STOCHASTIC PROCESSES & DATABASE-DRIVEN MUSICOLOGY

JOHN ASHLEY BURGOYNE

Music Technology Area
Department of Music Research
Schulich School of Music

McGill University · Montréal · Québec

October 2011

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

Clever yes
but mongrel
statistics are with us

— *Karen Mac Cormack, 'Multi-Mentional'*

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

FOR MORE THAN A DECADE, music information science and musicology have been at what Nicholas Cook has described as a 'moment of opportunity' for collaboration on database-driven musicology. The literature contains relatively few examples of mathematical tools that are suitable for analysing temporally structured data like music, however, and there are surprisingly few large databases of music that contain information at the semantic levels of interest to musicologists. This dissertation compiles a bibliography of the most important concepts from probability and statistics for analysing musical data, reviews how previous researchers have used statistics to study temporal relationships in music, and presents a new corpus of carefully curated chord labels from more than 1000 popular songs from the latter half of the twentieth century, as ranked by *Billboard* magazine's Hot 100 chart. The corpus is based on a careful sampling methodology that maintained cost efficiency while ensuring that the corpus is well suited to drawing conclusions about how harmonic practises may have evolved over time and to what extent they may have affected songs' popularity. This dissertation also introduces techniques new to the musicological community for analysing databases of this size and scope, most importantly the Dirichlet-multinomial distribution and constraint-based structure learning for causal Bayesian networks. The analysis confirms some common intuitions about harmonic practises in popular music and suggests several intriguing directions for further research.

RÉSUMÉ

Depuis plus d'une décennie, la science de l'information de la musique et la musicologie sont à ce que Nicholas Cook décrit comme « un moment clé » en ce qui concerne une collaboration pouvant mener à une réelle science de la musique fondé sur l'analyse de large quantité de données. Toutefois, la littérature comporte rélativement peu d'exemples d'outils mathématiques qui conviendraient à l'analyse des données qui, comme les données musicales, ont des dépendances temporelles, et il y a très peu de bases de données qui contiennent des informations avec la richesse sémantique intéressant d'ordinaire les musicologues. Cette thèse assemble une bibliographie des concepts les plus importants de la probabilité et de la statistique pour analyser les données musicales, revisite la manière dont les chercheurs précédents se servaient de la statistique pour étudier les rapports temporels, et présente un nouveau corpus soigneusement préparé contenant les transcriptions d'accords pour plus de 1000 chansons populaires de la deuxième moitié du xx^{e} siècle, figurant du « Hot 100 » de la revue *Billboard*. Le corpus résulte d'une méthodologie d'échantillonnage qui optimise les coûts et s'assure que le corpus conviendrait à tirer des conclusions montrant comment les pratiques harmoniques ont pu évoluer au fils du temps et dans quelle mesure elles peuvent avoir une incidence sur la popularité des chansons. Cette thèse introduit aussi quelques techniques qui sont nouvelles en musicologie pour analyser les bases de données d'une telle taille et d'une telle portée ; les plus importantes parmi ces techniques sont la distribution multinomiale de Dirichlet et l'apprentissage via contraintes de structure des

réseaux causaux de Bayes. L'analyse confirme quelques intuitions courantes concernant les pratiques harmoniques de la musique populaire et suggère quelques voies intéressantes pour la recherche à venir.

PREFACE

The work in this thesis benefited from the direct help of a number of parties. Foremost, I want to thank my advisor, Prof. Ichiro Fujinaga, heartily and sincerely for having been more involved and more supportive than most graduate students can imagine. Prof. Fujinaga was also seminal in helping me to secure funding from the Social Sciences and Humanities Research Council of Canada (SSHRC) to develop the *Billboard* corpus. Prof. Jonathan Wild was a co-applicant on the grant and was invaluable in helping to develop the file format for transcribing songs, in auditioning and training the annotators, and in helping to maintain quality throughout the process. Dr. Rhonda Amsel provided guidance in developing the sampling methodology and provided the critical insight that adjacent chart slots were probably interchangeable.

The project also obviously could not have happened without the jazz musicians who worked as chord annotators and some administrative support in managing so many people. Andrew Hankinson, Jessica Thompson, Will Carroll, and Alastair Porter helped in varying degrees to get the web site running and keep it working. Reiko Yamada helped me to acquire the audio and manage the auditions; she also did about half of the reconciliation work when there were differences of opinion between the two annotators. Tristan Paxton did the other half of the reconciliation work, and was a notably prolific transcriber himself. Mireille Boily and Paul Van Dyk were the other two annotators who transcribed hundreds of songs each, but we also appreciated the contributions from Eric Couture, Ted Crosby, Robert

Jordon, Jason Stillman, Joel Kerr, Taylor Donaldson, Jared Greeve, Patrick Hart, Mark McDonald, Marie-Claire Durand, Andrew Urbina, and Ben Henriques. Elizabeth Llewellyn and Mikaela Miller time-aligned all of these annotations. Gabriel Vigliensoni and Rena Raghunanan provided invaluable support managing the payroll.

Although I am not enough of a jazz musician to be qualified to do annotations myself, at least not for the more difficult songs in the corpus, I was responsible for all design decisions and coordinated the team and its technological tools throughout its compilation of the corpus. I spent a good deal of time standardising the corpus after the fact, however, and all of the code for parsing and working with it, as well as the code for generating confidence intervals after Bailey, Agresti, and Coull, is my own (Bailey 1980; Agresti & Coull 1998).

Finally, I wish to thank my labmates and friends, Andrew Hankinson, Johanna Devaney, Gabriel Vigliensoni, Jason Hockman, Alastair Porter, and Greg Burlet for keeping our lab one of the most pleasant workplaces at McGill, even when a very stressed dissertation-writing student was working alongside them.

NOTATIONAL GLOSSARY

NUMBER THEORY

$\mathbf{n}!$ factorial, $\prod_{i=1}^{n} i$

$\Gamma(x)$ the gamma function, $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$

$\binom{n}{k}$ number of distinct combinations of k objects from a pool of n objects, $\binom{n}{k} = \frac{n!}{n!(n-k)!}$

LOGICAL RELATIONS

$\triangleq$ definition

$\wedge$ logical conjunction

$\vee$ logical disjunction

SET THEORY & ANALYSIS

$\varnothing$ the empty set

$\triangle^n$ the n-simplex, i.e., $\{\boldsymbol{\pi} \in \mathbf{R}^{n+1} : \sum_{i=0}^{n} \pi_i = 1 \wedge \forall_i \pi_i \geq 0\}$

$\aleph_0$ the cardinality of $\mathbf{N}$

(a, b) the open interval between a and b, $\{x : a < x < b\}$

$[a, b]$ the closed interval between a and b, $\{x : a \leq x \leq b\}$

A a set

$\mathrm{A} \times \mathrm{B}$ the Cartesian product of sets A and B, $\{(a, b) : a \in \mathrm{A} \wedge b \in \mathrm{B}\}$

$|\mathrm{A}|$ the cardinality of a set A

$\mathscr{A}$ a family of sets A

$\mathfrak{B}(\mathrm{A})$ the Borel σ-algebra of a set A

$\mathfrak{c}$ the cardinality of $\mathbf{R}$

$\mathbf{N}$ the set of natural numbers $\{0, 1, \ldots\}$

$\mathfrak{P}(\mathrm{A})$ the power set (i.e., the family of all subsets) of a set A

$\mathbf{R}$ the set of real numbers

$\mathfrak{T}$ a topology (family of open sets)

$\mathbf{x}$ the vector $\{x_1, x_2, \ldots\}$

GRAPH THEORY

$\pi(v)$ the set of all parents of a vertex v

$\rho(v)$ the set of all descendants of a vertex v

D a set of all directed edges $v \to w$ in a graph

U a set of all undirected edges $v - w$ in a graph

V a set of all vertices in a graph

PROBABILITY & MEASURE THEORY

δ_ω the Dirac (point-mass) measure with respect to an outcome ω

Ω an outcome (sample) space, i.e., the set of all possible atomic outcomes

$\mathbf{E}$ the expected value

$\mathfrak{E}$, $\mathfrak{F}$, $\mathfrak{G}$ σ-algebras

$F_X(x; \boldsymbol{\theta})$ the cumulative distribution function with parameters $\boldsymbol{\theta}$ for a real-valued random variable X

$f_X(x; \boldsymbol{\theta})$ the probability density function or mass function with parameters $\boldsymbol{\theta}$ for a random variable X

$\mathbf{P}$ a probability measure

X a random variable

$X^{(t)}$ a random variable corresponding to time t in a stochastic process

$X^{1:t}$ the set of random variables $\{X^{(1)}, X^{(2)}, \ldots, X^{(t)}\}$

$\mathbf{X}$ a random vector $\{X_1, X_2, \ldots, X_n\}$

STATISTICS

$\hat{\theta}$ an estimate of a parameter θ

$L(y_0, y)$ a loss function for a classifier choosing y when the correct classification is y_0

$\mathscr{L}_{\mathrm{X}}(\boldsymbol{\theta}; x)$ the likelihood function of a random variable X

$$\mathscr{L}_{\mathrm{X}}(\boldsymbol{\theta}; x) \triangleq f_{\mathrm{X}}(x; \boldsymbol{\theta})$$

$\ell_{\mathrm{X}}(\boldsymbol{\theta}; x)$ the log likelihood function of a random variable X

$$\ell_{\mathrm{X}}(\boldsymbol{\theta}; x) \triangleq \log \mathscr{L}(\boldsymbol{\theta}; x)$$

1 DATABASE-DRIVEN MUSICOLOGY

In 2005, Nicholas Cook opened his invited talk to the Sixth International Conference on Music Information Retrieval (ISMIR) by noting that 'we stand at a moment of opportunity' for historical musicologists and music information scientists to work together to revitalise the sub-discipline of machine-assisted empirical musicology (Cook 2005). Two great obstacles have delayed realising this moment of opportunity. One has been the difficulty these two groups have communicating with each other: the needs and jargon of musicologists are alien to most music information scientists, who tend to originate from computer engineering, and engineers' statistical models are opaque to most musicologists, who do not normally acquire sophisticated training in mathematics. The other is that it is nearly impossible to undertake large-scale empirical research in music at the moment given the current paucity of organised, machine-interpretable data at the levels of reduction of most interest to musicians: notes, harmonies, and phrases. We need tools, preferably automated tools, to help us compile and distribute such databases of music.

1.1 SUGGESTIVE VS. PSYCHOLOGICAL MUSIC THEORIES

Cook is a strong proponent of empirical methods in musicology, having argued that:

> there is no useful distinction to be drawn between empirical and nonempirical musicology, because there can be no such

> thing as a truly non-empirical musicology; what is at issue is the extent to which musicological discourse is grounded on empirical observation, and conversely the extent to which observation is regulated by discourse.
>
> (Cook & Clarke 2004)

In some sense, music theory is this discourse that may regulate observation (Wiggins, Müllensiefen & Pearce 2010), but it has proved to be a frustrating field to define precisely. Jean-Jacques Nattiez's seminal *Musicologie générale et sémiologie* posited six different families of music analysis (1987, pp.175–79), but David Temperley has argued that in practise, most research in music theory falls into just two of these families. One, which Temperley describes as 'psychological', 'attempts to describe listeners' unconscious mental representations of music' (1999, p. 68) and corresponds to Nattiez's family 3 ('deductive esthesics'); the other, which Temperley describes as 'suggestive', has a didactic aim to encourage listeners to hear music differently and corresponds to Nattiez's family 4, 'inductive esthesics'.

Of these two approaches, the psychological one has been more traditionally associated with empiricism, fruitfully borrowing many of the techniques developed for the social sciences. In particular, as any issue of *Music Perception* will show, the psychological approach to music theory has made great use of statistical methods and probability. There is no inherent reason, however, that suggestive music theories cannot or should not use statistics. Leonard Meyer, usually cited as one of the founding figures of the psychological approach to music theory, was just as strong a supporter using empirical methods and statistics for suggestive approaches:

> 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. That is, instead of employing informal impressions of the relative frequencies of casually defined traits (in the case of Wagner's music, for instance, what actually is the relative frequency of deceptive cadences? in what ways is their occurrence correlated with that of bar forms? how exactly is chromaticism defined? and so on), 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. I should add that I have no doubt about the value of employing computers in such studies, 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 – something that most of us seldom do. (Meyer 1989, p. 64)

In short, Meyer argued for what I shall call *database-driven musicology:* a process of musicological inquiry that begins with the compilation of large corpora of musical data and then uses these data as a means to quantify baseline musicological principles. Perhaps more importantly, database-

driven methods can quantify precisely how the more notable examples from musical history deviate from these baselines and thus help musicologists hypothesise why.

The field was not quick to adopt this suggestion of Meyer's. In 1999, David Huron argued at the Ernest Bloch lectures of the University of California, Berkeley, that musicology needed to make a transition, as psychology had, from operating as a strictly 'data-poor' field to one that recognised places where it had the potential to become data-rich (Huron 1999). As noted above, this moment of opportunity was still yet to be realised in 2005. Only recently has the field begun to open in this direction (Anagnostopoulou & Buteau 2010), but one of the fundamental obstacles remains that there simply are not enough well-curated databases available that are usable for studying questions of musicological interest.

1.2 AN INTRODUCTORY EXAMPLE: BRAHMS OP. 51, NO. 1

Two broad sets of tasks are associated with databases for empirical research in music: one set for creating such databases and another for making optimal use of them. Creating a database is generally a process of *classification*, i.e., reducing music data from the forms in which it is most readily available, such as recordings or scores, to forms that are more semantically meaningful, such as chords or contrapuntal schemata. Once the database exists, researchers normally want to use techniques of *inference* to draw musicological conclusions from it. One could reasonably presume that classification and inference would be independent, given that they seem to be separated in time – figure 1.1 illustrates this apparent structure of

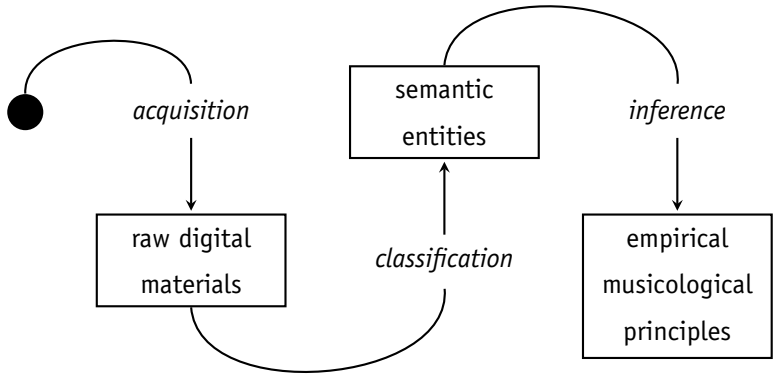


Figure 1.1 · Apparent dependencies in large-scale empirical music research. This representation suggests that classification and inference are completely separated in time, but in practise, they can and should inform each other.

dependencies – but these sets of tasks are more intimately linked than they might seem. Their interaction is one of the riper areas for collaboration between engineers and musicologists.

It is possible and indeed traditional to undertake both classification and inference by hand. Such approaches are well-suited to tightly-specified questions such as close analyses of a particular works, e.g., Alan Forte's famous analysis (1983) of the first movement of Brahms's String Quartet in C Minor (op. 51, no. 1). This analysis begins by outlining a formal methodology for a classification task to reduce a single movement to a collection of pitch-based *motives*:

1. The motive is primarily an intervallic event, distinct from any particular pitch manifestation.
2. The original pitch or pitch-class representation of a motive

is of singular importance, however, and we call this referential function of a particular form *pitch-specific* which means that the motive designated consists of the same pitches as the original form of the motive; and the term *pitch-class specific* means that the motive is recurring, but at some level of octave transposition with respect to the original form.

3. The boundary interval of the motive is its most salient feature. The internal structure is variable or may even be absent in some representation of the motive. The boundary interval may undergo octave inversion, or – more appropriate to Brahms's usage – expansion by octave displacement, as major third and minor sixth.

4. A motive may be transformed without losing its basic identity. The transformations which Brahms uses are retrograde, inversion, and retrograde inversion. He also uses a transformation which I will call *minor to major* or *major to minor*, depending upon the circumstances.

(p. 474)

Forte continues by cataloguing all of the motives and transformations he can find in the movement, or in other words, Forte classifies all of the motives in the movement manually, yielding eleven distinct sets, each set containing a motive in prime form and possibly one or more transformations. Finally, Forte examines how the instances of each set of motives change throughout the piece in order to draw conclusions about its structure, a manual form of

inference. Among his more notable conclusions are the claims that each of two sets of motives, the so-called α- and σ-motives, are defining structures within the movement.

David Huron (2001) hypothesised that Forte's α-motive was too general to be able to distinguish the first movement of op. 51, no. 1, from Brahms's output in general. To test this hypothesis, he compared the prevalence of the α-motive in this movement to its prevalence in the first movements of Brahms's other two string quartets (op. 51, no. 2, and op. 67). Undertaking even this relatively modest comparison by hand, however, would have crossed an unappealing threshold to tedium: calculating the empirical prevalence of α-motives relative to all others in op. 51, no. 1, alone involves classifying 7 045 pairs of melodic intervals, and across all three first movements, there are 21 155 such pairs to consider (¶ 32). Huron used a computer and his own Humdrum Toolkit (1995) to extract descriptive statistics from machine-readable encodings of the movements in question and concluded that Forte's α-motive is no more prevalent in the first movement of op. 51, no. 1, than it is in the first movements of Brahms's other quartets. Huron's analysis further suggests that Forte's decision to downplay the importance of rhythmic features during classification (note that rhythm appears nowhere in the passage quoted above, although Forte does comment upon it to some extent in the text that follows it) limits the range of inferences that are available afterward. Certain combinations of pitch motives and rhythmic patterns prove to have more discriminatory power than the α-motive alone, and moreover, Forte's assumption of transformational equivalence (assumption 4) appears questionable, at least with with respect to the α-motive.

Huron employed computers to avoid a prohibitive number of tedious calculations that would have been necessary to answer his musical question. There are many other questions that could have been asked, however: what, for example, of the other ten set of motives in Forte's analysis? Are there notable motivic patterns that are absent from the analysis? As computational power has improved, it has become feasible to allow computers even to broaden the list of possible questions beyond the scope of what a single human researcher or group of researchers could undertake alone. As part of a special issue of the *Journal of Mathematics and Music* dedicated specifically to computational approaches to analysing op. 51, no. 1, Darrell Conklin (2010) has applied techniques from data mining to list all possible motivic patterns according to Forte's classification methodology and to rank them by the extent to which they distinguish this movement from the first movements of Brahms's other quartets. Consistent with Huron's findings, Conklin's analysis show that Forte's α set collectively is a poor distinguishing feature, but overall, the analysis validates Forte's classification. At least one element of each of Forte's eleven sets of motives appear among the most prominent distinguishing patterns, and the list is headed by an instance of the σ-motive, to which Forte's analysis ascribes particular importance. The computer is also able to find three motivic patterns that are absent from both Forte's analysis and Huron's, including one that appears to be of significant structural importance (a descending minor second followed by an ascending major third).

Forte's own conclusion, however, suggests that he was looking less for features that made the first movement of op. 51, no. 1, unique and more for general conclusions about Brahms's music:

> In the work which is the subject of this analysis specific pitch classes and dyads serve throughout to initiate motions, to terminate them, and to refer to musical events already completed or forthcoming. Although this feature is by no means restricted to the music of Brahms, the elegance and subtlety with which Brahms negotiates motivic relations leave him without peer. (pp. 501–2)

In order to test the assertion that Brahms is 'without peer' in his treatment of musical motives, one would need to apply the classification methodologies of Forte, Huron, or Conklin to a much larger selection of examples from Brahms's output and an equally large selection of music from Brahms's potential peers. All of these selections would need to be machine-readable, and that presents another limitation: encoding musical scores in machine-readable formats is itself a manual and time-consuming process. In order to arrive at a sufficiently large amount of data for inference, not only must the process of classification be automated, classification must also start not from machine-readable format that has been encoded by a human in semantic terms but rather from a digital acquisition process that is capable of treating large quantities of documents quickly and at relatively low cost (MacMillan, Droettboom & Fujinaga 2002; Baird 2003; Bruder et al. 2003). Figure 1.1 includes this critical step in large-scale empirical research in music.

Large-scale acquisition processes like these are typically bound to a physical reality rather than a semantic one. Musical scores, for example, will in most cases be scanned to raw digital images that represent the amount of light reflected from each of many small regions of the page. Musical

sound is usually recorded or transferred to digital audio, which stores a representation of a sound wave or waves to be retransmitted to a human ear. For most questions of interest, the gap between these physical and semantic realities is too great to describe with absolute precision, and so one must accept that automatic classification processes will operate with some degree of uncertainty. Depending on the details of the questions asked and the processes used, that uncertainty may or may not propagate to the inference step. Either way, two complementary goals emerge: one to understand how much uncertainty is inherent in a particular classification or inference procedure and the other to minimise or at least control that uncertainty to the greatest extent possible.

One of the easiest ways to reduce the uncertainty (and thus improve the accuracy) of an automated classification procedure is to improve its quantitative model of the semantic space, e.g., the temporal relations among harmonies. The best way to learn a quantitative model of the semantic space is from inference, and the larger the database, the more precise that inference can be. A virtuous cycle emerges whereby larger databases enable more accurate automatic classifiers, which lessens the burden of generating still larger databases, which then enable even more accurate automatic classifiers, and so on. Given the dearth of good data at the present, most projects in database-driven musicology need to find a way onto this cycle.

1.3 CONCLUSION AND CHAPTER OUTLINE

The goal of this dissertation is to show how to drive a project onto that virtuous cycle and, for one important musicological problem, lift the problem up

to that virtuous cycle. In particular, it will investigate how hypotheses about musical structure can be encoded directly into a flexible, high-performing statistical framework known as the *graphical model*. Graphical models enable a nearly one-to-one correspondence between hypotheses about the data (e.g., that dominant chords tend to lead to tonic chords) and the mathematical apparatus, a correspondence that should enable freer communication between researchers with more musical training and those with more statistical training. One of the most popular types of graphical model can also encode and help to uncover causal relationships, which make these models especially rich tools for collaborative inquiry. Moreover, one can use existing databases to train graphical models, and their accuracy and precision of graphical models improve with size of the training database. As such, graphical models can be employed to start the virtuous cycle necessary to build large databases cost-effectively.

Given that this thesis is written for evaluation in a music department, I have done my best to assume minimal background in machine learning, which cannot be taken for granted even among music technologists. There is a glossary listing the most important mathematical notations used throughout the documents, and chapter 2 is a review of the most relevant concepts from probability theory and statistics. Surprisingly, I found that there was no one source or even small set of sources that brought together the right combination of material on probability, causality, stochastic processes, and classification for effective work in database-driven musicology, and I believe that collecting these concepts in a single chapter is perhaps one of the most important contributions of this thesis. Nonetheless, I must acknowledge that readers with a weaker mathematical background will

find this chapter to be a difficult read.

On the other hand, again given that this thesis is written for evaluation in a music department, I have taken for granted a knowledge of basic music-theoretical concepts up to the level of an intermediate course in harmony. Some interested parties may not in fact have that background, and I direct such readers to standard textbooks such as Edward Aldwell and Carl Schachter's (2003) or Robert Gauldin's (1997). Chapter 3 brings the thesis back to a more musicological domain, comprehensively reviewing how the statistical principles presented in the preceding chapter have been applied, implicitly or explicitly, to musicological problems. It begins with examples of inference alone and then engages the larger set of literature involved in classification problems that engage time-dependent structures in music.

Chapter 4 presents an example of using these techniques to enable serious database-driven musicology for an active area of musicological research: harmony in popular music. Leading a team of researchers, I was able to produce a corpus of popular harmony of unprecedented size, scope, and detail. This chapter describes the process of generating that corpus and presents the first results of using it for statistical inference.

The dissertation closes with a brief summary of the issues presented and an outline of the many possible areas for future work.

2 STOCHASTIC PROCESSES

PROBABILITY, CAUSALITY, AND STATISTICS underly most principled approaches to database-driven musicology, just as they do most data-oriented research in the social sciences. Probability quantifies uncertainty, and as discussed in the previous chapter, there is almost always uncertainty in empirical research. Sometimes raw probabilities are meaningful on their own, but often they are more interesting when interpreted relative to a causal structure that allows one to consider questions of what probabilities would be (or would have been) if one could take (or could have taken) certain actions to intervene in the process. Statistical methods are the tools to extract information about probabilities and causal relationships from a limited set of observations, such as a database of musical selections and their properties.

This chapter begins with an overview of some basic concepts in probability and their philosophical underpinnings (§ 2.1), which are necessary for understanding most of the mathematics throughout the document. It continues with a description of graphical models (§ 2.2), a common tool for working with complicated probabilistic relationships and linking them to notions of causality. Music unfolds over time, and so another section (§ 2.3) describes how to apply graphical models in temporal contexts. The chapter concludes with a survey of historical approaches to estimation and classification that are relevant to those that have been applied to music in the past as well as to the new approaches considered later in this thesis (§ 2.4).

2.1 BASIC PROBABILITY THEORY

Probability theory is a dense subject, and there is neither the space nor the need to treat it comprehensively here. Daphne Koller and Nir Friedman have written a very concise introduction that is targeted specifically toward the types of statistical models that appear in this thesis and in machine learning more generally (2009, pp. 15–34). Larry Wasserman included a somewhat longer and more general introduction at the beginning of his high-level survey of the most common topics in statistical inference (2004, pp. 3–46). Jeffrey Rosenthal has prepared a book-length overview with more mathematical rigor (2000), which serves as a lighter introduction to more comprehensive references in the field, e.g., Patrick Billingsley's (1995) or Sidney Resnick's (1999). The outline here will overview the foundational concepts, probability spaces and random variables, and provide some examples of how they might apply to musical domains. It concludes with a discussion of the mainstream interpretations of probability and where each may be appropriate for database-driven musicology.

✹ *Probability Spaces*

Formal theories of probability begin with the concept of a *sample space* Ω, which is a (possibly infinite) set of *outcomes* ω. Collectively, these outcomes represent the universe of all conceivable combinations of phenomena under study that are logically consistent. For musicological purposes, a very simple sample space might be the set of all diatonic pitch classes, $\Omega \triangleq \{\mathrm{C}, \mathrm{D}, \mathrm{E}, \mathrm{F}, \mathrm{G}, \mathrm{A}, \mathrm{B}\}$. A more realistic sample space, musically speaking,

might contain all conceivable fragments of music together with the details of their conception and composition, all combined with all conceivable representations of each of these musical elements in digital form together with the details of how those representations came or could come to light.

The sample space Ω is never used directly: it is only necessary to know that some theoretical sample space Ω exists. Instead, probability theory concerns itself with a family $\mathcal{F}$ of *events*, which are sets of outcomes in Ω. The family of events must be a *σ-algebra* or *σ-field*, which means that it must contain both $\varnothing$ and Ω and must also be closed under complementation, countable union, and countable intersection. The smallest σ-algebra for any sample space Ω is the trivial $(\varnothing, \Omega)$. When the cardinality of the sample space is finite or countably infinite ($|\Omega| \leq \aleph_0$), the largest σ-algebra is the power set $\mathcal{P}(\Omega)$, the family of all possible subsets. When Ω is uncountably infinite ($|\Omega| > \aleph_0$) and endowed with some topology (a family $\mathcal{T}$ of so-called *open sets* such that $\varnothing \in \mathcal{T}$, $\Omega \in \mathcal{T}$, and $\mathcal{T}$ is closed under infinite union and finite intersection), a common choice is the *Borel σ-algebra* $\mathcal{B}(\Omega)$, which is the closure of $\mathcal{T}$ under complementation, countable union, and countable intersection. For any sample space Ω endowed with a topology $\mathcal{T}$, $\mathcal{B}(\Omega)$ is the smallest σ-algebra that contains $\mathcal{T}$ (see Rudin 1976, p. 309, for one statement of this result, although it is a standard one). Despite its relatively small size, $\mathcal{B}(\Omega)$ is a popular choice because it tends to be easier to reason about it than it is to reason about alternatives and because in most cases it contains every set of practical interest (see Saxe 2002, pp. 57–58, for one explanation for this preference). Under the natural topologies, the Borel σ-algebra $\mathcal{B}(\Omega)$ of a sample space is equivalent to $\mathcal{P}(\Omega)$ when Ω is finite or countably infinite; if Ω is $\mathbf{R}$, then $\mathcal{B}(\Omega)$ is the σ-algebra

generated from the closure under complementation, countable union, and countable intersection of all open intervals $\{(a,b) : a \in \mathbf{R} \wedge b \in \mathbf{R} \wedge a \leq b\}$. Bruno de Finetti has proved that the algebraic properties of the σ-algebra alone are sufficient to derive all of the properties of Ω that are necessary for formal probability theory (1972, p. 72), and so when conceiving of a particular application of probability theory, it may make more sense to start by thinking of the σ-algebra rather than the sample space.

Returning to the simple example where $\Omega \triangleq \{C, D, E, F, G, A, B\}$, different choices of $\mathcal{F}$ allow for different types of musical questions. With a finite Ω like this one, the interpretation of an event $A \in \mathcal{F}$ is often the logical disjunction of its component outcomes $\omega \in A$ (de Finetti 1972, pp. 69–72), e.g., $\{C, E, G\}$ would be interpreted as a diatonic pitch class that could be one of C, E, or G. If the relevant question were whether a diatonic pitch class belonged to a C-major chord, then, a good choice of $\mathcal{F}$ might be $\{\varnothing, \{C, E, G\}, \{D, F, A, B\}, \Omega\}$. The Borel σ-algebra $\mathcal{B}(\Omega)$, by contrast, would allow questions about any combination of diatonic pitch classes.

The final foundational concept in formal probability theory is the *probability measure* itself, denoted here as $\mathbf{P}$, which is a function $\mathbf{P} : \mathcal{F} \to \mathbf{R}$ from the σ-algebra to the real numbers. The probability measure $\mathbf{P}$ is restricted such that $0 \leq \mathbf{P}(A) \leq 1$ for all $A \in \mathcal{F}$, $\mathbf{P}(\varnothing) = 0$, $\mathbf{P}(\Omega) = 1$, and $\mathbf{P}$ is *countably additive:*

$$\mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbf{P}(A_i) \tag{2.1}$$

where $A_1, A_2, \ldots$ are disjoint relative to Ω and all members of $\mathcal{F}$. By way

of illustration, a valid $\mathbf{P}$ for the C-major example above might be

$$\mathbf{P}(A) \triangleq \begin{cases} 0 & \text{if } A = \varnothing, \\ 1/3 & \text{if } A = \{C, E, G\}, \\ 2/3 & \text{if } A = \{D, F, A, B\}, \text{ and} \\ 1 & \text{if } A = \Omega. \end{cases} \tag{2.2}$$

On the other hand,

$$\mathbf{P}(A) \triangleq \begin{cases} 0 & \text{if } A = \varnothing, \\ 1/4 & \text{if } A = \{C, E, G\}, \\ 1/2 & \text{if } A = \{D, F, A, B\}, \text{ and} \\ 1 & \text{if } A = \Omega \end{cases} \tag{2.3}$$

would be invalid because $\{C, E, G\} \cup \{D, F, A, B\} = \Omega$ but $\mathbf{P}(\{C, E, G\}) + \mathbf{P}(\{D, F, A, B\}) = 1/4 + 1/2 = 3/4 \neq 1$. Collectively, triples $(\Omega, \mathcal{F}, \mathbf{P})$ are known as *probability triples* or *probability spaces*.

This formulation of probability just given is commonly attributed to Andrey Kolmogorov (1933) and is standard in mathematical texts. It is not entirely without controversy, particularly with respect to the notion of countable additivity. Countable additivity renders probability measures considerably easier to handle mathematically, but it entails certain restrictions on probability measures that sometimes contravene common sense. Perhaps most notoriously, it is impossible to define a probability measure over the natural numbers that would assign equal probability to each number: because the cardinality of $\mathbf{N}$ is infinite, the probability of

any individual number would have to be zero, but then

$$\mathbf{P}(\mathbf{N}) = \mathbf{P}(\bigcup_{i=1}^{\infty}\{i\}) = \sum_{i=1}^{\infty}\mathbf{P}(\{i\}) = \sum_{i}^{\infty} 0 = 0\,, \tag{2.4}$$

which contradicts the axiom that the probability of the complete space equal unity. Analogous problems arise for uncountable spaces like $\mathbf{R}$. This contradiction poses particular problems in applications that require 'non-informative' probability measures (e.g., Bishop 1995, pp. 396, 408). De Finetti has argued at length that because of such problems, it is essential to derive a mathematical notion of probability that corresponds better to its 'basic spirit' (1972, pp. 87–113) and has sketched mathematical foundations for a theory he thought would do so (1972, pp. 129–40); nonetheless, such are the mathematical complexities of these theories that they have failed to thrive (Martellotti 2001).

Random Variables

Just as the sample space Ω is often unwieldy and replaced by the σ-algebra $\S$, even $\S$ can be too fine-grained to answer most questions. Although we were able to define simple and effective probability triples for the case of diatonic pitch classes, imagine trying to define one in the more realistic example where the sample space contains all conceivable fragments of music paired with all conceivable representations of them! *Random variables* are functions from probability spaces to other measurable spaces, and the mapping is normally chosen in to reflect the intuition behind a specific probabilistic question.

More formally, given a probability space $(\Omega, \S, \mathbf{P})$ and a *measurable space* $(\Psi, \mathfrak{E})$, meaning simply that $\mathfrak{E}$ is a σ-algebra over Ψ, a random variable X is a function $X : (\Omega, \S) \to (\Psi, \mathfrak{E})$. The function X is required to be *measurable*, meaning that the 'pre-image' $X^{-1}(E) \triangleq \{\omega : X(\omega) \in E\}$ of any set E in $\mathfrak{E}$ must be a member of $\S$. One also sometimes speaks of $\sigma(X)$, which is defined to be the smallest σ-algebra that contains all possible values of X^{-1}, i.e., the smallest σ-algebra containing $\{X^{-1}(E) : E \in \mathfrak{E}\}$. Much as with probability spaces themselves, it is sufficient to specify only the algebraic properties $\mathfrak{E}$, as a compatible Ψ may be derived from them; also much as with probability spaces themselves, Borel σ-algebras are common choices for $\mathfrak{E}$.

A random variable $X : (\Omega, \S) \to (\Psi, \mathfrak{E})$ can be seen as pushing the probability measure $\mathbf{P}$ to a new probability measure μ_X on $(\Psi, \mathfrak{E})$, where for any E in $\mathfrak{E}$,

$$\mu_X(E) \triangleq \mathbf{P}\left[X^{-1}(E)\right] \ . \tag{2.5}$$

Such a measure μ is sometimes known as the *law* for the random variable. When $\Psi = \mathbf{R}$ and $\mathfrak{E} = \mathfrak{B}(\mathbf{R})$ – the canonical choice – the *distribution function* $F_X : \mathbf{R} \to [0,1]$ of a random variable X is defined as follows:

$$F_X(x) \triangleq \mu_X\left[(-\infty, x)\right] \ . \tag{2.6}$$

A *density function* $f_X : \mathbf{R} \to \mathbf{R}^+$ for such a random variable, commonly known as a *continuous* random variable, is a real-valued function such that

$$\int_a^b f_X(x)\,\mathrm{d}x = \mu_X\left[(a,b)\right] \tag{2.7}$$

for any $a \leq b$; this definition implies that $F_X(x) = \int_{-\infty}^{x} f_X(x)\,\mathrm{d}x$. To be strictly correct, the density function is a non-negative real-valued function

such that $\mu_X(B) = \int_B f_X(x)\,\lambda(dx)$, i.e., a real-valued function the integral of which in Lebesgue's sense over a Borel set B is equivalent to the measure of B under the law μ_X; f_X need not be integrable in Riemann's sense (Rosenthal 2000, p. 55). This thesis shall only consider distribution functions that are integrable in Riemann's sense, however, and for any Riemann-integrable function f, $\int_a^b f(x)\,dx = \int_a^b f(x)\,\lambda(dx)$ for any $a \leq b$ (Rudin 1976, pp. 322–24). Furthermore, if the lower limit of a Riemann-integrable f_X is well-defined, i.e., F_X exists, the definition of f_X in Riemann's sense is sufficient to derive a unique extension to a definition in Lebesgue's sense (Resnick 1999, pp. 42–56).

When the cardinality of Ψ is finite or countably infinite ($|\Psi| \leq \aleph_0$), then X is commonly known as a *discrete* random variable, and a *mass function* $f_X : \Psi \to [0, 1]$ is defined more simply as

$$f_X(\psi) \triangleq \mu_X[\{\psi\}] \tag{2.8}$$

for all $\psi \in \Psi$. Density and mass functions are conceptually quite similar, hence the identical notation, but it is important to remember that while the values of mass functions are directly interpretable as probabilities, only integrals of density functions are subject to the same interpretation. It is also possible to generalise the notion of density and mass function to other types of random variables (see Rosenthal 2000, pp. 52–56, among others), but such generalisations are unnecessary for the purposes of this thesis.

What does this mathematical machinery provide in the end? Return for a moment to Forte's assertion that 'the elegance and subtlety with which Brahms negotiates motivic relations leave him without peer'. One simple domain that springs to mind from this statement is the idea of music by

Brahms's as distinguished from music by his potential peers, in other words, $\Psi \triangleq \{$Brahms composed x, one of Brahms's potential peers composed $x\}$ with the natural choice of $\mathfrak{E}$ being the power set $\mathfrak{P}(\Psi) = \{\varnothing$, Brahms composed x, one of Brahms's potential peers composed x, either Brahms or one of his potential peers composed (i.e., $\Psi)\}$. In order to make meaningful statements about probability on this domain, there needs to be an underlying probability domain $(\Omega, \mathfrak{F})$ with a mapping (random variable) $\mathrm{X} : (\Omega, \mathfrak{F}) \to (\Psi, \mathfrak{E})$ such that each of the elements of $\mathfrak{E}$ corresponds to an element of $\mathfrak{F}$; Ω in this case might, for example, be the set of all appearances of all musical motives in the music of Brahms and his peers with $\mathfrak{P}(\Omega)$ as the σ-algebra, and X would map each motive to Ψ according to whether Brahms had composed it or not. If more than one random variable were under consideration, then $\mathfrak{F}$ would furthermore need to contain elements that were sufficient to map to the σ-algebras of all considered random variables. This underlying σ-algebra $\mathfrak{F}$ needs be endowed with a countably additive measure $\mathbf{P}$, which defines the behaviour of a mass function f_{X}. That mass function renders the uncertainty in discussing Brahms's music as compared to that of his peers quantifiable, which in turn enables a principled approach to database-driven musicology.

✵ *Frequentist vs. Bayesian Interpretations*

Even after one has imagined an underlying domain of inquiry Ω and a collection of random variables that encapsulate the most relevant aspects of it, the question remains of how exactly one should interpret this countably additive measure $\mathbf{P}$. The informal description is simple: $\mathbf{P}$ represents

probability. But what is probability? Is it a property of the physical world, a mathematical formalism, or both? When used in the context of database-driven musicology, is our notion of probability the same as the *p*-values reported in articles from the experimental sciences? Should it be?

There are many competing viewpoints on how probability ought to be interpreted. Often, the debate is presented as two-sided, with *frequentists* on one side and *Bayesians* or *subjectivists* on the other, although there are other views. Alan Hájek's article in the *Stanford Encyclopedia of Philosophy* (2010) is a short, relatively accessible summary of the major streams of thought. De Finetti wrote a biased – he was a dedicated subjectivist – but rigorous account of the different interpretations and their ramifications (1972, pp. 67–113, 147–227). After retiring from a thirty-year career in the philosophy of mathematics, Patrick Maher recently released an incomplete draft of a book summarising his own views on the merits and drawbacks of the various interpretations (2010b). This book is more neutral and more accessible than de Finetti's, but unlike de Finetti, Maher does not seek to engage the measure-theoretic formulations just presented. As any of these sources show, the interpretation of probability is one of the great debates in the philosophy of mathematics, well beyond the scope of a single section of a doctoral thesis, but this section will seek to present at least one consistent interpretation of the probabilities that arise in empirical musicology in order to inform the theories and results throughout this thesis.

One of Maher's foundational arguments, derived from Rudolf Carnap (1950), is that there are two distinct concepts, or *explicanda*, that theories of probability seek to formalise (explicate):

> Suppose you know that a coin is either two-headed or two-tailed but you have no information about which it is. The coin is about to be tossed. What is the probability that it will land heads? There are two natural answers: (i) ½; (ii) either 0 or 1. Both answers are right in some sense, though they are incompatible, so 'probability' in ordinary language must have two different senses. I'll call the sense of 'probability' in which (i) is right *inductive probability* and I'll call the sense in which (ii) is right *physical probability*. (p. 1)

Maher describes inductive probability as 'logical' and physical probability as 'empirical' (p. 10); he also explains that inductive probability is the notion of probability that is relative to available evidence and has no dependence on the facts of the world, whereas physical probability is an immutable fact about the world that is independent of the evidence one may or may not have about it (Maher 2006). Maher also explains that although under completely determined conditions, physical probabilities are always zero or unity, in the case of incompletely specified conditions, physical probabilities can also take on intermediate values (2010b, pp. 9–10). In Maher's view, some of the debate over interpretations of probability arises from misunderstanding about which of these two *explicanda* a given interpretation is seeking to explicate.

For database-driven musicology, physical probability seems to be the most appropriate *explicandum*. In the case of Forte's question, for example, one would ideally want to know how frequently 'specific pitch classes and dyads' open, close, and recur among musical events in Brahms's music and

in the music of his contemporaries. In principle, this question is determined: Brahms and his contemporaries are dead, the space of all music written during his lifetime is large but finite, and hence, with access to all of that music, one could compute exact proportions. Given a particular bar of a particular piece, the probability of a given dyad would be zero or unity; stretched over the entirety of Brahms's œuvre, the probability would be its relative proportion. Inductive probability, in contrast, makes little sense: although it certainly could be logically consistent to state that the probability of a given dyad appearing in Brahms's œuvre is something other than their actual relative frequency inasmuch as it would be possible to make such a statement without openly contradicting the axioms of probability, it is hard to see how such statements could hold much musicological relevance.

These raw relative frequencies are not, however, the only notions of probability that arise in database-driven musicology. In practise, much of the music written during Brahms's lifetime has been lost, and a researcher may not have access to a corpus that includes all of Brahms's music that has survived. Researchers must instead make educated guesses based on subsets of the music in question, and using knowledge about how the subset was chosen, one can then attempt to quantify how uncertain those guesses may be. When quantifying this second level of uncertainty, which I shall call *observational uncertainty*, either of the two *explicanda* for probability are viable. These *explicanda* overlap with the major interpretations of probability, or as Maher calls them, *explicata*.

The frequentist's *explicatum*, most famously espoused by Richard von Mises (1957; 1964) and similar to Karl Popper's propensity theory, (1959), considers probabilities to be the limiting relative frequencies of events after

an infinite number of experiments. This interpretation of probability is the usual one in scientific literature. The *explicandum* is still physical probability, as it seeks to describe the counts of actual (or at least potential) experiments. The 'experiment' in the case of database-driven musicology is *sampling*, i.e., how to choose a subset of music to examine. Under this interpretation, it is especially important to choose a methodology for sampling that would converge to the relative frequencies of the entire œuvre; the most cogent criticisms of the frequentist interpretation question whether such a sampling procedure exists (e.g., Hájek 2009).

De Finetti, in contrast lobbied strongly for the Bayesian interpretation of probability, arguing forcefully (1970; 1972) that

PROBABILITY DOESN'T EXIST.

Such a statement is consistent with Maher's argument that the Bayesian *explicandum* is not physical probability but inductive, or logical, probability given a set of evidence (2010a). From this viewpoint, probability is often described as 'rational degree of belief', but such presentations are criticised for being unable to provide more than a tautology to define 'rational'. The dominant view of probability in machine learning is Bayesian, although there are prominent exceptions. In the context of database-driven musicology, this interpretation of probability would focus on articulating what would be logically consistent to believe about the relative frequencies of an œuvre given an observed subset of the œuvre and a set of working assumptions about it.

The interpretation of probability is simply not a solved problem: the major interpretations all have potentially valid criticisms that are yet to be de-

fended for the general case. For the underlying probability spaces $(\Omega, \mathfrak{F}, \mathbf{P})$ of database-driven musicology, it does seem that the only useful interpretation is the so-called 'finite frequentism' (Hájek 1997) of raw relative frequencies; the primary contributions in this thesis are techniques for how to uncover them. For observational uncertainty, however, there is no such obviously compelling interpretation. For consistency's sake, I prefer to use the 'infinite' frequentists' perspective, which shares the same *explicandum* as finite frequentism, when handling observational uncertainty. That stated, there is no formal reason not to take a Bayesian perspective to the observational uncertainty instead – many researchers in machine learning do – and where that perspective would lead to a substantive change in method or result, I shall note it.

2.2 GRAPHICAL MODELS AND CAUSALITY

Most interesting questions about probability spaces involve uncovering relationships among multiple random variables. For example, one might ask whether the presence of Forte's α-motive (say a random variable $X : \Omega \to \{\text{true}, \text{false}\}$ where $X(\omega)$ is true if and only if ω contains the α-motive) is a distinguishing feature of music by Brahms (say a random variable $Y : \Omega \to \{\text{true}, \text{false}\}$ where $Y(\omega)$ is true if and only if ω is a complete piece of music by Brahms). The notions of *conditional probability* and *independence* facilitate working with multiple random variables. *Graphical models* are tools for encoding conditional probabilities and independence relationships among groups of random variables efficiently. This section outlines the basics of conditional probability, independence, graphical models, and

how all of these relate to causality. As with the previous section, readers interested in more detail on the mathematics behind conditional probability and independence should consult references such as Rosenthal (2000), Billingsley (1995), or Resnick (1999). For more information on graphical models, Koller and Friedman's recent compendium (2009) is an excellent resource. For more detail on causality, Judea Pearl's landmark monograph (2009) is the most comprehensive single-point reference, although the reader will need to follow some of the other citations there and in this section in order to learn about important competing theories.

✺ *Conditional Probability*

Conditional probability addresses questions of the form, 'What is the probability of Y given X?' In order to define this notion formally, it is necessary to define two other concepts first: *expected value* (or *expectation*) and the *indicator functions*.

An indicator function $\mathbf{I}_A : \Omega \to \{0, 1\}$ maps any outcome ω that is an element of A to unity and all others to zero. Indicator functions are useful because of the special mathematical properties of zero and unity.

The expected value of a real-valued random variable $X : (\Omega, \mathcal{F}) \to (\mathbf{R}, \mathcal{B}(\mathbf{R}))$ is, intuitively, an average value of the variable weighted by its density or mass function. If X is discrete, the expected value is defined in the obvious manner:

$$\mathbf{E}(X) \triangleq \sum_{i=1}^{\infty} x_i f_X(x_i) \ . \tag{2.9}$$

For continuous variables that are integrable in Riemann's sense, the defini-

tion is analogous:

$$\mathbf{E}(X) \triangleq \int_{-\infty}^{\infty} x f_X(x)\, dx \tag{2.10}$$

In the general case, one begins by defining expectation for *simple random variables*, viz., random variables that take only a finite number of values, in the same way as equation (2.9). This definition extends to random variables that take any number of non-negative values as the least upper bound over all simple random variables that are less than or equal to X: $\mathbf{E}(X) \triangleq \sup\{\mathbf{E}(Y) : Y \text{ is simple} \wedge Y \leq X\}$. For an arbitrary random variable X on $(\mathbf{R}, \mathfrak{B}(\mathbf{R}))$, one then defines two related random variables $X^+ \triangleq \max(X, 0)$ and $X^- \triangleq \max(-X, 0)$ and the expected value as $\mathbf{E}(X) \triangleq \mathbf{E}(X^+) - \mathbf{E}(X^-)$, which is often written $\int_\Omega X(\omega)\, \mathbf{P}(d\omega)$.

For a set of random variables $\{X_1 : (\Omega, \mathcal{F}) \to (\Psi_1, \mathcal{E}_1), X_2 : (\Omega, \mathcal{F}) \to (\Psi_2, \mathcal{E}_2), \ldots, X_n : (\Omega, \mathcal{F}) \to (\Psi_n, \mathcal{E}_n)\}$ (denoted with the shorthand $\mathbf{X}$ and sometimes known as a *random vector*), let $\mathcal{G}$ denote the smallest σ-algebra that contains the pre-images of all of their measurable sets, i.e, $\sigma\left[\bigcup_i \bigcup_{E \in \mathcal{E}_i} X_i^{-1}(E_i)\right]$. Formally speaking, the conditional probability of an event $A \in \mathcal{F}$ given $\mathbf{X}$, written $\mathbf{P}(A \mid \mathbf{X})$, is a random variable from $(\Omega, \mathcal{G}, \mathbf{P})$ to $([0,1], \mathfrak{B}([0,1]))$ such that for any $B \in \mathcal{G}$,

$$\mathbf{E}\left[\mathbf{P}(A \mid \mathbf{X}) \cdot \mathbf{I}_B\right] = \mathbf{P}\left(A \cap B\right) \ , \tag{2.11}$$

which by taking B to be Ω, implies that

$$\mathbf{E}\left[\mathbf{P}(A \mid \mathbf{X})\right] = \mathbf{P}(A) \ . \tag{2.12}$$

Intuitively speaking, the conditional probability of an event A defines a functional relationship between the values of the random variables $\mathbf{X}$

and the probability of A – which is why one speaks of the probability of A *given* a set of values of $\mathbf{X}$ – with the logical constraint that given ranges of possible values for $\mathbf{X}$, the expected value, i.e., weighted average, of $\mathbf{P}(\mathrm{A} \mid \mathbf{X})$ restricted to those ranges, i.e., $\mathbf{P}(\mathrm{A} \mid \mathbf{X}) \cdot \mathbf{I}_{\mathrm{B}}$, should be the same as the general probability of A restricted to outcomes that are consistent with the given ranges, i.e., $\mathbf{P}(\mathrm{A} \cap \mathrm{B})$.

One can also speak of the *conditional expectation* of a random variable Y given a random vector $\mathbf{X}$, written $\mathbf{E}(\mathrm{Y} \mid \mathbf{X})$, which is defined formally in a similar fashion to conditional probability: a $\mathcal{G}$-measurable random variable such that for any $\mathrm{B} \in \mathcal{G}$,

$$\mathbf{E}\left[\mathbf{E}(\mathrm{Y} \mid \mathbf{X}) \cdot \mathbf{I}_{\mathrm{B}}\right] = \mathbf{E}\left(\mathrm{Y} \cdot \mathbf{I}_{\mathrm{B}}\right) \ . \tag{2.13}$$

The intuitive interpretation is likewise analogous: the conditional expected value of a random variable Y defines a functional relationship between values of the random variables $\mathbf{X}$ and the random variable Y with the logical constraint that given ranges of possible values for $\mathbf{X}$, the weighted average of $\mathbf{E}(\mathrm{Y} \mid \mathbf{X})$ restricted to those ranges, i.e., $\mathbf{E}(\mathrm{Y} \mid \mathbf{X}) \cdot \mathbf{I}_{\mathrm{B}}$, should be the same as the general expectation of Y restricted to values that are consistent with the given ranges, i.e., $\mathbf{E}(\mathrm{Y} \cdot \mathbf{I}_{\mathrm{B}})$.

Using these notions, one can define the *conditional density function* of a continuous random variable Y given any other continuous random variable X to be a function $f_{\mathrm{Y}|\mathrm{X}} : \mathbf{R}^2 \to \mathbf{R}^+$ such that

$$\int_a^b \int_c^d f_{\mathrm{Y}|\mathrm{X}}(y \mid x) f_{\mathrm{X}}(x) \, \mathrm{d}y \, \mathrm{d}x = \mathbf{P}\left\{\mathrm{X}^{-1}\left[(a, b)\right] \cap \mathrm{Y}^{-1}\left[(c, d)\right]\right\} \ ; \tag{2.14}$$

the analogous replacements of integrals with direct probabilities yield conditional mass functions for discrete random variables Y or each type of

function when conditioning on a discrete X. As follows intuitively from that definition, the *joint density function* $f_{X,Y}$ of two random variables X and Y, is the product of a conditional density function with an ordinary density function:

$$f_{X,Y}(x,y) \triangleq f_{Y|X}(y \mid x) f_X(x) = f_{X|Y}(x \mid y) f_Y(y) \,. \tag{2.15}$$

This equation is a generalisation of what is commonly known as *Bayes' theorem.* If both X and Y are Riemann-integrable, it also follows from the definitions in this section that

$$\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x = f_Y(y) \quad \text{and} \quad \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y = f_X(x)\,, \tag{2.16}$$

which is sometimes known as *marginalisation* over the domain of integration (and is a generalisation of what is commonly known as the *law of total probability*).

Independence and Conditional Independence

When working with many random variables, it may not make sense to assume that every variable should be in a conditional relationship with the set of all other variables. Independence formalises the notion that random variables may not affect each other. As with conditional probability, one begins by defining independence over events $A \in \mathcal{F}$. Two events A and B, each members of $\mathcal{F}$, are independent if

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)\,; \tag{2.17}$$

it follows that for any finite family of events $\{A_1, A_2, \ldots, A_n\}$, all in $\mathcal{F}$, the events are mutually independent if

$$\mathbf{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbf{P}(A_i) \ . \tag{2.18}$$

Extending this definition to random variables, a set of random variables $\{X_1, X_2, \ldots, X_n\}$ are mutually independent if for any choice of events A_i such that $A_i \subset \sigma(X_i^{-1})$ for all i in $\{1, 2, \ldots, n\}$, the events $\{A_1, A_2, \ldots, A_n\}$ are mutually independent. More intuitively, knowing the value of one random variable in a set of independent random variables provides no information about the values of any other random variables in that set.

Both notions of independence extend to the conditional case in the natural way. Two events A and B, each members of $\mathcal{F}$, are conditionally independent given a set of random variables $\{X_1, X_2, \ldots, X_n\}$, denoted $\mathbf{X}$, if

$$\mathbf{P}(A \cap B \mid \mathbf{X}) = \mathbf{P}(A \mid \mathbf{X})\mathbf{P}(B \mid \mathbf{X}) \ . \tag{2.19}$$

A finite family of events $\{A_1, A_2, \ldots, A_n\}$, all in $\mathcal{F}$, of events are mutually independent given $\mathbf{X}$ if

$$\mathbf{P}\left(\bigcap_{i=1}^{n} A_i \mid \mathbf{X}\right) = \prod_{i=1}^{n} \mathbf{P}(A_i \mid \mathbf{X}) \ . \tag{2.20}$$

A set of random variables $\{Y_1, Y_2, \ldots, Y_m\}$ are mutually independent given $\mathbf{X}$ if for any choice of events A_i such that $A_i \subset \sigma(Y_i^{-1})$ for all i in $\{1, 2, \ldots, m\}$, the events $\{A_1, A_2, \ldots, A_n\}$ are conditionally independent given $\mathbf{X}$.

Bayesian Networks

Bayesian networks, also known as *directed graphical models*, are one way of representing the conditional dependencies among a group of random variables. They are one of two major sub-families of so-called graphical models. Graphical models get their name because they are based on graphs in the mathematical sense, or more formally, triples (V, D, U), where V is a finite set of elements known as the *vertices*; D is a finite set of ordered pairs of vertices, each of which is known as a *directed edge*; and U is a finite set of unordered pairs of vertices, each of which is known as an *undirected edge*. If D is empty but U is not, then the graph is known as an *undirected graph*. If U is empty but D is not, then the graph is known as a *directed graph*. If both U and D are non-empty, then the graph is known as a *partially-directed graph*.

If there is a directed edge $v \to w$ in D, one may say that v is a *parent* of w and that w is a *child* of v. If there is an undirected edge $v - w$ in U, one may say that v is a *neighbour* of w (or vice-versa). Frequently, one needs to reference the set of all parents of a vertex v; this set is denoted $\pi(v)$. In undirected graphs, one may also speak of *cliques*, which are sets of vertices such that every vertex in the clique has an edge connecting it to each other vertex in the clique.

Paths in mathematical graphs are sequences of edges $(v_i, w_i) \in D \cup U$ for $i \in \{1, 2, \ldots, n\}$ such that the internal endpoints join, i.e., $v_i = w_{i-1}$ for all $i > 1$. Paths are called *directed* if at least one of the component edges is directed i.e., $\exists j : v_j \to w_j \in D$. A directed path is called a *cycle* if $v_1 = w_n$. For any directed path, one may say that v_1 is an *ancestor* of w_n or that w_n is a

descendant of v_1. The set of all descendants of a variable v is denoted $\rho(v)$.

Bayesian networks are directed acyclic graphs where V is a set of random variables. For any vertex X of a Bayesian network, the graph implies that X is conditionally independent of its non-descendants $\{V - [\{X\} \cup \rho(X)]\}$ given its parents $\pi(X)$; note that $\pi(X)$ may be $\varnothing$. By equation (2.15), given a Bayesian network where $V = \{X_1, X_2, \ldots, X_n\} = \mathbf{X}$, the joint distribution function f_X factorises as

$$f_X = \prod_{i=1}^{n} f_{X_i|\pi(X_i)} \,. \tag{2.21}$$

These factorisations are very important in practise, both for interpreting the meaning of groups of random variables and for efficient computation during statistical inference.

As an example, consider the network in figure 2.1. It considers five random variables, each of which may be true or false: whether a piece was composed in the nineteenth century, whether a piece was written for a string quartet, whether a piece contains Forte's α-motive, and whether a piece contains Forte's σ-motive. Given only these five factors to consider, this network posits the following factorisation:

- the relative frequency of pieces containing Forte's α-motive depends only on whether a piece was composed by Brahms and whether it was written for string quartet (which implies that Brahms did treat the α-motive differently than other composers of the nineteenth century);
- the relative frequency of pieces containing the σ-motive likewise depends only on whether a piece was composed in the nineteenth

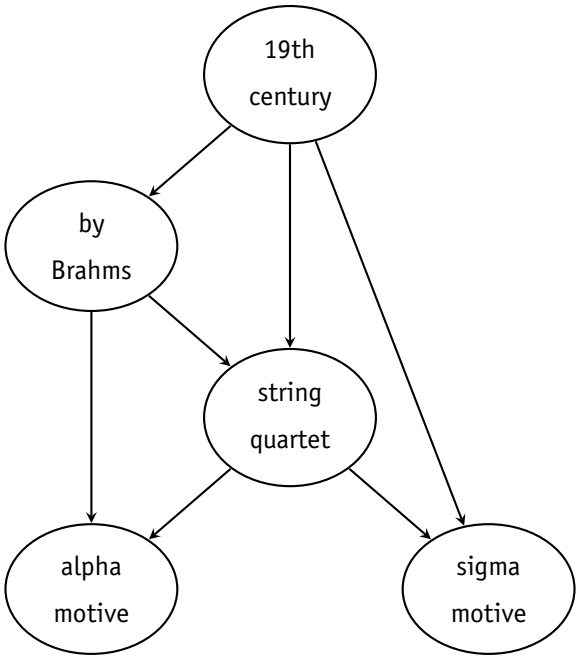


Figure 2.1 · A simple Bayesian network working on the themes of the sample analysis of Alan Forte. This network posits the presence and absence of specific relationships among the five variables.

century and whether a piece is a string quartet (in particular, knowing whether the piece is by Brahms is irrelevant);

› the relative frequency of string quartets depends both on whether a piece was composed in the nineteenth century and whether a piece was composed by Brahms, i.e., Brahms wrote significantly more or fewer string quartets proportional to his other works than the average composer from the nineteenth century; and

› the relative frequency of pieces by Brahms depends only on whether the piece was composed in the nineteenth century.

One could certainly question whether this particular configuration of dependencies is correct, and indeed, as later sections of this chapter will discuss, the choice of an appropriate network can be as important or even more important a question as the relative frequencies that such a network can help to uncover.

✳ *Causality*

The arrows in Bayesian networks are often thought to represent causal relationships, which can indeed be a very useful interpretation (Pearl 2009, pp. 21–26) – but *caveat emptor*! Dependence and conditional dependence are often known as *correlation*, and most scholars who have any experience with empirical research will have heard the maxim that 'correlation is not causation'. Bayesian networks are a good example: for most Bayesian networks, there are other networks that represent the same conditional dependencies but have arrows in different directions. In particular, any two graphs (V_1, D_1, U_1) and (V_2, D_2, U_2) represent identical sets of conditional dependencies if they share the same *skeleton*, i.e., the vertices and edges would be identical were all directed edges considered to be undirected edges, or

$$V_1 = V_2 \quad \text{and} \quad D_1 \cup U_1 = D_2 \cup U_2 \,, \tag{2.22}$$

and they share the same *v-structures*, i.e.,

$$\{(x, y, z) : x \to y \in D_1 \wedge z \to y \in D_1\} = \{(x, y, z) : x \to y \in D_2 \wedge z \to y \in D_2\} \,. \tag{2.23}$$

figure 2.2, for example, illustrates a network that represents the same conditional dependencies as the network in figure 2.1, because although the direction of the relationship between being composed during the nineteenth century and being composed by Brahms has been reversed, neither of these vertices were colliders in the original network. In fact, the situation is graver still: for two graphs to represent the same conditional dependencies, it is sufficient only for the *immoralities* of two graphs to be the same (Koller & Friedman 2009, pp. 68–78), i.e.,

$$\begin{aligned}\{(x,y,z) : x \to y \in \mathrm{D}_1 \wedge z \to y \in \mathrm{D}_1 \wedge x - y \notin \mathrm{D}_1 \cup \mathrm{U}_1)\} \\ = \{(x,y,z) : x \to y \in \mathrm{D}_2 \wedge z \to y \in \mathrm{D}_2 \wedge x - y \notin \mathrm{D}_2 \cup \mathrm{U}_2)\} \,. \quad (2.24)\end{aligned}$$

More intuitively, an immorality is a set of three nodes x, y, and z such that x and z are each parents of y but there is no edge connecting x and z. Neither of the graphs in figures 2.1 and 2.2 contain any immoralities, and so one could reverse the direction of any of the arrows without changing the conditional dependencies represented; figure 2.3 illustrates yet another Bayesian network that represents the same set of conditional dependencies. Correlation is not causation indeed!

If causation is not correlation, however, what is it? Pearl's landmark work presents one of the prevailing views (Pearl 2009), which he has proved to be mathematically equivalent to the other prevailing view, the Neyman-Rubin theory of potential outcomes (Spława-Neyman 1923; Rubin 2005; Wasserman 2004, pp. 251–61). Pearl's theory understands causality to be the effect of *interventions*, which are actions that change the underlying measure $\mathbf{P}$ of the probability space, or more formally, are functions from probability measures to other probability measures on the same space. More

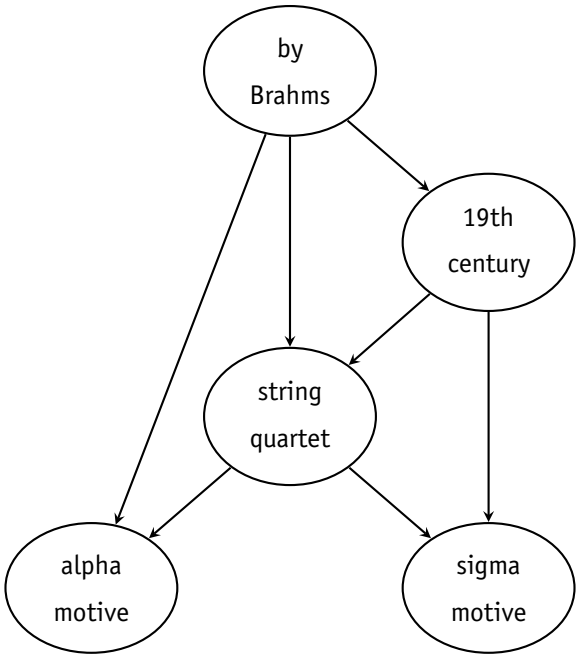


Figure 2.2 · A Bayesian network that represents an identical set of conditional dependencies to that of the network figure 2.1. The direction of the relationship between music by Brahms and music composed during the nineteenth century is undetermined because neither vertex is a collider.

specifically, interventions fix certain quantities and hold them constant, and the theory proposes a mathematical model, the *do-calculus*, for how exactly such interventions should alter probability distributions as represented with Bayesian networks.

In many cases, reasoning about interventions also involves reasoning about situations that are contrary to observed facts, e.g., the probability that Forte's α-motive would have appeared at a particular moment in a particular string quartet of Brahms if one had been able to intervene to force some other composer to write that moment instead of Brahms. Such questions

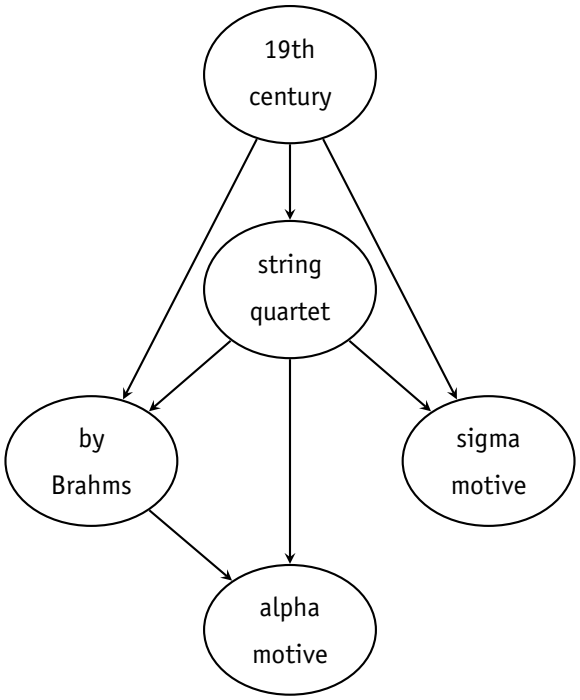


Figure 2.3 · Another Bayesian network the represents the same set of conditional dependencies as that of the network in figure 2.1. This network posits (implausibly) that the ensemble has a causal effect on the composer of a piece. Because all three networks represent the same conditional dependencies, one must choose among them based on common sense rather than empirical methods.

are known as *counterfactuals*, and they are of particular concern when dealing with retrospective data – including almost all of the data of concern in database-driven musicology – because it is impossible to intervene in the past. Even reasoning about prospective data, e.g., that of potential experiments, can involve counterfactuals because it is sometimes impossible or unethical to intervene in every way one might like. Contrary to what it may seem, counterfactuals do not entail any particular choice of *explicandum*

or *explicatum* of probability: Pearl himself is a subjectivist, but there is no necessary contradiction in considering what physical probabilities might be or what the result of frequentist experiments might be in the hypothetical worlds resulting from counterfactual interventions. Pearl is quite careful, in fact, to insist on a distinction between *statistical* concepts, which he defines to be those that can be derived from or apply to observed data only – a category that necessarily excludes counterfactuals – and *causal* concepts, which he defines to be all other constraints on and properties of causal models (2009, pp. 38–40).

In principle, there is no constraint on counterfactual probability spaces: because the events in question did not happen, one could assign them any consistent set of probabilities that one desired. The value in Pearl's *do*-calculus is that it proposes a carefully argued set of practical constraints on counterfactual distributions that make it possible to derive certain causal quantities from observed data alone. In Pearl's reckoning, such quantities are still considered causal, not statistical, because they cannot be computed without making assumptions about unobserved interventions.

The most useful aspect of the *do*-calculus for this thesis is the *back-door criterion* (Pearl 1993). Given a Bayesian network, consider some intervention that maps a probability measure $\mathbf{P}$ on $(\Omega, \S)$ to a new, possibly counterfactual, probability measure $\mathbf{P}^*$ on the same space such that the value of some random variable $X : (\Omega, \S) \to (\Psi, \mathfrak{E})$ is fixed to be x^*: i.e., $\mathbf{P}^*[X^{-1}(x^*)]$ is unity, or $X(\omega) = x^*$ *almost surely*. Suppose that one were interested in the density or mass function of some other random variable Y after this intervention. Denote the probability density functions pre-intervention as f and post-intervention as f^*. The back-door criterion states that if

one can find a set of random variables $\mathbf{Z}$ in the Bayesian network such that (1) no variable in $\mathbf{Z}$ is a descendant of X and (2) every path between X and Y in the skeleton of the network that containes an edge corresponding to a directed edge leading into X either (a) goes through at least one variable in $\mathbf{Z}$ that is not the 'collider' of a *v*-structure or (b) contains a *v*-structure for which neither the collider nor any of its descendants are in $\mathbf{Z}$, then it follows that

$$f_Y^*(y) = \int_{\mathbf{Z}} f_{Y|X,\mathbf{Z}}(y \mid x^*, \mathbf{z}) f_{\mathbf{Z}}(\mathbf{z}) \, d\mathbf{z} \,, \tag{2.25}$$

where the integral $\int_{\mathbf{Z}} d\mathbf{z}$ refers to the appropriate combination of multiple integrals and discrete sums over the ranges of all variables in $\mathbf{Z}$. A related criterion, the so-called front-door criterion, allows the computation of the effect of interventions from observed data in somewhat more complex Bayesian networks, and the complete *do*-calculus provides a number of more general simplification rules that can be used in arbitrary networks (Pearl 1995).

Simpson's paradox illustrates how essential causation can be for understanding correlation. Suppose, for example, that one were interested in comparing the relative frequency of some hypothetical motive in the string quartets and piano quintets of Brahms and his contemporary Camille Saint-Saëns; in particular suppose one wanted to know which composer used the motive more often. The outcome space Ω would be the set of all musical moments where the motive could theoretically have appeared in the quartets and quintets of these two composers; the set of random variables under consideration might be $X : \Omega \to \{\text{Brahms}, \text{Saint-Saëns}\}$ to represent the composer of a moment, $Y : \Omega \to \{\text{true}, \text{false}\}$ to represent whether the

Table 2.1 · A hypothetical example of Simpson's paradox

COMPOSER	QUARTETS	QUINTETS	OVERALL
Brahms	$\frac{1\,008}{14\,933} \approx 6.8\,\%$	$\frac{1}{7\,619} \approx 0.01\,\%$	$4.5\,\%$
Saint-Saëns	$\frac{713}{9\,832} \approx 7.3\,\%$	$\frac{38}{7\,323} \approx 0.52\,\%$	$4.4\,\%$
Both composers	$6.9\,\%$	$0.26\,\%$	$4.4\,\%$

Note: The numerators represent the total number of instances of a hypothetical motive in the string quartets and piano quintets of Brahms and Saint-Saëns; the denominators represent the total (hypothetical) number of places such a motive could have occurred. Looking at the quartets and the quintets individually, it would seem that Saint-Saëns used the motive more frequently, but overall, the opposite is true. The correct interpretation depends on which variables one believes to be the causes and which one believes to be the effects.

motive appears at a moment, and $Z : \Omega \to \{\text{quartet}, \text{quintet}\}$ to represent whether the moment arises in a quartet or a quintet. Table 2.1 presents hypothetical counts of the number of moments where the motive appears ($Y = \text{true}$) relative to the number of moments where it could possibly have occurred ($Y \in \{\text{true}, \text{false}\}$) for each composer, both separated by string quartet or piano quintet ($f_{Y|X,Z}$) and overall ($f_{Y|X}$). Looking at either the quartets or the quintets independently, it would seem that Saint-Saëns used the motive more frequently by a substantial margin:

$$f_{Y|X,Z}(\text{true} \mid \text{Saint-Saëns}, \text{quartet}) > f_{Y|X,Z}(\text{true} \mid \text{Brahms}, \text{quartet}) \quad (2.26)$$

and

$$f_{Y|X,Z}(\text{true} \mid \text{Saint-Saëns}, \text{quintet}) > f_{Y|X,Z}(\text{true} \mid \text{Brahms}, \text{quintet}) \,. \quad (2.27)$$

Overall, however, it is Brahms who used it more frequently:

$$f_{Y|X}(\text{true} \mid \text{Saint-Saëns}) < f_{Y|X}(\text{true} \mid \text{Brahms}) \,. \tag{2.28}$$

Simpson's paradox is the apparent paradox in situations like these, whereby the direction of an effect seems to be reversed after grouping according to an additional variable.

It is difficult to find a complete treatment of Simpson's paradox in the literature, although partial (and partially erroneous) explanations abound. Pearl has made perhaps the most thorough and accurate explanation with respect to the prevailing understanding of causality (2009, pp. 174–82), but his explication is entwined with his strictly Bayesian view of probability. Wasserman's explanation (2004, pp. 259–61), one of few of which Pearl approves, uses the potential-outcome framework and is thus more agnostic about the interpretation of probability, but it ends with a critical mistake whereby he presents a result as general that in fact assumes a particular causal relationship among variables. His mistake relates to the notion of *exchangeability*, which is an older theory for explaining Simpson's paradox (see de Finetti 1972, pp. 229–46).

The apparent paradox in Simpson's paradox (and Wasserman's mistake) arise from a crucial misinterpretation: equation (2.28) does not necessarily imply that Brahms *caused* the motive to be used more often, and the sub-counts by quartet and quintet are nearly irrelevant. Table 2.1 gives information about an underlying probability distribution $\mathbf{P}$, but a causal interpretation requires information about a counterfactual distribution $\mathbf{P}^*$ resulting from an intervention forcing the composer to be Brahms (or Saint-Saëns). Consider what those probabilities would be under the causal models

implied by the three networks presented so far, all of which can be calculated from observed data by way of the back-door criterion. In addition to the three random variables mentioned earlier, add a fourth random variable $W : \Omega \to \{19\text{th-century}, \neg 19\text{th-century}\}$. For the network of figure 2.1, $\{W\}$ is a back-door set: W is not a descendant of X, the variable of intervention, and because W is the only vertex in the network with an arrow pointing into X, W is necessarily on every path between X and Y that has an arrow pointing into X. By the back-door criterion, then,

$$\begin{aligned} f_Y^{\text{Brahms}}(y) = {} & f_{Y|X,W}(y \mid \text{Brahms}, 19\text{th-century}) f_W(19\text{th-century}) \\ & + f_{Y|X,W}(y \mid \text{Brahms}, \neg 19\text{th-century}) f_W(\neg 19\text{th-century}) \,. \end{aligned} \tag{2.29}$$

Because of the limitations on Ω, however, $f_W(19\text{th-century})$ is unity and $f_W(\neg 19\text{th-century})$ is zero, and so equation (2.29) reduces to

$$f_Y^{\text{Brahms}}(y) = f_{Y|X}(y \mid \text{Brahms}) \,, \tag{2.30}$$

from which it follows directly that $f_Y^{\text{Brahms}}(a) \approx 4.5$ percent and analogously that $f_Y^{\text{Saint-Saëns}}(a) \approx 4.4$ percent. For the network of figure 2.2, there are no arrows pointing into X, and so the empty set is a back-door set; that yields $f_Y^{\text{Brahms}}(y) = f_{Y|X}(y \mid \text{Brahms})$ directly. For the network in figure 2.3, however, the extra arrow into X requires a larger back-door set of $\{W, Z\}$, which means that, after reducing to account for the fact that all outcomes in Ω were composed during the nineteenth century,

$$\begin{aligned} f_Y^{\text{Brahms}}(y) = {} & f_{Y|X,Z}(y \mid \text{Brahms}, \text{quartet}) f_Z(\text{quartet}) \\ & + f_{Y|X,Z}(y \mid \text{Brahms}, \text{quintet}) f_Z(\text{quintet}) \,. \end{aligned} \tag{2.31}$$

The counts in table 2.1 give $f_Z(\text{quartet}) \approx 62$ percent and $f_Z(\text{quintet}) \approx 38$ percent, and so equation (2.31) simplifies to $f_Y^{\text{Brahms}}(a) \approx 4.2$ percent whereas the analogous computation for Saint-Saëns yields $f_Y^{\text{Saint-Saëns}}(a) \approx 4.7$ percent. Thus, $f_Y^{\text{Saint-Saëns}}(a) < f_Y^{\text{Brahms}}(a)$ under the causal assumptions of the networks in figures 2.1 and 2.2, but $f_Y^{\text{Saint-Saëns}}(a) > f_Y^{\text{Brahms}}(a)$ under the assumptions of figure 2.3. There is no paradox here – it should not be surprising that different causal assumptions would yield different probabilities – but it is sobering that such assumptions can change the direction of the effect. Wasserman's mistake is assuming that only the latter set of causal assumptions are possible (or equivalently, that different composers are exchangeable given the ensemble), which leads him to claim wrongly that for any two interventions x_1 and x_2 and variables Y and Z, if $f_{Y|Z}^{x_1}(y \mid z) > f_{Y|Z}^{x_2}(y \mid z)$ for some y and all z, then $f_Y^{x_1}(y) > f_Y^{x_2}(y)$.

More sobering still is the fact that although some causal relationships can be uncovered from observed data (see the following section), under the prevailing understandings of causality, it is generally impossible to recover all causal relationships from observed data. The choice of an appropriate Bayesian network must to some extent be guided by common sense and expert knowledge. For example, among the three networks considered in this section, only that of figure 2.1 is plausible: composers cannot will themselves into other centuries, as figure 2.2 implies; on the other hand, composers do choose to write for particular ensembles, not vice-versa, as figure 2.3 implies. In short, even in apparently quantitative, database-driven research, there is no substitute for qualitative knowledge about the problem domain, and conversely, researchers with strong qualitative knowledge about a problem domain are much better positioned to produce

trustworthy quantitative results.

Markov Networks

Sometimes, causality is irrelevant to the research question, and it can be simpler in such cases to work with models that explicitly ignore it. *Markov networks*, also known as *undirected graphical models* are a variant type of graphical model that can encode non-causal correlations among random variables. As the alternate name implies, Markov networks are based on undirected graphs. Any two vertices in Markov network may be connected with an edge (undirected graphs are by definition acyclic). For any two distinct vertices X and Y of a Markov network, and a set of other vertices $\mathbf{Z}$ containing neither X nor Y, if every path in the network between X and Y contains some node $Z \in \mathbf{Z}$, then the network implies that X and Y are conditionally independent given $\mathbf{Z}$, and thus $f_{X,Y|\mathbf{Z}} = f_{X|\mathbf{Z}} f_{Y|\mathbf{Z}}$. Because of the lack of directionality in a Markov network, there is no direct conversion from these independencies to a complete factorisation of the joint distribution function $f_{\mathbf{X}}$ of a Markov network where $V = \{X_1, X_2, \ldots, X_n\} = \mathbf{X}$ into independent conditional distributions, but Markov networks do correspond to a more general family of probability distributions known as *Gibbs distributions.* The joint density function of a set of random variables $\mathbf{X}$ that follow a Gibbs distribution can be represented

$$f_{\mathbf{X}} = \frac{\prod_{i=1}^{n} \phi_i}{\int_{\mathbf{X}} \prod_{i=1}^{n} \phi_i(x_i)\, \mathrm{d}\mathbf{x}}\,, \tag{2.32}$$

where each of the functions ϕ_i, known as *factors*, maps the values of the random variables in some clique of the network to a non-negative number.

Markov networks and Bayesian networks cannot in general represent the same sets of dependencies. The sole exception is for Bayesian networks that contain no immoralities, for which the equivalent Markov network has the same edges as the Bayesian network, only made undirected. Intuitively, this equivalence makes sense knowing that the direction of any arrow in a Bayesian network may be changed without affecting the underlying probability distribution unless that arrow is part of an immorality. When immoralities are present, the closest Markov-network approximation to the dependencies represented in a Bayesian network is the *moralised graph* of the Bayesian network, which adds an edge between any two nodes that have a common child (Koller & Friedman 2009, pp. 134–39). Figure 2.4 represents a Markov network that correspondes to an equivalent family of probability measures to those of the Bayesian networks in figures 2.1, 2.2, and 2.3. Because none of the Bayesian networks contain any immoralities, this Markov represents the same set of dependencies, but traditionally, it would represented as a Gibbs distribution over the three cliques,

$$f_{\mathrm{V,W,X,Y,Z}}(v,w,z,y,z) = \frac{\psi_1(w,x,z)\psi_2(x,y,z)\psi_3(v,w,z)}{\sum_{\mathrm{V,W,X,Y,Z}} \psi_1(w,x,z)\psi_2(x,y,z)\psi_3(v,w,z)}\,, \tag{2.33}$$

with the random variables taking the same labels as they would in the example of Simpson's paradox and an additional variable V added to represent whether the motive at a particular moment is Forte's σ-motive.

2.3 CLASSIFICATION AND STOCHASTIC PROCESSES

One of the defining features of music is that it unfolds over time, and so often, the random variables of interest in musical applications correspond

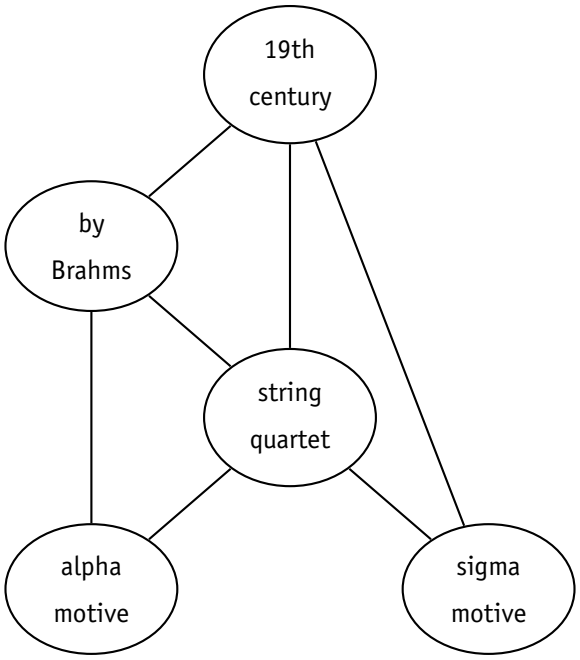


Figure 2.4 · A Markov network the is equivalent to the Bayesian networks of figures 2.1, 2.2, and 2.3. Because none these Bayesian networks contain any immoralities, the equivalence is exact.

to sequential points in time. This situation is a specific case of a *random field*, which is a random variable the domain of which is a probability space $(\Omega, \S, \mathbf{P})$ and the range of which is the space of all functions from a reference set T, endowed with a topology $\mathfrak{T}$, to some other topological space Ψ (Adler & Taylor 2007). When T represents time, a random field is usually known as a *stochastic process*, and it can be interpreted as a set of random variables $\mathrm{X}^{(t)} : (\Omega, \S) \to (\Psi, \mathfrak{E})$ for all $t \in \mathrm{T}$. This section describes stochastic processes as they can be used for database-driven musicology: it starts with general definitions, connects those definitions to classification problems as researchers tend to encounter them when constructing musico-

logical databases, and explains the broad strategies for learning them from data.

Stochastic Processes

This thesis exclusively considers *discrete-time* stochastic processes, where the reference set corresponds to $\mathbf{N}$, but there are also *continuous-time* stochastic processes, where the reference set would correspond to $\mathbf{R}^+$. In principle, all of the $X^{(t)}$ could be independent, but stochastic processes become interesting when they represent dependencies across different points in time. These types of dependencies are particularly significant for music: given that Forte's α-motive has appeared once in a piece, for example, does the probability that it will appear at some point later in the piece change, and is that change unique to music composed by Brahms?

In most cases, one assumes that the dependencies between different times follow some kind of fixed rule. In other words, we are interested in stochastic processes that represent the solution to a *dynamical system*. Dynamical systems are an enormous field of study in their own right – Anatole Katok and Boris Hasselblatt have written a standard but weighty reference (1995) – but the dynamical systems in this thesis are all quite simple. This thesis will consider stochastic process where each time t corresponds to a window of musical time, e.g., $t = 0$ corresponds to the first beat, $t = 1$ to the second beat, $t = 2$ to the third beat, etc., or $t = 0$ corresponds to the first 50 ms, $t = 1$ to the next 50 ms, etc. The random variables $X^{(t)}$ peek at that window of time for any outcome $\omega \in \Omega$ – which could be any piece or fragment of music together with some representation

for it – and yield the quantities under consideration for that window of time.

Like any other group of random variables, the question of independence is fundamental to stochastic processes. Strictly for reasons of computational tractability, all of the models in this thesis make some kind of *Markov assumption* that assumes that the random variables at any time point t are independent of all random variables from earlier times given the random variables between times $t - k$ and $t - 1$ for some small value of k. Using the notation $\mathrm{X}^{m:n}$ to represent the set $\{\mathrm{X}^{(m)}, \mathrm{X}^{(m+1)}, \ldots, \mathrm{X}^{(n)}\}$, these assumptions imply that

$$f_{\mathrm{X}^{(n)}|\mathrm{X}^{1:(n-1)}} = f_{\mathrm{X}^{(n)}|\mathrm{X}^{(n-k):(n-1)}} \,. \tag{2.34}$$

Discrete-time stochastic process with Markov assumptions are sometimes known as *Markov chains* Although there has been some practical success with Markov chains, Markov assumptions are in fact rather unmusical because they make it difficult to model long-range dependencies. There are tricks to work around this limitation (e.g., allowing the range of the $\mathrm{X}^{(t)}$ to include explicit information about $\mathrm{X}^{1:(t-1)}$), but there is nonetheless considerable musical interest in finding ways to avoid Markov assumptions.

⚙ *State-Space Models*

In the context of database-driven musicology, the random variable at each time point t is usually a compound random variable that encompasses a range of more specific random variables. For example, research involving harmony, researchers may want to be able to examine random variables

corresponding to the key, root, chord quality, and inversion at any point in time. As discussed in section 1.2, generating large-scale databases of music usually entails deriving such high-level labels from lower-level, physical information, such as statistics about an audio file at different points in time, which themselves can be considered as random variables. Because the lower-level random variables, often grouped together under the name *observations* and represented as vector-valued random variables $\mathbf{X}^{(t)}$, are often difficult to understand, a common simplifying assumption is to ignore their temporal dynamics and assume that they depend solely on the high-level random variables from the same window of time, which are often grouped together under the name *states* and represented as vector-valued random variables $\mathbf{Y}^{(t)}$. In other words,

$$f_{\mathbf{Y}^{(t)}|\mathbf{X}^{1:\infty},\mathbf{Y}^{1:(t-1)},\mathbf{Y}^{(t+1):\infty}} = f_{\mathbf{Y}^{(t)}|\mathbf{X}^{(t)}} \,. \tag{2.35}$$

Such a model is known as a *state-space model* or *state-observation model*. Many popular applications of state-space models involve states that are continuous random variables (see Shumway & Stoffer 2011, chap. 6), most famously the Kálmán filter (Kálmán 1960); musicological information, however, is almost always discrete, and so this thesis will consider only discrete-state state-space models.

Traditionally state-space models are used for three categories of tasks, all given the observations up to some time point t^*: *filtering*, which involves queries about $\mathbf{Y}^{(t^*)}$; *forecasting* or *prediction*, which involves queries about $\mathbf{Y}^{(t)}$ for some $t > t^*$; and *smoothing*, which involves queries about $\mathbf{Y}^{(t)}$ for some $t < t^*$. Prediction can be important for real-time applications, but when constructing databases, filtering and smoothing are usually the

only tasks of interest because one can always start from a complete set of observations. Specifically, one wants to learn enough about the density function $f_{\mathbf{Y}^{1:\infty}|\mathbf{X}^{1:\infty}}$ that one can predict an optimal sequence of high-level states given the low-level observations about any specific item of a large collection.

Classifiers and Consistency

In other words, for database-driven musicology, state-space models are important to the extent that they can be used as *classifiers.* Formally, a classifier is a measurable function $g : (\Phi, \mathfrak{G}) \to (\Psi, \mathfrak{E})$ from the range of a set of random variables $\mathbf{X} : (\Omega, \mathfrak{F}) \to (\Phi, \mathfrak{G})$ to the range of another set of random variables $\mathbf{Y} : (\Omega, \mathfrak{F}) \to (\Psi, \mathfrak{E})$ on a common probability space $(\Omega, \mathfrak{F}, \mathbf{P})$. The intuition behind this function is that it is predicting higher-level labels $\mathbf{y}$ given a number of lower-level observations $\mathbf{x}$. Classifiers are evaluated with respect to some other measurable *loss function* $L : (\Psi \times \Psi, \mathfrak{E} \times \mathfrak{E}) \to (\mathbf{R}, \mathfrak{B}(\mathbf{R}))$, where the σ-algebra $\mathfrak{E} \times \mathfrak{E}$ denotes the smallest σ-algebra over $\Psi \times \Psi$ that contains all *measurable rectangles*, i.e., $\{E_1 \times E_2 : E_1 \in \mathfrak{E} \wedge E_2 \in \mathfrak{E}\}$ where $A \times B$ is the Cartesian product of sets A and B, or $\{(a, b) : a \in A \wedge b \in B\}$. The loss function should reflect the degree of dissimilarity between the correct value and a prediction from a classifier.

Given the uncertain nature of classification, one normally defines a distinct random variable $R_g : (\Omega, \mathfrak{F}) \to (\mathbf{R}, \mathfrak{B}(\mathbf{R}))$ such that

$$R_g(\omega) \triangleq L\left(Y(\omega), g\left(X(\omega)\right)\right) . \tag{2.36}$$

This variable is sometimes known as *risk*, and in general, one wants a classifier with the smallest possible expected value of risk, $\mathbf{E}(R_g)$. With perfect knowledge of the conditional distribution function $f_{Y|X}$, it is possible to construct an optimal classifier g^* such that for any other classifier g, the expected risk $\mathbf{E}(R_{g^*}) \leq \mathbf{E}(R_g)$ (Devroye, Györfi & Lugosi 1996, p. 569). The expected risk of this optimal classifier is often known as the *Bayes risk*. Of course, one is normally interested in finding a classifier when it is impossible to have perfect knowledge of $f_{Y|X}$, and so it is necessary to develop some strategy for devising a classifier by other means. In a database-driven context, the normal strategy is to learn the best classifier possible from the data available, and with such a strategy, one normally wants a *consistent* rule for building classifiers (Devroye, Györfi & Lugosi 1996, pp. 2–3).

Informally, consistency of a rule for building classifiers from a database means that with a sufficiently large database, the expected risk of the classifier can be made arbitrarily close to the Bayes risk. Formally, defining consistency requires an understanding of *sequence spaces* and *product measure* over such spaces. Start with any probability triple $(\Omega, \mathcal{F}, \mathbf{P})$. Consider the sequence Ω^∞, which is a space that contains all outcomes of the form $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots)$, each of the ω_i, $i \in \mathbf{N}$, being members of the original outcome space Ω. Let the σ-algebra $\mathcal{F}^\infty$ be the smallest σ-algebra over Ω^∞ containing all measurable rectangles, analogous to the definition of $\mathcal{E} \times \mathcal{E}$ above. One can define a unique *product measure* $\mathbf{P}^\infty : \mathcal{F}^\infty \to [0, 1]$ over $(\Omega^\infty, \mathcal{F}^\infty)$ such that for all measurable rectangles $A_1 \times A_2 \times \cdots$, the product measure $\mathbf{P}^\infty = \prod_{i=1}^\infty \mathbf{P}(A_i)$ (Billingsley 1995, pp. 231–41). This product measure is meant to represent the idea of drawing an infinitely large sample of outcomes from the original outcome space. For any outcome $\boldsymbol{\omega} \in \Omega^\infty$, one can imagine the

first n elements to be a random sample of finite size – e.g., the n elements of a database. Let $\mathrm{R}_{g^{(n)}}$ represent the risk of using a classifier built from some rule on the first n elements of an outcome $\boldsymbol{\omega}$ to make a prediction on the next element ω_{n+1}. Such a rule is said to be consistent if

$$\lim_{n\to\infty} \mathbf{E}\big(\mathrm{R}_{g^{(n)}}\big) = \mathbf{E}\big(\mathrm{R}_{g^*}\big) \,. \tag{2.37}$$

An alternative way of describing this relationship is to say that the sequence of risk variables $\mathrm{R}_{g^{(n)}}$ *converge in probability* to the Bayes risk R_{g^*}. The limiting behaviour of consistent rules for classification means that with a sufficiently large database, any consistent rule is mathematically guaranteed to out-perform all non-consistent rules other than g^* itself.

✵ *Discriminative and Generative Training*

The obvious strategy for building a classifier, then, would seem to be finding a consistent rule for modelling $f_{\mathbf{Y}^{1:\infty}|\mathbf{X}^{1:\infty}}$ directly. This approach is known as *discriminative training*. At its best, it can yield excellent classifiers, but it suffers several limitations. It expressly avoids learning any distribution over the observations $\mathbf{X}^{1:\infty}$ or a joint density function $f_{\mathbf{X}^{1:\infty},\mathbf{Y}^{1:\infty}}$, which means that it is not possible to understand a discriminatively trained model in any kind of causal context or to use it to predict anything other than the explicit task for which it was trained; in particular, a discriminatively trained model for predicting high-level musicological labels will provide no insight into musicological queries about those high-level features. Because the resulting models are useless for causal queries, discriminative training is usually applied only to undirected models. It also tends to require a relatively

large set of training examples, which may be unavailable or expensive to produce.

Generative training is the alternative to discriminative training, and although it remedies many of the shortcomings of discriminative training, it has drawbacks of its own. Generative training starts from the principle that wherever $f_{\mathbf{X}^{1:\infty}} > 0$, it necessarily follows from the laws of total probability that the conditional density function

$$f_{\mathbf{Y}^{1:\infty}|\mathbf{X}^{1:\infty}} = \frac{f_{\mathbf{Y}^{1:\infty},\mathbf{X}^{1:\infty}}}{f_{\mathbf{X}^{1:\infty}}} \,. \tag{2.38}$$

By marginalisation, one can recover $f_{\mathbf{X}^{1:\infty}}$ from $f_{\mathbf{Y}^{1:\infty},\mathbf{X}^{1:\infty}}$, and so as an alternative to learning the conditional density directly, it suffices to learn the joint density $f_{\mathbf{Y}^{1:\infty},\mathbf{X}^{1:\infty}}$. It is also possible to use the joint density function to recover any other density function, conditional or not, over any desired combination of the $\mathbf{X}^{1:\infty}$ and $\mathbf{Y}^{1:\infty}$, and for this reason, generative training is a good choice for Bayesian networks that encode musicological and causal assumptions. Furthermore, because generative training encodes more constraints on the structure of the probability space than discriminatively training does (namely, the distribution of the observations), generative training requires fewer data than discriminative training. If these constraints are incorrect, however, a generatively trained model will be biased away from the true model in a way that a discriminatively trained model on a sufficiently large database would never be.

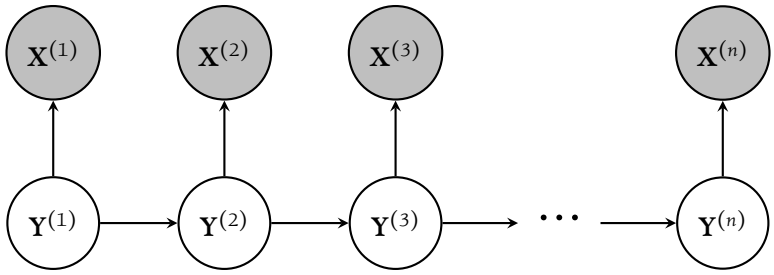


Figure 2.5 · A hidden Markov model (HMM)

2.4 COMMON MODELS FOR CLASSIFICATION ON STOCHASTIC PROCESSES

A wide variety of state-space models have been proposed in the literature. This section highlights some of the most important. Kevin Murphy's doctoral thesis remains the best resource for surveying state-space models that can be represented as Bayesian networks (2002, pp. 18–49), and there is a growing base of literature on 'linear-chain' Markov networks that are useful for applying discriminative training.

✺ *Hidden Markov Models*

The mainstay of discrete-state state-space models has long been the *hidden Markov model* (Baum & Petrie 1966; Baum et al. 1970), represented in figure 2.5; HMMs are most famous for their successes in speech recognition (Rabiner 1989), although as the next chapter will show, they have been applied broadly for musicological tasks as well. They make the standard state-space assumption that at any time t, the observations $\mathbf{X}^{(t)}$ depend only on the state $\mathbf{Y}^{(t)}$ at the same time point and that the state is the cause of the observation. HMMs also make the *first-order Markov assumption* that

at any time t, the state $\mathbf{Y}^{(t)}$ depends only the state at the immediately previous time, $\mathbf{Y}^{(t-1)}$. Being a Bayesian network, HMMs are normally trained generatively, and so one is seeking to learn a joint distribution

$$\widehat{f}_{\mathbf{X}^{1:\infty},\mathbf{Y}^{1:\infty}} = \widehat{f}_{\mathbf{Y}^{(1)}} \prod_{t=1}^{\infty} \widehat{f}_{\mathbf{Y}^{(t+1)}|\mathbf{Y}^{(t)}} \widehat{f}_{\mathbf{X}^{(t)}|\mathbf{Y}^{(t)}} \tag{2.39}$$

that matches the true joint distribution $f_{\mathbf{X}^{1:\infty},\mathbf{Y}^{1:\infty}}$ as closely as possible. One generally also assumes that the $\widehat{f}_{\mathbf{Y}^{(t+1)}|\mathbf{Y}^{(t)}}$ are identical for all t and that likewise the $\widehat{f}_{\mathbf{X}^{(t)}|\mathbf{Y}^{(t)}}$ are identical for all t. Running the stochastic process to infinity is a mathematical convenience to handle the natural variation in the length of musical pieces; the density functions are assumed to be zero at all time points after the end of a musical fragment corresponding to any specific outcome ω.

It may be easier to interpret HMMs in the context of an example. Consider the classification task treated later in this thesis: audio chord recognition. The outcome space Ω for this task might encompass all singles played on mainstream North-American radio stations in the latter half of the twentieth century; $\mathcal{F}$ would be $\mathcal{P}(\Omega)$. The time index t of the stochastic process would correspond to musical beats, the state variable $\mathrm{Y}^{(t)}$ would map any outcome $\omega \in \Omega$ to the (unknown) chord label for its t-th beat of ω, and the observation variables $\mathbf{X}^{(t)}$ would map any outcome $\omega \in \Omega$ to a vector of quantities derived from the audio signal at the t-th beat. Using a hidden Markov model for this process entails a belief that for any beat t, the only relevant knowledge for predicting the collective observations $\mathbf{X}^{(t)}$ is the active chord at beat t, and that likewise, the only relevant knowledge for predicting the chord at beat t is the chord on the immediately preceding beat.

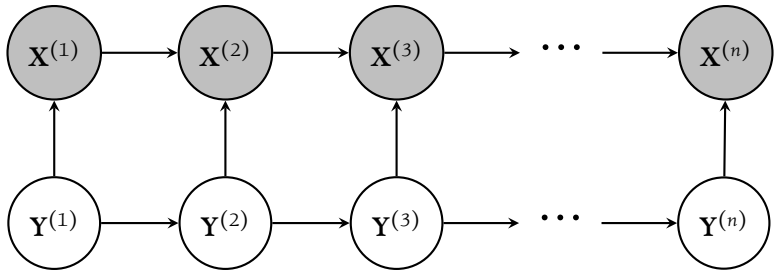


Figure 2.6 · An auto-regressive hidden Markov model

Simple HMMs like this example have been used in practise, but it is clear that they are quite an over-simplification. One common technique for making the model richer is to increase the order of the Markov assumption, e.g., to claim that some natural number k of previous chords are needed to predict the following chord. Another type of system, the *switching linear dynamical system*, allows $\widehat{f}_{\mathbf{X}^{(t)}|\mathbf{Y}^{(t)}}$ and $\widehat{f}_{\mathbf{Y}^{(t)}|\mathbf{Y}^{(t-1)}}$ to change in some defined way over time (Rosti & Gales 2003), which seems to be a better model in theory but has proved disappointing in practise (Layton 2006, p. 14).

Auto-regressive HMMs

The *auto-regressive hidden Markov model* relaxes a different assumption of the standard HMM and indeed of state-space models generally: namely, that the observations depend only on states. Auto-regressive HMMs add a time dependency for each observation to the previous observation (see figure 2.6); this extra dependency induces what would be called an auto-regressive filter among researchers in signal processing, hence the name (Murphy 2002, p. 23). This assumption makes good sense for audio signals, which one expects to evolve more smoothly than the discrete jumps of a state

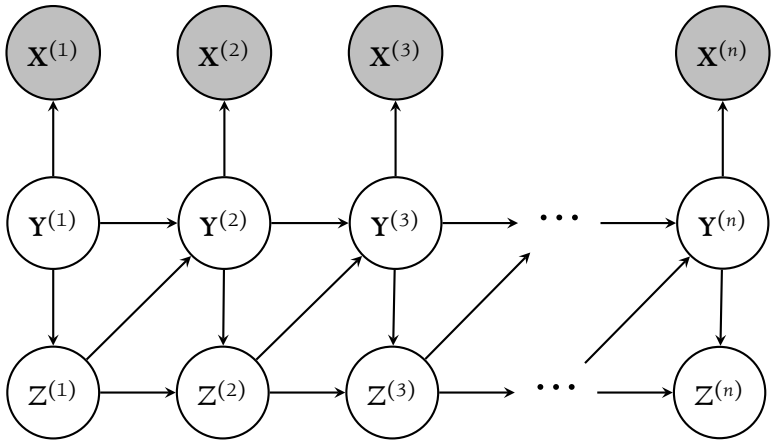


Figure 2.7 · A hidden semi-Markov model

space. *Buried Markov models* are a more elaborate variant of auto-regressive HMMs that allow a wider range of dependencies among observations (Bilmes 2003).

❋ *Hidden Semi-Markov Models*

Another problem with simple HMMs is that they are sharply limited in their ability to model how many time steps any given state will last. In a standard HMM, the duration of a state $\mathbf{y}$ is modelled only with the single conditional probability $\widehat{f}_{\mathbf{Y}^{(t)}|\mathbf{Y}^{(t-1)}}(\mathbf{y} \mid \mathbf{y})$. *Hidden semi-Markov models* add a third layer of random variables $Z^{1:\infty}$ that act as counters until a possible state change (see figure 2.7). The evolution of $Z^{(t)}$ is mostly deterministic: for any time t such that $Z^{(t-1)} > 0$, one has $Z^{(t)} = Z^{(t-1)} - 1$ and $\mathbf{Y}^{(t)} = \mathbf{Y}^{(t-1)}$. When $Z^{(t-1)} = 0$, then $\mathbf{Y}^{(t)}$ transitions as in a standard HMM and $Z^{(t)}$ takes on a new duration based on the new state (Levinson 1986). This duration can be modelled in as much detail as desired (e.g., one could try to incorporate

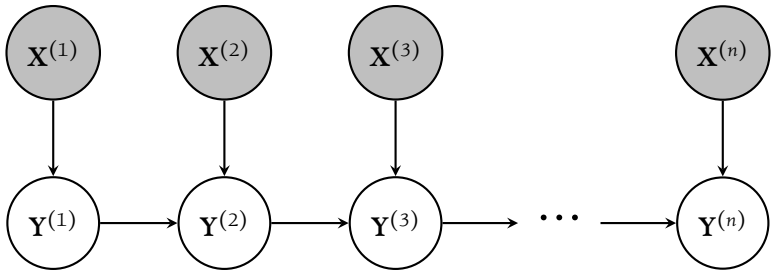


Figure 2.8 · A maximum-entropy Markov model (MEMM)

metric information when trying to predict the duration of a chord). *Segment* HMMS are an even more advanced way of modelling durations that allow every state to be associated with a randomly varying number of observations (Ostendorf, Digalakis & Kimball 1996).

Maximum-Entropy Markov Models

All of the models discussed so far have been generative models. As there became more interest in discriminative training for HMM-like models, one important step along the way was the *Maximum-entropy Markov model* (McCallum, Freitag & Pereira 2000). As can be seen in figure 2.8, maximum-entropy Markov models (MEMMS) are very similar to HMMS: only the direction of the dependencies between states and observations are reversed; in other words, the observations are thought to be causing the states. This assumption is in fact plausible for tasks like audio chord recognition if one thinks of harmonic analysis as a strictly *a posteriori* activity, although it has some curious implications. The joint distribution factorises as

$$\widehat{f}_{\mathbf{X}^{1:\infty},\mathbf{Y}^{1:\infty}} = \widehat{f}_{\mathbf{Y}^{(1)}|\mathbf{X}^{(1)}} \prod_{t=1}^{\infty} \widehat{f}_{\mathbf{Y}^{(t+1)}|\mathbf{X}^{(t+1)},\mathbf{Y}^{(t)}} \widehat{f}_{\mathbf{X}^{(t)}} \,, \tag{2.40}$$

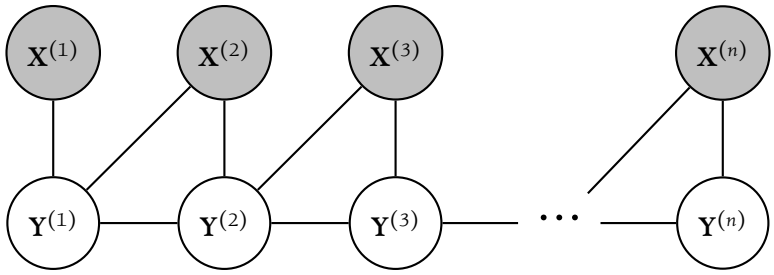


Figure 2.9 · A Markov network similar to an MEMM

which in particular implies that the state $\mathbf{Y}^{(t)}$ at some time t is independent of all observations that come later. That property can cause MEMMs to become stuck in a mis-classified state early on in the sequence and remain there despite evidence to the contrary later in the signal (Koller & Friedman 2009, pp. 952–53). The exciting property of MEMMs was that although like any Bayesian network, they trained generatively, the factorisation in equation (2.40) combined with the rule in equation (2.38) yields

$$\widehat{f}_{\mathbf{Y}^{1:\infty}|\mathbf{X}^{1:\infty}} = \widehat{f}_{\mathbf{Y}^{(1)}|\mathbf{X}^{(1)}} \prod_{t=1}^{\infty} \widehat{f}_{\mathbf{Y}^{(t+1)}|\mathbf{X}^{(t+1)},\mathbf{Y}^{(t)}} \,, \tag{2.41}$$

which is a Gibbs distribution on a Markov network like figure 2.9. The probability distributions represented by this Markov network are not exactly the same as the Bayesian network of the MEMM, but it is the closest possible approximation.

~ *Conditional Random Fields*

Unlike MEMMs, standard HMMs have no immoralities, and so it is possible to represent the same family of probability distributions as a Markov network

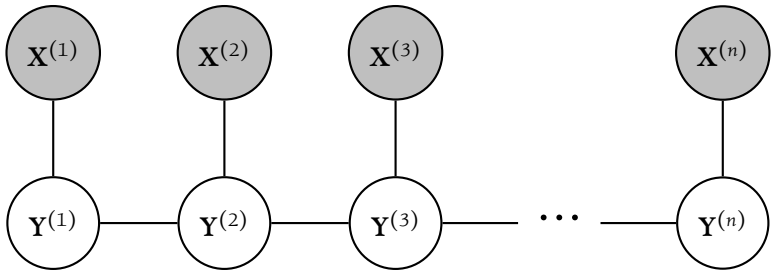


Figure 2.10 · A linear-chain conditional random field (CRF). The distribution represented is the same as that of a standard HMM, but the training strategy is discriminative rather than generative.

(see figure 2.10). This representation is known as a *linear-chain conditional random field*, and it is trained discriminatively, with all of the advantages (classification accuracy) and disadvantages (a need for large amounts of training data) that come with the discriminative approach (Lafferty, McCallum & Pereira 2001). Conditional random fields (CRFs) have been very successful for natural-language tasks (Sutton, forthcoming), but they have been little explored for musicological purposes. Given the relatively small size of data sets that had been available for musicological research until recently and the data hunger of discriminative training methods, this is perhaps understandable, but as the amount of data available for empirical research in music increases, discriminative approaches are well worth exploring. Not only have they demonstrated considerable improvements in performance over generative models, they have also led to promising links with classification approaches that historically had been unable to work for sequential data; *max-margin* HMMS (Taskar, Guestrin & Koller 2004), which combine CRFS with support vector machines (Hastie, Tibshirani & Friedman 2009, pp. 417–38), are among the most promising recent approaches

and are yet to be tried for musical applications.

Almost needless to say, there are many, many possibilities for choosing graphical models for the stochastic processes in database-driven musicology. Nonetheless, relatively few researchers have ventured away from fairly standard HMMs, as will be discussed in the next chapter.

3 MUSICOLOGICAL MARKOV CHAINS

Chapter 1 introduced database-driven musicology; chapter 2 described some of the mathematical tools that are necessary to engage it. This chapter will review how other researchers have applied these tools and similar tools to musical problems. It begins by presenting some seminal examples of corpus analysis on existing musical databases (§ 3.1), oriented toward those techniques of corpus analysis that are most useful as components of algorithms for labelling sequences. The next section (§ 3.2) describes the Music Information Retrieval Evaluation eXchange (MIREX), an annual competition that provides a good metric for what problems are important to researchers in music informatics. Section 3.3 reviews various approaches to solving these problems using state-space models, like those used in the remainder of this thesis. Readers familiar with David Temperley's *Music and Probability* (2007) will notice that there is an appreciable overlap between the references in this chapter and those in that volume, especially in the opening section, although Temperley's organisation and emphasis are rather different than my own.

3.1 MUSICOLOGICAL CORPUS ANALYSIS

Database-driven approaches to musicology have existed for more than a century. The earliest studies were certainly not described as such, and simply made general statistical observations about sets of music, e.g., that the frequency of melodic intervals is inversely related to their size (von Hornbostel 1906; Myers 1907; Watt 1924). Later, researchers presented

tables of statistics that had been computed by hand over relatively small corpora by current standards and cited those tables as evidence for their conclusions, which most commonly had to do with differences among musical styles or cultures. In the 1950s, a surge of interest in Claude Shannon and Warren Weaver's mathematical theory of communication (1949) prompted broader interest in these statistical approaches, and by the end of the 1960s, musicologists were able to use computers to facilitate their calculations (Babbitt 1965; Forte 1966a). In the 1980s, research in music psychology and music cognition introduced a new a kind of musicological corpus and a new approach to statistical analysis for musicology, and the efforts in this quarter to build computational models for human musical cognition laid an important foundation for the database-based classifiers that came later. Surprisingly, as music informatics developed throughout the 2000s, engineers made relatively little use of this body of research on musicological corpora, but the two communities seem to be slowly growing together.

Many of the database-driven analyses of musical corpora have been 'static' analyses: they do not attempt to reason about the temporal aspects of music. Lewis Lockwood and Arthur Mendel's groundbreaking work (1969) is a classic example. Working with the *Missa L'Homme armé super voces musicales* of Josquin des Prez, Lockwood and Mendel's analysis tabulates the proportion incomplete triads to complete triads as a means of distinguishing among mass movements but ignores the ordering of these sonorities. Any of the analyses of Brahms op. 51, no. 1 mentioned earlier (Forte 1983; Huron 2001; Conklin 2010) would be in a similar category, counting occurrences of different types and forms of motives and using these counts to classify

movements or composers. Other work, for example, Frauke Jürgensen's work with the Buxheim Organ Book (2005), draws conclusions about musical elements that depend on time in some way, e.g., the accidentals on notes where the melody changes direction, but does not link these elements to any general model of how music unfolds over time. Even David Huron's iconic analysis of the melodic arch in Western folk songs (1996) or Allen Forte's SNOBOL program for set-class analysis (1966b) would fall into this category. Static analyses can be of great musicological interest, but they fall out of the scope of this thesis: all of the classification approaches in this thesis, as described in section 2.4, rely on a direct understanding of temporal dependencies.

Of those musicological corpus analyses that do consider temporal dynamics, it is surprising how nearly universal is the notion of the Markov chain, especially considering that in addition to their inherently unmusical assumptions, Markov chains are also insufficient for learning hierarchical structures, which appear frequently in music theory (Chomsky 1957; Lerdahl & Jackendoff 1983; Dienes & Longuet-Higgins 2004; Rohrmeier 2011). Recall from equation (2.34) that a Markov chain is a discrete-time stochastic process that assumes that the random variables at each time point are independent of all others except a limited number of the immediately preceding random variables. Most often, the dependencies are restricted to the immediately preceding variables alone, usually known as *first-order* Markov assumption (although literature in psychology sometimes uses the term 'first-order' differently). For corpus analysis, using a Markov chain implies that there is an underlying domain of musicological inquiry Ω (e.g., the music of a particular composer or musical culture) along with an

appropriate σ-algebra $\mathcal{F}$ and probability measure $\mathbf{P}$, a meaningful series of random variables $Y^{(t)}$ that map each $\omega \in \Omega$ to musical notions in some semantic domain Ψ at discrete points in time (e.g., the successive pitches of a melody or harmonies of a piece of tonal music), and a corpus one can use to derive plausible estimates of $f_{Y^{(n)}|Y^{(n-k):(n-1)}}$, where k is the order of the Markov assumption. Sometimes k is assumed with no further comment (in particular, any study of melodic intervals with no other context inherently makes a first-order Markov assumption, $k = 1$), sometimes it is chosen to be whatever order is sufficient for distinguishing musical styles, and sometimes it is tested in its own right to see which value of k is most plausible given the corpus.

Although these formal mathematical structures necessarily underlie any study involving Markov chains, no musicologists have yet to my knowledge presented their results with respect to them. Even Temperley, who has rightly observed that probability can be a useful unifying viewpoint for interpreting many musicological studies (2007), stops well short of presenting the distinction between probability spaces and random variables. This omission is understandable given that no musicological research presents the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ or the random variables $Y^{(t)}$ fully and explicitly, but it makes it difficult to understand whether the corpora and techniques used to estimate $f_{Y^{(n)}|Y^{(n-k):(n-1)}}$ are appropriate and what conclusions are reasonable to draw from such estimates. Fortunately, the great majority of musicological studies with Markov chains fit one of a small number of basic patterns for Ω and the $Y^{(t)}$. The outcome spaces Ω, as mentioned above, typically comprise either the music of a single composer, culture, or traditional style or the music of a collection of a distinct number

of such composers, cultures, or styles; in the former case, researchers are normally seeking to describe a style, and in the latter, to compare or classify styles. The σ-algebra $\mathcal{F}$ is almost universally $\mathcal{P}(\Omega)$, and the probability measure $\mathbf{P}$ in most cases assigns equal probability to every set containing a single outcome $\omega \in \Omega$. The random variables $X^{(t)}$, as mentioned above, typically map each outcome ω to subsequent musical entities, most commonly the t-th notes in a melody or the t-th harmonies in a tonal composition. The techniques for estimating and working with $f_{X^{(n)}|X^{(n-k):(n-1)}}$, however, are much more varied and will be discussed as they arise below.

The remainder of this section presents specific examples of musicological corpus analysis from the literature. A history of studies invoking information theory comes first, as it motivated so much statistical work in music. Following this discussion, studies of melody and harmony each have their own section, due to the very large number of studies in each category. The section concludes by highlighting some of the most important corpus-based studies of other dynamic musical phenomena.

❧ *Information Theory*

As noted above, some of the earliest musicological corpus analysis that considered temporal dynamics arose from a wellspring of interest in Shannon and Weaver's mathematical theory of communication, also known as information theory (1949). Leonard Meyer was one of the first musicologists to consider how information theory might be applied to music (1957; 1962); in particular, Meyer observed that music might be profitably modelled under a first-order Markov assumption, like most of the models

presented in section 2.4. Most of these studies involve comparisons of *entropy*, which for a probability mass function f over a domain Ψ is defined as as

$$\mathrm{H}(f) \triangleq -\sum_{\psi \in \Psi} f(\psi) \log f(\psi) \,. \tag{3.1}$$

Entropy analysis can itself be static (e.g., Hiller & Bean 1966), but much of research on entropy and music has taken advantage of Meyer's observation that information theory and Markov models have a natural affinity.

A number of researchers have compared the entropy of pitch classes of melodies given the immediately preceding pitch class, i.e., the entropies of $f_{\mathrm{Y}^{(t)}|\mathrm{Y}^{(t-1)}}$ for each possible value of $\mathrm{Y}^{(t-1)}$, where the $\mathrm{Y}^{(t)}$ map melodies $\omega \in \Omega$ to their t-th pitch class (Pinkerton 1956; Youngblood 1958; Coffman 1992). The conditional mass functions $f_{\mathrm{Y}^{(t)}|\mathrm{Y}^{(t-1)}}$ reflect a first-order Markov assumption, of course, and in all of these corpus analyses, the authors estimated their values to be equivalent to the corresponding relative frequencies with which ordered pitch pairs appear in their corpora. This relative-frequency approach is probably the most common one for estimating mass functions in corpus analysis in general, and although it is always worth thinking for a moment about whether it applies in new situations, it most cases, such estimates are intuitively reasonable and have useful mathematical properties (see Doob 1934). Other researchers have used analogous approaches to study the entropy of melodic intervals (Hiller & Fuller 1967; Lewin 1968; Winter 1979; Snyder 1990), which amounts to the same Markov chain but with a more restrictive estimate of $f_{\mathrm{Y}^{(t)}|\mathrm{Y}^{(t-1)}}$ because it assumes that the interval pattern will be the same regardless of which pitch class is at time $t-1$. Elizabeth Margulis and Andrew Beatty

have performed one of the most thorough investigations of entropy as an analytical tool for melody, considering the temporal dynamics of pitch, interval, texture, contour, and duration in all voices of more than three hundred pieces, again using a relative-frequency approach (2008).

The relative-frequency approach to estimating mass functions is easiest to understand when the goals are 'suggestive' theories of music or style analysis (see § 1.1); psychological theories of music usually are looking to model expectation instead. Indeed, Mark Schmuckler has observed that most conventional theories of music have relied on some notion of expectation even in the absence of psychological theory or empirical data (1989). While some researchers have argued that relative frequencies should be good models for psychological expectation (see Simonton 1984), others have tried to model expectation more directly as psychological entropy: in the context of information theory, one interpretation of entropy is that it measures uncertainty or unpredictability, i.e., the inverse of the strength of expectation or *information* as defined in the mathematical theory of communication (Shannon & Weaver 1949). Following this interpretation, Leonard Manzara, Ian Witten, and Mark James ran a casino-style experiment on human subjects using betting behaviour to measure subjects' uncertainty at predicting the following note in a chorale melody and sought to explain the results using a more sophisticated notion of temporal dynamics, taking into account cadences (1992). Pushing this idea substantially further, Sarah Culpepper's recent thesis (2010) uses entropy as a tool for describing listeners' experiences as pieces unfold, in particular for highlighting salient discontinuities.

Melodic Markov Chains

Information theory is only one of many approaches that have been used to study melody as a Markov chain. In most cases, the sample space Ω of melodic studies represents melodies from multiple sources for those researchers interested in suggestive music theories, and for those interested in psychological theories, Ω represents melodies from a single culture. The $X^{(t)}$ all map these melodies to their t-th pitches or pitch classes, with only small variations in how to handle enharmonic equivalences. The techniques and degree of rigour in estimating the conditional mass functions $f_{Y^{(n)}|Y^{(n-k):(n-1)}}$ have varied substantially. For style analysis, which is discussed first, the relative-frequency approach is the most common; as with the entropy-based approaches, the most significant differences are usually whether researchers estimate values for all possible melodic transitions or whether they restrict the space by assuming that interval patterns will be the same regardless of starting note. For psychological music theories, however, which are discussed next, there is more variation.

The earliest work on melody that modelled melody with Markov chains, implicitly or explicitly, used small data sets of complete pieces and immediately recognised that the conditional mass functions had sharp peaks and thus great potential for describing and distinguishing musical styles (Watt 1924; Fucks 1962). This branch of research quickly moved on to larger data sets, but because working with large data sets was very labour-intensive, most of these studies used only the incipits of the pieces in each corpus. William Paisley made one of the earliest studies of transition frequencies in incipits, based on the first six notes of a large selection from a dictionary of

musical themes, with the goal of determining authorship (1964). Benjamin Suchoff did some early ethnomusicological work with melodic interval patterns in the first seven notes of a selection of folk songs from Béla Bartók's collection (1970). Somewhat later, and with stronger statistical tools, Alison Crerar conducted work on a series of incipits drawn from the work of five eighteenth-century composers (1985). Dean Simonton ran a series of studies using transition probabilities from the first six notes of melodies from a classic dictionary of musical themes (Barlow & Morgenstern 1948, 1950) to predict thematic fame and identify transhistorical trends (1980a; 1980b). Investigating several hypotheses about nationalism from a classic survey on Romantic chamber music, Fred Hofstetter made a more statistically sophisticated comparison of first- and second-order Markov chains of intervals (1979); the spirit of this study is similar to Lynn Trowbridge's study of composers' styles in fifteenth-century chansons (1985–86), although Trowbridge's statistics are considerably simpler. In a very creative analysis, David Huron linked the verbal 'qualia' that musicians use to describe different scale degrees in Western tonal music to a first-order Markov chain of scale degrees in Germanic folk melodies (2006, pp. 158–67). Jon Gillick, Kevin Tang, and Robert Keller have developed a system that can be used to work with such corpora in general, allowing for first-, second-, or third-order Markov chains to model melodies or melodic fragments (2010).

Although working with the conditional mass functions directly is the classical approach, some researchers have worked with cleverer derivatives. When the only goal is distinguishing among styles, for example, fully specified conditional mass functions involve considerably more parameters than are necessary; Jan Beran has suggested a novel approach derived from

the *stationary distributions* of melodic Markov chains (2004, pp. 175–84), i.e., the distribution for a Markov chain such that

$$f_{\mathbf{Y}^{(t+1)}|\mathbf{Y}^{(t)}} = f_{\mathbf{Y}^{(t)}|\mathbf{Y}^{(t-1)}} \tag{3.2}$$

(Billingsley 1995, pp. 124–31). Comparable to Margulis's thorough study of melody from the perspective of information theory, Darrell Conklin and Mathieu Bergeron have developed a data-mining system for melodies that uses a broad range of features, including directed and undirected intervals and contours (Conklin 2006; Conklin & Bergeron 2008); although their results are not presented as such, the system includes a wide variety of Markov-chain representations as special cases and is believed to have potential for achieving many of the benefits of full grammars with the underlying simplicity of Markov chains (Conklin & Witten 1995; Pearce & Wiggins 2004; Pearce 2005; Rohrmeier 2011). Panayotis Mavromatis has developed a system for analysing liturgical chants from the Greek Orthodox tradition that learns a considerably more elaborate Markov structure (in fact, a type of hidden Markov model) that is able to take account of differing melodic contexts (2006; 2009).

Another group of researchers have focused on the use of melody to study music perception. Diana Deutsch's early work focused on hierarchical models of melody (Deutsch 1980; Deutsch & Feroe 1981). Such models were the zeitgeist of the time, leading in music-theoretical quarters to Lerdahl and Jackendoff's landmark *Generative Theory of Tonal Music* (1977; 1983), which advocated strongly for a hierarchical, tree-based understanding of music. Mari Riess Jones also discussed hierarchical theories of melodic perception in her work, although with a more explicit connection to sim-

pler, Markov-style representations (1976). James Carlsen ran a series of experiments on the potential of melodic intervals to generate expectancy for the subsequent tone (Carlsen 1981; Unyk & Carlsen 1987), a very different approach to the counting approach that yields a second-order Markov model: the values of $f_{X^{(t)}|X^{(t-1)},X^{(t-2)}}$ are derived from subjects' ratings of how much different continuation pitches would be expected given a presented musical interval.

At this point, heated questions arose about whether and how humans learn Markov-chain statistics in music and whether and how humans might use such statistics when cognising music. One particularly lively debate about melodic Markov chains sets 'structural' and 'distributional' theories of key perception against each other. David Butler has been very critical of the classic 'key-profile' theory made famous by Carol Krumhansl and Edward Kessler, arguing that the temporal dynamics of melodies, particularly rare intervals, must play a critical role in identifying keys. (Krumhansl & Kessler 1982; Butler 1989; Brown, Butler & Jones 1994); Krumhansl's counterargument was that although temporal order is important, the temporally relevant factor in her experiments appeared to be relative tonal stability rather than the interval itself (1990b). David Temperley used results proving that humans are sensitive to transition probabilities to test how melodic transition probabilities related to key and key detection and found that the for key detection, such transitions are unnecessary (2008).

For other purposes, however, humans are clearly sensitive to transition probabilities, as Krumhansl herself admits (1990a, pp. 111–37). Piet Vos and Jim Troost used Markov representations of melodic intervals to generate melodies that followed the normal statistical patterns of Western music

and another set of melodies that reversed the direction of all intervals, and subjects were able to identify the difference (1989). Jenny Saffran and colleagues later showed that human infants and adults are sensitive to changes of first-order Markov probabilities even in random melodic sequences (Saffran et al. 1999). More specifically, their experiments tested whether of *implicit learning* of such statistics is possible: subjects were exposed to random sequences of pitches obeying a defined set of transition probabilities and then tested to see whether they could identify which of two new examples matched the 'language' to which they had been exposed. Zoltán Dienes and Christopher Longuet-Higgins explored implicit learning in atonal music and found that although *n*-gram probabilities are certainly an important component of implicit learning, experienced listeners of atonal music are also able to identify other characteristics (Dienes & Longuet-Higgins 2004). Psyche Loui, David Wessel, and Carla Hudson Kam were able to show that even when using the Bohlen-Pierce scale (an artificial, non-Western scale), humans are able to learn first-order melodic grammars implicitly (2010).

Like Beran, music psychologists have also seen value in reducing the number of parameters necessary to describe melodic Markov chains. Eugene Narmour's implication-realisation theory of melodic expectation (1990; 1992) is a famous example of a simpler model that could underly a second-order Markov chain of the same type as Carlsen's experiments. As formalised in the later literature, Narmour's theory suggests that melodic Markov chains are governed by five principles: *registral direction*, specifically that small intervals imply continuation in the same direction and that large intervals imply a change of direction; *intervallic difference*, specifically that

small intervals imply more intervals of the same size whereas large intervals imply smaller intervals; *registral return* to tone no more than a whole tone away from the first note of a melodic interval; *proximity*, or a general preference for smaller intervals; and *closure*, or the pattern of a large interval followed by a smaller interval in a different direction (Schellenberg 1996). Three independent experiments confirmed that, combined with a simple model of tonality advocated by Krumhansl and Kessler (1982), the implication-realisation model was a good description of listener's expectations, and these experiments were reconfirmed on Carlsen's data (Carlsen 1981; Unyk & Carlsen 1987; Cuddy & Lunney 1995; Schellenberg 1996; Thompson, Cuddy & Plaus 1997). Nonetheless, these experiments also revealed potential for making the model even more parsimonious, ultimately yielding a three-variable model based only on a revised notion of proximity, a combination of pitch proximity and registral return, and a tonality factor (Schellenberg 1997). Given these three parameters and Schellenberg's formula, it is possible to generate a full set of conditional mass functions $f_{X^{(t)}|X^{(t-1)},X^{(t-2)}}$, although later studies, e.g., Pearce & Wiggins (2006), have found that just three parameters may be an over-simplification. Although the numbers of parameters vary, the same general ability to derive Markov chains is also true of other perceptually motivated models of melodic tension such as Fred Lerdahl's (2001) or Steve Larson's (2004).

Harmonic Markov Chains

With a few exceptions (e.g., Youngblood 1970), the information-theory movement had dealt little with harmony, but otherwise, harmony is second

only to melody in the number of treatments it has received under the formalism of Markov chains. Unlike melodic Markov chains, however, harmonic Markov chains have been used somewhat less for style analysis, and thus often the underlying sample space Ω is a unified repertory of tonal pieces both for psychologically and descriptively motivated researchers. Analogous to the case with melody, the random variables $X^{(t)}$ typically map each $\omega \in \Omega$ to their t-th harmony, although there is substantial variation in whether this should be done with respect to key and how elaborate the dictionary of allowable harmonies should be.

It is considerably more time-consuming to generate databases of harmonies than it is to generate database of melodies because harmonies must normally be derived by a human analyst and cannot be entered directly from a musical score. Because of this expense, until recently, fewer of the early Markov-chain treatments of harmony were strictly corpus analyses: more appeared in the context of musical classification problems that are discussed later in section 3.3. This subsection first treats what corpus analyses there have been, for classical music and then for popular music, followed by a discussion of how researchers in music cognition have handled harmony.

Perhaps even more than melody, many modern theories of harmony have been explicitly hierarchical, most famously Schenkerian analysis (Salzer 1952) and the Lerdahl-Jackendoff generative theory (Lerdahl & Jackendoff 1983; Lerdahl 2001). The first computerised system for harmonic analysis, designed by Terry Winograd, likewise followed a complex, hierarchical grammar inspired by models from linguistics (Winograd 1968), and similar models maintained currency in the computer-music field (Ulrich 1977; Meehan 1980; Steedman 1984; Terrat 2004) and music psychology

(Deutsch & Feroe 1981; Krumhansl 1990a). More traditional theories of harmony, however, dating back to Jean-Philippe Rameau (1722) and codified by Walter Piston for widespread use in North American pedagogy (1941), have focused on the frequency of intervals between successive roots. Composer James Gabura was one of the first researchers to use first-order Markov statistics on harmony as a type of style analysis, in combination with melodic intervals (1970). Using the prevailing key of the harmony at the start of each bar of every symphony – all tabulated by hand! – Cecil Marillier presented a statistical description of Haydn's tonal plans and how they evolved throughout his career (1983). Dmitri Tymoczko evaluated three theories of harmony: a Ramellian one based on movements of a fundamental bass (see Rameau 1722; Meeùs 2003), one based on Markov chains, and one based on Riemannian functions (see Riemann 1893), and concluded that the first-order Markov chain yielded the most accurate representation of harmony in Bach's chorales (Tymoczko 2003).

In popular music, Matthias Mauch and colleagues ranked all four-chord progressions in a corpus of Beatles and Real Book songs, which although it is not itself a Markov chain, could be converted into a first-, second-, or third-order chain if desired (Mauch et al. 2007). Using Christopher Harte's larger corpus of Beatles songs (2005), Ricardo Scholz, Emmanuel Vincent, and Frédéric Bimbot investigated which order of Markov assumption is necessary to explain harmony well (2009). In general, they found that at least a fourth-order model was necessary to model harmony well, but using various adjustments to strict relative-frequency estimation (see also Pickens 2003), they were able to find first-order Markov chains that worked well. Previous research on similar corpora had mostly restricted itself to

first-order models without smoothing and achieved more limited success (Pachet 1999; Papadopoulos & Peeters 2007). Mitsunori Ogihara and Tao Li explored different harmonic Markov chains for identifying jazz composers, again concluding that fourth-order models work best (Ogihara & Li 2008). Trevor de Clercq and David Temperley have undertaken the most substantial corpus analysis of harmony to date, using a collection of 100 songs selected from *Rolling Stone* magazine's '500 Greatest Songs of All Time' (de Clercq & Temperley 2011); this analysis provides relative frequencies in the corpus for a first-order Markov chain of chord roots and provided the basis for a more detailed investigation specifically of the use of sub-dominant chords in rock (Temperley 2011).

Although a number of experiments have suggested that Western listeners have internalised some kind of statistical understanding of harmony, relatively few studies can be reduced to any kind of Markov chain – or indeed, any kind of explicit statistical model. Mark Schmuckler was able to verify that Piston's model of harmonic progressions matches Western listeners' expectations fairly well, although his experiments also suggested that a first-order chain was insufficient to explain the experimental results completely, especially with respect to relatively rarer harmonic progressions (1989). David Smith and Robert Melara found that even novice listeners were sensitive to changes in harmony that seemed to violate the normal sequential grammar (1990), but the study was based on variations to a single archetypical harmonic progression. Erin Jonaitis and Jenny Saffran tested human potential for learning harmonic idioms implicitly and showed that even for completely artificial, unfamiliar harmonic idioms, humans are able to sensitise themselves to their statistical properties quite

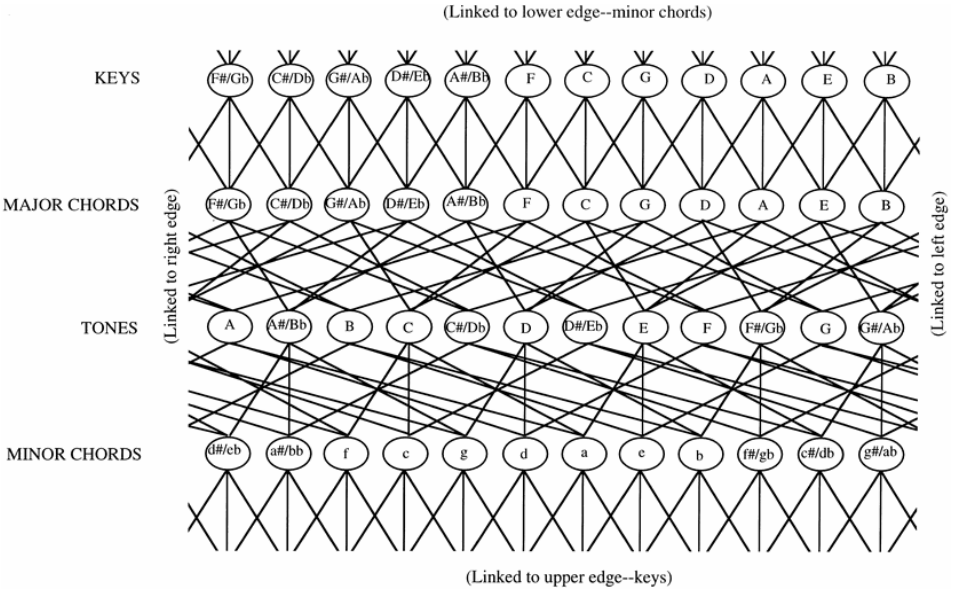


Figure 3.1 · Jamshed Bharucha's MUSACT model (Bharucha 1987, p. 16). This self-organising network receives individual tones as input (the third row) and spreads activations from these tones to chords (the second and fourth rows) and keys (the first row). Over time, these activations decay, and one can use the pattern of activations in the network to reflect degrees of expectation.

quickly (2009). Surprisingly, given the methodical checks in the literature on melody that human expectation aligns with the relative frequencies actually found in corpora, for harmony, the presumption seems to be that because experiments have shown that Western listeners are sensitive to harmony in general, their expectations must necessarily be identical to whatever statistical regularities can be found in Western harmony.

Instead, much of the psychological literature on harmony has focused on developing computational models for human perception. Jamshed Bharucha has been the most active researcher in this area, promoting the MUSACT model (Bharucha 1987). MUSACT is a type of hierarchical, self-

organising map (SOM) (see Kohonen 2001), illustrated in figure 3.1. The mathematical details of SOMs are immaterial for this thesis, but the core concept underlying them is that every input to the system (musical tones in the case of MUSACT, as illustrated in the third row of nodes in the figure) spreads a fixed degree of 'activation' to all other nodes connected to it, which in turn pass some activation along to the nodes connected to them, and so on; over time, these activations decay. The gradually decaying activations are meant to model long-term memory, and at any point in time, the relative activations of different nodes can be used as a model for expectation, e.g., after entering a sequence of harmonies into MUSACT, one could compare the activation levels of all of the chords (the second and fourth rows of nodes in the figure) to model expectations for which chord might come next. Bharucha himself has explored alternative models using similar approaches (Bharucha & Todd 1989; Bharucha 1999) as has Robert Gjerdingen in his *'l'ART pour l'art'* model (1989; 1990). Working with Bharucha and Emmanuel Bigand and inspired by yet another self-organising model of Niall Griffith (1994), Barbara Tillmann developed a self-organising computational model of harmonic perception, compared its performance to a large number of perceptual studies that had been performed previously, and found that the self-organising map behaved very similarly to subjects across the perceptual experiments (2000).

These results suggest that SOMs are appropriate statistical models for studying harmony, and they do have an advantage over Markov chains in that by decaying activations gradually, they are able to manage long-range temporal dependencies. Nonetheless, the parameters of SOMs are difficult if not impossible to interpret on their own, in stark contrast to the

relatively clean interpretations of Markov chains; moreover, almost all of the classification approaches discussed below rely on some kind of Markov-chain model, although it could be a very productive study to investigate whether any of these approaches could be adapted to use one of these SOMs instead. It is possible, of course, to check the activations after these networks after short or even single-chord sequences to derive estimates of the conditional mass functions $f_{Y^{(n)}|Y^{(n-k):(n-1)}}$ for a Markov chain, and in fact, Bharucha and Peter Todd have shown that self-organising neural networks learn transition probabilities effectively (1989). Furthermore, inspired by some early work in statistical ethnomusicology (Freeman & Merriam 1956), Krumhansl and colleagues tested behavioural experiments, statistical style analyses, and an SOM and concluded that all three approaches were comparable (2000).

Other Musical Markov Chains

Although melody and harmony have received the bulk of database-driven musicological treatments, a few other examples are worthy of note. Nigel Nettheim has observed that Knud Jeppesen's classic work on Palestrinian counterpoint is essentially statistical, although Jeppesen only presented it as such in a few tables, one analysing text setting and another analysing the number of dissonances in certain movements (Jeppesen 1923; Nettheim 1997). Rather than focusing on a single composer, Lesley Mearns, Dan Tidhar, and Simon Dixon have used first-order representations of contrapuntal techniques as a means of style analysis among seven Baroque composers (Mearns, Tidhar & Dixon 2010). In a similar vein, although not exactly

counterpoint, Ian Bent and John Morehen designed a classic computer program to solve text underlay problems in Renaissance music (1977–78). David Huron has conducted a number of creative studies over the years, many of which imply some kind of Markov chain. These studies have included patterns of musical entries (Huron 1990b), patterns in crescendos and decrescendos (Huron 1990a), and rhythmic syncopation (Huron & Ommen 2006). Timbre, always the ugly duckling in musicological studies, has been shown to be susceptible to implicit learning (Bigand, Perruchet & Boyer 1998; Tillmann & McAdams 2004). Finally, because they are generative models, Markov chains have also been of considerable interest for algorithmic composition (Ames 1989; Cope 1991, 2005),

It is also worth mentioning functional data analysis, a particular technique used for modelling continuous stochastic processes rather than discrete stochastic processes (Ramsay & Silverman 2005). Functional data analysis is far too complex a field to summarise here, but Bradley Vines, Regina Nuzzo, and Daniel Levitin have nicely described how to apply functional data analysis to music as an alternative to Markov chains (2005).

3.2 THE MUSIC INFORMATION RETRIEVAL EVALUATION EXCHANGE

All of the applications in section 3.1 presuppose the existence of a corpus, or database, for analysis. As explained in section 1.2, constructing large musical databases is much more practical when automatic tools can be used to label, or *classify*, digital representations of music in physical forms, most typically digital audio or scanned scores. As discussed in section 2.3, a classifier is usually seeking to model $f_{\mathbf{Y}^{1:\infty}|\mathbf{X}^{1:\infty}}$ for random variables $\mathbf{X}^{(t)}$ that

correspond to semantic qualities and $\mathbf{Y}^{(t)}$ that correspond to observable but less interpretable qualities of elements in a sample space Ω; when generating musical databases, Ω usually represents all pieces in a large genre or meta-genre of music together with low-level digital representations of those pieces, generally as audio, a symbolic format for notated music like MIDI, or as digital images of scores.

Although there are many sources of research on music classification, the foremost conference for this domain is the International Conference on Music Information Retrieval, commonly known as ISMIR, which has been held annually since the year 2000. After a few years, the ISMIR community sought to formalise some of the most fundamental tasks that would be treated at the conference and to develop a means of evaluating the performance of competing approaches to these tasks. A first attempt was made at the fifth ISMIR (2004) to address this issue, in the form of the Music Information Retrieval Audio Description Contest; its contests included genre classification, artist identification, melody extraction, tempo induction, and rhythm classification. Starting the following year, a more formal exchange, known as the Music Information Retrieval Evaluation eXchange (MIREX) began and has been held concurrently with the annual ISMIR ever since (Downie 2008).*

Table 3.1 lists all tasks that have been run at MIREX more than once. Dots in each column indicate that the task was run in a given year. The tasks fall into three broad categories. One category, including the two longest-running tasks, comprises classic tasks of information retrieval, following a

* http://www.music-ir.org/mirex/

Table 3.1 · Selected MIREX tasks

TASK	2005	2006	2007	2008	2009	2010	2011
INFORMATION RETRIEVAL							
Audio cover song identification		•	•	•	•	•	•
Query by singing/humming		•	•	•	•	•	•
Query by tapping				•	•	•	•
Symbolic melodic similarity	•	•	•			•	•
Audio similarity		•	•			•	•
SINGLE-POINT CLASSIFICATION							
Audio artist identification	•		•	•	•	•	
Audio genre classification	•		•	•	•	•	•
Audio mood classification			•	•	•	•	•
Audio tag classification				•	•	•	•
Key finding	•					•	•
Audio tempo estimation	•	•				•	•
Audio composer identification			•			•	•
SEQUENCE LABELLING							
Multiple f_0 estimation and tracking			•	•	•	•	•
Audio melody extraction	•	•		•	•	•	•
Score following		•		•	•	•	•
Audio chord estimation				•	•	•	•
Audio onset detection	•	•	•		•	•	•
Audio beat tracking		•			•	•	•
Audio structural segmentation						•	•

model of a user presenting a large database with some query and expecting a list of results that match that query or are in some way similar to it: audio cover song identification, query by singing or humming, query by tapping, symbolic melodic similarity, and audio similarity. A larger category involves single-point classification, where a single label is applied to an entire piece of music: audio artist identification, audio genre classification, audio mood classification, audio tag classification, audio tempo extraction, audio and symbolic key finding, and audio (classical) composer identification. The final category, containing as many tasks as the second, is the category of most interest to this thesis: classification tasks that involve classifying regions or moments of a piece of music as it unfolds over time.

This final category is of particular interest because it tends to represent the type of 'reduced' data that is of particular value for musicological research. One of the holy grails of music classification is the longest-running task in the category, multiple-f_0 (fundamental-frequency) tracking, which is a relatively small step away from transcribing full scores from audio files; it seeks to (1) estimate the active pitches sounding at each moment of an audio file, (2) track the overall contours of simultaneous melodic lines, and (3) distinguish changes in timbre. Many of the other MIREX tasks are simplifications of f_0 tracking: audio melody extraction simplifies multiple-f_0 tracking by tracking the prevailing melody only; score following simplifies multiple-f_0 tracking by providing a full score and looking only for a time-alignment with the audio; audio chord detection simplifies it by looking for chord labels only instead of complete transcriptions; audio onset detection by seeking only the beginning of each musical event, regardless of pitch. All of these tasks can benefit from knowledge about how the music in

question tends to unfold over time, or more formally, knowledge about the distribution of the random variables that will be the classification results. The same is also true of the two remaining tasks in this category, audio beat tracking and audio structure analysis, although both of these tasks have even higher-level goals than multiple-f_0 tracking and its relatives.

MIREX encompasses most of the historically significant tasks in music classification tasks, but there are several others worthy of mention. In most cases, they have been omitted from MIREX due to the lack of a common metric for evaluating their performance. A number of researchers have sought to generate chorale harmonisations, typically in the style of Johann Sebastian Bach, for a given melody. Another venerable problem is optical music recognition (OMR), which seeks to recover an electronic version of a musical score from digital images of that score.

The following sections will explore how different classification algorithms have been applied to the problems in this third category.

3.3 MUSICAL STATE-SPACE MODELS FOR CLASSIFICATION

As discussed in section 2.3, state-space models lend themselves particularly well to problems in musicological questions like those in this third category because they are designed to reduce observable, low-level features $\mathbf{Y}^{(t)}$ to semantically meaningful, high-level states $\mathbf{X}^{(t)}$. There are classifiers based on other types of models. Signal-processing techniques that rely on modelling the temporal dynamics of the low-level $\mathbf{Y}^{(t)}$ in audio have worked particularly well for onset detection and beat tracking in particular (Bello et al. 2005; McKinney et al. 2007). Other techniques, sometimes known as

'sliding window' methods (Dietterich 2002), classify each individual $\mathbf{X}^{(t)}$ only from their corresponding $\mathbf{Y}^{(t)}$; in other words, they ignore temporal relationships, assuming that $f_{\mathbf{X}^{(t)}|\mathbf{X}^{1:(t-1)},\mathbf{X}^{(t+1):\infty}} = f_{\mathbf{X}^{(t)}}$ for all $t \in \mathrm{T}$. These types of classifiers, however, are less interesting musicologically because they ignore the relationships among the $\mathbf{X}^{(t)}$, which are the only aspects of these models that are musically intelligible. This section reviews state-space models in the many ways they have been applied to classification tasks in music.

✷ *Rule-Based Systems*

Some state-space models for music classification eschew any notion of probability and work directly with a collection of heuristic rules. Rule-based systems are usually confined to working with symbolic data rather than audio or images due to the daunting complexity of audio and images. They also suffer from being guaranteed not to be consistent in the mathematical sense, and so as discussed in section 2.3, any system that learns from a database will theoretically outperform any competing rule-based system given a sufficiently large database (unless, of course, the heuristics happen to be perfectly correct and also happen to describe the underlying phenomenon completely). Nonetheless, due to their simplicity, rule-based systems have been historically popular.

Rule-based classifiers are particularly prominent in automatic systems for harmonic analysis in tonal music. Such systems generally start with a symbolic representation of a music score, e.g., a MIDI file, and return some kind of Roman-numeral analysis of the contents. The most famous

such system is probably David Temperley and Daniel Sleator's MELISMA analyser (Temperley & Sleator 1999; Temperley 2009), but there have been other notable examples over the past two decades (Maxwell 1992; Prather 1996; Taube 1999; Rowe 2001; Choi 2011). As languages for computer programming have evolved, it has even become possible to encode rules for harmonic analysis directly into the programming language itself (see Magalhães & de Haas 2011). Taking these tasks a degree further, Thomas Rocher and colleagues have built a completely rule-based system for harmonic analysis of audio, inspired by Fred Lerdahl's notions of pitch space (Lerdahl 2001; Rocher et al. 2010). David Cope's SPEAC system is another notable example in the category of rule-based analysis, an elaborate system of musical analysis oriented toward generating new compositions mimicking the styles of famous composers (2005, pp. 221–50).

A second area where rule-based classifiers have been notably popular is pitch spelling. Like rule-based classifiers for harmonic analysis, pitch spellers begin with symbolic representation of a musical score, specifically when those representations, like MIDI, neglect to distinguish enharmonically equivalent pitches. Pitch spellers seek to identify the correct spelling of ambiguous pitches. Temperley and Sleator included pitch spelling as part of the MELISMA analyser (Temperley 2001, pp. 115–36). Other well-known rule-based pitch spellers include that of Emilios Cambouropoulos, which is based on interval patterns (2003); David Meredith's ps13 algorithm (2006); and Elaine Chew and Yun-Ching Chen's algorithm based on the spiral array (Chew & Chen 2005; Meredith 2007).

Perhaps surprisingly, there has been relatively little work on finding consistent, database-based classifiers for either symbolic harmonic analysis

or pitch spelling, in contrast to the one other major application of rule-based state-space models in MIR, querying databases for melodies, where probabilistic methods now dominate. The early approaches to retrieving melodies used various string-matching techniques, including prefix or suffix trees (Blackburn & De Roure 1998; Chen et al. 2000) and approximate string matching (Ghias et al. 1995; McNab et al. 1996). The probabilistic replacements for these algorithms are discussed later in this section.

← *Smoothed Sliding-Window Systems*

So-called 'smoothed' sliding-window algorithms are a hybrid of between completely rule-based systems and fully probabilistic systems. Recall from the introduction to this section that a standard sliding-window approach derives a probabilistic, database-based classification rule for mapping low-level features to high-level labels but does so without regard to temporal dependencies. Smoothed sliding-window algorithms add some kind of rule-based 'smoother' to the results of sliding-window classification to add some notion of temporal dependency post-classification. Such methods will still not be consistent unless the outcomes in Ω do indeed behave exactly according to the smoothing rule, but again, because of their relative simplicity, smoothed sliding-window algorithms can be popular.

Audio melody extraction and audio chord recognition have been the most common applications for smoothed sliding-window algorithms in MIR. Jana Eggink developed a system for audio melody extraction that uses a series of preference rules rather than a Markov chain to link observations about the audio to temporal dynamics and found that adding

temporal information improved the system's performance substantially over an unsmoothed, strict sliding-window approach (2004). The next year, Rui Paiva, Teresa Mendes, and Amílcar Cardoso presented a similar system using the classic observation that smaller intervals are more common in Western melodies than larger intervals (2005). Ricardo Scholz and Geber Ramalho added smoothing to a classic sliding-window algorithm for chord recognition (Pardo & Birmingham 2002; Scholz & Ramalho 2008), again finding that the smoother improved results. In the same year, Johannes Reinhard, Sebastian Stober, and Andreas Nürnberger presented an alternative smoother and tested it with a number of local chord classifiers (2008).

Blackboard Systems

Several researchers have used *blackboard systems* (Engelmore & Morgan 1988) to attack polyphonic transcription. Blackboard systems can be strictly rule-based, but in practise, they also fall into a grey area between rule-based and probabilistic models. Like rule-based systems and smoothed state-space models, they have an intuitive appeal but are not in most cases mathematically consistent (Carver 1997). The basic concept of the blackboard system is that there is a virtual 'blackboard' where a variety of virtual 'experts' can propose partial, approximate solutions to sub-components of a larger problem. Blackboard systems allow the virtual experts to add or improve solutions to the blackboard repeatedly as well as 'erase' solutions that have come to seem too sub-optimal relative to the other solutions on the blackboard. These systems were popular in the 1980s and then gradually fell out of favour in preference to Bayesian and Markov networks, although

Norman Carver has argued strongly that blackboard systems can offer more flexibility than graphical models and are still well suited to large, complex problems where defining an accurate graphical model is difficult (1997).

Automatic music transcription is such a large, complex problem where defining an accurate graphical model can be difficult, and this area is where blackboard systems have been most used in MIR. Following up on several technical reports from the MIT Media Lab, Juan Bello, Giuliano Monti, and Mark Sandler presented a blackboard system for automatic music transcription at the first ISMIR (2000). The basic principle of the system was to maintain several candidate transcriptions on a virtual 'blackboard' for each moment in time. Masataka Goto's PreFest algorithm uses a similar system to extract the predominant melody and bass lines in popular music (2000; 2001), although he does not describe it as a blackboard system.

One notable application of blackboard systems outside of music transcription is a unique system for symbolic harmonic analysis from Takuya Yoshioka and colleagues (2004). Like Goto's PreFest system, it does not describe itself as a blackboard system, but the underlying concept is the same: based on a number of virtual experts, several candidate transcriptions are maintained as the system tracks a piece of music, and at the end of the piece, the most likely of the hypotheses is chosen as the best analysis.

← *Markov Chains for Music Retrieval*

In the musicological style analyses presented in section 3.1, one strategy was to learn individual Markov models for groups of pieces for which the categories are already known and to examine the differences between

the model parameters. A slight variant on this technique can be used for information retrieval, whereby one learns a Markov model for each document in a database and, given a query, returns the documents whose models are most consistent with the query. At the first ISMIR, Jeremy Pickens described a simplified version of such a system for melodies, although his system did not use all of the details of Markov models, rather rounding all positive probabilities to one (2000); with a large group of collaborators, Pickens later extended this technique for polyphonic music retrieval as well (Pickens et al. 2002). Using a similar approach, Stephen Downie and Michael Nelson performed a rigorous statistical analysis of various melodic Markov structures to find which worked best for music retrieval (Downie & Nelson 2000), extending work in this direction from Alexandra Uitdenbogerd and Justin Zobel (1999). Holger Hoos and colleagues developed a retrieval model using more proper first-order Markov chains (2001).

In a close variant, Jia-Lien Hsu, Chih-Chin Liu, and Arbee Chen used a structure known as the correlative matrix for music retrieval (2001). Although it is not presented as a Markov chain, these matrices keep track of Markov statistics and provide a natural way of averaging over Markov chains with different values of k. Ultimately, however, the more popular variants of these techniques move up a step in complexity to full hidden Markov models or related approaches, which are discussed below.

Hidden Markov Models

Although not quite as dominant as Markov chains are in basic musicological corpus analysis, there is no question that hidden Markov models (HMMS)

have a particularly important role in classifying musicological sequences. Section 2.4 provided a formal description of HMMs, and provided that this formalism accurately reflects the joint density $f_{\mathbf{X}^{1:\infty},\mathbf{Y}^{1:\infty}}$ of the random variables under consideration, for the first time among the models considered in this section, the standard techniques for learning the parameters of an HMM from data (namely, maximum likelihood) yield a mathematically consistent classification rule (Leroux 1992).

For a more intuitive illustration of when and how HMMs are useful, Wen-Huang Cheng's work with colleagues on segmenting wedding videos (2008) is an excellent example. Noting that traditional Western Christian weddings for heterosexual couples tend to consist of a subset a fairly standard set of components (see table 3.2), Cheng and colleagues sought an algorithm that could extract low-level, time-ordered audiovisual features from videos of wedding ceremonies and use them to segment and label the videos according to these standard components. Because almost all of these events tend to occur in a single location (the front of a church) with overlapping groups of people speaking or making music, it is difficult to derive any kind of audiovisual feature that would work on its own to distinguish all of the wedding events reliably; in other words, any sliding-window technique is likely to be disappointing. Much as with musical events, however, wedding events exhibit a significant dependence on temporal context, a dependence that Cheng and colleagues wanted to exploit to make segmenting the videos easier. The bride, for example, always enters after the groom, although the groom may or may not have entered with the main party; the choir may sing at a variety of moments, but it is more than twice likely to do so after the officiant has said something than immediately after the wedding kiss.

Table 3.2 · Taxonomy of wedding events

CODE	EVENT	DEFINITION
ME	Main Group Entering	Members of the main group walking down the aisle.
GE	Groom Entering	Groom (with the best man) walking down the aisle.
BE	Bride Entering	Bride (with her father) walking down the aisle.
CS	Choir Singing	Choir (with participants) singing hymns
OP	Officiant Presenting	Officiants giving presentations, e.g., invocation, benediction, and homily.
WV	Wedding Vows	Couple exchanging wedding vows.
RE	Ring Exchange	Couple exchanging wedding rings.
BU	Bridal Unveiling	Groom unveiling his bride's veil.
MS	Marriage License Signing	Couple (with officiants) signing the marriage license.
WK	Wedding Kiss	Groom kissing his bride.
AP	Appreciation	Couple thanking certain people, e.g., their parents or all participants.
ED	Ending	Couple (followed by the main group) walking back down the aisle.
OT	Others	Any events not belonging to the above, e.g., lighting a unity candle.

Source: Reproduced from Cheng et al. (2008, p. 1640).

Note: The main group indicates all persons, except the ones in GE and BE, who are invited to walk down the aisle, e.g., flower girls, ring bearers, groomsmen, bridesmaids, honorary attendants, officiants, etc.

Moreover, although audiovisual features on their own are insufficient to distinguish all types of wedding events, any given event is suggestive of certain audiovisual features: a single speaking voice, for example, while the officiant is presiding, and multiple voices singing during events featuring the choir. HMMs excel at exactly this type of problem: when low-level features on their own are expected to be insufficient for accurate classification (possibly only because better features are impractically difficult to obtain) but when information about time dependency may be sufficient to correct for their weaknesses.

≈

As noted above, simple Markov-chain techniques for retrieval eventually progressed to HMM techniques. Many of these techniques use a more restricted form of HMM known as dynamic time warping, which are treated below, but Riccardo Miotto and Nicola Orio have used standard HMMs in a multi-step audio retrieval processes. In a first version of the system, they used a separate audio segmenter to generate ground truth for an HMM much in the style of the wedding-video segmenter described above, and then the parameters of these HMMs were used for retrieval (2007). In a later refinement, these HMMs were the final selector after other features had narrowed a field of candidates (2008). Wryly noting that in most cases, 'Johnny can't sing', Colin Meek and William Birmingham used HMMs not for retrieval directly but to correct errors in the input to query-by-humming systems (2002). With Jonah Shifrin and Bryan Pardo, this team gradually refined this system (Shifrin et al. 2002; Shifrin & Birmingham 2003). Nonetheless,

they found that tools based on dynamic time warping proved somewhat more robust (Pardo, Shifrin & Birmingham 2003).

Cyril Joder, Slim Essid, and Gaël Richard showed that even for an apparently time-independent task, instrument recognition from audio, using HMMs to take into account time dependency can help (2009). Amaury Hazan and colleagues extended this idea somewhat further in developing an HMM for timbre grammars based on Mel-frequency cepstral coefficients (MFCCs), a popular psychoacoustic feature (2009). Independently, Emmanuel Vincent and Gautham Mysore have developed 'non-negative hidden Markov models' that undertake still more complete HMM-based model of temporal dynamics and spectral characteristics, typically oriented toward separating different instruments that have been mixed into a single stereo recording (Vincent 2006; Mysore 2010; Mysore, Smaragdis & Raj 2010). In a similar spirit, Michael Casey and Tim Crawford used HMMs to detect ornaments in Baroque lute music. The range of their $X^{(t)}$ was not a direct transcription but an automatically learned series of 40 states that, after training, would represent a reduction of the audio into a mid-level representation more reflective of the musical texture; points with more transitions between these mid-level states corresponded well to points of ornamentation.

≈

When HMMs first appeared at ISMIR, they were proposed as a general tool for audio segmentation (Batlle & Cano 2000) and achieved early successes for relatively simple problems, e.g., distinguishing those regions

of audio that contain the singing voice (Berenzweig & Ellis 2001). The challenge with using HMMs for segmentation, however, as noted earlier in section 2.4, is that by default, their models for the amount of time that a system remains in any given state is inflexible and not necessarily representative of musical reality. Suppose, for example, for some underlying outcome space Ω, the random variables $\mathbf{X}^{(t)}$ and $Y^{(t)}$ are mapped such that they correspond to the t-th measures of musical pieces represented by the outcomes $\omega \in \Omega$. Suppose that for each $\mathbf{Y}^{(t)}$, $Y_1^{(t)}$ maps each outcome to a type of high-level structural label (e.g., verse or chorus), $Y_2^{(t)}$ maps each outcome to the number of subsequent measures with the same structural label defined by $Y_1^{(t)}$, and $Y_3^{(t)}$ is a Boolean (true or false) flag marking whether $Y_1^{(t)}$ is the same as $Y_1^{(t-1)}$, i.e., false for the first t after each change of Y_1 and true otherwise. For the sake of example, suppose that $\mathbf{E}(Y_2^{(t)} \mid Y_1^{(t)}, Y_3^{(t)}) = 8$ when $Y_3^{(t)}$ is false and $Y_1^{(t)}$ is some particular label y_1; in other words, suppose that the particular structure labelled as y_1 lasts eight bars on average. Under these assumptions, the first row of table 3.3 illustrates the values of the conditional mass function $f_{Y_2^{(t)}|Y_1^{(t)},Y_3^{(t)}}$ at different values of y_2, i.e., the probability that y_1 will last y_2 bars using the default structures of an HMM. These structures imply a *geometric distribution*,

$$f_{Y_2^{(t)}|Y_1^{(t)},Y_3^{(t)}}(y_2 \mid y_1, \text{false}) = (1 - \pi_{y_1})^{y_2 - 1}\pi_{y_1}\,, \tag{3.3}$$

where π_{y_1} is the reciprocal of the average duration of label y_1 (⅛ in our example). The geometric distribution does not seem to reflect traditional understandings of how Western musical structures behave: on one hand, the distribution is weighted perhaps too much toward shorter durations, with a duration of a single measure being the most common, and the other

Table 3.3 · Using duplicated states to improve duration modelling in HMMs

STATES PER	PROBABILITY (%) OF LASTING N TIME UNITS											
TIME UNIT	1	2	3	4	5	6	7	8	9	10	11	≥ 12
1	12	11	10	8	7	6	6	5	4	4	3	23
2	4	8	9	10	10	9	8	7	6	5	4	20
4	1	3	7	10	12	12	11	10	8	7	5	14
8	0	0	3	7	12	15	16	14	11	8	5	8
16	0	0	0	3	10	18	22	19	14	8	4	3

Note: Rows may not sum to 100 due to rounding. The parameters of each distribution have been adjusted such that for each row, the expected value is exactly 8 time units.

hand, there also seems to be too much weight on long durations, with nearly a quarter of all phrases expected to be twelve measures or more. For musical segments, one would expect very small probabilities of short durations, a good peak near the expected duration, and a fall-off toward longer durations that starts sharp and becomes more gradual.

Having worked with a variety of HMM-based approaches as well as more general graphical models for music information retrieval, Raphael has identified the duration of segments to be one of the most important challenges for musical applications (Raphael 2006). Mark Levy and Mark Sandler have shown empirically that HMMs on their own model long segments quite poorly (Levy, Noland & Sandler 2007; Levy & Sandler 2008); to compensate, Samer Abdallah and colleagues have developed an elaborate Bayesian architecture over low-level HMMs in order to obtain more realistic segment durations (Abdallah et al. 2005, 2006). The discussion in section 2.4 presented hidden semi-Markov models as a more general solu-

tion to this problem, which is also the flavour of Raphael's own solution (see below). Levy and Sandler's approach uses a different strategy to improve the representation of duration within the HMM structure: by subdividing each of the possible labels in the range of the $Y_1^{(t)}$ into multiple steps (e.g., 'chorus$_1$', 'chorus$_2$', and 'chorus$_3$' rather than simply 'chorus'), it is possible to achieve a much more flexible space of distributions over duration that in particular is freer to target different levels of musical hierarchy.

The lower lines of table 3.3 illustrate the probabilities $f_{Y_2^{(t)}|Y_1^{(t)},Y_3^{(t)}}$ at different values of y_2, still with the constraint that Y_3 be false and that $\mathrm{E}(Y_2^{(t)} \mid Y_1^{(t)}, Y_3^{(t)}) = 8$ but with y_2 now represented by a sequence of multiple steps (and assuming for clarity of presentation that these steps are allowed to unfold sufficiently quickly that the complete sequence is achievable in the same amount of time as the original time step, although this change was unnecessary for Levy and Sandler's particular application). As the number of states increases, the expected distribution of durations tightens toward the expected value (8), with the probability of durations near the expected value steadily increasing and the probability of very short durations and the probability of very long distributions dropping toward the negligible. These distributions are perhaps a better match for how one expects long musical segments to unfold. As such, duplicating states is a common solution to the problem of representing duration with HMMS, and by loosening restrictions on these duplicated states, very general classes of distributions on duration are possible (Bilmes 2006) – indeed, managing this flexibility is the principle behind the success of Mavromatis's system for modelling Greek chant, described above (2006; 2009).

When HMMS are used for segmentation, they have the advantage of

being able to identify not only the segment boundaries but also the most appropriate label for each segment. Many other algorithms for segmentation return the segment boundaries only, in which case HMMs can also be useful after the fact to choose appropriate labels for the segments (Paulus & Klapuri 2009; Paulus 2010).

≈

Although segmentation was one of the first significant uses of HMMs at ISMIR, they are perhaps most dominant today in applications connected to Western tonal harmony. The most basic of these tasks is identifying key. The majority of key-finding algorithms do not use as detailed a notion of temporal dynamics as HMMs – e.g., the most famous key-finding algorithm, the Krumhansl-Schmuckler algorithm (Krumhansl 1990a, pp. 77–110), which uses only the relative duration of pitch classes – but several HMM-based alternatives have been proposed. Wei Chai and Barry Vercoe presented the first, using a hand-tuned HMM (2005). Katy Noland and Mark Sandler presented a more general system the next year that used HMMs to identify key and key changes throughout popular music (2006), as did Geoffroy Peeters (Peeters 2006).

Moving up a step in complexity of task, Christopher Raphael and Joshua Stoddard have used HMMs to analyse functional tonal harmony in MIDI files (Raphael & Stoddard 2003, 2004). In principle, the outcome space Ω is the space of all common-practise music (although due to the computational complexity of their algorithm, it was not possible to test on a corpus that would represent such a large space well), and the underlying Markov chain

advances such that the $\mathbf{X}^{(t)}$ and $\mathbf{Y}^{(t)}$ map to the t-th bar or half-bar in each piece $\omega \in \Omega$. The $\mathbf{X}^{(t)}$ map outcomes to all pitches that appear in the t-th bar or half-bar together with the beats on which they appear, and the $\mathbf{Y}^{(t)}$ take values in a space including all keys and standard harmonies (without respect to inversion). Randal Leistikow used a similar model to include harmony in a study of musical expectation in folk songs (2006), oriented toward learning musicological parameters rather than classification performance. Raphael and Stoddard did evaluate classification performance informally, describing their results as 'promising', but some earlier work from Dan Ponsford, Geraint Wiggins, and Chris Mellish on statistical models for harmony suggests that the Markov assumption may ultimately prove to be a serious limitation for any application of HMMs to harmony (1999).

Despite the potential limitations of harmonic Markov chains, the difficulty of combining harmonic analysis with audio modelling has made HMMs the standard technique for labelling chords in audio files. The earliest such applications used a hierarchical variant of the HMM, discussed below. Working with Lawrence Saul, I was one of the first researchers to apply classical HMMs to audio chord-recognition problem (Burgoyne & Saul 2005), and in the same year, Juan Bello and Jeremy Pickens presented a similar system (2005); all of us recognised the benefit of using musicological knowledge to hand-fix certain parameters, for the transition distributions $f_{\mathrm{Y}^{(t)}|\mathrm{Y}^{(t-1)}}$ in my case and the emission distributions $f_{\mathrm{X}^{(t)}|\mathrm{Y}^{(t)}}$ in the case of Bello and Pickens. In the following years, these structures became standard in the field. Benoit Catteau, Jean-Pierre Martens, and Marc Leman devised a system that incorporated musical knowledge for both transition and emission distributions while leaving a few parameters free for training (2007). Matti Ryynänen

and Anssi Klapuri used the larger Real World Computing (RWC) dataset (Goto et al. 2002, 2003) in an attempt to learn transition distributions with less need for musical knowledge (Ryynänen & Klapuri 2008). Kyogu Lee and Malcolm Slaney devised a technique for reducing the need for hand tuning by generating large amounts of training data by starting with MIDI files, using Temperley's MELISMA analyser to label their harmonies, and the generating audio from the same files to use for training; they tested this approach with standard HMM structures under variety of contexts, eventually accounting for key, genre, and specialised audio features (Lee & Slaney 2006, 2007, 2008; Lee 2008b, a), although this technique does suffer from learning the same kind of mistakes that the MELISMA analyser makes: its reported accuracy is only 85 percent.

Hélène Papadopoulos and Geoffroy Peeters explored a variety of methods for learning the parameters of HMMs for chord recognition, confirming that a combination of hand-tuning with music-theoretical knowledge and machine learning performs the best (2007); they later improved their system by using a beat detector to allow the Markov chain to evolve by beat rather than by a fixed duration of audio (2008), the first since Bello and Pickens to do so. Beat alignment became the new standard quickly. Björn Schuller and colleagues developed a simple, beat-aligned chord recogniser and compared models trained with musicological knowledge only to those learned from the data set (Schuller et al. 2009). Matthias Mauch and Simon Dixon had used the duration-modelling trick mentioned earlier, replaced each chord symbol with a three-state sequence (2008); like Papadopoulos, they eventually adapted their system by combining it with a beat detector and developing a detailed model for the conditional distributions $f_{\mathbf{Y}^{(t)}|\mathbf{Y}^{(t-1)}}$

that included bass notes, chord labels, keys, and metric position (Mauch 2010; Mauch & Dixon 2010b). Beat alignment, however, does not eliminate the problem of duration modelling, and there is room to explore this area further: one of the most recent studies on chord recognition found that correct duration models are in fact more important for good classification than correct statistics on chord changes (Pauwels & Martens 2010).

Finally, rather than recognising chords that are already written, some systems seek to generate plausible chords to accompany a monophonic melody. In the context of an HMM, this problem is almost identical to chord recognition or harmonic analysis: although the underlying probability space will include all plausible harmonisations and melodies rather than just pieces that already exist, the $\mathbf{X}^{(t)}$ and $\mathbf{Y}^{(t)}$ are still indexed by beat in most cases, and the $\mathbf{X}^{(t)}$ still take values on a space of chords (and sometimes keys). The $\mathbf{Y}^{(t)}$, however, rather than containing complete musical information, map simply to the individual notes of the melody. Randall Spangler, Rodney Goodman, and Jim Hawkins developed 'Bach in a Box', one of the first HMM-like systems for harmonisation in the style of Johann Sebastian Bach (1998); Uraquitan Cunha and Geber Ramalho developed a similar system at about the same time (1999). Moray Allan and Christopher Williams developed a system years later that uses classical HMMs more strictly (2005). Ching-Hua Chuan and Elaine Chew developed a hybrid system that combines elements of rule-based systems and elements of HMMs and can harmonise melodies in a wider variety of styles (2007). Ian Simon, Dan Morris, and Sumit Basu took this idea a step further and developed an HMM-based system that can generate an accompaniment for vocal melodies sung into a microphone (Simon, Morris & Basu 2008).

In contrast to harmony, most approaches to beat tracking and tempo tracking have relied on lower-level, often deterministic, techniques from signal processing rather than graphical models like HMMs. There are, however, several notable exceptions. Taylan Cemgil and colleagues designed a system for tracking tempo based on the Kálmán filter, a close relative of the HMM used when the $\mathbf{Y}^{(t)}$ are continuous rather than discrete – in this case, tempo as expressed in beats per minute rather than with discrete labels (Cemgil et al. 2000; Cemgil & Kappen 2003; Cemgil 2004; Kálmán 1960). Stephen Hainsworth and Malcolm Macleod presented a variation on this system using a slightly different filter (2003), and the approach is also similar to Dustin Lang and Nando de Freitas's system for beat tracking (Lang & de Freitas 2005). Christopher Raphael constructed another similar system that not only tracks tempo but also transcribes rhythms, rightly observing that these two tasks are linked (2001a). Nick Whiteley, Cemgil, and Simon Godsill later presented an even more elaborate Bayesian-style system for simultaneous tracking of tempo, rhythm, and meter that reduces to an HMM with a complex definition of transition probabilities (2006); in the same year, Anssi Klapuri, Antti Eronen, and Jaako Astola presented a more classical signal-processing model for a similar group of tasks also based on HMMs (2006).

Conceptually, these HMM-based models for tracking tempo and beat are a link to a different classical MIR task to which HMMs have been more often applied: score following. Score following attempts to align audio with a symbolic representation of the same music. Formally, the task operates

over a probability space where Ω includes all possible performances of a particular piece of music, quantised to some pre-specified degree of time precision and $\S$ would be the usual $\mathcal{P}(\Omega)$. Unlike many of the previous examples, however, $\mathbf{P}$ will almost certainly *not* assign each possible performance the same measure. Moreover, also unlike previous examples, a frequentist, physical-probability interpretation seems to make little sense for this task. The value of HMMs for score following is less in identifying the physical truth behind performances and more in making logically consistent assumptions about which tempi and what types of tempo variation are more likely than others – in other words, a Bayesian approach. The $\mathbf{X}^{(t)}$ map outcomes $\omega \in \Omega$ to audio observations at individual moments in a performance, and the $\mathbf{Y}^{(t)}$ take values over the series of relevant musical moments in the score. Because the score is known, the conditional distributions $f_{\mathbf{Y}^{(t)}|\mathbf{Y}^{(t-1)}}$ are mostly determined: they are only important as a means of modelling duration, as has been discussed extensively above.

Pedro Cano, Alex Loscos, and Jordi Bonada were the first researchers to apply HMMs to score following (1999), followed by a more elaborate HMM designed by Nicola Orio and François Déchelle (2001). Orio and Déchelle's system has been continually improved at the Institut de Recherche et Coordination Acoustique / Musique (IRCAM) ever since, although the underlying basis has remained some variant of HMM (Cont, Schwarz & Schnell 2005; Cont 2006; Montecchio & Orio 2009). To facilitate his 'Music Plus One' project, which generates accompaniments for musical soloists that respond to the soloists' tempo variations in real time, Raphael developed an independent HMM-based system to follow scores (1999; 2002b). Both of these systems face the usual challenges with modelling duration, and

later systems have attempted to sidestep the problem by defining a state for each beat or subdivision of the beat in a score rather than individual score elements (Pardo 2005; Jordanous & Smaill 2009). Score following is also very similar to the more general problem of aligning a score-like representation with an audio file; Paul Peeling, Taylan Cemgil, and Simon Godsill have used HMMs to solve this task (2007).

≈

The final area where HMMs have been used extensively in MIR is music transcription. There are two ways to consider the sample space for this problem, one of which lends itself more to a frequentist perspective and the other a Bayesian. In the more frequentist system, the sample space Ω would encompass all pieces of music in a particular corpus along with all *realised* recordings of them, the σ-algebra $\mathcal{F}$ would be the usual $\mathcal{P}(\Omega)$, and $\mathbf{P}$ would assign equal probability to all $\omega \in \Omega$. In the Bayesian perspective, the sample space Ω would encompass all pieces of music in a particular corpus with all *conceivable* recordings of performances them, the σ-algebra $\mathcal{F}$ might well again be $\mathcal{P}(\Omega)$, but because $|\Omega| \geq \aleph_0$, there could be no well-defined $\mathbf{P}$ that would assign equal probability to all $\omega \in \Omega$; some prior assumption on the form of $\mathbf{P}$ would be necessary and any learning could only be understood with respect to that assumption. Published research in this area often neglects to state which interpretation is preferable, which makes it difficult to assess how and where the results could be expected to generalise. Under either interpretation, however, the $\mathbf{X}^{(t)}$ and $\mathbf{Y}^{(t)}$ are similar to those from the score-following problem: the $\mathbf{X}^{(t)}$ correspond to

features computed from audio just as with score-following, and the $\mathbf{Y}^{(t)}$ correspond to events in a musical score, except that in this case, the score is not known in advance.

Unlike many other traditional tasks in MIR, the transcription problem lends itself to non-Western musics. Olivier Gillet and Gaël Richard used HMMs to label the types of drum strokes in Indian tabla music (2003). A couple of years later, Parag Chordia used a larger data set to compare a number of approaches and found that while on one hand, using an HMM harmed performance relative to sliding-window methods, on the other hand, the recognition rates overall were lower than those that Gillet and Richard had been able to achieve using HMMs (Chordia 2005). Gillet and Richard adapted their tabla-stroke detector for Western drumming (2004), still based on a first-order Markov chain of individual drum strokes; concurrently, Jouni Paulus and Anssi Klapuri developed a system based on Markov chains of measures ('words') with the order of the chain varying from 1 to 10 (2003).

As one might expect from the abundance of musicological studies of melody based on Markov chains, HMMs have been used extensively to transcribe melodies from audio. Kunio Kashino and Hiroshi Murase designed a system that functions as an HMM for audio melody extraction, using signal-processing techniques to generate observations from an audio signal and a first-order Markov chain based on melodic intervals and timbres to smooth these observations (1998). Adriane Durey and Mark Clements used an HMM trained for melodic transcription as a retrieval agent for returning audio files matching a symbolic melodic query (2001). Taking an explicitly Bayesian approach, Harvey Thornburg, Randal Leistikow,

and Jonathan Berger developed and described in considerable detail an HMM for melodic transcription with states that involve not only notes but also properties of the audio frame (2007). Likewise following a Bayesian approach, Emir Kapanci and Avi Pfeffer developed an HMM with an even more elaborate state structure to generate accurate musical scores from monophonic audio (2005); this model is notable for striving to transcribe rhythms correctly, whereas other systems seek only to identify the melodic notes along with their onset and offset times.

A few researchers have jumped directly to transcribing polyphony with HMMS. Raphael was the first to use HMMS for such a task, automatic transcription of piano music (2002a). Following a Bayesian approach, Taylan Cemgil built a more general system for polyphonic transcription based on the Kálmán filter (Cemgil 2004). Stanisław Raczyński and colleagues developed the most recent system, which uses HMM states that incorporate both harmony and sounding notes at each moment in time to yield a more accurate transcription (2010).

~ *Mixtures of* HMMS

In almost all cases, HMMS are used as the sole representation of the joint distribution function involved in musicological research problems. Occasionally, however, it is useful to consider *mixtures* of hidden Markov models. The premise behind a mixture model is that the observations may have been the result of one of a number of different stochastic processes, and there is uncertainty as to which of these processes produced the observation. Thus, in addition to the parameters of the HMMS themselves, mixture

models must also learn *mixture parameters*, the parameters of a probability mass function for a random variable Z that maps each outcome $\omega \in \Omega$ to the actual HMM that generated it. Yuting Qi, John William Paisley, and Lawrence Carin used a Bayesian approach to model individual pieces as mixture of hidden Markov models and to use the parameters of those models to compare pieces (Qi, Paisley & Carin 2007). Similar approaches have been used to segment vocal from non-vocal sections (Nwe & Wang 2004; Kan et al. 2008).

✵ *Hierarchical* HMMS

Mixture models allow for the possibility of different HMMs to describe an entire musical sequence. Another variant of the HMM, the *hierarchical* HMM allows a musical sequence to be described by a series of different HMMs, one following another. Figure 3.2 represents such a structure graphically. Due to limitations on space, time flows both vertically and horizontally. A higher-level Markov chain, denoted in the figure by $\mathbf{Z}^{(1:k)}$, runs from top to bottom. The values of these random variables are usually the labels of interest, and in the context of speech recognition, where hierarchical HMMs are common, the conditional mass function $f_{\mathbf{Z}^{(k)}|\mathbf{Z}^{(k-1)}}$ is commonly known as the *language model*. Each step of the $\mathbf{Z}$ Markov chain corresponds to multiple time steps in the underlying outcome ω; in fact, each possible value of the $\mathbf{Z}^{(k)}$ – say each $\psi \in \Psi$ – corresponds to an independent HMM with conditional distribution functions $f^{\psi}_{\mathbf{X}^{(t)}|\mathbf{Y}^{(t)}}$ and $f^{\psi}_{\mathbf{Y}^{(t)}|\mathbf{Y}^{(t-1)}}$. With good training data, there are special techniques to facilitate estimating $f_{\mathbf{Z}^{(k)}|\mathbf{Z}^{(k-1)}}$ and the $f^{\psi}_{\mathbf{X}^{(t)}|\mathbf{Y}^{(t)}}$ and $f^{\psi}_{\mathbf{Y}^{(t)}|\mathbf{Y}^{(t-1)}}$, but in fact hierarchical HMMs are technically

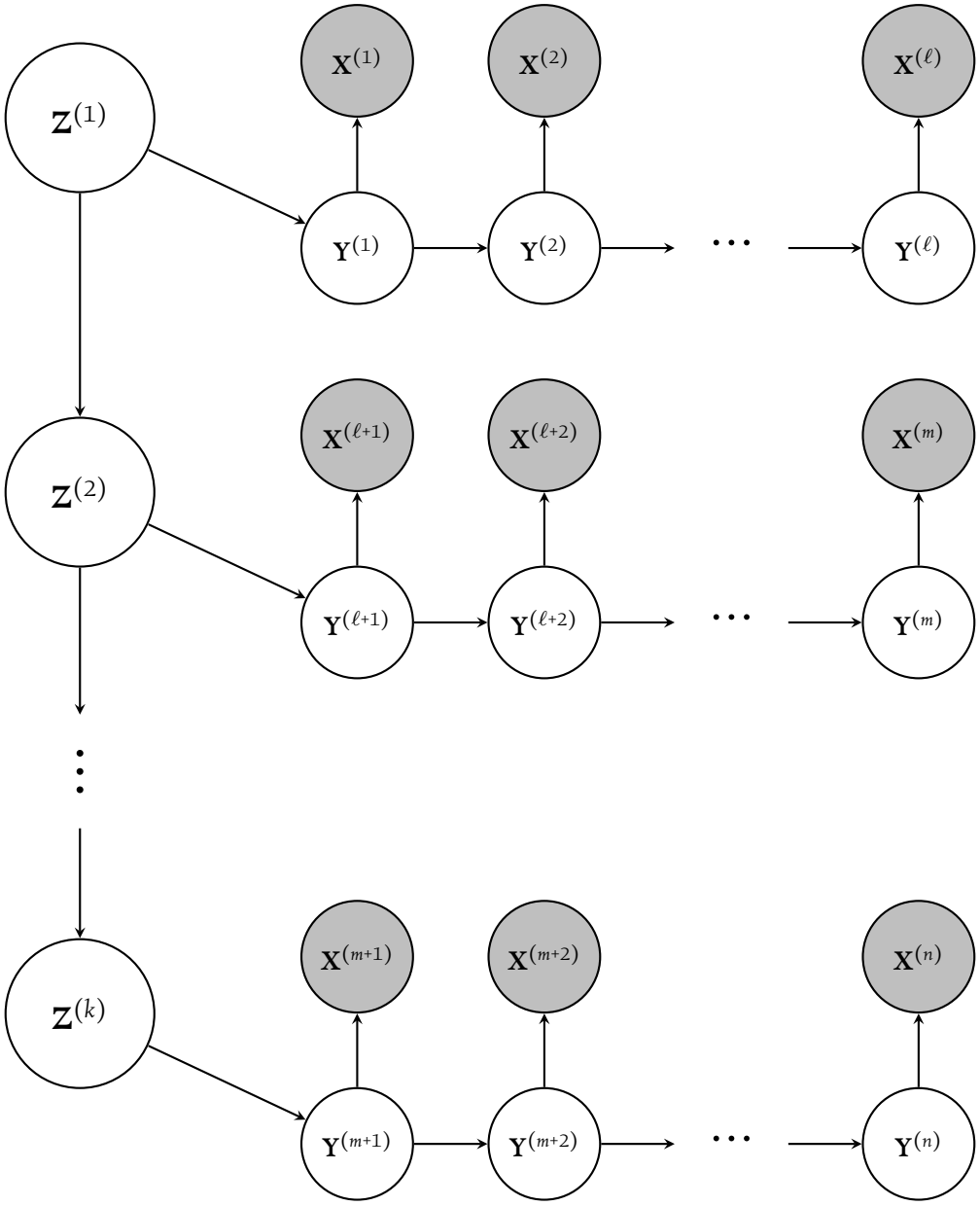


Figure 3.2 · A hierarchical HMM. A higher-level Markov chain of unobserved random variables **Z** controls a series of lower-level HMMs. It is possible to flatten this structure into a classical HMM, but with good data, there are techniques to train this hierarchical structure directly.

no different than ordinary HMMs. If one 'flattens' the $\mathbf{Z}^{(k)}$ and the $\mathbf{Y}^{(t)}$ into a combined series of random variables $\mathbf{Y}'^{(t)} = (\mathbf{Z}^{(k(t))}, \mathbf{Y}^{(t)})$, where $k(t)$ maps each time t to the corresponding value of k in the $\mathbf{Z}$ Markov chain, then one obtains a classical HMM defined by

$$f_{\mathbf{X}^{(t)}|\mathbf{Y}'^{(t)}} = f^{\mathbf{Z}^{(k(t))}}_{\mathbf{X}^{(t)}|\mathbf{Y}^{(t)}} \tag{3.4}$$

and

$$f_{\mathbf{Y}'^{(t)}|\mathbf{Y}'^{(t-1)}} = \begin{cases} f^{\mathbf{Z}^{(k(t))}}_{\mathbf{Y}^{(t)}|\mathbf{Y}^{(t-1)}} & \text{if } \mathbf{Z}^{(k(t))} = \mathbf{Z}^{(k(t-1))} \text{ and} \\ f_{\mathbf{Z}^{(k(t))}|\mathbf{Z}^{(k(t-1))}} & \text{otherwise.} \end{cases} \tag{3.5}$$

No information is lost in this conversion; the flattened model is mathematically equivalent to the hierarchical model (see Murphy 2002, pp. 28–41, for a more detailed formulation of this equivalence). Because of this equivalence and the prevalence of hierarchical models in speech recognition, many authors will refer to their models simply as HMMs even when, strictly speaking, they are hierarchical HMMs; other authors will refer to hierarchical HMMs as *model-discriminant* HMMs and classical HMMs as *path-discriminant* HMMs.

The earliest HMMs for chord recognition were hierarchical HMMs, beginning with the seminal research of Alexander Sheh and Dan Ellis on a small corpus of Beatles songs (2003). As discussed above, many later approaches to chord recognition used classical HMMs directly, but inspired by Sheh and Ellis, Namunu Maddage and colleagues have used hierarchical HMMs for chord detection as part of a larger system for semantic analysis of music (2004). Maksim Khadkevich and Maurizio Omologo used more sophisticated language modelling to improve performance under this technique (2009). Although further research is necessary, preliminary research

I conducted with Laurent Pugin, Corey Kereliuk, and Ichiro Fujinaga suggested that the techniques for training hierarchical HMMs yield superior classification results to those for classical HMMs (Burgoyne et al. 2007).

Hierarchical HMMs have also been applied to various transcription tasks. Haruto Takeda, Takuya Nishimoto, and Shigeki Sagayama used them for rhythm transcription (2004). Jouni Paulus and Anssi Klapuri improved their original drum transcriber, which was based on classical HMMs, by integrating a hierarchical model that modelled each individual drum stroke as its own HMM combined with a language model over stroke types (2007). For transcribing melodies from audio, Matti Ryynänen and Anssi Klapuri combined a first-order language model for melody with lower-level HMMs over the audio signal for each note (2006), a technique similar to that which Willie Krige, Theo Herbst, and Thomas Niesler used a couple of years later (2008). Ryynänen and Klapuri have also taken advantage of the particular algorithm used for decoding the maximally likely sequence of values of the $\mathbf{z}^{(k)}$ in hierarchical HMMs, the *token-passing algorithm* (Young et al. 2006, pp. 183–84), to use hierarchical HMMs for polyphonic transcription (Ryynänen & Klapuri 2005); because the token-passing algorithm makes it easy to 'black out' certain labels from being considered at particular points in time, by blacking out all previously identified notes and running the algorithm again, one can build a polyphonic transcription easily from an otherwise monophonic model.

Optical music recognition (OMR) had only rarely been treated with hidden Markov models of any variety before Laurent Pugin developed the Aruspix system, which features hierarchical HMMs, in 2006 (Kopec

& Chou 1996; Pugin 2006a, b).† Aruspix specialises in OMR for printed music from the Renaissance, although our group is currently extending the system to handle printed and handwritten plainchant notation as well. Working with Ichiro Fujinaga and me, Pugin developed new evaluation metrics for the system (Pugin, Burgoyne & Fujinaga 2007a) and used them to test improvements that allow users to tune the HMMs as they work in order to maximise accuracy on unseen books (Pugin et al. 2007; Pugin, Burgoyne & Fujinaga 2007b, c). This system was compared to Gamera, a successful OMR tool based on instance-based learning (Choudhury et al. 2000), and was found to be more accurate on Renaissance prints (Pugin et al. 2008). Inspired by the success of this work, Ana Rebelo, Artur Capela, and Jaime Cardoso recently compared an approach to OMR on common-practise music with HMMs to several other popular approaches to OMR (2010); hierarchical HMMs were less successful in these experiments, but the authors state that further tuning would be necessary to optimise the approach to common-practise rather than Renaissance music.

One final, rather creative, use of hierarchical HMMs in musicological research is Gabi Teodoru and Christopher Raphael's solution for pitch spelling, which models each melodic voice as an independent Markov chain dependent on a common, higher-level Markov chain of keys (2007).

Dynamic Time Warping

Dynamic time warping (DTW) is a technique for time-aligning two sequences using the well-known algorithmic paradigm of dynamic programming and

†http://aruspix.net

a distance metric known as the *Levenshtein distance* or *edit distance* (Kruskal 1983). DTW is often contrasted with HMMs as a different, computationally simpler approach to sequence alignment, but in fact, many common uses of DTW are equivalent to HMMs with particular constraints on the form of the conditional distributions $f_{\mathbf{X}^{(t)}|\mathbf{Y}^{(t)}}$ and $f_{\mathbf{Y}^{(t)}|\mathbf{Y}^{(t-1)}}$ (Juang 1984).

DTW has a long history in audio-score alignment, beginning with Roger Dannenberg's seminal paper at the 1984 International Computer Music Conference in Paris and continuing at the early ISMIR conferences (Mazzoni & Dannenberg 2001) as well as the early Joint ACM-IEEE Conferences on Digital Libraries (Hu & Dannenberg 2002). Ferréol Soulez, Xavier Rodet, and Diemo Schwarz developed an independent DTW system for audio-score alignment at IRCAM, using a customised measure of distance between notes in a score and an audio signal to find the best alignment possible (2003). As these techniques became established, Ning Hu and Roger Dannenberg explored techniques for improving DTW alignments of audio and scores (2003; 2005), and Bernhard Niedermayer developed a post-processing technique based on non-negative matrix factorisation in order to improve results (2009).

Score following is very similar to audio-score alignment, as mentioned above, but in general, DTW requires knowledge of the complete audio and the complete score, which poses challenges for score following, which needs to take place in real itme. Nonetheless, Simon Dixon developed an algorithm that can perform DTW in real time, allowing it to be used for live score following (2005), which Robert Macrae, working with Dixon, was able to improve a few years later (2010).

Meinard Müller, Frank Kurth, and Tido Röder developed a separate

DTW algorithm for audio-to-score alignment designed to reduce the extensive time and memory requirements (2004). Kurth, Müller, and colleagues developed this technology into an application called SyncPlayer that could present scores and audio to a user simultaneously (Kurth et al. 2005). In a very clever variant, Kurth, Müller and colleagues later used DTW on OMR output and audio files to synchronise score images with audio playback in SyncPlayer (Kurth et al. 2007; Fremerey et al. 2008; Fremerey, Müller & Clausen 2010). This technology was finally adapted to allow users to query databases based on selecting a few measures from a scanned score (Fremerey et al. 2009).

Database queries in general have been another popular use of DTW for MIR. Shai Shalev-Shwartz and colleagues used DTW to enable users to retrieve audio files from melodic queries (2002), and Norman Adams, Daniela Marquez, and Gregory Wakefield used it to match melodic queries to melodies in a database (2005). Bryan Pardo and Manan Sanghi explored various modifications of these techniques to make them feasible for polyphonic queries (2005). Cover-song detection can also be considered a type of polyphonic database query, and early attempts used DTW-aligned audio features that are closely associated with harmony (Ellis & Poliner 2007) and DTW on automatic chord transcriptions (Bello 2007).

Another final classical example of DTW is synchronising two performances of the same piece of music. Simon Dixon and Gerhard Widmer presented one of the earlier DTW algorithms for this purpose (2005). Meinard Müller, Henning Mattes, and Frank Kurth developed a considerably faster algorithm that achieves the same results as classical DTW (2005; 2006). Bernhard Niedermayer has worked with Widmer since to develop a vari-

ant of DTW that estimates note onsets more precisely (2010). There has also been some work aligning drum patterns in world music using DTW (Antonopoulos et al. 2007; Wright, Schloss & Tzanetakis 2008).

♆ *Semi-Markov Models*

As mentioned above, rather than using tricks like duplicating states to improve the ability of HMMs to model duration, one can also generalise the HMM slightly to the hidden semi-Markov model, which includes an explicit variable for the duration of each value of the $\mathbf{Y}^{(t)}$ (see § 2.4). Only two MIR studies to my knowledge invoke semi-Markov models explicitly: XiaoBing Liu, DeShun Yang, and XiaoOu Chen's model for classifying Chinese folk music (2008), and Arshia Cont's audio-score alignment system (2010), which uses hidden semi-Markov models as one component in a hybrid system. Although he never describes it as such, Christopher Raphael's system for aligning audio to scores is effectively a hidden semi-Markov model, however, with some specialised constraints on the conditional distributions $f_{\mathbf{Y}^{(t)}|\mathbf{Y}^{(t-1)}}$; it uses a novel tree-pruning method for computing the alignment that avoids the need to make many specific assumptions about the form of the distribution functions on state durations (2004; 2006). Raphael has also extended this technique in order to transcribe monophonic melodies into full scores (2005), like Kapanci and Pfeffer (2005).

✵ *Other Generative Graphical Models*

Section 2.4 mentioned switching linear dynamical systems briefly as another variant of the HMM. The key feature of switching linear dynamical systems

is that they allow for $f_{\mathbf{X}^{(t)}|\mathbf{Y}^{(t)}}$ and $f_{\mathbf{Y}^{(t)}|\mathbf{Y}^{(t-1)}}$ to vary over time. Switching linear dynamical systems are computationally expensive and have proved disappointing in other domains, but Taylan Cemgil, Bert Kappen, and David Barber have used them successfully for tempo tracking and polyphonic music transcription (Cemgil & Kappen 2002; Cemgil, Kappen & Barber 2003, 2006).

More general generative graphical models have also been applied to music. One of the most detailed is Christopher Raphael's Music Plus One project, mentioned earlier. The full system incorporates six different groups of random variables, comprising estimated onset times from the soloist (estimated by the HMM described earlier), tempo and rubato at each note, 'phantom' variables for accompaniment notes that do not align with solo notes, 'anchor' variables for those that do, 'sandwich' variables for rapid accompaniment notes between anchors, and of course, onset times for the accompaniment notes (2001b; 2001c). Another notable such multi-level generative model is Jean-François Paiement, Doug Eck, and Samy Bengio's system for modelling chord progressions (2005).

✳ *Neural Networks*

Neural networks are one of the classical approaches to statistical learning. Although there are parallels between their structure and that of a Bayesian network – indeed, it is even possible to formulate HMMs as neural networks (Niles & Silverman 1990) – they are normally used as deterministic approximators to complicated functions. Neural networks come in many forms and the readers is referred to a reference like Bishop (1995) for more

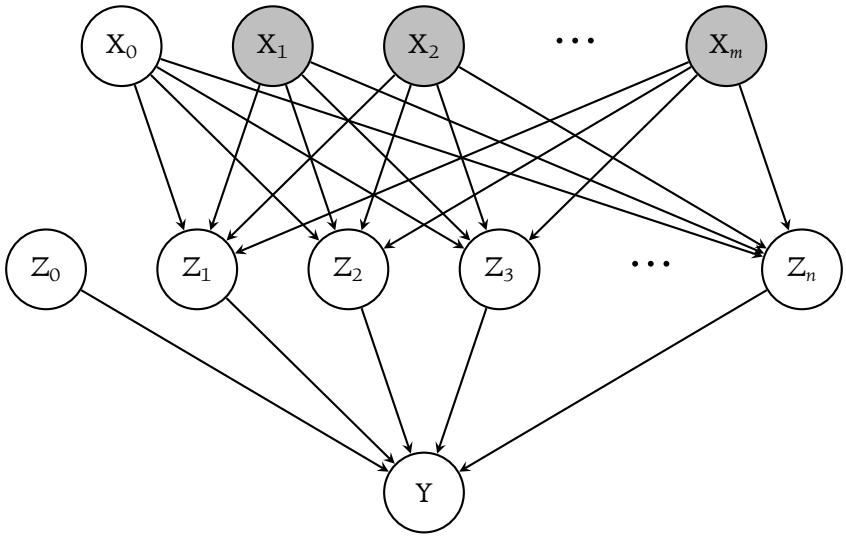


Figure 3.3 · A two-layer neural network. The top layer (**X**) represents an observed vector of features (in grey) with an extra, unobserved 'bias' term (X_0) and the bottom layer (Y) represents the desired output. The middle layer (**Z**) improves classification accuracy but is generally difficult to interpret in its own right.

detail. Their most typical form, however, illustrated in figure 3.3, does not allow for temporal dependencies. A vector of inputs $\mathbf{X}$ and an extra *bias* variable X_0 jointly determine each member a hidden layer of unobserved variables $\mathbf{Z}$, which together with a second bias variable Z_0 jointly determine the desired label Y. The values of the variables $\mathbf{Z}$ in the hidden layer are generally uninterpretable, which is one of the drawbacks of neural networks. This structure is also discriminative, not generative, and so given a label y, neural networks offer no way to predict what the inputs x to the network might have been.

With no means of modelling temporal dependency, one might think that neural networks would be confined to use in sliding-window tech-

niques. By using the outputs Y of the network for some number of previous points in time as part of the input vector $\mathbf{X}$ at the current point in time, one creates a simple kind of *recurrent neural network* that does represent temporal dependencies. Hermann Hild, Johannes Feulner, and Wolfram Menzel's HARMONET network for harmonising chorale melodies, for example, uses the previous three predicted chords as inputs, as well as information about the melody and some basic rhythmic information (Hild, Feulner & Menzel 1992). Another interesting experiment from this era showed that the values of the variables $\mathbf{Z}$ in the hidden layer of a recurrent neural network trained on Western tonal melodies statistically correlated with the first order Markov-chain transition probabilities (Stevens & Wiles 1994). Neural networks have not been as much in favour for sequence-related tasks in music in the past decade, with the one notable exception of Doug Eck and Jürgen Schmidhuber's system for improvising blues melodies, which used a type of recurrent neural network known as the *long short-term memory* recurrent network that seeks to maintain long-term structure (2002).

✳ *Other Discriminative Graphical Models*

Recall from section 2.3 that discriminative graphical models for classifying sequences seek to estimate the conditional distribution $f_{\mathbf{Y}^{(t)}|\mathbf{X}^{(t)}}$ to the exclusion of other distribution functions. Because of the amount of data required, discriminative approaches have been more popular for single-point classification problems like genre classification than they have for sequence classification. There are several intermediate cases, in fact, that are really smoothed sliding-window methods that use simple discriminative classifiers

like support vector machines (Cortes & Vapnik 1995) at each time point and then smooth them with HMMs. Graham Poliner and Dan Ellis have used such approaches for transcribing both monophonic and polyphonic music (Ellis & Poliner 2006; Poliner & Ellis 2007; Poliner 2008); similarly, for audio chord recognition, Xinglin Zhang and David Gerhard used neural networks to provide ranked lists of individual chords for each audio frame and a first-order Markov chain to smooth the results (2008).

As computational power has improved, however, there have been some attempts to perform full-fledged discriminative classification with musical sequences. Conditional random fields (CRFs), as mentioned above, are conceptually the most straightforward discriminative partner to the widely-employed, generative HMM. Victor Lavrenko and Jeremy Pickens laid the groundwork for this type of exploration by showing how random fields in general provide more flexible and musically sensible modelling possibilities than strict Markov chains for polyphonic music (2003); Pickens and Costas Iliopoulos applied this approach to polyphonic music retrieval from databases, analogously to approaches that used HMMs for the same purpose (2005). Pickens and Iliopoulos insisted that random fields promised to be useful for a number of other musical problems, a sentiment that Moray Allan and Christopher Williams echoed as they discussed future work based on their HMM for harmonising chorales (2005). Working with Laurent Pugin, Ichiro Fujinaga, and Corey Kereliuk, I was the first researcher to apply conditional random fields to audio chord recognition (2007). Our results for CRFs were disappointing relative to the best-tuned HMMs, but there were also limitations on our strategy for adapting the chord-recognition problem to CRFs that we intend to address in future

work. One other significant application of CRFs to music has appeared recently: Cyril Joder, Slim Essid, and Gaël Richard's system for aligning scores to audio (2010).

Although CRFs are the most popular discriminative models for classifying sequences in other fields, they are not the only such models. Two others of note have been applied to music. The *HMPerceptron* is an adaptation of the perceptron, a classic algorithm for training the simplest neural networks, to sequential data using a technique similar to that underlying CRFs (Collins 2002); Daniele Radicioni and Roberto Esposito have applied HMPerceptrons effectively for automatic harmonic analysis (2010). Shai Shalev-Shwartz, Joseph Keshet, Yoram Singer, and Dan Chazan used a discriminative model similar to Taskar, Guestrin, and Koller's max-margin Markov networks in order to align scores with audio, with results that significantly outperform methods based on HMMs (Shalev-Shwartz, Keshet & Singer 2004; Keshet et al. 2007; Taskar, Guestrin & Koller 2004).

3.4 SUMMARY

Despite their limitations, Markov chains have dominated almost all aspects of database-driven musicology that involve describing music as it unfolds in time. Traditional musicological corpus analysis that seeks to understand the temporal dynamics of music has sought primarily to uncover structures in melodic and harmonic sequences, often, although not always, just at the first order. Tools from MIR for building musicological databases with automated assistance also use Markov chains, most often in the form of state-space models, and strike at a broad array of problems. Many different state-space

models have been tried, but the HMM and its varieties has been strongly dominant. There is no reason to believe that the popularity of HMMs will end soon, but as the databases available for training automatic tools grow larger and computational power improves, discriminative models like CRFs and max-margin Markov networks show promise for the future.

4 THE BILLBOARD DATA SET

Having explored the mathematical foundations for database-driven musicology, how those foundations relate to understandings of the temporal dynamics of music that musicologists have reached through corpus analysis and psychological studies, and the assumptions that music information scientists have made about the temporal dynamics of music as they develop automatic classifiers for building larger musicological corpora, it is time to consider an example of how these techniques can be refined and explored with a new corpus. Under the direction of Profs. Ichiro Fujinaga and Jonathan Wild, I have led the development of a new corpus of harmonic transcriptions of popular American music from the latter half of the twentieth century, based on *Billboard* magazine's 'Hot 100' chart. It is a timely contribution to the musicological community, forming a good basis for comparison to the recently published RS 5 × 20 corpus (de Clercq & Temperley 2011) but representing a somewhat broader range of popular genres. Because these transcriptions have also been time-aligned with audio files, the corpus should also enable researchers in music information retrieval (MIR) to advance the state of the art in audio chord recognition substantially over the next few years. This chapter will motivate the need for this new corpus (§ 4.1), describe the process of constructing it (§ 4.2), present some descriptive statistics over its most salient features (§ 4.3), and describe a novel application of structure-learning techniques to establish which types of temporal relationships are most pertinent to the study of harmony in this repertory (§ 4.4).

4.1 WHY BUILD ANOTHER CORPUS OF CHORDS?

As noted in the previous chapter, corpora of harmony are much more expensive to generate than corpora of melody because at the present time, trained music theorists must undertake all of the harmonic analysis themselves. There are remarkably few corpora that can be used for any kind of computational analysis of harmony. For classical music, one of the mainstays has been the Kostka-Payne corpus, which David Temperley and Bryan Pardo released in 2001; based on the instructor's manual to a harmony textbook, the corpus contains 46 examples, all of between 8 and 20 bars, and 919 chords (Kostka & Payne 1995; Temperley 2001; Pardo & Birmingham 2002). Between 1994 and 2005, David Huron and Craig Sapp released a set of harmonic analyses of 69 of Bach's chorales in the Humdrum kern format (Huron 1995), all available as part of the KernScores database.* These two corpora have been useful for validating the results of rule-based algorithms and have borne some statistical analysis, but neither of them is of a size sufficient to support strong conclusions. A more recent addition to the field comes from Hitomi Kaneko, Daisuke Kawakami, and Shigeki Sagayama, who have recently provided harmonic analyses for all of the classical music in the RWC Music Database (Goto et al. 2002), although again, its size is limited: it contains just 50 pieces.†

The corpora of harmony in popular music have been somewhat larger. The transformative moment came in 2005, when Christopher Harte released transcriptions of the harmony of the complete recordings of the

*http://kern.ccarh.org/

†http://hil.t.u-tokyo.ac.jp/software/KSN/

Beatles (179 songs), complete with time markers corresponding to the audio (Harte et al. 2005). As discussed in the previous chapter, several research groups have used this corpus to draw some general conclusions about how to structure Markov chains for harmony, although to my knowledge, no thorough musicological analysis has yet been published. In 2009, the OMRAS2 project at Queen Mary, University of London, released another collection of audio-aligned chord transcriptions for 20 songs by Queen, 18 by Zweieck, and 14 by Carole King.‡ Excepting the Carole King transcriptions, which have been checked less thoroughly, these songs and the Beatles transcriptions have formed the basis for training and evaluating audio chord recognition systems since 2009, but at just over 200 songs, a larger corpus is still necessary in order to draw musicological conclusions about harmony and to teach computers to recognise harmony from audio, especially for chords that appear more rarely. Moreover, these transcriptions do not contain information about metric position, which somewhat restricts their use for musicological purposes. Most importantly, this corpus suffers from lack of diversity among the artists represented, which means that musicological conclusions based on this corpus may be biased and that automated systems trained on this corpus may perform poorly on broader bases of music. Trevor de Clercq and David Temperley's RS 5 × 20 corpus (2011) adds another 100 songs and that, in contrast, have been carefully selected to be well-distributed across different decades and artists and contain complete information about metric position, but these transcriptions have yet to be aligned with audio. There remains an important place for

‡http://isophonics.net/

a large, well-distributed corpus of popular harmony, complete with metric information and time markers aligning the corpus with commercially available recordings.

4.2 COLLECTING THE DATA

We sought to achieve several goals in designing a new corpus. One of our first goals was size: as noted above, even when combined, the existing corpora are not necessarily large enough to answer musicological queries with as much precision as one might like, although de Clercq and Temperley have made a very good start with the RS 5 × 20 corpus. Our second goal was diversity across artists and their relative dates of artistic activity, because again, although the RS 5 × 20 corpus made an excellent start, historically, corpora of harmony have been plagued by a lack of such diversity, especially for popular music. Our third goal was detail: Harte's transcriptions of the Beatles, in particular, include a good amount of information about advanced harmonic language, but of course restricted to the Beatles only; the RS 5 × 20 has broader coverage, but the harmonic analysis is somewhat higher-level. Akin to detail, in order to support engineering work as well as musicological work, we wanted all of the labels to be time-aligned with commercially available audio recordings. Finally, we wanted the corpus to be useful to as many researchers as possible. We chose to focus on Western popular music, which is a domain of growing interest among music theorists and has historically been dominant in research in MIR. We consulted with a statistician to ensure that given our resources, we were drawing a sample with the greatest possible capacity for both musicologists and engineers to

devise theories and systems that will generalise well to music outside the corpus. Finally, we wanted to draw music from a sufficiently wide time period that it would be possible to draw conclusions about how popular music has evolved over time.

This section describes the basis for our corpus, the *Billboard* Hot 100 chart, in more detail, explains our methodology for sampling from the chart, and outlines the process we used to transcribe the harmony in each song from our sample.

The Billboard *Hot 100*

In order to meet our requirements for diversity in popular music over a fairly wide span of time, we chose to base our corpus on *Billboard* magazine's Hot 100 chart.[§] *Billboard* has published the Hot 100 weekly since 4 August 1958. It lists and ranks the top 100 singles available for sale in the United States according to a formula that weighs the number of sales in American record stores, the amount of play time the single has received on major American radio stations, and more recently, the number of downloads of digital versions of the single. It replaced earlier charts that had appeared in *Billboard* previously, such as Best Sellers in Stores, Most Played by Jockeys, and Most Played in Jukeboxes; it was and is intended to be an industry-oriented, overall representation of the most popular music in the United States over a given week, and it is generally regarded to be the best available compilation for this purpose (Bradlow & Fader 2001). Several previous studies, in fact, have used the *Billboard* Hot 100 as the basis for statistical

[§] http://www.billboard.com/charts/hot-100

analyses of the popularity of individual singles and as a case study for developing statistical techniques for analysing regularly published rankings in general (Bradlow & Fader 2001; Giles 2007; Bhattacharjee et al. 2008).

What exactly does it mean, however, to sample from the space of what is popular? In mathematical terms, what constitutes the probability space $(\Omega, \S, \mathbf{P})$? We wanted each outcome $\omega \in \Omega$ to represent an experience whereby a particular person in the United States listened to a particular song at a particular time in a particular week; the natural σ-algebra $\S$ is the power set $\mathfrak{P}(\Omega)$ of all such experiences. That much is perhaps unobjectionable, but imagining an appropriate measure $\mathbf{P}$ is harder. Imagine a random variable X from Ω to the set of all days in our time period of study; we wanted f_X to be the same for every week, i.e., we wanted f_X to follow a *discrete uniform distribution*. Moreover, for all singleton sets of outcomes in the pre-images X^{-1} of each day, we wanted $\mathbf{P}$ to assign equal probability, i.e., we wanted all listening experiences in the United States in any given week within our time period of study to have equal probability. Because the outcomes ω are listening experiences rather than songs, the more popular a song – i.e., the more times more people listened to it – the greater its expected frequency in a random sample from this space.

It is not possible to synthesise this space of listening directly, however, and using the *Billboard* Hot 100 as a proxy for it has its drawbacks. The most obvious is that only the 100 most popular tracks appear each week, and so no music other than the very most popular songs will ever appear in a sample based on the Hot 100. A more subtle problem is that the Hot 100 does not provide any information about the relative popularity of individual songs other than their ranks; one knows that the single at

number 1 is more popular than the single at number 100, but there is no way to know how much more popular. Moreover, nobody could argue that the Hot 100 is a perfect representation even of what it purports to measure: the 100 most popular singles in the United States each week. Although the formula for computing the Hot 100 has always included the amount of playtime on radio stations and the volume of record sales, the details of this formula are obscure and have changed repeatedly over time. Furthermore, because of the importance of the Hot 100 to the industry, record labels have sought to manipulate the chart to their advantage. Most notoriously, due to the emphasis on playtime from radio stations, record labels would bribe stations to play particular songs in preference to others, an illegal practise known as *payola*; despite a serious federal crackdown in the early 1960s, payola and related manipulations have never been eliminated entirely (Coase 1979). As the corpus neared completion we received a comment from a scholar close to the industry that the *Billboard* 200, which tracks albums rather than singles, was somewhat less corrupted by payola and might have been a better basis for our sample; for future researchers interested in extending or complementing our corpus, this may well be good advice, as it would also complement the RS 5 × 20 corpus well.

Important as it may be to acknowledge and understand the limitations of any sample space and sampling methodology, we maintain that despite these limitations, a corpus based on the Hot 100 meets the needs of the research community. Using only 100 singles from each week does bias the sample toward the most popular songs relative to an ideal sample that could reach any song to which anybody in the United States listened, but

this bias seems relatively harmless for the purposes of drawing conclusions about popular music as a genre, which is by definition and name, popular. We considered using some kind of weighting to discount singles toward the bottom of the charts to compensate for the lack of information about relative popularity, but because there was no straightforward means to devise such a weighting and because the nature of the charts themselves already strongly emphasises the most popular singles (see below), we decided simply to weight each position on the chart equally. The recording industry's attempts to manipulate the chart are more problematic, but there can nonetheless be no doubt that, regardless of how one may judge the causal factors responsible for the amount time on the air devoted to the singles on the *Billboard* Hot 100, these songs were all played extensively and would have been very familiar to American audiences during their tenure on the chart. Excepting a few very similar charts like the *Billboard* 200, no other sources can claim even to try to represent historical listening habits of people in the United States.

❧ *Sampling the Charts*

Beyond the theoretical issues with using the Hot 100 as a proxy for our ideal probability space for popular music, we also needed to address a number of practical issues related to artefacts of the formula for computing the Hot 100, the limits on our resources, and ensuring that the sample would be usable even in the worst of all possible random draws.

The first issue involved choosing the start and end dates for music in the corpus. The start date was fairly easy to choose: 4 August 1958, or

the date of the first ever Hot 100 chart. The end date was a more difficult decision. We suspected that with the rise in the popularity of rap and hip-hop in the 1990s and 2000s, there would be a higher proportion of songs in a random sample for which harmonic analysis would be inappropriate. *Billboard* also made a particularly substantial change to the methodology for computing the Hot 100 in 1991. Prior to this change, information about the sales volume of singles had come from manual surveys; starting in December 1991, *Billboard* began using automatically collected data from Nielsen SoundScan.[‖] With the SoundScan data, singles began remaining on the Hot 100 for so much longer than before that *Billboard* eventually established limits on how long any one single could remain on the Hot 100 and added a new chart, the *Billboard* Recurrent Singles, for songs that were struck from the Hot 100 because of the new rule. Unlike the other changes in methodology over the years, *Billboard* considers this change to be so significant that they themselves attempt to correct for it when generating historical summaries like the 50th-anniversary charts (Billboard Magazine 2008). Because we had been considering setting an end date in the 1990s for the other reasons mentioned, we decided to draw sample for our corpus only through until the end of November 1991. To each of the 100 slots on each of the weekly charts from August 1958 through November 1991, we assigned equal probability for our sampling procedure.

Like any research team, we had to work within the limits of the materials and people we had available for this project as well as the amount of money we had to make new acquisitions. We knew that it would likely not be

[‖] http://nielsen.com/us/en/industries/media-entertainment.html

possible to obtain every single in a completely random sample, especially for older and relatively less popular music, and so before taking the sample, we consulted with several statisticians to develop a strategy for mitigating any further bias that these missing singles might cause. Our solution was to make a further simplifying assumption: that singles at adjacent ranks on a given chart are interchangeable for the purposes of our corpus: e.g., the difference in popularity between the singles ranked 27 and 28 on the chart in a particular week is slight enough that it should be allowable to use one single in place of the other. When unable to obtain the single at some target slot on a chart, we tried to obtain either the single in the slot directly above the target single on the same week's chart or the single in the slot directly below. If neither of those singles were available, then we tried the slots two above and two below the target slot. Only when all five singles were unavailable did we list the slot as unobtainable.

As stated above, one of the goals in generating the corpus was to be able to study how harmonic patterns in popular music evolved throughout the late twentieth century. With an unguarded random sample, however, there would have been a risk that every chart slot in the sample would have come from a short window of time rather than being distributed across the entire range of weeks; such a sample is not probable, but it is possible. In order to protect against this possibility, again in consultation with a statistician, we divided our date range into three chunks, 1958 to 1970, 1970 to 1979, and 1980 to 1991, and sampled from each chunk separately. In fact, in order to ensure that it would also be possible to test hypotheses about what might make some songs more popular than others, we sub-divided the chart slots in each of our existing three chunks into five further chunks, 1 to 20, 21

to 40, 41 to 60, 61 to 80, and 81 to 100, to yield a total of fifteen chunks overall, each sampled separately.

Figure 4.1 summarises the sampling algorithm from start to finish. After using *Billboard*'s web service to download the Hot 100 charts for the time period in question into a PostgreSQL database,[¶] we wrote a short Python script implementing the algorithm and creating a new table in the database with the results of the sample. We took a larger sample than necessary to allow for the possibility of future work, but for the purposes of our current experiments and this thesis, we used a sample of 2 000 songs total after all chunks had been recombined.

As illustrated in table 4.1, of the sample of 2 000 slots, we were able to obtain one of the five singles allowed by the sampling algorithm in 1 400 cases (70 percent): 501 target singles, 575 singles that were one slot above or below the target on the chart, and 324 singles that were two slots above or below the target on the chart. The table also shows that it was somewhat easier than average for us to acquire digital audio for songs from the 1970s and somewhat more difficult than average for us to locate digital audio from the 1950s and 1960s. More importantly, there is a clear trend in favour of finding songs near the top of the charts and against finding songs near the bottom of the charts, which means that our corpus is somewhat biased toward more popular songs.

Figure 4.2 confirms this bias. Part (a) illustrates the distribution of peak ranks (the highest rank a single ever achieves on the Hot 100) corresponding to all published chart slots during the time period covered by our corpus.

[¶] http://developer.billboard.com/

1. Divide the set of all chart slots into three eras:
    a) 4 August 1958 to 31 December 1969,
    b) 1 January 1970 to 31 December 1979, and
    c) 1 January 1980 to 30 November 1991.
2. Subdivide the chart slots in each era into five subgroups corresponding to quintiles on the chart:
    a) ranks 1 to 20,
    b) ranks 21 to 40,
    c) ranks 41 to 60,
    d) ranks 61 to 80, and
    e) ranks 81 to 100.
3. Select some fixed percentage of possible chart slots at random from each era-quintile pair.
4. For each selected chart slot:
    a) attempt to acquire the single at the target slot;
    b) if that fails, toss a virtual coin to choose between either the single directly above or directly below the target slot on the chart from the same week;
    c) if that fails, choose the single that was not selected by the coin toss in 4b;
    d) if that fails, toss a virtual coin to choose between either the single two ranks above or two ranks below the target single on the chart from the same week;
    e) if that fails, choose the single that was not selected by the coin flip in 4d; and
    f) if that fails, consider the chart position to be a missing data point.

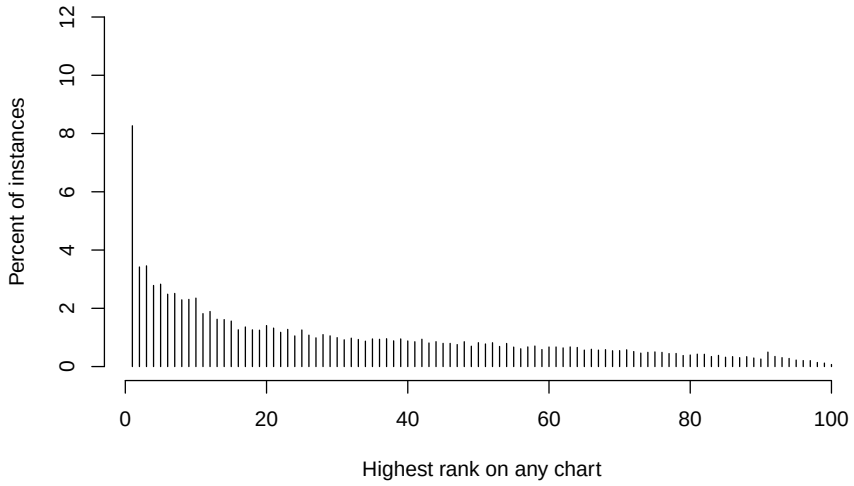
Figure 4.1 · Sampling algorithm for the *Billboard* Hot 100. The algorithm protects against sharply skewed random draws and, in an attempt to conserve resources in the case of music that may be expensive to obtain, assumes that the singles in adjacent chart slots are exchangeable.

Table 4.1 · Retrieval rates for audio in the *Billboard* sample

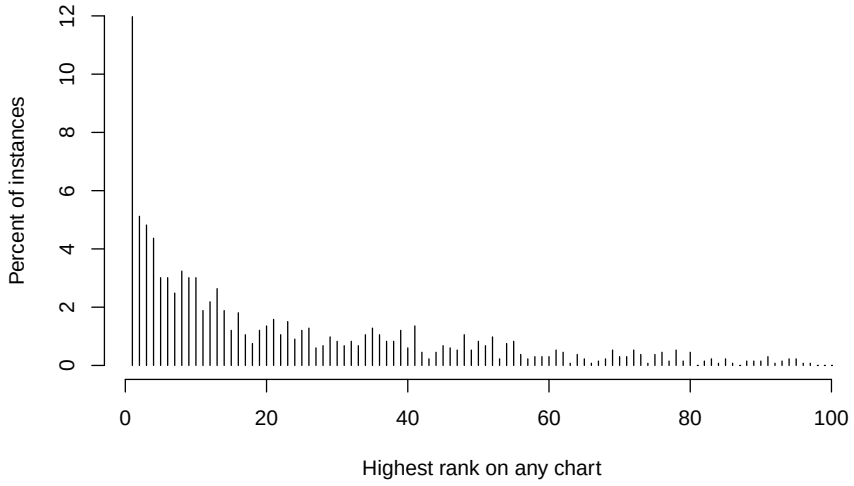
RANK	'50s/'60s	'70s	'80s/'90s	ALL
1–20	$\frac{101}{129} \approx 78\,\%$	$\frac{119}{134} \approx 89\,\%$	$\frac{104}{121} \approx 86\,\%$	$\frac{324}{384} \approx 84\,\%$
21–40	$\frac{101}{132} \approx 77\,\%$	$\frac{122}{135} \approx 90\,\%$	$\frac{103}{135} \approx 76\,\%$	$\frac{326}{402} \approx 81\,\%$
41–60	$\frac{88}{133} \approx 66\,\%$	$\frac{102}{133} \approx 77\,\%$	$\frac{88}{130} \approx 68\,\%$	$\frac{278}{396} \approx 70\,\%$
61–80	$\frac{67}{147} \approx 46\,\%$	$\frac{92}{131} \approx 70\,\%$	$\frac{85}{140} \approx 61\,\%$	$\frac{244}{418} \approx 58\,\%$
81–100	$\frac{64}{127} \approx 50\,\%$	$\frac{76}{127} \approx 60\,\%$	$\frac{88}{146} \approx 60\,\%$	$\frac{228}{400} \approx 57\,\%$
ALL	$\frac{421}{668} \approx 63\,\%$	$\frac{511}{660} \approx 83\,\%$	$\frac{468}{672} \approx 70\,\%$	$\frac{1400}{2000} \approx 70\,\%$

The distribution is not flat because the more popular the song, the longer it is likely to remain on the charts, rising and falling through different ranks. (The average number of weeks on the chart is ten, but for a top hit, it can be considerably longer.) Part (b) illustrates the analogous distribution of peak ranks over only the elements of our corpus. If the sampling algorithm were perfect, one would expect the shape to match the shape of part (a); in fact, it is slightly skewed toward the left, i.e., songs that have been more popular, due to the difficulty of locating audio for less popular singles. This bias reinforces our belief that it was better to weight all slots equally for sampling rather than adding extra assumptions penalising songs at the bottom of the charts to reflect an estimate of their relative popularity; ultimately, our inability to locate some of the less popular songs penalised them naturally.

Because of the nature of our sampling algorithm (and also our ideal probability space), it is possible for the same single to appear more than once in the sample. Of the 1 400 slots for which we were able to acquire a recording, there were only 1 084 unique singles: 218 singles appear twice,



(a) Complete *Billboard* Hot 100



(b) Our corpus

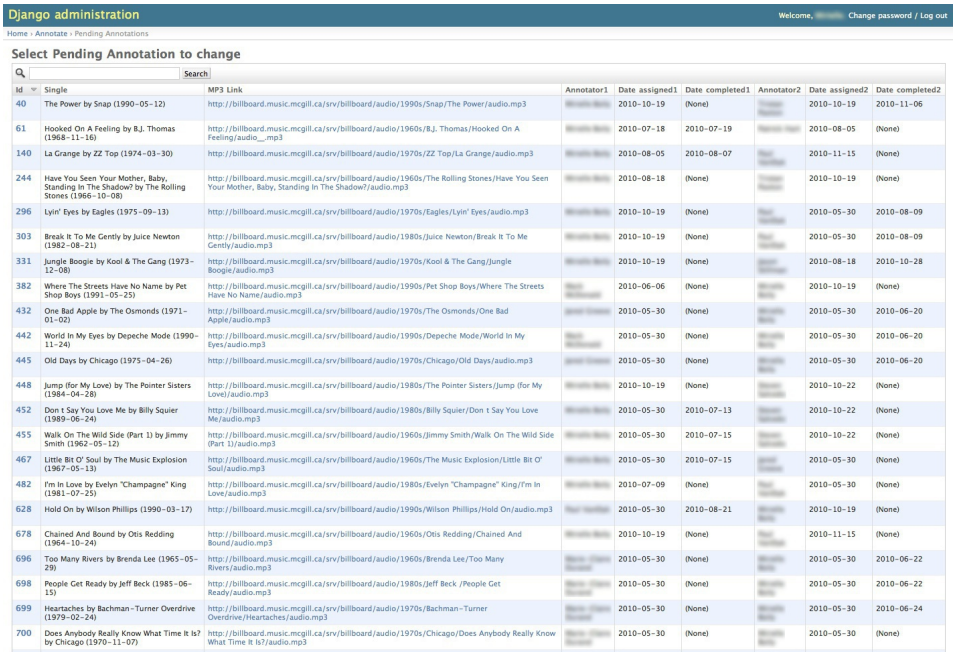
Figure 4.2 · Distribution of peak ranks in the *Billboard* Hot 100 over all published slots between August 1958 and November 1991 inclusive and over the slots in the corpus. Consistent with the pattern in songs for which we were able to acquire audio, the ranks in the corpus are somewhat biased toward more popular songs.

40 three times, and 6 four times; we weight these singles accordingly when using the corpus so as to obtain the most accurate statistics. Among these 1 084 unique singles, 529 unique artists are represented.

✳ *Transcribing the Sample*

After we had acquired as many audio files for our sample as we could find, we needed to transcribe the harmony for each song and time-align these harmonies with the audio. This process was surprisingly involved, stretching well over a year.

The first step was to develop a file format for transcriptions that would be readable and writable for both humans and machines. This format is detailed in appendix A. We also designed a web application using the Django toolkit as a front end and the same PostgreSQL database as a back end. The web application had three main components: a primary interface for accessing assigned songs, an upload page for each song, and a payroll system. The primary interface, illustrated in figure 4.3, provided annotators with a list of all songs to which they had been assigned, provided information about other annotators assigned to the same song in case of any questions, and allowed annotators to listen to their assigned songs. Once a transcription was completed, an annotator could upload the transcription directly to our database, using a page like that illustrated in figure 4.4; on this page, we also asked annotators to let us know how long they spent transcribing the song. The payroll system used the database to compute the appropriate weekly wage for all annotators and allowed them to print time sheets for the university payroll system. We also developed a variety of



Id	Single	MP3 Link	Annotator1	Date assigned1	Date completed1	Annotator2	Date assigned2	Date completed2
40	The Power by Snap (1990-05-12)	http://billboard.music.mcgill.ca/srv/billboard/audio/1990s/Snap/The Power/audio.mp3	[redacted]	2010-10-19	(None)	[redacted]	2010-10-19	2010-11-06
61	Hooked On A Feeling by B.J. Thomas (1968-11-16)	http://billboard.music.mcgill.ca/srv/billboard/audio/1960s/B.J. Thomas/Hooked On A Feeling/audio__.mp3	[redacted]	2010-07-18	2010-07-19	[redacted]	2010-08-05	(None)
140	La Grange by ZZ Top (1974-03-30)	http://billboard.music.mcgill.ca/srv/billboard/audio/1970s/ZZ Top/La Grange/audio.mp3	[redacted]	2010-08-05	2010-08-07	[redacted]	2010-11-15	(None)
244	Have You Seen Your Mother, Baby, Standing In The Shadow? by The Rolling Stones (1966-10-08)	http://billboard.music.mcgill.ca/srv/billboard/audio/1960s/The Rolling Stones/Have You Seen Your Mother, Baby, Standing In The Shadow?/audio.mp3	[redacted]	2010-08-18	(None)	[redacted]	2010-10-19	(None)
296	Lyin' Eyes by Eagles (1975-09-13)	http://billboard.music.mcgill.ca/srv/billboard/audio/1970s/Eagles/Lyin' Eyes/audio.mp3	[redacted]	2010-10-19	(None)	[redacted]	2010-05-30	2010-08-09
303	Break It To Me Gently by Juice Newton (1982-08-21)	http://billboard.music.mcgill.ca/srv/billboard/audio/1980s/Juice Newton/Break It To Me Gently/audio.mp3	[redacted]	2010-10-19	(None)	[redacted]	2010-05-30	2010-08-09
331	Jungle Boogie by Kool & The Gang (1973-12-08)	http://billboard.music.mcgill.ca/srv/billboard/audio/1970s/Kool & The Gang/Jungle Boogie/audio.mp3	[redacted]	2010-10-19	(None)	[redacted]	2010-08-18	2010-10-28
382	Where The Streets Have No Name by Pet Shop Boys (1991-05-25)	http://billboard.music.mcgill.ca/srv/billboard/audio/1990s/Pet Shop Boys/Where The Streets Have No Name/audio.mp3	[redacted]	2010-06-06	(None)	[redacted]	2010-10-19	(None)
432	One Bad Apple by The Osmonds (1971-01-02)	http://billboard.music.mcgill.ca/srv/billboard/audio/1970s/The Osmonds/One Bad Apple/audio.mp3	[redacted]	2010-05-30	(None)	[redacted]	2010-05-30	2010-06-20
442	World In My Eyes by Depeche Mode (1990-11-24)	http://billboard.music.mcgill.ca/srv/billboard/audio/1990s/Depeche Mode/World In My Eyes/audio.mp3	[redacted]	2010-05-30	(None)	[redacted]	2010-05-30	2010-06-20
445	Old Days by Chicago (1975-04-26)	http://billboard.music.mcgill.ca/srv/billboard/audio/1970s/Chicago/Old Days/audio.mp3	[redacted]	2010-05-30	(None)	[redacted]	2010-05-30	2010-06-20
448	Jump (for My Love) by The Pointer Sisters (1984-04-28)	http://billboard.music.mcgill.ca/srv/billboard/audio/1980s/The Pointer Sisters/Jump (for My Love)/audio.mp3	[redacted]	2010-10-19	(None)	[redacted]	2010-10-22	(None)
452	Don t Say You Love Me by Billy Squier (1989-06-24)	http://billboard.music.mcgill.ca/srv/billboard/audio/1980s/Billy Squier/Don t Say You Love Me/audio.mp3	[redacted]	2010-05-30	2010-07-13	[redacted]	2010-10-22	(None)
455	Walk On The Wild Side (Part 1) by Jimmy Smith (1962-05-12)	http://billboard.music.mcgill.ca/srv/billboard/audio/1960s/Jimmy Smith/Walk On The Wild Side (Part 1)/audio.mp3	[redacted]	2010-05-30	2010-07-15	[redacted]	2010-10-22	(None)
467	Little Bit O' Soul by The Music Explosion (1967-05-13)	http://billboard.music.mcgill.ca/srv/billboard/audio/1960s/The Music Explosion/Little Bit O' Soul/audio.mp3	[redacted]	2010-05-30	2010-07-15	[redacted]	2010-05-30	(None)
482	I'm In Love by Evelyn "Champagne" King (1981-07-25)	http://billboard.music.mcgill.ca/srv/billboard/audio/1980s/Evelyn "Champagne" King/I'm In Love/audio.mp3	[redacted]	2010-07-09	(None)	[redacted]	2010-05-30	(None)
628	Hold On by Wilson Phillips (1990-03-17)	http://billboard.music.mcgill.ca/srv/billboard/audio/1990s/Wilson Phillips/Hold On/audio.mp3	[redacted]	2010-05-30	2010-08-21	[redacted]	2010-10-19	(None)
678	Chained And Bound by Otis Redding (1964-10-24)	http://billboard.music.mcgill.ca/srv/billboard/audio/1960s/Otis Redding/Chained And Bound/audio.mp3	[redacted]	2010-10-19	(None)	[redacted]	2010-11-15	(None)
696	Too Many Rivers by Brenda Lee (1965-05-29)	http://billboard.music.mcgill.ca/srv/billboard/audio/1960s/Brenda Lee/Too Many Rivers/audio.mp3	[redacted]	2010-05-30	(None)	[redacted]	2010-05-30	2010-06-22
698	People Get Ready by Jeff Beck (1985-06-15)	http://billboard.music.mcgill.ca/srv/billboard/audio/1980s/Jeff Beck /People Get Ready/audio.mp3	[redacted]	2010-05-30	(None)	[redacted]	2010-05-30	2010-06-22
699	Heartaches by Bachman-Turner Overdrive (1979-02-24)	http://billboard.music.mcgill.ca/srv/billboard/audio/1970s/Bachman-Turner Overdrive/Heartaches/audio.mp3	[redacted]	2010-05-30	(None)	[redacted]	2010-05-30	2010-06-24
700	Does Anybody Really Know What Time It Is? by Chicago (1970-11-07)	http://billboard.music.mcgill.ca/srv/billboard/audio/1970s/Chicago/Does Anybody Really Know What Time It Is?/audio.mp3	[redacted]	2010-05-30	(None)	[redacted]	2010-05-30	(None)

Figure 4.3 · Screenshot of the primary web application for annotators. From this page, annotators could see a list of all of the work they had remaining, see who their partner for each song was in case of questions, and access the audio for files to which they had been assigned.

utility pages to help manage the project.

In April 2010, as we were finishing development of these technological tools, we contacted a group of seven graduate students in jazz performance who had been recommended to us by the head of our jazz department; many of these students perform popular music professionally in the Montréal area and throughout Canada. Four of these students were able to attend a training session during which they learned how to use the file format used for the project and annotated several test songs: 'You've Got a Friend', by Roberta Flack; 'An Innocent Man', by Billy Joel; 'I Don't Mind', by James Brown; 'Uneasy Rider', by the Charlie Daniels Band; and 'Hot Stuff', by the

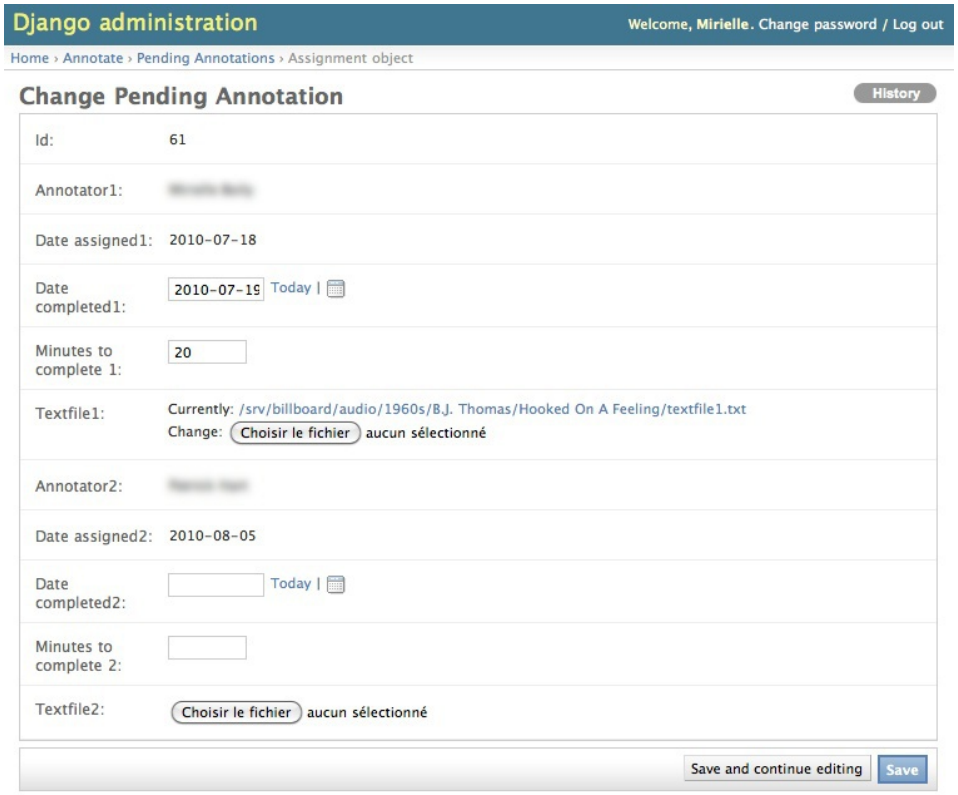


Figure 4.4 · Screenshot of the upload page for annotators. From this page, annotators could upload their transcription files to the main database and note the amount of time they had spent transcribing.

Rolling Stones. This original group of students recommended two more potential annotators, whom we asked to transcribe the same songs. After examining the curricula vitæ we received for a related project, we invited three other students to audition for the *Billboard* project as well, again with the same songs used at the original training session.

Prof. Wild reviewed all of the transcriptions from the auditions. The transcriptions of one of the annotators who had not been able to attend

the original training were too far from the other annotators' to continue, and this annotator was dismissed from the project. The remaining eight were all invited to a second training session in early May; seven were able to attend and discussed as a group the best solutions for all of the audition songs and how to handle various subtle musical cases in general. In June, we auditioned a further two annotators and accepted one of them, bringing our group to nine. Of these nine, however, one separated from the project shortly after beginning work, one had to be asked to leave due to low-quality annotations, and only two ever produced more than 40 annotations. In August, we auditioned six new annotators and hired four. Even among this group, however, one had to separate from the project shortly after beginning and only one ever produced a significant number of annotations. Between October and December, we sent ten more sets of audition materials, yielding three new annotators.

Clearly it was difficult to keep annotators working on the project. Surprisingly, money did not seem to be as large a motivator as we expected. The task requires a lot of concentration, and anecdotally, some annotators suggested that five or six songs a day was as many as they could handle, regardless of difficulty or rate of pay. The annotators had spent between 10 and 45 minutes per song to produce transcriptions for the audition materials. There was a lot of variation between songs, however, and the audition songs had been selected deliberately to be more difficult than the typical song from the charts. After some preliminary testing with another music student, it seemed reasonable to assume that annotators would take 15 minutes to transcribe songs on average with the training over. We set the original rate of pay at $4 per song, which would work out to $16 per

hour at first, and expected that as annotators became more experienced, the best annotators would reach 10 minutes per song and be earning $24 per hour. In July, responding to indirect complaints, we increased the rate to $5 per song for any week where the annotators did 30 songs or more, hoping that this would motivate annotators to work more. In September, however, the rate of annotation had again slowed to a crawl. We increased the rate again to $5 per song regardless of how many were done in a week. We also became more generous with 'bonus songs' added to the time sheets to ensure that everybody working earned at least $20 per hour, which seemed to be acceptable to all involved through until the end of the project.

Overall, the majority of songs took between 8 and 18 minutes to transcribe, with a median transcribing time of 12 minutes (and thus a median wage of $25 per hour). The most difficult songs, however, could take more than an hour. Fitting a series of log-linear models on the transcription data (Poisson regression, see McCullagh & Nelder 1989) suggests that, as expected, the song, annotator, and level of experience are all significant variables in predicting transcribing time. It also seems that there are significant differences in the effect of experience on each annotator: many improved their speed over time, but not all; on average, annotators improved their speed by a bit less than 10 percent for every 100 songs they transcribed. Figure 4.5 presents a box-and-whisker plot of the transcribing times for each annotator with the width of each bar proportional to the square root of the number of songs transcribed. There is variation with respect to the number of songs transcribed – three annotators carried the bulk of the load – and with respect to overall rate of work, but the overall pattern of most songs requiring less then 20 minutes with some exceptionally difficult

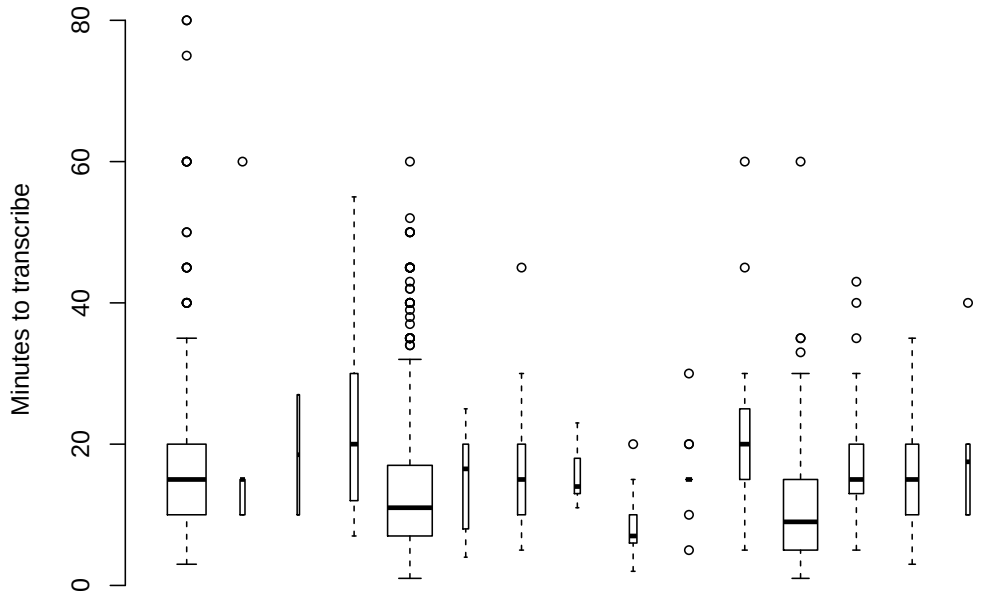


Figure 4.5 · Transcribing times for the corpus, separated by annotator. The width of each bar is proportional to the square root of the number of songs transcribed. Songs more than 1½ times the inter-quartile range away from the nearest quartile are marked as outliers. Although there is some variation among annotators, the overall statistic that most songs took between 8 and 18 minutes is clear, with a few exceptionally difficult songs as outliers.

outliers taking longer is clear.

In order to ensure accuracy, we asked two annotators to transcribe each song independently, although through the web application annotators were aware of their partners for each song and were free to communicate about their work if desired. A third meta-annotator with special training reconciled the differences between the two files. Chord transcriptions inevitably require questions of judgement and taste, and so in addition to correcting any actual mistakes, the meta-annotators needed to synthesise any difference in style between the two annotations, generally in favour of

the more detailed transcription. Finally, in partnership with the University of Southampton in the United Kingdom, we had a fourth annotator mark each transcription with key structural features (see Smith et al. 2011) and align the beginning of each phrase with audio. Considering the salaries and wages of all involved, it cost more than $20 each to arrive at this final, time-aligned file, but given the richness of the data, we believe that they have been worth the cost.

4.3 BASIC STATISTICS

One disadvantage of such a large corpus is that it becomes somewhat impractical to reflect on each detail of the corpus individually. It includes over 500 000 beats, which strains the limits of human ability to find patterns from inspection alone. Like De Clercq and Temperley's introduction to the RS 5 × 20 corpus (2011), this section presents summary statistics that help to understand the corpus and some of the basic properties of harmony in late-twentieth-century popular music.

Because it is the only other corpus that is comparable in scope, where possible, this section will also compare the summary statistics from the *Billboard* corpus with statistics from De Clercq and Temperley's corpus. They drew their corpus from *Rolling Stone* magazine's '500 Greatest Songs of All Time'. Unlike the *Billboard* Hot 100, which despite the attempted manipulations, sought to present the music to which the public was actually listening, the *Rolling Stone*'s compilation is a consensus of industry elites based more on perceived historical importance or quality. Thus, our more populist corpus and De Clercq and Temperley's corpus reflect two different

yet overlapping understandings of what constitutes popular music, and it is worthwhile to investigate both where they differ and were they do not. The only published results on De Clercq and Temperley's corpus are from the RS 5 × 20, which contained the 20 top-ranked songs from each of the 1950s, 1960s, 1970s, 1980s, and 1990s (de Clercq & Temperley 2011; Temperley 2011). Since the publication of these articles, they have released a larger corpus of 200 transcriptions in total, supplementing the 99 RS 5 × 20 songs (De Clercq and Temperley excluded one song from the 1990s due to insufficient harmonic content) with the 101 highest-ranked songs from 'Greatest Songs of All Time' that had not appeared in the original corpus.** This larger corpus is the basis for comparison throughout the section; I will denote it the *Rolling Stone* corpus and ours the *Billboard* corpus. The two corpora have 32 songs in common, which may form the basis for further study (see table 4.2).

→ *Multinomial and Dirichlet-Multinomial Distributions*

The first statistic that one might think to compute from a corpus of harmony is the distribution of the roots of all chords in terms of pitch classes. This concept seems simple, but that simplicity is deceptive: what kind of distribution is one seeking, exactly? For the populist *Billboard* corpus, the most plausible goal seems to be the relative frequencies of chord roots heard by an average listener over the period in question, i.e., the relative frequencies of chord roots weighted by popularity; for the *Rolling Stone* corpus, the most plausible goal would be the relative frequencies of chord

** http://theory.esm.rochester.edu/rock_corpus/

Table 4.2 · Songs common to both corpora

ARTIST	TITLE	YEAR
ABBA	Dancing Queen	1976
The B-52's	Rock Lobster	1980
The Beach Boys	God Only Knows	1966
The Beatles	A Hard Day's Night	1964
The Beatles	Help!	1965
The Beatles	I Saw Her Standing There	1964
David Bowie	Changes	1972
James Brown	I Got You (I Feel Good)	1965
The Byrds	Eight Miles High	1966
Johnny Cash	Ring of Fire	1963
Tracy Chapman	Fast Car	1988
Ray Charles	Georgia on My Mind	1960
Eric Clapton	Layla	1971
Patsy Cline	Crazy	1961
Cream	Sunshine of Your Love	1968
The Jacksons	I Want You Back	1969
Elton John	Your Song	1970
The Kingsmen	Louie Louie	1963
John Lennon	Imagine	1971
Otis Redding	(Sittin' on) the Dock of the Bay	1968
Otis Redding	I've Been Loving You Too Long (to Stop Now)	1965
The Rolling Stones	Honky Tonk Women	1969
The Ronettes	Be My Baby	1963
The Shirelles	Will You Love Me Tomorrow	1960
Simon & Garfunkel	The Sounds of Silence	1965
Sly & the Family Stone	Everyday People	1968
Bruce Springsteen	Born to Run	1975
Steppenwolf	Born to Be Wild	1968
Rod Stewart	Maggie May	1971
Ike & Tina Turner	River Deep, Mountain High	1966
U2	With or Without You	1987
Wilson Pickett	In the Midnight Hour	1965

roots weighted by perceived quality. Because the corpora selected complete songs rather than random selections of individual chords, however, it is impossible to estimate appropriate relative frequencies for either corpus without making some assumptions about how selecting songs interacts with the relative frequencies of chord roots.

The simplest assumption to make is that there is no interaction at all, i.e., that all songs have the same distribution of chord roots on average and that there is no systematic deviation from this average, or more formally, if X is a random variable mapping each outcome $\omega \in \Omega$ to its root and Y is a random variable mapping each outcome to its song, then $f_{\mathrm{X}|\mathrm{Y}} = f_{\mathrm{X}}$. Under this assumption, there is a twelve-dimensional parameter vector $\boldsymbol{\pi}$ representing the relative frequencies of each pitch class in the corpus, weighted for popularity or quality as appropriate to the corpus, with the constraint that $\sum_{i=0}^{11} \pi_i = 1$ so that

$$f_{\mathrm{X}}(i; \boldsymbol{\pi}) = \pi_i \tag{4.1}$$

is a valid probability mass function for $i \in \{0, 1, \ldots, 11\}$; such a random variable X is described as following a *categorical distribution.* Because of the assumption that the song is completely independent of X, the probability distribution function for random variables $\mathbf{Z}$, where each of the Z_i, for $i \in \{0, 1, \ldots, 11\}$, map outcomes ω to the total number of beats with pitch class i in the song $\mathrm{Y}(\omega)$. Strictly speaking, because of the assumption that the sample space Ω of all popular music is very large but finite,

$$f_{\mathbf{Z}|\mathrm{Y}}(\mathbf{z}, y; \boldsymbol{\psi}) = \frac{\prod_{i=0}^{11} \binom{m_i}{z_i} \psi_i^{z_i}}{\sum_{\{\mathbf{z}' : \sum_i z_i' = \mathrm{length}(y)\}} \prod_{i=0}^{11} \binom{m_i}{z_i'} \psi_i^{z_i'}} \tag{4.2}$$

where m_i is the total number of instances of chord root i in Ω for each $i \in \{0, 1, \ldots, 11\}$ and $\boldsymbol{\psi}$ is a vector of weights corresponding to popularity in the case of the *Billboard* corpus or quality in the case of the *Rolling Stone* corpus. This distribution is known as *Fisher's non-central multivariate hypergeometric distribution.* Its mathematical properties make it rather difficult to use: even when the parameters are known, the expected value of this distribution is usually only approximated (McCullagh & Nelder 1989, pp. 261–62). Fortunately, when the m_i are much larger than the z_i, which is certainly the case for the corpora under study here, the much simpler *multinomial distribution* is a close approximation:

$$f_{\mathbf{Z}|\mathrm{Y}}(\mathbf{z}, y; \boldsymbol{\pi}) = \frac{\mathrm{length}(y)!}{\prod_{i=0}^{11}(z_i!)} \prod_{i=0}^{11} \pi_i^{z_i} \,. \tag{4.3}$$

Using the relative frequencies of roots in the corpus to estimate π is equivalent to assuming this multinomial approximation on $f_{\mathbf{Z}|\mathrm{Y}}$ (McCullagh & Nelder 1989, pp. 164–74).

It is very easy to compute simple relative frequencies on a corpus, and so like many others, De Clercq and Temperley implicitly assume both the multinomial approximation and the independence of songs and root distributions in their analyses of the RS 5 × 20 corpus. Musically, however, it is rather implausible that songs and root distributions should be independent: it is exactly such a dependence that allows one to describe individual songs as having a particular harmonic feel. One might take this idea to its logical extreme and assume that there are no identifiable commonalities among songs whatsoever, namely that the distribution of chords in any one song is completely unique to the song itself. Such an assumption has its own

musical implausibilities, however, as it denies any notion of style. Ideally, one needs a compromise that models $f_{Z|Y}$ using more information about the song than its length.

The *Dirichlet-multinomial distribution*, also known as a *multivariate Pólya distribution* or *compound multinomial distribution*, is such a compromise (Mosimann 1962). Like the case where songs and root distributions are assumed to be independent, the Dirichlet-multinomial distribution presumes that the counts of roots in any given song do follow a multinomial distribution, but it allows each song to have its own multinomial distribution, and thus also its own parameter vector $\boldsymbol{\pi}_y$ characterising that multinomial distribution. The Dirichlet-multinomial distribution further assumes that the $\boldsymbol{\pi}_y$ cluster around an 'average' parameter vector $\boldsymbol{\pi}'$, which one can think of as characterising the multinomial distribution over the roots of a prototypical song in the corpus. The degree to which the actual songs deviate from this prototype is reflected by a *dispersion parameter* $\phi \in (0, 1)$. Low values of ϕ imply that there is little variation, i.e., that all of the $\boldsymbol{\pi}_y$ are very close to $\boldsymbol{\pi}'$ and that the simple multinomial approximation is less problematic; high values of ϕ, particularly those greater than 0.5, imply that the $\boldsymbol{\pi}_y$ may intuitively seem somewhat far from $\boldsymbol{\pi}'$. More specifically, the $\boldsymbol{\pi}_y$ are assumed to follow a *Dirichlet distribution:*

$$f_{\boldsymbol{\Pi}}(\boldsymbol{\pi}; \boldsymbol{\pi}', \phi) = \frac{\Gamma\left(\frac{1-\phi}{\phi}\right)}{\prod_{i=0}^{11} \Gamma\left(\frac{1-\phi}{\phi}\pi_i'\right)} \prod_{i=0}^{11} \pi_i^{\frac{1-\phi}{\phi}\pi_i'-1} \,. \tag{4.4}$$

The Dirichlet-multinomial distribution function is the expected value of all possible multinomial distribution functions when the parameters $\boldsymbol{\pi}_y$ of

the multinomial distributions obey a Dirichlet distribution:

$$f_{\mathbf{Z}|Y}(\mathbf{z}, y; \boldsymbol{\pi}', \phi) = \mathbf{E}_{\boldsymbol{\Pi}}\left[\frac{\text{length}(y)!}{\prod_{i=0}^{11}(z_i!)}\prod_{i=0}^{11}\pi_i^{z_i}\right] \tag{4.5}$$

$$= \int_{\triangle^{11}} \frac{\text{length}(y)!}{\prod_{i=0}^{11}(z_i!)}\prod_{i=0}^{11}\pi_i^{z_i} \cdot \frac{\Gamma\left(\frac{1-\phi}{\phi}\right)}{\prod_{i=0}^{11}\Gamma\left(\frac{1-\phi}{\phi}\pi_i'\right)}\prod_{i=0}^{11}\pi_i^{\frac{1-\phi}{\phi}\pi_i'-1}\,\mathrm{d}\boldsymbol{\pi} \tag{4.6}$$

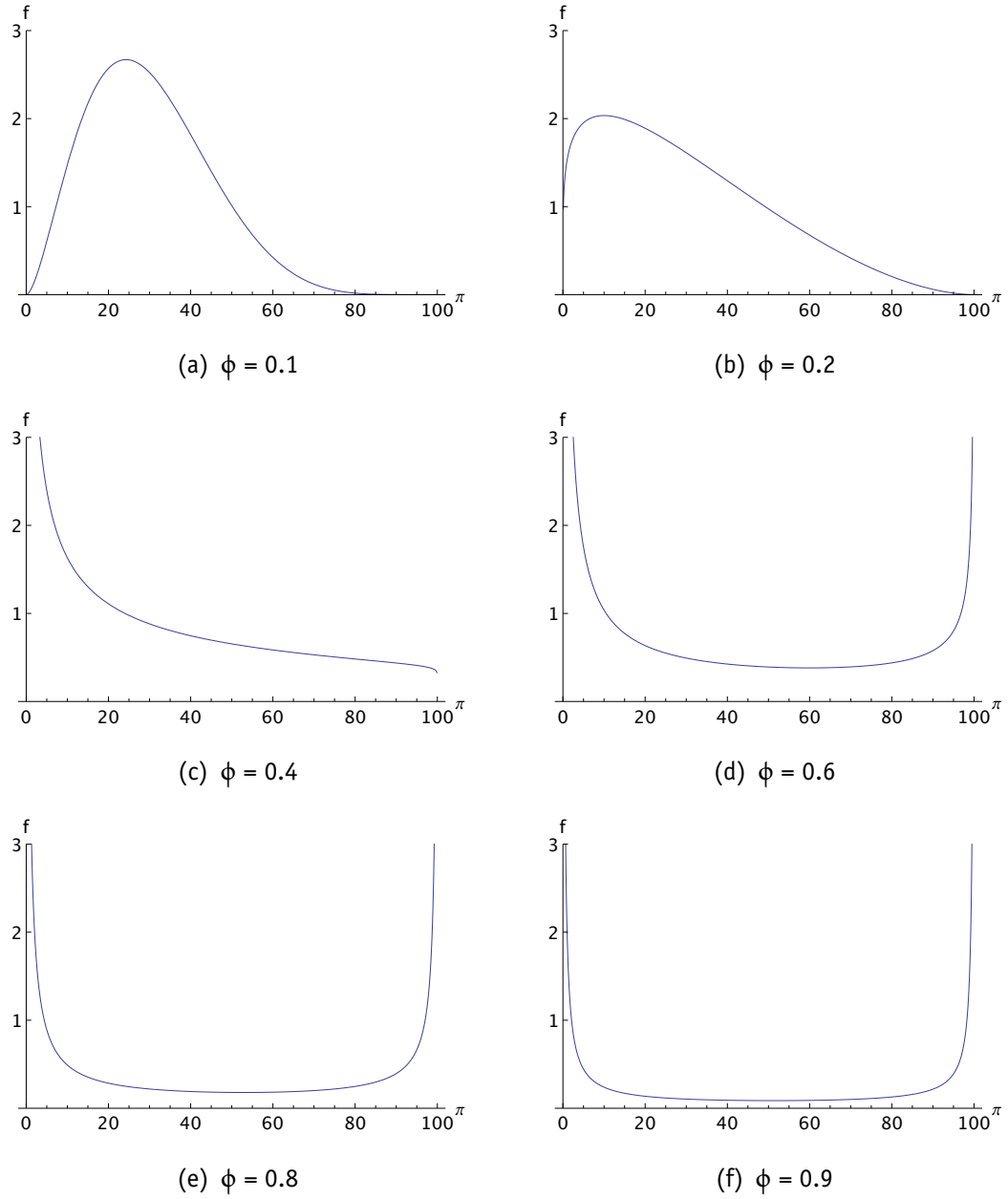
$$= \frac{\text{length}(y)!}{\prod_{i=0}^{11}(z_i!)} \cdot \frac{\Gamma\left(\frac{1-\phi}{\phi}\right)}{\prod_{i=0}^{11}\Gamma\left(\frac{1-\phi}{\phi}\pi_i'\right)} \cdot \int_{\triangle^{11}} \prod_{i=0}^{11}\pi_i^{z_i+\frac{1-\phi}{\phi}\pi_i'-1}\,\mathrm{d}\boldsymbol{\pi} \tag{4.7}$$

$$= \frac{\text{length}(y)!}{\prod_{i=0}^{11}(z_i!)} \cdot \frac{\Gamma\left(\frac{1-\phi}{\phi}\right)}{\prod_{i=0}^{11}\Gamma\left(\frac{1-\phi}{\phi}\pi_i'\right)} \cdot \frac{\prod_{i=0}^{11}\Gamma\left[z_i + \frac{1-\phi}{\phi}\pi_i'\right]}{\Gamma\left[\text{length}(y) + \frac{1-\phi}{\phi}\right]} \tag{4.8}$$

$$= \frac{\text{length}(y)!}{\prod_{i=0}^{11}(z_i!)} \cdot \frac{\prod_{i=0}^{11}\prod_{j=1}^{z_i}\left[\pi_i'(1-\phi)+\phi(j-1)\right]}{\prod_{j=1}^{\text{length}(y)}\left[1-\phi+\phi(j-1)\right]}\,. \tag{4.9}$$

Here, the notation $\int_{\triangle^{11}}\mathrm{d}\boldsymbol{\pi}$ is the multiple integral over all valid $\boldsymbol{\pi}$, i.e., $\int_0^1 \mathrm{d}\pi_0 \int_0^{1-\pi_0}\mathrm{d}\pi_1 \cdots \int_0^{1-\sum_{i=0}^{9}\pi_i}\mathrm{d}\pi_{10}$. Form (4.8) is the most common form of the Dirichlet-multinomial distribution, usually with the substitution $\alpha_i \triangleq \frac{1-\phi}{\phi}\pi_i'$ (which implies that $\frac{1-\phi}{\phi} = \sum_{i=0}^{11}\alpha_i$). This form is also the form used internally in most numerical optimisation packages. Bioinformaticians, in contrast, make more direct mathematical use of the dispersion parameter and thus sometimes prefer to use the equivalent form (4.9) instead (Curran et al. 1999; Paul, Balasooriya & Banerjee 2005; Tvedebrink 2010).

The Dirichlet distribution has its flaws as a choice for averaging over multinomial distributions, but it has proved remarkably difficult to find



(a) $\phi = 0.1$ (b) $\phi = 0.2$

(c) $\phi = 0.4$ (d) $\phi = 0.6$

(e) $\phi = 0.8$ (f) $\phi = 0.9$

Figure 4.6 · Marginal distribution functions of the Dirichlet distribution for any parameter π_i (measured in %) at selected values of ϕ and with $\pi_i' = 30\%$.

viable alternatives (for a good discussion of the these issues, see Aitchison 1982). The Dirichlet distribution also has some strong assets, however, particularly the dispersion parameter ϕ, which can characterise a variety of possible 'shapes' in the data. Figure 4.6 illustrates this effect. More specifically, this figure shows the marginal distribution functions for an arbitrary π_i (i.e., an arbitrary component of one of the $\boldsymbol{\pi}_y$ in the context of a Dirichlet-multinomial distribution) at several selected values of ϕ when π'_i has been fixed to be 30 percent. As ϕ increases, the distribution starts to skew in favour of smaller values of π_i and eventually splits into what is known as a *bimodal* distribution in which values of π_i near the nominally expected value of 30 percent are in fact much less likely than extreme values near 0 or 100 percent. One important message from this figure is that when ϕ is greater than about 0.5, one must be aware that although $\boldsymbol{\pi}'$ is technically a correct average, one should not expect any particular $\boldsymbol{\pi}_y$ to look anything like it.

Before using the multinomial and Dirichlet-multinomial models to investigate the *Billboard* and *Rolling Stone* corpora, one final note is necessary about how to estimate parameters like $\boldsymbol{\pi}$, $\boldsymbol{\pi}'$, and ϕ. There are many approaches to estimating parameters (see Wasserman 2004, chap. 9–10, or any other graduate text on statistics for an overview), but for this research, all parameters are approximated using the *maximum-likelihood* approach. The maximum-likelihood approach presumes that every instance in a corpus was drawn independently of all others (which is not strictly true of the *Rolling Stone* corpus given the rules for selecting songs, but it is impractical to try to correct for its bias). The *maximum-likelihood estimator* (MLE), often

denoted $\hat{\boldsymbol{\theta}}$, is the value of $\boldsymbol{\theta}$ that maximises the *likelihood function*

$$\mathscr{L}_{\mathrm{X}}(\boldsymbol{\theta}; \mathbf{x}) \triangleq \prod_i f_{\mathrm{X}}(x_i; \boldsymbol{\theta}) \,, \tag{4.10}$$

or equivalently maximises the *log-likelihood function*

$$\ell_{\mathrm{X}}(\boldsymbol{\theta}; \mathbf{x}) \triangleq \sum_i \log \mathscr{L}_{\mathrm{X}}(\boldsymbol{\theta}; \mathbf{x}) \,, \tag{4.11}$$

for a random variable X with a distribution function parameterised by a vector $\boldsymbol{\theta}$ and a corpus $\mathbf{x}$. In addition to $\hat{\boldsymbol{\theta}}$, frequentist statisticians will usually provide *confidence intervals* for the components of $\boldsymbol{\theta}$ at some level of confidence α. If the corpus x indeed was drawn from a sample space such that the distribution function of X is f_{X} for some parameter setting $\boldsymbol{\theta}^*$, then if one could compute confidence intervals an infinite number of times, the relative frequency of intervals that failed to include $\boldsymbol{\theta}^*$ would be α.

Finally, in the case of musical data, one must decide how to count the number of musical elements in a song. Many authors, including De Clercq and Temperley, simply count the number of unique appearances of a chord, regardless of duration, but it is unclear to what sample space Ω such an approach would correspond. There are a number of better-defined choices, but for the purposes of this research, each musical beat is counted as a distinct entity, e.g., a C major chord that lasted one bar in $\frac{4}{4}$ time would be counted as four instances of C. This choice corresponds as closely as practical to weighting the sample space to reflect the actual relative amount of time a listener would have heard one entity versus another.

Table 4.3 · Expected frequencies of absolute roots

	BILLBOARD				ROLLING STONE			
	DIR-MULTI		MULTI		DIR-MULTI		MULTI	
ROOT	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}$ (%)	C.I.	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}$ (%)	C.I.
C♯	2	2–2	2	2–2	4	2–6	4	3–4
F♯	4	3–5	4	4–4	5	3–7	4	4–5
B	7	6–7	7	6–7	7	5–9	5	5–5
E	10	9–11	12	12–12	11	9–14	13	12–13
A	13	12–14	**13**	13–13	**14**	11–17	**13**	13–14
D	13	12–14	**13**	13–13	13	10–17	11	11–12
G	**14**	12–15	12	12–12	13	9–16	12	12–12
C	12	11–13	12	12–12	10	7–13	9	9–10
F	10	9–11	9	9–9	8	6–10	9	9–10
B♭	7	6–8	7	7–7	6	4–8	8	7–8
E♭	5	4–5	5	5–5	4	3–6	6	5–6
A♭	4	3–5	5	5–5	5	3–6	6	5–6
	$\hat{\phi}$	C.I.			$\hat{\phi}$	C.I.		
	0.41	0.40–0.42			0.47	0.44–0.50		

Chord Roots

Table 4.3 presents $\hat{\boldsymbol{\pi}}'$ for the Dirichlet-multinomial model and $\hat{\boldsymbol{\pi}}$ for the simple multinomial model as percentages over the roots of all chords, reduced to their twelve-tone pitch class, in both the *Billboard* and *Rolling Stone* corpora. These correspond to the expected relative frequencies of each pitch class under the model, and the maximum expected frequency under

each model is marked as bold to aid reading. The table also presents $\hat{\phi}$ for the Dirichlet-multinomial models. Although the Dirichlet-multinomial distribution is more tractable than Fisher's non-central hypergeometric distribution, it is complex enough that there are no published methods to my knowledge for computing exact confidence intervals for its parameters, but it is possible to estimate *standard errors*, which are approximately distributed according to the well-known Gaussian distribution for sufficiently large samples, and derive approximate confidence intervals using them. The confidence intervals for the Dirichlet-multinomial parameters in table 4.3 and all other tables in this chapter use standard errors with a Bonferroni correction (Wasserman 2004, p. 166) such that approximately 19 times out of 20, *all* of the presented confidence intervals will contain the correct values of their corresponding parameters (without the Bonferroni correction, the confidence intervals would be narrower but one could only claim that *each* presented confidence interval would contain the correct parameter 19 times out of 20, which would imply that one would expect at least one of them to be wrong in any given table). The multinomial distribution is much simpler, but confidence intervals on multinomial distributions have received surprisingly little attention in the literature. There are several computationally intensive techniques for computing exact or near-exact confidence intervals on the parameters of a multinomial distribution (Sison & Glaz 1995; Hou, Chiang & Tai 2003) as well as more easily computed approximations (Goodman 1965; Bailey 1980). Table 4.3 and all other tables in this chapter use one of J. B. R. Bailey's methods to compute approximate confidence intervals on multinomial distributions (1980, eqs. 6 and 8′); this method is designed such that approximately 19 times out of 20, every

one of the confidence intervals will contain the correct value of its corresponding parameter without needing an extra Bonferroni correction. Both of these corpora are sufficiently large that the confidence intervals for the simple multinomial models are extremely tight, but for the more complex Dirichlet-multinomial model, the confidence intervals are wider.

The table shows that the roots are predictably skewed toward white notes, with perhaps slightly more emphasis on the flatter side of the circle of fifths for the *Billboard* corpus and slightly more emphasis on the sharper side of the circle of fifths for the *Rolling Stone* corpus. The differences between the Dirichlet-multinomial and multinomial estimates of the expected frequencies of roots are negligible; indeed, the confidence intervals overlap in all cases. The extra dispersion parameter one has in the Dirichlet-multinomial model tells a critical story, however, that the simple multinomial model misses entirely. As illustrated earlier in figure 4.6, at values of ϕ around 0.4, the roots in individual songs are not necessarily distributed near the expected values reflected in $\hat{\boldsymbol{\pi}}'$. A model with less dispersion is necessary to draw stronger conclusions about the harmonic style – and the simple multinomial models that are so often used in the literature are, in this case, misleading.

Standard music theory would suggest that key has a substantial effect on the distribution of pitch classes, and so it is the logical place to start when attempting to find a better-dispersed model. Table 4.4 fits the same models as those of table 4.3 but to the pitch classes of the acting tonics at every beat rather than the roots of the chords at every beat. Here, the simple multinomial fits are superficially plausible but extremely misleading. The great majority of popular songs visit only a single key and very few visit

Table 4.4 · Expected frequencies of different tonics

	BILLBOARD				ROLLING STONE			
	DIR-MULTI		MULTI		DIR-MULTI		MULTI	
ROOT	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}$ (%)	C.I.	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}$ (%)	C.I.
F♯	12	10–15	2	2–2	12	6–17	3	3–3
B	12	10–15	5	5–5	12	6–17	2	2–3
E	6	4–7	13	13–13	10	5–15	**17**	16–17
A	6	4–7	13	13–13	8	3–12	15	15–15
D	7	5–9	**15**	15–15	6	2–10	9	9–10
G	5	4–7	11	11–11	8	4–13	12	12–12
C	6	5–8	14	14–14	6	2–10	10	10–11
F	4	3–5	8	8–8	6	2–10	10	10–10
B♭	4	2–5	6	6–6	4	1–8	7	6–7
E♭	12	10–15	5	5–5	4	1–7	7	7–7
A♭	12	10–15	6	6–6	12	6–17	4	3–4
D♭	12	10–15	1	1–1	12	6–17	4	4–4
	$\hat{\phi}$	C.I.			$\hat{\phi}$	C.I.		
	0.93	0.92–0.93			0.92	0.90–0.94		

more than two, but the simple multinomial model presumes that songs are ready to modulate at any moment. Again, the extra dispersion parameter in the Dirichlet-multinomial model saves the day. For both corpora, $\hat{\phi}$ is very high, reflecting the fact that although there is much variation among songs, the chords in any given song are concentrated on just a few keys. With such an extreme dispersion parameter, however, one must interpret $\boldsymbol{\pi}'$

with caution: although it would seem that $\hat{\boldsymbol{\pi}}'$ implies a high frequency of remote keys in each corpora, in fact, these vectors illustrate that songs in keys near the centre of the circle of fifths are more likely to have passages in other keys, whereas songs in remote keys are less likely to modulate.

This example is primarily illustrative and confirmatory, as one does not really need to cite a ϕ-value to support that statement that most popular songs remain in more or less a single key throughout. There are other hidden dangers in table 4.4, however, stemming from the fact that keys do not follow the assumptions of a simple multinomial model at all, and yet it is exactly this model that one presumes when examining simple counts of chords. When using a statistical software package to do a simple multinomial analysis, one might well see confidence intervals like the ones presented in the table. These intervals are mathematically correct, but because the assumptions underlying the model are so incorrect, they are musicologically meaningless. Because almost all songs only explore one or two keys, when asking about the distribution of keys in the corpus, one is more likely thinking of the question of how likely it is that any given key will appear at some point in a song, not the probability that a chord chosen at random from any moment in time will be in a particular key. Table 4.5 presents these frequencies with approximate confidence intervals according to the method of Alan Agresti and Brent Coull (1998) such that any one of them will contain the correct frequency 19 times out of 20. Because each of them represents a single probability, the Bonferroni correction did not seem appropriate. Even without the Bonferroni correction, however, which usually widens confidence intervals considerably, the confidence intervals are much wider, especially for the smaller *Rolling Stone* corpus. The reduced

Table 4.5 · Expected proportion of songs visiting each tonic

	BILLBOARD		ROLLING STONE	
TONIC	f (%)	C.I.	f (%)	C.I.
F♯	2	1–3	5	3–9
B	5	4–7	4	2–8
E	13	12–15	**20**	15–27
A	12	11–14	16	11–21
D	**15**	14–17	12	8–17
G	12	11–14	16	12–22
C	14	13–16	12	8–17
F	9	8–11	12	9–18
B♭	8	7–10	9	6–14
E♭	6	5–7	8	5–13
A♭	7	6–8	4	2–8
D♭	1	0–2	5	3–9

precision is perhaps frustrating, but it is a more accurate reflection of the strength of conclusions one can draw from these corpora. The apparent dip for D-based keys in the *Rolling Stone* corpus, for example, is not statistically significant, and likewise, even the apparently heavy emphasis on E-based keys in the *Rolling Stone* corpus is only statistically distinguishable from the particularly unusual keys of F♯, B, A♭, and D♭. Certainly it is not possible to conclude whether there is any difference with respect to the distribution of keys between the *Billboard* and *Rolling Stone* corpora.

Ultimately, the most sensible manner of analysing chord roots is relative to key. Typically, this is done relative to the local key, and such was De

Table 4.6 · Expected frequencies of roots relative to overall tonic

	BILLBOARD				ROLLING STONE			
	DIR-MULTI		MULTI		DIR-MULTI		MULTI	
ROOT	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}$ (%)	C.I.	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}$ (%)	C.I.
♯IV	1	1–1	1	1–1	0	0–1	0	0–0
VII	1	1–1	0	0–0	0	0–1	0	0–0
III	3	2–3	3	3–3	2	1–2	2	2–2
VI	5	4–5	6	6–6	3	2–4	5	5–6
II	5	5–6	6	5–6	3	2–4	4	4–4
V	16	15–17	15	15–15	16	13–18	14	14–15
I	**41**	40–43	**40**	40–40	**48**	44–53	**46**	45–46
IV	19	18–20	18	18–18	21	17–24	19	18–19
♭VII	4	3–4	5	5–5	3	2–4	5	4–5
♭III	2	2–2	2	2–2	2	1–2	2	2–2
♭VI	2	2–3	3	3–3	1	1–2	2	2–3
♭II	1	1–1	1	1–1	1	0–1	0	0–0
	$\hat{\phi}$	C.I.			$\hat{\phi}$	C.I.		
	0.20	0.19–0.20			0.21	0.19–0.23		

Clercq and Temperley's analysis of the RS 5 × 20 corpus. John Snyder has shown that different strategies for simplifying with respect to key can cause substantially different statistical conclusions, however, and found that tracking scale degrees relative to the global key yields more accurate conclusions when faced with harmonic complexity (Snyder 1990). Table 4.6 shows the estimated parameters for the Dirichlet-multinomial and simple

multinomial models for these 'structural' roots. The dispersion parameters at 0.2 immediately show that this model is more reasonable. Predictably, the tonic, subdominant, and dominant have the overwhelming majority of the weight. Consistent with the relatively low dispersion parameter, the differences between the Dirichlet-multinomial and simple multinomial models are trivial in most cases, with only some minor differences for VI and ♭VII chords, although it is still high enough to suggest that the Dirichlet-multinomial model (and the wider confidence intervals that accompany it) is the more appropriate model. The differences between the corpora are minor, although it appears that *Billboard* corpus may have slightly more emphasis on the sharper side of the circle of fifths and very remote chords, perhaps indicative of a greater prevalence of modulations up a whole tone or semitone in pop relative to rock.

As De Clercq and Temperley did with the RS 5 × 20, it is logical to wonder whether these distributions change over decades or affect the chart position. The decade does have a statistically significant effect for both corpora, and table 4.7 presents the parameters of the Dirichlet-multinomial models fit to each decade for the *Billboard* corpus. The musicological story is consistent with what one would expect. Harmony in the 1950s was considerably simpler than it became in later decades, and the lower $\hat{\phi}$ also illustrates that there was less variation among songs. Harmonic complexity and diversity seems to have peaked in the 1970s, as reflected by the high value of $\hat{\phi}$. Another notable change in the 1970s is the increase in emphasis on the flatter side of the circle of fifths, a change that persists through until the end of the corpus. After a reduction in diversity in the 1980s, it appears that diversity and complexity come back in the 1990s, a pattern that is

Table 4.7 · Expected frequencies of roots relative to overall tonic, by decade (*Billboard* corpus only)

	1958–59		1960–69		1970–79		1980–89		1990–91	
ROOT	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}'$ (%)	C.I.
♯IV	1	0–2	1	1–1	1	1–1	1	1–1	1	0–1
VII	1	0–2	1	0–1	1	0–1	1	1–1	1	0–1
III	1	0–2	3	2–3	3	2–3	3	2–4	3	1–5
VI	3	1–5	5	4–6	5	4–6	4	4–5	6	3–9
II	3	1–5	5	4–6	4	3–5	6	5–7	5	3–8
V	23	17–29	18	16–20	14	13–16	16	14–18	17	12–22
I	47	39–55	44	41–47	40	38–43	41	38–44	37	29–45
IV	18	12–24	18	15–20	20	18–22	19	17–21	21	15–26
♭VII	1	0–2	2	2–3	5	5–6	4	3–4	4	2–6
♭III	1	0–2	2	1–2	2	2–3	2	2–3	3	1–4
♭VI	0	0–1	1	1–2	3	3–4	2	2–3	3	1–4
♭II	1	0–2	1	1–2	1	1–1	1	1–1	1	0–2
	$\hat{\phi}$	C.I.	$\hat{\phi}$	C.I.	$\hat{\phi}$	C.I.	$\hat{\phi}$	C.I.	$\hat{\phi}$	C.I.
	0.14	0.10–0.18	0.19	0.17–0.20	0.22	0.21–0.23	0.18	0.16–0.19	0.26	0.21–0.30

clear even with the relatively small selection of songs from the 1990s in the *Billboard* corpus because of the cut-off date in 1991. The pattern in the *Rolling Stone* corpus is similar.

For the *Billboard* corpus, it is also possible to assess whether the distribution of roots has any effect on chart popularity. The preliminary results were encouraging. After mapping each song to the highest quintile it ever achieved on the *Billboard* chart (1–20, 21–40, etc.), an analysis analogous to that for decades in table 4.7 also proved to be statistically significant, i.e., if roots in popular music weighted by popularity followed a Dirichlet-multinomial model that did not vary over decade, then the relative frequency of random corpora where the models by decade would fit as much better than the simple model from 4.6 as they do for the actual *Billboard* corpus would be less than 5 percent. Musicologically, however, the differences were trivial.

Chord Classes

One of the most important contributions of the *Billboard* corpus is that, unlike the *Rolling Stone* corpus, not only the roots but also the qualities of all of the chords have been curated. Moreover, these qualities include detailed information about upper extensions and inversions, comparable in detail only to Christopher Harte's annotations of the Beatles' œuvre. Excluding inversion, there are 104 distinct qualities represented in the corpus. Table 4.8 lists those that appear more than one thousand times along with their relative frequency over the entire corpus. Collectively, major chords, minor chords, and dominant seventh chords dominate the

Table 4.8 · Most frequent chord qualities in the *Billboard* corpus

QUALITY	FREQUENCY
maj	51.6
min	13.0
7	10.1
min7	8.4
maj7	2.6
5	2.1
[open bass]	2.0
add9	1.1
maj6	0.9
sus4	0.9
sus7	0.8
sus9	0.8
7(♯9)	0.7
min9	0.7
maj9	0.4
11	0.3
9	0.3
7(omit3)	0.3
13	0.2
min11	0.2
sus2	0.2
maj6(add9)	0.2
min6	0.2
dim	0.2

corpus, accounting for more than two thirds of all chords. As the table illustrates, however, with a corpus of this size, it is possible to extend previous investigations to consider the most important extended chords: certainly suspended chords and added ninths, and perhaps also added sixths (or thirteenths). With only a few exceptions – e.g., Ian Simon, Dan Morris, and Sumit Basu, who included suspended chords in an accompaniment-generation system for vocal melodies (2008) – researchers have ignored these extended chords, choosing instead to model augmented and diminished triads and the seventh chords thence derived. The frequencies in this corpus do not support this practise: of this entire category, only the diminished triad appears more than a thousand times, and even altogether, these chords comprise less than half a percent of the corpus.

Unfortunately, with so many possible chords, it is impractical to undertake a Dirichlet-multinomial analysis. One can make a step in this direction, however, by making use of Bayesian networks, as first seen in section 2.2. The challenge is determining what the structure of this network should be, ideally by gleaning much information from the corpus as possible. Learning the structure of Bayesian networks from data is a difficult task, and several classes of approaches have been tried; Koller and Friedman have an excellent discussion of the issues involved in their textbook on graphical models (Koller & Friedman 2009, chap. 18).

Because it is more compatible with a frequentist interpretation of probability than some of the other approaches, I chose to use the *constraint-based* family of approaches to structure learning (in contrast to the more subjectivist *score-based* family). Working on the principle that Bayesian networks are intended to represent independence relationships, constraint-based

methods rely on a battery of independence tests between all possible combinations of relevant random variables. Each of these tests is designed such that after an infinite number of them, the relative frequency of failures (variables deemed to have a relationship when in fact they are independent) would be less than or equal to some fixed probability α, traditionally 5 percent. Based on these tests, one constructs an undirected skeleton of the network, and then as many of these edges as can be unambiguously directed are directed consistent with the data (recall from section 2.2 that multiple networks can represent that same independence relationships). A very large number of independence tests are necessary when using constraint-based approaches (several thousand in the example in this chapter), and so it is also good practise to use a Bonferroni correction on α, usually quite a substantial correction, to ensure that the probability that the graph posits any false relationships is a false one is less than desired. All of the learned graphs in this chapter have been corrected such that at least 19 times out of 20, they would contain no incorrect edges if the underlying data were distributed according to a multinomial distribution.

The classical algorithm for constraint-based structure learning – and the easiest to understand – is the IC algorithm of Thomas Verma and Judea Pearl (Pearl 2009, pp. 49–54), highlighted in figure 4.7. The first step relies on a very large number – $O(2^n)$ – of independence tests among variables, which can make it impractical for all but very small numbers of variables. The most commonly used variant, known as the PC algorithm, seeks to optimise this first step so as to eliminate redundant independence tests (Spirtes, Glymour & Scheines 2000, pp. 84–90). More recent work in constraint-based structure learning has focused on learning the Markov blankets

1. For each pair of variables a and b in V, search for a set S_{ab} such that $(a \perp\!\!\!\perp b \mid S_{ab})$ holds in $\hat{P}$ – in other words, a and b should be independent in $\hat{P}$, conditioned on S_{ab}. Construct an undirected graph G such that vertices a and b are connected with an edge if and only if no set S_{ab} can be found.

2. For each pair of nonadjacent variables a and b with a common neighbor c, check if $c \in S_{ab}$. If it is, then continue. If it is not, then add arrowheads pointing at c (i.e., $a \rightarrow c \leftarrow b$).

3. In the partially directed graph that results, orient as many arrowheads as possible subject to two conditions: (i) any alternative orientation would yield a new v-structure; or (ii) any alternative orientation would yield a directed cycle.

(Pearl 2009, p. 50)

Figure 4.7 · The IC algorithm for learning the structure of Bayesian networks from data.

of individual nodes in the network first in order to reduce the number of independence tests more substantially (Margaritis 2003; Tsamardinos, Aliferis & Statnikov 2003; Yaramakala & Margaritis 2005). With only a slight degradation of accuracy, the Markov-blanket methods are much faster than the IC/PC algorithm, and as such, they were the basis for the graphs presented in this chapter.

In addition to the general problem of how to learn the structure of any Bayesian network, there is also the challenge of which variables one wishes to include in the network and what may already be known about causal relationships and potential causal relationships among them. For chords, even with detailed chord annotations like those in the *Billboard* corpus – indeed, perhaps especially with detailed annotations like those in

the *Billboard* corpus – it is quite difficult to determine what the most salient descriptors should be. In order to make as few assumptions as possible, and thus to learn as much from the corpus as possible, I chose to use variables fairly close to how extended chords are described.

third can take the values of major, minor, or absent. It reflects the quality of the third of the chord.

fifth can take the values of perfect, diminished, augmented (rare), or absent. It reflects the quality of the fifth of the chord.

seventh can take the values of major, minor, diminished, or absent. It reflects the quality of the seventh of the chord.

ninth can take the values of major, minor, augmented, or absent. It reflects the quality of the ninth or added second in a chord in most cases. This variable also comes into play in sus2 chords, where third will be absent and fifth will be perfect. The augmented ninth occurs almost exclusively in the context of the funk 7(♯9) chord.

eleventh can take the values of perfect, augmented, or absent. In general, the perfect eleventh is associated with an absent third in sus4 chords whereas the augmented eleventh is usually an addition for colour.

thirteenth can take the values of major, minor, or absent. A major thirteenth is the only addition in the very common maj6 and min6 chords, but it can also appear with other extensions.

relative.bass is the interval of the bass note of the chord relative to its root, following the pattern of Christopher Harte et al. (2005).

All of these variables, of course, take their meaning from a tonal context, which I encoded as three variables.

global.tonic is the pitch class of the tonic in the overall key of the piece.

local.tonic is the scale degree of the local tonic relative to global.tonic.

global.root is the scale degree of the root of the chord relative to global.tonic rather than local.tonic, a choice motivated as before by John Snyder's guidelines for musicological corpus analysis (1990).

One of the most vexing difficulties in choosing variables is the unavoidable possibility that what appears to be a causal relationship in the data is in fact the effect of some other *confounding variable* that one has neglected to consider or measure. Because of Simpson's paradox (see § 2.2), ignoring confounding variables can cause researchers to draw entirely incorrect conclusions in some cases. Worse still, there is no general statistical test for confounding – Pearl has a thorough proof of this fact entititled 'Why there is no statistical test for confounding, why many think there is, and why they are almost right' (2009, pp. 182–89) – although there are algorithms that can flag those areas of a learned Bayesian network that could possibly be affected by such variables, e.g., the IC*/FCI algorithm (Pearl 2009, pp. 52–54). To keep the analysis simple, however, there will be no complete analysis of possible confounding variables in this chapter; this is an important area of future work. Nonetheless, many likely confounders

are readily available in the *Billboard* corpus, and these have been included in all analyses in this chapter.

decade is the decade that the song made the *Billboard* charts, which can check for changes in style over time.

song is the actual song, in order to check for idiosyncracies that go beyond traditional style. Including song also compensates for our inability to use the full Dirichlet-multinomial model. The Dirichlet-multinomial models assumes that given a particular song, the data are distributed according to a multinomial distribution, and so by including `song` in the network, the Dirichlet assumption is unnecessary.

quintile is the highest quintile of chart ranks that the song ever achieved, which can check whether and how harmony may have affected popularity.

bar.of.phrase is the ordinal number of the bar within its phrase, up to a maximum of eight bars. Very few phrases in the corpus exceed eight bars, but for such phraes, all bars beyond the eighth are designated as '>8'.

bar.length is the length of the bar in beats, which stands as a proxy for the metre. This variable does not distinguish simple metres from compound metres; e.g., a $\frac{6}{8}$ bar will have `bar.length` 2.

beat.of.bar is the ordinal number of the beat within the bar. Temperley has recently shown that this variable can be quite important (2009),

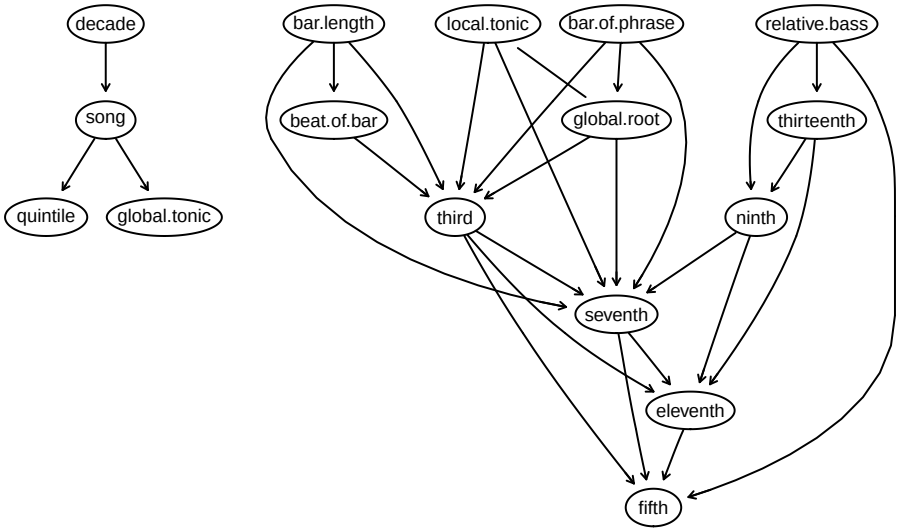


Figure 4.8 · A Bayesian network for popular chords. This network represents a single moment in time. The decade, song, global key, and peak quintile on the charts are independent of the rest of the network.

consistent with earlier work in the psychological literature (Jones, Boltz & Kidd 1982).

Figure 4.8 shows the resulting Bayesian network for popular harmony using the incremental-association algorithm (Tsamardinos, Aliferis & Statnikov 2003) on the *Billboard* corpus with a Bonferroni correction on the independence tests to ensure that the probability that the network contains an erroneous dependency is less than 0.05. This correction causes the algorithm to be quite conservative in adding edges, and so note carefully that there is no such guarantee that the network is not missing edges; rather, the absence of edges simply means that the data alone are insufficient to demonstrate a relationship between two variables. The most salient pat-

tern is the separation of decade, song, quintile, and global.tonic from all other variables. Again, although one must be wary of considering this separation definitive, it suggests that there is a somewhat stable musical style throughout the period, independent of particular decades, songs, or absolute keys. It also suggests that harmonic usage was not responsible for the popularity of songs in the period in question; either popularity is random, or there are other (confounding) factors dependent on the song that would be responsible. (Because we know exactly how the songs were selected and because they were selected at random, we know that in this case, confounding factors could not have a causal effect on the song itself.) The next notable pattern is that the algorithm was unable to determine the direction of the causal effect between local.tonic and global.root, and that both of these nodes have the same children: third and seventh. This pattern suggests that, contrary to Snyder, harmonic patterns in this style of music are based on local key without regard to the global harmonic structure; such a pattern is consistent with the well-known modulations up by tone or semitone for the final verse or chorus of a popular song, which are obviously not structural in the same way as a classical move to the dominant region or the relative major. This pattern is also consistent with recent empirical results on modelling chord sequences from Ricardo Scholz, Emmanuel Vincent, and Frédéric Bimbot (2009). Finally, it is notable that third and seventh together separate all of the other qualities of chords from the rest of the network. The corpus supports the traditional understanding in jazz theory that the third and the seventh are the defining components of a chord.

It is very tempting now to look at the right side of the figure and try to

develop an understanding of how all of the other components of chords relate to one another, at least as evidenced by the *Billboard* corpus. Here, however, one must remember that any Bayesian networks that share the same set of immoralities represent identical sets of conditional dependencies, and the only immorality among the set of chord components other than the root is third → seventh ← ninth. Moreover, detailed output from the incremental association algorithm shows that even the direction of the effect between seventh and ninth is not entirely certain. In short, one could re-orient almost any arrow among the harmonic components in figure 4.8 and have a model that the corpus supports just as well. After classifying chord types by their third and seventh (including the absence thereof), some combination of music theory and a more careful selection of variables would be necessary to unpack more information about the causal relationships, if any, among the harmonic components. This ambiguity is in sharp contrast to the middle of the figure, where there are also few immoralities, but the direction of causality is obvious from the nature of the entities themselves, e.g., the seventh of a chord certainly could not be a causative factor for the prevailing metre, and so one can confidently force the direction of the edge bar.length → seventh.

Finally, the edges from bar.length and beat.of.bar to third and seventh confirm Temperley's finding that metre has an important effect on harmony. Perhaps more interesting, however, is the edge from bar.of.phrase to global.root. This edge confirms the traditional idea that many phrases follow some kind of over-arching harmonic plan. It is particularly interesting that bar.of.phrase is the only variable for which the corpus finds evidence of a causal effect on global.root; there is only evidence for bar.length and

Table 4.9 · Expected frequencies of relative roots at different bars of the phrase (in %)

	BAR WITHIN PHRASE								
ROOT	1	2	3	4	5	6	7	8	>8
♯IV	0	0	0	0	0	0	0	0	0
VII	0	0	0	0	**1**	0	0	0	0
III	2	**3**	2	2	2	**3**	1	**3**	2
VI	5	7	5	5	5	**8**	5	4	2
II	5	5	**6**	5	**6**	5	4	2	4
V	10	17	14	21	20	19	14	**22**	8
I	**53**	34	41	35	37	34	39	39	**56**
IV	15	21	19	18	15	17	**23**	19	20
♭VII	3	6	5	6	6	5	**8**	5	5
♭III	2	**3**	2	2	2	2	1	1	1
♭VI	2	2	3	3	3	3	4	**5**	3
♭II	0	0	0	0	0	0	0	0	0

beat.of.bar affecting third and seventh. Table 4.9 breaks down the distribution of local roots on bar.of.phrase. The confidence intervals are all so tight – the largest is three percentage points wide – that they have been omitted so as to fit the table on a single page. The greatest values in each row are set in boldface. A few familiar patterns emerge. Phrases tend to begin with tonic chords, and very long phrases often end with extended tonic chords. The bump in the distribution of tonic chords in bar 3 reflects another well-known pattern of prolonging the tonic at the beginning of phrases. Dominant chords often occur in bars 4 and 8, reflecting that for phrases of typical length, it is common to end on a dominant – or more likely, elide with the tonic beginning the next phrase. Subdominant chords

are the most likely in bar 7, immediately preceding the bar where it is most likely to find dominant chords. Chords further away from the tonic on the circle of fifths tend to occur later and in longer phrases. In short, although there are certain clear differences such as the prevalence of ♭VII, common-practise norms also operate fairly clearly in the *Billboard* corpus. No other corpus to my knowledge provides as much information about phrases in popular music, and so uncovering further patterns is an exciting direction for future work.

4.4 TEMPORAL STRUCTURE

Western theories of harmony generally assume that chords operate under some kind of temporally inter-dependent structure, often described as *functional harmony*, and it is clear from the concepts that arise when trying to interpret table 4.9 that exploring traditional concepts in music theory would be easier with statistical models that accounted for time more directly than including variables like `bar.of.phrase` and `beat.of.bar`. This section explores several such models, starting like the previous section with some properties of the chord roots only, following the tradition of more recent corpus analyses of harmony like De Clercq and Temperley's. It has long been acknowledged that the notion of functional harmony assumes a causal structure on musical form (Zierolf 1983, pp. 125–31), and so the section will conclude with a discussion of two new Bayesian networks based on the *Billboard* corpus.

Pre- and Post-Tonic Distributions

One of the more discussed conclusions from De Clercq and Temperley's analysis of the RS 5 × 20 is that in rock, subdominant chords are more likely to precede tonic chords than dominant chords. Indeed, they find that the distribution of roots pre- and post-tonic are nearly identical. This behaviour is sharply contrary to traditional common-practise harmony. As discussed before, the *Billboard* corpus is based on a different, somewhat broader sample space Ω, and it does not support the same conclusion for popular music in general.

Table 4.10 lists the estimated $\hat{\pi}'$ for the relative roots of chords given that the immediately following chord is a tonic chord (i.e., the roots of pre-tonic chords). The interaction of decade and the behaviour proves to be both statistically and musicologically significant, and so the table provides a distinct set of parameters for each decade. The 1950s and 1960s show essentially common-practise patterns. Dominant (V) chords appear significantly more often than subdominant (IV) chords, especially in the 1950s, and there is very little use of chords other than the dominant and the subdominant, again particularly for the 1950s. Starting in the 1970s, there is increasing use of the subdominant chords pre-tonic, but never more than dominant chords; in fact, even with a corpus this large, as the wide and overlapping confidence intervals show, the data are insufficient to support a conclusion that the distribution of dominant and subdominant chords are different at all in the 1970s, 1980s, or 1990s (although that lack of distinction is itself a departure from common practise). Beginning in the 1970s, one also sees chords further from the tonic on the circle of fifths being used more often.

Table 4.10 · Expected frequencies of pre-tonic roots relative to the overall tonic, by decade (*Billboard* corpus only)

	1958–59		1960–69		1970–79		1980–89		1990–91	
ROOT	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}'$ (%)	C.I.
♯IV	1	0–3	1	0–1	0	0–1	0	0–0		
VII	0	0–2	1	0–2	1	0–2	0	0–1	2	0–4
III			1	0–2	2	1–2	2	1–3	1	0–2
VI	2	0–4	4	2–5	3	2–5	4	2–5	5	0–10
II	2	0–4	4	2–5	8	6–10	6	4–8	7	1–13
V	59	46–72	47	42–52	38	34–42	35	31–40	39	27–52
IV	34	21–46	35	31–40	35	31–39	32	27–36	29	18–40
♭VII	1	0–3	3	2–5	7	5–9	13	10–16	10	3–16
♭III			1	0–2	2	1–3	3	1–4	3	0–7
♭VI	1	0–2	1	0–2	2	1–3	4	2–5	3	0–7
♭II	1	0–3	1	0–1	1	0–2	1	0–2	2	0–4
	$\hat{\phi}$	C.I.	$\hat{\phi}$	C.I.	$\hat{\phi}$	C.I.	$\hat{\phi}$	C.I.	$\hat{\phi}$	C.I.
	0.33	0.20–0.47	0.45	0.41–0.50	0.46	0.42–0.49	0.54	0.51–0.58	0.62	0.53–0.71

Beginning in the 1980s, these explorations become rather weighted toward the flat side of the circle of fifths, with the ♭VII chord receiving particular emphasis; this departure from common-practise harmony is perhaps more striking than the emphasis on the subdominant chord. The other striking trend is the increase in dispersion parameters $\hat{\phi}$ as the decades progress, which is consistent with the overall trend of increasing harmonic diversity first identified in table 4.7. Unlike the earlier table, however, there is a sharp and statistically significant increase in the diversity of pre-tonic chords in the 1980s, possibly linked to the changing use of ♭VII; this is another trend worth of further inquiry. Moreover, all of the dispersion parameters in tables 4.10 are significantly higher than their counterparts in 4.7. There is more variation in harmonic practise preceding tonic chords than there is in the style overall.

Post-tonic, the distribution of dominant and subdominant chords in the *Billboard* corpus looks more like what one would expect from the common-practise era. In all decades but the 1950s, there is a statistically significantly lower expected frequency of dominant chords relative to subdominant chords, and even in the 1950s, there are simply not enough data to draw a conclusion. Throughout all decades, there is more of an emphasis on II and VI, both traditional pre-dominant chords, than in the pre-tonic distributions. Like the pre-tonic distributions (and also consistent with the pattern of roots overall from table 4.7), beginning in the 1980s, there is also emphasis on the flat side of the circle of fifths (largely at the expense of dominant chords). The pattern in dispersion parameters is the same as that of the pre-tonic distributions: all larger from their counterparts in table 4.7 and increasing at each decade with a substantial jump in the 1980s. The

Table 4.11 · Expected frequencies of post-tonic roots relative to the overall tonic, by decade (*Billboard* corpus only)

	1958–59		1960–69		1970–79		1980–89		1990–91	
ROOT	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}'$ (%)	C.I.	$\hat{\pi}'$ (%)	C.I.
♯IV	0	0–2	0	0–1	1	0–1	0	0–0		
VII	1	0–2	1	0–1	1	0–1	1	0–1	0	0–2
III	1	0–3	4	3–6	4	2–5	4	2–5	2	0–4
VI	8	2–13	8	6–11	7	5–9	8	6–10	8	3–14
II	5	0–9	8	5–10	9	7–11	6	4–8	9	3–14
V	35	24–46	22	18–25	17	14–20	16	13–19	17	9–24
IV	35	24–47	34	29–38	32	28–36	29	25–33	33	23–44
♭VII	0	0–2	5	3–6	7	5–8	9	6–11	6	2–11
♭III	1	0–3	2	1–3	3	2–5	4	2–5	3	0–7
♭VI			1	1–2	3	2–5	6	4–8	4	1–8
♭II	0	0–2	1	0–1	1	0–1	1	0–1		
	$\hat{\phi}$	C.I.	$\hat{\phi}$	C.I.	$\hat{\phi}$	C.I.	$\hat{\phi}$	C.I.	$\hat{\phi}$	C.I.
	0.28	0.19–0.37	0.40	0.37–0.43	0.44	0.42–0.47	0.52	0.49–0.56	0.50	0.42–0.58

post-tonic dispersion parameters do appear to be somewhat less than the pre-tonic, but there are not enough data in the corpus to know whether this difference is statistically significant.

✷ *Chord Transitions*

Again, the *Billboard* corpus includes more detail than any other corpus published to date, and this extra detail enables a richer analysis than has been possible before, particularly with respect to learning Bayesian networks. While it would be possible and indeed traditional to compute a full table of pre- and post- distributions for every chord, not just the tonic, such a table is only useful if there are no confounding variables affecting these distributions, and we have already seen that this is not the case. With a corpus this rich, it makes more sense to add a time lag to all of the measurable variables and learn a new Bayesian network to compare to the 'static' network of figure 4.8.

Figure 4.9 is just such a network. The variables are named as before, with the prefix prev. for the corresponding variables from the preceding beat. The differences between this network and that of 4.8 are a good lesson in how confounding variables can change the structure of a graph. Rather than decade, song, global.key, and quintile being separated from the remainder of the network, bar.of.phrase, bar.length, and beat.of.bar, are separated from the rest of the network, taking relative.bass with them. The song variable is seen to have an effect on eleventh, thirteenth, and relative.bass, suggesting that these colourations have more to do with the sound of a particular song than a global style. The global.tonic is seen to have an effect on the

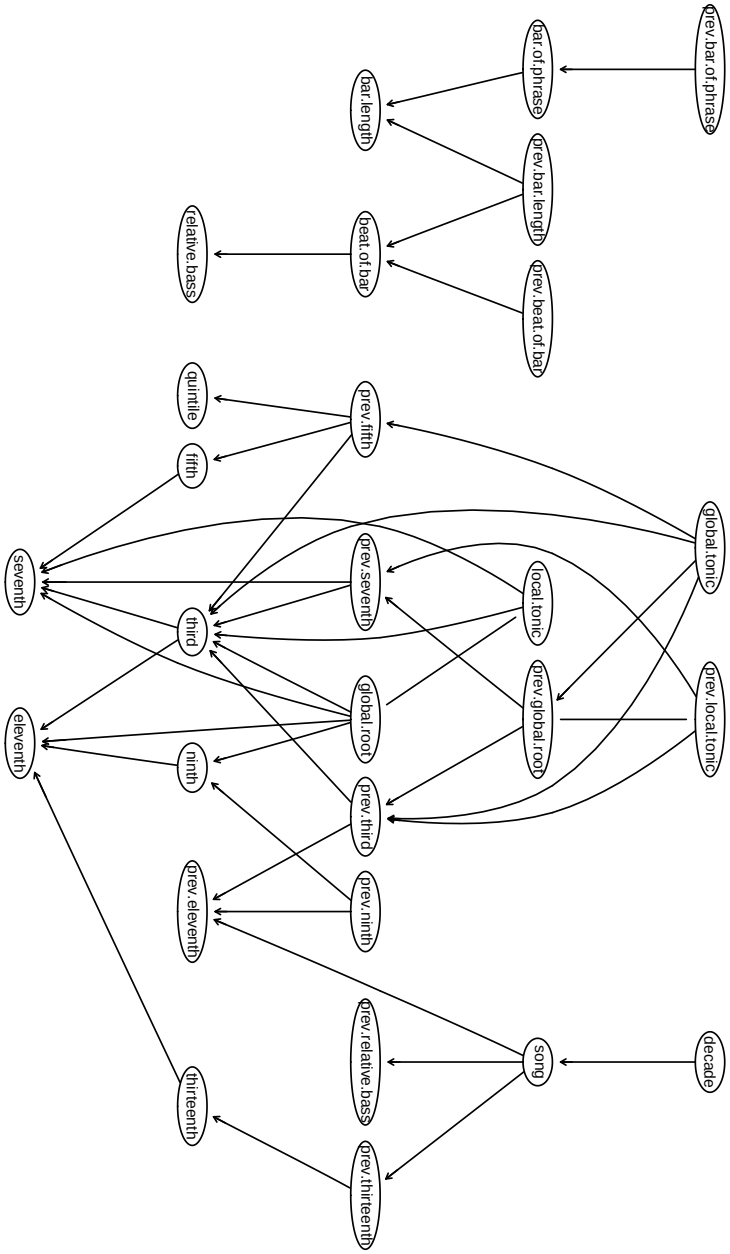


Figure 4.9 · A Bayesian network for popular chords with a one-beat time lag

transition between qualities of third. Perplexingly, the quality of fifth is seen to have a causal effect on popularity (quintile). With the exception of eleventh and relative.bass, all of the chord qualities are partly determined by the corresponding qualities on the preceding beat. Like the earlier network, local.tonic and global.root seem to be good candidates for being joined, but here, global.tonic is also seen to have an important effect, especially on the transition between prev.third and third with respect to prev.global.root, perhaps because of differences in the distribution of major and minor keys. Like the previous network, third and seventh are particularly seminal nodes, but they do not separate the network as cleanly as they did in the previous example.

Overall, the network in figure 4.9 seems difficult to use. Recall from section 3.3, however, that using a simple Markov assumption can cause problems, and by deriving dependencies based on the immediately preceding beat, this network is making just such an assumption. An alternative is to model duration separately and look for causal effects only where one chord changes to another – effectively, a semi-Markov model. Figure 4.10 uses such a model, and its patterns are considerably more familiar. Again, decade, song, and global.tonic are separated from the rest of the network, in addition to prev.local.tonic and local.tonic. Both of these separations are quite plausible, and as discussed before, suggest a common global harmonic style in all keys. Transitions between roots (prev.global.root to global.root) are an important aspect of this network, informed by prev.relative.bass and beat.of.bar. The quality of prev.third and third separate all other harmonic components from the remainder of the network. Of these harmonic components, all are again affected by their immediate predecessors. There is

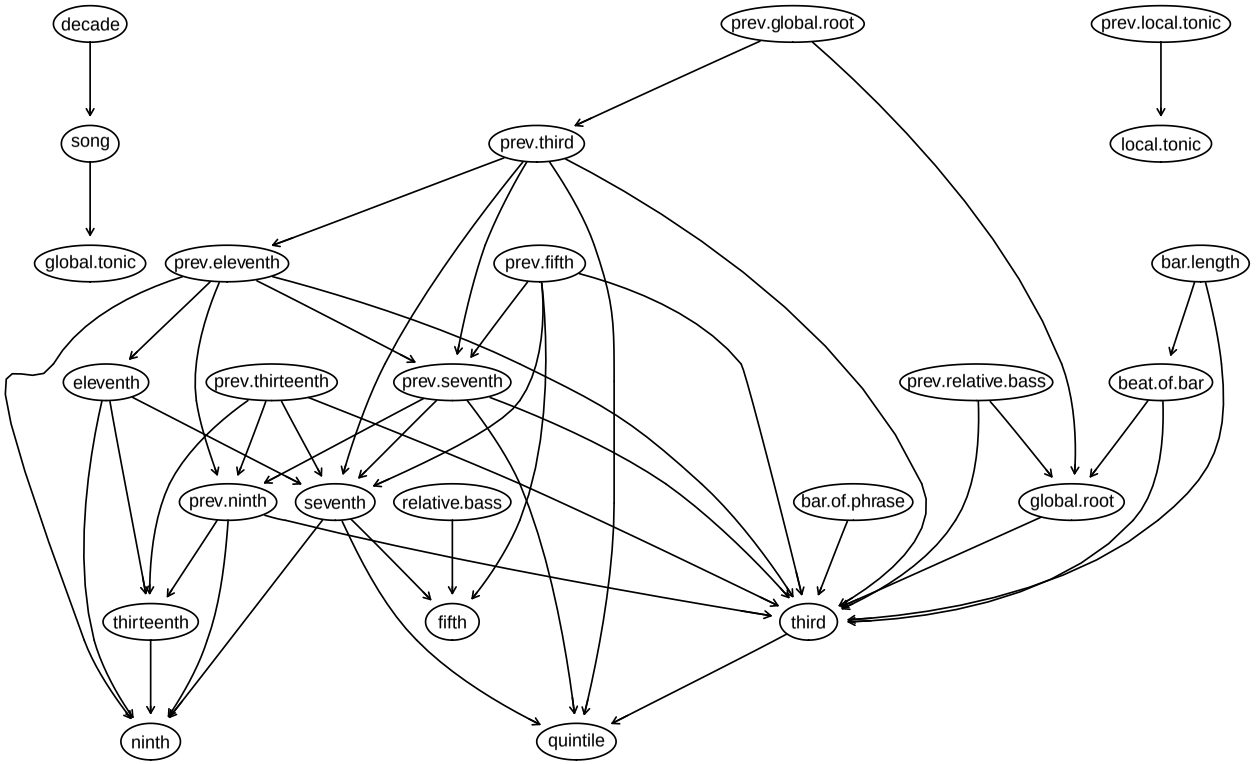


Figure 4.10 · A Bayesian network for transitions between popular chords. With information about the previous chord, even knowing the local key is unnecessary.

also a fascinating link between prev.third, third, prev.seventh, seventh, and quintile, suggesting that there are indeed aspects of harmonic style, and the operation of functional harmony in particular, that affected the popularity of songs in the United States in the late twentieth century.

More specifically, the network in figure 4.10 shows that given knowledge of the current and previous values of third and fifth, the data can support no other associations between quintile and any other variable. It is worth exploring, then, how exactly thirds and sevenths affect popularity. Table 4.12 lists the most common combinations. The first two columns list a hypothetical chord qualities for the previous and current chords, presuming the typical situations whereby the fifth is perfect and a missing third has been suspended; they imply nothing about the possible additions of ninths or thirteenths. The remaining columns list the expected percentage of instances in songs that reached major thresholds on the chart. In order to aid interpretation, the table also includes a baseline, which is average over all chord pairs in the corpus. It seems that relatively richer harmonies are good predictors of chart success: almost all combinations with major seventh chords show a substantial and statistically significant increase in the likelihood of reaching the top 20 relative to baseline, and most combinations with minor seventh chords also perform well. Surprisingly, most combinations with the dominant seventh perform significantly worse than baseline. With the structure of the networks relative to quintile varying so much across the three networks, one must be wary that there are confounding variables to consider, but as a working hypothesis, it does seem that relatively basic harmonic patterns had a meaningful effect on chart performance.

Table 4.12 · Relative frequency (%) of peak chart quintile given third and seventh

		≤ 20		≤ 40		≤ 60		≤ 80	
CHORD 1	CHORD 2	$\hat{\pi}$	C.I.	$\hat{\pi}$	C.I.	$\hat{\pi}$	C.I.	$\hat{\pi}$	C.I.
maj7	min7	77	73–80	87	84–89	94	92–95	99	98–100
maj	maj7	77	74–79	84	82–86	97	96–98	100	99–100
7	min	73	70–76	85	82–87	93	91–95	99	98–100
min7	min7	72	70–74	87	86–89	96	95–97	100	99–100
maj	min7	72	70–74	84	83–86	92	90–93	99	99–99
min7	maj	72	70–73	83	82–85	91	91–93	99	99–99
maj	sus7	71	68–74	88	86–90	93	91–94	99	98–100
maj7	maj	71	68–74	80	78–83	85	94–97	98	97–99
min	7	69	65–72	84	81–86	92	90–94	99	98–99
sus7	maj	68	65–72	88	85–90	95	93–96	99	99–100
min	maj	65	63–66	83	82–84	94	94–95	98	98–99
min7	maj7	64	61–67	81	78–83	90	88–92	99	98–99
sus	maj	63	61–65	88	86–90	93	92–95	99	98–99
maj	sus	63	61–65	86	84–88	92	91–94	97	96–98
maj	min	63	62–65	83	82–84	94	93–95	99	99–99
BASELINE		63	62–63	81	81–81	92	92–92	98	98–98
maj	maj	62	62–63	82	81–82	92	92–92	98	98–98
min	min	60	58–62	78	76–80	92	91–93	99	98–99
min7	7	58	55–61	81	78–83	93	91–95	99	98–99
maj	7	57	55–59	76	74–77	88	87–89	97	96–98
7	maj	56	54–58	76	74–78	87	86–88	97	96–98
min7	min	55	52–58	85	82–87	94	92–95	99	98–100
7	min7	52	50–55	77	75–79	91	89–92	98	98–99
5	5	50	48–51	70	68–71	87	86–88	100	100–100
7	7	45	43–47	69	68–71	83	81–84	92	91–93

The networks in figures 4.9 and 4.10 suffer from a weakness in the general incremental-association algorithm that prevents the algorithm from tying time states together – i.e., they do not guarantee that the relationships among the variable at the previous time step are identical to their relationships at the current time step. It would be possible to adapt the algorithm to be able to force consistency among such relationships, and likewise to use such an adaptation to explore dependencies at multiple time lags. Such an adaptation would be another excellent area for future work.

4.5 SUMMARY

Research in corpus-based analysis of harmony and automatic chord recognition have been hindered by a lack of detailed data on harmony. With a careful sampling methodology and a large group of expert musicians, we developed a corpus, the *Billboard* corpus, of unprecedented scope and detail. A corpus of this size and scope allows for more sophisticated statistical analysis than has been possible before. Using Dirichlet-multinomial models, it is possible to identify some of the basic properties of harmonic practise in popular music and how harmonic usage changed over time. Moreover, with a corpus of this size, structure-learning algorithms can identify some of the causal relationships among harmonic components and related musical features. Many of these results confirm traditional musical understandings, but by quantifying them, they provide a stronger basis for understanding exactly how and how much 'outliers' differ from the norm.

5 SUMMARY AND FUTURE WORK

After introducing database-driven musicology (chap. 1), this thesis presented a thorough overview of the mathematical apparatus necessary to conduct database-driven machinery well (chap. 2), reviewed how this machinery has been used for both inference and classification on musical sequences (chap. 3), and introduced a new corpus, the *Billboard* corpus, as a tool for future musicological research (chap. 4). Furthermore, this dissertation began exploring the types of information that one can extract from the *Billboard* corpus. Many of the conclusions confirm traditional understandings about harmony in popular music, but with this corpus, it has been possible to trace how some basic harmonic practises have evolved over time. These exploration used techniques that are new to the musicological and music information retrieval communities, and there are a great number of avenues for future work both with these techniques and with others.

5.1 SUMMARY OF CONTRIBUTIONS

The most significant contribution from this thesis is undoubtedly the *Billboard* corpus itself. As described in chapter 4, a lack of suitable data has seriously hindered database-driven work on harmony. This corpus is unprecedented in size, scope, degree of detail, and even sampling methodology. The corpus was designed so as to make it possible to draw conclusions about how harmonic practises have evolved over time and also how harmonic practises may have affected the popularity of songs in the latter half of the

twentieth century. As the tight confidence intervals throughout the chapter illustrate, its size allows researchers to draw meaningful conclusions even when multiple variables are interacting, e.g., decade and relative root. It encompasses a wider range of artists than any existing corpus, and within the research community, only the Beatles have previously enjoyed such careful attention to upper extensions. Moreover, because all of the annotations have been time-aligned with commercially available audio recordings, it is possible to use these annotations to improve the accuracy of audio chord recognition algorithms and test how well they can recover not only basic chords and roots but also upper extensions.

The analysis of the *Billboard* corpus introduced two techniques for corpus analysis to the musicological community. The first was the Dirichlet-multinomial distribution. In tabulating simple counts, previous researchers may not have realised that they were implicitly assuming a simple multinomial distribution over their data, an assumption that the analyses in chapter 4 suggest is rarely appropriate. In statistical terminology, this understanding would correspond to the commonly acknowledged reality that most proportions derived from data are 'over-dispersed' (see McCullagh & Nelder 1989, p. 183). The Dirichlet-multinomial not only allows for over-dispersion, it is able to quantify it, yielding a very useful extra point of information. In the context of the *Billboard* corpus, the dispersion parameter generally models the degree of harmonic diversity, and several of the analyses have been able to trace to what extent diversity increased over time.

For both the Dirichlet-multinomial and simple multinomial models, the analyses in chapter 4 are also rare in the field for presenting confidence

intervals on parameters. Long recognised as essential in the social sciences, confidence intervals are just as critical to the musicological community as databases grow and the number of variables under consideration grow. Without confidence intervals, it is impossible to know when the available data are sufficient to support a conclusion and when further data collection will be necessary. Fortunately, once on the virtuous cycle of database development whereby better inference can lead to better automatic classification, healthy collaborations between musicologists and engineers should make collecting more data easier than it has been previously.

The second technique new to musicological analysis was structure learning for Bayesian networks. These techniques are most valuable for uncovering causal relationships within the data. Although further research is necessary to identify descriptors for chords that may have clearer causal effects than the traditional music-theoretical breakdown, the analyses in chapter 4 were able to confirm and clarify hypotheses about the importance of considering relative or 'structural' roots and how metre affects choices of harmony. The most intriguing result suggested that certain harmonic colours had an effect on the popularity of songs independent of the particular decade.

Understanding the importance of statistical techniques like the Dirichlet-multinomial distribution or the causal interpretation of Bayesian networks requires a higher-level understanding of some of the mathematical and philosophical underpinnings of probability and statistics. Unfortunately, because this material is spread over so many disparate sources, it is difficult for researchers interested in database-driven musicology to gather the background necessary to conduct this type of research properly. As such, the

collection of material and citations in chapter 2 should also be a valuable contribution to the field.

5.2 FUTURE WORK

The analysis of the *Billboard* corpus in chapter 4 only broke the surface of questions that could be asked with the support of such rich data. The first question that strikes me, based on tables 4.7, 4.10, and 4.11, is whether it is possible to model changes in harmonic practise more finely than the level of the decade. Even without such an analysis, a musicological narrative for the changes in harmonic practise that arose in the 1970s would be a welcome addition to these tables. Likewise, rather than investigating only pre- and post-tonic chords, it would be well worth investigating changes in the behaviour of pre- and post-dominant chords.

There is also more work to be done in modelling chord qualities. As mentioned earlier, 104 distinct chord qualities appear in the corpus, a number which is somewhat unwieldy on its own. Moreover, the series of Bayesian networks presented in figures 4.8, 4.9, and 4.10 suggest that the traditional tertian breakdown of chord qualities may not be the most effective explanatory system for harmony in popular music. One direction for future work in this area would be to work collaboratively with music theorists to test alternative categorisations of the qualities in the corpus. A complement to this approach would be to use statistical techniques for *clustering* (see Hastie, Tibshirani & Friedman 2009) to group chord qualities that behave similarly (e.g., one would expect that a dominant ninth chord is no more or less likely to resolve to a tonic chord than a dominant seventh).

Clearer models of chord qualities could also help to explain the rather surprising results linking the popularity of songs with rather simple harmonic choices. Indeed, it is a sufficiently surprising conclusion that it would also be worth a more serious exploration of potential confounding variables of all kinds. The corpus contains not only harmonic information, but also a large amount of information about more general musical structure (e.g., the distribution of verses and choruses) that could just as plausibly be expected to have had a causal effect on popularity. These should be considered and run through structure-learning algorithms to see whether they alter the structure of the network significantly.

Finally, there is much more work to do in exploring how this data set may improve the state of the art in audio chord recognition. At first, one would try classical HMM-based models like Matthias Mauch's (Mauch & Dixon 2010b), but with the larger data set, it will also be important to explore whether discriminative models such as conditional random fields or max-margin networks could be more effective.

A CHORD TRANSCRIPTION FORMAT

In collaboration with Prof. Jonathan Wild, we developed a format for chord annotations that would be usable both by humans and by machines. Originally, we used a look-ahead left-to-right (LALR) parser (Aho, Sethi & Ullman 1986) generated with David Beazley's PLY package* to parse the files according to the grammar described below, although due to some challenges handling augmentation dots under this structure, we eventually replaced this parser with parser combinators (Frost & Launchbury 1989) built with the Daan Leijen and Paolo Martini's Parsec library† and a custom Haskell library of my own for managing music-theoretical concepts.

transcription → **title-line**
artist-line
metre-line
key-line
phrase-list
"##"

Each transcription begins with lines designating the title, artist, metre, and prevailing key of piece and closes with a doubled hash mark. In between, there is a list of phrases that constitute the transcription.

*http://www.dabeaz.com/ply

†http://hackage.haskell.org/package/parsec

title-line → "`#`" *string*

artist-line → "`#`" *string*

meter-line → "`#`" *integer* "`/`" *integer*

key-line → "`#key:`" *pitch-class*
∨ "`#key:`" *pitch-class mode*

The syntax for the special lines is fairly straightforward. All start with a hash mark followed by the title, artist, meter, or key in a standard format.

phrase-list → **phrase**
∨ **phrase phrase-list**
∨ **comment phrase-list**

comment → **key-line**
∨ **metre-line**
∨ "`#`" *string*

The phrases themselves are meant to correspond to mid-level units of musical structure. As an informal principle, we asked annotators to break phrases at any point one might realistically consider restarting a song during a rehearsal. Throughout the file, phrases may be interspersed with changes of key and metre as well as free comments. We encouraged annotators to use free comments in particular to denote major structural features such as verses, bridges, and choruses.

phrase → **bar-list** "`|`"
∨ **phrase** "`->`"
∨ **phrase** "`x`" *integer*
∨ **phrase** "`&pause`"

bar-list → **bar**
∨ **bar-list bar**

Each phrase is primarily a sequence of bars. There are three special designations, however, that may fall at the end of the line.

`->` is used to mark phrases that are elided. Each phrase always starts with its first complete bar. If this bar also functions as the close of the preceding phrase, the preceding phrase will end with the `->` designation.

`x` ***integer*** is an abbreviation denoting bars that are repeated wholesale. For example, a four-bar phrase consisting of solid G-major chords could be represented as `| G | x4`; alternating G-major and C-major chords could be represented as `| G | C | x2`. These abbreviations are not used when entire phrases repeat: instead, the entire line is recopied as many times as necessary.

`&pause` appears after phrases that are followed by a musical pause of undetermined length.

bar → **bar-start chord**
∨ **bar-start chord chord**
∨ **bar-start chord chord-or-dot chord-or-dot**
∨ **bar-start chord chord-or-dot chord-or-dot chord-or-dot**

bar-start → "|"
∨ "|" "(" *integer* "/" *integer* ")"

chord-or-dot → "."
∨ **chord**

Each bar is flanked by pipes ('|') to represent bar lines. Chords are annotated to the level of the beat, quarter notes in most simple metres and dotted quarter notes in most compound metres. In the most expanded form, each bar will contain either a chord symbol or a dot on each beat: a chord symbol on the first beat where a chord appears in each bar and a dot on each beat through which that chord continues. When the same chord lasts for the entire bar, the dots are sometimes omitted. In quadruple metres, the dots are also sometimes omitted in the very common pattern of a chord change on the third beat only.

chord → "NC"
∨ "[" *pitch-class* "]"
∨ *pitch-class chord-type*
∨ **chord** "/" *pitch-class*

Chords are represented by standard commercial chord symbols with the usual 'slash' notation for inversions. Unaccompanied bass notes are enclosed in square brackets.

Prof. Wild prepared a description of this format in more musical language for training annotators. This training document follows, including a number of his sample transcriptions.

The "lifting changes from pop songs" project

Since we need to develop statistics for how widespread various progressions are, we can't only pick great tunes to transcribe -- we have to get a random sampling from the charts. Consequently in this project you might find yourselves listening to a lot of mediocre music, though hopefully with a few isolated gems to make it worthwhile. We've tried to make these guidelines intuitive rather than rigorously spelled out, as spelling everything out rigorously for you will be too confining, and in any case there will always remain situations we haven't foreseen.

We need the chord progressions to be in a format that we can easily manipulate in a database, so we can search and analyse recurring patterns to track how they evolved over the decades. You'll produce a plain text file with the chords for each measure, and some other data. The beginning of the file will contain several lines of *metadata* (title, band, year etc.) -- these lines begin with a hash sign. One line of metadata is the time signature for the song, which we'll assume is 4/4 by default -- you only need to actually type anything in it if it's different from 4/4. Knowing the key of the song will be helpful too, so when the key is obvious, put that in too, in a line like `#key:Dm`. Some songs are too difficult to classify as in a particular key -- enter a question mark if this is the case. (Another possibility is to signal a tentative key with a question mark, if it seems like the best answer but there are still reasons to doubt it is a true key.) If the key changes for different sections of the song, please enter a new `#key:` line at each change.

Once the project is underway the lines of metadata that identify the song will already be there for you in a template file which you'll receive with the recording. The transcription file will always end with a double hash sign: `##`.

Measures

Measures are separated by vertical lines (aka "pipes", found above the backslash on most keyboards--it may look like a broken bar on some keyboards). You group measures on lines, and separate the lines according to the structure of the song -- usually when the song is simply organised into twos, fours, and eights the basic intuitive unit will be 4 bars, but there are many situations when 2 bars or 8 bars will be more appropriate. For the most part, individual lines of the file should correspond to fairly self-contained units; this is why you'd use an 8-measure line, for example, when four bars don't yet feel like a proper unit has been completed. The examples should make it clear. Basically you should start a new line at any point you can imagine could ever be a sensible point to start playing the song in a rehearsal.

The transcription of "Heart of Glass" shows places where the last measure of one 4-bar unit is also the first measure of a new 4-bar unit ("elision"). We prefer that the beginning of the new

unit be placed on a new line, meaning that the previous unit will appear to only have 3 measures. When you have to do this, use this two-character symbol after the last barline of the line: `->` which means that the progression from that line actually ends on the beginning of the following line.

"Heart of Glass" also has some truncated bars, with 3 beats. Put the time signature inside the barline at the point it changes, in parentheses, like this: `|(3/4) E|`. By default this means a change for only one measure. If a song has longer sections that remain in a different time signature, use a metadata line for each change of time signature.

Sometimes for short passages the local metre might temporarily sound different than the prevailing metre. Only change the metre if both (a) it is perceptually very clear it has to be changed and (b) the music does not "come out right in the end" if you don't change metre. For example just before the piano interlude in "All By Myself":

```
|F| Am7 |Cm/Eb D7(4-3)|Gm . Bbm A7|A7 Fm/Ab G7 . |
```

At the end it might sound like `|(2/4) Gm |(3/4) Bbm A7 . |(3/4) Fm/Ab G7 . |`, but since the whole thing fits just fine in 4/4, leave it in 4/4.

If there is a pause in the music (say, of a duration of at least one beat), use the following notation at the end of a line of the transcription: `&pause`. This happens for example before the last chorus of "All by myself". Elsewhere in the song there are brief rallentandos; these don't have to be signalled at all in the transcription.

To save you having to enter the same chord or chord progression many times, you can use the notation `x2, x4` etc. This applies to *everything* previous on the line, and shouldn't be followed by anything else on the same line. So in "I'm Your Boogie Man",

```
|Gm7|Gm7 C|x4
```

represents the eight-bar phrase, which we decide belongs on one line of the file:

```
|Gm7|Gm7 C|Gm7|Gm7 C|Gm7|Gm7 C|Gm7|Gm7 C|
```

Don't do this to repeat a whole line of transcription though, when the repeats belong on other lines - instead, cut and paste the original line however many times you need it.

Using this system for repetitions and elisions will sometimes end up looking awkward; that's ok. See this passage in Jackie Blue, where the last bar of a 16-bar section is replaced by four bars on the expected chord instead of just one bar:

```
|Ebm|Abm7|x7
|Ebm|->
|Abm7|x4
```

When there are two chords entered in a bar, the default interpretation will be to place them on beats 1 and 3. Syncopated anticipations (e.g. a new harmony on the last eighth) should be ignored; just put them on the next downbeat. When chords occur on other beats than 1 and 3, use dots to fill out the measure on each beat that doesn't have a chord change. For example: `| Dm7 G7 . . |` would indicate a measure that had chords on beats 1 and 2, and nothing new until the next measure. For the Latin pattern ♩.♩.♩, just un-syncopate it, e.g. `|Cmaj7 . Fm7 Bb7 |`. We won't ever have to consider any chords at the sub-quarter-note level.

If there's ever any ambiguity as to what constitues the quarter-note, use the back-beat (almost universally on 2 and 4) to help you decide.

Harmony

Each chord is separated from the chords before and after it either by a barline or a space. You don't need to leave spaces around a chord when it is the only thing in a measure (but it doesn't matter if you do put them in, or if you put more than one space in). Likewise for any other symbols, like a temporary key signature or a dot representing a beat with no change of harmony: they need to be separated from the other symbols by a space if there is no barline.

We want the transcribed changes to represent what is actually played on the recording, including what the main bass note is. As much as ordinary chord symbols can tell someone what to play to capture the recorded performance, that's what we'd like - not simply what the changes would be for the purposes of soloing over them. So a V chord in F major, if it's simply played as a triad, should be written as C, not C7, even though the implied background scale at this point would indicate a dominant seventh.

Many songs are entirely or almost entirely triadic, with barely a 7th, and will pose no problems whatsoever. But a few songs with more complex chords or progressions will require extra care. A difficult problem is how specific you should be with your chord labels. We'd like you to use the guideline that extensions to the basic triad should only be notated if they appear consistently in the instrumental parts. So for example don't notate an `E6` in "Heart of Glass" when the singer

has C# above the band's E major triad -- she always resolves it down to B within that chord, and there is no C# in the instrumental background. If extensions to the triad in the instrumental part are only transient, or they are only present in a solo instrumental line or vocal line, don't notate them. If you think the same music happens sometimes with and sometimes without the extensions or added notes, then consider them incidental, and leave them out. Only include them in the transcription if they are consistently present, and if they form part of the harmonic background, or if they are "iconic" in the context of that song. I have found it useful to imagine having to play a harmonic pad in the background that would do no violence to the song -- notate what the harmonic pad would be.

Here are some symbols we used in our own transcriptions; we can accomodate variant systems quite easily if you have a preference for something else (but please only use ascii characters so the text files remain as portable as possible).

```
C
Cm
Cdim
Caug or C+5
C7
Cm7
Cm7b5
Cdim7
Cmaj7
CmMaj7
C6
Cm6
C/E
C7/E
C/Bb
```

if the bass note is "passing" rather than an essential part of the harmony, in which case `C7/Bb` is acceptable.

Sometimes you'll have to decide between chords that contain the same notes, like `Cm6` and `Am7b5`. Sometimes the bass note helps; for example: `C Cm6 G`, versus `Am7b5 D7 G`. But sometimes it's hard to decide; e.g. in "All by Myself" `Bbm6/F` could have been `Gm7b5/F` or even `Gdim/F`. Ordinary 7th chords should be notated as 7th chords rather than triads plus a bass note, like in some pop fake books. That is, `Dm7` rather than `F/D` and `Fmaj7` rather than `Am/F`.

For chords without a 3rd, and chords with other extensions, we've used these:

`Csus` or `Csus4` ("sus" by itself will always imply a 4th present instead of a 3rd)
`C4-3` (a resolved suspension, *not* notated as two separate chords, like "`Csus C`")
`C7(4-3)` (here we need parentheses becase C74-3 would be difficult to read)
`Csus2` (by which we mean the 3rd is *replaced* by a 2nd or 9th, so C,D,G or C,G,D)
`Cadd2` or `Cadd9` (by which we mean the 2nd or 9th is *added* (and is substantially present in the instrumental parts, as explained above), so C,D,E,G)
`C7sus`
`C7sus2` or `Gm/C`
`C11` or `Bb/C` or `Gm7/C` or `C9sus` (=C,(G),Bb,D,F; the "`C11`" notation is a pop music convention and basically means `Bb/C`, sometimes with a G thrown in too, either in the top voice or elsewhere -- it doesn't mean a complete eleventh chord built of stacked thirds i.e. C-E-G-Bb-D-F)

And use higher extensions or other alterations as required -- when these occur, use jazz conventions (e.g. `E7#9`, not something like "`E7(add m3)`").

Whatever notation you use, never include any spaces within a chord symbol. This will help us to parse the file afterwards. You can use parentheses if you need to separate elements of the symbol, for example `C7(4-3)` instead of `C74-3` which might look confusing.

When we analyse the progressions afterwards, we'll sometimes collapse these chords into larger categories - for example we might have a `C7sus` category that includes more specific variants like `C7sus`, `Gm/C` and `C11`. But because it might be useful at some later point to have the separate information, try to include the specific chord used when possible.

When there is no chord present, use the symbol `NC`. For a single bass note with no chord above it, put the bass note in square brackets: `[C]`. If there is an open fifth, with no third present, use either `Cno3` , or for a power chord you could use `C5`. (If the third is present in the vocal part, but not in the instrumental part, you can indicate the quality of the triad as if it included the vocal note--see "All by myself" when the voice comes in, providing the third of the otherwise open 5th.)

If the bassline is active, with various melodic embellishments, try not to overanalyse: the bass line as defined by the changes you write need not always contain everything the bass player plays, especially when the bass part plays a melodic role. You would include the half note moves

in something like `|D D/F#|Gm Gm/F|C/E C7|F` , but arpeggiating between the root and the 5th isn't the kind of thing you'd need to notate. Selectively leaving out bass notes and testing for plausibility will help you decide what the "real" bass line is. See "Will you still love me tomorrow"--in the bars notated `G7/D`, the bass arpeggiates up to the root of the chord later in the measure. In this case it's pretty easy to say the D is more important as a bass note than the G, because:

- The D is on the downbeat
- The D is lowest in register
- The D connects the roots of the C and E chords before and after it

That's why I wrote G7/D instead of G7. But sometimes it will be less obvious than this and you'll have to weigh these and similar factors to judge what a hypothetical unembellished bassline should be.

For passages that are not well represented by chord symbols for whatever reason, include a descriptive comment, then simply enter the right number of barlines, with each measure containing a star to represent unspecified chords. This happens in "All by myself", for example:

```
|F| Am |Cm/Eb D7(4-3)|Gm . Bbm A7|A7 Fm/Ab G7 . |
#NB: 4 measures piano interlude
|*|*|*|*|
|Fno3|x2
```

Here it is not that the harmonic progression is impossible to identify, or even difficult to identify--rather that it seems the actual contrapuntal realisation of the piano part is essential to the music, and chord symbols would not capture enough of the performance to be useful.

If the whole song is poorly represented by chord symbols, don't even try. We won't be vetting songs before serving them to you, so you have to let us know when making a transcription would be pointless. This could be because the song is completely harmonically static-- "I'm Your Boogie Man" is borderline in this regard, for example. If you get a rap song with no chord progressions for most of the song, or a complex instrumental or vocal arrangement that you couldn't adequately perform a cover of using ordinary chords on a piano or guitar, then don't transcribe it.

To sum up, the transcriptions should be detailed when the details count musically, but you should leave off the details when they are not important musically. And you need to be vigilant for potential problem spots--we can always talk about those spots together to decide what to do, as long as you are always using your ears to notice potential problems in the first place, and you don't just put down the first thing that crosses your mind. Being aware of potential issues that we haven't thought of, so we can figure out together how to deal with them, is just as important as matching the same "answer" we imagined for each transcription.

Comments

If there is a fade out in the music, put a `#fade out` line before the first section of the transcription that begins to fade out--see all the examples appended.

If you want to put markers in the file to orient yourself while you're typing in the transcription, for example `#intro , #verse , #chorus , #solo , #bridge` ; you can, and please leave them in afterwards as they might be useful to us one day.

A line beginning with #NB: contains a comment. You can comment anywhere in the file; if something should be footnoted, you can use curly brackets enclosing a numeral as reference at the end of a line of transcription. For example in "I'm Your Boogie Man", some of the lines containing `|Gm7|x8` are followed by `{1}`. At the bottom of the file is the line:
`#NB {1}: alternatively, |Gm7 Bb C F|x8.`

Please don't overuse the practice of alternative readings in a song; we prefer that you think hard about which is better, and just go with that. In this case it seems the harmonic background, which is what we're really interested in, remains as Gm7, but the foreground chords in the horns could be important enough that you feel they need to be mentioned. Use a footnote in cases like this.

If anything in the songs seems interesting/perplexing/weird enough to draw our attention to, that would also be a good use of a comment in the file. Of course it will also be possible, especially early on in the project as we all figure out the best transcription practices to apply, to discuss with us, or with one another, about what seem to be the most useful solutions for novel or problematic situations.

Examples of transcriptions

```
#Heart of Glass
#Blondie
#4/4
#key: E
|NC|NC|NC|NC|
|E|E|E|E|
|E|C#m|C#m|E|
|E|C#|C#m|->
|E|E|E|E|
|E|C#m|C#m|E|
```

```
|E|C#|C#m|->
|E|E|
|A|A|E|E|
|A|A|F#|B|
|E|E|E|E|
|E|C#m|C#m|E|
|E|C#|C#m|->
|E|E|
|A|A|E|E|
|A|A|E|E|
|A|(3/4) A|E|(3/4) E|
|A|(3/4) A|E|E|
|A|A|E|E|
|A|A|F#|B|
|E|E|E|E|
|E|E|E|E|
|A|A|E|E|
|A|A|F#|B|
#fade out
|A|A|E|E|
|A|A|F#|B|
##

#Ordinary World
#Duran Duran
#4/4
#key: E? B?
|B|F#m7|Dsus2 A/C#|E|
|B|F#m7|Dsus2 A/C#|Am6/C|
|C#m|E F#sus7|
|C#m|E F#sus7|
|C#m|E F#sus7|
|C#m|E F#sus7|
|C#m|G#m|D#7|E|
|B|F#m7|Dsus2 A/C#|E|
|B|F#m7|Dsus2 A/C#|Am6/C|
|C#m|E F#sus7|
|C#m|E F#sus7|
|C#m|E F#sus7|
|C#m|E F#sus7|
|C#m|E F#sus7|
|C#m|G#m|D#7|E|
|B|F#m7|Dsus2 A/C#|E|
|B|F#m7|Dsus2 A/C#|E|
|B|F#m7|Dsus2 A/C#|E|
|B|F#m7|Dsus2 A/C#|Am6/C|
|C#m|E F#sus7|
|C#m|E F#sus7|
|C#m|E F#sus7|
|C#m|E F#sus7|
|B|F#m7|Dsus2 A/C#|E|
|B|F#m7|Dsus2 A/C#|E|
|B|F#m7|Dsus2 A/C#|E|
|B|F#m7|Dsus2 A/C#|E|
```

```
|B|F#m7|Dsus2 A/C#|E|
|B|F#m7|Dsus2 A/C#|E|
#fade out
|B|F#m7|Dsus2 A/C#|E|
##

#I'm your Boogie Man
#K.C. and the sunshine band
#4/4
#key: Gm
|Gm7|x8
|Gm7|x8
|Gm7|x8 #{1}
|Eb|F|
|Gm7|Gm7 C|x4
|Gm7|x8
|Gm7|x8 #{1}
|Eb|F|
|Gm7|Gm7 C|x4
|Gm7|x8
|Gm7|x8 #{1}
|Eb|F|
|Gm7|Gm7 C|x8
|Gm7|x12
|Gm7|x8 #{1}
#fade out
|Gm7|x8
##
#NB {1}: alternatively |Gm7 Bb C F| for each of the 8 bars in these lines.

#All by myself
#Eric Carmen
#4/4
#key: F
|Fno3|x2
|F|Bbm6/F F|Am7b5/Eb D7(4-3)|Gm Gm7b5/Bb|F/A . Gm7b5 C7/E|
|F|Bbm6/F F|Am7b5/Eb D7(4-3)|Gm Gm7b5/Bb|F/A F7/A D7 .|G7 Gm7b5/Db C7 .|
|F| Am7 |Cm/Eb D7(4-3)|Gm . Gm7b5/Db C7|
|F| Am7 |Cm/Eb D7(4-3)|Gm . Gm7b5/Db C7|
|F|Bbm6/F F|Am7b5/Eb D7(4-3)|Gm Gm7b5/Bb |F/A . Gm7b5 C7/E |
|F|Bbm6/F F|Am7b5/Eb D7(4-3)|Gm Gm7b5/Bb|F/A F7/A D7  .|G7 Gm7b5/Db C7 .|
|F| Am7 |Cm/Eb D7(4-3)|Gm . Gm7b5/Db C7|
|F| Am7 |Cm/Eb D7(4-3)|Gm . Bbm A7|A7 Fm/Ab G7 . |
#NB: 4 measures piano interlude
|*|*|*|*|
|Fno3|x2
|F|Bbm6/F F|Am7b5/Eb D7|Gm Gm7b5/Bb|F/A . Gm7b5 C7/E| &pause
|F| Am7 |Cm/Eb D7(4-3)|Gm . Gm7b5/Db C7|
|F| Am7 |Cm/Eb D7(4-3)|Gm . Gm7b5/Db C7|
|F| Am7 |Cm/Eb D7(4-3)|Gm . Gm7b5/Db C7|
#fade out
|F| Am7 |Cm/Eb D7(4-3)|Gm . Gm7b5/Db C7|
##
```

```
#The Ozark Mountain Daredevils
#Jackie Blue
#4/4
|Ebm7|Abm7|x2
|Ebm7|Abm7|x4
|G C|x2
|Dm|Cmaj7|
|G C|x2
|Dm|Cmaj7|
|Ebm7|Abm7|x4
|G C|x2
|Dm|Cmaj7|
|G C|x2
|Dm|Cmaj7|
|Ebm7|Abm7|x4
|Ebm7|Abm7|x7
|Ebm7|->
|Abm7|x4
|Ebm7|Abm7|x8
#fade out
|Ebm7|Abm7|x8
##

#Will you love me tomorrow
#The Shirelles
#4/4
#key: C
|C|x4
|C|C|F|G|
|C|C|G7/D|G7/D|
|E|E|Am|Am|
|F|G|C|C|
|C|C|F|G|
|C|C|G7/D|G7/D|
|E|E|Am|Am|
|F|G|C|C|
|F|F|Em|Em|
|F|F|C|C|
|F|F|Em|Em|
|Am|D7|F|G|
|C|C|F|G|
|C|C|G7/D|G7/D|
|E|E|Am|Am|
|F|G|C|C|
|C|C|F|G|
|C|C|G7/D|G7/D|
|E|E|Am|Am|
|F|G|C|C|
|F|G|C|C|
#fade out
|F|G|C|C|
##
```

B ACCESSING THE CORPUS

On 27 October 2011, at the ISMIR conference in Miami, Florida, we made approximately half of the *Billboard* transcriptions available to the public at http://www.billboard.music.mcgill.ca (Burgoyne, Wild & Fujinaga 2011). The released set includes annotations and features for 649 slots and comprises 545 distinct songs. We will release the remaining data progressively over the next two years in order to ensure that there are unseen data available for the MIR community to use for evaluating algorithms as part of MIREX or related events.

Each slot appears in the archive as a numbered folder containing three files:

echonest.json, which contains the output of the EchoNest analyzer version 3.01a,* the same as the Million-Song Dataset (Bertin-Mahieux et al. 2011);

nnls_chroma.csv, which contains the output of the Vamp plugin for computing non-negative-least-squares chroma (Mauch & Dixon 2010a), with default settings except for a rolloff of one percent, as recommended for pop; and

salami_chords.txt, which contains our annotations.

We provide echonest.json and nnls_chroma.csv as an aid to engineers interested in working with audio; copyright law prevents us from sharing

* http://developer.echonest.com/

the audio files directly. The annotation files are standardised somewhat relative to the raw format described in appendix A and contain considerable extra information about larger musical structures (see Smith et al. 2011).

Each annotation begins with a header including the title of the song (prefixed by `# title:`), the name of the artist (prefixed by `# artist:`), the metre (prefixed by `# metre:`), and the tonic pitch class of the opening key (prefixed by `# tonic:`). Similar metre and tonic comments may also appear in the main body of the annotations, corresponding to changes of key or metre. In some cases, there is no obviously prevailing key, in which case the tonic pitch class is denoted `?`.

The main body of each annotation consists of a single line for each musical phrase or other sonic element at a comparable level of musical structure, equivalent to the line breaks in the original annotations. Each line begins with a floating-point number denoting the timestamp of the beginning of the phrase (in seconds), followed by a tab character. There are special lines for silence at the beginning and end of the audio file and a special line for the end of the piece. The other lines continue with a comma-separated list of elements among the following.

Capital letters, possibly followed by an arbitrary number of primes, designate high-level musical structures. They appear at the beginning of each high-level musical segment and are presumed to continue until the next appearance of a capital letter. When two letters match, the two high-level segments are musically similar. Other than denoting similarity, the letters themselves have no intrinsic meaning, but for the letter `Z`. `Z` denotes non-musical passages in the audio such as

noise or spoken words.

Plain text strings denote more traditional names for musical structures, e.g., verse, chorus, and bridge. The vocabulary was semi-restricted, but annotators had the freedom to use whatever terms they felt were most appropriate for unusual contexts.

Chord annotations appear as series of bars flanked by pipes (|). A phrase may by followed by an x and an integer, which means that the phrase is repeated that number of times. A phrase may also be followed by an arrow (->), which is a musicological hint that the phrase is musically elided into the following phrase.

Leading instruments are noted in songs where there is a notable deviation from the norm of a leading vocal throughout the entire song. They appear as text strings preceded by a left parenthesis in the segment where the instrument comes to prominence and as text strings succeeded by a right parenthesis in the segment where that instrument fades from prominence. If an instrument is prominent for a single segment only, its name appears with both left and right parentheses.

The chord annotations are simplified to the beat level. All chord symbols follow the standard presented at ISMIR 2005 and used in MIREX since (Harte et al. 2005), with a few additions to the shorthand to facilitate the richness of these annotations: `5` for power chords, and `sus2`, `maj11`, `11`, `min11`, `maj13`, `13`, and `min13` for the corresponding chords in traditional jazz notation. An additional pseudo-chord type of `1` denotes bass notes

with no chord on top. To save space, repeated chords are denoted with a dot instead of the full chord name. To further save space, bars containing a single chord on all beats list the chord symbol only once; likewise, in quadruple metres ($\binom{4}{4}$ or $\binom{12}{8}$), bars with only two chords and the change on the third beat list those two chords with no dots. For brief changes of metre, the metre may appear in parentheses at the beginning of the bar rather than as a full metre comment.

Two non-chord symbols may appear within bars. For passages that were too musically elaborate to merit beat-level chord annotations, annotators sometimes filled the bar with an asterisk. For brief pauses of arbitrary length (often a single beat), annotators added a bar with the special annotation `&pause`.

A sample of one of these annotations (number 0040) follows.

```
# title: The Power
# artist: Snap
# metre: 4/4
# tonic: B

0.0           silence
0.255419501   A, intro, | F♯:1 | F♯:1 |, (synthesizer)
4.690045351   B, transition, | B:min | B:min | B:min | B:min |, (voice
13.467097505  C, pre-chorus, | B:min | B:min | B:min | B:min |
22.253061224  | F♯:min/11 | F♯:min/11 | F♯:min/11 | F♯:min/11 |
31.040884353  D, pre-verse, | B:min | B:min | B:min | B:min |
39.837687074  E, verse, | B:min | B:min | B:min | B:min |
48.650770975  | B:min | B:min | B:min | B:min |, voice)
57.449614512  | B:min | B:min |, (saxophone)
61.801496598  C, chorus, | B:min | B:min | B:min | B:min |, (voice
```

```
70.607346938   | F#:min/11 | F#:min/11 | F#:min/11 | F#:min/11 |
79.416938775   D, pre-verse, | B:min | B:min | B:min | B:min |, voice)
88.204285714   A, intro, | F#:1 | F#:1 |, (synthesizer)
92.587981859   F, chorus, | B:min | B:min | B:min | B:min |, (voice
101.429229024  | F#:min/11 | F#:min/11 | F#:min/11 | F#:min/11 |
110.207120181  | B:min | B:min | B:min | B:min |
119.009659863  | F#:min/11 | F#:min/11 | F#:min/11 | F#:min/11 |, voice)
127.770204081  G, interlude, | B:min | B:min | B:min | B:min |, (synthesizer)
136.602154195  D, pre-verse, | B:min | B:min | B:min | B:min |, (voice
145.363356009  D, pre-verse, | B:min | B:min | B:min | B:min |, voice)
154.178231292  G, interlude, | B:min | B:min |, (synthesizer)
158.551111111  E, verse, | B:min | B:min | B:min | B:min |, (voice
167.389705215  | B:min7 | B:min7 | B:min | B:min |, voice)
176.155147392  A, interlude, | F#:1 | F#:1 |, (synthesizer)
180.604058956  F, chorus, | B:min | B:min | B:min | B:min |, (voice
189.378888888  | F#:min/11 | F#:min/11 | F#:min/11 | F#:min/11 |
198.161428571  | B:min | B:min | B:min | B:min |
206.937981859  | F#:min/11 | F#:min/11 | F#:min/11 | F#:min/11 |
215.771088435  | B:min |, voice)
217.826394557  silence
220.891428571  end
```

BIBLIOGRAPHY

Abdallah, Samer A., Katy Noland, Mark B. Sandler, Michael Casey & Christophe Rhodes. 2005. Theory and evaluation of a Bayesian music structure extractor. In *Proceedings of the 6th International Conference on Music Information Retrieval*, edited by Joshua D. Reiss & Geraint A. Wiggins, pp. 420–25. London, England.

Abdallah, Samer A., Mark B. Sandler, Christophe Rhodes & Michael Casey. 2006. Using duration models to reduce fragmentation in audio segmentation. *Machine Learning* 65 (2–3): 485–515. doi:10.1007/s10994-006-0586-4.

Adams, Norman, Daniela Marquez & Gregory H. Wakefield. 2005. Iterative deepening for melody alignment and retrieval. In *Proceedings of the 6th International Conference on Music Information Retrieval*, edited by Joshua D. Reiss & Geraint A. Wiggins, pp. 199–206. London, England.

Adler, Robert J. & Jonathan E. Taylor. 2007. *Random Fields and Geometry*. Springer Monographs in Mathematics. New York, NY: Springer. doi:10.1007/978-0-387-48116-6.

Agresti, Alan & Brent A. Coull. 1998. Approximate is better than 'exact' for interval estimation of binomial proportions. *American Statistician* 52 (2): 119–26. doi:10.2307/2685469.

Aho, Alfred V., Ravi Sethi & Jeffrey D. Ullman. 1986. *Compilers: Principles, Techniques, and Tools*. Reading, MA: Addison-Wesley.

Aitchison, John. 1982. The statistical analysis of compositional data. *Journal of the Royal Statistical Society*, ser. B, 44 (2): 139–77. http://www.jstor.org/stable/2345821.

Aldwell, Edward & Carl Schachter. 2003. *Harmony and Voice Leading*. 3rd ed. South Melbourne, Australia: Thomson Schirmer.

Allan, Moray & Christopher K. I. Williams. 2005. Harmonising chorales by probabilistic inference. In *Advances in Neural Information Processing Systems*, edited by Lawrence K. Saul, Yair Weiss & Léon Bottou, vol. 17, pp. 25–32. Cambridge, MA: MIT Press.

Ames, Charles. 1989. The Markov process as a compositional model: A survey and tutorial. *Leonardo* 22 (2): 175–87. http://www.jstor.org/stable/1575226.

Anagnostopoulou, Christina & Chantal Buteau. 2010. Can computational music analysis be both musical and computational? *Journal of Mathematics and Music* 4 (2): 75–83. doi:10.1080/17459737.2010.520455.

Antonopoulos, Iasonas, Aggelos Pikrakis, Sergios Theodoridis, Olmo Cornelis, Dirk Moelants & Marc Leman. 2007. Music retrieval by rhythmic similarity applied on Greek and African traditional music. In *Proceedings of the 8th International Conference on Music Information Retrieval*, edited by Simon Dixon, David Bainbridge & Rainer Typke, pp. 297–300. Vienna, Austria.

Babbitt, Milton. 1965. The use of computers in musicological research. *Perspectives of New Music* 3 (2): 74–83. http://www.jstor.org/stable/832505.

Bailey, B. J. R. 1980. Large sample simultaneous confidence intervals for the multinomial probabilities based on transformations of the cell frequencies. *Technometrics* 22 (4): 583–89. http://www.jstor.org/stable/1268196.

Baird, Henry S. 2003. Digital libraries and document image analysis. In *Proceedings of the iapr 7th International Conference on Document Analysis and Recognition*. Edinburgh, Scotland. Keynote address.

Barlow, Harold & Sam Morgenstern. 1948. *A Dictionary of Musical Themes*. New York, NY: Crown. All themes in this collection have since been digitised and are available at http://www.multimedialibrary.com/barlow/index.asp.

———. 1950. *A Dictionary of Vocal Themes*. New York, NY: Crown.

Batlle, Eloi & Pedro Cano. 2000. Automatic segmentation for music classification using competitive hidden Markov models. In *Proceedings of the 1st International Conference on Music Information Retrieval*, edited by Donald Byrd. Plymouth, MA.

Baum, Leonard E. & Ted Petrie. 1966. Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics* 37 (6): 1554–63. http://www.jstor.org/stable/2238772.

Baum, Leonard E., Ted Petrie, George Soules & Norman Weiss. 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics* 41 (1): 164–71. http://www.jstor.org/stable/2239727.

Bello, Juan Pablo. 2007. Audio-based cover song retrieval using approximate chord sequences: Testing shifts, gaps, swaps, and beats. In *Proceedings of the 8th International Conference on Music Information Retrieval*, edited by Simon Dixon, David Bainbridge & Rainer Typke, pp. 239–44. Vienna, Austria.

Bello, Juan Pablo, Laurent Daudet, Samer A. Abdallah, Chris Duxbury, Mike Davies & Mark B. Sandler. 2005. A tutorial on onset detection in music signals. *ieee Transactions on Speech and Audio Processing* 13 (5): 1035–47. doi:10.1109/TSA.2005.851998.

Bello, Juan Pablo, Giuliano Monti & Mark B. Sandler. 2000. Techniques for automatic music transcription. In *Proceedings of the 1st International Conference on Music Information Retrieval*, edited by Donald Byrd. Plymouth, MA.

Bello, Juan Pablo & Jeremy Pickens. 2005. A robust mid-level representation for harmonic content in music signals. In *Proceedings of the 6th International Conference on Music Information Retrieval*, edited by Joshua D. Reiss & Geraint A. Wiggins, pp. 304–11. London, England.

Bent, Ian & John Morehen. 1977–78. Computers in the analysis of

music. *Proceedings of the Royal Musical Association* 104: 30–46. http://www.jstor.org/stable/766053.

Beran, Jan. 2004. *Statistics in Musicology*. Interdisciplinary Statistics. Boca Raton, FL: Chapman & Hall/CRC.

Berenzweig, Adam L. & Daniel P. W. Ellis. 2001. Locating singing voice segments within music signals. In *Proceedings of the ieee Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 119–22. New Paltz, NY. doi:10.1109/ASPAA.2001.969557.

Bertin-Mahieux, Thierry, Daniel P. W. Ellis, Brian Whitman & Paul Lamere. 2011. The million song dataset. In *Proceedings of the 12th International Conference on Music Information Retrieval*, edited by Colby Leider & Anssi P. Klapuri, pp. 591–96. Miami, FL.

Bharucha, Jamshed J. 1987. Music cognition and perceptual facilitation: A connectionist framework. *Music Perception* 5 (1): 1–30. http://www.jstor.org/stable/40285384.

———. 1999. Neural nets, temporal composites, and tonality. In *The Psychology of Music*, edited by Diana Deutsch, 2nd ed., chap. 11, pp. 413–40. New York, NY: Academic Press.

Bharucha, Jamshed J. & Peter M. Todd. 1989. Modeling the perception of tonal structure with neural nets. *Computer Music Journal* 13 (4): 44–53. http://www.jstor.org/stable/3679552.

Bhattacharjee, Sudip, Ram D. Gopal, James R. Marsden & Rahul Telang. 2008. A survival analysis of albums on ranking charts. In *Peer-to-Peer Video: The Economics, Policy, and Culture of Today's New Mass Medium*, edited by Eli M. Noam & Lorenzo Maria Pupillo, pp. 181–204. New York, NY: Springer. doi:10.1007/978-0-387-76450-4_8.

Bigand, Emmanuel, Pierre Perruchet & Maud Boyer. 1998. Implicit learning of an artificial grammar of musical timbres. *Cahiers de Psychologie Cognitive* 17 (3): 577–600.

Billboard Magazine. 2008. Hot 100 50th anniversary charts faq. http://www.billboard.com/specials/hot100/charts/hot100faq.shtml.

Billingsley, Patrick. 1995. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics, 3rd ed. New York, NY: Wiley.

Bilmes, Jeff A. 2003. Buried Markov models: A graphical-modeling approach to automatic speech recognition. *Computer Speech and Language* 17 (2–3): 213–31. doi:10.1016/S0885-2308(03)00010-X.

———. 2006. What hmms can do. *ieice Transactions on Information and Systems* e89-d (3): 869–91. doi:10.1093/e89-d.3.869.

Bishop, Christopher M. 1995. *Neural Networks for Pattern Recognition*. Oxford, England: Oxford University Press.

Blackburn, Steven & David De Roure. 1998. A tool for content based navigation of music. In *Proceedings of the 6th acm International Conference on Multimedia*, pp. 361–68. Bristol, England. doi:10.1145/290747.290802.

Bradlow, Eric T. & Peter S. Fader. 2001. A Bayesian lifetime model for the 'Hot 100' Billboard songs. *Journal of the American Statistical Association* 96 (454): 368–81. doi:10.1198/016214501753168091.

Brown, Helen, David Butler & Mari Riess Jones. 1994. Musical and temporal influences on key discovery. *Music Perception* 11 (4): 371–407. http://www.jstor.org/stable/40285632.

Bruder, I., A. Finger, A. Heuer & T. Ignatova. 2003. Towards a digital document archive for historical handwritten music scores. In *Digital Libraries: Technology and Management of Indigenous Knowledge for Global Access*, no. 2911 in Lecture Notes in Computer Science, pp. 411–14. Berlin, Germany: Springer. doi:10.1007/978-3-540-24594-0_41.

Burgoyne, John Ashley, Laurent Pugin, Corey Kereliuk & Ichiro Fujinaga. 2007. A cross-validated study of modelling strategies for automatic chord recognition in audio. In *Proceedings of the 8th International Conference on Music Information Retrieval*, edited by Simon Dixon, David Bainbridge & Rainer Typke, pp. 251–54. Vienna, Austria.

Burgoyne, John Ashley & Lawrence K. Saul. 2005. Learning harmonic relationships in digital audio with Dirichlet-based hidden Markov models. In *Proceedings of the 6th International Conference on Music Information Retrieval*, edited by Joshua D. Reiss & Geraint A. Wiggins, pp. 438–43. London, England.

Burgoyne, John Ashley, Jonathan Wild & Ichiro Fujinaga. 2011. An expert ground-truth set for audio chord recognition and music analysis.

In *Proceedings of the 12th International Conference on Music Information Retrieval*, edited by Colby Leider & Anssi P. Klapuri. Miami, FL.

Butler, David. 1989. 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 (3): 219–41. http://www.jstor.org/stable/40285588.

Cambouropoulos, Emilios. 2003. Pitch spelling: A computational model. *Music Perception* 20 (4): 411–29. doi:10.1525/mp.2003.20.4.411.

Cano, Pedro, Alex Loscos & Jordi Bonada. 1999. Score-performance matching using hmms. In *Proceedings of the International Computer Music Conference*, pp. 441–44. Beijing, China.

Carlsen, James C. 1981. Some factors which influence melodic expectancy. *Psychomusicology* 1: 12–29.

Carnap, Rudolf. 1950. *Logical Foundations of Probability*. Chicago, IL: University of Chicago Press.

Carver, Norman. 1997. A revisionist view of blackboard systems. In *Proceedings of Midwest Artificial Intelligence and Cognitive Science Society Conference*, pp. 18–23. Dayton, Ohio.

Catteau, Benoit, Jean-Pierre Martens & Marc Leman. 2007. A probabilistic framework for audio-based tonal key and chord recognition. In *Advances in Data Analysis*, edited by Reinhold Decker & Hans-J. Lenz,

Studies in Classification, Data Analysis, and Knowledge Organization, pp. 637–44. Berlin, Germany: Springer. doi:10.1007/978-3-540-70981-7_73.

Cemgil, Ali Taylan. 2004. Bayesian music transcription. Ph.D. thesis, Radboud University Nijmegen, Nijmegen, the Netherlands.

Cemgil, Ali Taylan & Hilbert J. Kappen. 2002. Tempo tracking and rhythm quantization by sequential Monte Carlo. In *Advances in Neural Information Processing Systems*, edited by Thomas G. Dietterich, Suzanna Becker & Zoubin Ghahramani, vol. 14, pp. 1361–68. Cambridge, MA: MIT Press.

———. 2003. Monte Carlo methods for tempo tracking and rhythm quantization. *Journal of Artificial Intelligence Research* 18 (1): 45–81.

Cemgil, Ali Taylan, Hilbert J. Kappen & David Barber. 2003. Generative model based polyphonic music transcription. In *Proceedings of the ieee Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 181–84. New Paltz, NY. doi:10.1109/ASPAA.2003.1285861.

———. 2006. A generative model for music transcription. *ieee Transactions on Audio, Speech and Language Processing* 14 (2): 679–94. doi:10.1109/TSA.2005.852985.

Cemgil, Ali Taylan, Hilbert J. Kappen, Peter Desain & Henkjan Honing. 2000. On tempo tracking: Tempogram representation and Kalman filtering. *Journal of New Music Research* 29 (4): 259–73. doi:10.1080/09298210008565462.

Chai, Wei & Barry Vercoe. 2005. Detection of key change in classical piano music. In *Proceedings of the 6th International Conference on Music Information Retrieval*, edited by Joshua D. Reiss & Geraint A. Wiggins, pp. 468–73. London, England.

Chan, Ching-Hua & Elaine Chew. 2007. A hybrid system for automatic generation of style-specific accompaniment. In *Proceedings of the 4th International Joint Workshop on Computational Creativity*. London, England.

Chen, Arbee L. P., Maggie Chang, Jesse Chen, Jia-Lien Hsu, Chich-How Hsu & Spot Y. S. Hua. 2000. Query by music segments. In *Proceedings of the International Conference on Multimedia and Expo*, vol. 2, pp. 873–76. New York, NY.

Cheng, Wen-Huang, Yung-Yu Chuang, Yin-Tzu Lin, Chi-Chang Hsieh, Shao-Yen Fang, Bing-Yu Chen & Ja-Ling Wu. 2008. Semantic analysis for automatic event recognition and segmentation of wedding ceremony videos. *ieee Transactions on Circuits and Systems for Video Technology* 18 (11): 1639–50. doi:10.1109/TCSVT.2008.2005608.

Chew, Elaine & Yun-Ching Chen. 2005. Real-time pitch spelling using the spiral array. *Computer Music Journal* 29 (2): 61–76. doi:10.1162/0148926054094378.

Choi, Andrew. 2011. Jazz harmonic analysis as optimal tonality segmentation. *Computer Music Journal* 35 (2): 49–66. doi:10.1162/COMJ_a_00056.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague, the Netherlands: Mouton.

Chordia, Parag. 2005. Segmentation and recognition of tabla strokes. In *Proceedings of the 6th International Conference on Music Information Retrieval*, edited by Joshua D. Reiss & Geraint A. Wiggins, pp. 107–14. London, England.

Choudhury, G. Sayeed, Tim DiLauro, Michael Droettboom, Ichiro Fujinaga, Brian Harrington & Karl MacMillan. 2000. Optical music recognition system within a large-scale digitization project. In *Proceedings of the 1st International Conference on Music Information Retrieval*, edited by Donald Byrd. Plymouth, MA.

Coase, Ronald H. 1979. Payola in radio and television broadcasting. *Journal of Law and Economics* 22 (2): 269–328. http://www.jstor.org/stable/725120.

Coffman, Don D. 1992. Measuring musical originality using information theory. *Psychology of Music* 20 (2): 154–61. doi:10.1177/0305735692202005.

Cohen, Joel E. 1962. Information theory and music. *Behavioral Science* 7 (2): 137–63. doi:10.1002/bs.3830070202.

Collins, Michael. 2002. Discriminative training methods for hidden Markov models: Theory and experiments with perceptron algorithms. In *Proceedings of the acl-02 Conference on Empirical Methods in Natural Language Processing*, pp. 1–8. Philadelphia, PA. doi:10.3115/1118693.1118694.

Conklin, Darrell. 2006. Melodic analysis with segment classes. *Machine Learning* 65 (2–3): 349–60. doi:10.1007/s10994-006-8712-x.

———. 2010. Distinctive patterns in the first movement of Brahms' String Quartet in C Minor. *Journal of Mathematics and Music* 4 (2): 85–92. doi:10.1080/17459737.2010.515421.

Conklin, Darrell & Mathieu Bergeron. 2008. Feature set patterns in music. *Computer Music Journal* 32 (1): 60–70. doi:10.1162/comj.2008.32.1.60.

Conklin, Darrell & Ian H. Witten. 1995. Multiple viewpoint systems for music prediction. *Journal of New Music Research* 24 (1): 51–73. doi:10.1080/09298219508570672.

Cont, Arshia. 2006. Realtime audio to score alignment for polyphonic music instruments, using sparse non-negative constraints and hierarchical hmms. In *Proceedings of the ieee International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 245–48. Toulouse, France. doi:10.1109/ICASSP.2006.1661258.

———. 2010. A coupled duration-focused architecture for real-time music-to-score alignment. *ieee Transactions on Pattern Analysis and Machine Intelligence* 32 (6): 974–87. doi:10.1109/TPAMI.2009.106.

Cont, Arshia, Diemo Schwarz & Norbert Schnell. 2005. Training ircam's score follower. In *Proceedings of the ieee International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 253–56. Philadelphia, PA. doi:10.1109/ICASSP.2005.1415694.

Cook, Nicholas. 2005. Towards the compleat musicologist? Invited talk, Sixth Annual Conference on Music Information Retrieval, London, England. http://ismir2005.ismir.net/proceedings/cook.pdf.

Cook, Nicholas & Eric F. Clarke. 2004. Introduction: What is empirical musicology? In *Empirical Musicology: Aims, Methods, Prospects*, edited by Eric F. Clarke & Nicholas Cook, pp. 3–14. New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195167498.003.0001.

Cope, David. 1991. *Computers and Musical Style.* No. 6 in The Computer Music and Digital Audio Series. Madison, WI: A-R Editions.

———. 2005. *Computer Models of Musical Creativity.* Cambridge, MA: MIT Press.

Cortes, Corinna & Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning* 20 (3): 273–97. doi:10.1007/BF00994018.

Crerar, M. Alison. 1985. Elements of a statistical approach to the question of authorship in music. *Computers and the Humanities* 19 (3). doi:10.1007/BF02259533.

Cuddy, Lola L. & Carole A. Lunney. 1995. Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Attention, Perception, and Psychophysics* 57 (4): 451–62. doi:10.3758/BF03213071.

Culpepper, Sarah Elizabeth. 2010. Musical time and information theory entropy. Master's thesis, University of Iowa, Iowa City, IA.

Cunha, Uraquitan Sidney & Geber Ramalho. 1999. An intelligent hybrid model for chord prediction. *Organised Sound* 4 (2): 115–19. doi:10.1017/S1355771899002071.

Curran, James M., Christopher M. Triggs, John Buckleton & B. S. Weir. 1999. Interpreting dna mixtures in structured populations. *Journal of Forensic Sciences* 44 (5): 987–95.

Dannenberg, Roger B. 1984. An on-line algorithm for real-time accompaniment. In *Proceedings of the International Computer Music Conference*, pp. 193–98. Paris, France.

de Clercq, Trevor & David Temperley. 2011. A corpus analysis of rock harmony. *Popular Music* 30 (1): 47–70. doi:10.1017/S026114301000067X.

de Finetti, Bruno. 1970. *Teoria delle probabilità: Sintesi introduttiva con appendice critica*. 2 vols. Torino, Italy: Giulio Einaudi. Translated by Antonio Machí and Adrian Smith as *Theory of Probability: A Critical Introductory Treatment*, 2 vols., Wiley Series in Probability and Mathematical Statistics (London, England: Wiley, 1994–95).

———. 1972. *Probability, Induction, and Statistics: The Art of Guessing*. Wiley Series in Probability and Mathematical Statistics. London, England: Wiley.

Deutsch, Diana. 1980. The processing of structured and unstructured tonal sequences. *Attention, Perception, and Psychophysics* 28 (5): 381–89. doi:10.3758/BF03204881.

Deutsch, Diana & John Feroe. 1981. The internal representation of pitch sequences in tonal music. *Psychological Review* 88 (6): 503–22. doi:10.1037/0033-295X.88.6.503.

Devroye, Luc, László Györfi & Gábor Lugosi. 1996. *A Probabilistic Theory of Pattern Recognition*. No. 31 in Applications of Mathematics. New York, NY: Springer.

Dienes, Zoltán & Christopher Longuet-Higgins. 2004. Can musical transformations be implicitly learned? *Cognitive Science* 28 (4): 531–58. doi:10.1016/j.cogsci.2004.03.003.

Dietterich, Thomas G. 2002. Machine learning for sequential data: A review. In *Structural, Syntactic, and Statistical Pattern Recognition*, no. 2396 in Lecture Notes in Computer Science, pp. 227–46. Berlin, Germany: Springer. doi:10.1007/3-540-70659-3_2.

Dixon, Simon. 2005. Live tracking of musical performances using on-line time warping. In *Proceedings of the 8th International Conference on Digital Audio Effects*, pp. 92–97. Madrid, Spain.

Dixon, Simon & Gerhard Widmer. 2005. match: A music alignment tool chest. In *Proceedings of the 6th International Conference on Music Information Retrieval*, edited by Joshua D. Reiss & Geraint A. Wiggins, pp. 492–97. London, England.

Doob, Joseph Leo. 1934. Probability and statistics. *Transactions of the American Mathematical Society* 36 (4): 759–75. doi:10.1090/S0002-9947-1934-1501765-1.

Downie, J. Stephen. 2008. The music information retrieval evaluation exchange (2005–2007): A window into music information retrieval research. *Acoustical Science and Technology* 29 (4): 247–55. doi:10.1250/ast.29.247.

Downie, J. Stephen & Michael Nelson. 2000. Evaluation of a simple and effective music information retrieval method. In *Proceedings of the 23rd Annual International acm sigir Conference on Research and Development in Information Retrieval*, pp. 73–80. Athens, Greece. doi:10.1145/345508.345551.

Durey, Adriane Swalm & Mark A. Clements. 2001. Melody spotting using hidden Markov models. In *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval*, edited by J. Stephen Downie & David Bainbridge, pp. 109–17. Bloomington, IN.

Eck, Douglas & Jürgen Schmidhuber. 2002. Finding temporal structure in music: Blues improvisation with LSTM recurrent network. In *Proceedings of the 12th ieee Workshop on Neural Networks for Signal Processing*, pp. 747–56. Martigny, Switzerland. doi:10.1109/NNSP.2002.1030094.

Eggink, Jana & Guy J. Brown. 2004. Extracting melody lines from complex audio. In *Proceedings of the 5th International Conference on Music Information Retrieval*, edited by Claudia Lomelí Buyoli & Ramon Loureiro, pp. 84–91. Barcelona, Spain.

Ellis, Daniel P. W. & Graham E. Poliner. 2006. Classification-

based melody transcription. *Machine Learning* 65 (2–3): 439–56. doi:10.1007/s10994-006-8373-9.

———. 2007. Identifying 'cover songs' with chroma features and dynamic programming beat tracking. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 1429–32. Honolulu, HI. doi:10.1109/ICASSP.2007.367348.

Engelmore, Robert & Anthony J. Morgan, eds. 1988. *Blackboard Systems*. Insight Series in Artificial Intelligence. Reading, MA: Addison-Wesley.

Forte, Allen. 1966a. Music and computing: The present situation. *Computers and the Humanities* 2 (1): 32–35. doi:10.1007/BF02402463.

———. 1966b. A program for the analytical reading of scores. *Journal of Music Theory* 10 (2): 330–64. http://www.jstor.org/stable/843247.

———. 1983. Motivic design and structural levels in the first movement of Brahms's String Quartet in C Minor. *The Musical Quarterly* 69 (4): 471–502. doi:10.1093/mq/LXIX.4.471.

Freeman, Linton C. & Alan P. Merriam. 1956. Statistical classification in anthropology: An application to ethnomusicology. *American Anthropologist* 58 (3): 464–72. doi:10.1525/aa.1956.58.3.02a00060.

Fremerey, Christian, Michael Clausen, Sebastian Ewert & Meinard Müller. 2009. Sheet music-audio identification. In *Proceedings of the 10th International Society for Music Information Retrieval Con-*

ference, edited by Keiji Hirata, George Tzanetakis & Kazuyoshi Yoshii, pp. 645–50. Tokyo, Japan.

Fremerey, Christian, Meinard Müller & Michael Clausen. 2010. Handling repeats and jumps in score-performance synchronization. In *Proceedings of the 11th International Society for Music Information Retrieval Conference*, edited by J. Stephen Downie & Remco C. Veltkamp, pp. 243–48. Utrecht, the Netherlands.

Fremerey, Christian, Meinard Müller, Frank Kurth & Michael Clausen. 2008. Automatic mapping of scanned sheet music to audio recordings. In *Proceedings of the 9th International Conference on Music Information Retrieval*, edited by Juan Pablo Bello, Elaine Chew & Douglas Turnbull, pp. 413–18. Philadelphia, PA.

Frost, Richard & John Launchbury. 1989. Constructing natural language interpreters in a lazy functional language. *Computer Journal* 32 (2): 108–21. doi:10.1093/comjnl/32.2.108.

Fucks, Wilhelm. 1962. Mathematical analysis of formal structure of music. *ire Transactions on Information Theory* 8 (5): 225–28. doi:10.1109/TIT.1962.1057746.

Gabura, A. James. 1970. Music style analysis by computer. In *The Computer and Music*, edited by Harry B. Lincoln, pp. 223–76. Ithaca, NY: Cornell University Press.

Gauldin, Robert. 1997. *Harmonic Practice in Tonal Music*. New York, NY: W. W. Norton.

Ghias, Asif, Jonathan Logan, David Chamberlin & Brian C. Smith. 1995. Query by humming – musical information retrieval in an audio database. In *Proceedings of the 3rd acm International Conference on Multimedia*, pp. 231–36. San Francisco, CA. doi:10.1145/217279.215273.

Giles, David E. 2007. Survival of the hippest: Life at the top of the Hot 100. *Applied Economics* 39 (15): 1877–87. doi:10.1080/00036840600707159.

Gillet, Olivier K. & Gaël Richard. 2003. Automatic labelling of tabla signals. In *Proceedings of the 4th International Conference on Music Information Retrieval*, edited by Holger H. Hoos & David Bainbridge. Baltimore, MD.

———. 2004. Automatic transcription of drum loops. In *Proceedings of the ieee International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 269–72. Montréal, QC. doi:10.1109/ICASSP.2004.1326815.

Gillick, Jon, Kevin Tang & Robert M. Keller. 2010. Machine learning of jazz grammars. *Computer Music Journal* 34 (3): 56–66. doi:10.1162/COMJ_a_00006.

Gjerdingen, Robert O. 1989. Using connectionist models to explore complex musical patterns. *Computer Music Journal* 13 (3): 67–75. http://www.jstor.org/stable/3680013.

———. 1990. Categorization of musical patterns by self-organizing neuronlike networks. *Music Perception* 7 (4): 339–69. http://www.jstor.org/stable/40285472.

Goodman, Leo A. 1965. On simultaneous confidence intervals for multinomial proportions. *Technometrics* 7 (2): 247–54. http://www.jstor.org/stable/1266673.

Goto, Masataka. 2000. A robust predominant-f_0 estimation method for real-time detection of melody and bass lines in CD recordings. In *Proceedings of the ieee International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 757–60. Istanbul, Turkey. doi:10.1109/ICASSP.2000.859070.

———. 2001. A predominant-f_0 estimation method for CD recordings: map estimation using EM algorithm for adaptive tone models. In *Proceedings of the ieee International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 3365–68. Salt Lake City, UT. doi:10.1109/ICASSP.2001.940380.

Goto, Masataka, Hiroki Hashiguchi, Takuichi Nishimura & Ryuichi Oka. 2002. rwc music database: Popular, classical, and jazz music databases. In *Proceedings of the 3rd International Conference on Music Information Retrieval*, edited by Michael Fingerhut, pp. 287–88. Paris, France.

———. 2003. rwc music database: Music genre database and musical instrument sound database. In *Proceedings of the 4th International Conference on Music Information Retrieval*, edited by Holger H. Hoos & David Bainbridge, pp. 229–30. Baltimore, MD.

Griffith, Niall. 1994. Connectionist visualisation of tonal structure. *Artificial Intelligence Review* 8 (5–6): 393–408. doi:10.1007/BF00849727.

Hainsworth, Stephen. 2003. Beat tracking with particle filtering algorithms. In *Proceedings of the ieee Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 91–94. New Paltz, NY. doi:10.1109/ASPAA.2003.1285827.

Hájek, Alan. 1997. 'Mises redux' – redux: Fifteen arguments against finite frequentism. *Erkenntnis* 45 (2–3): 209–27. doi:10.1007/BF00276791.

———. 2009. Fifteen arguments against hypothetical frequentism. *Erkenntnis* 70 (2): 211–35. doi:10.1007/s10670-009-9154-1.

———. 2010. Interpretations of probability. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, spring 2010 ed. http://plato.stanford.edu/archives/spr2010/entries/probability-interpret/.

Harte, Christopher A., Mark B. Sandler, Samer A. Abdallah & Emilia Gómez. 2005. Symbolic representation of musical chords: A proposed syntax for text annotations. In *Proceedings of the 6th International Conference on Music Information Retrieval*, edited by Joshua D. Reiss & Geraint A. Wiggins, pp. 66–71. London, England.

Hastie, Trevor, Robert Tibshirani & Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, 2nd ed. New York, NY: Springer. doi:10.1007/978-0-387-84858-7.

Hazan, Amaury, Ricard Marxer, Paul M. Brossier, Hendrik Purwins, Perfecto Herrera & Xavier Serra. 2009. What/when causal expectation modelling applied to audio signals. *Connection Science* 21 (2–3): 119–43. doi:10.1080/09540090902733764.

Hild, Hermann, Johannes Feulner & Wolfram Menzel. 1992. harmonet: A neural net for harmonizing chorales in the style of J. S. Bach. In *Advances in Neural Information Processing Systems*, edited by John E. Moody, Steve J. Hanson & Richard P. Lippmann, vol. 4, pp. 267–74. San Francisco, CA: Morgan Kaufmann.

Hiller, Lejaren & Calvert Bean. 1966. Information theory analyses of four sonata expositions. *Journal of Music Theory* 10 (1): 96–137. http://www.jstor.org/stable/843300.

Hiller, Lejaren & Ramon Fuller. 1967. Structure and information in Webern's Symphonie, op. 21. *Journal of Music Theory* 11 (1): 60–115. http://www.jstor.org/stable/842949.

Hofstetter, Fred T. 1979. The nationalistic fingerprint in nineteenth-century Romantic chamber music. *Computers and the Humanities* 13 (2): 105–19. doi:10.1007/BF02404506.

Hoos, Holger H., Kai Renz & Marko Görg. 2001. guido/mir – an experimental musical information retrieval system based on guido music notation. In *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval*, edited by J. Stephen Downie & David Bainbridge, pp. 41–50. Bloomington, IN.

von Hornbostel, Erich M. 1906. Phonographierte tunesische Melodien. *Sammelbände der Internationalen Musikgesellschaft* 8 (1): 1–43. http://www.jstor.org/stable/929108.

Hou, Chia-Ding, Jengtung Chiang & John Jen Tai. 2003. A family of simultaneous confidence intervals for multinomial proportions. *Computational Statistics and Data Analysis* 43 (1): 29–45. doi:10.1016/S0167-9473(02)00169-X.

Hsu, Jia-Lien, Chih-Chin Liu & Arbee L. P. Chen. 2001. Discovering nontrivial repeating patterns in music data. *ieee Transactions on Multimedia* 3 (3): 311–25. doi:10.1109/6046.944475.

Hu, Ning & Roger B. Dannenberg. 2002. A comparison of melodic database retrieval techniques using sung queries. In *Proceedings of the 2nd acm–ieee-cs Joint Conference on Digital Libraries*, pp. 301–7. Portland, OR. doi:10.1145/544220.544292.

———. 2005. A bootstrap method for training an accurate audio segmenter. In *Proceedings of the 6th International Conference on Music Information Retrieval*, edited by Joshua D. Reiss & Geraint A. Wiggins, pp. 223–29. London, England.

Hu, Ning, Roger B. Dannenberg & George Tzanetakis. 2003. Polyphonic audio matching and alignment for music retrieval. In *Proceedings of the ieee Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 185–88. New Paltz, NY. doi:10.1109/ASPAA.2003.1285862.

Huron, David. 1990a. Crescendo/decrescendo asymmetries in Beethoven's piano sonatas. *Music Perception* 7 (4): 395–402. http://www.jstor.org/stable/40285475.

———. 1990b. Increment/decrement asymmetries in polyphonic sonorities. *Music Perception* 7 (4): 385–94. http://www.jstor.org/stable/40285474.

———. 1995. *Unix Software Tools for Music Research: The Humdrum Toolkit Reference Manual*. Menlo Park, CA: Centre for Computer Assisted Research in the Humanities. http://www.humdrum.org.

———. 1996. The melodic arch in Western folksongs. *Computing in Musicology* 10: 3–23.

———. 1999. The new empiricism: Systematic musicology in a postmodern age. Ernest Bloch Lecture 3, University of California, Berkeley, California. http://www.music-cog.ohio-state.edu/Music220/Bloch.lectures/3.Methodology.html.

———. 2001. What is a musical feature? Forte's analysis of Brahms opus 51, no. 1 revisited. *Music Theory Online* 7 (4).

———. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.

Huron, David & Ann Ommen. 2006. An empirical study of syncopation in American popular music, 1890–1939. *Music Theory Spectrum* 28 (2): 211–31. doi:10.1525/mts.2006.28.2.211.

Jeppesen, Knud. 1923. *Palestrinastil med særligt henblik paa dissonansbehandlingen.* Copenhagen, Denmark: Levin & Munksgaard. http://hdl.handle.net/2027/heb.06336.0001.001. Translated by Margaret N. Hamerik as *The Style of Palestrina and the Dissonance* (London, England: Oxford University Press, 1927).

Joder, Cyril, Slim Essid & Gaël Richard. 2009. Temporal integration for audio classification with application to musical instrument classification. ieee *Transactions on Audio, Speech and Language Processing* 17 (1): 174–86. doi:10.1109/TASL.2008.2007613.

———. 2010. A conditional random field viewpoint of symbolic audio-to-score matching. In *Proceedings of the International Conference on Multimedia*, pp. 871–74. Florence, Italy. doi:10.1145/1873951.1874100.

Jonaitis, Erin McMullen & Jenny R. Saffran. 2009. Learning harmony: The role of serial statistics. *Cognitive Science* 33 (5): 951–68. doi:10.1111/j.1551-6709.2009.01036.x.

Jones, Mari Riess. 1976. Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review* 83 (5): 323–55. doi:10.1037/0033-295X.83.5.323.

Jones, Mari Riess, Marilyn Boltz & Gary Kidd. 1982. Controlled attending as a function of melodic and temporal context. *Attention, Perception, and Psychophysics* 32 (3): 211–18. doi:10.3758/BF03206225.

Jordanous, Anna & Alan Smaill. 2009. Investigating the role of score following in automatic musical accompaniment. *Journal of New Music Research* 38 (2): 197–209. doi:10.1080/09298210903180245.

Juang, Biing-Hwang. 1984. On the hidden Markov model and dynamic time warping for speech recognition – a unified view. *AT&T Bell Laboratories Technical Journal* 63 (7): 1213–43.

Jürgensen, Frauke. 2005. Accidentals in the mid-fifteenth century: A computer-aided study of the Buxheim organ book and its concordances. Ph.D. thesis, McGill University, Montréal, QC.

Kálmán, Rudolph Emil. 1960. A new approach to linear filtering and prediction problems. *Transactions of the ACME*, ser. D, 82: 35–45.

Kan, Min-Yen, Ye Wang, Denny Iskandar, Tin Lay Nwe & Arun Shenoy. 2008. Lyrically: Automatic synchronization of textual lyrics to acoustic musical signals. *IEEE Transactions on Audio, Speech and Language Processing* 16 (2): 338–49. doi:10.1109/TASL.2007.911559.

Kapanci, Emir & Avi Pfeffer. 2005. Signal to score music transcription using graphical models. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pp. 758–85. Edinburgh, Scotland.

Kashino, Kunio & Hiroshi Murase. 1998. Music recognition using note transition context. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, pp. 3593–96. Seattle, WA. doi:10.1109/ICASSP.1998.679655.

Katok, Anatole & Boris Hasselblatt. 1995. *Introduction to the Modern Theory of Dynamical Systems*. No. 54 in Encyclopedia of Mathematics and Its Applications. Cambridge, England: Cambridge University Press.

Keshet, Joseph, Shai Shalev-Shwartz, Yoram Singer & Dan Chazan. 2007. A large margin algorithm for speech-to-phoneme and music-to-score alignment. *ieee Transactions on Audio, Speech and Language Processing* 15 (8): 2373–82. doi:10.1109/TASL.2007.903928.

Khadkevich, Maksim & Maurizio Omologo. 2009. Use of hidden Markov models and factored language models for automatic chord recognition. In *Proceedings of the 10th International Society for Music Information Retrieval Conference*, edited by Keiji Hirata, George Tzanetakis & Kazuyoshi Yoshii, pp. 561–66. Tokyo, Japan.

Klapuri, Anssi P., Antti J. Eronen & Jaako T. Astola. 2006. Analysis of the meter of acoustic musical signals. *ieee Transactions on Audio, Speech and Language Processing* 14 (1): 342–55. doi:10.1109/TSA.2005.854090.

Kohonen, Teuvo. 2001. *Self-Organizing Maps*. No. 30 in Springer Series in Information Sciences, 3rd ed. Berlin, Germany: Springer.

Koller, Daphne & Nir Friedman. 2009. *Probabilistic Graphical Models: Principles and Techniques*. Adaptive Computation and Machine Learning. Cambridge, MA: MIT Press.

Kolmogorov, Andrey N. 1933. *Grundbegriffe der Wahrscheinlichkietsrechnung*. No. 2.3 in Ergebnisse der Mathematik und ihrer Grenzgebiete.

Berlin, Germany: Springer. Translated by Nathan Morrison as *Foundations of the Theory of Probability* (New York, N.Y.: Chelsea, 1950).

Kopec, Gary E. & Philip A. Chou. 1996. Markov source model for printed music decoding. *Journal of Electronic Imaging* 5 (1): 7–14. doi:10.1117/12.227527.

Kostka, Stefan & Dorothy Payne. 1995. *Tonal Harmony*. 3rd ed. New York, NY: McGraw-Hill.

Krige, Willie, Theo Herbst & Thomas Niesler. 2008. Explicit transition modelling for automatic singing transcription. *Journal of New Music Research* 37 (4): 311–24. doi:10.1080/09298210902890299.

Krumhansl, Carol L. 1990a. *Cognitive Foundations of Musical Pitch*. No. 17 in Oxford Psychology Series. New York, NY: Oxford University Press.

———. 1990b. Tonal hierarchies and rare intervals in music cognition. *Music Perception* 7 (3): 309–24. http://www.jstor.org/stable/40285467.

Krumhansl, Carol L. & Edward J. Kessler. 1982. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review* 89 (4): 334–68. doi:10.1037/0033-295X.89.4.334.

Krumhansl, Carol L., Pekka Toivanen, Tuomas Eerola, Petri Toiviainen, Topi Järvinen & Jukka Louhivuori. 2000. Cross-

cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition* 76 (1): 13–58. doi:10.1016/S0010-0277(00)00068-8.

Kruskal, Joseph B. 1983. An overview of sequence comparison: Time warps, string edits, and macromolecules. *siam Review* 25 (2): 201–37. http://www.jstor.org/stable/2030214.

Kurth, Frank, Meinard Müller, David Damm, Christian Fremerey, Andreas Ribbrock & Michael Clausen. 2005. SyncPlayer – an advanced system for multimodel music access. In *Proceedings of the 6th International Conference on Music Information Retrieval*, edited by Joshua D. Reiss & Geraint A. Wiggins, pp. 381–88. London, England.

Kurth, Frank, Meinard Müller, Christian Fremerey, Yoon-ha Chang & Michael Clausen. 2007. Automated synchronization of scanned sheet music with audio recordings. In *Proceedings of the 8th International Conference on Music Information Retrieval*, edited by Simon Dixon, David Bainbridge & Rainer Typke, pp. 261–66. Vienna, Austria.

Lafferty, John, Andrew McCallum & Fernando Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the International Conference on Machine Learning*, pp. 282–89. Williamstown, MA.

Lang, Dustin & Nando de Freitas. 2005. Beat tracking the graphical model way. In *Advances in Neural Information Processing Systems*, edited by Lawrence K. Saul, Yair Weiss & Léon Bottou, vol. 17, pp. 745–52. Cambridge, MA: MIT Press.

Larson, Steve. 2004. Musical forces and melodic expectations: Comparing computer models and experimental results. *Music Perception* 21 (4): 457–98. doi:10.1525/mp.2004.21.4.457.

Lavrenko, Victor & Jeremy Pickens. 2003. Polyphonic music modeling with random fields. In *Proceedings of the 11th ACM International Conference on Multimedia*, edited by Lawrence A. Rowe, Harrick M. Vin, Thomas Plagemann, Prashant J. Shenoy & John R. Smith, pp. 120–29. Berkeley, CA. doi:10.1145/957013.957041.

Layton, Martin. 2006. Augmented statistical models for classifying sequence data. Ph.D. thesis, Corpus Christi College, University of Cambridge.

Lee, Kyogu. 2008a. A system for acoustic chord transcription and key extraction from audio using hidden Markov models trained on synthesized audio. Ph.D. thesis, Stanford University, Stanford, CA.

———. 2008b. A system for automatic chord transcription from audio using genre-specific hidden Markov models. In *Adaptive Multimedia Retrieval: Retrieval, User, and Semantics*, edited by Nozha Boujemaa, Marcin Detyniecki & Andreas Nürnberger, no. 4918 in Lecture Notes in Computer Science, pp. 136–46. Berlin, Germany: Springer. doi:10.1007/978-3-540-79860-6_11.

Lee, Kyogu & Malcolm Slaney. 2006. Automatic chord recognition from audio using an HMM with supervised learning. In *Proceedings of the 7th International Conference on Music Information Retrieval*, edited by Kjell

Lemström, Adam Tindale & Roger B. Dannenberg, pp. 133–37. Victoria, BC.

———. 2007. A unified system for chord transcription and key extraction using hidden Markov models. In *Proceedings of the 8th International Conference on Music Information Retrieval*, edited by Simon Dixon, David Bainbridge & Rainer Typke, pp. 245–50. Vienna, Austria.

———. 2008. Acoustic chord transcription and key extraction from audio using key-dependent HMMs trained on synthesized audio. IEEE *Transactions on Audio, Speech and Language Processing* 16 (2): 291–301. doi:10.1109/TASL.2007.914399.

Leistikow, Randal J. 2006. Bayesian modeling of musical expectations via maximum entropy stochastic grammars. Ph.D. thesis, Stanford University, Stanford, CA.

Lerdahl, Fred. 2001. *Tonal Pitch Space.* Oxford, England: Oxford University Press.

Lerdahl, Fred & Ray Jackendoff. 1977. Toward a formal theory of tonal music. *Journal of Music Theory* 21 (1): 111–71. http://www.jstor.org/stable/843480.

———. 1983. *A Generative Theory of Tonal Music.* Cambridge, MA: MIT Press.

Leroux, Brian G. 1992. Maximum-likelihood estimation for hidden Markov models. *Stochastic Processes and Their Applications* 40 (1): 127–43. doi:10.1016/0304-4149(92)90141-C.

Levinson, S. E. 1986. Continuously variable duration hidden Markov models for automatic speech recognition. *Computer Speech and Language* 1: 29–45. doi:10.1016/S0885-2308(86)80009-2.

Levy, Mark, Katy Noland & Mark B. Sandler. 2007. A comparison of timbral and harmonic music segmentation algorithms. In *Proceedings of the ieee International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 1433–36. Honolulu, HI. doi:10.1109/ICASSP.2007.367349.

Levy, Mark & Mark B. Sandler. 2008. Structure segmentation of musical audio by constrained clustering. *ieee Transactions on Audio, Speech and Language Processing* 16 (2): 318–26. doi:10.1109/TASL.2007.910781.

Lewin, David. 1968. Some applications of communication theory to the study of twelve-tone music. *Journal of Music Theory* 12 (1): 50–84. http://www.jstor.org/stable/842886.

Liu, XiaoBing, DeShun Yang & XiaoOu Chen. 2008. New approach to classification of Chinese folk music based on extension of hmm. In *Proceedings of the International Conference on Audio, Language, and Image Processing*, pp. 1172–79. Shanghai, China. doi:10.1109/ICALIP.2008.4590068.

Loui, Psyche, David L. Wessel & Carl L. Hudson Kam. 2010. Humans rapidly learn grammatical structure in a new musical scale. *Music Perception* 27 (5): 377–88. doi:10.1525/mp.2010.27.5.377.

MacMillan, Karl, Michael Droettboom & Ichiro Fujinaga. 2002. Gamera: Optical music recognition in a new shell. In *Proceedings of the International Computer Music Conference*, pp. 482–85. Gothenburg, Sweden.

Macrae, Robert & Simon Dixon. 2010. Accurate real-time windowed time warping. In *Proceedings of the 11th International Society for Music Information Retrieval Conference*, edited by J. Stephen Downie & Remco C. Veltkamp, pp. 423–28. Utrecht, the Netherlands.

Maddage, Namunu, Changsheng Xu, Mohan S. Kankanhalli & Xi Shao. 2004. Content-based music structure analysis with applications to music semantics understanding. In *Proceedings of the 12th acm International Conference on Multimedia*, pp. 112–19. New York, NY. doi:10.1145/1027527.1027549.

Magalhães, José Pedro & W. Bas de Haas. 2011. Experience report: Functional modelling of musical harmony. Tech. Rep. UU-CS-2011-007, Department of Information and Computing Sciences, Utrecht University, Utrecht, the Netherlands.

Maher, Patrick. 2006. The concept of inductive probability. *Erkenntnis* 65 (2): 185–206. doi:10.1007/s10670-005-5087-5.

———. 2010a. Bayesian probability. *Synthese* 172: 119–27. doi:10.1007/s11229-009-9471-6.

———. 2010b. What is probability? Unpublished book draft, 27 August 2010. http://patrick.maher1.net/preprints/pop.pdf.

Manzara, Leonard C., Ian H. Witten & Mark James. 1992. On the entropy of music: An experiment with Bach chorale melodies. *Leonardo Music Journal* 2 (1): 81–88. http://www.jstor.org/stable/1513213.

Margaritis, Dimitris. 2003. Learning Bayesian network model structure from data. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA.

Margulis, Elizabeth Hellmuth & Andrew P. Beatty. 2008. Musical style, psychoaesthetics, and prospects for entropy as an analytic tool. *Computer Music Journal* 32 (4): 64–78. doi:10.1162/comj.2008.32.4.64.

Marillier, Cecil G. 1983. Computer assisted analysis of tonal structures in the Classical symphony. In *The Haydn Yearbook*, edited by H. C. Robbins Landon, I. M. Bruce & David Wyn Jones, vol. 14, pp. 187–200. Cardiff, Wales: University College Cardiff Press.

Martellotti, Anna. 2001. Finitely additive phenomena. *Rendiconti dell'Istituto di Matematica dell'Università di Trieste* 33: 201–49.

Mauch, Matthias. 2010. Automatic chord transcription from audio using computational models of musical context. Ph.D. thesis, Queen Mary, University of London, London, England.

Mauch, Matthias & Simon Dixon. 2008. A discrete mixture model for chord labelling. In *Proceedings of the 9th International Conference on Music Information Retrieval*, edited by Juan Pablo Bello, Elaine Chew & Douglas Turnbull, pp. 45–50. Philadelphia, PA.

———. 2010a. Approximate note transcription for the improved identification of difficult chords. In *Proceedings of the 11th International Society for Music Information Retrieval Conference*, edited by J. Stephen Downie & Remco C. Veltkamp, pp. 135–40. Utrecht, the Netherlands.

———. 2010b. Simultaneous estimation of chords and musical context from audio. *ieee Transactions on Audio, Speech and Language Processing* 18 (6): 1280–89. doi:10.1109/TASL.2009.2032947.

Mauch, Matthias, Simon Dixon, Christopher A. Harte, Michael Casey & Benjamin Fields. 2007. Discovering chord idioms through Beatles and Real Book songs. In *Proceedings of the 8th International Conference on Music Information Retrieval*, edited by Simon Dixon, David Bainbridge & Rainer Typke, pp. 255–58. Vienna, Austria.

Mavromatis, Panayotis. 2006. A hidden Markov model of melody production in Greek church chant. In *Music Analysis East and West*, no. 14 in Computing in Musicology, pp. 93–112. Cambridge, MA: MIT Press.

———. 2009. Minimum description length modelling of musical structure. *Journal of Mathematics and Music* 3 (3): 117–36. doi:10.1080/17459730903313122.

Maxwell, H. John. 1992. An expert system for harmonizing analysis of tonal music. In *Understanding Music with AI: Perspectives on Music Cognition*, edited by Mira Balaban, Kemal Ebcioğlu & Otte E. Laske, pp. 335–53. Cambridge, MA: AAAI Press.

Mazzoni, Dominic & Roger B. Dannenberg. 2001. Melody matching directly from audio. In *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval*, edited by J. Stephen Downie & David Bainbridge, pp. 17–18. Bloomington, IN.

McCallum, Andrew, Dayne Freitag & Fernando Pereira. 2000. Maximum entropy Markov models for information extraction and segmentation. In *Proceedings of the 17th International Conference on Machine Learning*, pp. 591–98. Stanford, CA.

McCullagh, Peter & John A. Nelder. 1989. *Generalized Linear Models*. No. 37 in Monographs on Statistics and Applied Probability, 2nd ed. Boca Raton, FL: Chapman & Hall/CRC.

McKinney, Martin F., Dirk Moelants, Matthew E. P. Davies & Anssi P. Klapuri. 2007. Evaluation of audio beat tracking and music tempo extraction algorithms. *Journal of New Music Research* 36 (1): 1–16. doi:10.1080/09298210701653252.

McNab, Rodger J., Lloyd A. Smith, Ian H. Witten, Clare L. Henderson & Sally Jo Cunningham. 1996. Towards the digital music library: Tune retrieval from acoustic input. In *Proceedings of the*

1st ACM International Conference on Digital Libraries, pp. 11–18. Bethesda, MD. doi:10.1145/226931.226934.

Mearns, Lesley, Dan Tidhar & Simon Dixon. 2010. Characterisation of composer style using high-level musical features. In *Proceedings of the 3rd International Workshop on Machine Learning and Music*, pp. 37–40. Florence, Italy. doi:10.1145/1878003.1878016.

Meehan, James R. 1980. An artificial intelligence approach to tonal music theory. *Computer Music Journal* 4 (2): 60–65. http://www.jstor.org/stable/3680083.

Meek, Colin & William P. Birmingham. 2002. Johnny can't sing: A comprehensive error model for sung music queries. In *Proceedings of the 3rd International Conference on Music Information Retrieval*, edited by Michael Fingerhut, pp. 124–32. Paris, France.

Meeùs, Nicolas. 2003. Vecteurs harmoniques. *Musurgia* 10 (3–4): 3–34.

Mendel, Arthur. 1969. Some preliminary attempts at computer-assisted style analysis in music. *Computers and the Humanities* 4 (1): 41–52. doi:10.1007/BF02393450.

Meredith, David. 2006. The *ps13* pitch spelling algorithm. *Journal of New Music Research* 35 (2): 121–59. doi:10.1080/09298210600834961.

———. 2007. Optimizing Chew and Chen's pitch-spelling algorithm. *Computer Music Journal* 31 (2): 54–72. doi:10.1162/comj.2007.31.2.54.

Meyer, Leonard B. 1957. Meaning in music and information theory. *Journal of Aesthetics and Art Criticism* 15 (4): 412–24. http://www.jstor.org/stable/427154.

———. 1989. *Style and Music: Theory, History, and Ideology*. Studies in the Criticism and Theory of Music. Philadelphia, PA: University of Pennsylvania Press.

Miotto, Riccardo & Nicola Orio. 2007. A methodology for the segmentation and identification of music works. In *Proceedings of the 8th International Conference on Music Information Retrieval*, edited by Simon Dixon, David Bainbridge & Rainer Typke, pp. 273–78. Vienna, Austria.

———. 2008. A music identification system based on chroma indexing and statistical modeling. In *Proceedings of the 9th International Conference on Music Information Retrieval*, edited by Juan Pablo Bello, Elaine Chew & Douglas Turnbull, pp. 301–6. Philadelphia, PA.

von Mises, Richard. 1957. *Probability, Statistics, and Truth*. London, England: George, Allen, and Unwin. Reprint, New York, NY: Dover, 1981.

———. 1964. *Mathematical Theory of Probability and Statistics*. New York, NY: Academic Press.

Montecchio, Nicola & Nicola Orio. 2009. A discrete filter bank approach to audio to score matching for polyphonic music. In *Proceedings of the 10th International Society for Music Information Retrieval Conference*,

edited by Keiji Hirata, George Tzanetakis & Kazuyoshi Yoshii, pp. 495–500. Tokyo, Japan.

Mosimann, James E. 1962. On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions. *Biometrika* 49 (1–2). http://www.jstor.org/stable/2333468.

Müller, Meinard, Frank Kurth & Michael Clausen. 2005. Audio matching via chroma-based statistical features. In *Proceedings of the 6th International Conference on Music Information Retrieval*, edited by Joshua D. Reiss & Geraint A. Wiggins, pp. 288–95. London, England.

Müller, Meinard, Frank Kurth & Tido Röder. 2004. Towards and efficient algorithm for automatic score-to-audio synchronization. In *Proceedings of the 5th International Conference on Music Information Retrieval*, edited by Claudia Lomelí Buyoli & Ramon Loureiro, pp. 365–72. Barcelona, Spain.

Müller, Meinard, Henning Mattes & Frank Kurth. 2006. An efficient approach to audio synchronization. In *Proceedings of the 7th International Conference on Music Information Retrieval*, edited by Kjell Lemström, Adam Tindale & Roger B. Dannenberg, pp. 192–97. Victoria, BC.

Murphy, Kevin Patrick. 2002. Dynamic Bayesian networks: Representation, inference, and learning. Ph.D. thesis, University of California, Berkeley, CA.

Myers, Charles S. 1907. The ethnological study of music. In *Anthropological Essays Presented to Edward Burnett Tylor*, edited by W. H. R. Rivers, R. R. Marett & Northcote W. Thomas, pp. 235–53. Oxford, England: Clarendon.

Mysore, Gautham J. 2010. A non-negative framework for joint modeling of spectral structure and temporal dynamics in sound mixtures. Ph.D. thesis, Stanford University, Stanford, CA.

Mysore, Gautham J., Paris Smaragdis & Bhiksha Raj. 2010. Non-negative hidden Markov modeling of audio with application to source separation. In *Latent Variable Analysis and Signal Separation*, no. 6365 in Lecture Notes in Computer Science, pp. 140–48. Springer. doi: 10.1007/978-3-642-15995-4_18.

Narmour, Eugene. 1990. *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago, IL: University of Chicago Press.

———. 1992. *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model*. Chicago, IL: University of Chicago Press.

Nattiez, Jean-Jacques. 1987. *Musicologie générale et sémiologie*. Musique / Passé / Présent. Paris, France: Christian Bourgois. Translated by Carolyn Abbate as *Music and Discourse: Toward a Semiology of Music* (Princeton, NJ: Princeton University Press, 1990).

Nettheim, Nigel. 1997. A bibliography of statistical applications in musicology. *Musicology Australia* 20: 94–106.

Niedermayer, Bernhard. 2009. Improving accuracy of polyphonic music-to-score alignment. In *Proceedings of the 10th International Society for Music Information Retrieval Conference*, edited by Keiji Hirata, George Tzanetakis & Kazuyoshi Yoshii, pp. 585–90. Tokyo, Japan.

Niedermayer, Bernhard & Gerhard Widmer. 2010. A multi-pass algorithm for accurate audio-to-score alignment. In *Proceedings of the 11th International Society for Music Information Retrieval Conference*, edited by J. Stephen Downie & Remco C. Veltkamp, pp. 417–22. Utrecht, the Netherlands.

Niles, Les T. & Harvey F. Silverman. 1990. Combining hidden Markov model and neural network classifiers. In *Proceedings of the ieee International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 417–20. Albuquerque, NM. doi:10.1109/ICASSP.1990.115724.

Noland, Katy & Mark B. Sandler. 2006. Key estimation using a hidden Markov model. In *Proceedings of the 7th International Conference on Music Information Retrieval*, edited by Kjell Lemström, Adam Tindale & Roger B. Dannenberg, pp. 121–26. Victoria, BC.

Nwe, Tin Lay & Ye Wang. 2004. Automatic detection of vocal segments in popular songs. In *Proceedings of the 5th International Conference on Music Information Retrieval*, edited by Claudia Lomelí Buyoli & Ramon Loureiro, pp. 138–45. Barcelona, Spain.

Ogihara, Mistunori & Tao Li. 2008. *n*-gram chord profiles for composer style identification. In *Proceedings of the 9th International Conference*

on Music Information Retrieval, edited by Juan Pablo Bello, Elaine Chew & Douglas Turnbull, pp. 671–76. Philadelphia, PA.

Orio, Nicola & François Déchelle. 2001. Score following using spectral analysis and hidden Markov models. In *Proceedings of the International Computer Music Conference*, pp. 151–54. Havana, Cuba.

Ostendorf, Mari, Vassilios V. Digalakis & Owen A. Kimball. 1996. From hmm's to segment models: A unified view of stochastic modeling for speech recognition. ieee *Transactions on Speech and Audio Processing* 4 (5): 360–78. doi:10.1109/89.536930.

Pachet, François. 1999. Surprising harmonies. *International Journal on Computing Anticipatory Systems* 4.

Paiement, Jean-François, Douglas Eck & Samy Bengio. 2005. A probabilistic model for chord progressions. In *Proceedings of the 6th International Conference on Music Information Retrieval*, edited by Joshua D. Reiss & Geraint A. Wiggins, pp. 312–19. London, England.

Paisley, William J. 1964. Identifying the unknown communicator in painting, literature and music: The significance of minor encoding habits. *Journal of Communication* 14 (4): 219–37. doi:10.1111/j.1460-2466.1964.tb02925.x.

Paiva, Rui Pedro, Teresa Mendes & Amílcar Cardoso. 2005. On the detection of melody notes in polyphonic audio. In *Proceedings of the 6th International Conference on Music Information Retrieval*, edited by Joshua D. Reiss & Geraint A. Wiggins, pp. 175–82. London, England.

Papadopoulos, Hélène & Geoffroy Peeters. 2007. Large-scale study of chord estimation algorithms based on chroma representation and hmm. In *Proceedings of the ieee International Workshop on Content-Based Multimedia Indexing*, pp. 53–60. Bordeaux, France. doi:10.1109/CBMI.2007.385392.

———. 2008. Simultaneous estimation of chord progression and downbeats from an audio file. In *Proceedings of the ieee International Conference on Acoustics, Speech, and Signal Processing*, pp. 121–24. Las Vegas, NV. doi:http://dx.doi.org/10.1109/ICASSP.2008.4517561.

Pardo, Bryan. 2005. Probabilistic sequence alignment methods for online score following of music performances. Ph.D. thesis, University of Michigan, Ann Arbor, MI.

Pardo, Bryan & William P. Birmingham. 2002. Algorithms for chordal analysis. *Computer Music Journal* 26 (2): 27–49. doi:10.1162/014892602760137167.

Pardo, Bryan & Manan Sanghi. 2005. Polyphonic musical sequence alignment for database search. In *Proceedings of the 6th International Conference on Music Information Retrieval*, edited by Joshua D. Reiss & Geraint A. Wiggins, pp. 215–22. London, England.

Pardo, Bryan, Jonah Shifrin & William P. Birmingham. 2003. Name that tune: A pilot study in finding a melody from a sung query. *Journal of the American Society for Information Science and Technology* 55 (4): 283–300. doi:10.1002/asi.10373.

Paul, Sudhir R., Uditha Balasooriya & Tathagata Banerjee. 2005. Fisher information matrix of the Dirichlet-multinomial distribution. *Biometrical Journal* 47 (2): 230–36. doi:10.1002/bimj.200410103.

Paulus, Jouni. 2010. Improving Markov model-based music piece structure labelling wiht acoustic information. In *Proceedings of the 11th International Society for Music Information Retrieval Conference*, edited by J. Stephen Downie & Remco C. Veltkamp, pp. 303–8. Utrecht, the Netherlands.

Paulus, Jouni & Anssi P. Klapuri. 2007. Combining temporal and spectral features in HMM-based drum transcription. In *Proceedings of the 8th International Conference on Music Information Retrieval*, edited by Simon Dixon, David Bainbridge & Rainer Typke, pp. 225–28. Vienna, Austria.

———. 2009. Labelling the structural parts of a music piece with Markov models. In *Computer Music Modeling and Retrieval: Genesis of Meaning in Sound and Music*, edited by Sølvi Ystad, Richard Kronland-Martinet & Kristoffer Jensen, no. 5493 in Lecture Notes in Computer Science, pp. 166–76. Berlin, Germany: Springer. doi:10.1007/978-3-642-02518-1_11.

Paulus, Jouni K. & Anssi P. Klapuri. 2003. Conventional and periodic *n*-grams in the transcription of drum sequences. In *Proceedings of the International Conference on Multimedia and Expo*, vol. 2, pp. 737–40. Baltimore, MD. doi:10.1109/ICME.2003.1221722.

Pauwels, Johan & Jean-Pierre Martens. 2010. Integrating musicological knowledge into a probabilistic framework for chord and key

extraction. In *Proceedings of 128th Convention of the Audio Engineering Society*. London, England. http://www.aes.org/e-lib/browse.cfm?elib=15383.

Pearce, Marcus T. 2005. The construction and evaluation of statistical models of melodic structure in music perception and composition. Ph.D. thesis, City University, London, England.

Pearce, Marcus T. & Geraint A. Wiggins. 2004. Improved methods for statistical modelling of monophonic music. *Journal of New Music Research* 33 (4): 367–85. doi:10.1080/0929821052000343840.

———. 2006. Expectation in melody: The influence of context and learning. *Music Perception* 23 (5): 377–405. doi:10.1525/mp.2006.23.5.377.

Pearl, Judea. 1993. Comment: Graphical models, causality, and intervention. *Statistical Science* 8 (3): 266–69. http://www.jstor.org/stable/2245965.

———. 1995. Causal diagrams for empirical research. *Biometrika* 82 (4): 669–88. http://www.jstor.org/stable/2337329.

———. 2009. *Causality: Models, Reasoning, and Inference*. 2nd ed. New York, NY: Cambridge University Press.

Peeling, Paul, Ali Taylan Cemgil & Simon Godsill. 2007. A probabilistic framework for matching music representations. In *Proceedings of the 8th International Conference on Music Information Retrieval*, edited by Simon Dixon, David Bainbridge & Rainer Typke, pp. 267–72. Vienna, Austria.

Peeters, Geoffroy. 2006. Musical key estimation of audio signal based on hidden Markov modeling of chroma vectors. In *Proceedings of the 9th International Conference on Digital Audio Effects*, edited by Philippe Depalle, Marcelo Wanderley, Vincent Verfaille & Andrey da Silva, pp. 127–31. Montréal, QC.

Pickens, Jeremy. 2000. A comparison of language modeling and probabilistic text information retrieval approaches to monophonic music retrieval. In *Proceedings of the 1st International Conference on Music Information Retrieval*, edited by Donald Byrd. Plymouth, MA.

———. 2003. Key-specific shrinkage techniques for harmonic models. In *Proceedings of the 4th International Conference on Music Information Retrieval*, edited by Holger H. Hoos & David Bainbridge. Baltimore, MD.

Pickens, Jeremy, Juan Pablo Bello, Giuliano Monti, Tim Crawford, Matthew Dovey, Mark B. Sandler & Donald Byrd. 2002. Polyphonic score retrieval using polyphonic audio queries: A harmonic modeling approach. In *Proceedings of the 3rd International Conference on Music Information Retrieval*, edited by Michael Fingerhut, pp. 140–49. Paris, France.

Pickens, Jeremy & Costas Iliopoulos. 2005. Markov random fields and maximum entropy modeling for music information retrieval. In *Proceedings of the 6th International Conference on Music Information Retrieval*, edited by Joshua D. Reiss & Geraint A. Wiggins, pp. 207–14. London, England.

Pinkerton, Richard C. 1956. Information theory and melody. *Scientific American* 194 (2): 77–87. doi:10.1038/scientificamerican0256-77.

Piston, Walter. 1941. *Harmony*. New York, NY: W. W. Norton.

Poliner, Graham E. 2008. Classification-based music transcription. Ph.D. thesis, Columbia University, New York, NY.

Poliner, Graham E. & Daniel P. W. Ellis. 2007. A discriminative model for polyphonic piano transcription. *eurasip Journal on Advances in Signal Processing* doi:10.1155/2007/48317.

Ponsford, Dan, Geraint A. Wiggins & Chris Mellish. 1999. Statistical learning of harmonic movement. *Journal of New Music Research* 28 (2): 150–77. doi:10.1076/jnmr.28.2.150.3115.

Popper, Karl. 1959. *The Logic of Scientific Discovery*. London, England: Hutchinson. Reprint, London, England: Routledge, 1992. Author's translation and expansion of *Logik der Forschung* (Vienna: Springer, 1934).

Prather, Ronald E. 1996. Harmonic analysis from the computer representation of a musical score. *Communications of the acm* 39 (12). doi:10.1145/272682.272716.

Pugin, Laurent. 2006a. Lecture et traitement informatique de typographies musicales anciennes : Un logiciel de reconaissance de partitions par modèles de Markov cachés. Ph.D. thesis, University of Geneva, Geneva, Switzerland.

———. 2006b. Optical music recognition of early typographic prints using hidden Markov models. In *Proceedings of the 7th International Conference on Music Information Retrieval*, edited by Kjell Lemström, Adam Tindale & Roger B. Dannenberg, pp. 53–56. Victoria, BC.

Pugin, Laurent, John Ashley Burgoyne, Douglas Eck & Ichiro Fujinaga. 2007. Book-adaptive and book-dependent models to accelerate digitization of early music. In *nips Workshop on Music, Brain, and Cognition*. Whistler, BC.

Pugin, Laurent, John Ashley Burgoyne & Ichiro Fujinaga. 2007a. Goal-directed evaluation for the improvement of optical music recognition on early music prints. In *Proceedings of the acm-ieee Joint Conference on Digital Libraries*, pp. 303–4. Vancouver, BC.

———. 2007b. MAP adaptation to improve optical music recognition of early music documents using hidden Markov models. In *Proceedings of the 8th International Conference on Music Information Retrieval*, edited by Simon Dixon, David Bainbridge & Rainer Typke, pp. 513–16. Vienna, Austria.

———. 2007c. Reducing costs for digitising early music with dynamic adaptation. In *Proceedings of the European Conference on Digital Libraries*, pp. 417–74. Budapest, Hungary.

Pugin, Laurent, Jason Hockman, John Ashley Burgoyne & Ichiro Fujinaga. 2008. Gamera vs. Aruspix: Two optical music recognition approaches. In *Proceedings of the 9th International Conference on Music*

Information Retrieval, edited by Juan Pablo Bello, Elaine Chew & Douglas Turnbull, pp. 419–24. Philadelphia, PA.

Qi, Yuting, John William Paisley & Lawrence Carin. 2007. Music analysis using hidden Markov mixture models. *ieee Transactions on Signal Processing* 55 (11): 5209–24. doi:10.1109/TSP.2007.898782.

Rabiner, Lawrence R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the ieee* 77 (2): 257–86. doi:10.1109/5.18626.

Raczyński, Stanisław A., Emmanuel Vincent, Frédéric Bimbot & Shigeki Sagayama. 2010. Multiple pitch transcription using dbn-based musicological models. In *Proceedings of the 11th International Society for Music Information Retrieval Conference*, edited by J. Stephen Downie & Remco C. Veltkamp, pp. 363–68. Utrecht, the Netherlands.

Radicioni, Daniele P. & Roberto Esposito. 2010. breve: An HMPerceptron-based chord recognition system. In *Advances in Music Information Retrieval*, edited by Zbigniew W. Raś & Alicja A. Wieczorkowska, no. 274 in Studies in Computational Intelligence, pp. 143–64. Berlin, Germany: Springer. doi:10.1007/978-3-642-11674-2_7.

Rameau, Jean-Philippe. 1722. *Traité de l'harmonie réduite à ses principes naturels*. Paris, France. Translated by Philip Gossett as *Treatise on Harmony* (New York, NY: Dover, 1971).

Ramsay, J. O. & B. W. Silverman. 2005. *Functional Data Analysis*. Springer Series in Statistics. New York, NY: Springer.

Raphael, Christopher. 1999. Automatic segmentation of acoustic musical signals using hidden Markov models. *ieee Transactions on Pattern Analysis and Machine Intelligence* 21 (4): 360–70. doi:10.1109/34.761266.

———. 2001a. Automated rhythm transcription. In *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval*, edited by J. Stephen Downie & David Bainbridge, pp. 99–107. Bloomington, IN.

———. 2001b. A probabilistic system for automatic musical accompaniment. *Journal of Computational and Graphical Statistics* 10 (3): 487–512. http://www.jstor.org/stable/1391101.

———. 2001c. Synthesizing musical accompaniments with Bayesian belief networks. *Journal of New Music Research* 30 (1): 59–67. doi:10.1076/jnmr.30.1.59.7121.

———. 2002a. Automatic transcription of piano music. In *Proceedings of the 3rd International Conference on Music Information Retrieval*, edited by Michael Fingerhut, pp. 15–19. Paris, France.

———. 2002b. A Bayesian network for real-time musical accompaniment. In *Advances in Neural Information Processing Systems*, edited by Thomas G. Dietterich, Suzanna Becker & Zoubin Ghahramani, vol. 14, pp. 1433–40. MIT Press.

———. 2004. A hybrid graphical model for aligning polyphonic audio with musical scores. In *Proceedings of the 5th International Conference on Music Information Retrieval*, edited by Claudia Lomelí Buyoli & Ramon Loureiro, pp. 387–94. Barcelona, Spain.

———. 2005. A graphical model for recognizing sung melodies. In *Proceedings of the 6th International Conference on Music Information Retrieval*, edited by Joshua D. Reiss & Geraint A. Wiggins, pp. 658–63. London, England.

———. 2006. Aligning music audio with symbolic scores using a hybrid graphical model. *Machine Learning* 65 (2–3): 389–409. doi:10.1007/s10994-006-8415-3.

Raphael, Christopher & Joshua Stoddard. 2003. Harmonic analysis with probabilistic graphical models. In *Proceedings of the 4th International Conference on Music Information Retrieval*, edited by Holger H. Hoos & David Bainbridge. Baltimore, MD.

———. 2004. Functional harmonic analysis using probabilistic models. *Computer Music Journal* 28 (3): 45–52. doi:10.1162/0148926041790676.

Rebelo, Ana, G. Artur Capela & Jaime S. Cardoso. 2010. Optical recognition of music symbols. *International Journal on Document Analysis and Recognition* 13 (1): 19–31. doi:10.1007/s10032-009-0100-1.

Reinhard, Johannes, Sebastian Stober & Andreas Nürnberger. 2008. Enhancing chord classification through neighbourhood histograms. In *Proceedings of the International Workshop on Content-Based Multimedia Indexing*, pp. 33–40. London, England. doi:10.1109/CBMI.2008.4564924.

Resnick, Sidney I. 1999. *A Probability Path*. Boston, MA: Birkhäuser.

Riemann, Hugo. 1893. *Vereinfachte Harmonielehre oder die Lehre von den tonalen Funktionen der Akkorde.* London, England: Augener. Translated as *Harmony Simplified, or the Theory of the Tonal Functions of Chords* (London, England: Augener, 1895).

Rocher, Thomas, Matthias Robine, Pierre Hanna & Laurent Oudre. 2010. Concurrent estimation of chords and keys from audio. In *Proceedings of the 11th International Society for Music Information Retrieval Conference*, edited by J. Stephen Downie & Remco C. Veltkamp, pp. 141–46. Utrecht, the Netherlands.

Rohrmeier, Martin. 2011. Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music* 5 (1): 35–53. doi:10.1080/17459737.2011.573676.

Rosenthal, Jeffrey S. 2000. *A First Look at Rigorous Probability Theory.* Singapore: World Scientific.

Rosti, Antti-Veikko & Mark John Francis Gales. 2003. Switching linear dynamical systems for speech recognition. Tech. Rep. 461, Cambridge University Engineering Department, Cambridge, England.

Rowe, Robert. 2001. *Machine Musicianship.* Cambridge, MA: MIT Press.

Rubin, Donald B. 2005. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association* 100 (469): 322–31. doi:10.1198/016214504000001880.

Rudin, Walter. 1976. *Principles of Mathematical Analysis*. International Series in Pure and Applied Mathematics, 3rd ed. New York, NY: McGraw-Hill.

Ryynänen, Matti P. & Anssi P. Klapuri. 2005. Polyphonic music transcription using note event modeling. In *Proceedings of the ieee Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 319–22. New Paltz, NY. doi:10.1109/ASPAA.2005.1540233.

———. 2006. Transcription of the singing melody in polyphonic music. In *Proceedings of the 7th International Conference on Music Information Retrieval*, edited by Kjell Lemström, Adam Tindale & Roger B. Dannenberg, pp. 222–27. Victoria, BC.

———. 2008. Automatic transcription of melody, bass line, and chords in polyphonic music. *Computer Music Journal* 32 (3): 72–86. doi:10.1162/comj.2008.32.3.72.

Saffran, Jenny R., Elizabeth K. Johnson, Richard N. Aslin & Elissa L. Newport. 1999. Statistical learning of tone sequences by human infants and adults. *Cognition* 70 (1): 27–52. doi:10.1016/S0010-0277(98)00075-4.

Salzer, Felix. 1952. *Structural Hearing: Tonal Coherence in Music*. New York, NY: Boni.

Saxe, Karen. 2002. *Beginning Functional Analysis*. Undergraduate Texts in Mathematics. New York, NY: Springer.

Schellenberg, E. Glenn. 1996. Expectancy in melody: Tests of the implication-realization model. *Cognition* 58 (1): 75–125. doi:10.1016/0010-0277(95)00665-6.

———. 1997. Simplifying the implication-realization model of melodic expectancy. *Music Perception* 14 (3): 295–318. http://www.jstor.org/stable/40285723.

Schmuckler, Mark A. 1989. Expectation in music: Investigation of melodic and harmonic processes. *Music Perception* 7 (2): 109–50. http://www.jstor.org/stable/40285454.

Scholz, Ricardo & Geber Ramalho. 2008. cochonut: Recognizing complex chords from midi guitar sequences. In *Proceedings of the 9th International Conference on Music Information Retrieval*, edited by Juan Pablo Bello, Elaine Chew & Douglas Turnbull, pp. 27–32. Philadelphia, PA.

Scholz, Ricardo, Emmanuel Vincent & Frédéric Bimbot. 2009. Robust modeling of musical chord sequences using probabilistic N-grams. In *Proceedings of the ieee International Conference on Acoustics, Speech, and Signal Processing*, pp. 53–56. Taipei, Taiwan. doi:10.1109/ICASSP.2009.4959518.

Schuller, Björn, Benedikt Hörnler, Dejan Arsic & Gerhard Rigoll. 2009. Audio chord labeling by musiological modeling and beat-synchronization. In *Proceedings of the ieee International Conference on Multimedia and Expo*, pp. 526–29. Cancun, Mexico. doi:10.1109/ICME.2009.5202549.

Shalev-Shwartz, Shai, Shlomo Dubnov, Nir Friedman & Yoram Singer. 2002. Robust temporal and spectral modeling for query by melody. In *Proceedings of the 25th Annual International* acm sigir *Conference on Research and Development in Information Retrieval*, pp. 331–38. Tampere, Finland. doi:10.1145/564376.564435.

Shalev-Shwartz, Shai, Joseph Keshet & Yoram Singer. 2004. Learning to align polyphonic music. In *Proceedings of the 5th International Conference on Music Information Retrieval*, edited by Claudia Lomelí Buyoli & Ramon Loureiro, pp. 381–86. Barcelona, Spain.

Shannon, Claude Elwood & Warren Weaver. 1949. *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.

Sheh, Alexander & Daniel P. W. Ellis. 2003. Chord segmentation and recognition using EM-trained hidden Markov models. In *Proceedings of the 4th International Conference on Music Information Retrieval*, edited by Holger H. Hoos & David Bainbridge, pp. 185–91. Baltimore, MD.

Shifrin, Jonah & William P. Birmingham. 2003. Effectiveness of hmm-based retrieval on large databases. In *Proceedings of the 4th International Conference on Music Information Retrieval*, edited by Holger H. Hoos & David Bainbridge, pp. 33–39. Baltimore, MD.

Shifrin, Jonah, Bryan Pardo, Colin Meek & William P. Birmingham. 2002. hmm-based musical query retrieval. In *Proceedings of the 2nd*

ACM–IEEE-CS *Joint Conference on Digital Libraries*, pp. 295–300. Portland, OR. doi:10.1145/544220.544291.

Shumway, Robert H. & David S. Stoffer. 2011. *Time Series Analysis and Its Applications*. Springer Texts in Statistics, 3rd ed. New York, NY: Springer. doi:10.1007/978-1-4419-7865-3.

Simon, Ian, Dan Morris & Sumit Basu. 2008. MySong: Automatic accompaniment generation for vocal melodies. In *Proceedings of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems*. Florence, Italy. doi:10.1145/1357054.1357169.

Simonton, Dean Keith. 1980a. Thematic fame and melodic originality in classical music: A multivariate computer-content analysis. *Journal of Personality* 48 (2): 206–19. doi:10.1111/j.1467-6494.1980.tb00828.x.

———. 1980b. Thematic fame, melodic originality, and musical zeitgeist: A biographical and transhistorical content analysis. *Journal of Personality and Social Psychology* 38 (6): 972–83. doi:10.1037/0022-3514.38.6.972.

———. 1984. Melodic structure and note transition probabilities: A content analysis of 15,618 classical themes. *Psychology of Music* 12 (1): 3–16. doi:10.1177/0305735684121001.

Sison, Cristina P. & Joseph Glaz. 1995. Simultaneous confidence intervals and sample size determination for multinomial proportions. *Journal of the American Statistical Association* 90 (429): 366–69. http://www.jstor.org/stable/2291162.

Smith, J. David & Robert J. Melara. 1990. Aesthetic preference and syntactic prototypicality in music: 'tis the gift to be simple. *Cognition* 34 (3): 279–98. doi:10.1016/0010-0277(90)90007-7.

Smith, Jordan B. L., John Ashley Burgoyne, David De Roure & J. Stephen Downie. 2011. Design and creation of a large-scale database of structural annotations. In *Proceedings of the 12th International Conference on Music Information Retrieval*, edited by Colby Leider & Anssi P. Klapuri, pp. 555–60. Miami, FL.

Snyder, John L. 1990. Entropy as a measure of musical style: The influence of a priori assumptions. *Music Theory Spectrum* 12 (1): 121–60. http://www.jstor.org/stable/746148.

Soulez, Ferréol, Xavier Rodet & Diemo Schwarz. 2003. Improving polyphonic and poly-instrumental music to score alignment. In *Proceedings of the 4th International Conference on Music Information Retrieval*, edited by Holger H. Hoos & David Bainbridge, pp. 143–48. Baltimore, MD.

Spangler, Randall R. & Rodney M. Goodman. 1998. Bach in a box – real-time harmony. In *Advances in Neural Information Processing Systems*, edited by Michael I. Jordan, Michael J. Kearns & Sara A. Solla, vol. 10, pp. 957–63. Cambridge, MA: MIT Press.

Spirtes, Peter, Clark Glymour & Richard Scheines. 2000. *Causation, Prediction, and Search*. 2nd ed. Cambridge, MA: MIT Press.

Spława-Neyman, Jerzy. 1923. Próba uzasadnienia zastosowań rachunku prawdopodobieństwa do doświadczeń polowych. *Roczniki Nauk Rol-*

niczych 10: 1–51. http://www.jstor.org/stable/2245382. Translated by D. M. Dabrowska and T. P. Speed as 'On the Application of Probability Theory to Agricultural Experiments: Essay on Principles', *Statistical Science* 5, no. 4 (1990): 465–72. Very often cited with the French title, 'Sur les applications de la théorie des probabilités aux expériences agricoles'.

Steedman, Mark J. 1984. A generative grammar for jazz chord sequences. *Music Perception* 2 (1): 52–77. http://www.jstor.org/stable/40285282.

Stevens, Catherine & Janet Wiles. 1994. Tonal music as a componential code: Learning temporal relationships between and within pitch and timing components. In *Advances in Neural Information Processing Systems*, edited by Jack D. Cowan, Gerald Tesauro & Joshua Alspector, vol. 6, pp. 1085–92. San Francisco, CA: Morgan Kaufmann.

Suchoff, Benjamin. 1970. Computer-oriented comparative musicology. In *The Computer and Music*, edited by Harry B. Lincoln, pp. 193–206. Ithaca, NY: Cornell University Press.

Takeda, Haruto, Takuya Nishimoto & Shigeki Sagayama. 2004. Rhythm and tempo recognition of music performance from a probabilistic approach. In *Proceedings of the 5th International Conference on Music Information Retrieval*, edited by Claudia Lomelí Buyoli & Ramon Loureiro, pp. 357–64. Barcelona, Spain.

Taskar, Ben, Carlos Guestrin & Daphne Koller. 2004. Max-margin Markov networks. In *Advances in Neural Information Processing Systems*,

edited by Sebastian Thrun, Lawrence K. Saul & Bernhard Schölkopf, vol. 16, pp. 25–32. Cambridge, MA: MIT Press.

Taube, Heinrich. 1999. Automatic tonal analysis: Toward the implementation of a music theory workbench. *Computer Music Journal* 23 (4): 18–32. http://www.jstor.org/stable/3680675.

Temperley, David. 1999. The question of purpose in music theory: Description, suggestion, and explanation. *Current Musicology* 66: 66–85.

———. 2001. *The Cognition of Basic Musical Structures*. Cambridge, MA: MIT Press.

———. 2007. *Music and Probability*. Cambridge, MA: MIT Press.

———. 2009. A unified probabilistic model for polyphonic music analysis. *Journal of New Music Research* 38 (1): 3–18. doi:10.1080/09298210902928495.

———. 2011. The cadential IV in rock. *Music Theory Online* 17 (1).

Temperley, David & Elizabeth West Marvin. 2008. Pitch-class distribution and the identification of key. *Music Perception* 25 (3): 193–212. doi:10.1525/mp.2008.25.3.193.

Temperley, David & Daniel Sleator. 1999. Modeling meter and harmony: A preference-rule approach. *Computer Music Journal* 23 (1): 10–27. http://www.jstor.org/stable/3680618.

Teodoru, Gabi & Christopher Raphael. 2007. Pitch spelling with conditionally independent voices. In *Proceedings of the 8th International Conference on Music Information Retrieval*, edited by Simon Dixon, David Bainbridge & Rainer Typke, pp. 201–6. Vienna, Austria.

Terrat, Richard G. 2004. Pregroup grammars for chords. In *Proceedings of the 5th International Conference on Music Information Retrieval*, edited by Claudia Lomelí Buyoli & Ramon Loureiro, pp. 250–53. Barcelona, Spain.

Thompson, William Forde, Lola L. Cuddy & Cheryl Plaus. 1997. Expectancies generated by melodic intervals: Evaluation of principles of melodic implication in a melody-completion task. *Attention, Perception, and Psychophysics* 59 (7): 1069–76. doi:10.3758/BF03205521.

Thornburg, Harvey, Randal J. Leistikow & Jonathan Berger. 2007. Melody extraction and musical onset detection via probabilistic models of framewise stft peak data. *ieee Transactions on Audio, Speech and Language Processing* 15 (4): 1257–72. doi:10.1109/TASL.2006.889801.

Tillmann, Barbara, Jamshed J. Bharucha & Emmanuel Bigand. 2000. Implicit learning of tonality: A self-organizing approach. *Psychological Review* 107 (4): 885–913. doi:10.1037/0033-295X.107.4.885.

Tillmann, Barbara & Stephen McAdams. 2004. Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30 (5): 1131–42. doi:10.1037/0278-7393.30.5.1131.

Trowbridge, Lynn M. 1985–86. Style change in the fifteenth-century chanson: A comparative study of compositional detail. *The Journal of Musicology* 4 (2): 146–70. http://www.jstor.org/stable/763793.

Tsamardinos, Ioannis, Constantin F. Aliferis & Alexander Statnikov. 2003. Algorithms for large scale Markov blanket discovery. In *Proceedings of the 16th International Florida Artificial Intelligence Research Society Conference*, pp. 376–81. St Augustine, FL.

Tvedebrink, Torben. 2010. Overdispersion in allelic counts and θ-correction in forensic genetics. *Theoretical Population Biology* 78 (3): 200–10. doi:10.1016/j.tpb.2010.07.002.

Tymoczko, Dmitri. 2003. Progressions fondamentales, fonctions, degrés : une grammaire de l'harmonie tonale élémentaire. *Musurgia* 10 (3–4): 35–64.

Uitdenbogerd, Alexandra & Justin Zobel. 1999. Melodic matching techniques for large music databases. In *Proceedings of the 7th* acm *International Conference on Multimedia*, vol. 1, pp. 57–66. Orlando, FL. doi:10.1145/319463.319470.

Ulrich, John Wade. 1977. The analysis and synthesis of jazz by computer. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, pp. 865–72. Cambridge, MA.

Unyk, Anna M. & James C. Carlsen. 1987. The influence of melodic expectancy on melodic perception. *Psychomusicology* 7 (1): 3–23.

Vincent, Emmanuel. 2006. Musical source separation using time-frequency source priors. *ieee Transactions on Audio, Speech and Language Processing* 14 (1): 91–98. doi:10.1109/TSA.2005.860342.

Vines, Bradley W., Regina L. Nuzzo & Daniel J. Levitin. 2005. Analyzing temporal dynamics in music. *Music Perception* 23 (2): 137–52. doi:10.1525/mp.2005.23.2.137.

Vos, Piet G. & Jim M. Troost. 1989. Ascending and descending melodic intervals: Statistical findings and their perceptual relevance. *Music Perception* 6 (4): 383–96. http://www.jstor.org/stable/40285439.

Wasserman, Larry. 2004. *All of Statistics*. Springer Texts in Statistics. New York, NY: Springer.

Watt, Henry J. 1924. Functions of the size of interval in the songs of Schubert and of the Chippewa and Teton Sioux Indians. *British Journal of Psychology* 14 (4): 370–86. doi:10.1111/j.2044-8295.1924.tb00150.x.

Whiteley, Nick, Ali Taylan Cemgil & Simon Godsill. 2006. Bayesian modelling of temporal structure in musical audio. In *Proceedings of the 7th International Conference on Music Information Retrieval*, edited by Kjell Lemström, Adam Tindale & Roger B. Dannenberg, pp. 29–34. Victoria, BC.

Wiggins, Geraint A., Daniel Müllensiefen & Marcus T. Pearce. 2010. On the non-existence of music: Why music theory is a figment of the imagination. *Musicæ Scientiæ* 14: 231–55.

Winograd, Terry. 1968. Linguistics and the computer analysis of tonal harmony. *Journal of Music Theory* 12: 2–49. http://www.jstor.org/stable/842885.

Winter, Keith. 1979. Communication analysis in jazz. *Jazzforschung* 11: 93–133.

Wright, Matthew, W. Andrew Schloss & George Tzanetakis. 2008. Analyzing Afro-Cuban rhythm using rotation-aware clave template matching with dynamic programming. In *Proceedings of the 9th International Conference on Music Information Retrieval*, edited by Juan Pablo Bello, Elaine Chew & Douglas Turnbull, pp. 647–52. Philadelphia, PA.

Yaramakala, Sandeep & Dimitris Margaritis. 2005. Speculative Markov blanket discovery for optimal feature selection. In *Proceedings of the 5th ieee International Conference on Data Mining*. Houston, TX. doi:10.1109/ICDM.2005.134.

Yoshioka, Takuya, Tetsuro Kitahara, Kazunori Komatani, Tetsuya Ogata & Hiroshi G. Okuno. 2004. Automatic chord transcription with concurrent recognition of chord symbols and boundaries. In *Proceedings of the 5th International Conference on Music Information Retrieval*, edited by Claudia Lomelí Buyoli & Ramon Loureiro, pp. 100–105. Barcelona, Spain.

Young, Steve, Gunnar Evermann, Mark Gales, Thomas Hain, Dan Kershaw, Xunying Liu, Gareth Moore, Julian Odell, Dave

Ollason, Dan Povey & Phil Woodland. 2006. The htk book (for htk version 3.4). Online documentation. http://htk.eng.cam.ac.uk/.

Youngblood, Joseph E. 1958. Style as information. *Journal of Music Theory* 2 (1): 24–35. http://www.jstor.org/stable/842928.

———. 1970. Root progression and composer identification. In *The Computer and Music*, edited by Harry B. Lincoln, pp. 172–78. Ithaca, NY: Cornell University Press.

Zhang, Xinglin & David Gerhard. 2008. Chord recognition using instrument voicing constraints. In *Proceedings of the 9th International Conference on Music Information Retrieval*, edited by Juan Pablo Bello, Elaine Chew & Douglas Turnbull, pp. 33–38. Philadelphia, PA.

Zierolf, Robert. 1983. Indeterminacy in musical form. Ph.D. thesis, University of Cincinnati, Cincinnati, OH.

COLOPHON

This thesis was designed and set in LaTeX using Peter Wilson's memoir package. The fonts may seem large, which is an artefact of McGill's regulations for formatting theses. These regulations require a twelve-point font but define that as printing ten to twelve characters per inch, which differs markedly from what has come to be the understanding of twelve-point fonts in a post-typewriter age.

The main body as well as the mathematical text are set in the Rimmer Type Foundry's Amethyst; I thank Robert Bringhurst's classic *The Elements of Typographic Style* for the recommendation. It was designed by Jim Rimmer, a well-loved Canadian type designer. Amethyst lacks a full set of Greek letters, and of the fonts I had available, the Greeks from Adobe's Arno were the best match. Arno also has a large set of text ornaments, which serve as the decorations for B-level headings. The sans-serif text is set in the International Typeface Corporations's Officina Sans, which I was happy to discover complements Amethyst so well. All of the black-letter mathematics is set in Linotype Duc de Berry, which is of a style known as *bastarda* that I think makes a reasonably readable compromise between a script font and the more traditional Fraktur.