Frédéric, Barbara Tillmann, and Charles Delbé. "Priming in Melody Perception: Tracking Down the Strength of Cognitive Expectations." *Journal of Experimental Psychology: Human Perception and Performance* 36, no. 4 (2010): 1016–1028.

Marmel, Frédéric, Barbara Tillmann, and W. Jay Dowling. "Tonal Expectations Influence Pitch Perception." *Perception and Psychophysics* 70, no. 5 (2008): 841–852.

Marsden, Alan. "Interrogating Melodic Similarity: A Definitive Phenomenon or the Product of Interpretation?" *Journal of New Music Research* 41, no. 4 (2012): 323–335.

Marston, Nicholas. Review of *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven*, by William E. Caplin. *Music Analysis* 20, no. 1 (2001): 143–149.

Martin, Frederick N., and Craig A. Champlin. "Reconsidering the Limits of Normal Hearing." *Journal of the American Academy of Audiology* 11 (2000): 64–66.

Martin, Nathan John. "Rameau and Rousseau: Harmony and History in the Age of Reason." PhD Dissertation, McGill University, 2008.

Martin, Nathan John, and Julie Pedneault-Deslauriers. "The Mozartean Half Cadence." In *What Is a cadence? Theoretical and Analytical Perspectives on Cadences in the Classical Repertoire*, edited by Markus Neuwirth and Pieter Bergé, 185–214. Leuven: Leuven University Press, 2015.

Matthews, Max V., John R. Pierce, and Linda A. Roberts. “Harmony and New Scales.” In *Harmony and Tonality*, edited by Johan Sundberg, 59–84. Stockholm: Royal Swedish Academy of Music, publ. no. 54, 1987.

McAdams, Stephen. “Music: A Science of the Mind?” *Contemporary Music Review* 2 (1987): 1–61.

———. “Psychological Constraints on Form-Bearing Dimensions in Music.” *Contemporary Music Review* 4 (1989): 181–198.

———. “Recognition of Sounds Sources and Events.” In *Thinking in Sound: The Cognitive Psychology of Human Audition*, edited by Stephen McAdams and Emmanuel Bigand, 146–198. Oxford: Oxford University Press, 1993.

———. “Spectral Fusion, Spectral Parsing and the Formation of Auditory Images.” PhD Dissertation, Stanford University, 1984.

———. “The Auditory Image: A Metaphor for Musical and Psychological Research on Auditory Organization.” In *Cognitive Processes in the Perception of Art*, edited by W. Ray Crozier and Antony J. Chapman, 289–324. Amsterdam: North-Holland, 1984.

McAdams, Stephen, and Josiane Bertoncini. “Organization and Discrimination of Repeating Sound Sequences By Newborn Infants.” *Journal of the Acoustical Society of America* 102, no. 5 (1997): 2945–2953.

McAdams, Stephen, and Carolyn Drake. “Auditory Perception and Cognition.” In *Stevens’ Handbook of Experimental Psychology*, edited by Hal Pashler and Steven Yantis, vol. 1: Sensation and Perception, 397–452. New York: Wiley, 2002.

McAdams, Stephen, Bradley W. Vines, Sandrine Vieillard, Bennett K. Smith, and Roger Reynolds. “Influences of Large-Scale Form on Continuous Ratings in Response to a Contemporary Piece in a Live Concert Setting.” *Music Perception* 22, no. 2 (2004): 297–350.

McCreless, Patrick. “The Hermeneutic Sentence and Other Literary Models of Tonal Closure.” *Indiana Theory Review* 12 (1991): 35–73.

McDermott, Josh H., Alan F. Schultz, Eduardo A. Undurraga, and Ricardo A. Godoy. “Indifference to Dissonance in Native Amazonians Reveals Cultural Variation in Music Perception.” *Nature* 535 (2016): 547–550.

McDermott, Josh, Andriana J. Lehr, and Andrew J. Oxenham. "Individual Differences Reveal the Basis of Consonance." *Current Biology* 20 (2010): 1035–1041.

McKay, Cory, and Ichiro Fujinaga. "Style-Independent Computer-Assisted Exploratory Analysis of Large Music Collections." *Journal of Interdisciplinary Music Studies* 1, no. 1 (2007): 63–85.

Meriam, Alan P. *The Anthropology of Music*. Evanston, IL: Northwestern University Press, 1964.

Meyer, Charles F. *English Corpus Linguistics: An Introduction*. New York: Cambridge University Press, 2002.

Meyer, Leonard. "A Universe of Universals." *The Journal of Musicology* 16, no. 1 (1998): 3–25.

Meyer, Leonard B. "Emotion and Meaning in Music." PhD Dissertation, The University of Chicago, 1954.

———. *Emotion and Meaning in Music*. Chicago: University of Chicago Press, 1956.

———. *Explaining Music: Essays and Explorations*. Berkeley: University of California Press, 1973.

———. "Meaning in Music and Information Theory." *The Journal of Aesthetics and Art Criticism* 15, no. 4 (1957): 412–424.

———. *Music, the Arts, and Ideas*. Chicago: The University of Chicago Press, 1967.

———. "Nature, Nurture, and Convention: The Cadential Six-Four Progression." In *The Spheres of Music*, 226–263. Chicago: The University of Chicago Press, 2000.

———. *Style and Music: Theory, History, and Ideology*. Philadelphia, PA: University of Philadelphia Press, 1989.

———. *The Spheres of Music*. Chicago: The University of Chicago Press, 2000.

Meyer, Max Friedrich. "Experimental Studies in the Psychology of Music." *The American Journal of Psychology* 14, nos. 3/4 (1903): 192–214.

Mihaescu, Radu, Dan Levy, and Lior Pachter. "Why Neighbor-Joining Works." *Algorithmica* 54, no. 1 (2009): 1–24.

Miles, Jeremy, and Mark Shevlin. *Applying Regression and Correlation*. London: Sage Publications, 2001.

Miller, George. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review* 63 (1956): 81–97.

Minsky, Marvin. "A Framework for Representing Knowledge." In *The Psychology of Computer Vision*, edited by P. H. Winston, 211–277. New York: McGraw-Hill, 1975.

Miranda, Robbin A., and Michael T. Ullman. "Double Dissociation Between Rules and Memory in Music: An Event-Related Potential Study." *Neuroimage* 38, no. 2 (2007): 331–345.

Mirka, Danuta. "Absent Cadences." *Eighteenth Century Music* 9, no. 2 (2012): 213–235.

Moffat, Alistair. "Implementing the PPM Data Compression Scheme." *IEEE Transactions on Communications* COM-38, no. 11 (1990): 1917–1921.

Moore, Brian C. J. "Pitch Perception." In *An Introduction to the Psychology of Hearing*, 2nd ed., 115–149. New York: Academic Press, 1982.

Moore, Brian C. J., Robert W. Peters, and Brian R. Glasberg. "Thresholds for the Detection of Inharmonicity in Complex Tones." *Journal of the Acoustical Society of America* 77, no. 5 (1985): 1861–1867.

Müllensiefen, Daniel, and Marc Pendzich. "Court Decisions on Music Plagiarism and the Predictive Value of Similarity Algorithms." *Musicae Scientiae* Discussion Forum 4B (2009): 257–295.

Müller, Mira, Lea Höfel, Elvira Brattico, and Thomas Jacobsen. "Aesthetic Judgments of Music in Experts and Laypersons—An ERP Study." *International Journal of Psychophysiology* 76 (2010): 40–51.

Narmour, Eugene. "Analyzing Form and Measuring Perceptual Content in Mozart's Sonata K. 282: A New Theory of Parametric Analogues." *Music Perception* 13, no. 3 (1996): 265–318.

———. *Beyond Schenkerism*. Chicago: The University of Chicago Press, 1977.

———. *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago, IL: University of Chicago Press, 1990.

———. *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model*. Chicago, IL: University of Chicago Press, 1992.

Neely, James H. "Semantic Priming and Retrieval from Lexical Memory: Evidence for Facilitatory and Inhibitory Processes." *Memory & Cognition* 4, no. 5 (1976): 648–654.

———. "Semantic Priming and Retrieval from Lexical Memory: Roles of Inhibitionless Spreading Activation and Limited-Capacity Attention." *Journal of Experimental Psychology: General* 106, no. 3 (1977): 226–254.

Neisser, Ulric. *Cognition and Reality*. San Francisco: W. H. Freeman and Company, 1976.

———. *Cognitive Psychology*. New York, NY: Taylor & Francis, 1967/2014.

Nettheim, Nigel. "A Bibliography of Statistical Applications in Musicology." *Musicology Australia* 20 (1997): 94–106.

Neuwirth, Markus. "*Fuggir la cadenza*, or the Art of Avoiding Cadential Closure: Physiognomy and Functions of Deceptive Cadences in the Classical Repertoire." In *What Is a Cadence? Theoretical and Analytical Perspectives on Cadences in the Classical Repertoire*, edited by Markus Neuwirth and Pieter Bergé, 117–156. Leuven: Leuven University Press, 2015.

———. "Recomposed Recapitulations in the Sonata-Form Movements of Joseph Haydn and his Contemporaries." PhD Dissertation, Leuven University, 2013.

Noorden, Leon van. "Temporal Coherence in the Perception of Tone Sequences." PhD Dissertation, Institute for Perception Research, 1975.

Okada, Akinori. "An Asymmetric Cluster Analysis Study of Car Switching Data." In *Data Analysis: Studies in Classification, Data Analysis, and Knowledge Organization*, edited by Wolfgang Gaul, Otto Opitz, and Martin Schader, 495–504. Berlin, Germany: Springer Berlin Heidelberg, 2000.

Okada, Akinori, and Takeyoshi Iwamoto. "University Enrollment Flow Among the Japanese Prefetures." *Behaviormetrika* 23, no. 2 (1996): 169–185.

Omigie, Diana, Marcus T. Pearce, and Lauren Stewart. "Tracking of Pitch Probabilities in Congenital Amusia." *Neuropsychologia* 50 (2012): 1483–1493.

Omigie, Diana, Marcus T. Pearce, Victoria J. Williamson, and Lauren Stewart. "Electrophysiological Correlates of Melodic Processing in Congenital Amusia." *Neuropsychologia* 51 (2013): 1749–1762.

Otsuka, Asuka, Shinya Kuriki, Noboru Murata, and Toshikazu Hasegawa. "Neuromagnetic Responses to Chords are Modified by Preceding Musical Scale." *Neuroscience Research* 60 (2008): 50–55.

Palmer, Caroline, and Carol L. Krumhansl. "Mental Representations for Musical Meter." *Journal of Experimental Psychology: Human Perception and Performance* 16, no. 4 (1990): 728–41.

Parncutt, Richard. "A Perceptual Model of Pulse Salience and Metrical Accent in Musical Rhythms." *Music Perception* 11, no. 4 (1994): 409–464.

———. *Harmony: A Psychoacoustical Approach*. Berlin: Springer-Verlag, 1989.

Parncutt, Richard, and Albert S. Bregman. "Tone Profiles Following Short Chord Progressions: Top-Down or Bottom-Up?" *Music Perception* 18, no. 1 (2000): 25–57.

Patel, Aniruddh D. *Music, Language, and the Brain*. Oxford: Oxford University Press, 2008.

Patel, Aniruddh D., Edward Gibson, Jennifer Ratner, Mireille Besson, and Phillip J. Holcomb. "Processing Syntactic Relations in Language and Music: An Event-Related Potential Study." *Journal of Cognitive Neuroscience* 10, no. 6 (1998): 717–733.

Pearce, Marcus T. "The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition." PhD Dissertation, City University, London, 2005.

Pearce, Marcus T., Darrell Conklin, and Geraint A. Wiggins. "Methods for Combining Statistical Models of Music." In *Computer Music Modelling and Retrieval*, edited by U. K. Wiil, 295–312. Heidelberg, Germany: Springer Verlag, 2005.

Pearce, Marcus T., Daniel Müllensiefen, and Geraint A. Wiggins. "The Role of Expectation and Probabilistic Learning in Auditory Boundary Perception: A Model Comparison." *Perception* 39 (2010): 1367–1391.

Pearce, Marcus T., María Herrojo Ruiz, Selina Kapasi, Geraint A. Wiggins, and Joydeep Bhattacharya. “Unsupervised Statistical Learning Underpins Computational, Behavioural, and Neural Manifestations of Musical Expectation.” *NeuroImage* 50 (2010): 302–313.

Pearce, Marcus T., and Geraint A. Wiggins. “Auditory Expectation: The Information Dynamics of Music Perception and Cognition.” *Topics in Cognitive Science* 4 (2012): 625–652.

———. “Expectation in Melody: The Influence of Context and Learning.” *Music Perception* 23, no. 5 (2006): 377–405.

———. “Improved Methods for Statistical Modelling of Monophonic Music.” *Journal of New Music Research* 33, no. 4 (2004): 367–385.

Peebles, Crystal. “The Role of Segmentation and Expectation in the Perception of Closure.” Dissertation, Florida State University, 2011.

Pinkerton, Richard C. “Information Theory and Melody.” *Scientific American* 194, no. 2 (1956): 77–86.

Piston, Walter. *Harmony*. 3rd ed. New York: W. W. Norton & Company, 1962.

Plantinga, Judy, and Laurel J. Trainor. “Memory for Melody: Infants Use a Relative Pitch Code.” *Cognition* 98 (2005): 1–11.

Plomp, Reinier, and Willem J. M. Levelt. “Tonal Consonance and Critical Bandwidth.” *Journal of the Acoustical Society of America* 38, no. 4 (1965): 548–560.

Posner, Michael I. “Empirical Studies of Prototypes.” In *Noun Classes and Categorization*, edited by Colette Craig, 53–61. Amsterdam: John Benjamin, 1986.

Potter, Keith, Geraint A. Wiggins, and Marcus T. Pearce. “Towards Greater Objectivity in Music Theory: Information-Dynamic Analysis of Minimalist Music.” *Musicae Scientiae* 11, no. 2 (2007): 295–324.

Poulin-Charronnat, Benedicte, and Emmanuel Bigand. “The Influence of Voice Leading on Harmonic Priming.” *Music Perception* 22, no. 4 (2005): 613–627.

Povel, Dirk Jan. “Internal Representation of Simple Temporal Patterns.” *Journal of Experimental Psychology: Human Perception and Performance* 7, no. 1 (1981): 3–18.

Pylyshyn, Zenon. *Computation and Cognition*. Cambridge, MA: MIT Press, 1985.

Quené, Hugo, and Huub van den Bergh. “Examples of Mixed-Effects Modeling with Crossed Random Effects and with Binomial Data.” *Journal of Memory and Language* 59 (2008): 413–425.

Quinn, Ian. “Are Pitch-Class Profiles Really Key for Key.” *Zeitschrift der Gesellschaft der Musiktheorie* 7 (2010): 151–163.

———. “Fuzzy Extensions to the Theory of Contour.” *Music Theory Spectrum* 19, no. 2 (1997): 232–263.

———. “General Equal-Tempered Harmony (Introduction and Part I).” *Perspectives of New Music* 44, no. 2 (2006): 114–158.

———. “General Equal-Tempered Harmony: Parts 2 and 3.” *Perspectives of New Music* 45, no. 1 (2007): 4–63.

———. “Listening to Similarity Relations.” *Perspectives of New Music* 39, no. 2 (2001): 108–158.

Quinn, Ian, and Panayotis Mavromatis. “Voice-Leading Prototypes and Harmonic Function in Two Chorale Corpora.” In *Mathematics and Computation in Music*, edited by Carlos Agon, Moreno Andreatta, Gérard Assayag, Emmanuel Amiot, Jean Bresson, and John Mandereau, 230–240. Heidelberg: Springer, 2011.

Rameau, Jean-Phillippe. *Treatise on Harmony*. Translated by Phillip Gossett. New York: Dover, 1971. Originally published as *Traité l'harmonie* (Paris: Ballard, 1722).

Randel, Don. “Emerging Triadic Tonality in the Fifteenth Century.” *The Musical Quarterly* 57, no. 1 (1971): 73–86.

Rasch, Rudolf A. “The Perception of Simultaneous Notes such as in Polyphonic Music.” *Acustica* 40 (1978): 21–33.

Ratner, Leonard G. *Classic Music: Expression, Form, and Style*. New York: Schirmer Books, 1980.

———. *Harmony: Structure and Style*. New York: McGraw-Hill, 1962.

Ratz, Erwin. *Einführung in die musikalische Formenlehre: Über Formprizipien in den Inventionen und Fugen J. S. Bachs und ihre Bedeutung für die Kompositionstechik Beethovens*. 3rd ed. Vienna: Universal, 1973.

Reber, Arthur S. "Implicit Learning and Tacit Knowledge." *Journal of Experimental Psychology: General* 118, no. 3 (1989): 219–235.

———. *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*. Oxford: Oxford University Press, 1993.

———. "Implicit Learning of Artificial Grammars." MA Thesis, Brown University, 1965.

———. "Implicit Learning of Artificial Grammars." *Journal of Verbal Learning and Verbal Behavior* 6 (1967): 855–863.

———. "The Cognitive Unconscious: An Evolutionary Perspective." *Consciousness and Cognition* 1, no. 2 (1992): 93–133.

Richards, Mark. "Closure in Classical Themes: The Role of Melody and Texture in Cadences, Closural Function, and the Separated Cadence." *Intersections: Canadian Journal of Music* 31, no. 1 (2010): 25–45.

Rockstro, William S. "Cadence." In *Grove's Dictionary of Music and Musicians*, edited by J. A. Fuller Maitland, 1:434–441. London, 1908-1910.

Roese, Neal J., and Jeffrey W. Sherman. "Expectancy." In *Social Psychology: Handbook of Basic Principles*, 2nd ed., edited by Arie W. Kruglanski and E. Tory Higgins, 91–115. New York: The Guilford Press, 2013.

Rohrmeier, Martin, and Ian Cross. "Statistical Properties of Tonal Harmony in Bach's Chorales." In *Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC)*, edited by Ken'ichi Miyazaki, Yuzuru Hiraga, Mayumi Adachi, Yoshitaka Nakajima, and Minoru Tsuzki, 619–627. Sapporo, Japan, 2008.

———. "Tacit Tonality: Implicit Learning of Context-Free Harmonic Structure." In *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music*, edited by Jukka Louhivuori, Tuomas Eerola, Suvi Saarikallio, Tommi Himberg, and Päivi-Sisko Eerola, 443–452. Jyväskylä: University of Jyväskylä, 2009.

Rohrmeier, Martin, Patrick Rebuschat, and Ian Cross. "Incidental and Online Learning of Melodic Structure." *Consciousness and Cognition* 20 (2011): 214–222.

Rosch, Eleanor. "Cognitive Reference Points." *Cognitive Psychology* 7 (1975): 532–547.

———. "Principles of Categorization." In *Cognition and Categorization*, edited by E. Rosch and B. B. Lloyd, 28–48. Hillsdale, NJ: Erlbaum, 1978.

Rosch, Eleanor H. "Natural Categories." *Cognitive Psychology* 4 (1973): 328–350.

Rosch, Eleanor, and Carolyn B. Mervis. "Family Resemblances: Studies in the Internal Structure of Categories." *Cognitive Psychology* 7 (1975): 573–605.

Rosen, Charles. *The Classical Style: Haydn, Mozart, Beethoven*. New York: Norton, 1972.

Rosner, Burton, and Eugene Narmour. "Harmonic Closure: Music Theory and Perception." *Music Perception* 9, no. 4 (1992): 383–412.

Rothstein, William. *Phrase Rhythm in Tonal Music*. New York: Schirmer Books, 1989.

Rumelhart, David E. "Notes on a Schema for Stories." In *Representation and Understanding: Studies in Cognitive Science*, edited by Daniel G. Bobrow and Allan M. Collins, 211–236. New York: Academic Press, 1975.

———. "Schemata: The Building Blocks of Cognition." In *Theoretical Issues in Reading Comprehension*, edited by Rand J. Spiro, Bertram C. Bruce, and William F. Brewer, 33–58. Hillsdale, NJ: Erlbaum, 1980.

Rumelhart, David E., and Andrew Ortony. "The Representation of Knowledge in Memory." In *Schooling and the Acquisition of Knowledge*, edited by Richard C. Anderson, Rand J. Spiro, and William E. Mongague, 99–135. Hillsdale, NJ: Erlbaum, 1977.

Saffran, Jenny R. "Statistical Language Learning: Mechanisms and Constraints." *Current Directions in Psychological Science* 12, no. 4 (2003): 110–114.

Saffran, Jenny R., Elizabeth K. Johnson, Richard N. Aslin, and L. Newport Elissa. "Statistical Learning of Tone Sequences By Human Infants and Adults." *Cognition* 70 (1999): 27–52.

Saffran, Jenny R., Seth D. Pollak, Rebecca L. Seibel, and Anna Shkolnik. "Dog is a Dog is a Dog: Infant Rule Learning is Not Specific to Language." *Cognition* 105, no. 3 (2007): 669–680.

Saitou, Naruya, and Masatoshi Nei. "The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees." *Molecular Biology and Evolution* 4, no. 4 (1987): 406–425.

Sanguinetti, Giorgio. *The Art of Partimento: History, Theory, and Practice*. Oxford, UK: Oxford University Press, 2012.

Saslaw, Janna. "Forces, Containers, and Paths: The Role of Body-Derived Image Schemas in the Conceptualization of Music." *Journal of Music Theory* 40, no. 2 (1996): 217–243.

Sattath, Shmuel, and Amos Tversky. "Additive Similarity Trees." *Psychometrika* 42, no. 3 (1977): 319–345.

Schank, Roger C., and Robert P. Abelson. *Scripts, Plans, Goals and Understanding: An Enquiry into Human Knowledge Structures*. Hillsdale, NJ: Erlbaum, 1977.

Schellenberg, E. Glenn. "Expectancy in Melody: Tests of the Implication-Realization Model." *Cognition* 58 (1996): 75–125.

———. "Simplifying the Implication-Realization Model of Melodic Expectancy." *Music Perception* 14, no. 3 (1997): 295–318.

Schellenberg, E. Glenn, Emmanuel Bigand, Benedicte Poulin-Charronnat, and Catherine Stevens. "Children's Implicit Knowledge of Western Music." *Developmental Science* 8, no. 6 (2005): 551–566.

Schmalfeldt, Janet. "Cadential Processes: The Evaded Cadence and the 'One More Time' Technique." *Journal of Musicological Research* 12, no. 1 (1992): 1–52.

———. *In the Process of Becoming: Analytic and Philosophical Perspectives on Form in Early Nineteenth-Century Music*. New York: Oxford University Press, 2011.

Schmuckler, Mark A. "Expectancy Effects in Memory for Melodies." *Canadian Journal of Experimental Psychology* 51, no. 4 (1997): 292–306.

Schmuckler, Mark A. "Expectation in Music: Additivity of Melodic and Harmonic Processes." PhD Dissertation, Cornell University, 1988.

———. "Expectation in Music: Investigation of Melodic and Harmonic Processes." *Music Perception* 7, no. 2 (1989): 109–150.

Schnupp, Jan, Israel Nelken, and Andrew King. *Auditory Neuroscience: Making Sense of Sound*. Cambridge, MA: The MIT Press, 2011.

Schoenberg, Arnold. *Fundamentals of Musical Composition*. London: Faber & Faber, 1967.

———. *Harmonielehre*. Translated by Roy Carter. Berkeley and Los Angeles, CA: University of California Press, 1978.

———. *The Musical Idea and the Logic, Technique, and Art of Its Presentation*. Translated by Patricia Carpenter and Severine Neff. New York: Columbia University Press, 1995.

Schubert, Emery. "Measurement and Time Series Analysis of Emotion in Music." PhD Dissertation, University of New South Wales, 1999.

Sears, David. "The Perception of Cadential Closure." In *What is a Cadence? Theoretical and Analytical Perspectives on Cadences in the Classical Repertoire*, edited by Markus Neuwirth and Pieter Bergé, 251–283. Leuven: Leuven University Press, 2015.

Sears, David, William E. Caplin, and Stephen McAdams. "Perceiving the Classical Cadence." *Music Perception* 31, no. 5 (2014): 397–417.

Selfridge-Field, Eleanor, ed. *Beyond MIDI: The Handbook of Musical Codes*. Cambridge, MA: MIT Press, 1997.

Sethares, William A. "Consonace-Based Spectral Mappings." *Computer Music Journal* 22, no. 1 (1998): 56–72.

Shannon, Claude. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27, no. 3 (1948): 379–423.

Shannon, Claude E. "Prediction and Entropy of Printed English." *Bell System Technical Journal* 30 (1951): 50–64.

Shepard, Roger N. "Psychophysical Complementarity." In *Perceptual Organization*, edited by Michael Kubovy and James R. Pomerantz, 279–341. Hillsdale, N.J.: Erlbaum, 1981.

Simon, Herbert A., and Richard K. Sumner. "Pattern in Music." In *Machine Models of Music*, edited by Stephan M. Schwanauer and David A. Levitt, 83–110. Cambridge, MA: MIT Press, 1993.

Simons, M., H. Ney, and S. C. Martin. "Distant Bigram Language Modeling Using Maximum Entropy." In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, 787–790. Washington, DC: IEEE Computer Society Press, 1997.

Simonton, Dean. "Computer Content Analysis of Melodic Structure: Classical Composers and Their Compositions." *Psychology of Music* 22, no. 1 (1994): 31–43.

———. "Melodic Structure and Note Transition Probabilities: A Content Analysis of 15,618 Classical Themes." *Psychology of Music* 12, no. 3 (1984): 3–16.

Sloboda, John A. *The Musical Mind: The Cognitive Psychology of Music*. Oxford: Oxford University Press, 1985.

Smith, Barbara H. *Poetic Closure: A Study of How Poems End*. Chicago: The University of Chicago Press, 1968.

Smith, Bennett K. "PsiExp: An Environment for Psychoacoustic Experimentation Using the IRCAM Musical Workstation." Paper presented at the Society for Music Perception and Cognition, 1995, Berkeley, CA.

Sneath, Peter H. A., and Robert R. Sokal. *Numerical Taxonomy: The Principles and Practice of Numerical Classification*. San Francisco: Freeman, 1973.

Snyder, Bob. *Music and Memory: An Introduction*. Cambridge, MA: The MIT Press, 2000.

Snyder, John L. "Entropy as a Measure of Musical Style: The Influence of A Priori Assumptions." *Music Theory Spectrum* 12, no. 1 (1990): 121–160.

Spitzer, Michael. "The Retransition as Sign: Listener-Oriented Approaches to Tonal Closure in Haydn's Sonata-form Movements." *Journal of the Royal Musical Association* 121, no. 1 (1996): 11–45.

Squire, Larry R. "Mechanisms of Memory." *Science* 232, no. 4758 (1986): 1612–1619.

Steege, Benjamin. *Helmholtz and the Modern Listener*. Cambridge: Cambridge University Press, 2012.

Stefanowitsch, Anatol, and Stefan Th. Gries. "Collostructions: Investigating the Interaction of Words and Constructions." *International Journal of Corpus Linguistics* 8 (2003): 209–243.

Steinbeis, Nikolaus, Stefan Koelsch, and John A. Sloboda. "The Role of Harmonic Expectancy Violations in Musical Emotions: Evidence from Subjective, Physiological, and Neural Responses." *Journal of Cognitive Neuroscience* 18, no. 8 (2006): 1380–1393.

Studier, James A., and Karl J. Keppler. "A Note on the Neighbor-Joining Algorithm of Saitou and Nei." *Molecular Biology and Evolution* 5, no. 6 (1988): 729–731.

Symons, James. "Temporal Regularity as a Key to Uncovering Statistically Significant Schemas in an Eighteenth-Century Corpus." *Paper Presented at the Society for Music Theory*, 2012, New Orleans, LA.

Takeuchi, Akinobu, Takayuki Saito, and Hiroshi Yadohisa. "Asymmetric Agglomerative Hierarchical Clustering Algorithms and Their Evaluations." *Journal of Classification* 24 (2007): 123–143.

Takumi, Satoshi, and Sadaaki Miyamoto. "Top-down vs Bottom-up Methods of Linkage for Asymmetric Agglomerative Hierarchical Clustering." In *International Conference on Granular Computing*, 459–464. IEEE, 2012.

Tan, Norma, Rita Aiello, and Thomas G. Bever. "Harmonic Structure as a Determinant of Melodic Organization." *Memory and Cognition* 9, no. 5 (1981): 533–539.

Tekman, Hasan Gürkan, and Jamshed J. Bharucha. "Implicit Knowledge Versus Psychoacoustic Similarity in Priming of Chords." *Journal of Experimental Psychology: Human Perception and Performance* 24, no. 1 (1998): 252–260.

———. "Time Course of Chord Priming." *Perception and Psychophysics* 51 (1992): 33–39.

Temperley, David. "Pitch-Class Distribution and the Identification of Key." *Music Perception* 25, no. 3 (2008): 193–212.

———. Review of *Music in the Galant Style*, by Robert O. Gjerdingen. *Journal of Music Theory* 50, no. 2 (2006): 277–290.

———. *The Cognition of Basic Musical Structures*. Cambridge, MA: The MIT Press, 2004.

———. "What's Key for Key? The Krumhansl-Schmuckler Key-Finding Algorithm Reconsidered." *Music Perception* 17, no. 1 (1999): 65–100.

Temperley, David, and Leigh VanHandel. "Introduction to the Special Issue on Corpus Methods." *Music Perception* 31, no. 1 (2013): 1–3.

Tenenbaum, Joshua B., Charles Kemp, Thomas L. Griffiths, and Noah D. Goodman. "How to Grow a Mind: Statistics, Structure, and Abstraction." *Science* 331 (2011): 1279–1285.

Terhardt, Ernst. "Pitch, Consonance, and Harmony." *Journal of the Acoustical Society of America* 55 (1974): 1061–1069.

———. "The Concept of Musical Consonance: A Link between Music and Psychoacoustics." *Music Perception* 1, no. 3 (1984): 276–295.

Terhardt, Ernst, Gerhard Stoll, and Manfred Seewann. "Algorithm for Extraction of Pitch and Pitch Salience from Complex Tonal Signals." *Journal of the Acoustical Society of America* 71, no. 3 (1982): 679–688.

———. "Pitch of Complex Signals According to Virtual-Pitch Theory: Tests, Examples, and Predictions." *Journal of the Acoustical Society of America* 71 (1982): 671–678.

Tillmann, Barbara, Jamshed J. Bharucha, and Emmanuel Bigand. "Implicit Learning of Tonality: A Self-organizing Approach." *Psychological Review* 107, no. 4 (2000): 885–913.

Tillmann, Barbara, and Emmanuel Bigand. "A Comparative Review of Priming Effects in Language and Music." In *Language, Vision and Music*, edited by Paul McKevitt, Seán Ó Nualláin, and Conn Mulvihill, 231–240. John Benjamins, 2002.

———. "Global Context Effect in Normal and Scrambled Musical Sequences." *Journal of Experimental Psychology: Human Perception and Performance* 27, no. 5 (2001): 1185–1196.

———. "Musical Priming: Schematic Expectations Resist Repetition Priming." In *Proceedings of the 8th International Conference on Music Perception & Cognition*, edited by Scott D. Lipscomb, Richard Ashley, Robert O. Gjerdingen, and Peter Webster, 674–676. Adelaide, Australia: Causal Productions, 2004.

Tillmann, Barbara, and Emmanuel Bigand. “Musical Structure Processing After Repeated Listening: Schematic Expectations Resist Veridical Expectations.” *Musicae Scientiae*, 2010, 33–47.

Tillmann, Barbara, Emmanuel Bigand, Nicolas Escoffier, and Philippe Lalitte. “The Influence of Musical Relatedness on Timbre Discrimination.” *European Journal of Cognitive Psychology* 18, no. 3 (2006): 343–358.

Tillmann, Barbara, Emmanuel Bigand, and Francois Madurell. “Local versus Global Processing of Harmonic Cadences in the Solution of Musical Puzzles.” *Psychological Research* 61 (1998): 157–174.

Tillmann, Barbara, Petr Janata, and Jamshed J. Bharucha. “Activation of the Inferior Frontal Cortex in Musical Priming.” *Cognitive Brain Research* 16 (2003): 145–161.

Tillmann, Barbara, Petr Janata, Jeffrey L. Birk, and Jamshed J. Bharucha. “Tonal Centers and Expectancy: Facilitation or Inhibition of Chords at the Top of the Harmonic Hierarchy.” *Journal of Experimental Psychology: Human Perception and Performance* 34, no. 4 (2008): 1031–1043.

Tillmann, Barbara, Petr Janata, Jeffrey Birk, and Jamshed J. Bharucha. “The Costs and Benefits of Tonal Centers for Chord Processing.” *Journal of Experimental Psychology: Human Perception and Performance* 29, no. 2 (2003): 470–482.

Tillmann, Barbara, and Frédéric Marmel. “Musical Expectations Within Chord Sequences: Facilitation Due to Tonal Stability Without Closure Effects.” *Psychomusicology* 23, no. 1 (2013): 1–5.

Tillmann, Barbara, and Stephen McAdams. “Implicit Learning of Musical Timbre Sequences: Statistical Regularities Confronted with Acoustical (Dis)Similarities.” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30, no. 5 (2004): 1131–1142.

Tillmann, Barbara, Isabelle Peretz, Emmanuel Bigand, and Nathalie Gosselin. “Harmonic Priming in an Amusic Patient: The Power of Implicit Tasks.” *Cognitive Neuropsychology* 24, no. 6 (2007): 603–622.

Tillmann, Barbara, and Bénédicte Poulin-Charronnat. "Auditory Expectations for Newly Acquired Structures." *The Quarterly Journal of Experimental Psychology* 63, no. 8 (2010): 1646–1664.

Tirovolas, Anna, and Daniel J. Levitin. "Analysis of Empirical Articles in Music Perception." *Music Perception* 29, no. 1 (2011): 23–36.

Toiviainen, Petri, and Tuomas Eerola. "Autocorrelation in Meter Induction: The Role of Accent Structure." *Journal of the Acoustical Society of America* 119, no. 2 (2006): 1164–1170.

Toiviainen, Petri, and Carol L. Krumhansl. "Measuring and Modeling Real-Time Responses to Music: The Dynamics of Tonality Induction." *Perception* 32 (2003): 741–766.

Trainor, Laurel J., and Sandra E. Trehub. "Key Membership and Implied Harmony in Western Tonal Music: Developmental Perspectives." *Perception and Psychophysics* 56, no. 2 (1994): 125–132.

Tversky, Amos. "Features of Similarity." *Psychological Review* 84, no. 4 (1977): 327–352.

Tversky, Amos, and J. Wesley Hutchinson. "Nearest Neighbor Analysis of Psychological Spaces." *Psychological Review* 93, no. 1 (1986): 3–22.

Tymoczko, Dmitri. Review of *Music21: a Toolkit for Computer-aided Musicology*, by Michael Cuthbert. *Music Theory Online* 19, no. 3 (2013).

Vallières, Michel. "Beginnings, Middles, and Ends: Perception of Intrinsic Formal Functionality in the Piano Sonatas of W. A. Mozart." PhD Dissertation, McGill University, 2011.

Vallières, Michel, Daphne Tan, William E. Caplin, and Stephen McAdams. "Perception of Intrinsic Formal Functionality: An Empirical Investigation of Mozart's Materials." *Journal of Interdisciplinary Music Studies* 3, nos. 1-2 (2009): 17–43.

VanHandel, Leigh. "The Role of Phrase Location in Key Identification by Pitch Class Distribution." In *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*, edited by Emilios Cambouropoulos, Costas Tsougras, Panayotis Mavromatis, and Konstantinos Pastiadis, 1069–1073. Thessaloniki, Greece: School of Music Studies, Aristotle University of Thessaloniki, 2012.

Venables, William N., and Brian D. Ripley. *Modern Applied Statistics with S*. New York: Springer, 2002.

Villarreal, Eduardo G.A., Elvira Brattico, Sakari Leino, Leif Østergaard, and Peter Vuust. "Distinct Neural Responses to Chord Violations: A Multiple Source Analysis Study." *Brain Research* 1389 (2011): 103–114.

Vos, Piet G., and Jim M. Troost. "Ascending and Descending Melodic Intervals: Statistical Findings and Their Perceptual Relevance." *Music Perception* 6, no. 4 (1989): 383–396.

Vos, Piet, and Dennis Pasveer. "Goodness Ratings of Melodic Openings and Closures." *Perception and Psychophysics* 64, no. 4 (2002): 631–639.

Ward, Joe H. "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association* 58, no. 301 (1963): 236–244.

Weber, Reinhard. "The Continuous Loudness Judgement of Temporally Variable Sounds with an 'Analog' Category Procedure." In *Fifth Oldenburg Symposium on Psychological Acoustics*, edited by A. Schick, J. Hellbrück, and R. Weber, 267–294. Oldenburg: BIS, 1991.

Webster, James. "Binary Variants of Sonata Form in Early Haydn Instrumental Music." In *Internationaler Joseph Haydn Kongress Wien 1982*, edited by Eva Badura-Skoda, 127–135. Munich: Henle, 1986.

Weiser, Margaret. "Rating Cadence Stability: The Effects of Chord Structure, Tonal Context, and Musical Training." PhD Dissertation, McMaster University, 1992.

Welch, Bernard L. "On the Comparison of Several Mean Values: An Alternative Approach." *Biometrika* 38, nos. 3/4 (1951): 330–336.

West, Brady T., Kathleen B. Welch, and Andrzej T. Galecki. *Linear Mixed Models. A Practical Guide using Statistical Software*. Boca Raton: Chapman & Hall/CRC, 2007.

White, Christopher W. "An Alphabet-Reduction Algorithm for Chordal n-Grams." In *Proceedings of the 4th International Conference on Mathematics and Computation in Music*, 201–212. Springer, 2013.

———. "Some Statistical Properties of Tonality, 1650-1900." PhD Dissertation, Yale University, 2013.

Wild, Jonathan. "A Review of the Humdrum Toolkit: UNIX Tools for Musical Research, created by David Huron." *Music Theory Online* 2.7 (1996).

Wingfield, Paul. "Beyond 'Norms and Deformations': Towards a Theory of Sonata Form as Reception History." *Music Analysis* 27, no. 1 (2008): 137–177.

Winkler, István, Elena Kushnerenko, János Horváth, Rita Čeponienė, Vineta Fellman, Minna Huotilainen, Risto Näätänen, and Elyse Sussman. "Newborn Infants Can Organize the Auditory World." *Proceedings of the National Academy of Sciences* 100, no. 20 (2003): 11812–11815.

Witten, Ian H., and Timothy C. Bell. "The Zero-Frequency Problem: Estimating the Probabilities of Novel Events in Adaptive Text Compression." *IEEE Transactions on Information Theory* 37, no. 4 (1991): 1085–1094.

Witten, Ian H., Leonard C. Manzara, and Darrell Conklin. "Comparing Human and Computational Models of Music Prediction." *Computer Music Journal* 18, no. 1 (1992): 70–80.

Wright, James K., and Albert S. Bregman. "Auditory Stream Segregation and the Control of Dissonance in Polyphonic Music." *Contemporary Music Review* 2 (1987): 63–92.

Yang, Albert C.-C, C.-K Peng, H.-W Yien, and Ary L. Goldberger. "Information Categorization Approach to Literary Authorship Disputes." *Physica A: Statistical Mechanics and its Applications* 329, nos. 3-4 (2003): 473–483.

Yates, Graeme K. "Cochlear Structure and Function." In *Hearing*, edited by Brian C. J. Moore, 41–74. San Diego: Academic Press, 1995.

Youngblood, Joseph. "Style as Information." *Journal of Music Theory* 2, no. 1 (1958): 24–35.

Zbikowski, Lawrence M. *Conceptualizing Music: Cognitive Structure, Theory, and Analysis*. Oxford: Oxford University Press, 2002.

Zipf, George Kingsley. *Human Behavior and the Principle of Least Effort*. Oxford, UK: Addison-Wesley, 1949.

Ziv, Naomi, and Zohar Eitan. "Themes as Prototypes: Similarity Judgments and Categorization Tasks in Musical Contexts." *Musicae Scientiae* Discussion Form 4A (2007): 99–133.