On the perception of musical form and the *shouga* of traditional Japanese music

Tanor Bonin

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

August 2019

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

Table of Contents

Abstract

Music appears in the mind of the listener exhibiting a distinct *form*. The current scientific literature on music perception generally regards the subjective experience of musical form as a psychological representation of an originally physical stimulus event. But in the act of musical creation this relationship is reversed: musical art objects (*materials*) are understood to be physical representations of the abstract forms (*musical images*) initially conceived by the creative subject (e.g., composer, improviser, performer). Taking together these diametric perspectives of the musical arts and the musical sciences, the psychological *images* and physical *materials* of musical experience are understood to relate in a dynamic confluence. A scientific investigation of the experience of musical form thus entails an investigation of the perceptual capacity to differentiate, coordinate, and relate various *physical* and *psychological* entities within the unified and coherent experience of the listening subject. The phenomenologically oriented question, "How does a musical form appear to the listener?" therefore implies the psychologically oriented question, "What are the perceptual processes underlying the listener's ability to systematically associate physical materials and psychological images?"

Chapter 1 begins the dissertation with a theoretical conceptualization of the perception of musical form, the complementarity between scientific and artistic investigations of music, and the creative process by which musicians represent *musical images* in the form of empirical, perceptible art objects. I summarize these considerations with a dual-process model in which the so-called *musical image* characterizing the internal state of mind of the subject is both representative and generative of external, empirical art objects.

Chapter 2 investigates the contribution of the phase correspondence between metrical and melodic dimensions of the musical stimulus to the perception of Japanese and Western musical

style. Previous research has demonstrated the contribution of musical tonality to the perceptual distinctiveness of Japanese and Western musical styles. The three experiments in this Chapter further demonstrate that the characteristic musical contour profiles of the traditional Japanese and Western musical styles are both perceptible by listeners of various cultural backgrounds and make additive, independent contributions to the perception of musical form.

Chapter 3 investigates a cross-modal musical transliteration device known as the *shouga* (唱歌). The *shouga* is an embodied phonetic system used to represent and articulate the intended form of a given piece of music, with particular *idioms* associated with particular formal contours. The studies of this Chapter are predicated on the observation that the empirical art object, as a consequence of human creativity, is a formal representation of the mental state of the perceiving subject who produced them. In the special case afforded by the *shouga*, expert musicians have created one art object (the *shouga*) to represent, or articulate, their phenomenological impression of another art object (the perceived form of the music itself). Empirical observations of the parametric correspondences between these two objects allows the psychologist to draw inferences regarding how the subjective impression of musical form may be related to the physical art object that elicits this impression. Specifically, the three studies of this Chapter concern the exploration and statistical parameterization of structural correspondences between the *shouga* and the seminal solo koto work *Rokudan no Shirabe* (六段の調). Results indicate considerable structural correspondence between the macro- and micro-structure of the *shouga* and the tonal, rhythmic, and metric dimensions of *Rokudan no Shirabe.*

Chapter 4 provides perceptual data regarding the ability of non-Japanese listeners who are not trained in this music to distinguish true *shouga* sequences from arbitrarily constructed strings of Japanese syllables (foils) in association with the structure of koto melodies. If the *shouga* is a coherent and intelligible representation of musical form, these participants should be able to recognize the semblance, or likeness, between the musical melody and its intended *shouga* sequence. The results of this final experiment indicate that the vast majority of non-Japanese listeners can detect the true *shouga* above chance rates, and that the majority of true *shouga* sequences used within the experiment were discernible from their foils at rates exceeding those predicted by chance.

Chapter 5 concludes the dissertation with a general discussion of the implications of the presented findings for future research into musical form, musical creativity, and the mental capacity to process abstract dynamic systems.

Résumé

Titre : La perception de la forme musicale et du *shouga* de la musique japonaise traditionnelle

La musique apparaît dans l'esprit de l'auditeur sous une forme distincte. La littérature scientifique actuelle sur la perception de la musique considère généralement l'expérience subjective de la forme musicale comme une représentation psychologique d'un événement stimulant d'origine. Mais dans l'acte de création musicale, cette relation est inversée : les objets d'art musical (*matériaux*) sont considérés comme des représentations physiques des formes abstraites (*images musicales*) initialement conçues par le créateur (par exemple, compositeur, improvisateur, interprète). En combinant ces perspectives diamétrales d'art musical et de science musicale, les images psychologiques et les matériaux physiques sont considérés comme liés par une confluence dynamique dans le contexte de l'expérience musicale. Une étude de la forme musicale coïncide ainsi avec une étude de la capacité subjective de différencier, coordonner et relier diverses entités physiques et psychologiques au sein de l'expérience unifiée et cohérente de l'auditeur. La question phénoménologique « Comment une forme musicale apparaît-elle à l'auditeur ? » implique donc la question psychologique « Quels sont les processus de perception sous-tendant la capacité de l'auditeur d'associer systématiquement des matériaux physiques et des images psychologiques ? »

Le chapitre 1 commence la thèse par une conceptualisation théorique de la perception de la forme musicale, de la complémentarité entre les recherches scientifiques et artistiques sur la musique et du processus de création par lequel les musiciens représentent des *images musicales* sous la forme d'objets d'art empiriques et perceptibles. Je résume ces considérations avec un modèle à double processus dans lequel la soi-disant image musicale caractérisant l'état d'esprit interne du sujet est à la fois représentative et générative d'objets d'art empiriques externes.

Le chapitre 2 examine l'apport de la correspondance de phase entre les dimensions métrique et mélodique du stimulus musical à la perception des styles musicaux japonais et occidentaux. Des recherches antérieures ont démontré la contribution de la tonalité musicale au caractère distinctif des styles musicaux japonais et occidentaux. Les trois expériences de ce chapitre démontrent également que les profils de contour musicaux des styles musicaux traditionnels japonais et occidentaux sont à la fois perceptibles par les auditeurs de divers contextes culturels et apportent une contribution supplémentaire et indépendante à la perception de la forme musicale.

Le chapitre 3 examine un dispositif de translittération musicale intermodal connu sous le nom de *shouga* (唱歌). Le *shouga* est un système phonétique incorporé utilisé pour représenter et articuler la forme voulue d'un morceau de musique donné, avec des idiomes particuliers associés à des contours formels spécifiques. Les études de ce chapitre reposent sur l'observation que l'objet d'art empirique, conséquence de la créativité humaine, est une représentation formelle de l'état mental de l'auditeur qui l'a produit. Dans le cas particulier du *shouga*, des musiciens experts ont créé un objet d'art (le *shouga*) pour représenter ou articuler leur impression phénoménologique d'un autre objet d'art (la forme perçue de la musique elle-même). Les observations empiriques des correspondances paramétriques entre ces deux objets permettent au psychologue de tirer des conclusions sur la manière dont l'impression subjective de la forme musicale peut être liée à l'objet d'art physique qui provoque cette impression. Plus précisément, les trois études de ce chapitre concernent l'exploration et la paramétrisation statistique des correspondances structurelles entre le *shouga* et l'œuvre séminale pour koto solo, *Rokudan no Shirabe* (六段の調). Les résultats indiquent une correspondance structurelle considérable entre la macro et la microstructure du *shouga* et les dimensions tonales, rythmiques et métriques de *Rokudan no Shirabe.*

Le chapitre 4 fournit des données perceptives sur la capacité des auditeurs non japonais sans formation avec cette musique à distinguer les véritables séquences *shouga* des chaînes de syllabes japonaises (leurres) construites arbitrairement, en association avec la structure des mélodies koto. Si le *shouga* est une représentation cohérente et intelligible de la forme musicale, ces participants devraient être en mesure de reconnaître l'apparence, ou la ressemblance, entre la mélodie musicale et la séquence *shouga* correspondante. Les résultats de cette dernière expérience indiquent que la grande majorité des auditeurs non japonais peuvent détecter les véritables *shouga* à un taux au-dessus du hasard et que la majorité des séquences de *shouga* réelles utilisées au cours de l'expérience étaient discriminables de leurs leurres à des taux supérieurs à ceux prédits par le hasard.

Le chapitre 5 conclut la thèse par une discussion générale sur les implications des résultats présentés pour les recherches futures sur les formes musicales, sur la créativité musicale et sur la capacité mentale de traiter des systèmes dynamiques abstraits.

Acknowledgments

Thank you to Professor Stephen McAdams for supervising this dissertation and for facilitating the practical and intellectual risks I felt it necessary to take. Thank you to Professor MATSUNAGA Rie for supervising my research in Japan at Kanagawa University. Thank you sensei ANDO Masateru for accepting me as one your students, teaching me the koto, lending me an instrument, and encouraging my investigations of the *shouga.*

Special thanks to the Canadian National Sciences and Engineering Research Council (NSERC) and the Japanese Society for the Promotion of Sciences (JSPS) for sponsoring my graduate work. Thank you to the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) and to McGill University for sponsoring my research and travel.

Thank you Takeru HIRATA, Dr Shoko KASUGA, Yamato SUZUKI, Reo OTAWA, Nick Westbrook, and Tatsuya NAKANISHI for your hospitality and companionship during my visit to Japan. Thank you for your help with translation and for putting me in contact with your Japanese friends and colleagues on behalf of my research once I had returned to Canada.

Thank you to my MPCL lab mates Jason Noble, Moe Touizrar, Kit Soden, Behrad Madahi, Erica Huynh, Lena Heng and Manda Fischer. It was a pleasure to move through graduate life alongside you. Thank you to Dr Collins Oghor for putting me up on your couch in my early days at McGill.

Thank you to Braden Flynn, Ephraim Hahn and Nicolas Galli for keeping me connected to the music, even when that meant contending with impossible demands. Thank you to Professor Martha de Francisco and to Andrew McLaughlin for supporting my musical intuition. Thank you to Professor Daniel Smilek for helping me to see the big picture beyond the big picture.

Thank you to the staff, faculty, and students of B21 for cultivating an academic environment that allowed my working process to thrive. Thank you to Professor Ollivier Dyens, Jhave, Anita, and Rebecca for supporting my creative work and facilitating the translation of my ideas into a format which would become this dissertation. Special thanks to Damián Arteca for exhausting my trains of thought. This document would still be a work in progress were it not for our conversations.

Thank you Dr Jared Boasen and Professor Kyoko Hashimoto for the great company and for connecting me to the Japanese culture within Montreal. A special thank you to Satomi Boasen for translating the experimental forms for Chapter 4 into Japanese, and to Kyla Nadorozny for translating the abstract of this dissertation into French.

Finally, thank you to my family. Thank you to my sister for holding down the fort. Thank you to my brother for training me up. Thank you to my mother for watching me enter the dark. And thank you to my father for encouraging me to keep going.

Contributions to Knowledge

The present dissertation provides four substantial contributions to the field of psychology.

The first is the development of a theoretical model for the conceptualization of the perception of musical form. Critically, this model incorporates both the creative and receptive capacities of the listener to engage with musical forms. This model is developed according to the notions of *musical image* and *perceptual geometry,* presented in Chapter 1.

The second is an empirical manuscript, providing the first evidence of the contribution of melodic contour to the perceptual distinctiveness between *Japanese* and *Western* musical styles. In particular, this work reveals that Japanese, Western, and non-Japanese non-Western listeners are capable of perceiving nuanced modifications to the phase relations among the local pitch height maxima and minima, implied harmonic progression, and implied meter of melodic sequences. Finally, these perceived differences are systematically and reliably associated with particular musical forms, identified by listeners as differentially characteristic of the Japanese and Western musical styles. This work is presented as Chapter 2.

The third contribution to knowledge, presented in Chapter 3, is an extensive corpus analysis of the structural correspondence between the *shouga* and melody of the *soukyoku* (koto music) composition *Rokudan no Shirabe.* It is the first extensive analysis of its kind, indicating that the phonetic structure of a traditional Japanese musical transliteration (the *shouga*) is systematically organized with respect to the tonal, metric, and rhythmic dimensions of the musical sound it represents.

The fourth and final contribution (Chapter 4) is an empirical manuscript providing the first experimental investigation of *shouga* perception. The results of this experiment provide evidence that naïve listeners, unable to speak or understand the Japanese language and untrained in traditional Japanese music, are capable of distinguishing true *shouga* sequences from random combinations of Japanese speech sounds in association with the original *soukyoku.*

Contributions of Authors

I wrote the manuscripts for Chapters 1–5 of this dissertation. I conceived of the present research program, developed the theoretical framework, experimental designs, and experimental apparatus for each of these chapter, and analyzed the experimental data for Chapters 2–4. Professor Stephen McAdams supervised this dissertation in full, participated in the experimental design and plans for data analysis, and edited the manuscripts for Chapters 1–5. Professor Rie Matsunaga supervised the research conducted at Kanagawa University presented as Experiment 1 in Chapter 2. Bennet K. Smith programmed and implemented the experimental apparatus for Chapter 4.

The following chapters of this thesis are in preparation for submission for publication:

Chapter 2: Bonin, T., Matsunaga, R. & McAdams, S. (in preparation). Cross-cultural perception of Japanese- and Western-typical melodic contour profiles.

Chapter 3: Bonin, T. & McAdams, S. (in preparation). 箏の唱歌. Structural relationships between the *shouga* and melody of *Rokudan no Shirabe* (六段の調).

Chapter 4: Bonin, T. & McAdams, S. (in preparation). 箏の唱歌. Naïve listeners' perception of formal analogy between the *shouga* and melody of *Rokudan no Shirabe* (六段の調).

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 1 – Theoretical considerations for the scientific investigation of the perception of musical form: *The musical image* and *perceptual geometry*

Introduction

Music appears in the mind of the listening subject exhibiting a distinct *form.* This *form* is amenable to identification, abstract parametric manipulations, and comparison with other more or less similar *forms.* Such properties suggest that subjective experience of musical *form* can be modelled in terms of an abstract *psychological entity* situated within an intelligible *psychological domain.* Furthermore, this *psychological domain* must be in some sense systematically associated with the *physical domain* in which musical sounds are created, measured, and reproduced. Indeed, both the music perception sciences and musical arts rely on the implicit assumption that the subjective impression of musical *form* can be systematically investigated and associated with the physical world; such a correspondence would be impossible if either the psychological domain or the musical *form* was itself unintelligible or incomprehensible. Despite this systematic correspondence, the subjective experience of musical *form* cannot be effectively described in terms of physical materials, and the physical constitution of music cannot be effectively described in terms of subjective experience. Thus, both the scientist and the artist are in the intermediate position of trying to intelligibly associate abstract mental impressions with concrete physical materials—the scientist by way of analysis and interpretation, the artist by way of synthesis and creation. Likewise, the listener, experiencing the music itself, is in the profound and somewhat ineffable position to directly observe the confluence of *material* and *immaterial things,* otherwise considered by all accounts to be inhabiting entirely distinct *physical* and *psychological* domains.

The present psychological dissertation is concerned with the perception of musical *form*. This introductory chapter is dedicated to theoretical considerations for the scientific investigation of musical *form.* In particular, this chapter will focus on how the implicit intelligibility of the relation between the subjective impression of musical *form* (*the musical image*) and physical materials (*the musical signal*) can be explicitly articulated within a unified conceptual framework appropriate for the scientific investigations of music cognition and perception. Borrowing from and extending the theoretical precedents of the existing psychological literature in this research domain, the present framework will be operationalized in terms of two notions: the *musical image* and *perceptual geometry.* The *musical image* is the discrete musical *idea* whose *form* is perceptible to the subject and can be systematically associated with *musical signals* within the physical domain. *Perceptual geometry* is proposed as a model for describing the general psychological process by which the conscious subject identifies, scales, and relates distinct *percepts* (such as *musical images*) within the psychological domain of experience. I will conclude the chapter with an overview of how these theoretical considerations can be used to orient the reader with respect to the experimental studies presented in the following data chapters of this dissertation.[1]

The Musical Image

The *musical image* is the basic perceptual unit of *music*. It can be conceptualized as a finite element within the conscious psychological domain of *perception*.[2, 3] It is the means by

---

[1] I also include two tangentially related appendices for the interested reader. The first is a brief review of the psychology and neurophysiology of creative activity. The second is a brief review of the development of the scientific epistemology of the 20th century as it applies to the investigation of abstract conceptual frameworks as empirical facts worthy of scientific investigation. I present four abstract axioms and argue that these axioms apply equally well within the context of the epistemological challenge facing contemporary science as within the technical domain of artistic creativity.

[2] "The psychological domain" is the theoretical construct necessary to make sense out of musical phenomena that remain unexplained in terms of the theoretical construct "the physical domain."

[3] *Perception* is the psychological consolidation of precipitant organismic processes into a phenomenologically distinct element of experience (i.e., the process by which perceptual images are rendered within the conscious experience of the organism; cf., Dewey, 1896).

which musical perception may be said to exhibit *form.*[4] The musical image is therefore a percept. A *percept* is a concept developed as a consequence of the process of perception (Colman, 2009). A *concept* is an abstract mental picture of a group or class of objects formed by combining all their aspects (Margolis & Laurence, 2018). This notion of a *percept* as *an abstract mental picture* is indicated by the use of the term *image*.[5] That this image is associated with music is indicated by the use of the term *musical.* Stated intuitively and concisely, the musical image is an *idea.* The original Greek *idea* means *pattern* or *form,* and derives from the base *idein,* meaning *to observe (perceive).* Thus, the musical image is an observable form of music appearing to consciousness. In contemporary scholastic vernacular, the musical image is a *psychological entity* (Margolis & Laurence, 2015). It is a *thing* possessing distinct *form*, operative within and confined to the psychological domain of perception.[6]

The musical image refers to the *psychological aspect* of the *total musical event* (i.e., music *in toto*)*,* which comprises both a psychological and a physical domain (see Figure 1.1). The psychological domain may be considered the "ideal pole" of music while the physical domain is considered its "material" counterpart. Thus, the *musical image* is the projection of the music within the psychological domain as its ideal aspect, while the *musical signal* is the projection of the music within the physical domain as its material aspect.[7]

---

[4] See third axiom, Appendix 1B

[5] The concepts of "image," "picture," and "observation," etc., need not—and in the case of music, *should not*—be constrained by implicit association with the notions of "inertia/timelessness/non-dynamic" or "3-dimensional spatial" associated with the colloquial use of these terms as referring to *visual* phenomena.

[6] The entity here named *musical image* has received various names and variously similar and dissimilar definitions within the literature (see below), but central to all definitions is the notion of the musical image as a percept (see above).

[7] For the sake of lexical continuity within this chapter I ask for the reader's goodwill in allowing the use of the term "material" in reference to the *physical stimulus*, despite the well-established fact within the physical sciences that a physical stimulus is a complex interaction of *matter* and *energy*, with the former aspect sometimes referred to uniquely as "material."

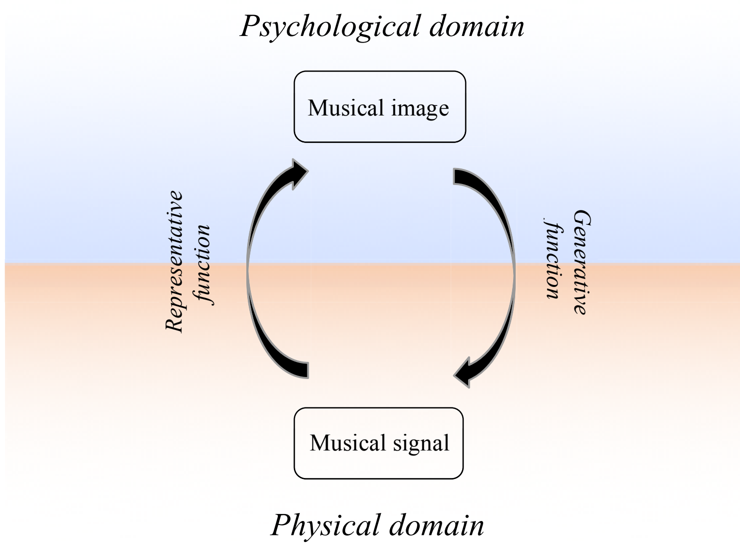

Fig. 1.1 A schematic representation of the *total musical event,* in which the *musical image* exhibits both generative and representative functions.

This notion of the musical image is central to many works in both music cognition science and the musical creative process. Speaking simplistically and for the sake of brevity, musical science and musical art reveal two distinct but interrelated functions of the musical image, each corresponding with a distinct unidirectional relationship between the ideal and material aspects of the total musical event. In the context of this thesis, these are called the *representative* and *generative functions* of the musical signal image (Figure 1.1). The former is best understood through an extraverted attitude toward the musical event and is typical of the scientific discipline, while the latter is best understood through an introverted attitude toward the musical event and is typical of the artistic discipline (Jung, 1922/1923).[8]

[8] Personal correspondence with artists and scientists leaves little room for the naïve unilateral interpretation that in fact the *artistic* and *scientific* approaches to music are entirely distinct. In actuality, scientific work requires a degree of creativity and imagination and artistic work requires a degree of analysis and measurement. Nevertheless, the *artistic* and *scientific* attitudes may be summarized schematically as by Jung: "… active phantasy and ideation being the principal attribute of the artistic mentality, the artist is not merely a *representor:* he is also a *creator,* hence essentially an *educator,* since his works have the value of symbols that trace out the line of future empirical investigation and development," (Jung, 1922/1923, p. 580). "… the [introverted] artistic attitude, oriented by the idea, must tend towards monism. The idea has always a hierarchical character to the introvert, whether it be gained

In adopting an extraverted attitude to music, science emphasizes the primacy of the physical domain within the total musical event, and thereby the *representative function* of the musical image. Accordingly, the musical image in this scheme is conceptualized as *the psychological representation of a physical event*, and the musical signal that precipitates this image is called the (*physical*) *stimulus*. The representative function of the musical image is therefore the capacity of the musical image to stand as the psychological representation of the physical source object within the total musical event. For example, Schaeffer (1966/2017) defines his *sound object* as "sound itself, considered as *sound,* and not the material (instrument or some sort of device) that produces it," (p. 8), McAdams (1984) defines the *auditory image* as "a psychological representation of a sound entity exhibiting an internal coherence in its acoustic behaviour," (p. 291) and Bregman (1990) defines the *auditory stream* as "the perceptual unit that represents a single happening. … play[ing] the same role in auditory mental experience as the object does in visual… acting as a center for our [psychological] description of an acoustic [physical] event," (pp. 10–11).

It is in this physically reductive sense that the scientific works of the extant music cognition literature[9] prioritizes rational description of the musical stimulus within the total musical event. These works provide empirical descriptions of the physical, physiological, and cognitive-perceptual processes *leading up to* a listener's conscious experience of music (Dowling, 1999). The seminal works in this field require listeners to complete various

by abstraction from representations and concrete concepts, or whether it has an a priori existence as unconscious form. … This way of thinking is mutually opposed to the extraverted attitude in a remarkable way: It is equally clear that the extraverted attitude oriented by the object [as in physical science] must always incline to a majority of principles (pluralism), since the multiplicity of objective qualities entails also a plurality of concepts and principles without which a suitable interpretation of the nature of the object cannot be gained. The monistic tendency belongs to the introverted attitude, the pluralistic to the extraverted," (Jung, 1923, p. 396).

[9] See e.g., Krumhansl & Kessler (1982) for seminal work on tonality; Dowling (1978) for that on melodic contour; McAdams et al. (1995) for timbre; and Large & Jones (1999) on temporality

discrimination tasks (*goodness of fit judgments* between two stimuli, such as in the probe tone method, e.g., Krumhansl & Kessler, 1982; *dissimilarity judgment* between two stimuli, e.g., McAdams et al., 1995; and *target detection tasks*, in which a particular stimulus [known as 'oddballs' if only implicitly specified by the task design or 'targets' if explicitly specified by the task design] must be identified amongst a stimulus array, as in e.g., Large & Jones, 1999), as well as memory recognition and recollection tasks that probe the internal structure of unconscious memory and perceptual systems (e.g., Dowling, Tillman & Ayers, 2001). The ideal pole opposed to the rational (parameterized) material pole of the extraverted attitude is inherently irrational (unparameterized; Jung, 1922/1923). That is, *ideas* are undifferentiated with respect to the physical dimensionality of a *material* framework. The ideal reality of the musical image (i.e., the psychological aspect of the music) is therefore considered arbitrary[10] (cf., the 'black box' of cognitive psychology) in scientific works partial to the rational descriptions of the physical domain of music.

In contrast, by adopting an introverted attitude toward the musical event, the artist emphasizes the primacy of the psychological domain within the total musical event, and thereby the *generative function* of the musical image (Figure 1.1). The musical image in this scheme is conceptualized as the *psychological source of a physical event,* and the physical signal that follows this source is called the *artwork, artefact, art object, technique* or *behaviour*.[11] The generative function of the musical image is therefore the capacity of the musical image to serve as a psychological point of reference for the creation and manipulation of a physical musical signal within the total musical event. For example, Richard Wagner observed, "Imagination

[10] Underlying the conflation of "subjective" with "arbitrary" implicit to the majority of scientific psychological discourse.
[11] For the reader's reference, I most frequently refer to this construct as the "art object" in the context of the empirical manuscripts presented in Chapters 2–4.

creates reality," (Brown, 2015). Juno award winning composer and instrumentalist Matthew Good described his creative process in the following manner: "The inspiration is like a breeze passing through bedsheets hanging out on a laundry line, and then my job with the guys is to go recreate the ripples on the blanket first created by the wind. My best song ever was maybe 50% successful in this," (CJXY FM radio interview, 2013). And in a musician's clinic, drum and bass pioneer JoJo Mayer explains, "What is a paradiddle? It's a rudiment. It's a *shape*. It's an exercise. When you practice that paradiddle, you have a *concept* of how to execute it. You're gonna get to a certain level where you can execute this *structure*. Eventually you can get into the detail of it… all those inner beats become musical *entities*, and *not just mechanical texture*—'This is dependent on that'—but what I learned is that *you can uncouple inner hearing from what [you're] playing physically*.[12] [My] inner hearing started to *evolve* and anything that you do you have to be able to hear it first.[13] … A lot of drummers come to me and want to increase speed but they can't *hear music* at that speed. *It's not a technical but conceptual limitation.* Simple devices or meditations have been really opening up the way I behave behind my kit and the type of decisions that I make," (Mayer, 2016, emphasis mine).

It is in this idealistically reductive sense that artistic works prioritize rational description of the musical image within the total musical event. The effective creation of artworks (e.g., sound, score, stage) and execution of creative techniques (e.g., behaviours, speech, imagination) requires an awareness and parametric differentiation of the hierarchical character of the musical image itself within the psychological domain of perception in order that the physical musical signal can be constructed, evaluated, and revised with respect to the original source *idea*.[14] This

---

[12] See Appendix 1B
[13] Cf., Wagner in Brown (2015)
[14] See Appendix 1A: "A brief review of the psychology and neurophysiology of creative activity"

parameterization results in the musical image as the active principle by reason of its own structural integrity, which compels the individual to seize and shape the physical materials of the musical event: Any *symphony*, before it is “a symphony” in the material, stimulating, audible, sense of a musical *signal,* is a *musical image* accessible to the composer within the psychological domain. The material pole opposed to the rational ideal pole of the introverted attitude is inherently irrational (Jung, 1922/1923), i.e., *materials* are undifferentiated with respect to the psychological dimensionality of an *ideal* framework. The acoustic reality of the musical signal (i.e., the physical aspect of the music) is therefore considered arbitrary in creative works partial to the rational description of the psychological domain of music. The physical dimensionality of musical signals is relevant to the generative function of music only insofar as it affords an array of material possibilities from which the artist may select the most appropriate representations of the precipitant *idea* (see Figure 1.2 for the respective black boxes[15] of the artistic and scientific investigations of music). In this way, the artistic and scientific approaches to the dynamic relation between the physical and psychological aspects of music can be considered complementary perspectives regarding the *form,* structure, and function of musical perception (Figure 1.3).

[15] Within scientific and engineering contexts, the “black box” is a device, system, or object that can be viewed in terms of its inputs and outputs without any knowledge of its internal workings

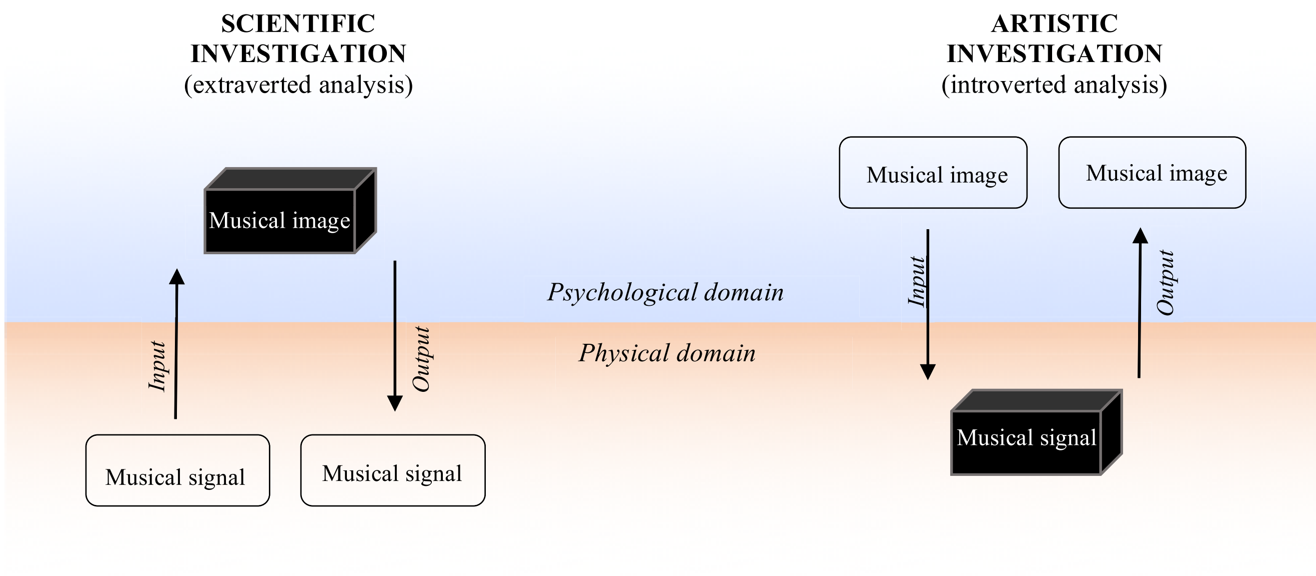

Fig. 1.2 The respective black boxes of the artistic and scientific investigations of the relation between the *musical image* and the *musical signal.*

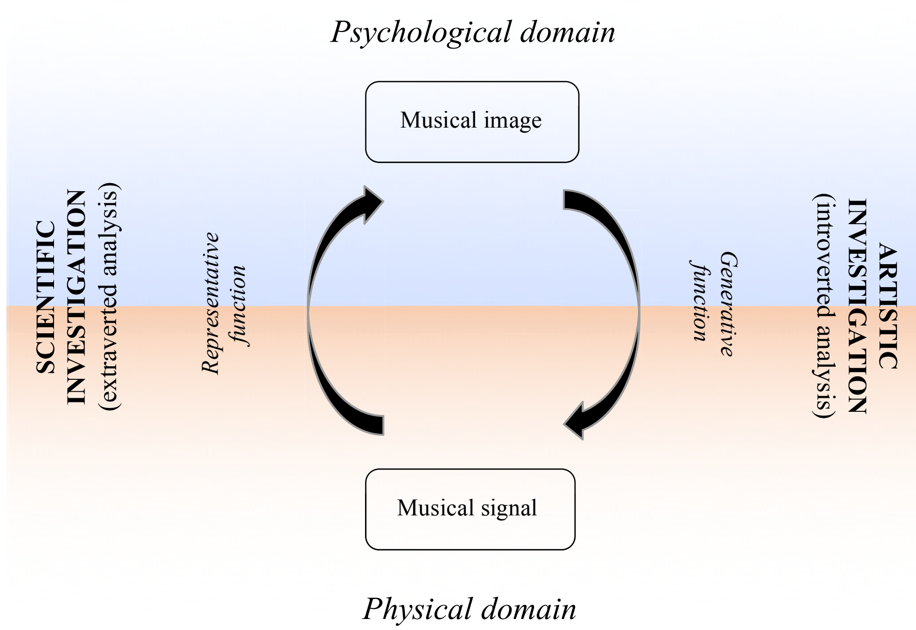

Fig. 1.3 A schematic representation of the stereotypically scientific and artistic investigations of the *musical image* in the context of *the total musical event.*

For the optimal creative process, a compensatory extraverted attitude must therefore be taken to facilitate the representative function of the musical image. This attitude is often facilitated by producers, recording engineers, and other technical professionals, though artists are of course capable of considering both attitudes in their own right (see footnote 8). I suggest an analogous compensatory attitude is conducive to an optimal *scientific* investigation of the musical image. Specifically, if a complementary introverted attitude were taken toward the understanding of the *generative function* of musical images, the resultant scientific models of musical images would offer a more comprehensive understanding of the coordination between the *psychological* and *physical* domains in the creative context of the total musical event. The generative function of the musical image requires the perceptual capacity of the subject to identify, scale, and relate distinct *percepts* (such as *musical images*) within the psychological domain of experience. I propose a model to facilitate the conceptualization of these perceptual processes according to the notion of *perceptual geometry.*

Perceptual Geometry

Perceptual geometry is the rational description of percepts. Perceptual geometry is a scheme of classification, like spatial geometry.[16] Perceptual geometry corresponds both with the

[16] The application of geometric concepts to the structure of mental states, "schemata," and their associated imagery is not unprecedented. Shepard (1964, 1982) and Krumhansl and Kessler (1982) applied the geometric structures of the *helix* and the *torus* to provide rational descriptions of the structure of pitch perception, as a function of chroma and octave, and key space, respectively. This thesis differs from these works in that it does not attempt to generate metaphorical associations between the structure of mental states and the structure of spatial geometric forms. Instead this work is concerned with the rational principles fundamental to geometric thinking, for which the forms of spatial geometry provide only conceptual analogies, not exemplary cases. Some criticisms of the aforementioned models (derived partially through misunderstanding, e.g., Butler, 1989, 1990a, 1990b) could have been disarmed or circumvented completely if such spatial metaphors—which necessarily entail logical and verification inaccuracies when applied to psychological phenomena—were not used in the first place. It is noted that the use of spatial metaphors provides a superficial (extraverted) impression of precision that is in some cases preferable to the "vagueness" of the purely theoretical (introverted) descriptions, such as those provided here. The benefit of the latter process with which this thesis is concerned, however, is in no way metaphorical, since it describes a rational scheme for a discussion of the structure of internal states. It cannot be overstated that none other than an introverted

unconscious capacity of the subject to *identify* distinct percepts (e.g., musical images, see following section), and the associated conscious capacity of the subject to ascertain the relative *sameness* or *difference* (*likeness*) between these percepts, by which finite moments of experience may be *characterized* and *related*.[17]

From a rational perspective, the impression of *likeness* relating two percepts is proportional to the structural coordination of their forms; when one entity is subjectively experienced as being *like* an other entity, it may be said to exhibit a *sameness of structure* with that other. We then operationally define *structure* as follows: The structure of an entity is a distinct *set of empirical ratios*[18] characterizing the relations among the constituent parts of the entity. In order for two percepts to exhibit a common structure, they must share a distinct set of empirical ratios relating their respective constituent parts to one another. Stated inductively: In the case of like percepts, the components of one percept are so structured within their respective dimensionality as to produce a form which also emerges from the arrangement of the components of the other percept within their respective empirical dimensionalities. This structural correspondence is the basis of analogy (e.g., McGilchrist, 2009), metaphor (Lakoff & Johnson, 1980) and, according to some authors, any and all thought whatsoever (e.g., Wittgenstein, 1921/1922; Hofstadter & Sander, 2013).

---

analytical attitude can provide a rational description of the intrinsically internal entities such as percepts. It is the application of extraverted analogies (e.g., spatial forms or acoustic correlates) to the description of internal states (e.g., musical images) that is metaphorical, just as the application of introverted analogies (e.g., feelings or concepts) to external states (e.g., acoustic signals) is metaphorical. The charge of "vague metaphor" is entirely warranted if the extraverted analogies for psychological states provided by the mathematical-geometric models of musical science or the aesthetic-narrative models of the musical art are not subjected to an introverted analysis of their respective psychological structures. But in and of itself the term *metaphorical* is not to be confused with *irrational,* and *irrational* is not to be confused with *erroneous,* although metaphorical schemes will inevitably lead to irrational theoretical constructs that will necessarily fail within (or without!) a rationally constrained framework. The contemporary scientific literature does not possess a rational framework for the description of internal states such as musical images. This framework is posited here under the name *perceptual geometry.*

[17] See fourth axiom

[18] "arrangement" or "constellation"

Thus, the components of a percept perceived to be *like* another are co-ordinated such that the resultant affine structure finds empirical analogy in that resulting from the co-ordination of the components of the other percept. The *likeness* of the two percepts is consciously accessible to the subject as the impression of *common form*; the *structural sameness* is not, it is unconscious.[19] The structural sameness can be known only through inductive generalization (as by *perceptual geometry*) of the psychological *spaces* and *dimensionality* in which the two percepts are consciously formed. Perceptual geometry is therefore a rational abstraction of the forms of percepts, and, in this way, an investigation of the unconscious percept-structuring tendencies of the psyche. Having been abstracted in this way, the resultant structural descriptions of percepts may then be related to empirical measurements of, for example, physical stimuli. This possibility, of particular interest to the experimental psychologist, is explored in the empirical chapters (2, 3, and 4) of this dissertation.

Expert artists are those who have liminal, pre-verbal, and sometimes verbal conscious access to this largely tacit knowledge (cf., Polanyi, 1966; Gill, 2015) and can translate this awareness into artistic work, just as expert mathematicians and physicists are those who have access to liminal, pre-verbal, and sometimes verbal conscious access to spatial geometry.[20] In this regard, the geometric proofs of the geometer are the behavioural analogues of the artistic works of the artist, each reflecting a common process of intellectual abstraction of perceived forms. Critically, these unconscious faculties do not need to be explicated in order for musical perception to operate effectively, in exactly the same way that one needn't know the mathematical proofs of spatial geometry for spatial perception to operate effectively. In the absence of explicit intellectual abstraction, the structural identity of two musical images is

[19] cf., Wittgenstein (1921/1922; 2.033), "Form is the possibility of structure."
[20] See Appendix 1A: "A brief review of the psychology and neurophysiology of creative activity"

variously approximated through explicit and implicit means. Objectively, this correspondence is available through physical parameterization of distinct art objects, verbal reports of perceiving subjects, and common cognitive-behavioural adaptations to the two objects (e.g., Neumann, 1955/1972). Subjectively, the correspondence is available as the perception of common form, which can be approximated by a listener's sense of *coherence* (McAdams, 1984), *consonance* (Bonin et al., 2016), and multimodal coordination of the experienced physicality, physiology, and sonority of a piece of music (Ando, 1986). It is inversely proportional to the *dissimilarity* and *dissonance* among musical images (e.g., Krumhansl, 1990; McAdams et al., 1995, Bonin & Smilek, 2016), and to the *unlikeness* relating an art object (or performance) to the intention of the artist (cf., Good, 2013; Condo, 2017).

Thus, perceptual geometry is not itself measured. It is the rational description of the perceptual images that *are* measured. The ineffability of perceptual geometry is identical to that of spatial geometry; squares and tori are not themselves measured—physical distances are measured, and the relative positions prescribed by these distances are captured in the structural concepts of, for example, squares and tori. It would be erroneous to critique the utility of spatial geometry on the basis of one's inability to find "a square" in the physical world (they don't exist as such; Mach, 1906/2011, see Appendix 1B). The utility of spatial geometry is that it provides reliable structural classifications of physical phenomena. Likewise, the utility of conceptualizing psychological systems in terms of perceptual geometry is that the concept emphasizes the uniform parameterization of psychological entities (e.g., *musical images*, see following section) by which the relations among their endlessly distinct *forms* can be made intelligible. The geometry of psychological *things* is not a participant in the empirical realm of perception (it is not perceptible to the conscious subject): Perceptual geometry is the logical ordering principle presupposed by the rational description of psychological *things*. Thus in the same way Mach (1906/2011; see Appendix 1B)[21] described the correspondence between geometric schemes and physical objects appearing to the subject, the same may be said of the relation between geometric schemes and psychological entities appearing to the subject.

This notion of perceptual geometry finds analogous predecessors in earlier authors' definitions of *schema* (Colman, 2009),[22] *form* (Spencer-Brown, 1969),[23] and *structure* (Wittgenstein, 1921/1922).[24] The utility of addressing this common notion with the term *geometry* is the resultant connotative emphasis on the intrinsically *rational* nature of the schema, form, and structure of the percept according to which the subject is capable of making systematic changes to the material substrate (as in the creation of art objects). The term geometry ("a branch of mathematics that deals with the measurement, properties, and relationships of points, lines, surfaces, solids, and higher dimensional analogs; broadly: the study of properties of given elements that remain invariant under specified transformations"[25]) entails reference to the classically accepted epitome of rational constructs (mathematics). If asked by a scientist which elements of a musical image (or other percept) correspond *specifically* with "points," "lines," "surfaces," "solids," and "higher dimensional analogs," of material entities, the creative artist is inclined to say, "It depends." But this relativity is functionally equivalent, psychologically

---

[21] "The question whether a given *physical* [sic.] object is a straight line or the arc of a circle is not properly formulated. … The question is simply whether the object… conforms better to the one concept than to the other…."

[22] "*A plan, diagram, or outline, especially… of some aspect of experience*, based on prior experience and memory, structured *in such a way as to facilitate perception*, cognition, the *drawing of inferences*, or the *interpretation* of new information," (*schema*) (emphasis mine).

[23] "We take as given the idea of distinction and the idea of indication *and that we cannot make an indication without drawing a distinction. We take, therefore, the form of distinction for the form.* Call the space cloven by any distinction, together with the entire content of the space, the form of the distinction. Call the form of the distinction *the form,"* (p. 1) (emphasis mine).

[24] "That *the elements* of the picture [image] are *combined with one another in a definite way… this connexion of the elements of the [image] is called its structure,* and the possibility of this structure is called the form of the [image]," (2.15) (emphasis mine).

[25] "Geometry." *Merriam-Webster.com*. Merriam-Webster, n.d. Web. Accessed 26 Aug. 2018.

speaking, to the prevailing ambivalence of the scientist regarding which subjective states are to be *specifically* associated with a particular material object: “It depends.”

“There is something intrinsically irritating in this conflict of attitude, and at bottom, this is the cause of the most heated and futile scientific discussions,” (Jung, 1922/1923, p. 386). But if both parties can accept the relativity with which the other regards their particular perspective concerning the common goal (i.e., the investigation of the relation between materials and ideas), they may each benefit from the insights of the other. This mutual understanding may be facilitated by a generalized definition of three more geometric concepts in the context of this thesis: *objects, dimensions,* and *spaces.*

An *object* is an entity within a parameterized set whose aspects or *features* serve as the basis of its relations[26] to the other entities within the *space*.[27]

A *dimension* is a measurable extent among the *objects* of a parameterized set. Dimensions express the featural or aspectual differentiation of *objects*, and constitute the means of rationally explicating the parameterization of the *space* in which *objects* are situated.[28] A *dimension* is that aspect common to every member of a set of *objects* that makes their differentiation and relation intelligible.

A *space* is a parameterized set of objects. *Objects* can be of any number of classes; they can be unqualified entities within the set or they can be functions of other *object-spaces*, entire subspaces (subsets of parameterized objects and their relations) of other *object-spaces*, etc. In any case, differentiated *objects* are assigned a distinct *point* (finite position) within the *space*.

---

[26] One of these relations is its *non-ness,* or *unlikelihood*, with other objects, which constitutes the basis of the object’s differentiation and identification.
[27] cf., axioms, Appendix 1B: An *object* is to a *thing* as a *subject* is to *consciousness.*
[28] See James (1909); Whitehead (1925); Appendix 1B

The nature of the relations among those points constitutes the parameterization, or *structure,* of the *space*. *Geometry* is the rational description of these relations and the identification of their structures. *Dimensions* serve to metrically delineate this parameterization and its description. It is within a common *space* and with respect to common *dimensions* that *objects* may be said to *coordinated.*

Thus, we can operationally define the musician's creative process within the model of *perceptual geometry* in the following way: By the application of introverted analysis, the subject construes the *musical image* as an *object* situated within the *dimensionality* of the prevailing perceptual *space.* It is according to this *dimensionality* that 1) the relative positions of the constituent elements of the *form* of the *musical image* are observed, 2) the boundaries of the *form* are construed with respect to the surrounding *space,* and 3) the subject is capable of altering its vantage point of observation with respect to the *object.*

The *generative function* of the musical image allows the original *idea* (*i*) conceived by the subject to be translated into a material substrate. This substrate then acts as a stimulus, which, by the activity of the *representative function,* produces a secondary musical image (*i'*) with its respective form appearing to the subject. The geometric comparison of the *forms* of these two images allows the subject to make further alterations to the structure of the material substrate; this cycle is reiterated until the musical images *i* and *i'* are perceived to exhibit a perceptual *likeness.* Mastery of the creative process is thus as much a consequence of the subject's capacity to parameterize the perceptual space and to systematically orient themselves with respect to it as it is of the manual and technical skills necessary to systematically alter the material substrate.[29]

[29] See Appendix 1A: "A brief review of the psychology and neurophysiology of creative activity"

Modelling the artistic investigation of *musical images* in this way reveals the fundamentally analogous use of *reason* as applied by the scientist to the description and interpretation of *musical signals* and the artist to the synthesis and creation of *musical images.* Taken as a discrete and parameterized *object* within the psychology domain*,* the *musical image* can thereby be incorporated into a generalized scientific framework concerning the interactions between physical and psychological entities. It is hoped that this conceptualization will help to facilitate scientific investigations of human creativity and of the systematic relation between the *musical images* perceived by the subject and the *musical signals* resulting from the alteration of physical materials.

Summary

According to the foregoing account, the *musical image* characterizing the internal state of mind of the conscious subject is both *representative* and *generative* of external, empirical art objects. This means music ("the total musical event") cannot be comprehensively understood solely in terms of a *tabula rasa* epistemology that regards the subjective experience as a merely "imperfect representation" of the physical world.[30] An essential element of the subjective experience of music is the creative engagement with *ideas* and other abstract mental states. This is true even for listeners not involved in the original creation of a musical work; the form and content of free associations produced by individual listeners presented with the same musical excerpts share robust thematic and conceptual analogies (Bonin, Degtyareva & Hahn, 2019; see also Noble, 2018). In this sense, the empirical art object may just as readily be considered an "imperfect representation" of the psychological world. Thus, the experience of music is more

[30] See Appendix 1B

appropriately characterized by the subject's coordinated observation of both *psychological* and *physical* entities. At the most practical level, this means that descriptions of a listener's experience of music in terms of the musical image situated within the psychological domain or in terms of the musical signal situated within the physical domain provide complementary insights into the perception of musical form. Psychological sciences interested in the cognition and perception of music must therefore address and contend with this creative and ideal aspect of musical experience. It is on behalf of these investigations that the notions of *musical image* and *perceptual geometry* have been developed.

Applications to the following empirical research chapters

Taking together the respective works of the artist and scientist, the "directionality" of the relation between the psychological and physical domains of music is seen as a matter of operative definitions and is not an empirical reality of the musical image itself (Figure 1.3). This perspective is virtually unstated in the existing music cognition literature, and as a result the scientific work progresses with a unilateral extraverted bias (Jung, 1922/1923) that considers the musical image only in terms of its capacity to represent physical properties of musical events (see, e.g., the seminal works of music perception and cognition research referenced in footnotes 9 and 16). The result is a paucity of scientific studies concerned with the subjective parameterization of musical images within the psychological domain (Godøy, 2019). While an extraverted bias is appropriate and maximally useful in purely physical sciences such as acoustics,[31] the same cannot be said of psychological science, which concerns itself with the

[31] And largely to the same extent with digital signal processing and music information retrieval, although these processes may ostensibly be informed by psychological research and enhanced by perceptual modelling (e.g., Alías, Socoró and Sevillano, 2016).

balanced interplay between psychological and physical states. In this latter case, a purely extraverted (i.e., *physical*) rationale distorts the general theoretical conceptualization of the relation between the physical and psychological aspects of musical experience. Correcting this bias necessitates the consideration of the complementary introverted rationale central to musical experience. I have suggested *perceptual geometry* as one potential model by which these considerations can be formulated within a scientific context.

The premise of this thesis is that the introverted, artistic attitude that regards musical images *as discrete, objective things,* may reveal unique insights to the scientific investigation of music perception. Although the explicit communication of introverted analysis is notoriously difficult (see Appendix 1A), insights gleaned from the artistic disciplines can be used to develop scientific modes of reasoning (cf., Whitehead, 1925; Appendix 1B) conducive to the description and interpretation of the subjective experience of musical form. For example, the proposition that the musical signal is a representation of the musical image originating in the mind of the composer suggests that structural correspondence among the empirical parameters of expertly crafted musical signals can be interpreted by psychologists as reflecting a coherent subjective experience of musical *form.* That is, the physical art object can be understood to be *intentionally* created with respect to perceptible dimensions of the associated *musical image*. This perspective offers researchers an alternative methodological approach to the investigation of musical form as compared with direct subjective reports or dissimilarity judgments: If the art object is understood to be a reflection of a coherent subjective experience of musical *form,* then it can be empirically evaluated for systematic or recurrent structural motifs. These motifs could then be abstracted and modelled with exploratory statistical or computational techniques or presented to listeners for complementary subjective report or dissimilarity judgments. The question then becomes whether such a concept can be operationally implemented within an empirical framework. That is, can an experimental design be conceived such that a reasonable interpretation of the data is only possible in light of the *generative* and *representative* functions of the *musical image* and the perceptual processing of musical form as formulated in the proposed model of *perceptual geometry*?

The following data chapters of this dissertation explore this question with three empirical investigations. The first data chapter (Chapter 2) concerns the perceptual ability of Japanese, Western, and non-Japanese non-Western participants to discriminate melodic sequences on the basis of whether they are more representative of a Japanese or Western musical style. Before beginning the experiment, we presented Japanese musicians with traditional Japanese melodies and asked them what would have to be done in order to make the melody sound “more Western.” The general artistic feedback was that the contour profiles (the sequence of rising and falling pitches) within these melodies would need to be “straightened out.” We then asked a musician to systematically manipulate the relative phase correspondence between the respective metrical and melodic dimensions of each melody, such that the local contour maxima and minima corresponded with strong metric downbeats. This objective manipulation was taken as the structural proxy for the perception of “straightened out” melodic sequences. We then transposed the original Japanese melodies and the altered Western-contoured variants into a stereotypical Western scale, and presented them to our participants. Previous research has demonstrated that musical scale contributes to the perceptual distinctiveness of Japanese and Western musical styles. Results from three experiments in Chapter 2 demonstrate furthermore that the characteristic musical contour profiles of the traditional Japanese and Western musical styles are perceptible by listeners of various cultural backgrounds and make additive, independent contributions to the perception of musical form. Thus, the artistic insight that the perceived cultural style of a musical form can be systematically altered on the basis of their respective temporalities was effectively implemented within, and corroborated by, an empirical investigation.

The second and third data chapters (Chapters 3 and 4) investigate a cross-modal musical transliteration device known as the *shouga* (唱歌). The *shouga* is an embodied musical transliteration used to represent and articulate the intended form of a given piece of music, with particular *idioms* associated with particular melodic fragments. The studies of these chapters are predicated on the observation that the empirical art object, as a consequence of human creativity, is a formal representation of the mental state of the artist who produced them. In the special case afforded by the *shouga*, expert musicians have gone on to develop not only the musical art object (the music itself) but a descriptive system to articulate the intended subjective impression of musical form perceived by the listener when presented with the musical art object. Empirical observations of the parametric correspondences between these two objects allows the psychologist to draw inferences regarding how the subjective impression of musical form may be related to the physical art object that elicits this impression. Specifically, the three studies of Chapter 3 concern the exploration and statistical parameterization of structural correspondences between the *shouga* and the seminal solo koto work *Rokudan no Shirabe* (六段の調). Results indicate considerable structural correspondence between the macro- and micro-structure of the *shouga* and the tonal, rhythmic, and metric dimensions of *Rokudan no Shirabe.*

Chapter 4 extends this work by providing perceptual data regarding the ability of untrained non-Japanese listeners to distinguish true *shouga* sequences from arbitrarily constructed strings of Japanese syllables (foils) in association with the structure of koto melodies. If the *shouga* is a coherent and intelligible representation of musical form, these participants should be able to recognize the semblance, or likeness, between the musical melody and its intended *shouga* sequence. The results of this final experiment indicate that the vast majority of non-Japanese listeners can detect the true *shouga* above chance rates, and that the majority of true *shouga* sequences used within the experiment were discernible from their foils at rates exceeding those predicted by chance.

Conclusion

In this chapter I have argued for the necessity of a perceptual science of musical *form* to consider an operational definition of the psychological domain in which *musical images* are experienced. I suggested that the alternative attitudes characterizing the scientific and artistic approaches to musical investigations can provide complementary insights into the development of a psychology of music perception and outlined a candidate research program by which some of these theoretical considerations may be operationalized in an empirical context. The results of these empirical investigations are presented in the following data chapters.

References

Alías, F., Socoró, J., & Sevillano, X. (2016). A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds. *Applied Sciences*, *6*(5), 1–44.

Ando, M. (1986). 生田流の箏曲 (*Koto of the Ikuta School*). H. Yoshikawa (Ed.). Tokyo, JP: Kodansha.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211-227.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617-645.

Bergson (1907/1911). *Creative evolution* (A. Mitchell, Trans.). New York, NY: Henry Holt and Company.

Bernstein (1976). Music. In Ratcliffe, S. (Ed.), *Oxford Essential Quotations*. Retrieved 13 Aug. 2018, from https://www.oxfordreference.com/view/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00007743.

Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nature Neuroscience*, *2*(4), 382–387.

Boasen, J., Takeshita, Y., Kuriki, S., & Yokosawa, K. (2018). Spectral-spatial differentiation of brain activity during mental imagery of improvisational music performance using MEG. *Frontiers in Human Neuroscience*, *12*(156), 1–12.

Bohr, N. (1960). *Atomic physics and human knowledge.* New York, NY: John Wiley & Sons.

Bonin, T., Degtyareva, K., Hahn, E. (2019). Investigating the content and form of musical imagery evoked by excerpts of Mahler's 5th symphony. Paper presented at the CIRMMT-OICRM-BRAMS Symposium. May 9–10, 2019. Montreal, CA.

Bonin, T., & Smilek, D. (2016). Inharmonic music elicits more negative affect and interferes more with a concurrent cognitive task than does harmonic music. *Attention, Perception, & Psychophysics*, *78*(3), 946-959.

Bonin, T., Trainor, L. J., Belyk, M., & Andrews, P. W. (2016). The source dilemma hypothesis: Perceptual uncertainty contributes to musical emotion. *Cognition*, *154*, 174-181.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT press.

Brown, D. (2015). Richard Wagner. *The Esoteric Quarterly, 10*(4), 55–83.

Butler, D. (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), 219-241.

Butler, D. (1990a). A study of event hierarchies in tonal and post-tonal music. *Psychology of Music*, *18*(1), 4-17.

Butler, D. (1990b). Response to Carol Krumhansl. *Music Perception*, *7*(3), 325-338.

Cabanac, M. (1996). On the origin of consciousness, a postulate and its corollary. *Neuroscience & Biobehavioral Reviews*, *20*(1), 33-40.

Cabanac, M. (1999). Emotion and phylogeny. *Journal of Consciousness Studies*, *6*(6-7), 176-190.

Campbell, J. (1949). *The hero with a thousand faces.* New York, NY: Pantheon Books.

Colman, A. M. (2009). *A dictionary of psychology*. Oxford, UK: Oxford University Press.

Condo, G. (2017). *The way I think.* Interview for Louisiana Museum of Modern Art. Soho, New York.

Cosmides, L., & Tooby, J. (1994). Better than rational: Evolutionary psychology and the invisible hand. *The American Economic Review*, *84*(2), 327-332.

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. *Handbook of Emotions*, *2*(2), 91-115.

Descartes, René (1644/1983). *Principia philosophiae* (V. Rodger & R.P. Miller., Trans.). Dordrecht, NL: Reidel.

Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review*, *3*(4), 357-370.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*(4), 341-354.

Dowling, W. J. (1999). The development of music perception and cognition. In D. Deutsch (Ed.), *The psychology of music* (2nd Ed.) (pp. 603-625). San Diego, CA: Academic Press.

Dowling, W. J., Tillman, B., & Ayers, D. F. (2001). Memory and the experience of hearing music. *Music Perception, 19*(2), 249-276.

Galileo (1623/1957) *The assayer.* In S. Drake (Ed.), *Discoveries and opinions of Galileo.* New York, NY: Doubleday.

Gendlin, E.T. (1970). The significance of felt meaning. In R. Cormier, E. Chinn & R.H. Lineback (Eds.), *Encounter: An introduction to philosophy* (pp. 561–566). Glenview, IL: Scott, Foresman & Co.

Gibson, J. J. (1960). The concept of the stimulus in psychology. *American Psychologist*, *15*(11), 694-703.

Gibson, J. J. (1975). Events are perceivable but time is not. In J. T. Fraser & N. Lawrence (Eds.), *The study of time II* (pp. 295-301). New York, NY: Springer-Verlag.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin Harcourt.

Gill, S. P. (2015). *Tacit engagement: Beyond interaction.* New York, NY: Springer Cham.

Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für mathematik und physik*, *38*(1), 173-198.

Gödel (1961/1995). The modern development of the foundations of mathematics in light of philosophy. In S. Feferman, J. Dawson, S. Kleene, G. Moore, R. Solovay & J. van Heihenoort (Eds.), *The collected papers of Kurt Gödel. III: Unpublished essays and lectures.* Oxford, UK: Oxford University Press.

Godøy, R. I. (2019). Musical shape cognition. In M. Grimshaw-Aagaard, M. Walther-Hansen & M. Knakkergaard (Eds.), *The Oxford Handbook of Sound and Imagination II* (pp. 237–258). New York, NY: Oxford.

Good, M. (2013). Interview for CJXY FM radio, June 27, 2013. Hamilton, ON.

Gray, J. A. (1982). Précis of The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system. *Behavioral and Brain Sciences*, *5*(3), 469-484.

Hofstadter, D., & Sander, E. (2013). *Surfaces and essences: Analogy as the fuel and fire of thinking*. New York, NY: Basic Books.

James, W. (1909). *The meaning of truth*. Mineola, NY: Dover Publications.

Jung, C. G. (1922/1923). *Psychological types: Or the psychology of individuation.* (H.G. Baynes, Trans.). New York, NY: Harcourt, Brace & Company.

Jung, C. G. (1944/1968). *Psychology and alchemy* (R.F.C. Hull, Trans). Princeton, NJ: Princeton University Press.

Jung, C. G. (1951/1959). *Aion: Researches into the phenomenology of the self.* (R.F.C. Hull, Trans.). Princeton, NJ: Princeton University Press.

Jung, C. G. (1956/1970). *Mysterium coniunctionis: An inquiry into the separation and synthesis of psychic opposites in alchemy*. (R.F.C. Hull, Trans.). Princeton, NJ: Princeton University Press.

Kant, I. (1783/1902). *Prolegomena to any future metaphysics that will be able to present itself as a science*. (P. Carus, Trans.). Chicago, IL: Open Court Publishing Company.

Köhler, W. (1947/1970). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York, NY: W.W. Norton & Company.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York, NY: Oxford University Press.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*(4), 334-368.

Laban, R. (1960). *The mastery of movement.* London, UK: MacDonald & Evans.

Lakoff, G. (1987). *Women, fire, and dangerous things.* Chicago, IL: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119-159.

Lecoq, J. (2001). *The moving body (Le corps poétique): Teaching creative theatre.* New York, NY: Routledge

Levenson, R. W. (1999). The intrapersonal functions of emotion. *Cognition & Emotion*, *13*(5), 481-504.

Lévy-Bruhl, L. (1938). *L'expérience mystique et les symboles chez les primitifs*. Paris, FR: Alcan.

Locke, J. (1690). *An essay concerning human understanding. Book II: Ideas.* Retrieved from https://www.earlymoderntexts.com/assets/pdfs/locke1690book2.pdf. Accessed July 1, 2019.

Mach, E. (1906/2011). *Space and geometry: In the light of physiological, psychological, and physical inquiry* (T.J. McCormack, Trans.). Ocean Shores, WA: Watchmaker Publishing.

Margolis, E. & Laurence, S. (2018). Concepts, *The Stanford Encyclopedia of Philosophy,* Edward N. Zalta (ed.), Retrieved from https://plato.stanford.edu/archives/sum2019/entries/concepts/. Accessed May 1, 2018.

Margolis, E. & Laurence, S. (2015). *The conceptual mind: New directions in the study of concepts*, Cambridge, MA: MIT Press.

Mayer, J. (2016, November). *JoJo Mayer - PASIC16.* Workshop presented at PASIC 2016. Indianapolis, IN.

McAdams, S. (1984). The auditory image: A metaphor for musical and psychological research on auditory organization. In W. R. Crozier & A. J. Chapman (Eds.), *Cognitive processes in the perception of art* (pp. 289-323). Amsterdam, NL: North-Holland.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*(3), 177-192.

McGilchrist, Iain (2009). *The master and his emissary.* New Haven, CT: Yale University Press.

Meier, C. A. (Ed.). (2001). *Atom and archetype: The Pauli Jung letters, 1932-1958* (Vol. 631). Princeton, NJ: Princeton University Press.

Näätänen, R. & Gaillard, A.W.K. (1983). The orienting reflex and the N2 deflection of the event-related potential (ERP). In A.W.K. Gaillard & W. Ritter (Eds.), Tutorials in event related potential research: Endogenous components (pp. 119-141). Amsterdam: North Holland.

Neumann, E. (1949/1970). *The origins and history of consciousness* (R.F.C. Hull, Trans.). Princeton, NJ: Princeton University Press (Bollingen)

Neumann, E. (1955/1972). *The great mother* (R. Mannheim, Trans.). Princeton, NJ: Princeton University Press.

Neumann, E. (1959). *Art and the creative unconscious.* London, UK: Routledge & Kegan Paul

Noble, J. (2018). *Perceptual and semantic dimensions of sound mass* (Doctoral dissertation, McGill University, Montreal, QC).

Peterson, J. B. (1999). *Maps of meaning: The architecture of belief.* London, UK: Routledge.

Peterson, J. B. (2000). *The pragmatics of meaning.* Retrieved from https://semioticon.com/frontline/jordan_b.htm. Accessed June 1, 2018.

Peterson, J. B., & DeYoung, C. G. (2000). Metaphoric threat is more real than real threat. *Behavioral and Brain Sciences*, *23*(6), 992-993.

Piaget, J. (1971a). *Biology and knowledge: An essay on the relations between organic regulations and cognitive processes* (B. Walsh, Trans.). Chicago, IL: University of Chicago Press.

Piaget, J. (1971b). *The theory of stages in cognitive development*. In D.R. Green, M.P. Ford, & G.B. Flamer (Eds.), *Measurement and Piaget* (pp. 1–11). New York, NY: McGraw-Hill.

Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures* (A. Rosin, Trans.). New York, NY: Viking.

Polanyi, M. (1966). *The tacit dimension.* Chicago, IL: The University of Chicago Press.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3-25.

Razran, G. (1961). The observable and the inferable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychological Review*, *68*(2), 81-147.

Rubia, K., Smith, A. B., Brammer, M. J., & Taylor, E. (2003). Right inferior prefrontal cortex mediates response inhibition while mesial prefrontal cortex is responsible for error detection. *Neuroimage*, *20*(1), 351-358.

Rubia, K., Smith, A. B., Brammer, M. J., Toone, B., & Taylor, E. (2005). Abnormal brain activation during inhibition and error detection in medication-naive adolescents with ADHD. *American Journal of Psychiatry*, *162*(6), 1067-1075.

Rumsfeld, D. H. (2002). *DoD news briefing – Secretary Rumsfeld and Gen. Myers.* February 12, 2002, 11:30 AM EDT. Retrieved from https://archive.defense.gov/Transcripts/Transcript.aspx?TranscriptID=2636. Accessed Aug 13, 2019.

Russell, B. (1967). *Autobiography*. Oxford, UK: Routledge.

Seligman, P. (1974). *Being and not-being: An introduction to Plato's Sophist*. The Hague, NL: Nijhoff.

Schaeffer, P. (1966/2017). *Treatise on musical objects: An essay across disciplines.* (J. Dack & C. North, Trans.). Oakland, CA: University of California Press.

Scheirer, E. D. (2000). *Music-listening systems* (Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA).

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America,* 1964, *36,* 2346-2353.

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89, 305–333.

Spencer-Brown, G. (1969). *Laws of form*. London, UK: Allen & Unwin.

Stetson, C., Fiesta, M. P., & Eagleman, D. M. (2007). Does time really slow down during a frightening event? *PloS One*, *2*(12), e1295.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J.H. Barkow, L. Cosmides, J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (19–136). New York, NY: Oxford University Press.

Whitaker, A. K. (Ed.). (1996). *Plato's Parmenides*. Indianapolis, IN: Hackett.

Whitehead, A. N. (1925). *Science and the modern world*. Cambridge, UK: Cambridge University Press.

Whitehead, A.N. (1929). *Process and Reality. An Essay in Cosmology.* Cambridge, UK: Cambridge University Press.

Whitehead, A. N. & Russell, B. (1910), *Principia mathematica,* Volume 1. Cambridge, UK: Cambridge University Press.

Whitehead, A. N. & Russell, B. (1912), *Principia mathematica,* Volume 2. Cambridge, UK: Cambridge University Press.

Whitehead, A. N. & Russell, B. (1913), *Principia mathematica,* Volume 3. Cambridge, UK: Cambridge University Press.

Wigner, E. P. (1961), Remarks on the mind-body question. In: I. J. Good (Ed.), *The Scientist Speculates* (pp. 171–184). London, UK: Heinemann

Wittgenstein, L. (1921/1922). *Tractatus logico-philosophicus* (C.K. Ogden, Trans.). London, UK: Routledge & Kegan Paul.

Zhang, Y., Chen, G., Wen, H., Lu, K. H., & Liu, Z. (2017). Musical imagery involves Wernicke's area in bilateral and anti-correlated network interactions in musicians. *Scientific Reports*, *7*(1), 7:17066. doi: 10.1038/s41598-017-17178-4

Appendix 1A: A brief review of the psychology and neurophysiology of creative activity

The scientific understanding of musical creativity may be facilitated through considerations of the evolutionary (e.g., Levenson, 1999; Tooby & Cosmides, 1992; Cabanac, 1996), neurophysiological and behavioural (Gray, 1982; Blood, Zatorre, Bermudez & Evans, 1999; Cabanac, 1999), and cognitive-cybernetic (Gibson, 1960; Piaget, 1977; Lakoff 1987; Cabanac, 1996; Barsalou, 2008) processes that coordinate the sensory, emotional, perceptual, and cognitive modalities of an individual organism within a creative act of consciousness. This review is intended to provide a cursory overview of these processes, which are sometimes only partially communicable by artists themselves.

The liminal communicability of artistic processes and creative experience is due in part to the fact that the psychological materials of creative experience are not fully accessible to verbal and conceptual articulation. Given the fact that creative experience is characterized by the breach of consciousness by previously unknown entities (e.g., Jung, 1944/1968, 1951/1959; Neumann, 1955/1972), this may not be so surprising. Indeed, the very task of the artist is to *further pursue* the source and substance of these unknown entities in order to formulate a coherent (intelligible) and meaningful (relevant) framework in which they can be associated with previously known entities.[32] This is to say that the operational realm of creative work is the boundary between "the known" and "the unknown."[33] Creative work is predicated on the insight that "the unknown" is in fact a constrained and *known* construct; a *known unknown,* to use Donald Rumsfeld's (2002) terminology.[34] The apparent paradox of this statement is clarified by stating instead that "the

[32] cf., Leonard Bernstein: "Music . . . can name the unnameable and communicate the unknowable." (Ratcliffe, 2016)
[33] Or "consciousness and the unconscious" in psychological terminology
[34] Conscious materials deemed "unknown" must be some extent known if they have been distinguished from "known" materials.

unknown" *thing* exhibits an *incomprehensible* form, (cf., Peterson, 2000). Thus, although the *form* of a *thing* termed "unknown" is insufficiently parameterized with respect to the context in which it is situated, this form is nevertheless susceptible to conscious manipulation and exploration by virtue of the same psychological faculties[35] of rational construal that allow for the manipulation of "known" i.e., *sufficiently parameterized* entities. The *creative act* is to pursue and intentionally engage in the natural psychological processes and generate comprehensible formulations of the incomprehensible entities that continually impinge on consciousness.[36] In effect, it is an expansion of an individual's dominion of consciousness.

Biologically speaking, the individual's initial encounters with unknown entities are characterized by the behavioural orienting reflex to "errors" and "anomalies," (Razran, 1961; Näätänen & Gaillard, 1983), autonomic motivational states that direct the individual's energies towards or away from certain metabolic processes (Levenson, 1999; Tooby & Cosmides, 1992; Cabanac, 1999) and neurophysiological activation of the right parahippocampal (Blood, Zatorre, Bermudez & Evans, 1999), mesial (Rubia, Smith, Brammer & Taylor, 2003), and prefrontal (Rubia, Smith, Brammer, Toone & Taylor, 2005) cortices that underlie the perceptual and conceptual updating processes. These neurobehavioural processes are coupled with heightened states of attention (e.g., Posner, 1980; Stetson, Fiesta & Eagleman, 2007) and modified affective valence and intensity (Levenson, 1999; Peterson, 1999; Blood, Zatorre, Bermudez & Evans, 1999) that underlie and in some cases characterize the conscious experience of an individual's encounter with the unknown. Thus, the experience of *incomprehensibility* is constrained by patterned behavioural, motivational, neurological, and experiential responses within the

---

[35] The perceptual faculties that demarcate the boundaries and forms of conscious materials; see section "Axioms and Background."
[36] Bergson (1907/1911)

psychophysiology of the individual. These responses lay the initial groundwork for the *parameterization* of the subject's perception of "the unknown."[37] Critically, the same physiological processes respond to material and ideal anomalies (Peterson & DeYoung, 2000), indicating that the neurophysiology of the human organism regards physical and psychological parameter spaces as functionally analogous (Peterson, 2000) and differentiates and refines these parameter spaces using common perceptual processes (e.g., Boasen et al., 2018; Zhang et al., 2017). From this initial point of departure, the individual begins to construct comprehensible[38] perceptual representations of the material and ideal aspects of reality.

The *forms* of the resultant images[39] possess affine structures constrained by the dimensionality of the prevailing space of consciousness in which the incomprehensible entity was rendered comprehensible (cf., the assumptions of *perceptual geometry*). The artist who explores this process therefore becomes aware of the constraints that sensory, emotional, perceptual, and cognitive systems place on coherent subjective experience and the possibility (and plausibility) of subjective states that could emerge through the manipulation and coordination of these processes. Behavioural patterns serve as the fundamental mode of characterizing and translating unknown and unparameterized sense data into comprehensible percepts and situating the individual in space (e.g., Mach, 1906/2011; Gibson, 1960, 1979; Piaget, 1971b). Motor patterns and postures are associated with a limited range of *affects* and *motivational states* (Laban, 1960; Lecoq, 2001). The emotional vector serves to determine how the anomalous unknown information is to be parsed into existing knowledge structures (Lakoff,

---

[37] cf., "High-resolution levels of behavioural operation constitute sub-elements of low-resolution conceptualizations…and are governed (sequenced, hierarchically rank-ordered, and evaluated) in consequence of their relevance as affordances, obstacles, or irrelevances in relation to those lower-resolution conceptualizations," (Peterson, 2000).

[38] "Known"

[39] In the case of music experience, a *musical image.*

1987; Mach, 1906/2011), and whether it is necessary to construct new knowledge structures to accommodate or reorient the individual's interpretive framework (the processes of *assimilation,* and *accommodation,* respectively, resulting collectively in an *equilibration* of cognitive structures, Piaget, 1971a, 1971b, 1977). In turn, these emotional dimensions serve to further enhance the individual's comprehension of objects by positioning them in a basic motivation-vectoring notion of *time* (Laban, 1960). Through this incorporation of new information and the associated compensations made within the entire conceptual scheme, the individual experiences a *perceptual stream* of *functionally and structurally discrete entities* (Gibson, 1975, 1979; Barsalou, 1983). These percepts entail distinct, bounded, and identifiable forms, the differentiation and relation of which serve as the basis of a coherent consciousness of reality.

If an individual willingly engages in the creative act, they are thereby taking it upon themselves—implicitly or explicitly—to seek out and master this process of perceptual image formation. Such a task relies heavily on the introverted analysis of the ideal entities of conscious experience.[40] If one becomes an expert at observing this process, they can create material substrates (art objects, words, behaviours) whose function is to systematically interact with this process in other individuals for the purpose of *reinstating* the modes of experience that result from this procedural cascade. Indeed, such an expert is capable of engaging in this process without any explicit knowledge of the underlying physical, physiological, or psychological processes that mediate or substantiate its success. Considering the life histories of the greatest artistic minds of human history; it may even be reasonably stated that the best artists are those who are *not* concerned with these scientific descriptions and are instead fully engrossed in this creative act (Neumann, 1959).

---

[40] Or the "empirical investigation of phenomenological ontology".

The foregoing sampling and considerations of the physiological, neurological, and psychological underpinnings of creativity, knowledge formation, and perception has intended to reveal the *psychological* complexity of the artist's task (to say nothing of the technical and aesthetic challenges!). In this way, it has hopefully facilitated some understanding of the difficulty artists often exhibit when expressing their process and experience. The "richness," or *parametric complexity*, of creative experience resulting from an introverted analysis of perceptual images often resists the traditionally "rational" verbal and symbolic descriptions of science, since these are designed to coherently describe a particular system of *external* phenomena that may possess only a partial or functionally limited resemblance with the structure of the images themselves. In the case of introverted analysis, the rational system subsumes any number of subjective associations with any number of physical systems under a coherent model of *psychological organization*. Hence the stereotypical tendency on the one hand of the artist to speak in metaphors and analogies (which seek to elucidate the subjective coordination implicit between symbolic schemes), and on the other of the "tortured artist" faced with the dilemma of *either* concisely *or* entirely expressing their insights with the expressive tools of "the real world."

The artist is consequently left to liminal and pre-verbal modes of codified description (Laban, 1960; Gendlin, 1970; Peterson, 1999, pp. 11, 159; Lecoq, 2001), and the details of artistic works are likewise best understood through comparison with one's own liminal and pre-verbal experiences of the art object (see Jung, 1923; Campbell, 1949; Neumann, 1955/1972, 1959), rather than through merely arbitrary metaphorical or symbolic reductions which further interfere with a direct understanding of the initial psychological state (see esp. Jung, 1951/1959; Scheirer, 2000). Artistic discussions are for this reason profoundly implicit, allowing vast amounts of unspoken (and in many ways unspeakable) information to unfold through various metaphorical schemes and processes whose realizations approximate the Platonic ideals they subserve. One fundamental challenge facing the scientific investigation of the creation and perception of artworks is therefore the formulation of a conceptual framework that allows researchers to explicate these implicit processes in a way that maintains the coherence of the direct subjective experience of the individual. Only then can rational (and replicable) claims be made regarding the psychology of creative activity.

Appendix 1B: A brief review of the *idea* as an empirical fact of 20th century scientific investigation

The following is brief review of the development of the scientific epistemology of the 20th century as it applies to the investigation of abstract conceptual frameworks as empirical facts worthy of scientific investigation. I present four abstract axioms and argue that these axioms apply equally well within the context of the epistemological challenge facing contemporary science as within the subjective act of artistic creation.

Axioms

The primary axiom is the assumption that consciousness is coincident with the existence of some *thing*; where there is no *thing* whatsoever, there is no consciousness whatsoever, and where there is no consciousness whatsoever, there is no *thing* whatsoever.[41, 42] The coincidence of consciousness and *thing* is called a *moment.*

The second axiom is that a *thing* is finite, i.e., a *thing* is *bounded*. The finitude of a *thing* implies its differentiation from other *things*, from the prevailing "context" in which the *thing* exists, or from "nothing" (see footnote 42). Rationally parameterized, the boundary of a *thing* therefore exists within a *dimensionality* by which the *thing* and its surroundings are simultaneously differentiated from and related to one another.[43]

---

[41] As we will see, this is not a solipsistic position but rather an empirically grounded position which acknowledges that whatever does or does not exist beyond the realm of *things* available to conscious perception is by definition inaccessible to consciousness (and therefore unascertainable by scientific investigation). Cf., e.g., Kant's *noumenal* vs *phenomenal* realms; the distinction between a Wittgensteinian "fact" and a Wittgensteinian "thing"; etc.

[42] One may readily object, "but I have certainly been aware of 'nothing.'" According to this axiom, if consciousness observes "nothing," then "nothing" is the name (sign) ascribed to a *thing* which *is* naught*,* i.e., the sign "nothing" signifies a thing that exists *as*-not an other *thing* signified by an other sign*.*

[43] 縁起; *engi;* "mutual arising;" *pratītyasamutpāda* of the Buddhist canon

The third axiom states that the *appearance*, or manifestation of the *thing* to consciousness, is its *form.* The *form* of a *thing* serves the *identity* of the *thing as such*.[44] Thus, the *form* of a *thing* is contingent upon the *identity* of consciousness with a finite system of relations.

The fourth and final axiom states that the *likeness* of *things* is a function of the extent to which their *forms* correspond with a common identity of consciousness.

The* idea *as an empirical fact of 20th century scientific investigation

The twentieth century witnessed a remarkably pervasive recognition of the relativity between a finite set of observed *things* and the finite state of observation by which the *things* are observed (see footnote 43). This recognition developed in parallel within scientific fields as diverse as psychology (e.g., Jung, 1951/1959; Neumann, 1949/1970; Köhler, 1947/1970), mathematics (e.g., Spencer-Brown, 1969; Gödel, 1961/1995), and physics (e.g., Bohr, 1960; Wigner, 1961). This was not necessarily a welcome arrival in the history of empirical thought, as it coincided with the dissolution of the absolute frames of reference (according to concepts of e.g., *time*, *space*, *energy*, *matter,* etc.) necessary to the scientific models of 17th and 18th century Enlightenment science (see Whitehead, 1925; below). The Enlightenment thinkers attempted to do away with any and all “subjective interpretations” of their empirical observations, but in so doing inadvertently assumed an absolute equation between the scientific vantage point and the conceptual framework (“subjective interpretation”) according to which a given empirical observation is made. By the twentieth century, the fallibility of this assumption was becoming more and more evident, and the decoupling of the scientific attitude from an absolute frame of reference was found to be inevitable. The associated paradigm shift is perhaps most succinctly

---

[44] i.e., the form of the *thing* is its appearance to the consciousness in relation to which the *thing* coexists

exemplified in the trajectory from Whitehead and Russell's (1910, 1912, 1913) *Principia Mathematica,* which sought to derive all of mathematics from a common set of axioms, and the resultant *incompleteness theorems* of Gödel (1931), which demonstrated by the same logic that such an absolute derivation was in fact impossible. But this progression of thought was in no way restricted to mathematics and continues to pose considerable challenges for contemporary science. Though it is easy enough to discuss the linearity of this development after the fact, the first recognitions of this progression of scientific reason were the result of scientists contending with the very edge of established knowledge, and much of these insights were initially rejected.

For example, Ernst Mach (1838–1916), the polymath empiricist responsible for discoveries as wide-ranging as the psychophysiological differentiation of sensory arrays (cf., Mach bands) and the physical-acoustic discovery of the speed of sound (cf., Mach speed), was chastised for his pioneering assertions that the vague concepts of "matter" and of "atoms" were patently unscientific. Though his opinion was unpopular among his contemporaries, his views in *Space and Geometry* (1906/2011) foreshadowed the critical issue that was to come into full force in the following century:

> The question whether a given *physical* [sic] object is a straight line or the arc of a circle is not properly formulated. A stretched chord or a ray of light is certainly neither the one nor the other. The question is simply whether the object so spatially reacts [so reacts to spatial apprehension] that it conforms better to the one concept than to the other, and whether with the exactitude which is sufficient for us and obtainable by us it conforms at all to any geometric concept (pp. 136–137).

A few years later, William James' (1909) response to absolutist distortions of his pragmatic philosophy develops the same idea proposed by Mach, with a more explicit consideration for the psychological root of these distortions:

> Let me give the name 'vicious abstractionism' to a way of using concepts which may be thus described: We conceive a concrete situation by singling out some salient or important feature in it, and classing it under that; then, instead of adding to its previous characters all the positive consequences which the new way of conceiving it may bring, we proceed to use our concept *privately;*[45] reducing the originally rich phenomenon to the naked suggestions of that name abstractly taken, treating it [the actual situation] as a case of 'nothing but' that concept, and acting as if all the other characters from out of which the concept is abstracted were expunged. Abstraction functioning in this way becomes a means of arrest far more than a means of advance in thought. It mutilates things; it creates difficulties and finds impossibilities[46]; and more than half the trouble that metaphysicians and logicians give themselves over the paradoxes and dialectic puzzles of the universe may, I am convinced, be traced to this relatively simple source. The viciously privative employment of abstract characters and class names is, I am persuaded, one of the great original sins of the rationalistic mind (pp. 110–111).

---

[45] i.e., removed from and in the absence of its normal situation; James does not mean to imply the social connotations of "privacy" in the "possessive" or "personally exclusive" sense; cf., his use of the term *privative* (def. "marked by the absence, removal, or loss of something quality or attribute normally present") below.

[46] By "viciously" reducing the actual situation in its entirety to a "mere instance" of the impoverished and privately employed concept, which in fact corresponds only with a subset of the actual situation's "salient features," the concept is then erroneously assumed to represent the actual situation, i.e., in its entirety: Such a misuse of thought then leads the thinker the indignant proclamation of "paradoxes," as if it were the somehow by the fault of *reality itself* that the actual situation does not correspond with the impoverished concept which in fact only corresponds with a finite and subset abstraction of that situation. See Plato's *Parmenides* (Whitaker, 1996).

In 1921 Wittgenstein's *Tractatus Logico-Philosophicus,* introduced by Bertrand Russell, provides the first formal demonstration of the inefficacy of Russell's conceptual scheme (as embodied in, e.g., *Principia Mathematica*). Russell's scheme necessitated that the *name* of a sign correspond with an absolute *form* of the *thing* signified by that sign. This assumption led to what is known as "Russell's paradox," which arises by considering the proposition "The set of all sets that are not members of themselves." This proposition uses the same *name* ("set") to refer to two actually different *things,* each with their respective *forms* within the total rational scheme of the proposition. Because Russell assumed, however, that all instances of the same name within a proposition refer to the same *thing* (and therefore to the same logical *form* within the proposition), he encounters a paradox, in which such a "set" could only be a member of itself if and only if it were not a member of itself.[47] Wittgenstein resolves this paradox succinctly by decoupling the actual *things* represented by a logical proposition from the *names* by which they are represented. Thus, the *form* of a logical function (the appearance of the *thing*) is contingent upon its distinct identity and relations within the total context of the logical proposition; not with its "name," which may be irrationally applied to the representation of multiple *things* within the same framework, even if that framework is considered "rational" by its author. The work remains one of the most enigmatic demonstrations of pure logic, but Russell himself acknowledged Wittgenstein's genius in his introduction to the *Tractatus*, and in his autobiography notes that Wittgenstein's criticism of his work was "an event of first-rate importance" in his life, as he thereby saw that he "could not hope ever again to do fundamental work in philosophy," (Russell, 1967, p. 66).

[47] Russell assumes that because *in the ideal case* of a rational proposition the same name is only ever used to refer to the same thing, that therefore *it actually does* in any actual case presumed by the author to be a formulation a rational proposal. This is the privative development of concept resulting in the "vicious abstraction" of actual cases to "nothing but" instances of the ideal case warned against by James (1909).

Wittgenstein (1921) writes:

> 3.33 In logical syntax the meaning [*form*] of a sign ought never to play a rôle; it must admit of being established without mention…; it ought to presuppose *only* the description of the expressions [*things*].
>
> 3.331 From this observation we get a further view into Russell's *Theory of Types*. Russell's error is shown by the fact that in drawing up his symbolic rules he has to speak of the meaning of the signs.
>
> 3.332 No proposition can say anything about itself, because the propositional sign cannot be contained in itself (that is the "whole theory of types" [sic.]).
>
> 3.333 A function cannot be its own argument, because the functional sign already contains the prototype of its own argument and it cannot contain itself. If, for example, we suppose that the function $F(fx)$ could be its own argument, then there would be a proposition '$F(F(fx))$,' and in this the outer function F and the inner function F must have different meanings; for the inner has the form $\emptyset(fx)$, the outer the form $\psi(\emptyset(fx))$. Common to both functions is only the letter 'F', which by itself signifies nothing [i.e., has no *form* within the same rational framework as that of the *forms* to which it refers]. This is at once clear, if instead of '$(F(F(u))$' we write '$\exists\emptyset$: $F(\emptyset u).\emptyset u = Fu$.'[48] Herewith, Russell's paradox vanishes.

The elegance of Wittgenstein's solution comes at the cost of recognizing that *finite* and *localized things* are inextricably bound to *finite* and *localized* moments of observation according

[48] i.e., [This is at once clear, if instead of '$(F(F(u))$' we write '$F(\emptyset(u))$.']; In lay terms, Don't use the same name to refer to different *things* within the same conceptual framework.

to a particular conceptual framework.[49] That is, the observable relations between *things* and their surroundings are rational (comprehensible) only according to a particular frame of reference (the *dimensionality*[50]), and the reality of a particular observation being situated within a particular dimensionality cannot be altered by the mere use of a common name[51] which in some remote moment of observation[52] refers to another *thing* situated with respect to its respective frame of reference. Thus the assumption of the Enlightenment sciences that an invariant, absolute frame of reference accounts for all of the varied observable *things* was no longer tenable.

Four years later, and more than two decades after publishing the *Principia Mathematica* with his student, Russell, Whitehead (1925) addresses this assumption of an absolute frame of reference for rational observations with an explicit and cogent statement in his lectures on *Science and the Modern World.*

> There persists… a fixed scientific cosmology which presupposes the ultimate fact of an irreducible brute matter, or material, spread throughout space in a flux of configurations. In itself such a material is senseless, valueless, purposeless. It just does what it does do, following a fixed routine imposed by external relations which do not spring from the nature of its being. It is this assumption that I call "scientific materialism." Also it is an assumption which I shall challenge as being entirely unsuitable to the scientific situation at which we have now arrived. It is not wrong if properly construed… [i.e., as] abstracted from the complete circumstances in which they occur… But when we pass beyond the abstraction, either by more subtle employment of our

[49] See second axiom, above; see footnote 43
[50] See second axiom, above; see section "Perceptual Geometry," below
[51] Again see Plato's Parmenides (Whitaker, 1996); cf., footnote 46
[52] Even if this remote observation is physically located a fraction of a second later and on the other side of a parenthesis, cf., F(F(fx))

senses, or by the request for *meanings* and for coherence of[53] thoughts, the scheme breaks down at once. …

So long as any theory of space, or of time, can give a meaning, either absolute or relative, to the idea of a definite region of space, and of a definite duration of time, the idea of simple location has a perfectly definite meaning. This idea is the very foundation of the seventeenth century scheme of nature. Apart from it, the scheme is incapable of expression. I shall argue that among the primary elements of nature as apprehended in our immediate experience, there is no element whatever which possesses this character of simple location. It does not follow, however, that the science of the seventeenth century was simply wrong. I hold that by a process of constructive abstraction we can arrive at abstractions which are the simply-located bits of material, and at other abstractions *which are the minds included in the scientific scheme* [sic.]. Accordingly, the real error is what I have called The Fallacy of Misplaced Concreteness.

The advantage of confining attention to a definite group of abstractions is that you confine your thoughts to clear-cut definite things, with clear-cut definite relations. Accordingly, if you have a logical head, you can deduce a variety of conclusions respecting the relationships between abstract entities. Furthermore, if the abstractions are well-founded, that is to say, if they do not abstract from everything that is important to experience, the scientific thought which confines itself to these abstractions will arrive at a variety of important truths relating to an experience of nature. …

The disadvantage of exclusive attention to a group of abstractions, however well-founded, is that, by the nature of the case, you have abstracted from the remainder of

[53] i.e., for comprehensible formal coordination among

> things. In so far as the excluded things are important in your experience, your modes of thought are not fitted to deal with them. You cannot think without abstractions; accordingly, it is of the utmost importance to be vigilant in critically revising your *modes* of abstraction... [with respect to your momentary frame of reference]. …
>
> My point is that a further stage of provisional realism is required in which the scientific scheme is recast, and founded upon the concept of *organism* (pp. 22, 72–73).

Whitehead (1929) would go on to develop this thesis in another of his seminal works, *Process and Reality,* in which he argued that reality consists of fundamentally dynamic *processes* rather than static material objects, with the empirically visible "irreducible materials" of experience being the consequence of *a priori* interactions among these dynamic substrates:

> We diverge from [the Enlightenment thought of] Descartes [1644] by holding that what he has described as primary *attributes*[54] of physical bodies, are really the forms [cf., *appearance*] of internal relationships *between* actual occasions [cf., *things*]. Such a change of thought is the shift from materialism to Organic Realism, as a basic idea of physical science (p. 471).

Thus we can frame the epistemological problem in the following way: In any moment of observation, the observed *things*[55] are *absolutely observed*—i.e., there is nothing present in the finite measurement other than what is present in the finite measurement—yet the moment of observation is itself *finite, relatively* situated among alternative moments of observation, with the finite *things* appearing to a particular moment of observation contingent upon the finite system of

---

[54] cf., also Galileo's (1623/1957) and Locke's (1690) *primary qualities.*
[55] cf., axioms 1, 2 and 3

concepts according to which the observation is made.[56] This consideration of the fundamental interaction between the *observation* and the *act of observing* became paramount among the sciences of the 20th century.[57] The act of observing *per se* has come into the purview of the scientific disciplines (Wigner, 1961); the scientific understanding of a *thing* has itself become a *thing* to be scientifically understood.

How is this *moment of observation* which relates empirical facts to *understanding* to be parameterized and conceptualized within the scientific (i.e., *knowledge generating*) scheme? As Whitehead (1925) indicates, "by a process of constructive abstraction we can arrive at abstractions which are the simply-located bits of material, and at other abstractions *which are the minds included in the scientific scheme*," (p. 72). This consideration arguably remains the basic theoretical challenge still facing contemporary science: "How are materials related to concepts?"

One formulation of this problem can be found in the psychophysical conceptualizations of the relation between the perceiving *subject* and the perceived *object.*[58] An extensive investigation of this topic took place between the Nobel prize-winning quantum physicist Wolfgang Pauli and the founder of analytical psychology Carl Jung from 1932 until Pauli's death in 1958 (Meier, 2001). After 21 years of an exceedingly erudite correspondence, Jung writes to Pauli on March 7 of 1953:

---

[56] See footnote 43

[57] Even among the most classically "hard" sciences of physics, see Bohr (1960); Wigner (1961).

[58] For the present purposes, the term *subject* is used in its original scientific sense to refer to the conscious entity within the *consciousness-thing* dichotomy (See first axiom). It is not intended to "reduce" the entire individual in which the consciousness is situated to this singular role; it is intended to singularly refer to this particular capacity of the total individual to be conscious. This is in recognition of the fact that the theory and empiricism of this work is restricted by its scientific viewpoint from addressing the total individual, and instead refers only to the delimited *subject* of its investigation. "Participant" or "listener" will be used when the entire individual is in question, as in, e.g., "participant demographics".

Without wishing to cast aspersions on Bohr's originality,[59] I should nevertheless like to point out that Kant had already demonstrated the necessary antinomy of all metaphysical statements. Of course, this also applies to statements concerning the unconscious, in that the latter is in itself non-ascertainable.[60] As such, it can either be 'a potential being' or 'nonbeing'.[61] I would, however, place these last two concepts in the category of metaphysical judgments, where in fact all concepts of 'being' belong. Aristotle was not able to create sufficient distance from the influence of Plato to see the merely postulated character of his concepts of 'being.'

In that 'spiritualism' and 'materialism' are statements on Being, they represent metaphysical judgments. They are only admissible [in scientific inquiry] as necessary elements in the process of apperception; namely, as the labeling of categories of… [observed entities], such as 'that is of mental (or spiritual) origin' or 'that is of physical (or material) origin.'[62] Metaphysical judgment, however, always places an element of the [observer's] psyche in an external location, thus preventing a union of Idea and Matter. Only in a third medium (in the *τρίτον είδος* ['third kind'] of Plato…) can the union of the two spheres take place, where both Spirit and Matter are removed from their 'in and for itself being'[63] and adapted to this third medium—namely, the *psyche of the observer* [sic.]. Nowhere else but in the psyche of the individual can the union be completed and the essential identity of Idea and Matter be experienced and perceived. I view metaphysical judgments—forgive this heresy—as a relic of the primitive *participation*

[59] Regarding his concept of *complementarity* among mutually exclusive quantum mechanical states
[60] See first axiom, see footnote 41
[61] See first and second axioms
[62] cf., Seligman (1974) regarding the Eleatic canon of Parmenides' *way of truth* presented in *On Nature* (c. 515 BC): "The decision rests in this: It is or it is not (fr. VIII, 15 f.). It is, and it is impossible for it not to be (fr. II). It is impossible that things that are-not are (paraphrased from fr. VI and VII)."
[63] Kant's *an sich* (cf., *Prolegomena,* 1783/1902)

mystique [Levy-Bruhl, e.g., 1938], which forms the main stumbling block to the attainment of an individual consciousness. … [This] stumbling block… is [the recognition] that the opposition is not *physis* versus *psyche*, but *physis* versus *pneuma* [*spirit*/*idea*], with *psyche* the medium between the two. In recent history, the spirit has been brought into the *psyche* and been identified with the function of the intellect.[64] In this way, the spirit has virtually disappeared from our field of vision and been replaced by the *psyche*; we find it difficult to attribute to the spirit an autonomy and reality that we ascribe to matter without a moment's hesitation. … Metaphysical judgments [such as these] lead to one-sidedness such as spiritualization or materialization, for they take a more or less large or significant part of the psyche and situate it either in Heaven or in earthly things, and then it can drag the whole person along with it, thus depriving him of his middle position.

If, in epistemological self-limitation we characterize Spirit and Matter 'in and for itself' as nonascertainable, this does not detract in any way from their metaphysical Being, for it is absolutely impossible for us even to approach it. But we have prevented the projection of the psychic into an external location[65], thus promoting the integration of the wholeness of man.

The psyche as *τρίτον είδος* and as a medium participates in both Spirit and Matter. I am convinced [in this sense] that… the psyche… is partly of a material nature. The archetypes, for example, are *ideas* (in the Platonic [Spiritual] sense) on the one hand, and yet are directly connected with physiological processes on the other. … It is part of the

[64] cf., the epistemological assumption of the Enlightenment sciences, introduced above and discussed further below

[65] i.e. precluded the possibility of James' "vicious abstractionism" or Whitehead's "fallacy of misplaced concreteness"

nonascertainability of their being that… [the metaphysical concepts of Idea/Spirit and Matter] cannot be situated in place.[66] This is particularly [evident in] the case of the archetype of wholeness—that is, of the Self.[67] It is the One and the Many, *ἓν τὸ πᾶν* [wherein and everything]. As you rightly say, the wholeness of man holds the middle position, namely between the *mundus archetypus* [ideal manifestation of the world], which is *real,* because it acts, and the *physis* [material manifestation of the world], which is just as *real,* because it acts. The principle of both, however, is unknown and therefore not ascertainable. Moreover, there are grounds for supposing that both are just different aspects of one and the same principle; hence the possibility of setting up identical or parallel physical and psychological propositions on the one hand, and on the other the psychological interpretability of religious revelations [i.e., parallel spiritual and psychological propositions]. (Theologians have the same resistance to psychologists as physicists, except that the former believe in *Spirit* and the latter in *Matter*.[68])

The fact that on the whole our views coincide is very pleasing to me, and I am very grateful to you for presenting your opinions in such detail (Meier, 2001, pp. 100–101).

Pauli (1953) responds a few weeks later, March 31 of that same year:

I believe in fact—not as dogma but as a working hypothesis—in the essential identity (*homo usia*) of the *mundus archetypus* and *physis* as you formulate it on p. 6 of

[66] i.e., within an empirical observation or conceptual framework as a finite and distinct *thing*
[67] The Self is an exceedingly nuanced empirical construct developed in Jung's writings between 1913 and his final publication *Mysterium Coniunctionis* (1956/1970); it is admitted by Jung only as an impoverished frame of reference with respect to which empirical observations of the fundamentally ineffable "centre" or "core" of *the psyche* could be coordinated; cf., aniconistic conceptions of the godhead.
[68] cf., Whitehead (1925) on the prevailing scientific cosmology of absolute matter; above

your letter. If this hypothesis is valid—and the possibility of physical and psychological parallel statements support this—then it must be expressed conceptually. In my view, this can happen only by means of concepts that are neutral in relation to the opposition *psyche-physis.*[69] Now in fact such concepts do exist—namely, mathematical ones: The existence of mathematical ideas that can also be applied in physics seems to me possible only as a consequence of that *homo-usia* of the *mundus archetypes.* At this point, the archetype of *number* always comes into operation… No one is likely to say that *the object of mathematics is psychic,* for a distinction has to be made between mathematical concepts [*ideas*] and the experience of mathematicians (which certainly takes place in their psyches). On the other hand, it seems to me important that the... [psychological] background of the number concept should not be neglected. (Among mathematicians themselves there was for quite a while an odd tendency to degrade mathematical statements into mere tautologies [cf., Russell and Whitehead's *Principia*]. This endeavor seems to have failed [cf., Wittgenstein's and Gödel's solutions], since it was not possible to understand the consistency of mathematics in this way.) It is this number archetype that ultimately makes possible the application of mathematics in physics [i.e., the ability of *ideas* to correspond with *materials*]. On the other hand the same archetype has a connection to the *psyche* (cf., trinity, quaternity, mantic, etc.), so that here I feel lies a crucial key for a conceptual expression of the *homo-usia* of *physis, psyche,* and *spirit* (ideas, etc.) (Meier, 2001, pp. 106, 107).

What these thinkers are contending with is the logical necessity that the *subject*—the consciousness of the individual—cannot be construed as existing within the same domain as

[69] cf., Whitehead (1925), above; Wigner's twin paradox (1961)

either the physical objects (*materials*) or the conceptual objects (*ideas*) appearing to it. That is to say, *ideas* stand apart from the consciousness of the individual in the same way that *materials* stand apart from the consciousness of the individual. This proposition is incompatible with the *tabula rasa* epistemology of the Enlightenment sciences, which conceptualized ideas as equivalent with the subjective consciousness of the individual,[70] and reduced both of these entities to "imperfect impressions" of material objects.[71,72] We still find vestiges of this rationale in contemporary colloquialisms. For example, one often hears the claim that "ideas" are "in the head" in the same way that a person's "consciousness" is "in the head." But if an *idea* presents itself to a person's consciousness then this *idea* cannot logically be construed as being "the same" as the person's consciousness. In this experience, the *idea* is an *object* appearing to the *consciousness* that is the *subject.*[73]

Jung and Pauli note that this distinction between the consciousness of the individual and the ideas presenting themselves to the individual's consciousness was historically maintained in Western philosophy by employing the metaphysical notion of *Spirit* as distinct from the metaphysical notion of *Psyche.* These constructs allowed for a rational means of distinguishing between the respective domains in which *ideas* and a person's *consciousness* may be said to exist. Furthermore, this scheme included a third factor, *Physis,* according to which material *things* could be understood to exist within its respective domain, distinct from *Psyche* and *Spirit*. The authors reason that *ideas* (*concepts, images,* etc.) will need to be relieved of their equation

[70] cf., Jung (in Meier, 2001), quoted above: "In recent history, the spirit has been brought into the psyche and been identified with the function of the intellect. In this way, the spirit has virtually disappeared from our field of vision and been replaced by the *psyche*; we find it difficult to attribute to the spirit an autonomy and reality that we ascribe to matter without a moment's hesitation." (p. 100)
[71] See Descartes (1644/1983); Galileo (1623/1957) and Locke (1690); see footnote 54
[72] See Tooby & Cosmides (1992); Cosmides & Tooby (1994); Cosmides & Tooby (2000) for thorough and explicit criticisms of the *tabula rasa* model as an inadequate model of contemporary psychology.
[73] see axioms

with the consciousness of the perceiving subject if they are to find their appropriate place within a contemporary scientific epistemology. This is the proposition of an *objective idea.* To reframe the metaphysical statements of the ancient Greeks according to the epistemological development of the 20th century, these authors suggest the following empirical proposition: Both *ideal things* and *material things* partake of the role of a perceived "object" with respect to the perceiving "subject" of the individual's consciousness. Thus, Jung and Pauli's answer to the question, "How are materials related to concepts?" is "As distinct *objects*[74] according to the finite moment of observation[75] by the *subject*."

The foregoing considerations are particularly important to the scientific investigation of human creativity and the perception of musical form. In the creative act, an artist very often experiences the need to "come to grips" with a particular idea. Furthermore this *idea* is taken to be the original *thing* and point of reference of which the material *thing* is only an "imperfect representation." In this way, the postulate of an objective *idea* is central to the artistic experience. The creative act therefore requires a conceptualization of the objectivity of both materials and ideas—only if ideas are considered distinct *things* can their relative *forms* be ascertained, and only if the *forms* of ideas are ascertained can the *likeness* among them and their respective associations with particular *materials* be reasonably interpreted. The artistic experience therefore provides an exemplary case of the analogous function of *materials* and *ideas* as *objects*. While the existence of objective materials is intuitive to the contemporary sciences on the basis of the well-established epistemology of the Enlightenment, the postulate of objective *ideas* is

[74] cf., *things*
[75] cf., *consciousness*

noticeably absent within the scientific disciplines. A candidate scientific model of objective ideas may therefore be established on the basis of the creative process of artists.[76]

[76] See Appendix 1A: “A brief review of the psychology and neurophysiology of creative activity”

Chapter 2 - Cross-cultural perception of Japanese- and Western-typical melodic contour profiles

Abstract

We investigate whether the contour of a musical melody influences a listener's perception of its culture of origin. Eighteen (18) original Japanese folk melodies (Matsunaga et al., 2012, 2014) were transformed according to a 2 (*Tonality;* Japanese, Western) x 2 (*Contour;* Japanese, Western) matrix design, producing four stimulus variants of each of the 18 original melodies (Japanese tonality x Japanese contour [*JJ*]; Japanese tonality x Western contour [*JW*]; Western tonality x Japanese contour [*WJ*]; Western tonality x Western contour [*WW*]) resulting in 72 stimuli in total. Participants were sequentially presented with each stimulus randomly and without replacement and asked to rate the perceived musical style on a 7-point Likert scale, with the leftward bound semantically anchored as "Very Western," (「とても西洋的」) the rightward bound semantically anchored as "Very Japanese," (「とても日本的」) and the mid-point semantically indicating "cultural ambivalence" (「日本的でも西洋的でもない」). Experiment 1 presents Japanese listeners' perception of these stimuli (N = 122); Experiment 2 presents the perceptions of Western listeners (N = 210); and Experiment 3 presents the perceptions of Non-Japanese non-Western listeners (N = 62). The results of all three experiments indicate the same perceptual effect: Contour profiles serve as an independent predictor of listeners' judgments of the style of a musical melody. Our results indicate that culturally musical styles are determined not only according to well-known and previously documented perceptual schemata for musical tonality, but also to the characteristic temporal organization of musical tones throughout the duration of a melody. We discuss empirical and theoretical implications of our results for future music perception research.

Introduction

This work originally began as a collaborative investigation of the influence of musical timbre on the perception of Japanese and Western tonalities. Matsunaga, Abe and Yokosawa (2012, 2014) have previously demonstrated that distinct regions of the right inferior frontal gyrus are responsible for processing Japanese and Western tonalities, as indexed by the locus of an early right anterior negativity (mERAN) evoked by tonal oddballs in Japanese and Western melodic sequences (Koelsch, 2009). They also demonstrated that the effect is mediated by musical experience—the more musical experience one has, the more cortical overlap one observes in the tonal processing of Japanese and Western musics. This effect may be understood as an optimization of a general *musical tonality* schema that, with experience, comes to abstract tonal structures of any sort, resulting in the enhanced detection of contextual violations. Given that the perception of distinct musical tonalities between Japanese and Western musics is influenced by experience, and that musical timbre is known to influence the perception of tonality (Warrier & Zatorre, 2002; Paraskeva & McAdams, 1997; Russo & Thompson, 2005; Bonin, Trainor, Belyk & Andrews., 2016) we sought to determine whether Japanese and Western tonal sequences performed in culturally distinct timbres would influence the mERAN index of perceptual sensitivity to musical tonality.

In modifying the stimuli previously developed by Matsunaga et al. (2012, 2014) for use in these experiments, we noticed that, despite being transposed from a Japanese key into a Western key, such "Western tonality" variants of traditional Japanese folk melodies did not, in their own right, sound very "*Western*". More specifically, the melodies did not seem to "move" (cf., Zuckerkandl, 1973; Shove & Repp, 1995; Eitan & Granot, 2006) in the same way Western melodies tend to move, resulting in a strange "temporal impression" (or *temporality*) compared to that of prototypical Western tonal art music. Upon further reflection, we began to entertain the notion that these "Westernized" stimuli still possessed an abstract "Japanese" temporality. We then sought to parameterize these stimuli to determine which elements could have been contributing to this residual "Japanese" impression of time. Since these melodies were monophonic, isochronous, and possessing no metrical accents, we were left to conclude that the strange temporality of these stimulus variants was being conveyed through the interval patterns retained in the transposition of the traditional Japanese melodies. More precisely, we identified the progression of relative pitch heights throughout the duration of the melody, or the melodic *contour*.

We consulted with Japanese and Western musicians to determine how they would make these traditionally Japanese sequences, performed in a Western key "sound more Western, changing only the rising and falling sequences within the melody," and the primary feedback we received is that the melody would need to be "straightened out." Working with the musicians to further define what it meant to "straighten out" a melody, two principles became clear, although the specific parametric manipulations required to "Westernize" a particular melody were specific to each individual case: First, melodies in which the local pitch maxima and minima coincided regularly and consistently with strong pulses of the implied meter (beats 1 and 3 in the present case) were said to be "*straighter*" and "*more Western*" than those that did not. Second, melodies in which the implied harmonic transitions[77] coincided regularly and consistently with strong metric pulses were understood to be straighter and more Western than those that did not. Thus,

[77] Here, in speaking of both Japanese and Western musics, we refer to *harmony* in the broad sense of "the heard relation among musical sounds," structurally mediated through a tonal hierarchy, and not in the narrow sense constrained to the specific system extensively developed in Western music to allow or forbid particular relations between musical chords and tones (Rich, 2019). Within the Japanese Ritsu scale used in the following experiments, the 1^{st}, 5^{th} and 4^{th} scale degrees are the most stable tones of the scale and serve equivalent tonal functions as the tonic, dominant, and subdominant functions of Western music (Rose & Kapuscinsky, 2013).

the relative variability of the *phase alignments* among the metric, harmonic, and pitch height cycles was determined to be a critical factor distinguishing the *Japanese* and *Western* melodic contours. Whereas Western melodies were said to possess more homogenous, *monophasic* contour sequences, Japanese melodies by comparison were determined to possess a *polyphasic* contour, in which the phase relations among the metric, harmonic, and pitch height cycles vary throughout the melodic progression. Several previous research findings corroborate these artistic insights, leading further plausibility to the hypothesis that culturally specific melodic contour profiles contribute to the perceptual discrimination of Japanese and Western musical styles.

First, listeners' perceptions of musical form are influenced by pitch height dynamics within a melodic sequence. Interpreting continuous response data, Krumhansl (1996), Granot & Eitan (2011), Farbood (2012) have each found that progressions and transitions in local pitch height predict listeners' perception of structural nodes within a dynamic musical sequence. In particular, these studies have determined that local pitch maxima are associated with points of musical "tension," implying an impending return to a point of relative stasis, whereas local pitch minima are associated with musical "relaxation," associated with both the conclusion-arrival and initiation-departure of a musical phrase.

Second, perceptions of musical form are influenced by the implied harmonic motion of a melodic sequence. Cuddy, Cohen and Mewhort (1981) found that participants perceived melodic sequences possessing a "prototypic" harmonic progression as exhibiting more *musical structure* compared with those that did not. The authors created as their prototypic stimulus a 7-note tone sequence exhibiting "a I-V-I progression… [with] the leading note of the scale [VII] mov[ing] directly to the tonic." As comparisons, 31 melodic variations were created to alter, in varying degrees, the implied harmonic progression, diatonicism, and contour, of the prototypic sequence.

Results indicated that the highest ratings of musical structure were perceived when a melodic sequence incorporated the I-V-I progression, when the contour of the sequence was simpler, and when the penultimate and final tones of the sequence were the leading note and the tonic, respectively.

In a related study, Bigand (1990) created two melodic stimulus groups: a “true” melodic family (TF) and a “false” melodic family (FF). In the TF, all constituent melodies shared the same implicit harmonic progression, despite distinct surface rhythms and contours among them. In the FF, the surface rhythms and contours were again distinct, but the melodies did not share the same harmonic progression. Participants were trained on a TF or FF stimulus set. At test, they were presented with a randomized series of intermixed training stimuli and test foils and asked to identify which of the melodies they had previously heard in the training set. Participants trained on a TF set were more likely to identify training stimuli and reject test foils than those who were trained on an FF set. Interestingly, most successful participants completed this task intuitively, without explicit knowledge of which perceptual parameters were informing their judgments. Thus, the tones of a melodic sequence inherit the abstract temporal “structure” of the harmonic sequences they imply, and these structures are perceptually discernible by listeners, if only accessible to subjective report through intuitive appraisal of one’s experience.

The limited capacity listeners have to describe their perception of musical contour may relate to a third line of relevant research. Namely, the developmental primacy of melodic contour perception. That is, contour sensitivity reflects one of the earliest and most fundamental auditory processes listeners use to parse musical stimuli. Prior to the capacities for linguistic speech, relative interval identification, or musical key discrimination, human infants as young as 8 months will discriminate differences in melodic contour sequences (Trehub, Bull & Thorpe, 1984; Trehub & Hannon, 2006). Indeed, it has been suggested by some authors that this sensitivity to musical-auditory stimuli may arise from a common *musilinguistic* disposition that underlies the infants' ability to respond to the emotive cues conveyed through the rising and falling contours of its caregiver's voice (often referred to as *motherese*; Fernald, 1985). Consistent with this hypothesis are the findings that contour perception is both a universal auditory faculty (e.g., Hauser & McDermott, 2003), that contour continues to facilitate the perception and memory of musical stimuli[78] in adulthood (Dowling & Fujitani, 1971; Dowling, 1978; Dowling & Harwood, 1986), and that an adult's perceptual schema for musical contour reflects the normative linguistic and musical auditory patterns found in an infant's culture (Peng et al., 2010; Liu, 2013).

Particularly relevant to the present investigation involving the transposition of *Japanese-* and *Western-* melodic contour profiles between *Japanese* and *Western* key spaces are the findings of Dowling (1978) and Dowling and Fujitani (1971). Early work by Dowling and Fujitani (1971) investigated the influence of melodic contour in the absence of a prevailing tonal context. Specifically, the authors measured the rate of recognition for atonal standard melodies on the basis of whether or not the test-comparison melody possessed the same contour and began on the same pitch as the standard melody. The results indicated that contour was critical for the discrimination of transposed melodic sequences (starting on a different note) but played a lesser role in the discrimination of test-comparison melodies beginning on the same initial pitch (i.e., when the two stimuli were not transposed).

This experiment was replicated by Dowling (1978) with *tonal* (diatonic) melodies, thereby investigating the influence of contour on melody recognition in tonal contexts. In this

[78] particularly in the case of novel, unfamiliar stimuli

experiment, subjects found it easier to reject atonal test-comparison melodies when they were preceded by a tonal standard melody, regardless of whether the test and standard melodies began on the same note. However, subjects found it "extremely difficult"—performing at chance levels—when required to distinguish between standard tonal melodies and exact transpositions of these melodies into new tonal keys, as well as between standard tonal melodies and test melodies in which the same contour sequence was shifted throughout the same key as the standard. This result illustrates the dynamic relationship between contour and tonality.

Taken together, the reviewed works provide considerable evidence in favour of the prediction that the pitch height and implied harmonic motion influence the perception of musical contour, and that contour interacts with musical tonality in order to influence the perception and recognition of musical melodies. The question now remains whether or not traditional Japanese and Western melodies possess perceptually distinct contour profiles, whether these profiles can be operationally defined in terms of *polyphasic* and *monophasic* contour sequences, and whether these sequences contribute to listeners' judgments of the culture of origin of a given melodic sequence. We test this possibility in three experiments involving Japanese, Western, and non-Japanese non-Western listeners.

In the present experiments, traditional Japanese melodies (Matsunaga et al, 2012, 2014) were transformed to produce four stimulus variants. The first variant [*JJ*] was the original, unaltered Japanese melody, exhibiting a Japanese tonality and a Japanese contour profile. The second variant [*WJ*] was the JJ variant transposed into a Western key, exhibiting a Western tonality and a Japanese contour profile. The third variant [*WW*] was "straightened out" to exhibit both a Western key and a Western contour profile. Finally, the fourth variant [*JW*] was the WW variant transposed into the original Japanese key to exhibit a Japanese tonality and a Western contour profile. Participants of these experiments were sequentially presented with each of the stimuli randomly and without replacement and asked to rate the perceived musical style of each melody on a 7-point Likert scale, with the leftward bound semantically anchored as "Very Western" (「とても西洋的」), the rightward bound semantically anchored as "Very Japanese," (「とても日本的」), and the mid-point indicating "cultural ambivalence" (「日本的でも西洋的でもない」). To the extent that melodic contour serves as an independent diagnostic criterion of cultural origin, listeners should rate culturally homogenous melodies (i.e., JJ and WW variants) as more characteristic of their culture of origin than their culturally heterogeneous counterparts (i.e., JW and WJ stimulus variants). If the hypothesized distinction between Japanese polyphasic contour profiles and Western monophasic contour profiles does not bear any perceptual relevance, then participants should rate the JJ and JW as equivalently representative of Japanese style on the basis of their common Japanese tonality and the WW and WJ as equivalently representative of Western style on the basis of their common Western tonality, with any perceptible differences in contour failing to contribute systematically to a sense of musical style.

Experiment 1 – Perception of Japanese listeners

Methodology

Participants

One hundred and twenty-eight senior undergraduate students (76 males; average 20.4 years old) from Kanagawa University participated in the experiment as part of an in-class demonstration of music perception and experimental psychology. The data from three Chinese exchange students and three Japanese participants who did not complete the response form were excluded from the final analysis (but see Experiment 3), yielding a final sample of 122 Japanese participants (N = 122; 73 males; average 20.3 years old). Participants were not selected on the basis of prior musical experience, but 50 participants (41%) reported prior experience with an instrument, with an average of 5.36 years (range 1–15 years) of musical experience among them. This study was certified for ethical compliance by the McGill University Research Ethics Board II and all participants provided informed consent.

Apparatus

A pen-and-paper questionnaire was manually distributed, containing as its first section a demographic questionnaire (the results of which are presented in the *Participants* section above) and as its second section 72 enumerated instances of a 7-point Likert scale. The lower bound of the Likert scale (numerically coded as "1") was semantically anchored as "Very Western style," (「とても西洋的」); the upper bound ("7") was semantically anchored as Very Japanese style," (「とても日本的」); and the mid-point of the scale ("4") was semantically anchored as "Neither Japanese style nor Western style" (「日本的でも西洋的でもない」).

Stimuli

Eighteen *original melodies* were selected from a previous experimental corpus of monophonic isochronous Japanese folk melodies performed in the Ritsu scale (do, re, fa, sol, la) and a duple meter (Matsunaga et al., 2012, 2014). By restricting the stimuli to monophonic and isochronous melodies, we were able to isolate the respective effects of tonality and contour as strictly as possible. As mentioned above, each original melody was transformed three times according to a 2 (*Tonality;* Japanese, Western) x 2 (*Contour;* Japanese, Western) matrix design (see Table 2.1), resulting in four stimulus variants (Japanese tonality x Japanese contour [*JJ*]; Japanese tonality x Western contour [*JW*]; Western tonality x Japanese contour [*WJ*]; Western tonality x Western contour [*WW*]) of each of the eighteen original melodies, for a total of 72 stimuli.

Table 2.1 The transformation scheme used to produce the four stimulus variants of each of the eighteen original melodies.

	Japanese Tonality (*Ritsu*)	Western Tonality (*C major*)
Japanese Contour (*Polyphasic*)	JJ	WJ
Western Contour (*Monophasic*)	JW	WW

The *Tonality* dimension was treated as a categorical binary (see Matsunaga et al., 2012, 2014), with a *Japanese* pole represented by the Ritsu scale and a *Western* pole represented by the C-major scale (do, re, mi, fa, sol, la). To be clear, the intention of the *Tonality* dimension was not to represent a comprehensive or exhaustive representation of Japanese and Western tonality *in universa,* but to serve as a simplified and controlled independent variable that would reliably produce a perceptual impression of “Japanese tonality” and “Western tonality” within the listener

(Matsunaga et al., 2012, 2014).[79] Likewise, the *Contour* dimension was treated as categorical binary, with a *Japanese* pole represented by a *polyphasic* contour and a *Western* pole represented by a *monophasic* contour.

All stimuli were created using the Cubase 9 digital audio workstation (Steinberg Media Technologies GmbH, Hamburg). One of the musicians consulted in our preparations (see Introduction) was particularly interested in the experiment and offered to create the stimuli for us according to the tonality and contour principles outlined above. First, the eighteen *original melodies* were recorded verbatim into Cubase as MIDI data, constituting the JJ stimulus variants. As mentioned previously, these melodies were performed in the Japanese Ritsu scale with their original, characteristically Japanese polyphasic contours. Next, copies of these MIDI sequences were pasted to independent playback channels. These copies were subjected to the first manipulation of *Tonality,* now possessing a Western tonality and a Japanese contour and constituting the second set of stimulus variants [WJ]. This was accomplished by our musician replacing some instances of *do*, *re*, *sol*, and *la* within their respective original Japanese melodies with median (*mi*) and leading tone (*ti*) scale degrees in accordance with prototypical harmonic and voice leading sequences of classical Western art music. All *Tonality* manipulations were made while maintaining the melodic contour of the respective JJ stimuli, such that the only factor differentiating the JJ and WJ counterparts was their "*tonality"*, as manifest in the traditional Japanese Ritsu and traditional Western C-Major scales, respectively.

Third, copies of the WJ stimuli were copied to independent playback channels and subjected to a *Contour* manipulation, resulting in the WW stimulus variants. As outlined in the

---

[79] Critically, Matsunaga and colleagues (2012; 2014) have previously demonstrated that although the Ritsu scale degrees are *nominally* subsumed within the Western diatonic scale (i.e., the former set of tones constitutes a subset of the latter), they are not *perceived* as such, with listeners clearly distinguishing the respective tonal qualities of the Ritsu and C-major scales.

introduction, these manipulations were made on a case-by-case basis, according to which specific note events of a given WJ stimulus deviated from the desired monophasic contour profile typical of Western classical art music. The three contour parameters available for manipulation were the local pitch maxima and minima, the implied harmonic transitions, and the implied metric pulses within the melody. The WJ melodies were altered as minimally as possible to create a WW variant in which the temporal cycles of the local pitch maxima and minima, the implied harmonic transitions, and the implied metric pulses were aligned. The result was a reduction in the variability of the phase relations among the three cyclical temporal dimensions, and stimulus variants that sounded most typical of the classical Western musical style.

The resulting 18 WW stimulus variants were then submitted to the second *Tonal* manipulation, in which each stimulus was transposed back into the Japanese Ritsu scale to create the fourth and final JW stimulus variants. As in the first tonal manipulation of the JJ stimuli, all *Tonality* manipulations of the WW stimuli were made while maintaining the melodic contour of the respective WW stimuli, such that the only factor differentiating the WW and WJ counterparts was their culturally specific "*tonality,*" according to the scale in which they were performed. Figure 2.1 presents the four stimulus variants JJ, JW, WJ, WW of the first traditional Japanese melody as a visual aid to the interpretation of the stimulus design protocol; the *polyphasic* and *monophasic* contour profiles in particular are intuitively accessible through the visual representation of the musical score.

Fig. 2.1 JJ, JW, WW, and WJ variants of traditional Japanese melody #1

Procedure

Participants were informed verbally (in Japanese) that they would be presented with a sequence of 72 short musical excerpts, and that their task was to indicate on the corresponding instances of the Likert scale (1–7) their impression of how strongly the excerpt reflected a Western style (西洋的) or Japanese style (日本的) or whether it was culturally ambivalent (日本的でも西洋的でもない; *"indicative of n/either*[80]*"*). Four practice trials were presented to the participants to solicit any questions and provide clarification, though none were posed. The 72 stimuli of the experiment proper were randomly sampled and presented over the classroom's PA loudspeaker system without replacement at a comfortable and clearly audible listening level. A

[80] N.b. According to the ethos underlying the Japanese language there is no difference between the statement "indicative of either" or "indicative of neither"; This is because the polarity between Japanese and Western styles suggested by the question "Is this Japanese or Western?" implies that *Japanese* is-not *Western* and *Western* is-not *Japanese*. The statement 日本的でも西洋的でもない literally translated into English is "Japanese style and Western style is-not", i.e., "this stimulus does not afford the possibility of discriminating between Japanese and Western styles." This can be equivalently translated into English as "this stimulus could be arbitrarily considered *either* Japanese or Western" or "this stimulus is considered *neither* Japanese nor Western."

brief two-minute rest intervened between the 36th and 37th trials of the experiment proper. Data are presented in a recoded version of the Likert scale, with –3 corresponding to “Very Western”, +3 corresponding to “Very Japanese” and 0 corresponding to cultural ambivalence.

Results

Figure 2.2 presents the mean perceived musical style for each of the JJ, JW, WJ, and WW stimulus variants provided by the Japanese participants. The mean *melodic style* rating for each of the four stimulus variants provided by each participant was submitted to a two-way repeated measures ANOVA with Tonality (Japanese, Western) and Contour (Japanese, Western) submitted as the two within-subjects factors. The ANOVA confirmed the main effects of Tonality, $F(1, 121) = 203.12$, $p < 0.001$ and Contour, $F(1, 121) = 29.57$, $p < 0.001$, indicating that participants were perceptually receptive to both the musical scale and melodic phase manipulations used to create the four stimulus variants. There was no interaction between Tonality and Contour, indicating that the effects of Contour were equally present for melodies played in both the Japanese and Western scales, $F(1,121) = 0.83$, $p = 0.364$.

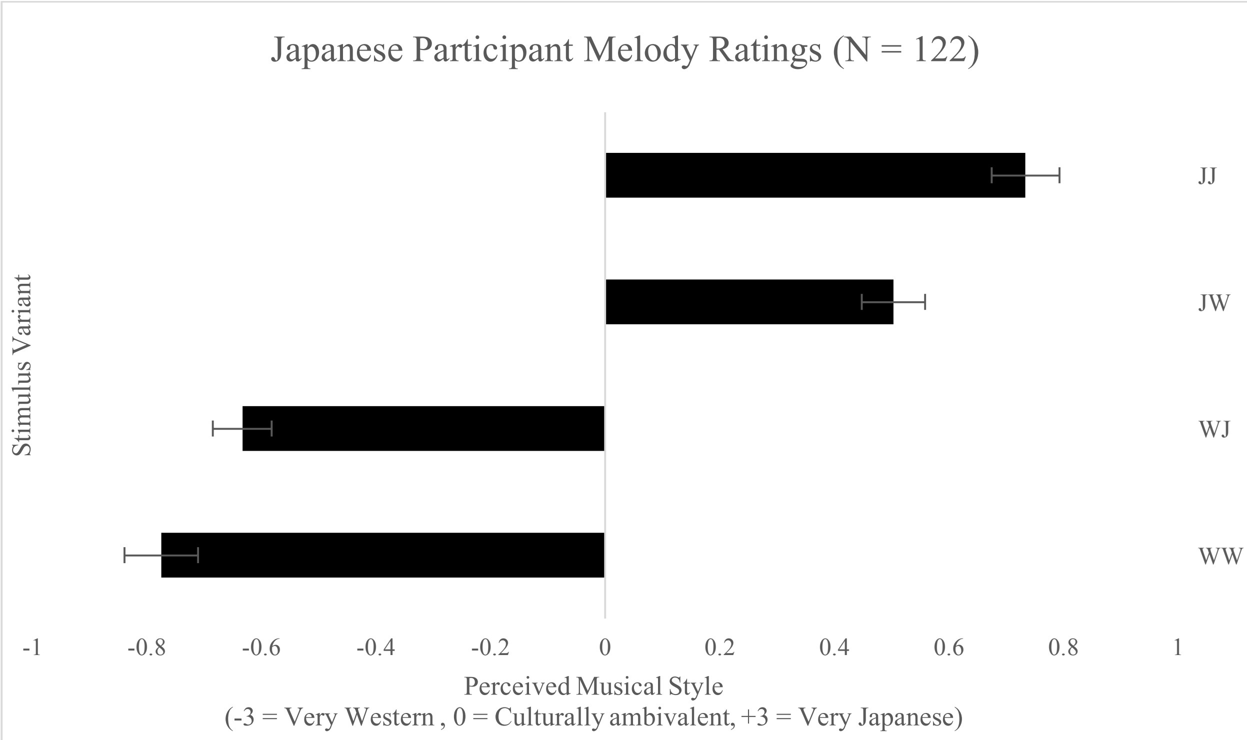

Fig. 2.2 Average *musical style* rating for each of the JJ, JW, WJ, and WW stimulus variants provided by the Japanese participants (N = 122). Error bars indicate standard error.

Summary and Discussion

Analyses of our Japanese participants' musical style ratings indicate that both tonality and contour contribute to a melody's perceived culture of origin. Stimuli possessing only a Japanese or Western tonality were not rated as stylistically characteristic of their musical traditions as those stimuli which also possessed the characteristic contour profiles of their musical tradition. These effects were observed regardless of the fact that participants had no knowledge of which stimulus parameters were being manipulated throughout the 72 randomly presented listening trials. The most culturally diagnostic stimuli (JJ) and (WW) received average absolute scores of 0.8 out of 3. Given that the JJ and WW stimuli are rated as equally diagnostic of their respective cultures, and that the JJ stimuli are original melodies drawn from a corpus of

Japanese folk songs (Matsunaga et al., 2012, 2014), this result likely reflects a response bias not to use the extreme ends of the Likert scale, rather than an indication that the JJ and WW stimuli were not intrinsically characteristic of their alleged culture of origin. That is, if for example the WW stimuli were rated as much less Western than the JJ stimuli were rated Japanese, it may be reasonably suspected that the experimental manipulations used to produce the WW stimuli were invalid; but instead these Japanese listeners seem to be using a constrained subset of the rating scale as a response bias. It is also possible that the stimuli selected for the experiment are not indicative of their respective alleged musical styles. The relative merit of these two interpretations can be adjudicated on the basis of the participant response averages presented in the following experiments—if the remaining samples report equally low average values then this may be interpreted in favour of the notion that the stimuli were ineffectively generated. If instead the average ratings produced by additional participant samples are nearer to the extreme values (–3 and 3) then this suggests a response bias particular to the present sample. This effect will be considered in the general discussion. Finally, the absence of a statistical interaction between the Tonality and Contour dimensions indicates that musical contour plays an equally important role in the perception of both Japanese- and Western-typical stimuli. Given that Japanese listeners are exposed to both traditional Japanese and traditional Western music from an early age (Tokita, 2014; Matsunaga et al., 2014) this result may reflect, in part, a perceptual specialization unique to the cultural experiences of Japanese listeners. We investigate this possibility in the following experiments by replicating this study with Western (Experiment 2) and non-Japanese non-Western (Experiment 3) participant demographics.

Experiment 2 – Perception of Western listeners

Methodology

Participants

Two hundred and ten undergraduate students (N = 206; 115 females; average age = 20.2 years) from McGill University participated in the experiment as part of an in-class demonstration of music perception and experimental psychology. Thirty (30) of these participants were of a European nationality, while the remaining 180 students were of North American nationality (i.e., not including Latin or Middle American countries). Participants were not selected on the basis of prior musical experience, but 155 participants (73.8%) reported prior experience with an instrument, with an average of 5.85 years (range 1–16 years) of musical experience among them. This study was certified for ethical compliance by the McGill University Research Ethics Board II and all participants provided informed consent.

Apparatus

The apparatus was the same as described in Experiment 1, with two exceptions: First, the lower bound of the 7-point Likert scale (1) read "Very Western", the upper bound (7) read "Very Japanese", and the mid-point (4) read "Equally Japanese and Western."[81] Second, the stimuli were played at a comfortable and clearly audible listening level over a different lecture hall PA system.

[81] Although the literal English translation of「日本的でも西洋的でもない」as "*N/either* Japanese style n/or Western style" may appear different from the chosen English phrase "Equally Japanese and Western," the desired sentiment of "*cultural ambivalence*," conveyed by the Japanese statement 「日本的でも西洋的でもない」 is more concisely conveyed by the English statement "Equally Japanese and Western," than by the literal English translation of the original Japanese statement.

Stimuli

The stimuli were the same as described in Experiment 1.

Procedure

Participants were informed verbally that they would be presented with a sequence of seventy-two short musical excerpts, and that their task was to indicate on the corresponding instances of the Likert scale (1–72), their impression of how strongly the excerpt reflected a Western style, Japanese style, or whether the melody was characteristic of n/either style (*"culturally ambivalent"*). Four practice trials were presented to the participants to solicit any questions and provide clarification. The 72 stimuli of the experiment proper were randomly sampled and presented over the classroom's PA loudspeaker system without replacement. A brief two-minute rest intervened between the 36th and 37th trials of the experiment proper.

Results

Figure 2.3 presents the mean Western participant perceptions of musical style for each of the JJ, WJ, JW, and WW stimulus variants. The mean *melodic style* rating for each of the four melody types provided by each participant was submitted to a two-way mixed ANOVA with Tonality (Japanese, Western) and Contour (Japanese, Western) submitted as the two within-subjects factors, and Nationality (European, North American) as between-subjects factors. The omnibus ANOVA revealed no interaction between any of the main effects or interactions with Nationality, and so the entire Western participant dataset will be considered together for the remainder of the analyses. The repeated measures contrasts indicated a main effect of Tonality, $F(1,208) = 418.64$, $p < 0.001$ and a main effect of Contour, $F(1,208) = 105.62$, $p < 0.001$, indicating that participants were sensitive to both the musical scale and melodic phase manipulations employed to create the desired stimulus variants. The interaction between Tonality and Contour was significant, $F(1,208) = 5.98$, $p = 0.015$, indicating that for these Western participants the effect of Contour was more pronounced for melodies presented in Western Tonality (C major scale; WW = –1.20, WJ = –0.87, repeated measures t-test $p < .001$) compared to those presented in Japanese Tonality (Ritsu scale; JJ = 1.25, JW = 1.06, repeated measures t-test $p < 0.001$). This effect is consistent with the influences of enculturation and prior musical experience on perceptual sensitivity (Matsunaga et al., 2014; Ayari & McAdams, 2003).

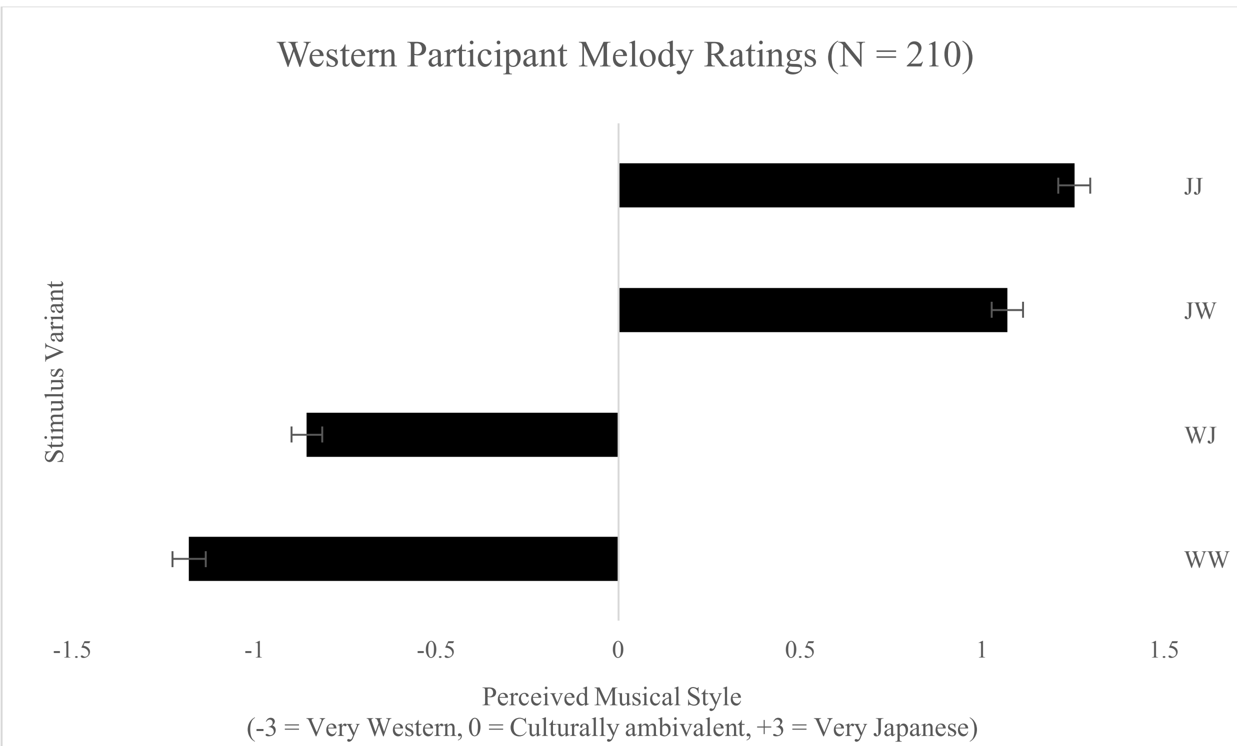

Fig. 2.3 Average *musical style* rating for each of the JJ, JW, WJ, and WW stimulus variants provided by the Western participants (N = 210). Error bars indicate standard error.

Summary and Discussion

Replicating the results produced by Japanese participants in Experiment 1, the observed main effects of both Tonality and Contour indicate that Western listeners also attend to both the musical scale and the sequence of contours present in a melodic stimulus to determine its cultural style. Critically, the main effect of Contour in particular indicates that it is not merely the pitch context of monophonic isochronous sequences that determines a melody's cultural style: the temporal organization of this pitch content throughout the duration of the sequence is also critical to the perceptual evaluation of a melody's culture of origin.

Finally, we found a statistically significant interaction between Tonality and Contour, indicating that Western listeners were more sensitive to contour manipulations made to melodies presented in C major compared to those presented in Ritsu. This result may be explained in terms of differentially developed perceptual schemas: While the typical Western listener has considerable experience with Western melodies, the same cannot be said of their experience with Japanese melodies. With more experience listening to Western melodies, Western listeners are better able to detect contour sequences which are not typical of a C major tonal context. An enculturation account might predict that with more exposure to Japanese melodies (from an early age and throughout life) Western listeners would demonstrate an equivalent sensitivity to Japanese melodic contour sequences. Nevertheless, the main effect of Contour indicates that Western listeners are capable of discerning characteristic Japanese contour profiles from deviant profiles, despite their lack of experience with traditional Japanese music. We return to this point in the general discussion. To round out our experimental investigations, in Experiment 3 we evaluate the perceptual impressions of listeners raised in neither a Japanese nor a Western cultural tradition.

Experiment 3 – Perception of Non-Japanese, Non-Western listeners

Methodology

Participants

Sixty-three intermediate undergraduate exchange students (20 males; average age = 20.4 years) from McGill University participated in the experiment as part of an in-class demonstration of music perception and experimental psychology. All participants were born and raised outside of Japan or a North American or European nation. Four (4) participants did not complete the response form and so were not included in the analyses. However, we also included the data from the three (3) complete response forms provided by the Chinese exchange students of Kanagawa University collected in Experiment 1. Therefore, the current study was conducted on a total sample size of 62 non-Japanese non-Western participants. Participants were not selected on the basis of prior musical experience, but 45 of these participants (72.5%) reported prior experience with a musical instrument, with an average of 6.53 (range 1–16) years of musical experience among them. This study was certified for ethical compliance by the McGill University Research Ethics Board II and all participants provided informed consent.

Apparatus

The apparatus was identical to that used in Experiment 2.

Stimuli

The stimuli were the same as those presented in Experiments 1 and 2.

Procedure

The procedure was identical to that performed in Experiment 2.

Results

Figure 2.4 presents the mean perceived musical style provided by the Non-Japanese Non-Western participants for each of the JJ, WJ, JW, and WW stimulus variants. The mean *melodic style* rating for each of the four melody types provided by each participant was submitted to a two-way repeated measures ANOVA with Tonality (Japanese, Western) and Contour (Japanese, Western) submitted as the two within-subjects factors. The analyses indicated a main effect of Tonality, $F(1,61) = 207.75$, $p < 0.001$, a main effect of Contour, $F(1,208) = 55.53$, $p < 0.001$, again indicating a perceptual sensitivity to the musical scale and musical contour manipulations used to generate our stimulus variants. Finally, the interaction between Tonality and Contour was significant, $F(1,61) = 22.3$, $p < 0.001$, indicating that Non-Japanese Non-Western listeners are more sensitive to the influence of Contour in a Western Tonal context (C major scale; WW = –1.18, WJ = –0.69, repeated measures t-test $p < 0.001$) compared to a Japanese Tonal context (Ritsu scale; JJ = 1.09, JW = 0.98, repeated measures t-test $p < 0.001$).

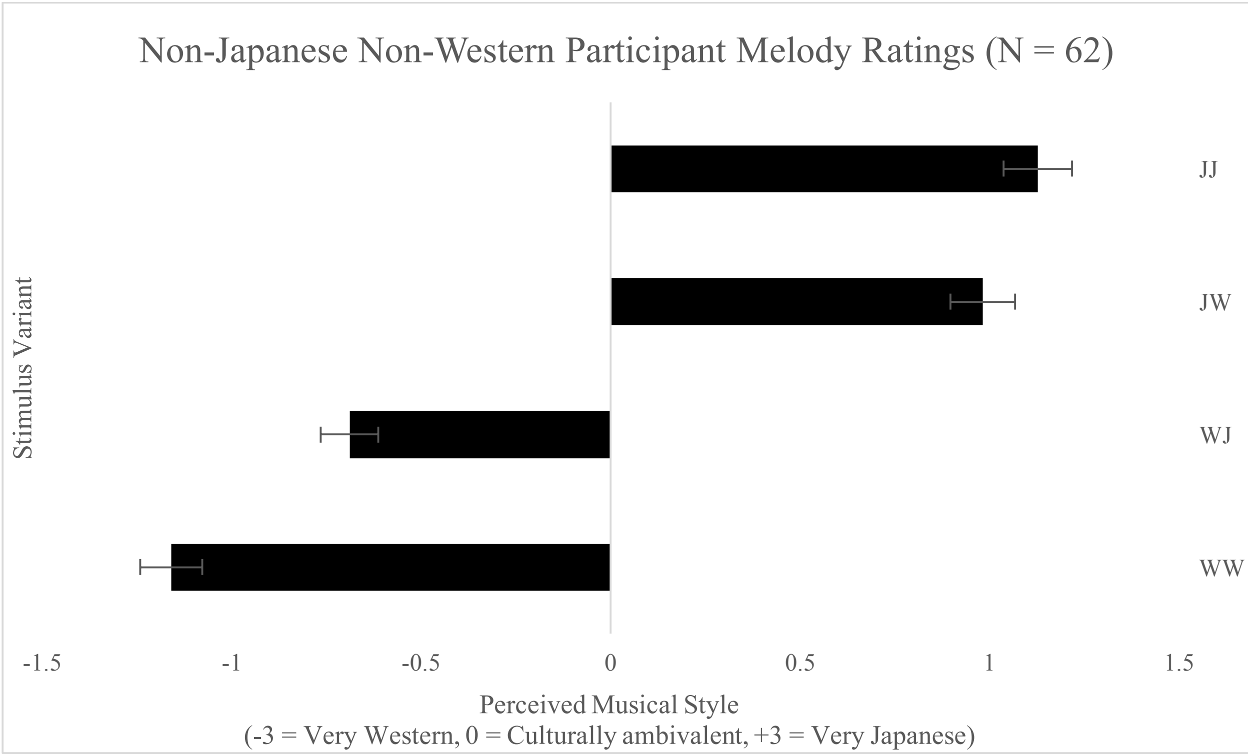

Fig. 2.4 Average *musical style* rating for each of the JJ, JW, WJ, and WW stimulus variants provided by the Non-Japanese Non-Western participants (N = 62). Error bars indicate standard error.

Summary and Discussion

A third study of the perceived musical style of melodic sequences again revealed contributions of both Tonality and Contour. Specifically, listeners in this sample also made use of both the musical scale and the sequence of relative intervals within a melodic stimulus to inform their judgments of musical style. Critically, this third study was conducted with non-Japanese non-Western listeners who were raised neither in Japan nor in a European or North American culture, indicating that while the perception of musical style is certainly influenced by musical experience, there also exists a universal (cross-cultural) perceptual sensitivity to both the tonal and temporal structures of musical melodies. This perceptual capacity allows listeners from varied cultural backgrounds to make comparable and equivalent musical style judgments.

Finally, analysis of the musical style judgments produced by non-Japanese non-Western listeners in this third sample also revealed a statistically measurable interaction between Tonality and Contour: atypical contour sequences were easier to detect when they were played in C major than when they were played in Ritsu. Given the pervasiveness of Western culture in the modern age this result is perhaps not surprising—non-Japanese non-Western persons are more likely to be familiar with traditional Western music than with traditional Japanese music on the basis of mere exposure.

General Discussion

Although previous research has typically investigated the temporal organization of music in terms of rhythm and meter, we had found in a series of pilot experiments (outlined in the Introduction) that isochronous monophonic tonal sequences could be made to sound more or less characteristic of a particular musical style by altering its *contour profile.* Specifically, we found that while prototypically Western melodic sequences possess a *monophasic* contour profile—in which the phase relationships between the local pitch maxima and minima, implied harmonic shifts, and implied metrical boundaries are maintained throughout the duration of the melodic sequence—prototypically Japanese melodic sequences possess a *polyphasic* contour profile—in which the phase relationships between the local pitch maxima and minima, implied harmonic shifts, and implied metrical boundaries are shifted throughout the duration of the sequence. The primary goal of the present research was to implement this serendipitous artistic insight into a formal scientific experiment to determine whether traditional Japanese and classical Western musical styles might be perceptually characterized in terms of their respective contour profiles.

Previous work has emphasized the contribution of Japanese and Western scales to the perception of a melody's culture of origin (e.g., Matsunaga et al., 2012, 2014). We reasoned that if melodic contour makes an independent contribution to the perception of a melody's musical style, then culture of origin judgments should reflect the manipulation of melodic contour profiles: Melodic stimuli possessing both a typical Japanese Ritsu tonality and a typical Japanese polyphasic contour profile (JJ) should be rated as most exemplary of Japanese musical style. Likewise, melodic stimuli possessing both a typical Western C major tonality and a typical Western monophasic contour profile (WW) should be rated as most strongly Western. In contrast, a melody performed in the Japanese Ritsu tonality but possessing a monophasic contour profile (JW), or a melody performed in the Western C major tonality but possessing a polyphasic contour profile (WJ), should be rated less exemplary of the Japanese and Western musical styles, respectively. If in fact contour does not contribute to the perception of musical style, then the JJ and JW stimuli should be rated as equally Japanese and the WW and WJ stimuli should be rated as equally Western.

The experimental results were all consistent with the hypothesized effect of musical contour. Across three different samples we found that Japanese ($N = 122$), Western ($N = 210$), and non-Japanese non-Western ($N = 62$) listeners all rated JJ stimuli as more Japanese than JW stimulus variants and WW stimuli as more Western than WJ stimulus variants. This main effect of Contour indicates that the sequence of relative intervals within a melodic sequence is central to the listener's evaluation of the differences in Japanese and Western musical style. The average values for the style ratings of the JJ and WW stimuli (presumably the most "Japanese" and "Western") were intermediate with respect to the extremes of the rating scale. This effect is fairly common in the case of repeated measures on Likert scales and has been associated with response bias (see Cheyne et al., 2016). It is also worth reiterating that participants were not made aware of the experimental manipulations used to create these stimulus variants, and that all of the stimuli were randomly presented. It is not implausible that a naïve listener would withhold strong judgments of musical style when presented with a listening situation in which unknown melodies differing only in key or contour are presented one after the other. Indeed, this particular aspect of the current listening situation may also help to explain the relative strengths of the Tonality and Contour effects: The former factor is particularly salient in an extended series of monophonic isochronous stimuli sharing identical rhythmic or timbral cues. In contrast, the latter is accomplished by only subtle changes in the phase relationship between the local maxima and minima of the melodic contour and the implied meter of the sequence. The variability and uncertainty characterizing the participants' task may therefore explain both the response bias not to use the entire Likert scale and to distinguish melodies more readily from one another on the basis of Tonality rather than on the basis of Contour. That is, the relative influence of Tonality and Contour on the perception of Japanese and Western musical styles may be further investigated with additional experimental paradigms that reduce the variability of the listening task and attempt to orient the listeners' attention more directly to the temporality of the musical melodies. Blocking the presentation of melodies composed in a Japanese or Western tonality would be one simple modification to the present design that could isolate the effect of contour within a given tonal context. In light of these considerations, the fact that the contour effect persists even under the present experimental conditions is a testament to the impressive sensitivity of these listeners to the subtle changes in the contour profiles (temporal structures) of musical melodies.

In the absence of previous research into the specific contour profiles of traditional Japanese and Western music we relied on consultation and artistic insights. In the present design, we operationalized the insight that Western melodies are "straightened out" compared to

Japanese melodies by modifying the phase relations among the tonal, harmonic, and metric dimensions of the latter stimuli such that the local pitch maxima and minima and implied harmonic shifts within the melodic contour corresponded with metrically strong positions. We referred to these two types of melodic sequences accordingly, describing the Japanese melodic contour profiles as *polyphasic* and the Western melodic contour profiles as *monophasic.* The choice of the term *polyphasic* was intended to reflect the characteristic sensation of the implied temporal multithreading that one experiences listening to these monophonic traditional Japanese melodies. That is, in the same way that an "implied polyphony" is associated with, for example, the solo string compositions of Bach (Davis, 2006), in which the listener experiences a sense of multiple concurrent voices interacting throughout a single temporal trajectory, the listener of traditional Japanese music experiences a sense of multiple "musical timelines," interwoven throughout the expression of a single melodic voice. As the perceptual judgments provided by three independent samples of almost 400 participants align with this initial artistic insight, our results indicate that, at least from a psychological perspective, the phase relations between melodic, metric, and harmonic-functional musical cycles are important determinates of the Japanese and Western musical styles.

Future research seeking to further specify the unique contour profiles of these cultures could explore any statistical regularities in the *polyphasic* contour profiles of traditional Japanese melodies—How often and in what melodic contexts do the metric and harmonic phrases align? How often do local pitch minima or maxima correspond with strong metrical boundaries? How often do harmonic shifts correspond with local pitch minima and maxima? Does a structural principle exist that constrains how these various melodic parameters fall in and out of phase with each other? If such questions are answered and the contour patterns of Japanese music could be systematically abstracted, a sister set of experiments could be conducted in which traditional Western folk melodies are transformed to sound more Japanese. These findings would also lead to more computationally sophisticated (stylistically informed) musical manipulations. To this end the computational models of Schmuckler (2010) may prove particularly useful.

A final point worth theoretical consideration is the influence of a melody's tonal context on the detection of atypical melodic contour profiles. Japanese listeners could distinguish typical contours from atypical contours equally well in both Japanese and Western tonal contexts. In contrast, Western and non-Japanese non-Western listeners—each with more exposure to Western music than to Japanese music—found it easier to discern atypical contour profiles when the melody was presented in a Western tonal context compared to when it was presented in a Japanese tonal context. This finding is consistent with previous research on the *bi-musicality* of Japanese listeners (Tokita, 2014; Matsunaga et al., 2012, 2014) and with the influence of cultural and experiential factors on the perception of musical style. A post-hoc omnibus 3-way mixed ANOVA with Nationality as the between-subjects factor and Contour and Tonality as within-subjects factors confirmed that the interaction Nationality, Tonality, and Contour ($F = 7.05$, $p < 0.001$) was significant. However, the specific details of the present results also raise an interesting question regarding the cross-cultural perception of novel, unfamiliar musics: Although Western and non-Japanese non-Western listeners found it *easier* to distinguish typical from atypical contour profiles in a Western compared to a Japanese tonal context (i.e., there was a greater d' in the former context than in the latter context), they were nevertheless capable of correctly distinguishing typical (original) Japanese contour profiles from atypical (Western-modified) contour profiles presented in the traditional Japanese Ritsu scale. That is, Western and non-Japanese listeners still demonstrated a perceptual sensitivity to style-specific Japanese melodic contours in the absence of any prior experience with the Japanese musical style. In the absence of cross-cultural experience with traditional Japanese music, how are these foreign listeners capable of making a correct judgment regarding the relationship between melodic contour and Japanese musical style?

We suggest these results indicate that the question is improperly formulated: Participants are not "listening to" culturally specific contour or tonal profiles. They are listening to *music* (Schaeffer, 1966/2017). This music exhibits a particular *form* that is perceptible to the listener and exhibits a coherence within the "mind's ear" (cf., McAdams, 1989). This coherence may be abstractly parameterized in terms of the empirical dimensions of contour and tonality. But to the direct experience of the listener, the forms of the music possess a particular phenomenological identity. As indicated in the introduction, perceptual sensitivity to musilinguistic (sonic) contour is developed at a very young age. Before an individual is even conscious of the distinction between "musical sounds" and "non-musical sounds" they are picking up on structural regularities in the way the pitch content of sounds rises and falls (e.g., Trehub et al., 1984). This abstract temporality begins to delineate predictable *forms* in the cadence and the trajectory of various pitch sequences. As many sounds emerge from nonmusical settings, these forms may also be conditioned by physical (material and energetic) as well as physiological (organismic) constraints on the ways in which pitch sequences can be patterned across time (Bregman, 1990; Bonin et al., 2016). Thus, particular pitch sequences within a musical context may be conditioned by the interaction of the available interval vectors afforded by the tonality of the prevailing musical context (Huron, 1994) and the ecologically conditioned schema the listener has for sonic temporality. That is, we expect pitches to be systematically organized not only in a "vertical" spectral domain, such as a scale or harmony, but also *throughout time.* This temporal dimension of sonic organization is central to the perception of musical form. It may be the case that particular cultural styles are the result of cultivating a distinct subset of these potential musical forms. And therefore, the tonal and temporal dimensions of the musical art object (considered distinct by virtue of abstraction) are in actual fact developed in tandem and in coordinated self-restriction as a musical style develops. Thus, when a listener hears a particularly coherent musical form, they associate this with a stronger impression of a particular musical style (JJ or WW), even if this style is foreign to them.

This universal sensitivity to musical form (cf., Mungan, Yazıcı & Kaya, 2017; Iversen, Patel & Ohgushi, 2008) may be the basis upon which individuals can learn and, in some cases, even master performance traditions originating from foreign cultures. It also suggests the potential of a fundamentally common psychological basis for the creation and perception of music (Huron, 2001; Fitch, 2006). Indeed, real-world experience with cross-cultural music performance and perception tends to indicate that the boundaries constraining the individual's ability to experience, explore, and enjoy music are more theoretical than they are empirical.

Author note This research was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Alexander Graham Bell Canada Graduate Scholarship (CGS-D) and a Japan Society for the Promotion of Sciences (JSPS) Doctoral fellowship to TB, and FRQSC (2014-SE-171434) and NSERC (RGPIN 2015-05280) grants and a Canada Research Chair (950-223484) to SM. Thank you Professor Jerry Cain for facilitating our experiment in his undergraduate course at McGill University.

References

Ayari, M., & McAdams, S. (2003). Aural analysis of Arabic improvised instrumental music (taqsīm). *Music Perception, 21*(2), 159-216.

Bigand, E. (1990). Abstraction of two forms of underlying structure in a tonal melody. *Psychology of Music*, *18*(1), 45-59.

Bonin, T. L., Trainor, L. J., Belyk, M., & Andrews, P. W. (2016). The source dilemma hypothesis: Perceptual uncertainty contributes to musical emotion. *Cognition*, *154*, 174-181.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Cheyne, J. A., Bonin, T., Wright, C., Carriere, J. S., Danckert, J., & Smilek, D. (2016). "You're on ten, where can you go from there?" Tufnel problems in repeated experiential judgments. *Consciousness and Cognition*, *42*, 311-324.

Cuddy, L. L., Cohen, A. J., & Mewhort, D. J. K. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(4), 869-883.

Davis, S. (2006). Implied polyphony in the solo string works of JS Bach: A case for the perceptual relevance of structural expression. *Music Perception*, *23*(5), 423-446.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*(4), 341-354.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *The Journal of the Acoustical Society of America*, *49*(2B), 524-531.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition*. San Diego, CA: Academic Press.

Eitan, Z., & Granot, R. Y. (2006). How music moves: Musical parameters and listeners images of motion. *Music Perception*, *23*(3), 221-248.

Farbood, M. M. (2012). A parametric, temporal model of musical tension. *Music Perception*, *29*(4), 387-428.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*(2), 181-195.

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, *100*(1), 173-215.

Granot, R. Y., & Eitan, Z. (2011). Musical tension and the interaction of dynamic auditory parameters. *Music Perception*, *28*(3), 219-246.

Hauser, M. D., & McDermott, J. (2003). The evolution of the music faculty: A comparative perspective. *Nature Neuroscience*, *6*(7), 663-668.

Huron, D. (1994). Interval-class content in equally tempered pitch-class sets: Common scales exhibit optimum tonal consonance. *Music Perception*, *11*(3), 289-305.

Huron, D. (2001). Is music an evolutionary adaptation?. *Annals of the New York Academy of Sciences*, *930*(1), 43-61.

Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *The Journal of the Acoustical Society of America*, *124*(4), 2263-2271.

Koelsch, S. (2009). Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. *Psychophysiology*, *46*(1), 179-190.

Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception*, *13*(3), 401-432.

Liu, C. (2013). Just noticeable difference of tone pitch contour change for English-and Chinese-native listeners. *The Journal of the Acoustical Society of America, 134*(4), 3011-3020.

Matsunaga, R., Yokosawa, K., & Abe, J. I. (2012). Magnetoencephalography evidence for different brain subregions serving two musical cultures. *Neuropsychologia*, *50*(14), 3218-3227.

Matsunaga, R., Yokosawa, K., & Abe, J. I. (2014). Functional modulations in brain activity for the first and second music: a comparison of high-and low-proficiency bimusicals. *Neuropsychologia*, *54*, 1-10.

McAdams, S. (1989). Psychological constraints on form-bearing dimensions in music. *Contemporary Music Review*, *4*(1), 181-198.

Mungan, E., Yazıcı, Z. F., & Kaya, M. U. (2017). Perceiving boundaries in unfamiliar Turkish makam music: Evidence for gestalt universals? *Music Perception*, *34*(3), 267-290.

Paraskeva, S., & McAdams, S. (1997, September). Influence of timbre, presence/absence of tonal hierarchy and musical training on the perception of musical tension and relaxation schemas. In T. Rikakis (Ed.), *Proceedings of the 1997 International Computer Music Conference* [CD-ROM]. Thessaloniki, Greece: Aristotle University.

Peng, G., Zheng, H. Y., Gong, T., Yang, R. X., Kong, J. P., & Wang, W. S. Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics,* 38(4), 616-624.

Russo, F. A., & Thompson, W. F. (2005). An interval size illusion: The influence of timbre on the perceived size of melodic intervals. *Perception & Psychophysics*, *67*(4), 559-568.

Schaeffer, P. (1966/2017). *Treatise on musical objects: An essay across disciplines.* (J. Dack & C. North, Trans.). Oakland, CA: University of California Press.

Schmuckler, M. A. (2010). Melodic contour similarity using folk melodies. *Music Perception*, *28*(2), 169-194.

Shove, P., & Repp, B. H. (1995). Musical motion and performance: Theoretical and empirical perspectives. In J. Rink (Ed.) *The practice of performance: Studies in musical interpretation* (pp. 55-83). Cambridge, UK: Cambridge University Press.

Tokita, A. (2014). Bi-musicality in modern Japanese culture. *International Journal of Bilingualism*, *18*(2), 159-174.

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development*, 821-830.

Trehub, S. E., & Hannon, E. E. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition*, *100*(1), 73-99.

Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception & Psychophysics*, *64*(2), 198-207.

Zuckerkandl, V. (1965/1973). *Sound and symbol: Music and the external world.* Trans. W. R. Trask. Princeton, NJ: Princeton University Press.

Chapter 3 - 箏の唱歌. Structural correspondence between the *shouga* and melody of *Rokudan no Shirabe* (六段の調)

Abstract

In the musical scores of the traditional Japanese *Ikuta-ryu* (生田流), one finds alongside typical symbolic representations of tonal, metric, and rhythmic musical structure, a phonetic transcription of the intended morphology of musical phrases. This transcription is known as the *shouga* (唱歌). The *shouga* is comprised of Japanese speech sounds (*morae*), concatenated into various sequences constructed as an objective expression of the subjective impression of musical form. As such the *shouga* provides researchers with an unparalleled opportunity to study the perceptual relationship between the subjective impression of musical form and the objective parameters of the musical stimulus. We present three corpus-analytic studies of the structural correspondence between the *shouga* and melody of the seminal *soukyoku* (箏曲; music for koto) work *Rokudan no Shirabe* (六段の調). Study 1 investigates several vowel coding schemes as a follow-up to the only previous corpus analysis of the *shouga* (Hughes, 2000). The basic assumption of each vowel coding scheme is that the sequence of vowels found within the mora of the *shouga* reflect coherent pitch contour changes within the referent melody. Study 2 presents the results of exploratory statistical analyses of the non-arbitrary structural relationships between the *shouga* and the melody, despite this topology being remarkably dynamic. Study 3 presents parametric analyses describing the *pitch, interval, scale degree, rhythm,* and *meter* states associated with common *shouga* idioms identified by expert practitioners. The results of all three studies provide convergent evidence of the non-arbitrary structural correspondence between the *shouga* and the musical melody of *Rokudan no Shirabe*. We discuss the implications of these results for future perceptual studies of the *shouga* and of musical form more generally.

Introduction

Background

The *Ikuta-ryū* (生田流) is a school of traditional Japanese music. It was founded by IKUTA Kengyo (生田検校, 1666–1716) in the interest of coordinating performances of the *jiuta* (地唄; folk songs) of the Osaka region, resulting in a close collaboration between koto and shamisen music. In the musical scores of the Ikuta school, one finds alongside typical symbolic representations of tonal, metric, and rhythmic musical structure, a phonetic transcription of the intended morphology of musical phrases (Figure 3.1). This transcription is known as the *shouga* (唱歌). The *shouga* is comprised of Japanese speech sounds (*morae*), concatenated into various sequences. The structure of the *shouga* is unique to each song and the instrument on which it is performed. Ando (1986, p. 198) notes:

> There is a way to sing the sounds and renditions of musical instruments—such as "ko-ro-ri-n shya-n" and "sa-ra-ri-n"—from the old days of koto as well as the "kuchishamisen" of the shamisen. The *shouga* originated as a pedagogical technique for the instruments of Gagaku [noble court music]. For example, in the first exemplar [above; *ko-ro-ri-n shya-n*], the beginning of the musical phrase is a strong sound, the second part is a weaker rolling sound, and so on, *which is very similar to the morae used in the shouga*. Certain shouga phrases accompany the music depending on how the musical sounds are connected, that is, musical context, or depending on the manual technique used to play the music. (Emphasis mine.)

Thus, *shouga* sequences are employed on the basis of their *perceived similarity* to the musical phrases they represent—the sounds of the *shouga* and the sounds of the melody present a perceptual isomorphism to the listening subject. Indeed, there is a distinct phenomenological emphasis in the expert discussions and explanations of the structure and function of the *shouga.* This mimetic device is just as concerned with representing the subjective perception of *likeness* among musical sounds as with the onomatopoetic, or strictly *aural* representation of the musical sound, or *art object* (Ando, 2017). This joint consideration of the combined subjective-objective structure of the music is evident in the interchangeable use of subjective and objective terms by authors describing the form and function of the *shouga.* Terms such as “motif,” “structure,” and “sound,” are often used in tandem with descriptions of musical “aura,” “form,” and “feeling”. Mananoi and Akira (1978, p. 10) describe further:

> Shouga are generally used as an abstract means when it is desired to show the contents of a piece of music in a way other than performing the song. If you want to reproduce the approximate shape of the music, it may be more convenient to reproduce it by voice, rather than having to play regularly, especially if it is a matter of how to create the desired song-sounds on the instrument. … In contrast to the solmization [solfège] method, the [*shouga*] phrase *chi-ra-ro-ru-ro* used in shamisen or *to-chi-chi-ri-chi-n* and *ko-ro-ri-n shya-n* of the koto display the melody by imitating the sound as well as the musical sense[82] of the sound.

[82] 感性; *kansei;* “sensitivity,” “sensibility,” “impression”

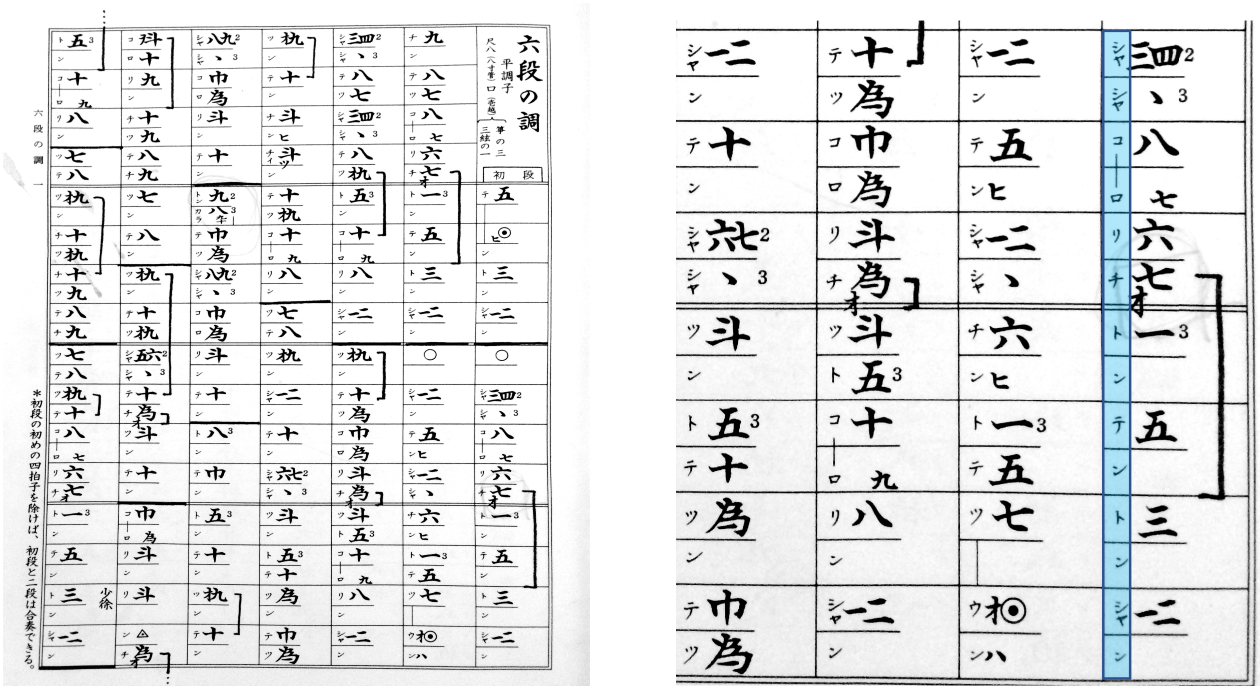

Fig. 3.1 The score and *shouga* of the *Ikuta-ryu*. The first image on the left presents the Ikuta score for 初段 (*shodan*; "First step"). The score is read from top to bottom and right to left. Large *kanji* characters indicate string number, presenting melodic pitch in a tablature. Horizontal bars indicate duple meter pulse divisions. The *shouga* is presented in the smaller *katakana* script to the left of each column, highlighted in the second image.

Consequently, the external structure and internal coherence of the *shouga* are unlike those of other music solmization techniques, such as the pitch systems of Western *solfège* or Chinese *gōngchě*: These latter systems—which were not developed to represent the musical form perceived by the listening subject—involve the assignment of speech sounds to single absolute parameter states of the musical art object (e.g., *do re* and *mi* correspond, respectively and only, with the first, second, and third degrees of the musical scale they represent). Furthermore, the abstract signs of these systems bear no perceptual relevance to the objective musical parameters they signify, in the same way the word "dog" bears no perceptual relevance to the furry, four-legged, fetch-loving animal to which it refers[83]. In contrast, the phenomenological grounding of

[83] A *symbol* in the vernacular of Peircean semiotics

the *shouga* entails the coordinated use of speech sounds to re*present* the abstract form perceived by the listener when listening to the corresponding music. Since more and less similar internal states of the listener may be elicited by different empirical parameters of the musical sound, a single mora within the *shouga* is associated with multiple absolute pitch, interval, contour, and rhythm states within the total musical art object. "But this does not mean that *tsu-n* [for example] is used randomly. Rather, different melodic events require a common function of the same mora, often understood unconsciously" (Mananoi & Akira, 1978, p. 36). Instead of a static mapping system relating the constituent mora of the *shouga* with particular musical parameter states, different mora strings, or *idioms*, are used to indicate perceptually distinct musical structures, which may be modulated or "transposed" throughout the various empirical parameters of the art object.

> In the unitary structure of Japanese music, the associations of a certain sound often serve as a kind of idiom, and there are also cases where the whole music is made up of the combination of many kinds of idioms. In other words, music is not composed of every single sound, but composed of hundreds and hundreds of units. For example, in Nohkan [transverse bamboo flute music], one does not consider individual mora such as "o" "hya" and "ra" and their association with specific pitches and rendition techniques, but rather with the resulting construct that has been created in the particular combination of these sound-associations in the idiom "o-hya-ra". *It can be said that a certain shouga idiom is related to the fact that the unit structure of the music is reminiscent of a sound quality, which is a timbral [phenomenological] consideration.*" (Mananoi & Akira, 1978, p. 14, emphasis mine).

To make matters still subtler, the *shouga* does not rely solely on acoustic mimesis to represent the subjective impression of musical form. The subjective impressions of proprioceptive and motoric events play a critical role in the *shouga's* representation of the ideal musical sound (Giolai, 2019). As Hughes (2000, p. 94) notes, the *shouga* is only fully comprehensible through its orality: “To fully experience the impact of the syllables, one must sing or recite them, preferably aloud but at least in one's head. Such systems have often come to be written down as well… but even in these cases their oral use is likely to continue in parallel.” Thus, the internal logic of the *shouga* operates through the abstraction and cross-modal relation of various perceptual qualia. Giolai (2019) argues that this “embodied” cognition characterizes “the physiologically antecedent level of affect,”—and emphasizes that such psychophysiological processes influence the conscious experience of the subject without themselves being expressed in consciousness. The embodied, cross-modal, and tacit cognition of the *shouga* is consistent with the myriad findings of implicit cross-modal associations between musical form and the subjective impressions of form derived from other sensory modalities (e.g., Schubert et al., 2019; Bonin, Degtyareva & Hahn, 2019; Eitan & Granot, 2006; Marks, 2000; Zbikowski, 1997). This coincidence may provide further insight into the understanding of expert musicologists and practitioners that the function of the *shouga* is to convey the ideal subjective impression of musical form elicited by the acoustic art object. In summary, where the syllables of solfège communicate *what* to play—in the sense of specifying the objective structure of the musical sound in question—, the idioms of the *shouga* convey *how* to play—in terms of the categorical and qualitative likeness perceived by the listening subject to exist among different musical sounds. Where solfège provides an objective, structural abstraction of the music, *shouga* provides a subjective, formal abstraction of the music.

By virtue of this representational technique, a *shouga* idiom may be transposed throughout the objective parameters of the music, coordinated with the musical phrases it represents: *Ko-ro-ri,* for example, represents the same melodic motif regardless of whether that motif starts on G3, D4, or A5, regardless of whether the precise intervals characterizing the movements from *ko* to *ro* and *ro* to *ri* are [-2, -3] semitones or [-2, -4] semitones, and regardless of whether the motif appears in its prototypical elongated-triplet rhythm or in a duple variation in which all notes of the motif are performed in the same rhythm, etc. (see Figure 3.2 for an example of two rhythmic variants of *korori*). Thus, the *Ikuta-ryu* has developed in the *shouga* a representational object whose *structural* differentiation corresponds with the *formal* delineations characterizing the perceptual image in the mind of the listening subject—a mimetic articulation of the subjective experience of musical form.[84] Furthermore, the temporal structure of this object is coordinated within the empirical domain with respect to the temporal development of the discrete subjective impressions of musical form elicited by the original musical sound. As such the *shouga* provides music psychologists with a unique and arguably unprecedented opportunity to investigate the perceptual processes underlying the relations between subjective impression of musical form and the objective parameters of musical structure.

Fig. 3.2 Dotted (left) and straight (right) rhythmic variations of the idiom *ko-ro-ri* (コロリ). The melodic line written in each staff is the koto music to be performed according to *Ikuta-ryu* (生田流演奏). The *katakana* sequence beneath it is the corresponding *Ikuta-ryu shouga* (生田流唱歌), with *romaji* transcriptions presented at the bottom for ease of interpretation.

[84] An *icon* in the vernacular of Peircean semiotics

Previous scholarly treatments of the shouga

Halliwell (1994) provides an ethnographic description of the traditional Japanese teaching philosophy *nusumeba ii* (盗めばいい; “steal it, if you can”). Although he disclaims, “A full discussion of *shôga* in *koto* music, which is fundamental to an understanding of the musician's perception, is beyond the scope of this paper,” (p. 37) he aptly notes that in the context of traditional Japanese music, “Perception is largely unmediated by explicit, formalized "theory."… [So although] terminology, and the ways in which language is used, can provide the researcher with certain essential clues about the ways in which music is conceived[, n]on-language mediation … is more important in lessons, and also can tell us a great deal about what is important musically. … [T]he oral representation of instrumental sound *(shôga)* is especially important in mediating perception.” (p. 45)

David Hughes (1989, 1991, 2000) has conducted the most extensive ethnomusicological research of the *shouga* and other so-called “acoustic-iconic mnemonic systems” throughout various Japanese, far-Eastern, and other world musical traditions. Hughes (2000), Nelson (2008a, b), and Giolai (2019) all indicate that the consonant components of the mora within the *shouga* indicate phrasing and performance techniques, while vowel components reflect pitch relations among adjacent notes and within an idiom. Hughes (2000) in particular notes a series of important insights regarding the patterned relationship between the ordinal rank of ascending F2 formants within the Japanese vowels found within the *shouga* and the musical contour of the corresponding local pitch trajectories (described in further detail in Study 1). He also presents the only empirical demonstration of this correspondence, demonstrating in a corpus of 330 cases that the contour of a *Noh* flute melody rises and falls according to the relative height of the F2 vowel formants within the *shouga.* In so doing, Hughes (2000) has presented the first empirical demonstration that the sequencing of morae within the *shouga* is non-arbitrarily associated with the sequencing of contours within the associated musical melody. No research has yet tested the generalizability of these insights regarding the relationship between the musical melody and the vowels found in the *shouga,* although the expert performance treatise of Mananoi and Akira (1978) also presents an ordinal *vowel coding scheme* identical to one of the "F2 systems" presented in Hughes (2000). Study 1 is conducted as a replication and extension of the work of Hughes (2000) and Mananoi and Akira (1978) to a new corpus of *soukyoku* music.

Critically, Giolai (2019) emphasizes that while acoustic and phonetic mimesis are fundamental to the internal logic of the *shouga*, so too are the embodied perceptions of physical force, muscular tension, and affective sensibility (see also Mananoi & Akira, 1978). The concomitant influence of these two "codes" means that one cannot fully predict the structural coherence between the *shouga* and the music it represents solely on the basic of auditory mimesis. In the absence of any operationally defined models of this interaction, Study 2 is an exploratory analysis intended to reveal general statistical indices of the structural correspondence between the *shouga* and the musical object it represents.

Finally, despite the fact that the structural correspondence between the *shouga* and musical sound is "often understood unconsciously," several critical *shouga* idioms have been identified by expert practitioners (the focus of Study 3). Ando's (1986) text "*Soukyoku of the Ikuta-ryu*," discusses the *shouga* of the *Ikuta-ryu* directly. Sensei ANDO Masateru (安藤政輝) is the youngest and last-living student of MIYAGI Michio (宮城道雄, 1894–1956), father of modern koto music. Additional treatment of *Ikuta-ryu no shouga* is available in the essays accompanying a rare set of traditional music archives (*"Kuchishouga taikei: Solmization of Japanese instruments"*) by Mananoi and Akira (1978). Both of these publications are available only in Japanese, but they provide invaluable insight into the various *soukyoku shouga* idioms, as well as artistic descriptions of the critical musical parameters or melodic figures associated with these idioms. In the summer of 2017, the first author had the opportunity to study *soukyoku* under the direct instruction of Ando-sensei. The following research derives first and foremost from the knowledge he managed to steal during his time as Ando-sensei's student.

The present investigation

The purpose of this work is to investigate structural relationships between the *shouga* and the musical melody of the seminal koto work *Rokudan no Shirabe*. The *koto* (箏) is a 13-string zither and the national instrument of Japan. It was first introduced to the archipelago in the early Nara period (8th century). The silk strings of the koto are activated by three ivory plectra, located on the thumb, index, and middle digits. *Rokudan no Shirabe* (六段の調; "Song of Six Steps") is the earliest extant work written for solo koto, or *soukyoku* (箏曲). It was composed in the 18th century by YATSUHASHI Kengyo (八橋検校, 1614–1685; Ando, 1986).

Rokudan no Shirabe is written in a duple meter and performed in *hirajoshi* (平調子), or the "everyday, standard" tuning of the koto. This tuning is pentatonic (D-Eb-G-A-Bb) and derived from a combination of the *Gagaku* court modes of *ryo* and *ritsu*, as well as the indigenous tonal systems later categorized as the *yo* and *in* scales (Figure 3.3; Malm, 2015). In ascending order from the first to the thirteenth, the strings of the koto in *hirajoshi* are tuned to D4, G3, A3, Bb3, D4, Eb4, G4, A4, Bb45, D5, Eb5, G5, A5 (Ando, 1989). There are two psychologically relevant notes with respect to this tuning sequence. First, the lowest string (physically most distal from the performer) is tuned to the tonic, with the first and fifth string tuned in unison. Furthermore, a common idiom *shyan* is often used to begin and end musical phrases and corresponds with the subdominant dyad produced by activating the two bottom strings, D4 and G3 together. This "inverted dominant" relation gives the impression of a "phantom" or missing fundamental D3. Therefore, when a performer plays the lowest string they are also playing the most stable scale degree of *hirajoshi,* and when activated in conjunction with the second lowest string, G3, as in the idiom *shyan,* activating the lowest string furthest from the body also creates the deepest perceived tone.

Fig. 3.3 *Hirajoshi* of *Rokudan no Shirabe*

Second, from the second string upward, the pitches of the pentatonic scale are repeated at octave intervals every five strings. That is, the second and seventh string are tuned to G ($\hat{4}$; G3 and G4 respectively), the third and eighth are tuned to A3 ($\hat{5}$) and A4 ($\hat{5}$), etc. The physical distribution of this tuning scheme is particularly relevant to interpreting the embodied contributions to the performer's understanding of some of the *shouga* idioms: For example, *to-te* (トテ) indicates a melodic motif that jumps between octaves at the same chroma and is performed by activating the lower *to* pitch with the plectrum of the middle finger and the higher *te* pitch with that of the thumb. As a result, longer sequences (such as *to-te-to-to-te-to-te*) induce a manual, proprioceptive impression of a varied rhythmic oscillation emerging from the coordinated action of the two outstretched digits, as well as a strangely "translucent" auditory image resulting from the relegation of the contribution of *pitch height* to the perception of tonal space, emphasizing a more abstract *chromatic relation* across registers (cf., Shepard, 1982). Finally, the tonality of *hirajoshi* is described as a pentatonic variation of its *gagaku* (*ryo, ritsu*) and indigenous (*yo, in*) predecessors. In hirajoshi, as in the *ryo* and *ritsu* scales of *gagaku,* the three most important pitches ("pillar tones") are defined as the "tonic," the "dominant," and "sub-dominant" in descending degrees of tonal stability (Adriaansz, 1973; Rose & Kapuscinski, 2013; Malm, 2015). In the *hirajoshi* of *Rokudan no Shirabe*, these tones are D ($\hat{1}$), G ($\hat{4}$), and A ($\hat{5}$).

No empirical research has yet studied the *shouga* of *Rokudan no Shirabe,* or any *soukyoku* repertoire for that matter, from the *Ikuta-ryu* or otherwise. Neither has there yet been a deliberately psychological interpretation of the *shouga* research that does exist. The present research will progress through three studies of the structural correspondence between the *shouga* and music of *Rokudan no Shirabe* and conclude with a discussion of the implications for psychological research into the perception of musical form. In Study 1 we investigate the extent to which the *vowel coding schemes* described by Hughes (2000) and Mananoi and Akira (1978) predict the structural correspondence between the *shouga* and melody of *Rokudan no Shirabe.* In Study 2 we present a series of statistical analyses that explore supplementary indices of non-arbitrary structural relations between the *shouga* and the melody of *Rokudan no Shirabe.* Finally, in Study 3, we consolidate the definitions of the various *shouga* idioms presented in Ando (1986) and Mananoi and Akira (1978) to determine a set of robust idioms within the *soukyoku shouga.* We then investigate the distribution of these idioms across various empirical parameters of *Rokudan no Shirabe*.

Study 1 – Replication and Extension of Hughes (2000)

Background

As noted in the Introduction, previous research has indicated that the structures of the *shouga* and the musical melody represented by the *shouga* are coordinated such that ordinal motion between adjacent vowel components of the *shouga* correspond with ordinal motion between adjacent pitches in the musical melody. The only empirical demonstration of this fact to date was produced by Hughes (2000), who argued that the efficacy of this mapping is due to the acoustic-iconic mimesis between the rising and falling contours of the pitch content of the musical melody on the one hand and the rising and falling of the second vowel formants of the *shouga* on the other.

Formants are the resonant frequencies of the vocal tract, which produce patterns of relatively high and low energy throughout the frequency spectrum of a vocalization. These relative energy patterns are used for the perceptual categorization of vocal sounds, such as the vowels of formal speech, which allows for the intelligibility of speech signals across various absolute pitch (fundamental frequency) contexts influenced by sex, age, emphasis, emotionality, etc. The second formant (F2) is audibly salient during, for example, whispering or whistling (Hughes, 2000). The Japanese vowels /a/ (ア), /i/ (イ), /u/ (ウ), /e/ (エ), and /o/ (オ) can be assigned ordinal ranks in terms of the relative heights of their second formants. The resulting sequence, *IEAUO* was forwarded by Hughes (2000) as a prime example of a cross-culturally pervasive phenomenon among musical mnemonics to which he refers as "F2 systems." In the present work, we will test the applicability of this F2 system in the context of *Rokudan no Shirabe* alongside two other candidate *vowel coding schemes*.

The vowel coding schemes

Central to each vowel coding scheme is the assumption that the sequence of vowels found within the mora of the *shouga* reflect coherent pitch contour changes within the referent melody. This is operationalized by assigning ordinal ranks to each of the vowels and associating the sequence of these ordinal ranks with the sequence of "ascents" and "descents" (contours) within the musical melody.

IEAOU: Mananoi and Akira (1978) and Hughes (2000) describe the vowel coding scheme *IEAOU*—with /i/ corresponding to the highest pitch within a musical contour and /u/ corresponding to the lowest. For example, according to this scheme, a *shouga* sequence *chi-te-tsu* should correspond with a continually descending melodic contour; the *shouga* sequences *chi-tsu-te* and *te-to-chi* should correspond with V-shaped melodic contours, which first descend and then ascend; and *ton-kara-te* starts low and ascends.

IEAUO (F2): Hughes (2000) described the *IEAUO* vowel coding scheme in his treatment of "F2 systems." In this scheme, ranks are produced according to descending values of the second formant for each vowel. As the names imply, the IEAUO scheme differs from the IEAOU scheme only in terms of the relative ranks of /u/ and /o/, with /o/ taking the lowest relative pitch position in the former and /u/ taking that spot in the latter. Hughes (2000, p. 100), interpreting the results of his analysis of the *shouga* for Noh flute (能管), notes: "The vowels [o] and [u] are not clearly ranked in relation to each other: they seem to share the bottom rung." For this reason, we include both schemes.

AIUEO: *AIUEO* is the vowel sequence found in the mora columns of the Japanese *hiragana* and *katana* syllabaries. The sequence is analogous to that of the English recitation of vowels "A, E, I, O, U," with the further specification that *all* Japanese speech sounds are organized into these vowel columns: When Japanese speakers learn the *kana* syllabaries, they first learn "a, i, u, e, o," then "ka, ki, ku, ke, ko," "sa, shi, su, se, so," "ta, chi, tsu, te, to," etc. Therefore, from the psychological perspective, the fluent Japanese speaker possesses a robust auditory schema (influencing both perception and conceptualization) for the sequence /a/-/i/-/u/-/e/-/o/. According to this scheme, /a/ is associated with a "beginning," "departing," or "initiation" of an audible sequence or *phrase,* whereas /o/ is associated with a "cadence," "closing," or "anchoring." We include this scheme to consider a culturally determined schematic influence on the structure of the *shouga,* as well as to heed the indications of Mananoi and Akira (1978), Hughes (2000), and Giolai (2019) that the *shouga* can only be fully understood in terms of its *orality* and that it is primarily understood (learned, rehearsed, and performed) *unconsciously.* If the shouga is determined in part by an embodied combination of perceptual factors—ranging from the auditory to the proprioceptive to the conceptual—then one of the primary oral schemes for Japanese speech (AIUEO) is arguably worth investigating.

Corpus, Analysis Procedures, and Results

The corpus

A database was constructed to uniquely represent every distinct musical event (sounded or unsounded) of *Rokudan no Shirabe*. The database included the following dimensions along which each event was coded: Event Number, Movement Number (段; *dan*), Mora, Consonant [component of the mora], Vowel [component of the mora], Pitch [of note] 1 [of a dyad], Chroma 1, Scale Degree 1, MIDI 1, String [on which Pitch] 1 [was played], Pitch 2, Chroma 2, Scale Degree 2, MIDI 2, String 2, Elapsed Time [since previous event], Absolute Metric Position [within the movement], Relative Metric Position [within the bar].

There were 1289 events in total, twelve of which were Start (6) and Stop (6) indicators for each of the six movements of the piece, leaving 1277 musical events for analysis. Four hundred and forty-four (444) of these cases were either the mora ン (*'n',* indicating a pitch resonance, 389), the *shouga* sequence ヨーイ (*'yo-i',* signifying a pulsed rest; 10), or another breath, phrasing, or sustain mark possessing no pitch (45). Seven hundred and thirty-four (734) of these 1277 cases comprised the *basic pitch events* (Mananoi & Akira, 1978) of the melody of *Rokudan no Shirabe*. Each of these was represented within the *shouga* by a mora including both consonant and vowel components.

The remaining 99 events were pitched events signified by a mora possessing only a vowel component (ア， イ， ウ， エ， オ). Each of these vowel-only events were ornamental pitch bends, continuing an expression within the melodic line whose base pitch had earlier been indicated by a mora possessing the same vowel as well as a consonant component (e.g., *chi-i,* where *chi* is the basic pitch and *i* is the ornamental pitch bend elaboration). If a pitch bend was used to create one of the basic pitch events in the melody (i.e., if it was not an ornament), this event was indicated as a basic pitch event by a mora possessing both a consonant and vowel component. Therefore, all single vowels indicated a pitch bend, but not all pitch bends were indicated by a single vowel.

Although ornamental events are essential to the musical style of the *soukyoku* and *shouga* of the *Ikuta-ryu*, the scope of the current research is primarily concerned with basic relationships that might be revealed between the structure of the *shouga* and that of the melodic surface of the music it represents. Consistent with this scope, Mananoi and Akira (1978) note that the *shouga* is more concerned with representing the feeling relations among the basic idioms and motifs rather than creating an abstract codification of each musical event considered in isolation.[85] Given that 1) ornamental pitch bends are extensions of the base melodic structure, 2) the shouga is concerned primarily with the basic structural components of the music rather than every note considered in isolation, and 3) including the ornaments would interfere with a direct replication and extension of the previous work conducted by Hughes (2000), we decided not to include these pitch bends in our current analysis. Nevertheless, the events are coded independently in the database and can be returned to for further investigation in future work.

Likewise, in constraining our analyses in the following studies to consider only the pitch, interval, and contour relations among musical events, we acknowledge that our analyses do not address the entire form and internal logic of the *shouga*, which correspond with many more musical parameters than these. Timbral analyses of the different mora sounds and onomatopoetic associations (e.g., *shyu* indicates a swift scraping of the ivory plectrum of the middle finger along the two lowest strings of the koto), mechanical and geometric analyses of the string activation patterns (e.g., *ko-ro-ri* is always produced by playing three adjacent strings in a row with a specific rhythm and dynamic; Ando, 1986), as well as the position of silences and resonances within the musical surface, are all interesting facets of Japanese *soukyoku* and the *koto shouga* developed to represent it. It is hoped that future research questions concerning these topics will be facilitated by the variable scheme provided in the current database. For the present analysis, however, we evaluate the 734 basic pitch events of *Rokudan no Shirabe*.

---

[85] "In general, the shouga does not address such small variations and ornaments. This is because the shouga fulfills its purpose by casting the essence of a melody." (Mananoi & Akira, 1978, p. 15)

Analyses

We conduct two statistical analyses of the relative predictability of three candidate *vowel coding schemes*: I) The proportion of vowel-contour pairs within *Rokudan no Shirabe* predicted by each of the three vowel coding schemes, II) The maximum sequence length for which the predictions of each vowel coding scheme apply.

I. Computing the proportion of vowel-contour pairs within Rokudan no Shirabe predicted by each of the three vowel coding schemes

I.1. Procedure

Following Hughes (2000), we first report the proportion of vowel-contour correspondence explained by each of the vowel coding schemes. This was accomplished for each scheme by comparing the observed melodic contour (rising [+], falling [-], or flat [=]; N = 733) relating two adjacent tones within the melody with the hypothetical melodic contour predicted by the corresponding mora events within the *shouga*, according to ordinal vowel ranks proposed by that scheme. When the observed contour matched the predicted contour of a given vowel coding scheme, it was scored as a "hit". Otherwise it was scored as a "miss." For example, the tones of the melodic sequence G4–G5 are related by the contour [+]. If the *shouga* sequence corresponding with this melodic sequence was *tsu-ko,* this would be scored as a "hit" for the IEAOU scheme—which predicts that pitch events represented by /o/ kana should be higher than pitch events represented by /u/ kana—but a "miss" for the IEAUO and AIUEO schemes, which predict the opposite relative pitch-height relation. Proportion correct was computed for each vowel scheme by dividing the number of hits it produced by the total number of melodic contours (N).

I.2. Results

The frequencies for each contour-vowel mapping and the hypothetical contours predicted by each vowel coding scheme are presented in Table 3.1.We find that any particular vowel coding scheme explains at most 56.6% of the mappings between the contours of the musical melody and the vowel sequences of the *shouga* in *Rokudan no Shirabe* (Table 3.2). Although the numbers do not exceed 90% (cf., Hughes, 2000), the cross-culturally observable and acoustically grounded vowel coding schemes IEAUO and IEAOU explain an impressive proportion of the shouga-melody mappings found in the current corpus specifically dedicated to the shouga of *Rokudan no Shirabe*. The culturally specific AIUEO vowel coding scheme explains a fair proportion of the shouga-melody couplings in its own right (43.5%), but when included alongside the strongest vowel coding scheme (IEAOU; 56.6%) only 62 unique instances (8.5% of the total corpus) are additionally explained (IEAOU + AIUEO; 65.1%). That is, when the contours of the musical melody are evaluated for correspondence with *either* the IEAOU or the AIUEO schemes, only 62 additional cases are found to correspond with the AIUEO scheme that are not already accounted for by the IEAOU scheme.

A few systemic exceptions to all vowel coding schemes are worthy of note. In particular, every vowel scheme predicted the opposite trend of a robust effect found for the a-o (N = 7) and a-u (N = 13) vowel sequences: 20/20 of the corresponding melodic contours rose [+], although each scheme predicted they should fall [-]. Similarly, of the 49 observed contours corresponding with the e-u vowel sequence, 26 rose [+] and 23 fell [-], therefore rendering the predictions of the vowel coding schemes obsolete. Finally, the majority of i-i (78%), o-o (96.5%) and u-u (58%) vowel sequences (N = 115) corresponded with rising [+] or falling [-] contours, in contrast with the predictions of each vowel scheme that the contour should be flat [=].

Table 3.1 Frequencies for each contour mapped onto a particular vowel duple

Vowel-duple	Observed Contour			Hypothetical Contour		
	+	=	-	IEAUO	IEAOU	AIUEO
a-a	5	37	2	=	=	=
a-e	30	0	0	+	+	-
a-i	12	0	2	+	+	-
a-o	7	0	0	-	-	-
a-u	13	0	0	-	-	-
e-a	1	0	23	-	-	+
e-e	2	5	0	=	=	=
e-i	20	0	4	+	+	+
e-o	12	0	38	-	-	-
e-u	26	0	23	-	-	+
i-a	0	0	16	-	-	+
i-e	0	1	28	-	-	-
i-i	14	4	0	=	=	=
i-o	1	1	24	-	-	-
i-u	11	0	39	-	-	-
o-a	1	1	11	+	+	+
o-e	40	0	0	+	+	+
o-i	1	0	69	+	+	+
o-o	13	3	69	=	=	=
u-a	0	0	12	+	+	+
u-e	31	0	16	+	+	-
u-i	12	0	1	+	+	+
u-o	23	0	17	-	+	-
u-u	5	5	2	=	=	=

Table 3.2 Proportion of vowel-contour pairs explained by each of the vowel coding schemes (N = 733 duples)

Vowel coding scheme	Proportion of shouga-contour duples explained (Hits)
IEAUO	0.558 (409)
IEAOU	0.566 (415)
AIUEO	0.435 (246)
IEAOU + AIUEO	0.651 (477)

II. Determining the maximum sequence length for which the predictions of each vowel coding scheme apply

The results of analysis I indicate there is more to the internal logic of the *shouga's* representation of musical form than can be effectively captured with a vowel coding scheme of static (state-invariant) ordinal pitch ranks.

II.1. Procedure

Extending the previous work of Hughes (2000), we identify the longest *shouga* sequences for which the respective predictions of the three vowel coding schemes apply. These sequences were determined by querying the database for incrementally longer hit sequences (i.e., by iteratively increasing the size of the sampling window) until there were no hit sequences found for a given tuple size. Specifically, a sliding window was used to identify all of the size-*t* vowel tuples within the *shouga.* For each vowel tuple, a *hypothetical contour tuple* was created for each of the vowel coding schemes. Then, an additional sliding window was used to identify the actual (*observed*) size-(*t*-1) contour tuple within the melody corresponding to each vowel tuple within the *shouga*. If the hypothetical contour tuple predicted by a given vowel coding scheme matched the observed contour sequence, it was deemed *coherent* and scored as a hit.

For example, the size-6 *shouga* sequence *to-ka-ra-te-to-te* represented by the vowel tuple [o, a, a, e, o, e] is predicted by both the IEAUO and IEAOU coding schemes to correspond with the size-5 contour sequence [+, =, +, -, +], but predicted by the AIUEO coding scheme to correspond with the sequence [+, =, -, -, +]. If the observed sequence was [+, =, +, -, +], a "hit" would be scored for the IEAUO scheme and the IEAOU scheme, and a "miss" would be scored for the AIUEO scheme. If the observed sequence was [+, =, -, -, +], the IEAUO and IEAOU schemes would score a "miss" and the AIUEO scheme would score a "hit". If the observed sequence was neither, e.g., [+, +, -, =, +], all schemes would score a "miss".

If, for a given vowel coding scheme, at least one vowel tuple of a given size *t* was found to produce a hit, then the size of the sampling window was increased by 1 (*t*+=1). Instead, if a given vowel coding scheme produced no hits at sample size *t*, then no coherent mappings between the *shouga* and the melody were found, and the maximum length for which a coherent mapping did exist according to the predictions of that vowel coding scheme was reported as length *t*-1.

II.2. Results

Another informative statistic describing the relative predictive capacity of the vowel coding schemes concerns the largest tuple size for which the mora patterns within the *shouga* and the pitch patterns within the melody remain coherent according to each of the schemes. The results are summarized in Table 3.3. The IEAUO and IEAOU schemes produced the same largest-string *shouga* sequence with a length of 13 mora.[86] Interestingly, although the AIUEO vowel coding scheme explained the least variability among the *shouga*-melody relations, it produced the longest coherent *shouga* sequence of any coding scheme (N = 16).[87] Inspection reveals that this advantage is due to the fact that the AIUEO scheme ranks */e/* as corresponding to lower relative-pitch values than */u/,* whereas the two former schemes do not. Thus, extending the previous work of Hughes (2000), the present analysis indicates that the vowel structure of the *shouga* varies coherently with the contour of extended melodic sequences.

[86] It is worth acknowledging that the *shouga* may rightly be interpreted as a typical solfège given the specific pitch-mora mappings in the present case considered in isolation, but further examples will discount this possibility.
[87] An exemplary contradiction to the solfège interpretation

Table 3.3 Maximum tuple size for which the *shouga*-melody relation remains coherent according to each of the vowel coding schemes. Pitch heights are reported as MIDI numbers

IEAUO	
Max Tuple Size (N = 1)	13
Shouga Sequence	*ri-te-shya-shya-te-shya-shya-te-shya-shya-te-shya-shya*
Pitch Tuple	[51, 50, 38, 38, 50, 38, 38, 50, 38, 38, 50, 38, 38]
Contour Tuple	[-, -, =, +, -, =, +, -, =, +, -, =]
IEAOU	
Max Tuple Size (N = 1)	13
Shouga Sequence	*ri-te-shya-shya-te-shya-shya-te-shya-shya-te-shya-shya*
Pitch Tuple	[51, 50, 38, 38, 50, 38, 38, 50, 38, 38, 50, 38, 38]
Contour Tuple	[-, -, =, +, -, =, +, -, =, +, -, =]
AIUEO	
Max Tuple Size (N = 1)	16
Shouga Sequence	*te-to-to-te-to-te-tsu-to-te-to-to-te-to-te-chi-te*
Pitch Tuple	[62, 50, 50, 62, 50, 62, 65, 55, 67, 55, 55, 67, 55, 67, 70, 69]
Contour Tuple	[-, =, +, -, +, +, -, +, -, =, +, -, +, +, -]

Discussion

The results of Study 1 indicate that the vowel sequences of the *shouga* alone predict more than half of the variability in the contour sequences of the melody in *Rokudan no Shirabe*. Given that the only other empirical evidence for these schemes is based on the monophonic flute and percussion musics of Japan (Hughes, 1991, 2000), it is particularly remarkable that the vowel coding schemes described in Hughes (2000) and Mananoi and Akira (1978) apply so robustly in this corpus of koto music. Indeed, Mananoi and Akira (1978, p. 36) note that “In the case of the koto, the precise use of the syllabary sounds is not as clear…” This may be due to the nearly 3-octave range of the instrument, the ability to perform rapidly changing and precise pitch sequences by virtue of the 13 independently tuned strings coupled with the speed and dexterity of the fingers, and the relative structural complexity characterizing the interwoven melodic phrases and gestures of *soukyoku*. Still, the current study finds a strong association between the ordinal sequence of vowels within the IEAUO and IEAOU vowel coding schemes and the contour sequences of the musical melody.

While the present data are impressive, they are considerably less decisive than the frequencies previously presented by Hughes (2000; >90%). To this end we recognize several other musical parameters and *shouga* properties that may contribute to the intelligibility of the *shouga* transliterations. First of all, although contour is essential, it is not the only melodic parameter with which the *shouga* is concerned. Registral, harmonic-functional, intervallic, rhythmic, and metric considerations are some of the additional objective musical factors contributing to a common subjective impression of musical form among various motifs.

Second, the vowel coding schemes do not take into account the consonant components of the *shouga.* In light of more recent research indicating the essential influence of tacit and emobodied factors on the intelligibility of the *shouga* (e.g., Giolai, 2019), it may well be that the consonants not only contribute to timbral acoustic-iconic (onomatopoetic) representations of musical sound (as proposed by Hughes) but also to metaphorical and conceptual analogies between the sound and feel (together encapsulated in the *orality*) of the *shouga* and, for example, the perceived musical tension of the corresponding melody. Although the current research is not directly concerned with the proprioception of the *shouga,* should such embodied factors contribute to their structural coherence,[88] they may reasonably be expected to contribute to the statistical regularities observed in the empirical structures of the *shouga* and melody with which these analyses are concerned. In Study 2 we explore further statistical indices of non-arbitrary structural correspondence between the *shouga* and the musical melody it represents.

[88] The original author of the *shouga* was, after all, a musician with a body.

Study 2 – Exploratory analyses of the structural relationship between the *shouga* and melody of *Rokudan no Shirabe*

Background

In this Study, we investigate the database for complementary statistical indices of non-arbitrary structure characterizing the *shouga* and its relation to the musical melody. These analyses do not rely on the vowel coding schemes discussed above, though theoretical and empirical overlap may be observed. The motivation to conduct these analyses derives from the acknowledgment that the mental processes underlying the efficacy of the s*houga* operate largely unconsciously and through the performer's direct engagement with the musical and phonetic materials (Giolai, 2019; Hughes, 2000; Ando, 1986; Akira & Mananoi, 1978). This direct experience is thus mediated not only by an explicit knowledge, but also by a tacit knowledge (see Polanyi, 1966) of the embodied and dynamic isomorphism between different sound structures. It is this tacit dimension that escapes verbal explication. There may, therefore, be greater statistical regularity among the measurable structural relationships between *shouga* and melody than performers or researchers are capable of identifying or articulating through explicitly descriptions of the empirical structure of the *shouga* alone.

Simply stated, conscious theorizing may not reveal all of the psychologically relevant structural correspondences between the *shouga* and the music it represents, because much of the perceptual relationship between these two objects is available only as a tacit knowledge that cannot easily be articulated with formal signs (Polanyi, 1966; Gill, 2015). Statistical analyses can therefore be leveraged to reveal hidden large-scale trends in the structural correspondence between the *shouga* and melody of musical works such as *Rokudan no Shirabe*. Salient relationships revealed by these analyses may then inform subsequent theoretical and empirical analyses of the *shouga*. In the ideal case, these exploratory analyses will provide useful guides for further research dedicated to understanding the *shouga's* capacity to represent musical form.

Corpus, Analysis Procedures, and Results

The corpus

The corpus was the same as that used in Study 1.

Analyses

The analyses are grouped into two primary categories: I) Statistics describing the non-arbitrary use of the mora syllabary within the *shouga,* and II) Statistics describing the structural relationship between the *shouga* and the musical melody.

I. Statistics describing the non-arbitrary use of the katakana syllabary within the shouga

In analysis I.1 and I.2 we investigate the statistical distribution of Japanese speech sounds within the *shouga* of *Rokudan no Shirabe.* In analysis I.1, we investigate prevalence of unique speech sounds observed within the *shouga* compared the total number of unique speech sounds found within the Japanese language. In analysis I.2, we investigate the sequencing of these speech sounds throughout the *shouga,* comparing the prevalence of repeated *shouga* sequences with the rates predicted by chance.

I.1. Non-arbitrariness characterizing the mora syllabary of the shouga

I.1.1. Hypothesis

According to the null hypothesis (and the claims previously reported and countered by Hughes, 2000) that the *shouga* is a nonsensical concatenation of Japanese speech sounds, we should find that the number of unique mora and the relative frequencies of these mora within the *shouga* are equivalent to the frequencies predicted by a random sample of Japanese speech sounds.

I.1.2. Procedure

First, we compare the frequency of unique mora speech sounds observed within in the *shouga* with the number of possible morae within the Japanese syllabary. Second, we compare the respective frequencies of these morae with the frequencies predicted by chance using a binomial test.

I.1.3. Results - Non-arbitrariness characterizing the mora syllabary of the shouga

There are 150 possible morae in the *katakana* syllabary, but only 23 (15.3%) unique morae are used within the entire length of *Rokudan no Shirabe* (see Background of Study 1 above). If we restrict our consideration to only those morae associated with the *basic melodic structure* (Mananoi & Akira, 1978) with which the current study is concerned, this number drops to only 16 morae (10.7%) employed to produce a coherent representation of the basic melodic form (Table 3.4). As observed in the table, the highest frequency of a single mora within the corpus is 152 (*te*). The random chance probability of sampling 1 of 16 possible morae 152 times in 734 trials is $p < 1.29E\text{-}25$.

Table 3.4 Frequency of mora (1-gram tuples) within the corpus (N = 734 events)

Mora	Frequency
te（テ）	152
tsu（ツ）	118
shya（シャ）	99
ri（リ）	72
to（ト）	70
ko（コ）	69
ro（ロ）	69
chi（チ）	64
ra（ラ）	5
ka（カ）	3
rya（リャ）	3
shyu（シュ）	3
re（レ）	2
sa（サ）	2
yu（ユ）	2
ru（ル）	1

I.2. Non-random frequencies characterizing the prevalence of unique size-t mora tuples

I.2.1. Hypothesis

If, according to the null hypothesis, the *shouga* is a nonsensical concatenation of Japanese speech sounds, we should find that the frequency of mora tuples should be equivalent with those predicted by a random sampling of the possible combinations of the mora found within the *shouga.*

I.2.2. Procedure

We identify the frequency of repeated mora tuples of size *t* within the corpus. We then identify the longest *shouga* sequence repeated within the corpus, and calculate the probability that a sequence of this length would be repeated by chance according to Mittledorf's (2003) solution for the probability of *r* repeated sequences in an *n-size* sample from a universe of N possible states (see Appendix 3A).

I.2.3. Results - Non-random frequencies characterizing the prevalence of unique size-t mora tuples

We find non-random frequencies when comparing the actual size-*t* mora tuples within the *shouga* compared to the possible number of size-*t* mora tuples available by chance (Table 3.5): There are 256 possible unique mora-duples (16^2), but only 56 (41.2%) of these are found in the corpus—even though the corpus contains a total of 733 duple sequences. As tuple sizes increase beyond t = 2, the ratio between the number of possible unique tuple frequencies and the length of the corpus becomes exponentially large, indicating that there are far more unique tuple sequences of length-t than can be observed in the corpus. By random chance, therefore, the corpus should exhibit very few duplicate tuples (i.e., the actual tuple sequences within the corpus should all be unique). In contrast to this null hypothesis, we observe repeated tuple sequences up to length t = 16 (see Table 3.5).

Table 3.5 Observed frequencies of unique size-t mora sequences compared with the possible number of unique size-t mora sequences in *Rokudan no Shirabe*

Tuple size (t)	Number of t-mora tuples in *Rokudan no Shirabe* (Length = 734 mora)	Possible unique t-mora tuples (16^t)	Actual number (% corpus) of unique t-mora tuples in *Rokudan no Shirabe*
2	733	256	56 (41.2)
3	732	4 096	145 (20.1)
4	731	65 536	270 (36.9)
5	730	1 048 576	402 (55.1)
6	729	16 777 216	530 (72.7)
7	728	2.68E+8	611 (83.9)
8	727	4.24E+9	659 (90.7)
9	726	6.87E+10	689 (94.9)
10	725	1.10E+12	705 (97.2)
11	724	1.76E+13	711 (98.2)
12	723	2.81E+14	716 (99.0)
13	722	4.50E+15	718 (99.4)
14	721	7.21E+16	718 (99.6)
15	720	1.15E+18	718 (99.7)
16	719	1.84E+19	718 (99.9)
17	718	2.95E+20	718 (100.0)

We attempted to calculate the probability of such a length-t sequence re-occurring by chance according to Mittledorf's (2003) solution (see Appendix 3A), but could not find access to a powerful enough calculator: However, given that 1.84E19! * 719! in the numerator more or less cancels out the (1.84E19-719+1)! * 717! in the denominator, we are left with a probability of approximately 719*718 divided by (2*1.84E18^719), which tends to an infinitesimally small number. The repeated 16-mora sequence is presented in Table 3.6. Even more remarkable is the fact that this repeated mora tuple is found to coincide with *the same pitch tuple* in both instances, despite the fact that its contours violate all of the extant vowel coding schemes. Furthermore, this pitch sequence is itself found only twice in the corpus and corresponds only with the particular mora tuple in question. These results provide the first indication of a non-arbitrary structural relationship between the *shouga* and melody of *Rokudan no Shirabe.*

Table 3.6 The longest repeated *mora* tuple in Rokudan no Shirabe

	No Coding Scheme
Size of Largest Repeated Mora Tuple (N = 2)	16
Corresponding *Shouga* Sequence	*ko-ro-ri-chi-to-te-to-shya-shya-te-shya-shya-chi-to-te-tsu*
Corresponding Pitch Tuple(s)	[57, 55, 51, 55, 38, 50, 45, 38, 38, 50, 38, 38, 51, 38, 50, 55] (N = 2)
Corresponding Contour Tuples(s)	[-, -, +, -, +, -, -, =, +, -, =, +, -, +, +] (N = 2)

II. Statistics describing the structural relationship between the shouga and the musical melody

In the following analyses, we explore further statistical descriptions of the structural relationship between the *shouga* and the musical melody. These statistics are intended to guide further developments in the theoretical understanding of the representational capacity of the *shouga* specifically, and the perception of musical form more generally. We know from Mananoi and Akira (1978) and Ando (1986) that the morae of the *shouga* are not exclusively related to single pitch values within the musical melody. In the absence of this "typical" solmization logic, there are several candidate melodic parameters which may correspond with the *shouga* more intelligibly: melodic *interval, relative interval,* and *contour* are all reasonable and practical alternative parameters to investigate. As we have seen, *contour,* or "the pattern of ups and downs" within the melody (Dowling, 1978), has received the most attention to this point. The results of Study 1, alongside the works of Hughes (2000) and Nelson (2008a, b), indicate some structural correspondence between the *vowel* sequences within the *shouga* and the *contour* sequences within the musical melody. As discussed in Study 1*,* however, this model predicts only about half of the structural mappings between *shouga* and melody within *Rokudan no*

Shirabe. This result leaves open two questions. The first is whether the consonant components of the *shouga* make an additional contribution to the differentiation of melodic sequences. For this reason, we use *mora* (including both consonant and vowel components), instead of *vowel,* as the *shouga* parameter in the following analyses.

The second concerns whether *contour* is a sufficient index of melodic form. Melodic contour, according to its current operational definition (cf., Dowling, 1978) conveys only the direction of a given pitch vector and does not take into account the relative magnitude of the distances (or intervals) between adjacent pitches. As a result, this melodic dimension may be unnecessarily limited in its ability to capture the relationship between *shouga* and melody. It is plausible that the subjective impressions of musical form are more nuanced than those implied by the coarse index of whether pitches are indefinitely "higher" or "lower" than one another. Indeed, we know from auditory scene analysis research that the subjective impression of the number of sound sources (streams) elicited by a monophonic tone sequence is influenced by the magnitude of the intervals among the notes (Bregman, 1990). That is, auditory perception is qualitatively differentiated on the basis of interval magnitude thresholds. This possibility is worth particular consideration in light of the knowledge that the *shouga* is not simply a string of individual mora one after the other but is rather a sequencing of *idioms,* which represent qualitatively distinct subjective impressions as the musical form develops across time. Thus, what is needed is a melodic parameter that is more abstract and less specific than the explicit musical *pitches* of the musical melody on the one hand, and less abstract and more specific than the sequence of *contours* describing the musical melody on the other. The two "intermediate" parameters we will consider are the *interval* and *relative interval* descriptions of melodic sequences.

Melodic *interval,* considered only in the quantitative sense, is a direct size-(t-1) reflection of a size-t melodic *pitch* tuple: Everything one needs to know about the intervals of a specific melodic sequence can be gleaned from the pitch content of that melodic sequence. This is because *interval* describes both the direction and magnitude of a pitch change. If one knows the pitches defining a melodic sequence, then one knows which intervals relate the change from one pitch to the other. In this sense, the *interval* parameter may appear redundant. But the value of the *interval* is that it abstracts pitch movements from the literal (observed) pitches of a specific melodic sequence. *Interval* is therefore a structural melodic parameter capable of differentiating among specific pitch movements regardless of where this movement occurs within pitch space (like a first derivative of pitch space). This capacity of interval to provide a pitch-generalized description of pitch relation is the basis of exact melodic transposition (musical perception by *relative pitch relations*). Thus, the two melodic sequences D4–G4–A4–D4 and G4–C5–D5–G4 are composed of totally distinct pitch sequences[89] but totally identical interval sequences[90], and are perceived by listeners as formally analogous. Likewise, expert performers indicate that the idioms of the *shouga* are transposed throughout various absolute states of the musical surface in order to represent common musical "functions" and forms. Thus, the structural relation between the *shouga* and the melody may be mediated more precisely by characteristic *interval sequences* of a given *shouga* idiom than by the over-specified literal pitches of the melody. Said differently, mora sequences that correspond with different melodic *pitch* sequences may be meaningfully associated under the same statistical index in terms of a common melodic *interval* sequence.

The relative frequencies of unique *pitch* and *interval* sequences associated with a given *shouga* sequence within the melody also serve as a meaningful statistical index of the extent to

[89] [50, 55, 57, 50] vs. [55, 60, 62, 55]
[90] [+5, +2, -7]

which *shouga* sequences are associated with formally analogous musical motifs: If the frequencies of unique *pitch* sequences and unique *interval* sequences associated with a given *shouga* sequence are basically equivalent, this would indicate that there is no systematic association between a given *shouga* sequence and a particular musical interval. But if the frequencies of unique *pitch* sequences associated with a given *shouga* sequence are markedly higher than the number of unique *interval* sequences associated with a given *shouga* sequence, this result would indicate that the interval parameter is effectively consolidating common pitch motion patterns associated with a given *shouga* sequence, i.e., that a given *shouga* sequence is associated with a characteristic melodic interval range. In the latter case, the *interval* parameter may be regarded as a useful statistical tool for tracking the structural correspondence between a given *shouga* idiom and the musical melody as the idiom is transposed throughout the melodic pitch space.

From the statistical analysis perspective, one limitation of the *interval* parameter is that although it abstracts *pitch* relations, it, like *pitch,* corresponds with a very specific literal value. Quantitatively, the interval sequence [-2, -3] is just as computationally distinct from [-2, -4] as it is from [-2, +9]. Qualitatively, however, there is a clear perceptual similarity between [-2, -3] and [-2, -4] that is not shared with the sequence [-2, +9]. This consideration is particularly important given that the *shouga* is representing common subjectively perceived forms, not literal objective states (as in *solmization*)—especially when one considers that as a common musical form (or melodic motif) is transposed throughout the pentatonic pitch space of *hirajoshi,* the literal intervals defining this motif will shift.

Contour, which retains only the sign component of the interval vector, has been used to approximate this categorical similarity in the past (e.g., Dowling, 1978; Hughes, 2000). But what about sequences where the contour is retained, but the relative interval sizes are varied? For example, [-5, -2] has a qualitatively distinct form from that of [-2, -5] despite the fact that they share the same contour [-, -]. In the first case, the first interval is larger than the second. In the second case, the second interval is larger than the first. Given the expressed intention of the *shouga* to represent the subjective impression of musical form, an ideal melodic parameter would be able to distinguish between the relative magnitudes within an interval sequence on the basis of their perceived categorical similarity and difference. Such a project is deserving of an extensive psychophysical research program in its own right. For the present purposes, we will approximate this ideal parameter through the implementation of the purely objective *relative interval* (or *ordinal interval ranks*) parameter.

Just as the *interval* parameter abstracts pitch movements from the literal pitches of a particular melodic sequence, the *relative interval* parameter abstracts *interval movements* from the literal intervals of a particular melodic sequence. The so-called *relative interval* parameter generalizes the interval relations within a sequence by normalizing the individual interval events within a sequence with respect to the local maxima and minima within that sequence. This parameter therefore consolidates interval sequences that share the same sequence of relative magnitudes and directions. Continuing with the same example used to demonstrate the abstraction from pitch to interval, we return to the two melodic sequences D4–G4–A4–D4 and G4–C5–D5–G4 that share the same *interval* sequence [+5, +2, -7]. Assigning ordinal ranks to the intervals within this sequence yields the *relative interval* sequence [3, 2, 1]. Just as the alternative coding of *pitches* as *intervals* is redundant when comparing identical pitch sequences, the alternative coding of *intervals* as *relative intervals* is redundant with respect to identical interval sequences. But the utility of the *relative interval* parameter is its capacity to reflect qualitatively familiar melodic movements across different literal interval sequences by maintaining the ordinal ranks of the interval magnitudes and directions within the total sequence. Thus, the *relative interval* sequence [3, 2, 1] common to both D4–G4–A4–D4 and G4–C5–D5–G4 is also shared with melodic variants such as D4–A4–Bb4–D4 and D4–G4–Bb4–D4, defined by the distinct interval sequences [+7, +2, -9] and [+5, +3, -9], respectively. As a result, the *relative interval* parameter is intended to capture the common interval *trends* among the various absolute *interval* sequences associated with a particular *shouga* sequence. As such it should reflect an "intermediate" level of resolution concerning the structural relations between the *shouga* and melody as compared with the under-specified, more abstracted *contour* parameter on the one hand, and the over-specified, less abstracted *interval* parameter on the other. That is, two instances of the same melodic *contour* tuple may be meaningfully distinguished from one other on the basis of their respective *relative interval* sequences, while distinct *interval* tuples may be meaningfully associated with one another on the basis of their common *relative interval* sequence.

One obvious limitation of this *relative interval* parameter is that it allows an ostensibly endless combination of pitch sequences to be treated as equivalent in terms of their ordinal ranks, though they may be clearly perceived as qualitatively distinct. For example, the [+5, +2, -7], [+7, +2, -9], and [+5, +3, -9] interval sequences reported above, which all share a qualitatively similar interval sequence, share their common *relative interval* sequence [3, 2, 1] with the interval sequence [+2, + 1, -13]. This latter *interval* sequence is obviously qualitatively distinct from the former set, which would apparently undermine the original intention for the development of such a variable. However, such an interpretation ignores several critical facts. First, in terms of a methodological precedent, the exact same limitation is true of *contour,* which has been used extensively and effectively by music perception researchers in the last 40 years (at least; McAdams, 1987). Second, the quantifiable pitches and intervals within *Rokudan no Shirabe* are *not* limitless. Third, the qualitatively distinct forms within the piece are *also* not limitless. They reflect a common theme, a cultivated artistic expression of a particular musical idea, or one might even say a *thesis,* in the parlance of the Hegelian dialectic. While it is easy to be carried away by the apparent infinities (countable or otherwise) *available* to researchers through quantitative analysis of art objects, to do so is to ignore the reality of the constraints posed by the object itself. In this case, such constraints are the result of a creative process mediated by subjective intention and perception. Finally, to reiterate, the idioms and sequencing of the *shouga* used to represent the form of this object are expressly intended to convey the common musical forms perceived by the subject. Therefore, while a *relative contour* sequence associated with a particular *shouga* sequence *could* be associated with a limitless set of variant *interval* sequences, this possibility is virtually negligible in the present context. For these reasons, the *relative interval* parameter is anticipated to do more good than harm for the present purposes, in that it will 1) consolidate similar *interval* sequence variations of a common musical motif that may be associated with a particular *shouga* sequence, 2) effectively differentiate these qualitatively similar *interval sequence* variations from an even larger set of starkly different *interval* sequences (such as retrograde or shuffled permutations of the same intervals), and 3) effectively differentiate multiple instances of the same *contour* sequence associated with a particular *shouga* sequence that may be characterized by different *relative interval* magnitudes. The validity of this prediction is of course to be demonstrated by the appropriate selection of statistical analyses and their complementary results.

Taken together, the foregoing considerations yield severable tenable predictions. We will address three: First, given the intent of the *shouga* to represent common subjectively perceived musical forms and the perceived similarity of transposed intervals, *shouga* sequences should be more closely associated with characteristic *interval* sequences transposed throughout pitch space than with the literal *pitch* sequences themselves. Second, given that several *interval* sequences can be considered as structural variants of the same melodic form despite their distinct literal values, *shouga* sequences should be more closely associated with *relative interval* sequences than with *interval* sequences. A corollary to this prediction is that to the extent the *relative interval* parameter can distinguish between two instances of the same *contour* interval differing in their relative interval magnitudes, more unique *relative interval* sequences should be observed to correspond with a particular *mora* sequence compared to the number of unique hypothetically under-specified *contour* sequences associated with the same *mora* sequence. Finally, in spite of the ostensibly limitless interval sequences with which a particular *shouga* sequence might be associated, the psychologically informed prediction is rather that *mora* sequences will be associated with constrained interval ranges (in terms of both interval direction and magnitude). To explore these possibilities, we present two complementary descriptive statistics. The first is the weighted-average number of unique *pitch, interval, relative interval,* and *contour* tuples associated with a given *mora* tuple of size *t*.[91] The second is the weighted-average melodic interval associated with each possible *mora* duple found within the corpus.

[91] Where the average number of *unique* tuples of a given [melodic parameter] corresponds with the average number of *distinct [melodic parameter] states* associated with a given *mora* tuple of size *t*.

II.1. Weighted-average frequency of unique pitch, interval, relative interval, *and* contour *tuples mapping onto a single* mora *tuple*

II.1.1. Hypotheses

The first analysis (II.1) is directed toward the investigation of the first and second predictions: If the *pitch* parameter of the musical melody is over-specified in its association with the idioms of the *shouga* relative to the *interval* parameter, then we should expect to find that the frequency of unique *pitch* tuples (P) associated with a given *mora* tuple (M) is greater than the frequency of unique *interval* tuples (I) associated with the same *mora* tuple, on average (P(M)> I(M)). Similarly, if the literal *interval* parameter of the musical melody is over-specified in its structural relation with the *shouga* as compared with the *relative interval* parameter, then the frequency of unique *interval* tuples associated with the average *mora* tuple should be greater than the frequency of unique *relative interval* tuples (O) associated with the same *mora* tuple, on average (I(M) > O(M)). Finally, if the *relative interval* parameter of the musical melody is more precise than the *contour* parameter of the musical melody, then more unique *relative interval* tuples should be observed to correspond with a single *mora* tuple compared with the number of unique *contour* tuples. Stated colloquially, more unique *pitch* tuples are expected to be associated with the average *mora* tuple compared with the number of unique *interval* tuples, which are in turn expected to be more frequent than the number of unique *relative interval* tuples associated with the average *mora* tuple. In addition, the number of unique *relative interval* tuples associated with the average *mora* tuple should be greater than the number of unique *contour* tuples (O(M) > C(M)). Thus, taken together, the hypothesis predicts that P(M) > I(M) > O(M) > C(M).

II.1.2. Procedure

The standard equations used for the calculation of weighted means (and standard deviations) are included in Appendix 3B.

The following procedure was replicated for each tuple size t = [1, 17].[92] A sliding window of size t was used to search through every length-t event in *Rokudan no Shirabe* (N = 734-(t-1)). Every *shouga* sequence of length t was parameterized in terms of a corresponding t-gram *mora* tuple (e.g., the *shouga* sequence *ko-ro-ri* was coded as the *mora* tuple [ko, ro, ri]). Every melodic sequence was parameterized in terms of a *pitch, interval, relative interval,* and *contour* tuple. A *pitch* tuple consisted of the pitch height MIDI values drawn directly from the database. An *interval* tuple consisted of positive and negative integers reflecting the direction (+, -, =) and magnitude (in semitones) between adjacent pitch values within the melodic sequence. The *relative interval* tuples recoded each of the *interval* tuple values in terms of their ordinal ranks relative to the other interval values within the melodic sequence: Lowest relative intervals (largest negative values) corresponded to lowest ordinal ranks, highest relative intervals (largest positive values) corresponded to highest ordinal ranks, and equivalent relative intervals shared the same rank. The *contour* tuples consisted of rising [+], falling [-], and equivalent [=] value indicators reflecting the direction between adjacent pitch values within the melodic sequence.

For example, the 3-tone melodic sequence A4–C5–D4 occurs 5 times within the corpus. This sequence was represented during analysis by the *pitch* tuple [57, 60, 50], the interval tuple [+3, -10], the *relative interval* tuple [2, 1], and the contour tuple [+, -]. As another example, the 5-tone melodic sequence A5–G5–Eb5–D5–C5 occurs 3 times within the corpus. This sequence

[92] The set of tuple sizes was dictated by the range in which a non-unique tuple mapping could be found.

was represented with the 5-gram *pitch* tuple [69, 67, 63, 62, 60], the 4-gram *interval* tuple [-2, -4, -1, -2], the 4-gram *relative interval* tuple [2, 1, 4, 2], and the 4-gram *contour* tuple [-, -, -, -].[93]

Every time a unique *mora* tuple was encountered, it was appended to an empty array, and a frequency counter for that tuple (*mora tuple frequency*) was increased by one. For every subsequent instance of the same tuple, the corresponding frequency counter was increased by one, but the tuple was not again appended to the array. This array thus constituted the bins of the subsequent *[Melodic parameter] x Mora* histograms. Explicitly, four such arrays were created per tuple size *t*: A *Pitch* x *Mora* (P x M) array, an *Interval* x *Mora* (I x M) array, a *Relative interval* x *Mora* (O x M) array, and a *Contour* x *Mora* (C x M) array.

Each of these histograms was populated by using each of the unique *mora* tuples to query the database. Each time a given mora *tuple* was encountered within the database, the corresponding *pitch, interval, relative interval,* and *contour* tuples were extracted. If a given [melodic parameter] tuple had not yet been observed to correspond with the *mora* tuple used to query the database, the *unique [melodic parameter] frequency counter* for that melodic parameter sequence was increased by one; if it had previously been encountered, it was ignored. For example, the *shouga* sequence *te-tsu-to* was represented by the 3-gram *mora tuple* [te, tsu, to] and occurred 8 times within the corpus. It was associated with 6 different *pitch* tuples, 4 different *interval* tuples, 1 *relative interval* tuple, and 2 different *contour* tuples. As another example, the *shouga* sequence *tsu-te-ko-ro-ri* represented by the 5-gram *mora tuple* [tsu, te, ko,

[93] Note that the length of a given *pitch* tuple (*t*) is identical to that of the corresponding *mora* tuple, while the *interval, relative interval* and *contour* tuple lengths (*t*-1) are shorter by one index. Furthermore, the *interval* and *contour* tuples were not computed for tuple sizes of $t = 1$, as they convey a relation between adjacent musical events, the *relative interval* tuple was not computed for $t < 3$ because an *interval* tuple must consist of more than 1 interval item ($t > 2$) in order for the intervals to be assigned relative ranks. In the Tables and description of the Results section below, we refer only to the tuple size *t* of the *mora* tuples for the sake of concisely articulating the relationship between the *shouga* and each of the melodic parameters. This value is understood, however, to correspond with different tuple sizes of the respective melodic parameter tuples (*t* or *t*-1), depending on which parameter is concerned.

ro, ri] occurs 9 times within the corpus. It corresponded with 7 different *pitch* tuples, 4 different *interval* tuples, 5 different *relative interval* tuples, and 2 different *contour* tuples.

The resultant *unique [melodic parameter] frequencies* constituted the *counts* (x) while the *mora tuple frequencies* divided by the total number of unique mora sequences of size t were used as weights (w) for the calculation of the *weighted mean frequency* ($\bar{x}$) of unique *pitch, interval, relative interval,* and *contour* tuples mapping onto a single t-gram *mora* tuple. Unique morae tuples of size t with greater frequencies within the corpus contributed more weight to the estimate of the average number of unique [melodic parameter tuples] associated with a given mora of size t than did those mora tuples with lesser frequencies. For example, the 5-gram *mora* tuple [to, shya, chi, te, tsu] only occurs once out of N = 730 5-gram mora tuples, and therefore the frequency of unique *[melodic parameter]* tuples associated with this mora tuple contribute less to the estimate of unique *[melodic parameter]* tuples associated with the *average* 5-gram *mora* tuple than, for example, those associated with the 5-gram mora tuple [shya, shya, te, shya, shya], which occurred 11 times.

Calculating the inverse frequencies

For the sake of completeness, we also calculated the inverse unique mapping frequencies. That is, acknowledging the previously recognized fact that “different melodic events require a common function [as represented by] the same mora,” (Mananoi & Akira, 1978, p. 36), we also computed the average number of unique *shouga* sequences mapping onto a given melodic parameter state. Although this analysis does not directly address the hypotheses outlined above, we include them for the interested reader. The procedures for querying the corpus, developing the histograms, and calculating the respective weighted-average frequency and weighted-standard deviation of the mean for all tuple sizes t = [1, 17] were all equivalent to those described above. The only difference in the case of this complementary analysis was that unique *[melodic parameter]* tuples were first appended to an empty array to create the bins of the subsequent *unique mora x [melody parameter]* histograms. These were then populated with frequencies by querying the database for unique *shouga* tuples associated with each *[melodic parameter]* state. The resulting arrays were referred to as the *Mora* x *Pitch* (M x P) array, the *Mora* x *Interval* (M x I) array, the *Mora* x *Relative Interval* (M x O) array, and the *Mora* x *Contour* (M x C) array. The associated weighted-average frequencies M(P), M(I), M(O), and M(C) for each tuple size *t* are presented in Appendix 3C.

II.1.3. Results – Weighted-average frequencies of unique pitch, interval, relative interval, *and* contour *tuples mapping onto a given* mora *tuple*

The results are presented in Table 3.7. Consistent with the first hypothesis, we find that P(M) > I(M), O(M) > C(M) for 4 <= *t* < 9. Indeed, there are no exceptions to this pattern at any tuple size within this range. From *t* >= 9, identity between the *pitch* and *interval* parameters is observed. All four parameters reach identity at *t* = 12. Concerning the specific relation between *interval* (I) and *pitch* (P), we find that the average number of unique *interval* tuples associated with a single *mora* tuple is smaller than the average number of unique *pitch* tuples associated with a single *mora* tuple (P(M) > I(M)) for all possible tuple sizes *t* = [2, 8] until *interval* and *pitch* reach identity at *t* = 9. The specific relation between *interval* tuples and *relative interval* (O) tuples follows the same trend, with I(M) > O(M) for all possible tuple sizes *t* = [3, 11] until *relative interval* and *interval* reach identity at *t* = 12. Finally, concerning the specific relation between the *relative interval* (O) and *contour* (C) parameters, we find that O(M) > C(M) for tuple sizes *t* = [4, 9], indicating that the average 3-gram *mora* sequence is associated with a greater number of distinct *contours* than it is with structural variants of a given *contour,* but for *t*

>= 4, the *relative interval* parameter effectively distinguishes among the interval variants of melodic sequences exhibiting the same *contour*.

Table 3.7 Weighted average frequency (*SD*) of unique *Pitch* (P), *Interval* (I), *Relative Interval* (O), and *Contour* (C) tuples mapping onto unique *Mora* (M) tuples of length *t*

Mora Tuple Size (t)	N	P(M) (*SD*)	I(M) (*SD*)	O(M) (*SD*)	C(M) (*SD*)
1	734	7.72 (1.73)			
2	733	7.38 (3.69)	3.71 (2.57)		1.41 (0.51)
3	732	4.97 (2.76)	3.33 (2.06)	1.36 (0.57)	1.48 (0.69)
4	731	3.10 (1.93)	2.45 (1.40)	1.61 (0.87)	1.41 (0.63)
5	730	2.05 (1.36)	1.82 (1.10)	1.52 (0.86)	1.27 (0.52)
6	729	1.44 (0.75)	1.38 (0.68)	1.28 (0.60)	1.14 (0.39)
7	728	1.20 (0.43)	1.19 (0.42)	1.16 (0.40)	1.07 (0.25)
8	727	1.11 (0.32)	1.10 (0.32)	1.09 (0.30)	1.05 (0.22)
9	726	1.05 (0.21)	1.05 (0.21)	1.04 (0.20)	1.03 (0.16)
10	725	1.01 (0.13)	1.02 (0.13)	1.01 (0.10)	1.01 (0.10)
11	724	1.01 (0.09)	1.01 (0.09)	1.01 (0.07)	1.01 (0.07)
12	723	1 (0.00)	1 (0.00)	1 (0.00)	1 (0.00)
13	722	“	“	“	“
14	721	“	“	“	“
15	720	“	“	“	“
16	719	“	“	“	“
17	718	“	“	“	“

Incidentally, this analysis also revealed that although the *mora* tuples are entirely unique by the time the tuple size is increased to 17 (cf., analysis 1.2, Table 3.5), there remains a duplicate *contour* tuple of this length. That contour tuple, with its respective *shouga* sequences and *pitch* tuples is presented in Table 3.8. One will notice in the inverse weighted-average frequencies (M(O) vs M(C), t = 17) presented in Appendix 3C that the *relative interval* parameter (O) effectively distinguishes between these two relative interval magnitude variants of the same *contour* tuple.

Table 3.8 The longest repeated contour tuple found in *Rokudan no Shirabe* (N = 2)

No Coding Scheme	
Size of Largest Repeated Contour Tuple (t-1)	16
Contour Tuple	[-, -, +, -, +, -, -, =, +, -, =, +, -, +, +, -] (N = 2)
Corresponding *Shouga* Sequence(s)	*ko-ro-ri-chi-to-te-to-shya-shya-te-shya-shya-chi-to-te-tsu-to* *ko-ro-ri-chi-to-te-to-shya-shya-te-shya-shya-chi-to-te-tsu-shya*
Corresponding Pitch Tuple(s) (Length = t)	[57, 55, 51, 55, 38, 50, 45, 38, 38, 50, 38, 38, 51, 38, 50, 55, 46] [57, 55, 51, 55, 38, 50, 45, 38, 38, 50, 38, 38, 51, 38, 50, 55, 45]

II.2. Analysis of weighted-average interval size for each mora-*duple and* vowel-*duple sequence found within the* shouga *of* Rokudan no Shirabe

II.2.1 Hypotheses

Analysis II.2 addresses whether the various mora pairs within the *shouga* are identifiable in terms of characteristic interval ranges. If mora pairs are characterized by particular interval trends, then we should find that the majority of intervals are associated with a distinct interval range as indexed by relatively low variability (weighted standard deviation) in the interval sizes associated with the mora pair. If these mora pairs are considered as proxies for the "building blocks" of the *shouga idioms* (or particular morphological impressions), then we might also expect to find that the mora pairs are associated with a particular interval direction, i.e., that they are not observed to be multiply associated with positive, negative, and equivalent (+, -, =) contours.

Finally, this study will address whether the *relative interval* parameter used in analysis II.1 can be meaningfully interpreted: If the standard deviation of the average interval associated with a particular *mora* duple is demonstrably constrained, then this variability can reasonably be interpreted as the result of "structural variants of the same melodic form." That is, if the *mora* duples are associated with constrained interval ranges, then this suggests a greater possibility of perceived categorical semblance among the various exemplars. In this case, the *relative interval* parameter reported in analysis II.1 can reasonably be interpreted as consolidating qualitatively similar interval tuples. In the alternative case, however, that the standard deviation of the average interval associated with a particular *mora* duple is highly variable, then the *relative interval* parameter used in analysis II.1 should be dismissed, as this result would demonstrate that the supposed "categorical variations" of a central *interval* tendency cannot reasonably be argued for (or measured) in light of the fact that the average *mora* duple is associated with such a wide range of interval magnitudes. Likewise, if most *mora* pairs are found to be associated with a specific contour, this would also corroborate the hypothesis that distinct *shouga* sequences are associated with qualitatively distinct musical forms, whereas the opposite result in which most *mora* pairs are found to be associated with multiple contours would again undermine the interpretability of the *relative pitch* parameter. Analysis II.2 provides a general overview of how closely the various *shouga* sequences are associated with distinct interval ranges and will be complemented by the case-by-case in-depth analyses of the well-known *shouga* idioms described by Mananoi and Akira (1978) and Ando (1986) in Study 3.

For the sake of continuity with the previous work presented by Hughes (2000) and the first study of the current work, we conduct two variants of this analysis: once using the *mora* parameter and the second time using the *vowel* parameter. Comparing the results of these two analyses will address the more general question concerning whether the consonant components of the *shouga* contribute to its structural coordination with the musical melody. We anticipate that this is true and expect to find a greater specificity (reduced weighted standard deviation, reduced liability to change sign) in the characteristic intervals of *mora* duples than in the characteristic intervals of *vowel* duples.

II.2.2. Procedure

II.2.2.1. Using mora as the shouga *parameter*

We calculate the weighted average and standard deviation of the interval size for each of the 56 distinct 2-gram *mora* tuples found in *Rokudan no Shirabe* (cf., Table 3.5, above). Each of the 56 distinct mora duples was used to query the database. When the mora tuple was encountered in the database, the associated *interval* value (1-gram tuple *x*) was extracted. As in analysis II.1, the weights for each frequency estimate were proportional to the ratio of the frequency of a particular observation (in this case, the frequency of a specific interval value associated with a given mora duple) and the total frequency of that mora duple within the corpus (see Appendix 3B).

II.2.2.2. Using vowel *as the* shouga *parameter*

The procedure was identical to that in II.1, except that the corpus was queried for each of the 24 distinct 2-gram *vowel* tuples found in *Rokudan no Shirabe.*[94]

II.2.3. Results - *Weighted-average interval size for each* mora-*duple and* vowel-*duple sequence found within the* shouga *of* Rokudan no Shirabe

The results of analysis II.2 are presented in Tables 3.9 and 3.10. As displayed in Table 3.9, the weighted-average intervals associated with each of the distinct *mora* duples in *Rokudan*

[94] Though there are 25 possible vowel combinations, the vowel sequence *o-u* is never present.

no Shirabe range in size and contour from –23.5 semitones (*tsu-shyu, SD* = 1.5) to +31 semitones (*shya-sa, SD* = N/A; *shyu-ko, SD* = 0.00). Given this extensive range of interval magnitudes, it is truly remarkable how constrained the interval sizes are for a given duple—only 6 of the 56 duples (10.7%) possess a standard deviation greater than one semitone, and of these, only four duples (*te-tsu,* mean = 0.15, *SD* = 0.31; *shya-shya,* mean = 0.18, *SD* = 0.18, *ri-ko,* mean = 0.8, *SD* = 1.45, and *te-te,* mean = 1.8, *SD* = 1.23) are liable to exhibit different contours within the corpus.

These results are in contrast to the same statistics calculated for the 24 distinct *vowel* duples within the corpus, displayed in Table 3.10. Here, the average interval magnitudes range from -14.94 (i-a, *SD* = 0.59) to 18.23 (a-u, *SD* = 0.92). Only 12.5% of these duples possess a standard deviation greater than 1 semitone, but more than 20% of them are liable to be found associated with 2 or more melodic contours within the corpus. Indeed, there are no flat [=] contours represented in the weighted means of the vowel set, indicating that every repeated vowel sequence is associated at least once with non-flat melodic contours, contrary to the predictions of the vowel coding schemes (see Study 1).

Table 3.9 Weighted-average interval range (in semitones) for each of the 56 mora duples in *Rokudan no Shirabe*

Average Interval	Shouga Duple	N	Standard Deviation	Average Interval	Shouga Duple	N	Standard Deviation
-23.50	*tsu-shyu*	2	1.50	0.15	*te-tsu*	47	0.31
-20.00	*ri-shyu*	1	0	0.18	*shya-shya*	38	0.18
-15.21	*ri-shya*	14	0.70	0.80	*ri-ko*	5	1.45
-15.15	*chi-to*	13	0.23	1.15	*tsu-te*	47	0.09
-13.33	*ri-to*	3	5.45	1.61	*tsu-ko*	23	0.12
-13.00	*chi-shya*	2	0.00	1.80	*te-te*	5	1.23
-13.00	*tsu-shya*	12	0.83	1.87	*ri-tsu*	15	0.31
-12.00	*sa-ra*	2	0.00	2.00	*ka-ra*	3	0.91
-11.61	*te-shya*	23	0.55	2.00	*tsu-rya*	1	0
-9.93	*tsu-to*	14	0.44	2.00	*tsu-tsu*	8	0.38
-9.78	*te-to*	37	0.57	2.00	*yu-rya*	1	0
-6.00	*ra-ri*	2	0.00	2.60	*tsu-chi*	10	0.17
-5.80	*to-shya*	10	1.15	2.92	*te-chi*	24	0.59
-4.00	*yu-chi*	1	0	2.92	*te-ko*	13	0.74
-3.97	*chi-tsu*	32	0.15	3.00	*re-tsu*	2	0.00
-2.84	*ko-ro*	69	0.02	3.57	*ri-chi*	14	0.19
-2.72	*ro-ri*	69	0.03	4.00	*ru-tsu*	1	0
-2.00	*rya-yu*	2	0.00	9.67	*ra-te*	3	1.22
-1.05	*ri-te*	19	0.08	12.00	*te-sa*	1	0
-1.00	*chi-ko*	5	0.00	12.00	*to-chi*	1	0
-1.00	*chi-te*	10	0.00	12.00	*to-ko*	13	0.00
-0.67	*to-ka*	3	0.30	12.00	*to-te*	40	0.00
0.00	*chi-chi*	2	0.00	13.00	*shya-ko*	7	0.94
0.00	*ri-ri*	1	0	14.67	*shya-te*	27	0.41
0.00	*rya-rya*	1	0	16.83	*shya-chi*	12	1.37
0.00	*te-re*	2	0.00	18.23	*shya-tsu*	13	0.92
0.00	*to-to*	3	0.00	31.00	*shya-sa*	1	0
0.00	*tsu-ru*	1	0	31.00	*shyu-ko*	3	0.00

Table 3.10 Weighted-average interval range for each of the 24 vowel duples in *Rokudan no Shirabe*

Average Interval	Vowel Duple	N	Standard Deviation
-14.94	i-a	16	0.59
-13.00	u-a	12	0.83
-10.63	e-a	24	1.10
-9.15	i-o	26	0.89
-6.48	e-o	50	0.57
-4.62	o-a	13	0.92
-2.51	o-i	70	0.21
-2.46	i-u	50	0.41
-2.25	u-u	12	2.69
-1.03	i-e	29	0.05
-0.47	o-o	85	0.15
-0.23	u-o	40	0.93
0.27	e-u	49	0.30
0.45	a-a	44	0.78
1.15	u-e	47	0.09
1.29	e-e	7	0.91
2.00	u-i	13	0.51
2.78	i-i	18	0.20
2.92	e-i	24	0.59
12.00	o-e	40	0.00
13.00	a-o	7	0.94
13.57	a-i	14	2.06
14.17	a-e	30	0.47
18.23	a-u	13	0.92

Discussion

The analyses of Study 2 provide several convergent insights regarding the structural coordination between the *shouga* and the musical melody. Starting with analyses I.1 and I.2 we find clear evidence of the assertions made by expert performers and ethnomusicologists that practitioners of the *shouga* are not arbitrarily babbling along to the sound of the music—either with a single repeating mora (“lalala”) or with whatever speech sounds happen to come to mind.

Instead we find measurable biases in the sounds used within the corpus, their respective frequencies, and the repetition of extended sequences (up to 16 mora long) throughout the corpus. The probability of encountering these structures through a purely random permutation of Japanese speech sounds is virtually nil. Finally, the fact that the longest repeated mora sequence is found to correspond with the same melodic sequence in both cases served as an initial indication that a coherent and recurrent mapping exists between *shouga* and melody. This coherence extends beyond the predictions made by the extant vowel coding schemes considered in Study 1. However implicit its internal logic may be, the *shouga* is evidently the product of an intelligent and intentional representation of musical form.

In analyses II.1 we investigated the average frequency of unique *pitch, interval, relative interval,* and *contour* tuples associated with a single *mora* tuple. The results indicate that all four parameters of the musical melody are empirically differentiable in their structural relation to the *shouga.* The observed relative frequencies among these parametric relations are consistent with the hypothesized results, namely that the *interval* parameter effectively consolidates similar *pitch* movements associated with particular *shouga* sequences, whereas the *relative interval* parameter effectively consolidates similar *interval* sequences associated with a given *shouga* sequence. These relations between *pitch, interval*, and *relative interval* were maintained up to sequences 8 events long.

It is intriguing that the statistical distinction between *pitch, interval,* and *relative interval* sequences persists until melodic sequences reach a length of 9. Recall that *Rokudan no Shirabe* is written in a duple meter. Although temporal parameters such as rhythm and duration were not coded in the current analysis of the melodic structure, sequences up to 8 events long are relevant to the consideration of musical phrasing in duple meter regardless of their absolute duration.

That is, binary sequences 2, 4, 6, and 8 events long are common musical phrase lengths in Japanese music (Tokita & Hughes, 2008), regardless of whether they are performed on quarter note, eighth note, or sixteenth note pulses. The fact that distinct structural relations between *shouga* and the *pitch, interval,* and *relative interval* parameters are maintained up to sequences of 8 events long may provide a further indication of how highly refined the *shouga*'s representation of musical form really is. Further research involving the subjective and expert segmentation of coherent musical phrases could be used to further explore this possibility (cf., Ayari & McAdams, 2003). These segmentations could then be used in a computational model to delineate perceptually relevant sequence boundaries within the objective musical parameters of the database. Within these boundaries, the correspondence between state transition probabilities of the melody and *shouga* could be used to develop more subtle models of the structural coordination between the *shouga* and melody. Such results would also reveal general trends concerning the level of abstraction of the musical melody with which the *shouga's* internal logic is concerned.

Finally, the *relative interval* parameter was observed to effectively differentiate the relative interval magnitude variants of repeated under-specified *contour* sequences 4–9 events long. An interesting exception to the general relation between the *relative interval* and *contour* parameters was the finding that the shortest *shouga* sequences (3 events long) were associated with more unique *contour* parameter states than *relative interval* states. This result indicates that the average 3-gram *shouga* sequence is associated with more than one unique *contour* state. This finding is explored in further detail in the *idiom*-specific analyses of Study 3. Beyond this exception, however, it appears that longer *shouga* sequences are more robustly associated with a particular *contour* sequence, and furthermore that these contour sequences are canalized in the musical structure as a set of related *relative interval* variations. Indeed, the *relative interval* parameter appears to be an effective and meaningful parameter for the differentiation of more subtle variations in the contour of a musical melody than the basic operational definition implied by the *contour* parameter.

To this end, we found through analysis II.2 a remarkably constrained relation between specific mora duples and their associated interval ranges. In contrast with the coarse impression provided by the weighted-average number of unique *contour* tuples associated with 2-gram mora sequences (cf., analysis II.1), the vast majority (N = 50/54) of mora duples were not only associated with a single melodic contour, but also a very precisely constrained interval range: 48 out of 54 mora duples had interval magnitude standard deviations smaller than one semitone. This finding is quite remarkable, as it indicates most directly of all the findings that the mora sequences of the *shouga* are associated with very particular movements within the musical melody. In comparing the average interval ranges of *vowel* duples and *mora* duples we also found evidence that the consonant components of the *shouga* are meaningful not only in terms of acoustic-iconic representation of musical timbres, but also of typical melodic parameters such as *interval.*

The results of these analyses are all consistent with the artistic understanding that the structural delineations of the *shouga* reflect subjectively differentiated musical forms rather than objectively differentiated musical parameter states. Much perceptual research needs to be done to determine which of and to what extent the various melodic parameters are associated with listeners subjective judgments. Such studies have the potential to reveal the psychological processes underlying the subjective experience of musical form, specifically, and the perception of structural analogy, category, and resemblance (“similarity) more generally. To this end, the current results indicate clearly that *shouga* idioms are repeated across interval, relative interval, and contour patterns throughout the music. A more detailed analysis of the association between particular *shouga* idioms and the various melodic parameters of the musical object is presented in Study 3.

Study 3 – Parametric analyses of robust *shouga* idioms identified by expert pracitioners

Background

Here we conduct an in-depth investigation of the structural relations between various melodic parameters and the *shouga* idioms identified by expert musicians. As the results of Studies 1 and 2 have indicated, the *shouga* exhibits a complex, dynamic, and uncertain structural relationship with the musical melody it represents when considered in terms of objective and statistical musical parameters. Nevertheless, expert practitioners are intuitively sensitive to all of these parametric constraints and can detect when a given idiom is incorrectly associated with a particular musical phrase (Ando, 1986; 2017). That is, although the individual differences among *shouga* are too complex to be modelled by simple statistical measures, they are nevertheless intelligible and coherent to the perceiving subject on a case-by-case basis. Therefore, in the following analysis we treat each of the well-known *soukyoku* idioms independently, comparing and contrasting their most characteristic and robust associations with the various musical parameters. There are no pointed hypotheses in this regard, other than our expectation of finding that each idiom is particularly well differentiated on some melodic parameters while being relatively "unrelated" to others, and that the set of these parametric relations will change on an individual basis. The results are intended to convey to the reader an intuitive understanding of the musicality of the *shouga* in descriptive form.

Analysis Procedures

First, we searched through the treatises on *soukyoku shouga* of the *Ikuta-ryu* written by Mananoi and Akira (1978) and Ando (1986). We found all instances in which a *shouga* idiom was mentioned (N = 15) and then organized these references into *idiom*-based categories (see

Appendix 3D for the raw text excerpts). From these categories we extracted five critical *soukyoku shouga* idioms: *ko-ro-ri, chi-tsu-te, sa-ra-ri(n), shya-shya, to-te,* and *te-to.* We compared these idioms with the frequencies for 2-gram and 3-gram mora tuples in the database for *Rokudan no Shirabe* and discovered two additional highly prevalent mora-duples: *tsu-te,* and *te-tsu,* which were appended to the list. Therefore, our sample amounted to seven distinct *shouga* idioms.

For each idiom we conducted five analyses and produced a corresponding graphical representation of the results:

1. a representation of the unique pitch sequences associated with a given idiom, and their respective frequencies;
2. a normalized relative-interval sequence plot, in which all of the idiom interval sequences were presented with the local minimum set to "0" and the local maximum set to its integer value (in semitones) above this local minimum. Intermediate pitches (for *ko-ro-ri, chi-tsu-te,* and *sa-ra-ri*) were plotted according to their relative interval between the local maximum and local minimum;
3. a histogram of the sequences of scale degrees on which the idiom could be found;
4. an analysis of the various rhythms characterizing the idiom, and their respective frequencies;
5. a histogram of the metric positions on which the idiom began (i.e., where the first mora of the idiom was located within the musical meter).

Results

Ko-ro-ri

All 69 instances of *ko-ro-ri* follow the same contour [-,-]. Figure 3.4 shows the seven distinct pitch sequences that follow the *ko-ro-ri* idiom in *Rokudan no Shirabe,* and their respective frequencies. Figure 3.5 demonstrates that these seven pitch sequences can be reduced to three specific interval sequences. It is worth noting, however, that the third of these interval sequences ([-2, -3], N = 1) is a structural variant of the most common interval sequence ([-3, -3], N = 40), the two sequences differing only in their first interval magnitude by one semitone as a natural consequence of the tuning system used for *hirajoshi.* (*ko-ro-ri* is created by performing three adjacent strings; Ando 1986, p. 158–159; Appendix 3D.) With this in mind, it is perhaps more appropriate to conceptualize the seven *ko-ro-ri* pitch sequences as reducible to two different *relative-interval* sequences.

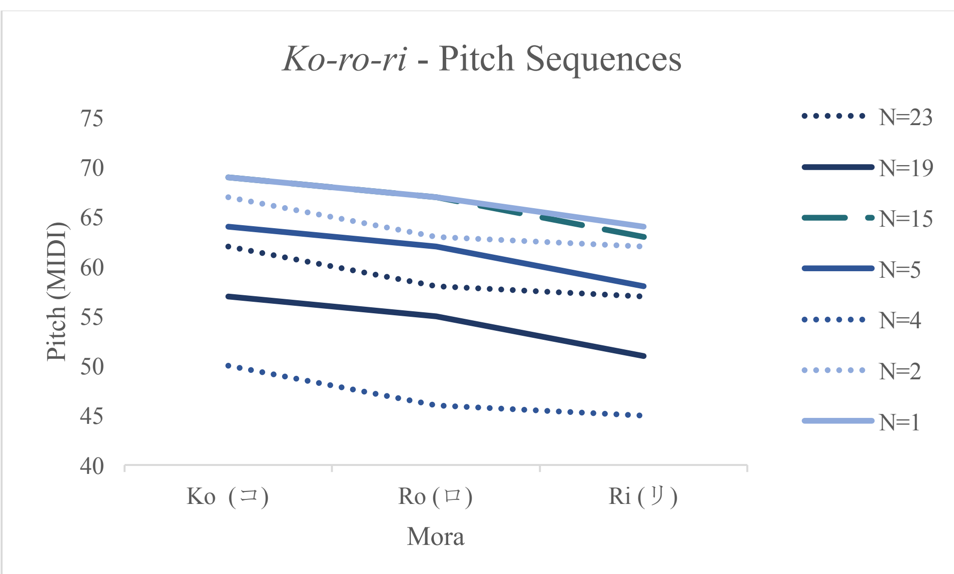

Fig. 3.4 *Ko-ro-ri* – Pitch sequences (N = 69)

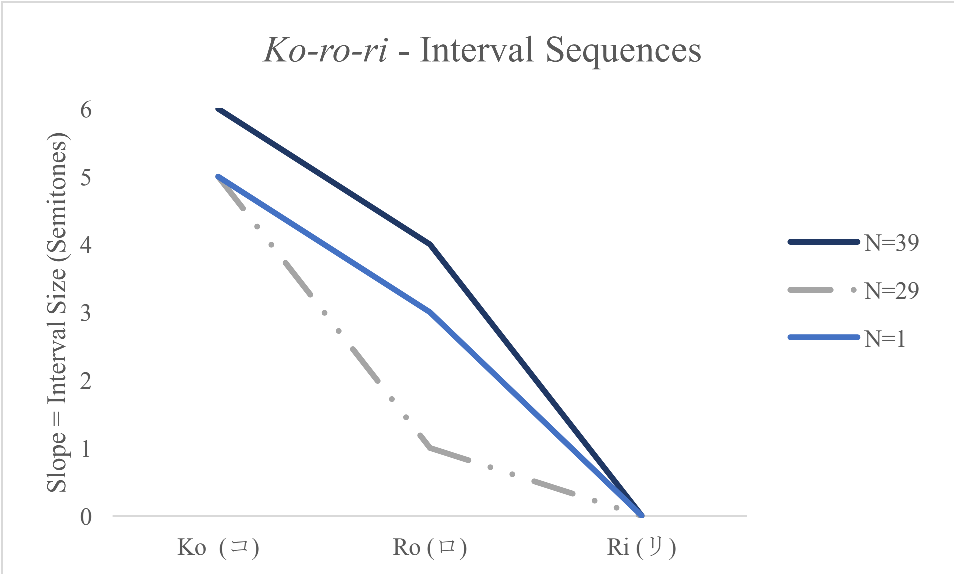

Fig. 3.5 *Ko-ro-ri* - Interval sequences (N = 69)

Figure 3.6 illustrates a common melodic motion profile between the mora of *ko-ro-ri*. Specifically, one notices that the most prevalent sequences (N = 61) are from the fifth scale degree (A; コ) through the fourth (G; ロ) to the second (Eb; リ), or from first scale degree (D; コ) through the sixth (Bb; ロ) to the fifth (A; リ). The remaining alternatives (N = 8) are either tonicized transpositions of these patterns ($\hat{2} - \hat{1} - \hat{6}$; $\hat{4} - \hat{2} - \hat{1}$) or variants by accidental (i.e., "sharp 2" ($\hat{5} - \hat{4} - \#\hat{2}$ compared to $\hat{5} - \hat{4} - \hat{2}$). The fact that *ko-ro-ri* reliably terminates on unstable tones may be posited as a partial psychophysical basis for the "questioning movement" associated with this idiom, on the basis of a heightened musical tension at the end of the corresponding melodic motif (cf., Mananoi & Akira, p. 37; Appendix 3D). Finally, we find that in all 69 cases, the motion is from a relatively more stable position to a relatively less stable position within the melody (Rose & Kapuscinski, 2013, 2014). Thus, the *i* of *ri* may indicate not necessarily a peak in melodic contour (as suggested by the vocal coding schemes in Study 1) but a peak in melodic *tension*, which would explain why this corpus "contradicts" the evidence for the strength of the vowel coding schemes previously provided by Hughes (2000). We return to this point in the General Discussion.

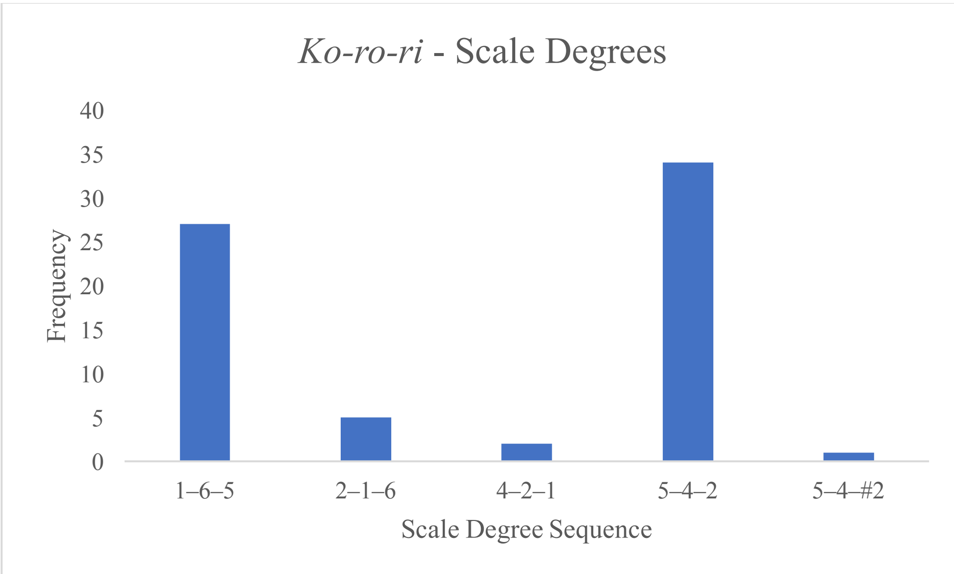

Fig. 3.6 *Ko-ro-ri* - Scale degrees (N = 69)

Figure 3.7 presents the rhythmic profiles generated by the relative durations of the mora within *ko-ro-ri.* By far the most common rhythmic variation (55/69) includes a triplet-eighth note relation between *ko-* (2/3) and *–ro* (1/3) while *ri* is held for an entire eighth note. The next most common rhythm is three eighth note durations in series (N = 7) and the remaining 7 cases are context-dependent phrase elongations or contractions.

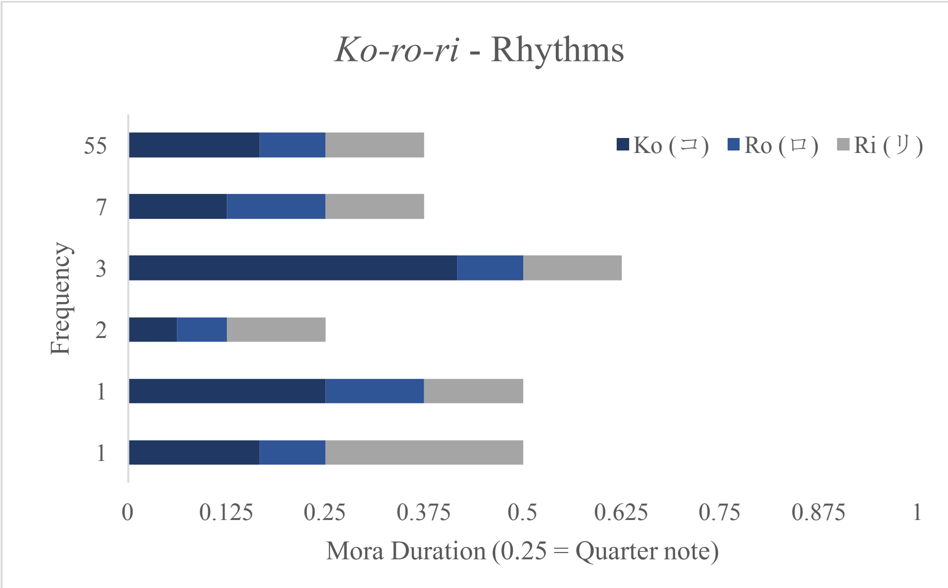

Fig. 3.7 *Ko-ro-ri* - Rhythms (N = 69)

Finally, Figure 3.8 presents the metric positions at which *ko-ro-ri* can be found, indicating the motif is most commonly located on quarter note downbeats (67/69 instances), with the two exceptions falling on eighth note off-beats surrounding beat 4.

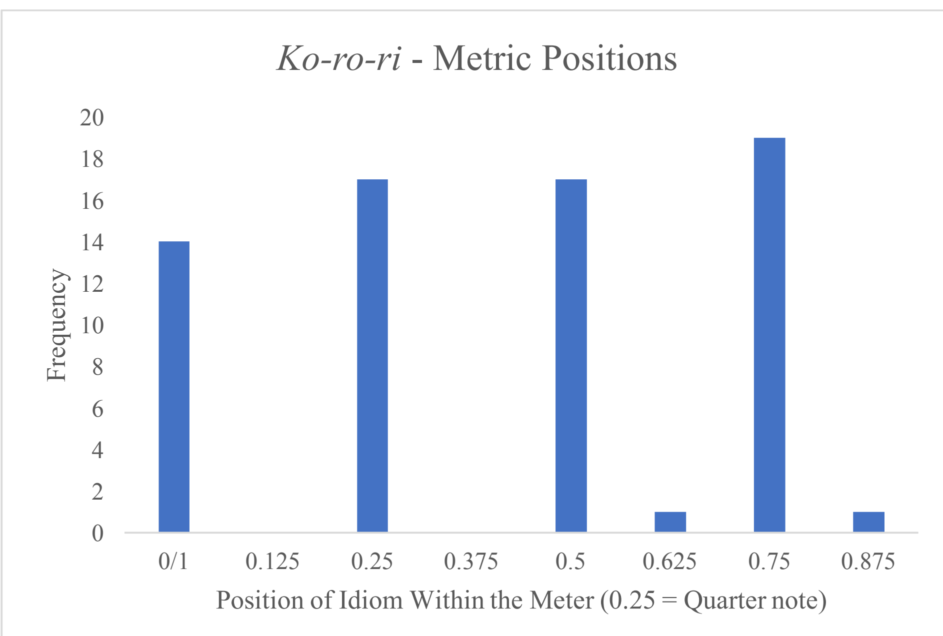

Fig. 3.8 *Ko-ro-ri* - Metric positions (N = 69)

Chi-tsu-te

Three quarters of the six distinct *chi-tsu-te* pitch sequences exhibit a [-,-] contour (N = 16) while the remaining four sequences exhibit three different variations of a [-,+] contour (Figure 3.9). Figure 3.10 illustrates that all 16 pitch sequences of the falling contour can be expressed by the same interval sequence [-4, -1], and most interestingly, this interval sequence is identical to the second-most-common interval sequence in *ko-ro-ri.* (Recall from Ando**,** 1986, p. 198, the fact that a pitch sequence possibly represented by *ko-ro-ri* "may also be written as *chi-tsu-te*, depending on the musical context.") The interval sequence [-4, -1] is represented in each graph with the same line style (dash-dot-dot) for ease of visual comparison. Note that although the interval sequence occurs frequently within both idioms, it occurs at different *relative* frequencies for each (*i.e.,* [-4, -1] is the second-most-common interval sequence in *ko-ro-ri,* but the most common interval sequence in *chi-tsu-te*). This difference in relative statistical distribution may contribute to the user's implicit ability to determine the appropriate musical context when interpreting or rehearsing the *shouga.*

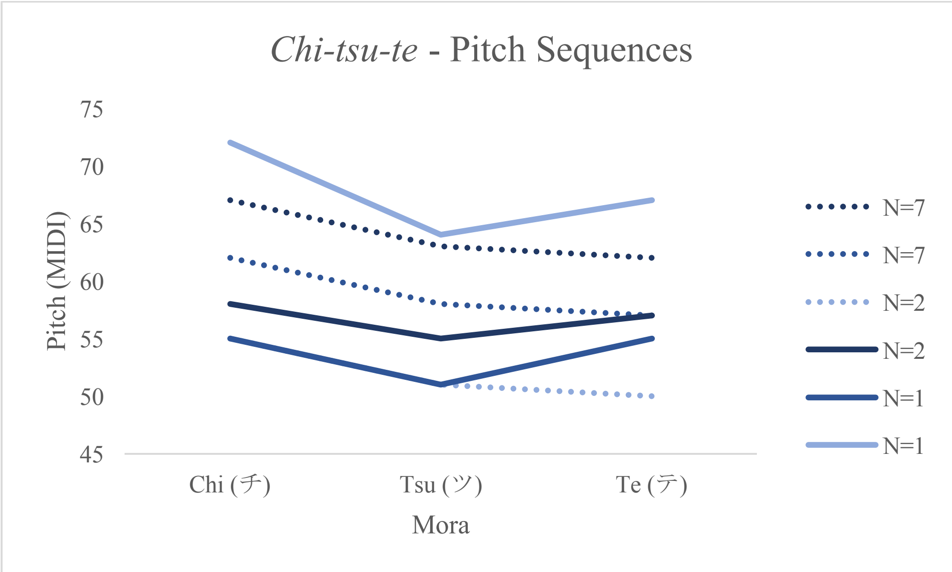

Fig. 3.9 *Chi-tsu-te* - Pitch sequences (N = 20)

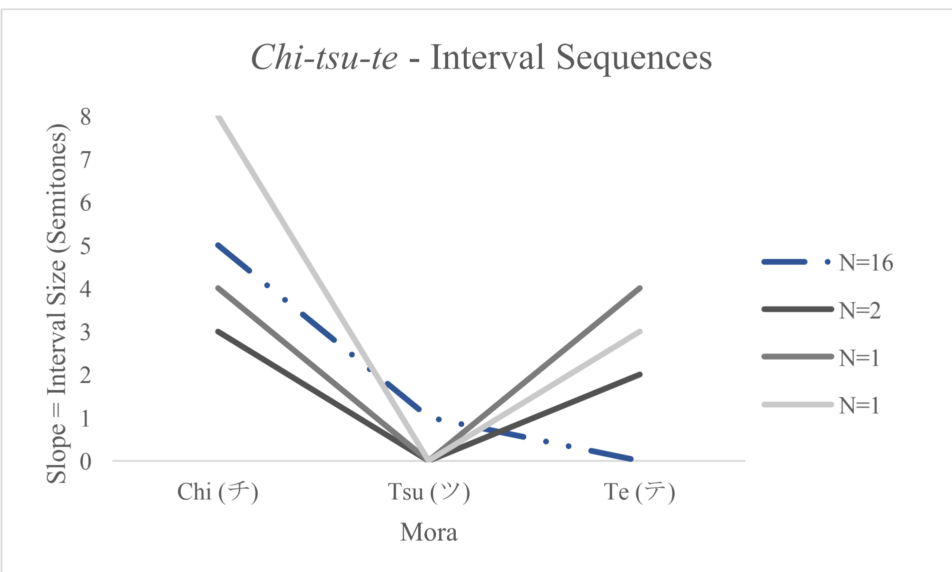

Fig. 3.10 *Chi-tsu-te* - Interval sequences (N = 20)

Further similarities and differences between *ko-ro-ri* and *chi-tsu-te* are found in their respective distributions among the scale degrees of *hirajoshi.* The two most common scale degree sequences for *chi-tsu-te* are $\hat{1}-\hat{6}-\hat{5}$ and $\hat{4}-\hat{2}-\hat{1}$ (75%), both of which were also found for *ko-ro-ri* (Figure 3.11). However, the sequences are found in vastly distinct relative frequencies between the two idioms: $\hat{4}-\hat{2}-\hat{1}$ is the most common scale degree sequence for *chi-tsu-te* (45%), compared to $\hat{1}-\hat{6}-\hat{5}$ (35%). In contrast, $\hat{4}-\hat{2}-\hat{1}$ was found only twice for *ko-ro-ri* (3%) compared to $\hat{1}-\hat{6}-\hat{5}$, being the second-most-common scale degree sequence for *ko-ro-ri* and making up 39% of all cases. Therefore *chi-tsu-te* is found most commonly on scale degrees that are found almost anomalously for *ko-ro-ri,* an idiom which shares a common characteristic interval sequence.

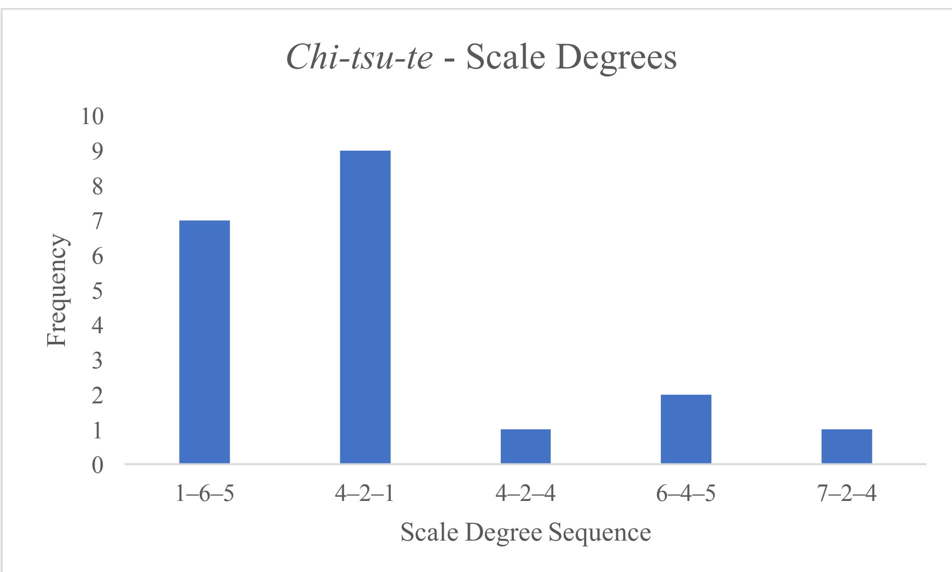

Fig. 3.11 *Chi-tsu-te* - Scale degrees (N = 20)

Figure 3.12 presents the most common rhythm for *chi-tsu-te* (75%)—quarter-eighth-eighth—with *chi* being held for twice as long as *tsu* or *te.* Again, therefore, we find a distinction between *ko-ro-ri* and *chi-tsu-te* on the basic of their rhythmic profiles. In Figure 3.13 we see the metric positions of *chi-tsu-te.* The most common positions are on beat 1 or beat 3. In contrast to *ko-ro-ri* (3%), 30% of all *chi-tsu-te* idioms are found on eighth-note off-beats.

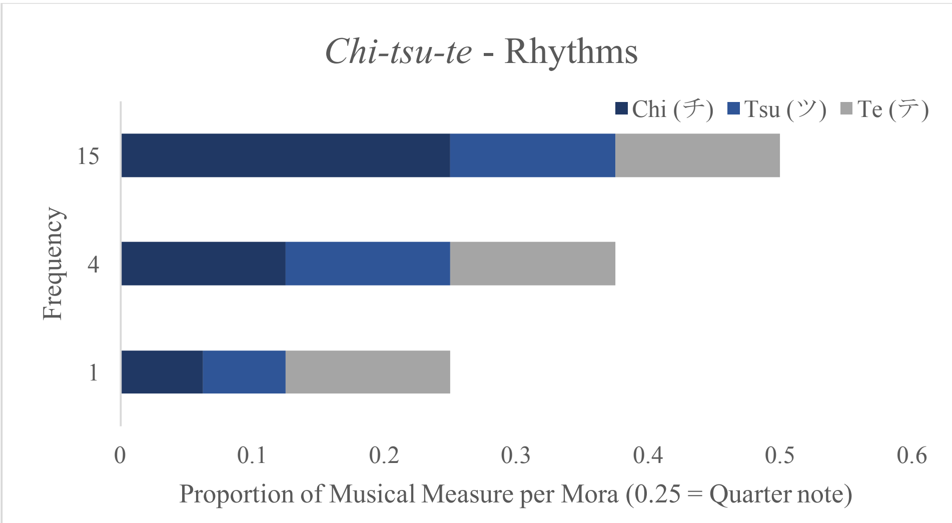

Fig. 3.12 *Chi-tsu-te* - Rhythms (N = 20)

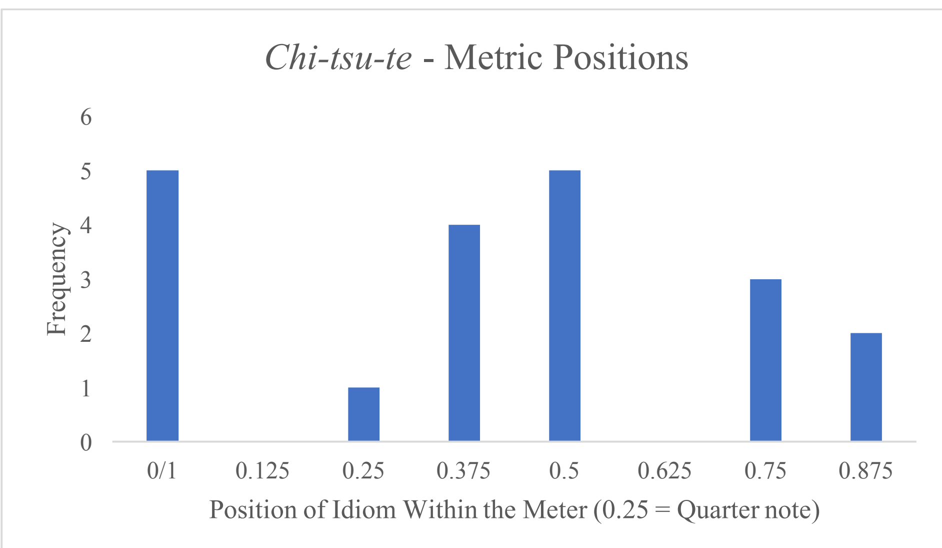

Fig. 3.13 *Chi-tsu-te* - Metric positions (N = 20)

Comparing the respective melodic parameter states of *ko-ro-ri* and *chi-tsu-te* one quickly recognizes the inherent complexity of attempting to describe the many *quantifiable* differences associated with the *qualitatively* distinct musical forms so readily identified by the *shouga* idioms. Nevertheless, the evidence of parametric distinctiveness and robustness among these idioms is impressive and clear.

Sa-ra-ri(n)

There are only two instances of the *shouga* idiom *sa-ra-ri(n)* in *Rokudan no Shirabe.* Despite the lack of statistical frequency within the corpus, this glissando motif is perceptually unmistakable within the piece upon listening. The parametric similarity between these two cases corroborates the evidence provided by direct subjective report.

Figures 3.14, 3.15, and 3.16 indicate that the same pitch, interval, and scale degree sequences are used for both instances of *sa-ra-ri(n).* Incidentally, this idiom is a good example of the descriptive limitations incumbent upon the decision not to include the vowel-only morae in the current analytical scheme. The *glissando* characteristic of this idiom is only fully articulated when considering the descending pitches indicated by a series of *a'*s between the three basic pitch structure morae. Nevertheless, we discover the interesting fact that *Sa* ($\hat{5}$) and *ra* are separated from each other by one descending octave (compatible with the understanding that "*ra, ri, ru, re, ro* represent a repetition of the same note [chroma]" (Ando, 1986, p. 198), and *ra* and *ri* by a diminished fifth, ending on $\hat{2}$ (Eb). Note also the similarity of the melodic functions of *ri* in both *ko-ro-ri* and *sa-ra-ri*—in both cases the idiom ends on a point of high musical tension, being resolved by subsequent melodic progression. In the case of *sa-ra-ri,* the first instance in the third movement (*sandan*) is followed by *to* on scale degree $\hat{1}$ (D). The second instance, which serves to end the piece as the final motif of the sixth movement (*rokudan*), is followed by *shyan* (an octave dyad on $\hat{1}$). This $\hat{5}$ – $\hat{2}$ – $\hat{1}$ melodic motion is starkly reminiscent of a Western-style V- I cadence.

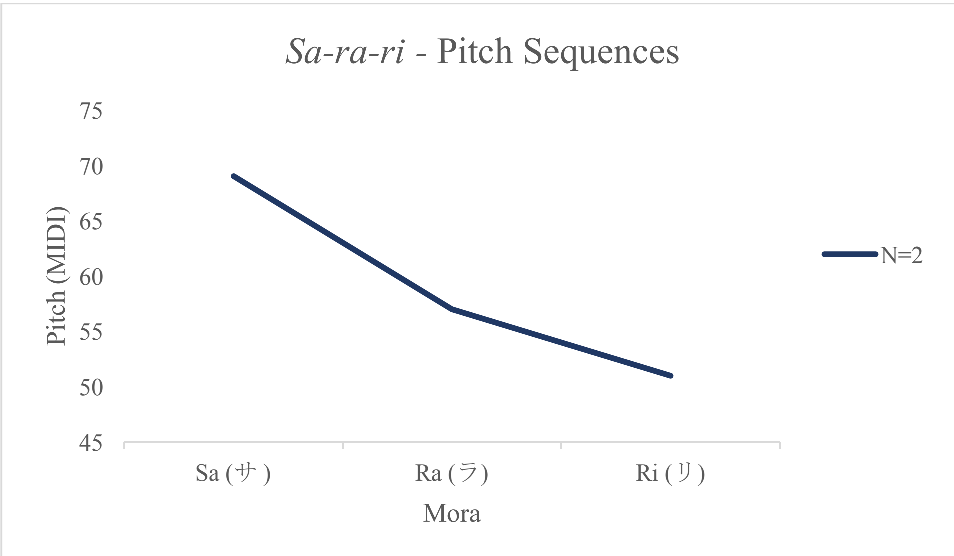

Fig. 3.14 *Sa-ra-ri* - Pitch sequences (N = 2)

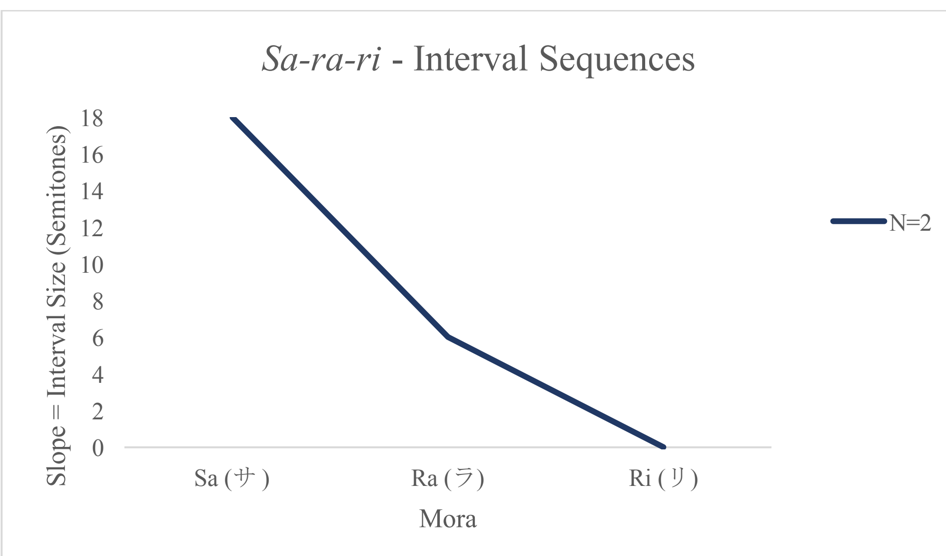

Fig. 3.15 *Sa-ra-ri* - Interval sequences (N = 2)

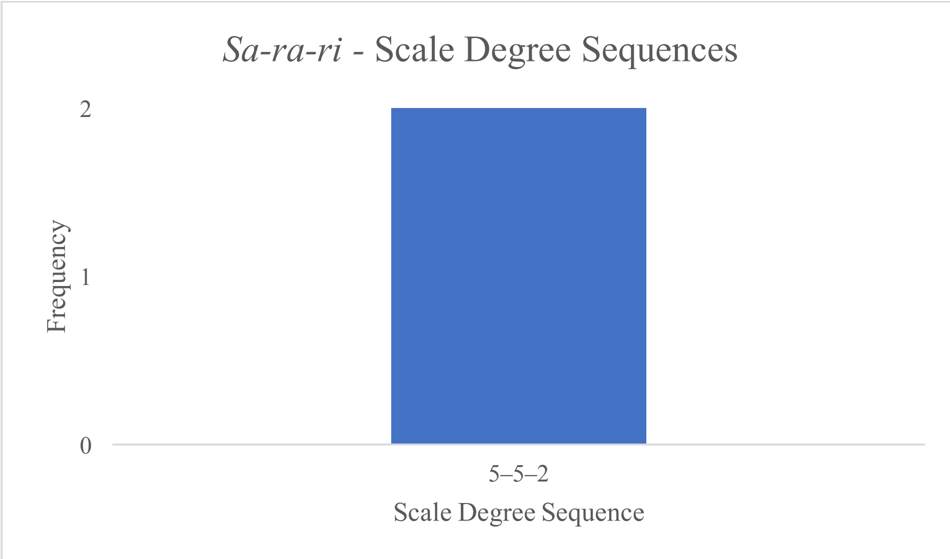

Fig. 3.16 *Sa-ra-ri* - Scale degrees (N = 2)

Sa-ra-ri starts on beat 2 in the third movement and on beat 3 in the sixth movement (Figure 3.17) and lasts for the same total duration of two beats in both cases. The rhythm is slightly different between the two, however: The length of *ra* is slightly truncated—consequently placing a stronger emphasis on *ri,* which is then held with a fermata before concluding with *shyan*—in the sixth and final movement (Figure 3.18).

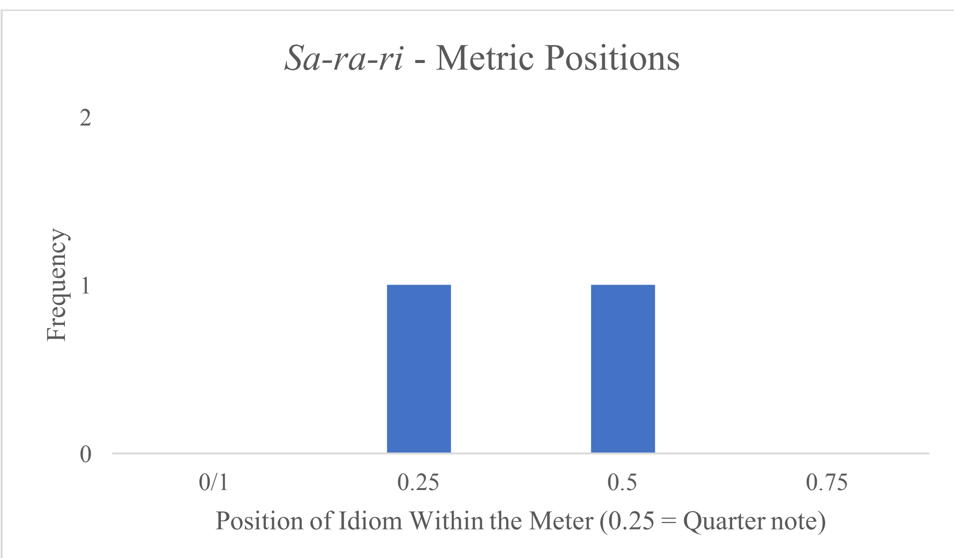

Fig. 3.17 *Sa-ra-ri* - Metric positions (N = 2)

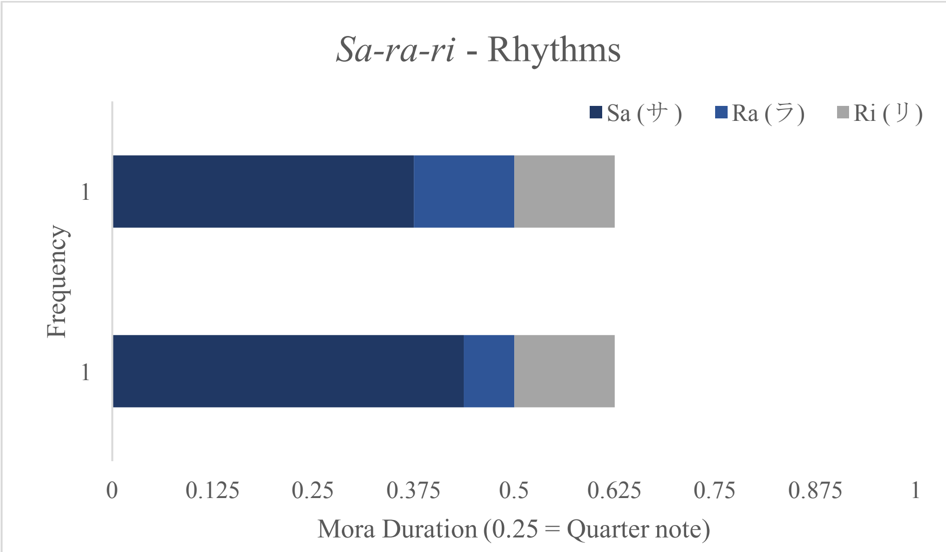

Fig. 3.18 *Sa-ra-ri* - Rhythms (N = 2)

Shya-shya

Figures 3.19 and 3.20 indicates that 97% of all *shya-shya* duples represent a repeated pitch pattern. The exception occurs in one of nine cases within the corpus where the first *shya* of the duple is in fact the end of one musical phrase, and the second *shya* is the beginning of the next. Consistent with Ando's (1986) report that the role of *shya-shya* is to indicate phrase boundaries, the overwhelming majority (87%) of *shya-shya* scale degree sequences were either on scale degree $\hat{1}$ ($\hat{1}$ – $\hat{1}$, N = 23) or $\hat{5}$ ($\hat{5}$ – $\hat{5}$, N = 10), with four of the five exceptional instances characterized by the familiar tonicizations previously encountered when analyzing *ko-ro-ri* ($\hat{2}$ – $\hat{2}$, N = 2; $\hat{4}$ – $\hat{4}$, N = 2) and the remaining case characterizing the just-described transition between the end of one musical phrase and the beginning of the next ($\hat{1}$ – $\hat{5}$) (see Figure 3.21).

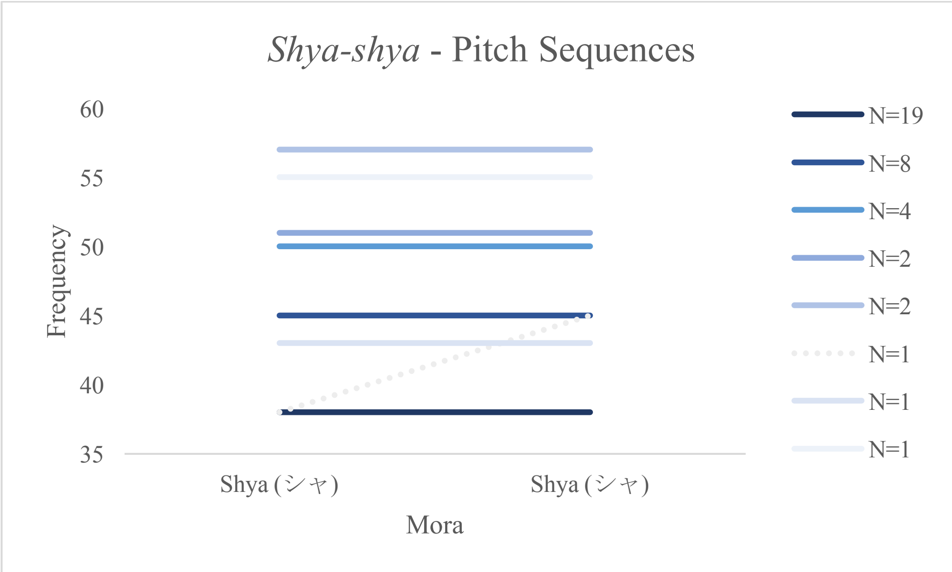

Fig. 3.19 *Shya-shya* - Pitch Sequences (N = 38)

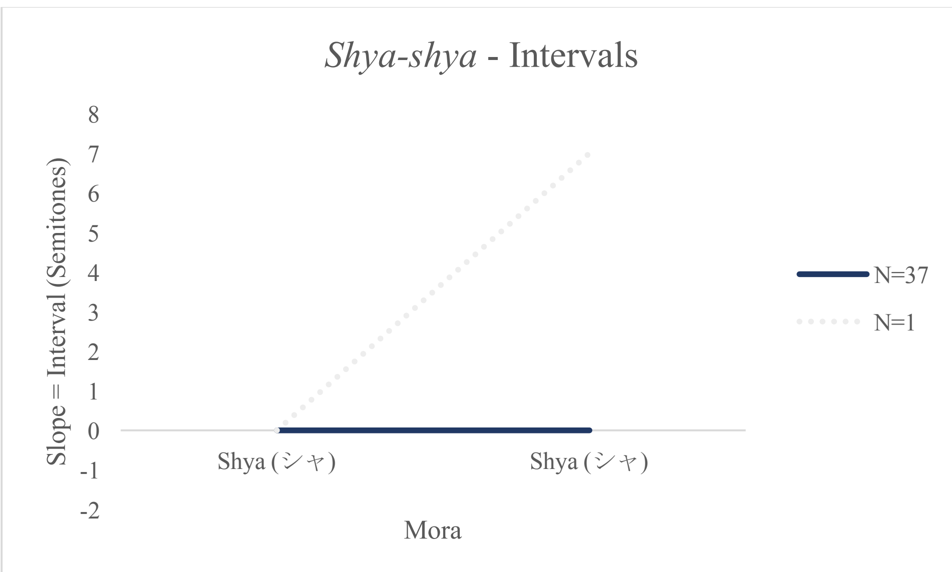

Fig. 3.20 *Shya-shya* – Intervals (N = 38)

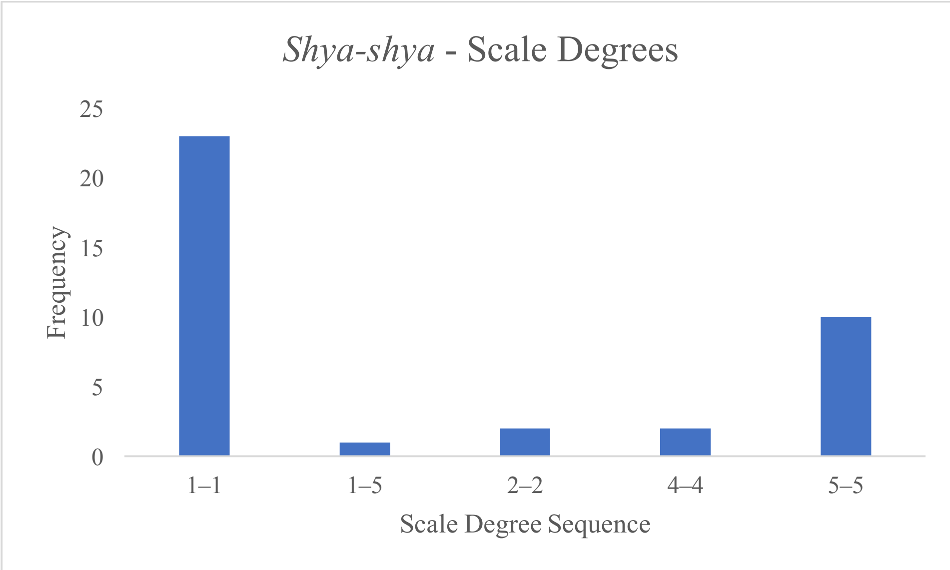

Fig. 3.21 *Shya-shya* - Scale degrees (N = 38)

Twenty-nine of the 38 *shya-shya* duple instances exhibit the prototypical eighth note-eighth note rhythm of *shya-shya,* while the remaining 9 are characterized by a quarter note-eighth note rhythm, coinciding with the 9 instances of *shya-shya* occurring across a melodic phrase boundary (Figure 3.22). Finally, *shya-shya* is only ever found on quarter note down-beats, although the relative frequencies among down-beats are not what one would expect from a schema for classical Western monophonic melodies: *shya-shya* occurs on beats 3 and 4 68% of the time, and more often on beat 3 than on beat 1 (which occurs 24% of the time; Figure 3.23). These frequencies are exemplary, however, of the *polyphasic* relationship between musical phrases and musical meters in traditional Japanese melodies. That is, one of the characteristic stylistic elements of traditional Japanese music such as *Rokudan no Shirabe,* is the "decoupling" of the varied-length melodic phrases from the equal-length metric boundaries, or the deviation of melodic-metric phase relations from simple integer ratios (Bonin, Matsunaga & McAdams, in preparation; see Chap. 2).

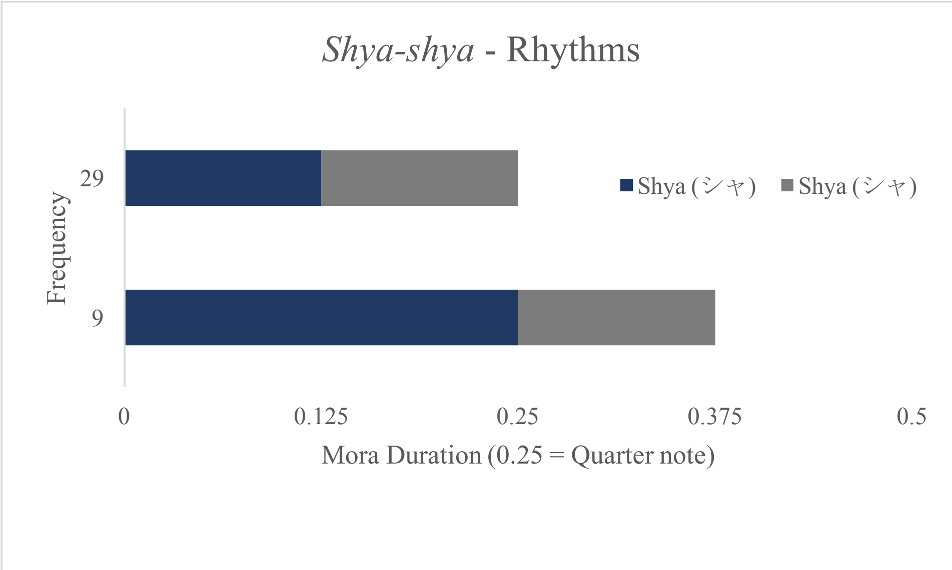

Fig. 3.22 *Shya-shya* - Rhythms (N = 38)

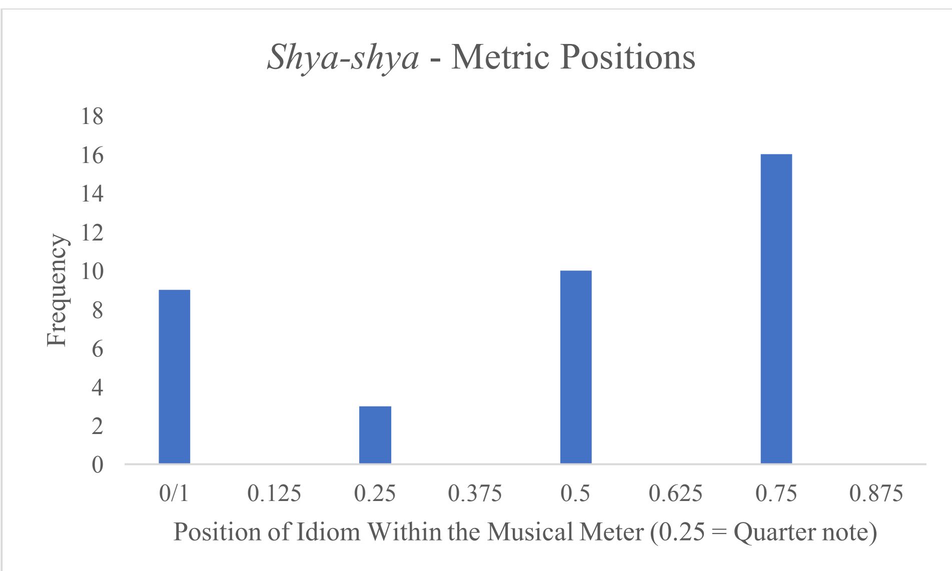

Fig. 3.23 *Shya-shya* - Metric Positions (N = 38)

To-te

Consistent with the insight of Mananoi and Akira (1978, p. 36), all 6 pitch sequence variants of the 40 *to-te* events prescribe movement from a lower to a higher octave (Figures 3.24

and 3.25). This motif is found exclusively on scale degrees $\hat{1}$ (N = 21), $\hat{5}$ (N = 10), and $\hat{4}$ (N = 9), in descending frequency among the most stable scale degrees (Figure 3.26). The exclusive mapping of *to-te* onto scale degrees $\hat{1}$, $\hat{5}$, and $\hat{4}$ is consistent with the qualitative understanding that *to* serves an "anchoring" function; a local tonic around which its respective phrase orbits.

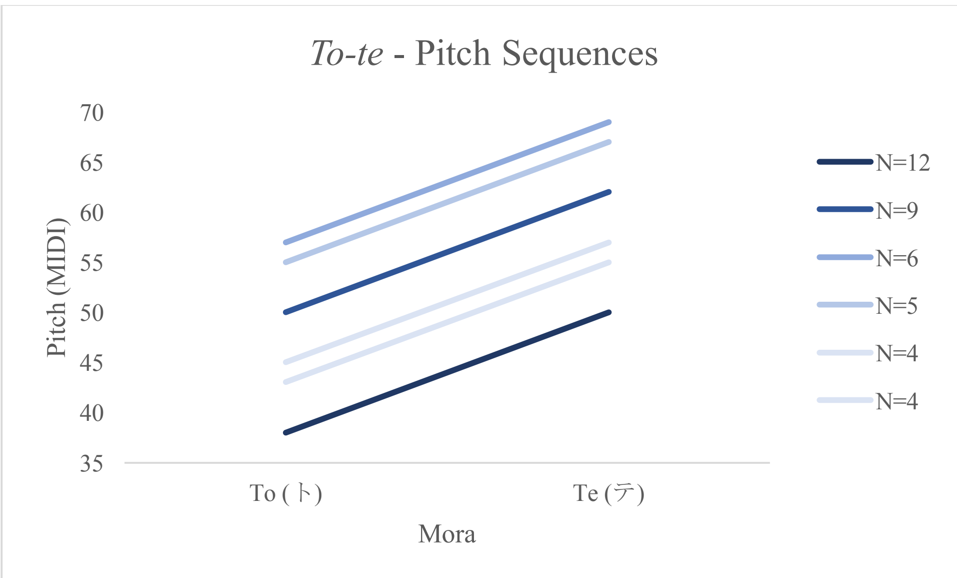

Fig. 3.24 *To-te* - Pitch sequences (N = 40)

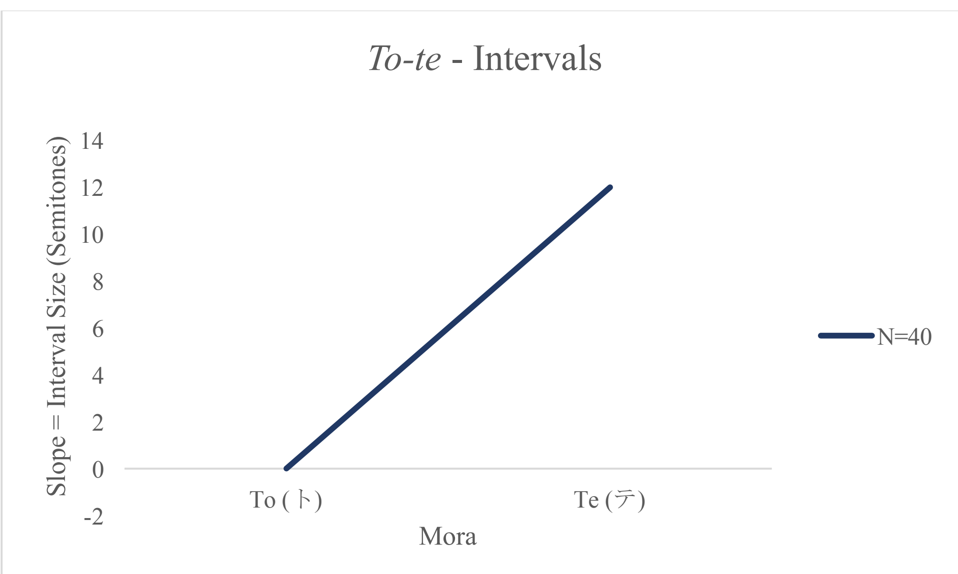

Fig. 3.25 *To-te* - Intervals (N = 40)

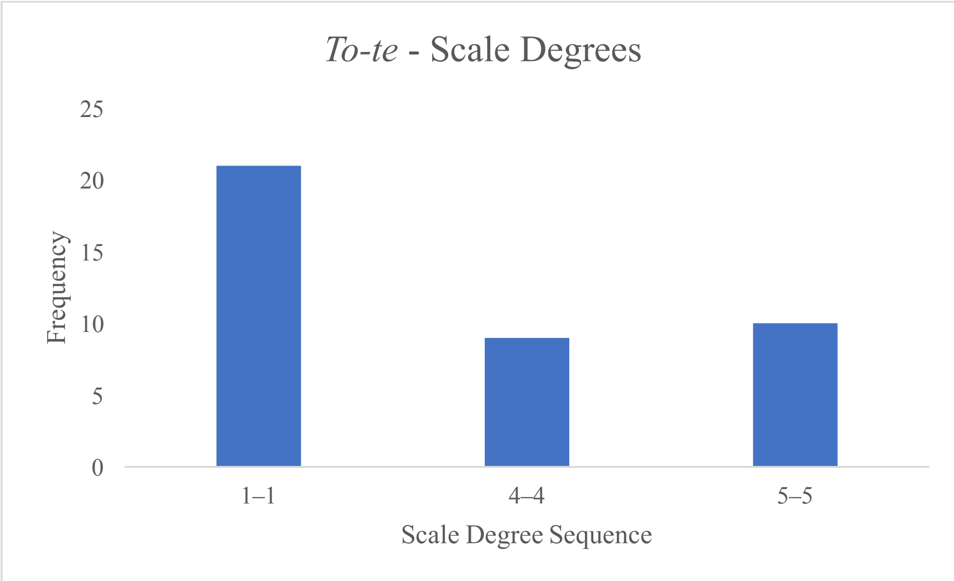

Fig. 3.26 *To-te* - Scale degrees (N = 40)

Finally, 37 cases (92.5%) of *to-te* occur with an equal-footed binary rhythm, comprising either two eighth notes (N = 31) or two quarter notes (N = 6). The three exceptions share a common rhythm in which *to* falls on an eighth note, and *te* is held twice as long on a following quarter note (Figure 3.27). Figure 3.28 demonstrates that *to-te* begins most often on quarter note downbeats (38/40), with beat one being the most common of these (N = 14). The two exceptions fall on the eighth note off-beats surrounding beat 3.

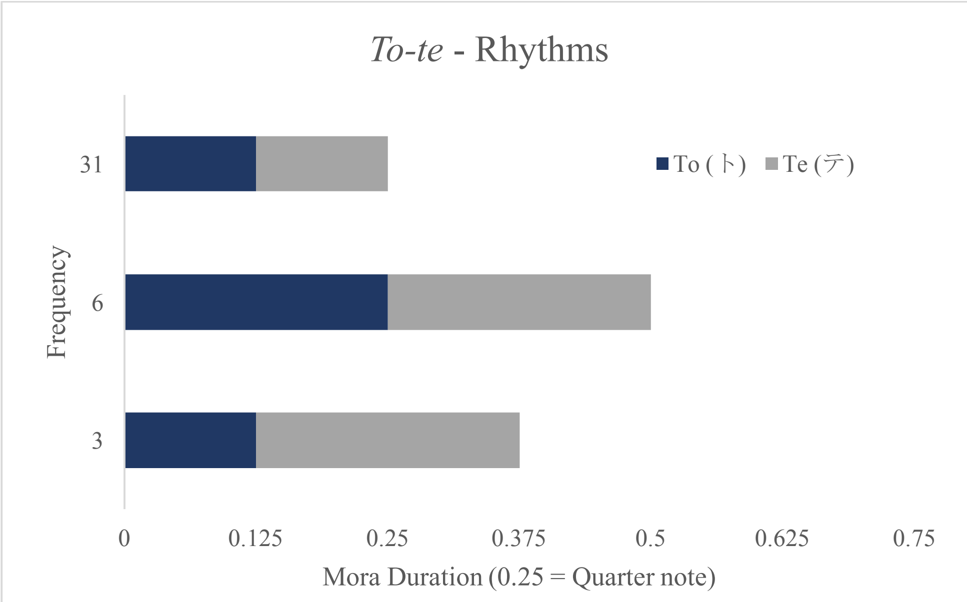

Fig. 3.27 *To-te* - Rhythms (N = 40)

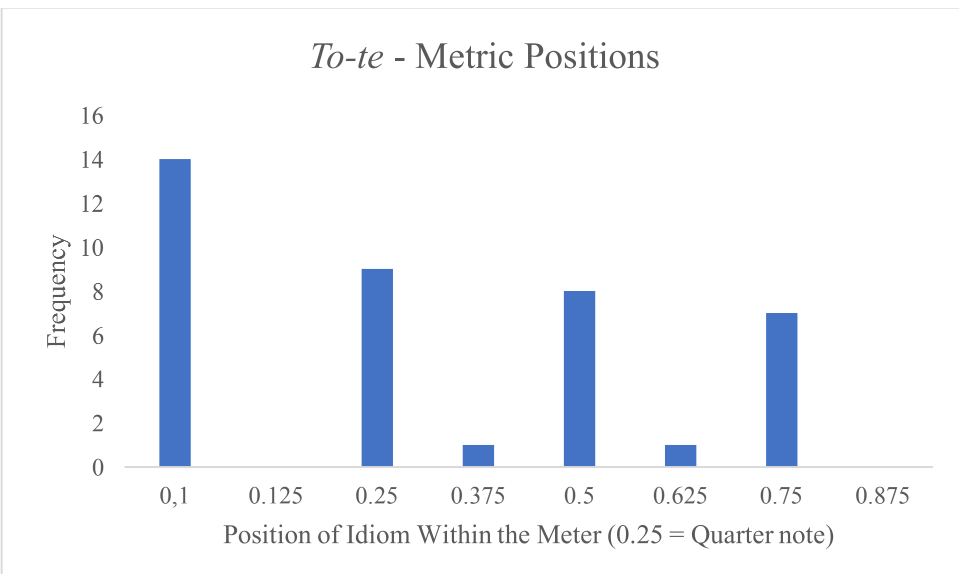

Fig. 3.28 *To-te* - Metric positions (N = 40)

Te-to

In contrast to the tonal simplicity of its retrograde counterpart *to-te,* the 37 instances of *te-to* are found across 16 different pitch sequences, 11 different scale degree sequences, and 10 different interval sequences (Figures 3.29, 3.31, and 3.30). A critical common denominator, however, is the single common descending contour [-] which characterizes all but one instance of *te-to.* This exceptional case rises in the melody from D4 to G4 ($\hat{1}$ – $\hat{4}$), and in the subsequent event rises again to G5, represented by *to-te.* As implied by the variability in interval sequences, only seven of these cases repeat at the descending octave ($\hat{1}$ – $\hat{1}$, N = 5; $\hat{4}$ – $\hat{4}$, N = 2). Thus, considered in its own right, *te-to* has the function of "descending," or "returning" to the local melodic anchor within the musical phrase. If we consider that *te-to* is most often paired with *to-te* in longer, more complex melodic sequences such as *to-te-to-to-te-to* or *to-te-to-te-te-to* that vary in melodic contour, and that *to* only moves to *te* by the octave, then *te-to* may be also understood in terms of a "pivoting" function, which allows the extended motifs to move through *chroma space* despite spanning multiple octaves. This latter interpretation is more closely aligned with the perceptual impression made by such sequences.

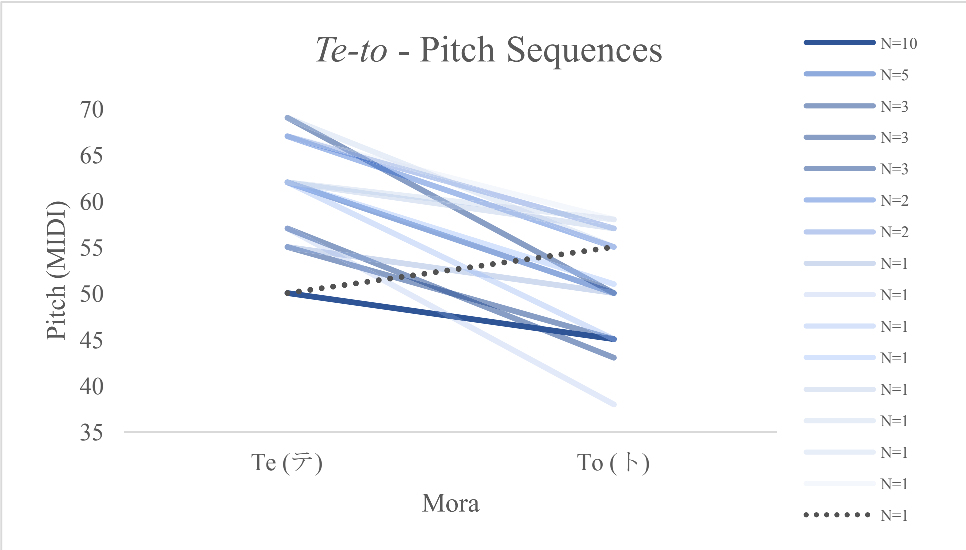

Fig. 3.29 *Te-to* - Pitch sequences (N = 37)

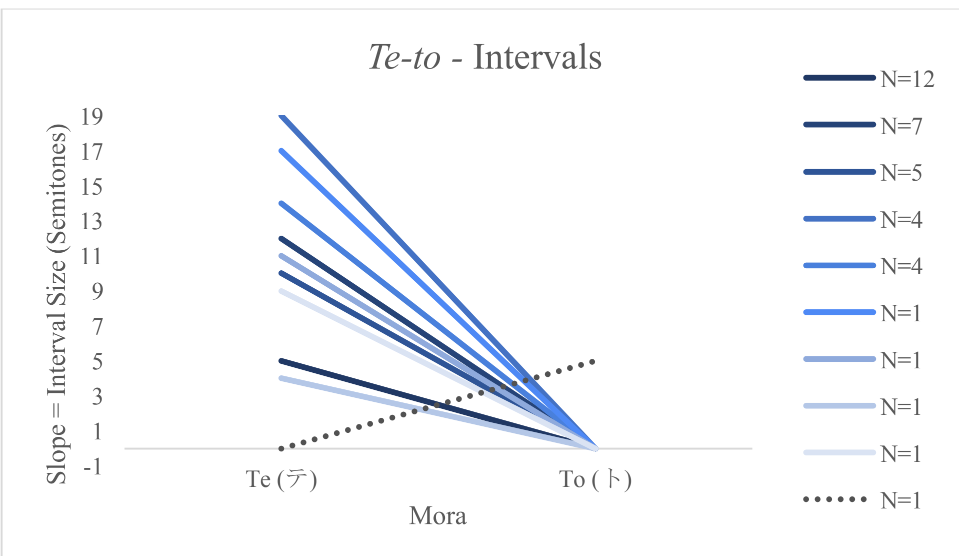

Fig. 3.30 *Te-to* - Intervals (N = 37)

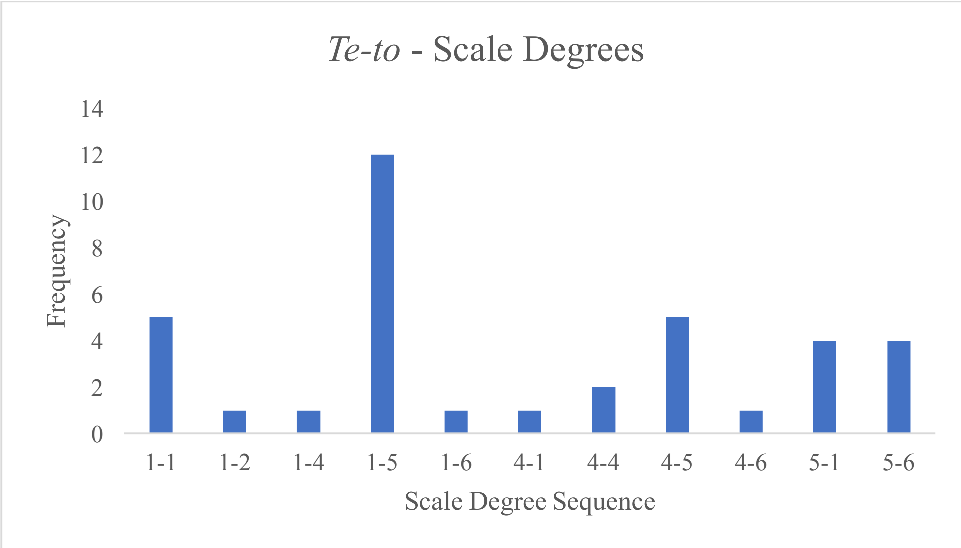

Fig. 3.31 *Te-to* - Scale degrees (N = 37)

A seemingly lone similarity between *te-to* and its retrograde counterpart is its rhythmic profile (Figure 3.32). Ninety-two percent of *te-to* sequences (compare with 92.5% of *to-te* sequences) are found with an equal-footed binary rhythm, either composed of two eighth notes (N = 28) or two quarter notes (N = 6). Three exceptions are comparable with the three total exceptions of *to-te: te* is held twice as long on a quarter note while *to* falls on an eighth note. The final exception maintains the same metric proportions: *te* is held for an eighth note, followed by *to* played for half the duration as a sixteenth note. *Te-to* is most commonly found on beats 2 and 4 (73%; Figure 3.33), which is what one would expect if *te-to* often follows and precedes *to-te* in more complex sequences, given that *to-te* begins most often on beat 1.

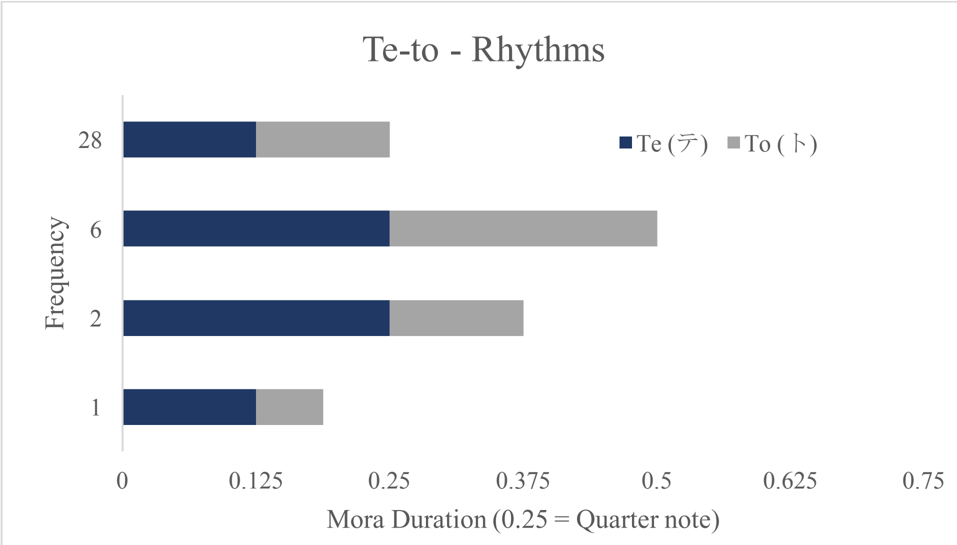

Fig. 3.32 *Te-to* - Rhythms (N = 37)

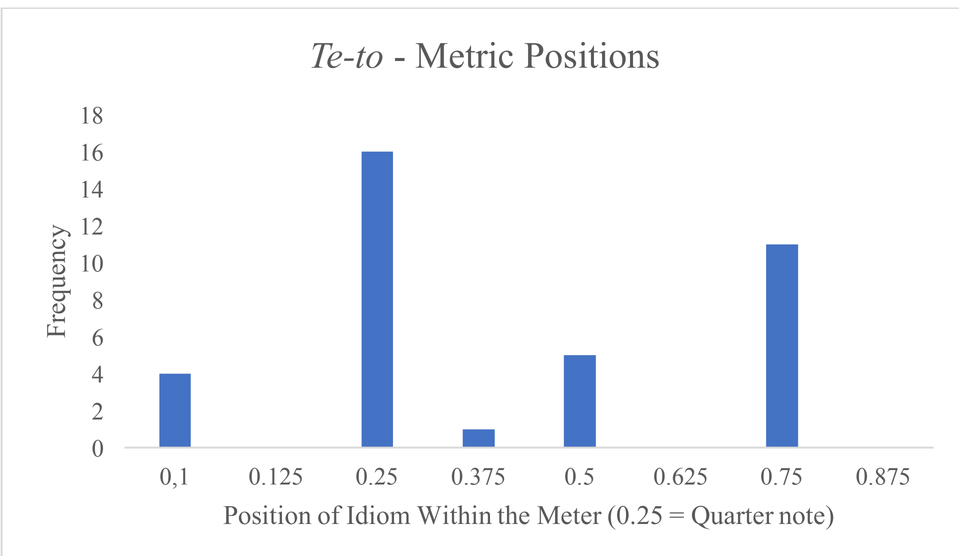

Fig. 3.33 *Te-to* - Metric positions (N = 37)

Te-tsu

Te-tsu is the first duple of the duple pair *te-tsu, tsu-te* which occurred most frequently in the present corpus analysis despite receiving no explicit theoretical treatment from previous performers or music researchers. The following parametric analyses may substantiate why a definitive statement about the relation between these *shouga* duples and the musical surface has not yet been formulated (and whether or not they may appropriately be considered *idioms* in their own right).

As Figures 3.34 and 3.35 indicate, *te-tsu* is found on 15 pitch sequences and 9 interval sequences that rise virtually as often as they fall: There are 8 falling pitch sequences, 5 falling interval sequences, 7 rising pitch sequences, and 4 rising interval sequences. Furthermore, *te-tsu* is found on 9 different scale degree sequences (Figure 3.36). But in fact, it is here where we catch a first glimpse of the role *te-tsu* might play: Across all 47 instances and 9 different scale degree sequences of *te-tsu, all* melodic motion begins with *te* on either the tonic, dominant, or subdominant scale degree, and transitions through *tsu* to a relatively less stable position within the melody. Recall also from Studies 1 and 2 that the longest sequences for which a coherent mapping between *shouga* and melodic contour were found when /*u*/ was mapped to an intermediate location between /a/ and /i/ on the upper end and /e/ and /o/ on the lower end (such as in, e.g., AIUEO). Finally, consider the sound made when pronouncing *tsu*— particularly in contrast to *chi, te, to,* and *ko,* by which it is most often preceded and followed: Although a timbral analysis of the respective mora sounds was not explicitly conducted in the current study, we can immediately recognize the characteristic fricative—a turbulent airflow—of *tsu* which is not produced by the most common adjacent mora. Taken together, these considerations imply that perhaps the role of *tsu* in describing the musical contour is not correctly conceptualized in terms of its ordinal rank within an existing vowel coding scheme, but instead as a *transitive* function, a "going across; Latin *transit*" within the melody, from more structurally stable tones to less stable tones. We will see in the following analysis of *tsu-te* that, indeed, in keeping with this interpretation, the sequential relation of relative scale degree stability is reversed between the two tones.

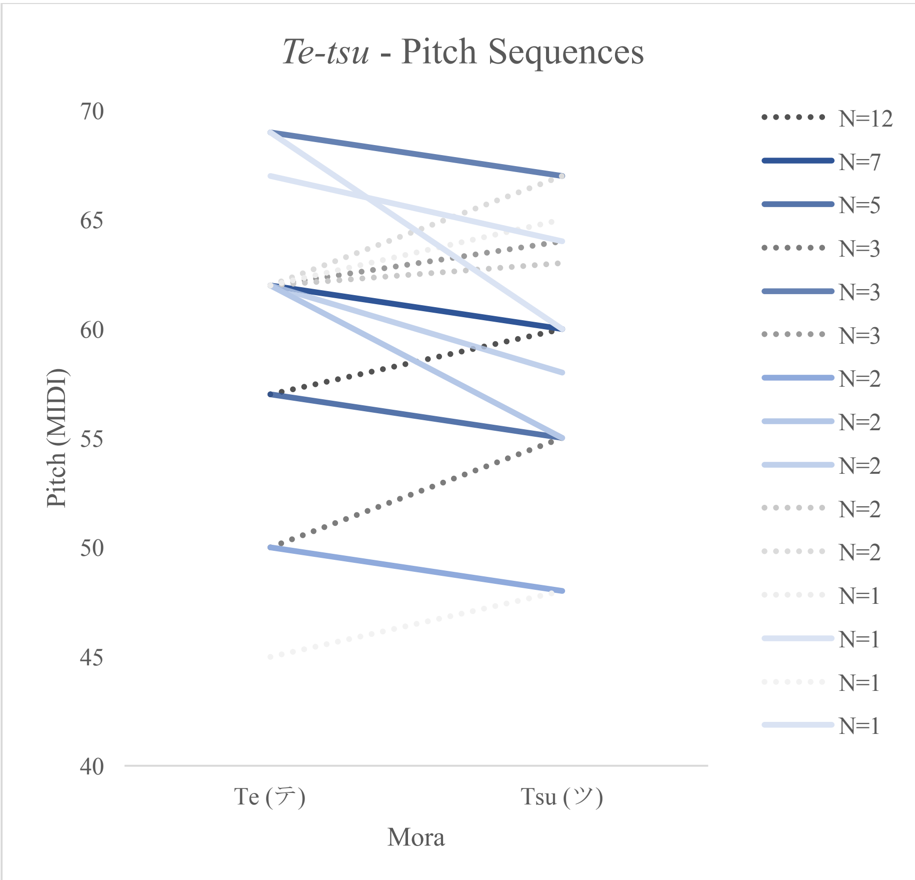

Fig. 3.34 *Te-tsu* - Pitch sequences (N = 47)

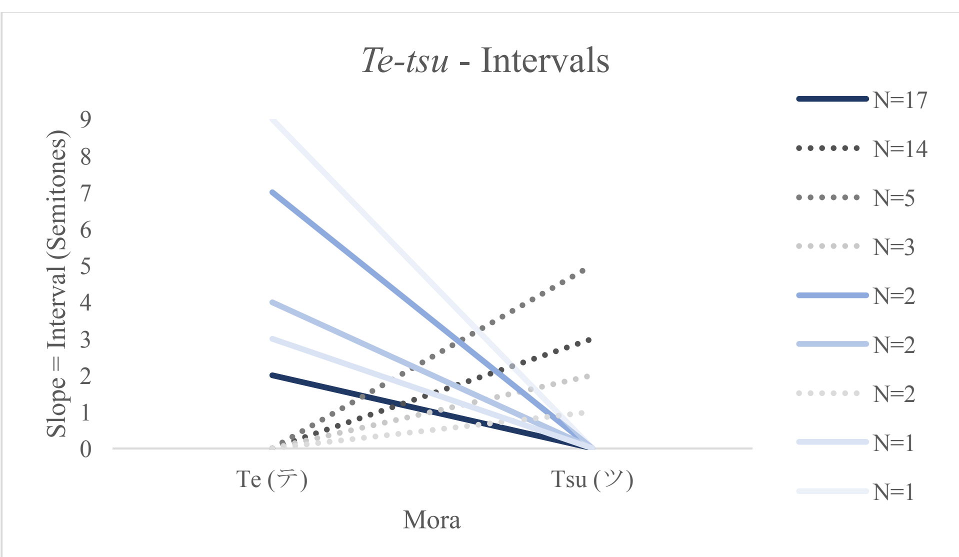

Fig. 3.35 *Te-tsu* - Intervals (N = 47)

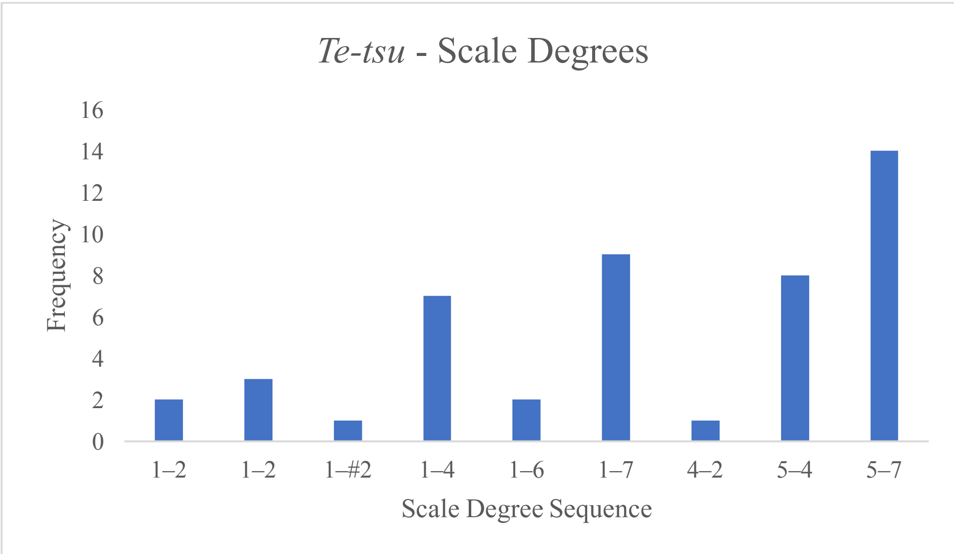

Fig. 3.36 *Te-tsu* - Scale Degrees (N = 47)

Te-tsu is also found sporadically throughout various positions of the musical meter (Figure 3.38). It is most commonly found on beats 1 (28%) and 3 (23%). The remaining 49% are strewn throughout quarter-note downbeats and eighth-note off-beats, although never on the eighth note occurring between beats 3 and 4; whether there is any structural significance to this lone omission escapes our current interpretation but may be addressed with larger corpus analyses in the future. Among all the variability of the other melodic parameters, *Te-tsu* does possess a clear rhythmic characteristic: more than 80% of all cases exhibit an equal-footed eighth note-eighth note rhythm (Figure 3.37).

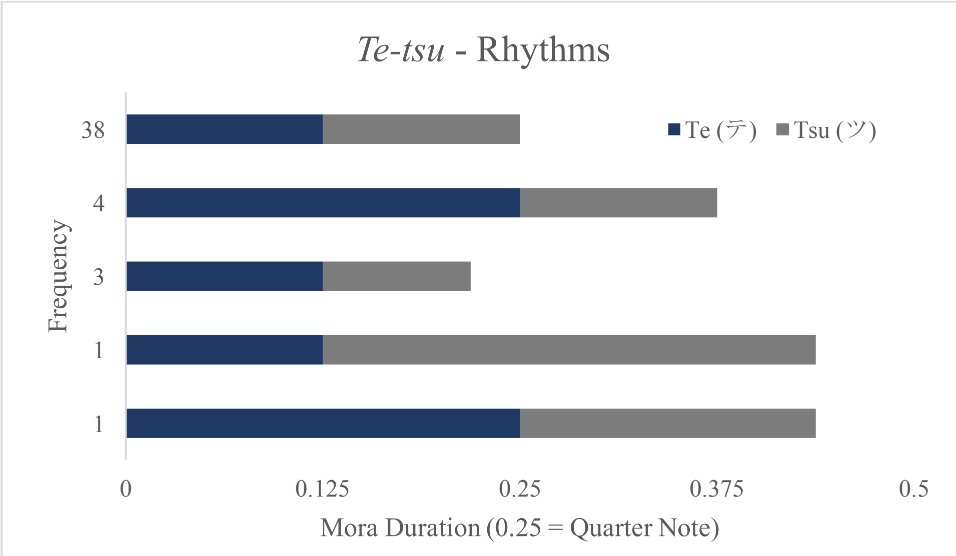

Fig. 3.37 *Te-tsu* - Rhythms (N = 47)

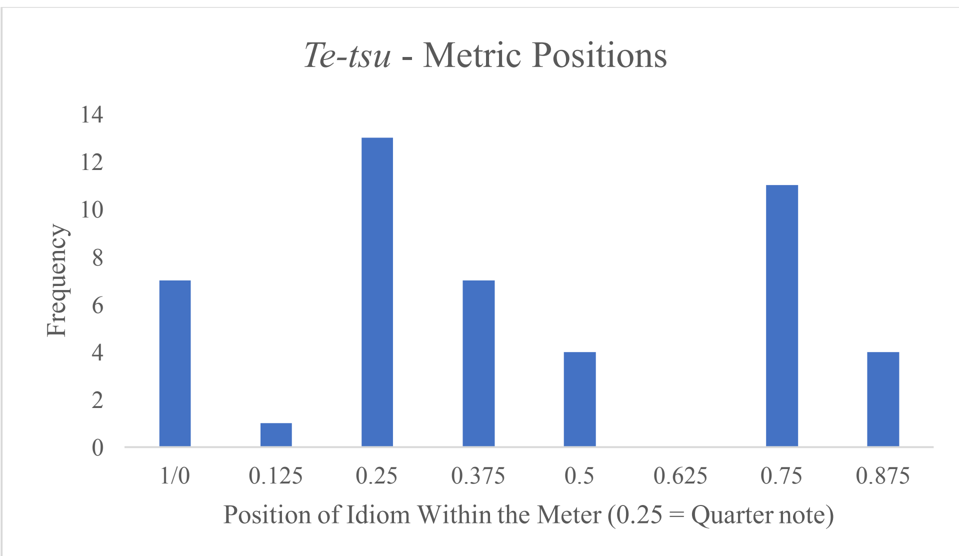

Fig. 3.38 *Te-tsu* - Metric positions (N = 47)

Tsu-te

Similar to its *te-tsu* retrograde counterpart, *tsu-te* (N = 47) exhibits a considerable variability in the number of distinct pitch sequences (7 rising and 3 falling; Figure 3.39). Interestingly, however, these pitch sequences converge onto a considerably reduced set of intervals (Figure 3.40): There are three rising contours, the most common rises 2 semitones and accounts for 26 of the 47 cases (55%). There is only one descending interval, which falls by a semitone and accounts, as the second-most-prevalent interval in the set, for 34% of all cases (N = 16). As alluded to in the previous analysis of *te-tsu,* we find that in every one of the 47 cases of *tsu-te,* the melodic motion is from a relatively less stable scale degree to a relatively more stable scale degree and one of the three pillar tones $\hat{1}$, $\hat{4}$, or $\hat{5}$ (Figure 3.41). This finding again corroborates the suggested *transitive,* "going across" function of *tsu,* especially when considered in conjunction with the notably smaller magnitude and variability of the intervals passing from *tsu* to *te* as compared to the motion from *te* to *tsu*: The notion that a passing tone should arrive at a nearby chordal tone is strongly reminiscent of the Western voice-leading principle that emphasizes motion toward the "nearest chordal tone."

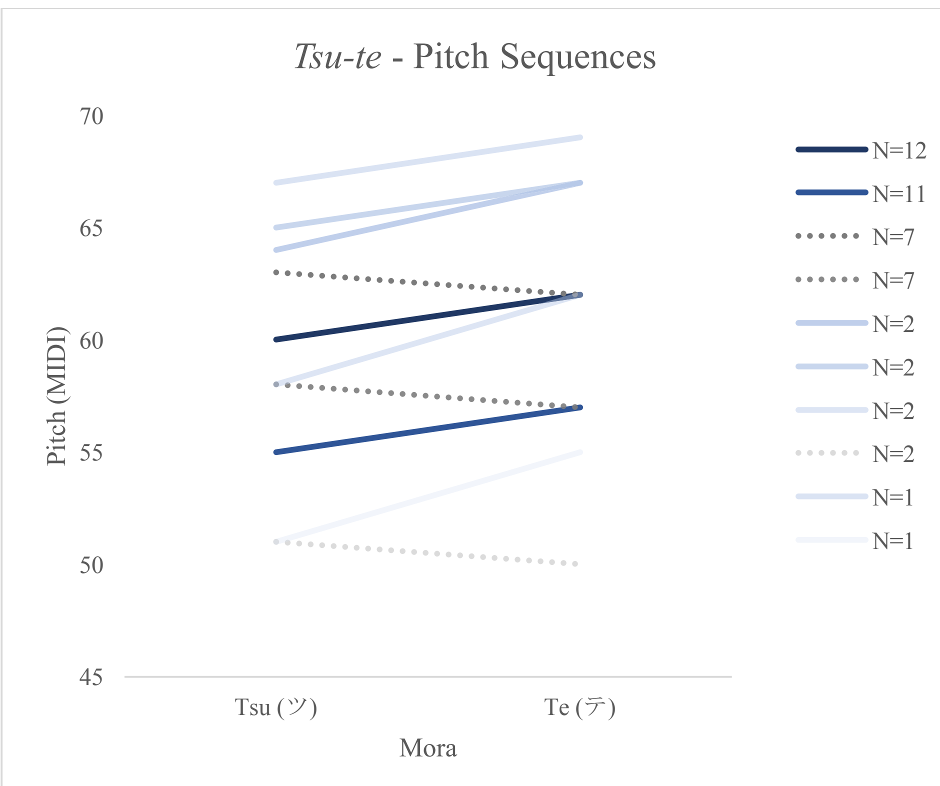

Fig. 3.39 *Tsu-te* - Pitch Sequences (N = 47)

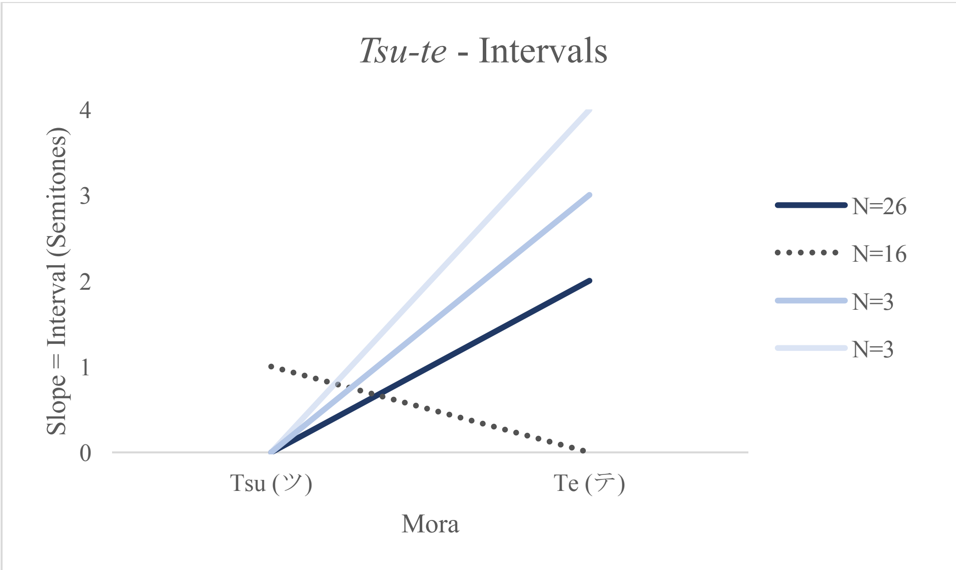

Fig. 3.40 *Tsu-te* - Intervals (N = 47)

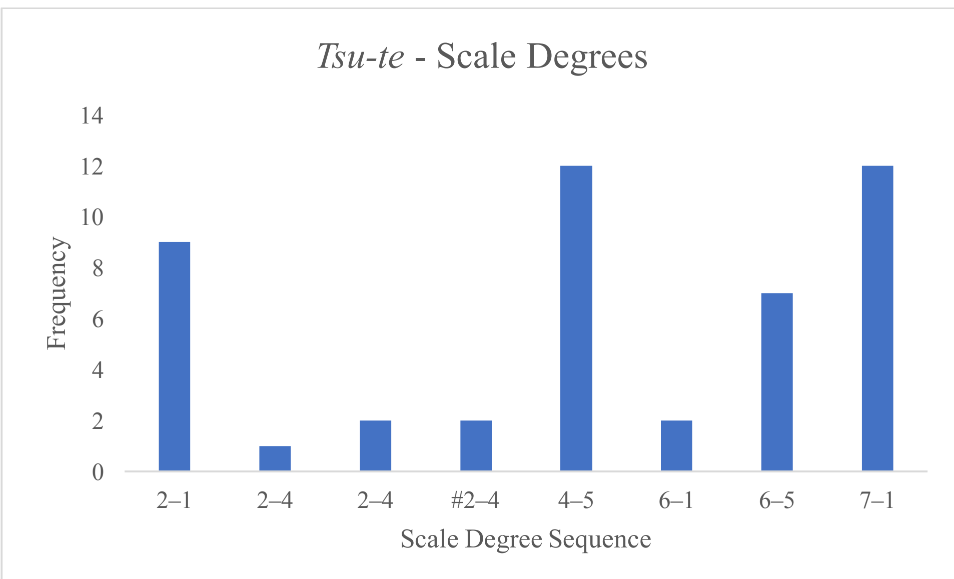

Fig. 3.41 *Tsu-te* - Scale degrees (N = 47)

Finally, *tsu-te* is comparable to *te-tsu* in terms of its relatively definite rhythmic profile (>80% of all cases possessing an eighth note-eighth note rhythm; Figure 3.42) and its sporadic placement within the musical meter (Figure 3.43), although *tsu-te* is found more consistently and exclusively on quarter note downbeats (83%, compared to *te-tsu,* 73%).

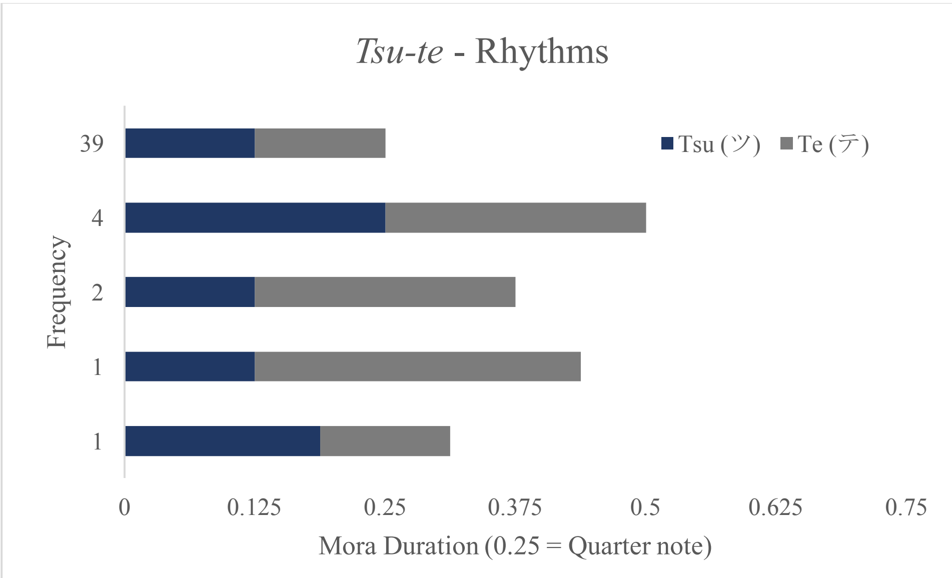

Fig. 3.42 *Tsu-te* - Rhythms (N = 47)

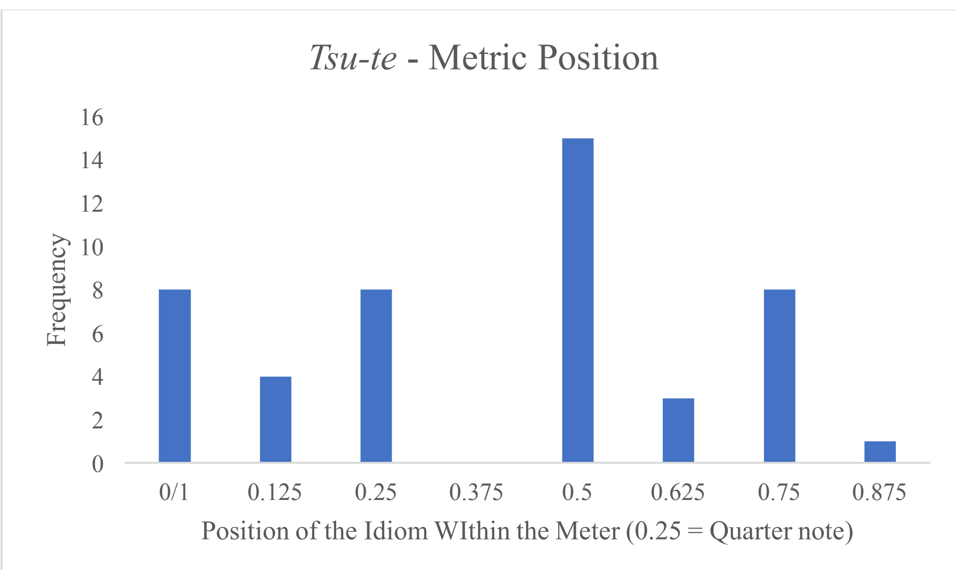

Fig. 3.43 *Tsu-te* - Metric positions (N = 47)

Discussion

No two idioms submitted to the parametric analyses of Study 3, were found to be alike. In some circumstances, two idioms may be observed to "co-inhabit" the same tonal, rhythmic, or metric states. But even in these cases, there was always a salient determining factor to distinguish the two idioms: [95] In the case of *ko-ro-ri* and *chi-tsu-te,* for example, which shared a common interval sequence [-4, -1], the relative frequencies by which they associated with this sequence was distinct; while [-4, -1] was the most typical interval progression of *chi-tsu-te* it was only of secondary prevalence in the case of *ko-ro-ri,* whose most common interval [-2, -4] is clearly distinct. Although *to-te* and *shya-shya* were both characterized through the repetition of stable scale degrees $\hat{1}$, $\hat{4}$, and $\hat{5}$, *to-te* was always associated with upward octave motion, while *shya-shya* repeated the same pitch at unison. *Tsu-te* and *te-tsu* were both associated with a vast and seemingly unintelligible count of unique scale degrees, and yet *tsu-te* always moved from a less stable scale degree to a pillar tone $\hat{1}$, $\hat{4}$, or $\hat{5}$, while *te-tsu* always moved in the opposite direction, from the more stable $\hat{1}$, $\hat{4}$, or $\hat{5}$ to a less stable tone. Further, these salient discriminating factors were all perceptually relevant, consistent with the idea that the *shouga* is used to articulate, or re-*present*, the musical form perceived by the subject when listening to the original melody. Future computational research may seek to incorporate data markers which identify the positions of these and other idioms throughout the various musical parameters of interest. Such tools would serve to make the statistical analysis of musical objects more reminiscent of the perceptual process undertaken by human listeners.

---

[95] This is not to mention the stark distinctions between the two idioms found along the other melodic dimensions

General Discussion

The *shouga* operates on the basis of its perceived similarity with the music it represents. *Shouga* sequences are constructed such that the perceptual images they elicit from the listener are formally analogous to the perceptual images elicited by corresponding musical phrases. Thus, we may say that the *shouga* is an objective expression of the subjective impression of musical form. To attempt to understand the relationship between *shouga* and music is therefore to acknowledge an essential psychological factor in the production and perception of music.

The musical composition is a result of the human creative process. In the creative process, the subjective experience of ideas, sensations, and emotions (*psychological entities*), for example, are translated into objective material analogies. Creative activity necessitates the recognition that the internal experience of a human being does not exist as a pure void (*nihilo*), formless morass, or incessant flux of nonsensical half-thoughts. Neither can the psyche be regarded as a mere repository of superficial or imperfect representations of physical objects first encountered in the outside world.[96] To the contrary, the experiential salience and empirical integrity of subjective states *per se* very often serves as the initial motivation to create. In the case of creative *expertise*—in which the creator is reliably and effectively translating their subjective insights into an analogous material object—the resultant art object serves as an empirical index of the internal state of its creator (Condo, 2017). Such objects (as in the case of, e.g., the *shouga* of the *Ikuta-ryu*) corroborate the psychological insight that the internal states of the human subject can be rationally parameterized and therefore scientifically investigated.

[96] This is not to say that the composer's prior experiences of the outside world do not inspire or influence the composition and complexion of the newly created art object.

In the contemporary perceptual sciences, the virtually ubiquitous definition of percepts and mental images is that they are *representations* of the external state of the object.[97] Although this definition is undoubtedly appropriate in many cases, such as basic psychophysical research, it precludes music perception research from addressing some of the most fundamental questions of musical experience, because it is only partially true: The mental images of the subject's consciousness are representative of objective states, but they are also *generative* of external objective states. This latter function is central to the creative process. During the creative process, it is the external object which is the representation of the internal state of the subject. Indeed, much of the artist's work is to cultivate the technical skills necessary to be able to *craft* the external medium of their expression in such a way that exemplifies the desired result which first presents itself "from the inside," (Neumann, 1959). The task of the artist is to translate the *idea* into a tangible manifestation within the sensible world. This process necessitates that there be an internal logic, or *ratio,* among the perceptual images, or internal mental states, experienced by the subject; if there weren't, the artist could not find anything to improve about their work. But it is precisely because of this subjectively experienced ratio among mental states that the human subject can meaningfully speak of *forms* common to distinct objective structures. As Wittgenstein (1921; 2.033) stated so simply and directly, "Form is the possibility of structure." It is this perceptual capacity to detect formal isomorphism that underlies the representational efficacy of the *shouga* as a musical transliteration device.

Typical solmization techniques, such as solfège, are used to refer to the parameter states of the art object itself. Thus, they are always at least one degree removed from the mental states that produced the original art object. Telling a student that the desired sequence is "do re *fa,* not

[97] The use of the term "image" is indicative of this ethos.

do re mi," does not communicate anything at all about the formal consequences of mistaking *mi* for *fa,* though it does inform the student how to modify their objective performance. In contrast, instructing the student of *shouga,* "no no, it's not *chi tsu* te, it's *chi tsu* to," conveys something of the formal distinction which results in the mind of the listener from these two object manipulations. Because the forms to which the *shouga* refers can be associated with various objectively unrelated structural states, the *shouga* has been unduly criticized as "nonsense," (see Hughes, 2000, 1991, 1989). While it is certainly true that the *shouga*'s preoccupation with the internal experience of the listener makes its objective structure elusive to statistical and empirical analysis, this is no reason to presume that the *shouga* lacks coherence or intelligibility. The studies of the present work were directed toward the statistical and empirical demonstration of this intelligibility.

David Hughes (2000) has conducted pioneering work to this same end. In Study 1 we replicated and extended the results of Hughes (2000), who found that the mora sequences of the *shouga* correspond with the contour sequences of the melody according to the rising and falling F2 values of the Japanese values used within the *shouga.* In addition to the strict F2 ordering of the Japanese values as IEAUO, we also included the acoustically informed variant IEAOU previously noted by Mananoi and Akira (1978) and Hughes (2000), and the culturally conditioned sequence found in the Japanese kana syllabary columns AIUEO. We referred to these variants all together as *vowel coding schemes.*

We found considerable evidence of structural correspondence between the *shouga* of *Rokudan no Shirabe* and the vowel coding schemes IEAOU and IEAUO, and to a lesser degree AIUEO. All three sequences predicted coherent melodic sequences up to 15 events long. The most predictive vowel coding scheme, IEAOU, explained an impressive 56.6% of all vowel sequence-contour pairs, although this result was not nearly as conclusive as the rates of >90% reported by Hughes (2000). We addressed the remaining ~50% of unexplained variance by suggesting that the vowel coding schemes do not take into account the consonant components of the *shouga,* nor can they account for the cross-modal associations of musical form sourced from, for example, proprioceptive feedback of the playing techniques, muscular tension in the vocal tract used to produce the *shouga* sounds (whether in audible or sub-vocal speech, see McGuire et al., 1996).

To address the possibility of implicit influence from these cross-modal domains resulting from the creative process (see also Giolai, 2019), we conducted a series of exploratory statistical analyses in Study 2. In the first half of Study 2 we found clear evidence of non-arbitrary frequencies characterizing the use and prevalence of different Japanese speech sounds within the *shouga* (analysis I.1)*,* as well as the number of repeated mora sequences throughout the *shouga* (analysis I.2). The probability of these structures occurring by random chance is negligible, providing further evidence of the implicit coherence of the *shouga.*

In the second half of Study 2, we conducted two exploratory analyses to evaluate the claim that the *shouga* is primarily concerned with common musical *forms,* rather than explicit literal structures or specific objective parameter states of the musical art object. In the first analysis (II.1), we reasoned that if the *shouga* sequences are not randomly scattered throughout the melodic pitch space but rather systematically transposed according to common formal motifs, then the number of unique melodic pitch sequences associated with a given *shouga* sequence should be greater than the number of unique melodic interval sequences, as distinct pitch sequences are perceived to be more common if they share a common interval sequence. If a given *shouga* sequence represents a particular musical form, then the interval sequences with which it is associated should effectively consolidate the number of less-abstract pitch sequences with which it is associated. If instead the *shouga* sequences are randomly distributed throughout the musical pitch space, then the number of pitch and interval sequences associated with a given *shouga* sequence should be the virtually identical, where any two pitch sequences would not be systematically associated by a common interval sequence. To further refine this measurement, we also included a further-abstracted *relative pitch* parameter, which maintained the ordinal ranks of the interval magnitudes throughout a given sequence to capture common melodic "gestures" which, despite appearing in the musical object with various literal interval states, nevertheless convey common musical form trajectories or morphologies. We made the same predictions with respect to the relation between the relative frequencies of unique relative interval sequences and interval sequences as with the relative frequencies of unique interval sequences and unique pitch sequences: namely, that if *shouga* sequences are intended to convey common musical forms, then the *relative interval* parameter should consolidate the slight variations in literal interval sequences under a common *relative interval* sequence, while leaving clearly distinct interval sequences (with different contours) untouched. Finally, we reasoned that by including interval magnitude ranks in the operational definition of the relative interval parameter we should be able to effectively distinguish between two interval sequences that share the same *contour* but possess clearly different slopes, which we consider more closely aligned with the artistic sentiment of musical "contour." The results of this analysis were consistent with our hypotheses and the claim that the *shouga* is a coherent representation of the listener's perception of musical form.

In the second analysis of part two of Study 2 we investigated this claim further by analyzing the average interval size and direction associated with each of the 56 unique mora pair sequences and each of the 25 unique vowel pair sequences found in the *shouga* of *Rokudan no Shirabe* (analysis II.2). We found remarkably constrained interval sizes associated with each of the unique mora pairs (the vast majority characterized by a standard deviation less than one semitone in magnitude), and furthermore that 52 out of 56 (92.9%) of all mora pairs were most often associated with only one melodic contour sequence. We found that, contrary to the predictions of each of the vowel coding schemes, every repeated vowel sequence (e.g., o-o, i-i) was associated at least once with non-flat melodic contours, contrary to the predictions of the vowel coding schemes.

Taken together, the results of the second half of Study 2 demonstrate the importance of *interval* and *relative interval* in parameterizing the relationship between *shouga* and *melody*. Specifically, we saw that the *relative interval* of melodic sequences was more closely associated with unique *shouga* idioms than was the *contour* of these melodies, as previously investigated by Hughes (2000). Furthermore, we saw in the final analysis of Study 2 that *shouga* idioms may be meaningfully distinguished from one another in terms of the respective interval ranges found in the corresponding melodic motifs. That is, *shouga* idioms are characterized by a constrained range of interval magnitudes, such that, for example, *shya-shya* and *ko-ro-ri* are distinguished from one another not only in terms of their specific pitch or rhythmic content, but also by the magnitude and direction of melodic movement associated with each idiom. Further psychological studies should investigate the interval and relative interval thresholds associated with melodic motifs perceived to exhibit a common musical form (cf., e.g., Eiting, 1984). Such research would also be of primary interest to music information retrieval researchers, who have already conducted extensive research on this and similar problems (e.g., Sears, 2008; van Kranenburg, Volk & Wiering, 2013; Rao et al., 2014). Finally, extending the previous work emphasizing the relation between the vowels and the *shouga* and the contours of the musical melody, we find that the consonant components of the mora are also systematically associated with various melodic parameter states.

In Study 3 we conducted an in-depth analysis of the melodic *pitch, interval, scale degree, rhythm,* and *meter* parameters associated with robust *shouga* idioms identified by expert practitioners (Mananoi & Akira, 1978; Ando, 1986). We found the parametric distribution of each of the idioms to be clearly distinguished from one another, although the ways in which these parameters were distinguished was highly dynamic and contingent on case-by-case comparisons. In some cases, two idioms may be observed to correspond with common states on one or two parameters, but even here the relative distribution of the idioms across these states were varied, and the differentiation afforded by this variation was further bolstered by the idioms' distinctiveness on the other musical parameters. That is, the structural relations between *shouga* and melody are dynamic as a function of musical idioms. Each finite idiom of this dynamic topology (morphology) is locally coherent. But the logical contingencies characterizing the parametric relations between the *shouga* and the melody according to one idiom cannot be readily predicted by the parametric relations of another idiom within the same global topology of the complete *shouga*. Future research should investigate whether naïve and untrained listeners can perceive this dynamic structural mapping in the absence of any knowledge of the composer's intentions to represent the musical form through the sounds of the *shouga.*

To this point, a final fact made clear by the results of the present work is how complex a purely "objective" measurement of musical structure would have to be in order to effectively articulate all of the subtle structural nuances that are so essential to the perception of musical form and processed so readily and intuitively by the perceiving subject. To reframe the sentiments of Halliwell (1994), although the descriptive statistics presented throughout this study of objective musical parameters can provide researchers with certain essential clues about the ways in which music and the *shouga* are perceived, the implicit, subjective, and direct experience of the art itself is indispensable to a true understanding of the relations between its psychological forms on the one hand and physical structures on the other, as well to the enjoyment and appreciation of both the beauty and intelligibility of the musical tradition.

Conclusion

The present analyses have provided several convergent indications of a clearly complex but just as clearly non-arbitrary structural correspondence between the *shouga* of the *Ikuta-ryu* and the melody of *Rokudan no Shirabe.* It is hoped that this work will inspire further research into the psychological insights the traditional Japanese musical ethos and techniques such as the *shouga* has to offer to the musical and perceptual sciences.

Author note This research was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Alexander Graham Bell Canada Graduate Scholarship (CGS-D) and a Japan Society for the Promotion of Sciences (JSPS) Doctoral fellowship to TB, and FRQSC (2014-SE-171434) and NSERC (RGPIN 2015-05280) grants and a Canada Research Chair (950-223484) to SM. A profound thank you to sensei ANDO Masateru for accepting TB as one of his musical pupils, for references to related scholarly work, and for encouraging this research.

References

Adriaansz, W. (1973). *The Kumiuta and Danmono traditions of Japanese koto music*. Los Angeles, CA: University of California Press.

Ando, M. (1986). 生田流の箏曲 (*Koto of the Ikuta School*). H. Yoshikawa (Ed.). Tokyo, JP: Kodansha.

Ando, M. (1989). Koto scales and tuning. *Journal of the Acoustical Society of Japan (E)*, *10*(5), 279-287.

Ando, M. (2017). Personal communication. Jun–Dec, 2017. Setagaya, JP.

Ayari, M., & McAdams, S. (2003). Aural analysis of Arabic improvised instrumental music (taqsīm). *Music Perception, 21*(2), 159-216.

Bonin, T., Matsunaga, R. & McAdams, S. (in preparation). Cross-cultural perception of Japanese- and Western-typical melodic contour profiles.

Bonin, T., Degtyareva, K., Hahn, E. (2019). Investigating the content and form of musical imagery evoked by excerpts of Mahler's 5th symphony. Paper presented at the CIRMMT-OICRM-BRAMS Symposium. May 9–10, 2019. Montreal, CA.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Condo, G. (2017). *The way I think.* Interview for Louisiana Museum of Modern Art. Soho, New York.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*(4), 341.

Eitan, Z., & Granot, R. Y. (2006). How music moves: Musical parameters and listeners images of motion. *Music Perception*, *23*(3), 221-248.

Eiting, M. H. (1984). Perceptual similarities between musical motifs. *Music Perception*, *2*(1), 78-94.

Gill, S. P. (2015). *Tacit engagement: Beyond interaction.* New York, NY: Springer Cham.

Giolai, A. (2019). Sensing the music: oral mnemonics as a technique of affective sensitization in Japanese “court music”. *Asian Anthropology*, 1-19. doi: 10.1080/1683478X.2019.1629472

Halliwell, P. (1994). Learning the koto. *Canadian University Music Review/Revue de musique des universités canadiennes*, *14*, 18-48.

Hughes, D. W. (1989). The historical uses of nonsense: Vowel-pitch solfège from Scotland to Japan. In M Philipp (Ed.), *Ethnomusicology and the historical dimension* (pp. 3–18). Ludwigsburg, Germany: Philipp Verlag.

Hughes, D. W. (1991). Oral mnemonics in Korean music: Data, interpretation, and a musicological application. *Bulletin of the School of Oriental and African Studies*, *54*(2), 307-335.

Hughes, D. W. (2000). No nonsense: The logic and power of acoustic-iconic mnemonic systems. *British Journal of Ethnomusicology*, *9*(2), 93-120.

Malm, W.P. (2015). Japanese music. In *Encyclopedia Britannica.* Retrieved from https://www.britannica.com/art/Japanese-music. Accessed July 15, 2019.

Mananoi, Y. & Akira, S. (1978). 口唱歌大系,日本 の楽器のソルミゼーション (*Kuchishouga taikei: Solmization of Japanese instruments*). Tokyo, JP: CBS-Sony.

Marks, L. E. (2000). Synesthesia. In E. A. Cardeña, S. J. Lynn, & S. C. Krippner (Eds.), *Varieties of anomalous experience: Phenomenological and scientific foundations*. Washington, DC: American Psychological Association.

McAdams, S. (1987). Music: A science of the mind? *Contemporary Music Review*, *2*(1), 1-61.

McGuire, P. K., Silbersweig, D. A., Murray, R. M., David, A. S., Frackowiak, R. S. J., & Frith, C. D. (1996). Functional anatomy of inner speech and auditory verbal imagery. *Psychological medicine*, *26*(1), 29-38.

Mittledorf, J. (2003). *Probability of duplicate pairs.* Retrieved from http://mathforum.org/library/drmath/view/62941.html. Accessed June 1, 2019.

Nelson, S. G. (2008a). Court and religious music (1): History of gagaku and shōmyō. In A. M. Tokita and D. W. Hughes (Eds.), *The Ashgate research companion to Japanese music* (pp. 35–48). Farnham, UK: Ashgate.

Nelson, S. G. (2008b). Court and religious music (2): Music of gagaku and shōmyō. In A. M. Tokita and D. W. Hughes (Eds.), *The Ashgate research companion to Japanese music* (pp. 49–76). Farnham, UK: Ashgate.

Neumann, E. (1959). *Art and the creative unconscious.* London, UK: Routledge & Kegan Paul.

Polanyi, M. (1966). *The tacit dimension.* Chicago, IL: The University of Chicago Press.

Rao, P., Ross, J. C., Ganguli, K. K., Pandit, V., Ishwar, V., Bellur, A., & Murthy, H. A. (2014). Classification of melodic motifs in raga music with time-series matching. *Journal of New Music Research*, *43*(1), 115-131.

Rose, F. & Kapuscinski, J. (2013). Orchestration in gagaku music. Retrieved from https://ccrma.stanford.edu/groups/gagaku/theory/pitch-en.html. Accessed May 1, 2019.

Rose, F. & Kapuscinski, J. (2014). Japanese traditional orchestral music: The correlation between time and timbre. Retrieved from http://www. jaroslawkapuscinski.com/pdf/japanese-traditional-orchestral-music.pdf. Accessed May 1, 2019.

Schubert, E., Murari, M., Rodà, A., Canazza, S., Da Pos, O., & De Poli, G. (2019). Verbal and cross-modal ratings of music: Validation and application of an icon-based rating scale. *i-Perception*, *10*(3), 1–21. doi: 10.1177/2041669519852643

Sears, David R. W. 2008. Processing surprise tension in tonal melodies. *Inquiry Research Journal, 9,* 16–23.

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89, 305–333.

Tokita, A., & Hughes, D. W. (2008). *The Ashgate research companion to Japanese music*. Farnham, UK: Ashgate.

van Kranenburg, P., Volk, A., & Wiering, F. (2013). A comparison between global and local features for computational classification of folk song melodies. *Journal of New Music Research*, *42*(1), 1-18.

Wittgenstein, L. (1921). *Tractatus logico-philosophicus*. (Trans. CK Ogden, 1922). London: Routledge & Kegan Paul.

Zbikowski, L. M. (1997). Conceptual models and cross-domain mapping: New perspectives on theories of music and hierarchy. *Journal of Music Theory*, *41*(2), 193-225.

Appendix 3A: Mittledorf's (2003) solution for the probability of *r* repeated sequences in a sample of *n* events from a universe of N possible states

$$\frac{\mathrm{N!\,n!}}{(\mathrm{N-n+r})!\ (\mathrm{n-2r})!\ \mathrm{r}!\ 2^{r}\ \mathrm{N}^{n}}$$

Where:

N = the number of possible states in the universe

[possible unique *t*-mora tuples]

n = the size of the sample taken from the universe of possible states

[number of *t*-mora tuples in *Rokudan no Shirabe*]

r = the number of repeated sequences

[r = 1for *t* =16]

Compare with the results presented in Table 3.5, *t* = 16:

$$\frac{1.84\mathrm{E}19!\,719!}{(1.84\mathrm{E}19-719+1)!\ (719-2)!\ \ 2\ \ 1.84\mathrm{E}19^{719}}$$

Appendix 3B: Standard equations for the computation of weighted means and weighted standard deviations

Weighted mean

$$\bar{x}_w = \frac{\sum_{i=1}^{n}(x_i * w_i)}{\sum_{i=1}^{n} w_i}$$

Weighted st dev

$$\sigma_w = \sqrt{\frac{\sum_{i=1}^{n} w_i * (x_i - \bar{x}_w)^2}{\frac{(N' - 1)\sum_{i=1}^{N} w_i}{N'}}}$$

Where:

$\bar{x}_w$ = the weighted mean

σ_w = the weighted standard deviation

n = the number of observations

i = the ith observation

x_i = the value of the ith observation

w_i = the weight of the ith observation

N'= the number of non-zero weights

Appendix 3C: Weighted average frequency (*SD*) of unique *Mora* (M) tuples of length *t* mapping onto a given *Pitch* (P), *Interval* (I), *Relative Interval* (O), and *Contour* (C) tuple

Mora Tuple Size (t)	N	M(P) (*SD*)	M(I) (*SD*)	M(O) (*SD*)	M(C) (*SD*)
1	734	5.07 (1.68)			
2	733	3.17 (1.59)	6.55 (2.69)		28.06 (5.74)
3	732	2.18 (1.45)	3.38 (1.96)	80.29 (13.67)	33.50 (11.37)
4	731	1.58 (0.91)	2.10 (1.35)	47.13 (16.65)	26.93 (10.26)
5	730	1.28 (0.62)	1.54 (0.98)	15.64 (8.63)	19.03 (8.59)
6	729	1.12 (0.35)	1.25 (0.62)	4.29 (2.80)	11.62 (5.99)
7	728	1.07 (0.26)	1.13 (0.42)	1.81 (1.06)	6.70 (3.94)
8	727	1.04 (0.20)	1.06 (0.26)	1.22 (0.47)	3.99 (2.50)
9	726	1.03 (0.18)	1.04 (0.19)	1.1 (0.33)	2.48 (1.58)
10	725	1.03 (0.16)	1.03 (0.16)	1.05 (0.22)	1.73 (1.00)
11	724	1.02 (0.14)	1.02 (0.14)	1.03 (0.20)	1.35 (0.67)
12	723	1.01 (0.12)	1.01 (0.12)	1.03 (0.19)	1.17 (0.41)
13	722	1.00 (0.09)	1.01 (0.09)	1.02 (0.14)	1.09 (0.28)
14	721	1.00 (0.00)	1.00 (0.00)	1.01 (0.10)	1.04 (0.21)
15	720	"	"	1.01 (0.07)	1.02 (0.14)
16	719	"	"	1.00 (0.05)	1.01 (0.09)
17	718	"	"	1.00 (0.05)	1.00 (0.05)
18	717	"	"	1.00 (0.00)	1.00 (0.00)

Appendix 3D: Raw text files used in the development of Study 3

1. Concerning the use of morae Chi, te, tsu, and to

Ando (1986, p. 198) *[t]chi, te, tsu, to*. If the sound is short, the sound remains represented by this single syllable, but if the note is sustained, the syllable "n" is added to extend the tone. In [a sequence such as] *chi-te-to*, *chi* and *te* are played using for the thumb while *to* is mainly performed using the middle finger and forefinger. A single chi (n), tsu (n), or te (n) is played with a strong thumb.

Mananoi & Akira (1978, p. 36) In other words, there is a fairly obvious tendency to sing the core tones in the melody of a koto as "ten" and the connection with the above-mentioned "ton" and "chi-n."

Chi

Mananoi & Akira (1978, p. 36) It can be said that *chi* is used for the high tone in the *phrase*. This may be a direct effect of the shouga chant, but it can also be said that it is a faithful reflection of the general tendency that "chin" represents the high-pitch while "tsun" and "ten" are higher than "ton".

Tsu

Mananoi & Akira (1978, p. 36) There are also exceptional cases where [tsu and te] are associated with higher or lower [tones], and the height cannot therefore be treated in general. Even with the shamisen, "tsun" and "ten" are not necessarily either high or low, but when organized in a sequence "tsunten" or "tentsun" its meaning is more readily understood. But on the koto height relationship is still not necessarily determined even if the two sounds are placed side by side. But this does not mean that "tsun" is used randomly, *but rather that different melodic events require a common function of the same mora, often performed unconsciously.* In other words, there is a fairly obvious tendency to sing the core tones in the melody of a koto as "ten" and the connection with the above-mentioned "ton" and "chi-n."

To-te

Mananoi & Akira (1978, p. 36) Ton is a kana sound often indicating a normal striking of an open string by the middle finger. In that case, it is often associated with the same note played by the thumb an octave above and represented by the kana "ten".

Mananoi & Akira (1978, p. 37) tontentontenten is a unique sound type. There is something unique to the accompaniment, and unique chords, with tones associated with "ton, ren, and ten"

2. Concerning the use of morae *Ra, ri, ru, re,* and *ro*

Ando (1986, p. 198) When you play the middle finger, forefinger, or one thumb weakly, the sound is to be played weakly is represented by a "ra" "ri" "ru" " re" or "ro". For example, you can hear in "ton ren ten" or "kara kara rin" that [the "r" sounds] are played more softly. To reflect [the technique] of the scooped plectrum, syllables like chi-ri, tsu-ru, te-re, to-ro, are used.

Mananoi & Akira (1978, p. 13) For particular "[plectrum] scooping" ornamental techniques "ren, rin, ron, and run" are used, e.g., *kororin*, *sararin*, etc. [These mora] represent a particular "back to back" continuity."

> [N.B., sometimes this "'back to back' continuity" may be associated with a correlated physical continuity, such as in the case where *re* represents a continuity of *te* in *te-re*, *ru* of *tsu* in *tsu-ru*, or *ri* of *chi* in *chi-ri* by activating the same physical string twice in a row. At other times, this "'back to back' continuity" is more obviously perceptual, reflecting the continuity of a musical contingency between melodic adjacent notes represented by the idiom—e.g., in the common *ko-ro-ri, ro* and *ri* are continuing the descent of the initial pitch *ko,* and so all three notes are intended to be heard as a singular musical expression, as opposed to the case where the same notes are represented by *chi-tsu-te* and are therefore intended to represent distinct musical events. The same continuity is expressed in the "ascending motif" idiom *ka-ra.* Therefore *ra, re, ri, ro*, are used to reflect a psychological "'back-to-back' continuity" among musical events, and is sometimes correlated with an empirical "back-to-back" event such as may be observed in the physical playing technique.]

3. Concerning robust *soukyoku* idioms

Ko-ro-ri

Ando (1986, p. 158) the beginning of a phrase is a strong sound, the weak part is a rolling sound, and so on, which is very similar to the syllables used in the *shouga*. There are certain types of accompaniment depending on how the sounds are connected, that is, context, or depending on the finger used to play the music.

Mananoi & Akira (1978, p. 37) In kororin, three adjacent strings are played sequentially descending with a thumb in a dotted rhythm. It can be said that it is a typical idiom of the koto, and even when it is not performed with a dotted rhythm it still descends with a questioning movement….

Ko-ro-rin vs Chi-tsu-ten

Ando (1986, p. 198) In the case of three consecutive descending notes in a melodic motif, it is often expressed as "kororin." In this case you can see that the first sound is strong and the second and third are weaker. However, care should be taken as it may also be expressed as "chi tsu ten" depending on the melodic context.

Ka & Ko

Mananoi & Akira (1978, p. 37) Katakana 'K" is found commonly in general onomatopoeia but is rare to begin a musical phrase.

Sa-ra-rin

Mananoi & Akira (1978, p. 37) In actual performance [of the *shouga*] it is common to sing in a narrow range, but [the corresponding musical event] is a long descending glissando.

Shya, shya-shya, rya, and chya

Ando (1986, p. 199) To indicate [the characteristic timbre of] two sounds playing simultaneously, *shya*, *rya*, and *chya* are used. It should be noted that *shya* is used equally to represent activation of the strings by one finger on two strings, two plectra (thumb and middle finger) simultaneously, or the forefinger plectrum followed by the middle finger plectrum sequentially [as in the case of *shya-shya*, see below], depending on the context. Shya is distinguished from the others because it is often repeated as shya-shya and occurs at the beginning and end of phrases. Rya often used when playing two consecutive strings with the thumb (note 11). Chya is often used when playing adjacent strings by plucking upward with the plectra.

4. Concerning the expression of melodic ornaments

Ando (1986, p. 199) In order to show the [ornamental] change of the resonance due to modifications of the pitch, the trailing [ornament] tones are used as follows: *chi-i*, *tsu-u*, *te-e*, *to-o*, *shya-a*

Mananoi & Akira, (1978, p. 15) In general, the shouga does not address such small variations and ornaments. This is because the shouga fulfills its purpose by casting the essence of a melody.

Chapter 4 - 箏の唱歌. Naïve listeners' perception of formal analogy between the *shouga* and melody of *Rokudan no Shirabe* (六段の調)

Abstract

The *shouga* (唱歌) is a performative transliteration of traditional Japanese music, comprised of systematically ordered Japanese speech sound (*mora*) sequences known as *idioms*. The *shouga* is used to indicate the intended perceptual form of a given piece of music, with the temporal progression of the *shouga*'s constituent idioms exhibiting a structural correspondence with the temporal progression of the listening subject's perception of the formal contours and motifs of a musical melody. The present experiment provides perceptual data regarding the ability of untrained non-Japanese listeners to distinguish true *shouga* patterns from arbitrarily constructed strings of Japanese morae (*foils*) in association with melodic excerpts of the seminal *soukyoku* (箏曲; koto music) work *Rokudan no Shirabe* (六段の調), by YATSUHASHI Kengyo (八橋検校). Sixty-four melodic phrases were extracted as *original excerpts* from the *Ikuta* scores of *Rokudan no Shirabe* based on the teachings of sensei ANDO Masateru (安藤政輝). Vocal performances of the corresponding target *shouga* phrases were recorded. Finally, vocal performances of *retrograde foils* were recorded. The retrograde foil of each true *shouga* sequence presented all of the original morae within the true *shouga* in reverse order, except for the phoneme ン (n), which retained its original position in the phonetic string. On each trial, participants were first presented the *original excerpt*. After this, participants (N = 40) clicked sequentially on two buttons, labelled "1" and "2," to play the two vocal renditions. The assignment of the true *shouga* and foil renditions to each of the playback buttons was hidden from participants and randomized between trials. Participants indicated which one they believed to be the *shouga* and clicked *Next* to advance to the next trial. Analysis of the results indicates

the capacity of untrained non-Japanese listeners to detect true *shouga* above chance rates. Implications for the psychological conceptualization and measurement of the subjective perception of musical form are discussed.

Introduction

Background

The perception and psychology of musical form

Music appears in the mind of the listener exhibiting a distinct *form*. As both colloquial understanding and expert analysis indicate, this form is dynamic. Musical forms are characterized by the impressions of continuity and change and are frequently related to the listener's abstract senses of space, time, relation, and contrast (see most recently Noble, 2018a; Godøy, 2019). To study these forms analytically in the context of perceptual science is to pose the question of how these forms are systematically structured and operative within the perceptual field of the listener. How are the varieties of musical form to be intellectually organized—delineated and categorized? This fundamental question has been posed under a variety of working models: Where does a musical form begin and end (e.g., Tenney & Polanksy, 1980)? How is one form to be distinguished from a concurrent or subsequent other (e.g., Bregman, 1990)? What are the basic perceptual dimensions along which the structures, identities, functions, and relations of musical forms might be reliably parameterized (e.g., McAdams, 1989)?

Such questions beckon the artistic concept of musical *contour.*[98] The contour of music, like the contour of sculpture, is the *boundary*—the curvature, figure, or outline—of its perceptible form. As we have briefly introduced, the forms of music, unlike the forms of sculpture, are dynamic. Therefore, the contours of music are not only *spatial,* in the metaphoric

[98] Not to be confused with the often-cited operational definition of melodic *contour* used in the music cognition literature as "the patterns of ups and downs in the musical melody," (cf., e.g., Dowling, 1978). We will return to this point in the general discussion.

sense that they differentiate multiple *simultaneously possible* states, they are also *temporal,* in the metaphoric sense that they differentiate multiple *sequentially possible* states. Thus, a musical contour prescribes formal delineations across simultaneously and sequentially perceived musical states. Perhaps unsurprisingly, articulating the direct subjective experience of musical contour in formal language is remarkably difficult. Perceptual scientists have grappled with this problem by devising a variety of empirical techniques. To investigate the temporal development of musical forms, some of the most informative techniques are those that require participants to provide repeated measures or continuous responses of their experience of musical form as a piece unfolds across time.

For example, Krumhansl (1996) conducted a perceptual analysis of Mozart's Piano Sonata in Eb Major (K. 282), which required participants to complete three response tasks. The first was to indicate where the temporal contours were (through a process known as segmentation), the second was to provide a continuous measure of experienced tension over the duration of the piece, and the third was to identify new musical ideas (or motifs) as they occurred throughout the piece. The results indicated that tension ratings were influenced by a complex array of objective factors, including local pitch maxima and minima in the melody, the density of notes and dynamics within the performance, harmony, and tonal functions. Temporal boundaries were found to coincide with peaks in musical tension and slower tempi, while new musical ideas co-occurred with lower levels of tension and intermediate tempi. Although the specific contributions of each objective parameter to the overall sense of musical tension could not be discerned, subjective ratings of musical tension were remarkably consistent within and between listeners. Furthermore, a follow-up study in which three performance variations constrained either 1) the tempo of the music, 2) the dynamics of the music, or both the tempo and dynamics of the music, found that the musical tension ratings provided by subjects for each of these performance variations differed only slightly, indicating that the *form* of the musical image is clearly dissociable from any one particular objective parameter state—even if these states were found to correlate in some cases with the subject's rating of musical form. Such a result is consistent with the Gestalt notion that the subjective perception of form cannot be reduced to the objective state of the stimulus. Several related studies include the work of Deliège (1989), Meyer (1996), Addessi and Caterina (2000), Douglas, Noble & McAdams (2016).

In one such related study, Ayari and McAdams (2003) investigated European and Arabic musicians' respective perceptions of a traditional Arabic *taqsīm*—a melodic improvisation which often foreshadows the characteristic features of a larger Arabic musical composition. The listening experiences of both Arab and European musicians were determined by three tasks: Identification, Segmentation, and Reduction. Identification was similar to the procedure in Krumhansl (1996), in which participants had to identify new musical ideas. In Ayari and McAdams (2003), the task involved pressing a key in the moment the listener detected a change in the musical idea, with the option of then providing a verbal description of the musical dimension (e.g., rhythm, melody, dynamics, etc.) that cued them to this change. As in Krumhansl (1996), segmentation involved the delineation of the temporal boundaries of a musical form "into chunks that were as musically coherent as possible," by pressing a key and stating aloud the defining musical aspect when a segment boundary was detected. These segments were refined across third and fourth subsequent listening trials. Finally, in the Reduction task, participants were asked to reduce each of the segments they had delineated in the segmentation task into a single, "generative" melodic figure or motif, either by singing or transcribing a figured "core structure."

Analyses of the European listeners' task performance revealed three general modes of perceptual sensitivity to the *taqsīm*: The first was a sensitivity to tonicity, reflected in the designation of pivot and resting tones, which were reported by European listeners to provide temporal landmarks throughout the piece on the basis of their relative tonal (in)stability. The second was to salient formal (i.e., "structural") and ornamental (i.e., "superficial") envelopes of the music. These were observed in European listeners' reports of the "repeated notes," "introduction," "cadence," or "re-exposition," etc., of particular musical phrase structures in the former case, and of melodic "contours," "trills," and "appoggiaturas" along the musical surface in the latter. Finally, European listeners demonstrated sensitivity to apparent modal shifts within the *taqsīm*. The majority of Western segmentations were made without verbal elaboration, but the available verbal reports relied strongly on descriptions of mode (spectral) and metrico-rhythmic (temporal) features, as did their subsequent reductions, which typically defined rhythmic patterns, melodic contours, or ornamentation, and thus generally reflected surface features of the music.

In contrast, and perhaps unsurprisingly, Arab listeners reported twice as many formal identifications as the European listeners, and these reports exhibited greater inter-rater reliability as well as closer correspondence with the theoretical formal analysis of the *taqsīm*. Furthermore, the character of these identifications was different from those of the European listeners. Thus, the European and Arabic listeners' experience of the *taqsīm* were both quantitatively and qualitatively distinct from one another. The authors reasoned that this process of reduction, typical of Arabic music listening, facilitates a hierarchical organization of the musical percept and improves intelligibility of the piece across its longer time spans. This proposition is aligned with models of Western musical organization and perception (e.g., Lerdahl & Jackendoff, 1983;

Krumhansl, 1990). The ability to interpret the hierarchical structuring of the *taqsīm* allowed Arabic listeners the attentional flexibility to engage with multiple levels of the musical form, with individual differences in the tendency to focus on either the local or global details nevertheless resulting in consistent formal identifications between participants at a given level of resolution. In contrast, this flexibility of attending was suggested to be relatively unavailable to the Western listeners who, in the absence of prior experience with the *taqsīm*, could not access the hierarchical structuring of the form necessary to make the appropriate formal reductions. Such findings are coincident with those of Drake and El Heni (2003), who found that French and Tunisian subjects could entrain to higher and more diverse metrical levels of their native music than they could to those of each other's (foreign) music. The ability to perceive musical form is contingent not only on the objective parameters of the musical stimulus but also on musical experience, mediated through (often implicit) theoretical knowledge of how the form is expected to progress.

To evaluate the extent to which culture-specific expertise influences the perception of musical form, Mungan, Yazıcı and Kaya (2017) replicated the design of Ayari and McAdams (2003) using more structured Arabic musical forms instead of the *taqsīm* improvisatory form. The authors found a higher rate of coincidence between Western and Arabic segmentations, and that musicians, regardless of their cultural disposition, outperformed non-musicians. Thus, the authors reasoned that while musical experience is a primary predictor of the complexity of the musical forms perceived by the subject, expertise with a particular musical tradition plays only a secondary role, with this factor being more influential in cases where the musics are unstructured and free-formed (as in the improvisations of the *taqsīm*). The results also indicated that the boundary segmentations produced by untrained subjects and expert raters alike correlated well with the melodic arcs, duration separations, or tonal "jumps" on the musical surface. Mungan et al. (2017) argue that this result supports "a universality claim" for the perception of world musics. Finally, the authors noted that there were many locations in the music that exhibited the same physical attributes listed above that nevertheless did not elicit demarcations from participants. This final result indicates that the subjective perception of musical form is not a matter of "rotely segmenting," the music but instead a dynamic process of its own, in which the salience of objective stimulus features interacts with concurrent psychological processes such as the flux of attentional deployment at the discretion of the subject's interest.

The foregoing examples demonstrate that the perception of musical form is associated with objective parameters of the musical stimulus (bottom-up factors), theoretical knowledge and prior experience with a musical tradition (top-down factors), and a third, much more elusive psychological factor. This third factor allows subjects to distinguish among the purely experiential factors of the musical form, resulting in the ability to, for example, perceive and report the same musical form despite significant changes in the objective parameter states of the musical stimulus (cf., Krumhansl, 1996) or to distinguish between identical objective parameter states such as rhythmic and melodic sequences on the basis of whether or not they contribute to core melodic motifs (cf., Ayari & McAdams, 2003) or formal boundaries (cf., Mungan et al., 2017). This psychological factor is irreducible to the other purely physical or knowledge-based parameters. It would appear to correspond more directly with the phenomenological qualia, or experiential content, of the musical percept *per se.*

1.1.2. Investigating the phenomenology of musical forms

A recent chapter by Godøy (2019) argues that this third factor has received relatively little attention and that the experiential content of the musical percept is perhaps best modelled in terms of *sound shapes.* The *shape cognition* central to this processing of musical forms is suggested to be amodal, in that it does not correspond with any one particular sensory modality and that cross-modal analogies can be found between the various perceptual domains. Godøy emphasizes in particular the experiential commonality between the temporal morphologies of musical forms and those of proprioceptive-motoric forms, or physical motion *gestures*. He writes,

> A possible answer to this question of… musical imagery could be that of shape cognition in music; the appearance and persistence of salient imagery of musical sound may be linked to *enacted* shape cognition, meaning that images of musical sound may be "brought to life" in our minds by active shape tracings. … The crucial element here would be shifting between images of motion and sound, with motion imagery triggering sound imagery, and conversely, sound imagery triggering motion imagery, all the time with shape cognition as the translating factor between motion and sound. … Shape cognition can become useful in a number of music-related tasks, in particular by bridging the gaps between "hard" numerical data and "soft" qualitative concepts. … Furthermore, these advantages of shape cognition are based on the fact that shapes are inherently holistic and extended, whereas more abstract symbols are inherently atomistic and local (Godøy, 2019, p. 254).

Regarding the descriptive challenge of measuring this shape cognition in action, Godøy (2019, p. 238) notes, "With the focus on abstract symbolic representations in Western musical culture, for instance, on discrete pitches and durations, concepts for continuous sound and body motion have been less developed… Handling shapes in research [with these]… conceptual and technological tools… is inherently challenging because shapes are *distributed;* that is, shapes are not reducible to singular values or abstract symbols," and concludes that advances in technology and scientific concepts will allow for a greater investigation of the shape cognition underlying the perception of musical forms. In her 2012 article, Stevens (2012) highlighted the need to diversify the empirical and theoretical approaches to music cognition research across cultures for exactly these reasons. Such research, she suggested, would not only determine which of the existing theoretical paradigms in experimental psychology generalize to non-Western demographics, but would also contribute novel insights to the Western music, psychology, and music perception literatures by introducing new conceptual paradigms according to the musical and philosophical traditions of non-Western cultures. To this end, we introduce in the present article not a new computational technology but rather a centuries old musical transliteration device known as the *shouga* (唱歌), from the teachings of the traditional Japanese *Ikuta-ryu* (生田流).

The shouga of Ikuta-ryu

The *shouga* is an embodied phonetic system used to represent and articulate the intended form of a given piece of music, with particular *idioms* associated with particular formal contours. The *Ikuta-ryu* is one of the primary schools of traditional *soukyoku* (箏曲) or Japanese koto music. It was founded by IKUTA Kengyo (生田検校, 1656–1715), with particular emphasis on the co-ordination of the koto and shamisen musics of the Osaka region. The origins of the *shouga* as found in the *Ikuta-ryu* stem from the rehearsal practices of *gagaku* (雅楽) court musicians of Japan (453 AD; Malm, 1983). The internal logic of the *shouga* is predicated on the fact that the empirical art object (the musical sound itself) is, as a consequence of human creativity, an objective representation of the internal states of the subject who produced them.[99] Such a conceptual paradigm has enabled the development of a highly refined and dynamic system of musical description (see Hughes, 2000; Giolai, 2019). Expert musicians are able to rehearse, teach, and communicate with others, the desired form of the music to be perceived by the listener through the articulation of the *shouga*. The *shouga* is comprised of systematically ordered Japanese speech sound (*mora*) sequences known as *idioms* (Mananoi & Akira, 1978). The temporal progression of the idioms within the *shouga* corresponds with the intended temporal progression of the *sound shapes*, which are, in the parlance of Godøy (2019), perceived by the subject when listening to the corresponding music. Several researchers have indicated the intuitive analogy between the perceptible form of the *shouga* and the perceptible form of the musical melody it represents.

Giolai (2019) makes particular mention of the embodied and cross-modal associations between the physical actions used to create the instrumental sounds of the music, the muscular tension resulting from the performance of the *shouga,* and the resulting formal tension in the musical percept. Hughes (2000) argues that while the *shouga* has come to be written down in recent years it will most likely continue as an aural and performative tradition, since the *orality,* or embodied experience, of the *shouga* phonemes is essential to fully comprehending their association with the music. Hughes (2000), Giolai (2019), and Nelson (2008a, b) all suggest that the pattern of vowels within the morae of the *shouga* may correlate with the rising and falling

[99] Such a proposition would appear to contradict the conceptual paradigm typical of contemporary perceptual science discourse, which regards the mental state of the subject as a "representation" of an objective occurrence that originates in the "outer," physical realm. This asymmetrical relation between subject and object is not recognized within the ethos of traditional Japanese performance arts, which considers the respective identities of subject and object to be inseparable from their relation to one another, and to reflect the opposite terms of a "mutual arising" (縁起; *engi,* cf., the Sanskrit *pratītyasamutpāda* of the original Buddhist canon).

pitch sequences in the musical melody (Bonin & McAdams, in preparation; see Chap. 3). Finally, Ando (1986), an expert koto performer, notes that the *shouga* idioms are to be selected and understood as a function of musical context, and Mananoi and Akira (1978) likewise caution against interpreting the *shouga* as a static system of abstract symbols used to represent objective musical parameter states,[100] highlighting the fact that the same speech sounds are found to be associated with multiple distinct musical parameter states because "different melodic events require a common function of the same [speech sound], often understood unconsciously," (p. 36). If the *shouga's* coherence and structural correspondence with musical forms is mediated by largely unconscious, intuitive, and embodied (i.e., *tacit,* cf., Polanyi, 1966; Gill, 2015) cognitive processes, to what extent is this implicit and subjective knowledge available to the direct experience of naïve listeners?

The present work

The present experiment provides perceptual data regarding the ability of untrained non-Japanese listeners to distinguish true *shouga* patterns from arbitrarily constructed strings of Japanese morae (*foils*) in association with melodic excerpts of the seminal *soukyoku* work *Rokudan no Shirabe* (六段の調) by YATSUHASHI Kengyo (八橋検校, 1614–1685).[101] The purpose of this experiment was to determine whether untrained listeners could discern true *shouga* phrases from foil sequences containing the same speech sounds (morae) in reverse order. If the functioning of the *shouga* is predicated on its perceptible morphological analogy with the musical stimulus it represents, then untrained listeners should be able to discern true *shouga*

[100] As in Western solfège; cf., the interpretation of Berger, 1965
[101] See Bonin and McAdams (in preparation; see Chap. 3) for in-depth parameterization of the musical melody and *shouga* of *Rokudan no Shirabe.*

sequences from foil sequences on the basis of their differential semblance, or "likeness," to the form of the musical melody. If instead the *shouga* is merely an arbitrary social convention, with the constituent speech sounds bearing no intrinsic formal analogy with the musical phrases they represent, then participants should perform at chance rates when attempting to distinguish true *shouga* sequences from foil sequences containing an arbitrary ordering of the same speech sounds. Should the results indicate that untrained listeners *can* detect the formal analogies between true *shouga* sequences and the musical sounds they represent, they would reveal a highly refined perceptual capacity to discern subjective impressions of musical form that applies *across* individual listeners. Future research of musical form may consequently uncover remarkable insights through further investigation of the perceptual and phenomenological bases of this ability.

Methodology

Participants

Forty (N = 40) North American (Canadian and American) adults (7 female; average age = 23.6 years) participated in this experiment. Thirty-six participants were students in an undergraduate music perception and cognition class who voluntarily took part in the study for extra course credit. The four remaining participants were volunteers interested in Japanese music. None of the participants had any prior training in Japanese music, and none of the participants could speak Japanese, although five (5) participants indicated on the demographic questionnaire that they were "familiar with some phrases, questions, and answers," and 10 (ten) participants responded Yes to the Yes/No question, "Do you watch anime with the original Japanese voice actors? (With or without foreign subtitles, but no foreign language dubs)," indicating that they had at least passive exposure to the speech sounds of the Japanese language. Participants were not selected on the basis of prior musical experience. However, 35 participants (87.5%) indicated some degree of formal musical training, with an average of 10.2 years (range 1–19 years) of musical training among them. This study was certified for ethical compliance by the McGill University Research Ethics Board II and all participants provided informed consent.

Apparatus

All recorded stimuli were captured using a rode NT-1A microphone and a quad core Apollo Twin MkII portable interface from Universal Audio with a minimum 6 dB FS of headroom. All stimuli were processed in Cubase 9 (Steinberg Media Technologies GmbH, Hamburg).

The experimental program was written in PsiExp (Smith, 1995) for Apple MacIntosh computers and uploaded with experimenter instructions to a publicly accessible server address. Participants were assigned a unique participant number by the experimenter, downloaded the program onto their local hard drives, and completed the experiment independently. Participants were instructed to complete the experiment wearing functional headphones (working in stereo with no audible artifacts) to facilitate immersion in the experiment and were requested to adjust their playback levels to a comfortable and clear listening level.

Participants provided signatures indicating informed consent before beginning the experiment. An online demographic questionnaire recorded participants' unique participant number, age, biological sex, number of years of musical training, and experience with Japanese language. All identifying information was stored independently of participants' performance data. All data were stored on a private lab server. This study was certified for ethical compliance by the McGill University Research Ethics Board II and all participants provided informed consent.

Stimuli

Sixty-four (64) *original excerpt* stimuli were melodic phrases excerpted from a solo koto (*soukyoku*) performance of *Rokudan no Shirabe* according to the teachings of sensei ANDO Masateru. Phrases varied in length from 4 to 20 morae, and the number of unique morae within each phrase ranged from 2 to 10 (see Appendix 4A). For each original excerpt, a *true shouga* and a *retrograde foil* were recorded. The *true shouga* and *retrograde foil* stimuli were sung with the same pitches and rhythms as the original excerpt. Each *true shouga* stimulus was a vocal recitation of the *shouga* corresponding with the melodic phrase of the *original excerpt* as notated in the *Ikuta-ryu* notation. Each *retrograde foil* was a vocal recitation of the true shouga *in reverse.* That is, the mora sequence of the true shouga stimulus were reversed to create the mora sequence of the retrograde foil stimulus. There were two exceptions to this general rule. The first concerned ン (n) morae, which retained their original position according to the true shouga sequence. For example, the true shouga mora sequence "*te-n-to-n-shya-n"* was contrasted with the retrograde foil "*shya-n-to-n-te-n".* The shouga sequence "*tsu-n-te-tsu-ko-ro-ri-chi"* was contrasted with the retrograde foil "*chi-n-ri-ro-ko-tsu-te-tsu".* The second concerned the continuation of vowels within the *shouga* in keeping with the melodic ornaments of the original excerpt: Basic pitch events within the original excerpt are always associated with a mora containing both a consonant and vowel component in the *shouga* (*cv mora*; Mananoi & Akira, 1978; Bonin & McAdams, in preparation; see Chap. 3). When a mora is found in the *shouga* containing only a vowel component (*v mora*), it corresponds with a melodic ornament such as a pitch bend or glissando within the original excerpt and continues the vowel found in the preceding cv mora. Thus, in creating the retrograde foil sequences, vowels were extended from the previous cv mora within the *foil* sequence, not within the original sequence; likewise, if a v mora was found in the *true shouga* sequence but did not correspond with a melodic ornament within the original excerpt, it was removed to maintain sequence length and internal coherence. For example, *te-e-n-to-n-shya-n* becomes *shya-a-n-to-n-te-n*. As another example, *chi-n-te-tsu-ko-ro-ri-chi-i-to-n* becomes *to-n-chi-ri-ro-ko-tsu-te-e-chi-n*. These decisions were made on the basis of initial pilot studies, which indicated that a purely retrograde foil of the true shouga stimulus without these modifications was too easily recognized as the foil.[102] The true shouga and retrograde foil stimuli corresponding to each original excerpt were all quasi-controlled for loudness by equating the RMS amplitudes of each sound file. The original excerpt, true shouga, and retrograde foil stimuli together comprised a *stimulus family.* There were 64 stimulus families (64 presented in the experiment proper) in total. Three stimuli were randomly selected for use in the three practice trials. The results of the practice trials were not submitted to the analysis of the final results. Table 4.1 presents four example stimulus families. The entire list of *shouga* strings and foils is presented in Appendix 4A, and the audio files for all stimulus families can be accessed at http://132.206.14.109/supplementaryMaterials/Bonin_PhDThesis_Chap4_Stimuli/.

[102] This preliminary finding in and of itself provides an insight into the non-arbitrary choice of speech sounds within the shouga. ン is used to indicate pitch resonance or sustains of previously articulated "basic pitch events" (Mananoi & Akira, 1978). In the completely retrograde foils of the pilot test, this "n" sound would sometimes correspond with the initial articulation of a "basic pitch event." (e.g., *n-shya-n-to-n-te,* from the first example above). Participants were quick to notice that something about this "acoustic-iconic mapping" (Hughes, 2000) was "not right." Why? One possibility may relate to the fact that the phoneme ン(*n*) is not aspirated; when creating the speech sound *n,* one can sustain the resonance of a previous pitch without interrupting the heard sound and felt pressure in the vocal tract. In closing the mouth for ン, the oral cavity also resonates more noticeably than for open vowel sounds, therefore giving a strong sense of *resonance.* In contrast, *t-* and *k-* morae such as *te, to*, *chi*, *ko* and *ka* are aspirated: to be articulated they must interrupt the flow of air and alter the pressure in the vocal tract. If a sequence of musical events to be represented by a sequence of speech events is characterized by a continuous pitch resonance, then a sequence of speech events that results in an interrupted pitch resonance should naturally be regarded as "less accurate," less desirable, less similar.

Table 4.1 The first four *shouga* sequences extracted from *Rokudan no Shirabe,* and their respective retrograde foils.

Stimulus Family	True *shouga* sequence	Retrograde foil sequence
1	テエントンシャン *te e n to n shya n*	シャアントンテン *shya a n to n te n*
2	シャシャコロリチイトン *shya shya ko ro ri chi i to n*	トチリロコシャアシャン *to chi ri ro ko shya a shya n*
3	チンテツコロリチイトン *chi n te tsu ko ro ri chi i to n*	トンチリロコツテエチン *to n chi ri ro ko tsu te e chi n*
4	シャンテエンシャシャチイン *shya n te e n shya shya chi i n*	チンシャアンシャテシャアン *chi n shya a n shya te shya a n*

Procedure

The program began with a block of three practice trials to familiarize the participants with the experimental procedure. Each block randomly presented one of the practice stimulus families without replacement. At the top of the visual display for each trial was the trial number. Beneath this was a button, centred in the screen which read “play original.” Beneath this button and to the left and right were two additional buttons, equally sized, labeled “1” and “2”, respectively. When clicked, each button would play either the true shouga or the retrograde foil stimulus. The true shouga and retrograde foil were randomly assigned to the two buttons on each trial, i.e., on some trials button 1 would play the true shouga stimulus while button 2 played the retrograde foil, while on other trials the reverse was true. Beneath the buttons 1 and 2 were small square boxes which, when clicked, would become “checked” or “unchecked”. If the box beneath button 1 or 2 was clicked, it would remain clicked until it was clicked again or the other button was clicked. Therefore, only one or none of the square boxes could be checked at the same time. Finally, beneath these two square boxes, centred at the bottom of the screen was a button which read “next,” (see Figure 4.1).

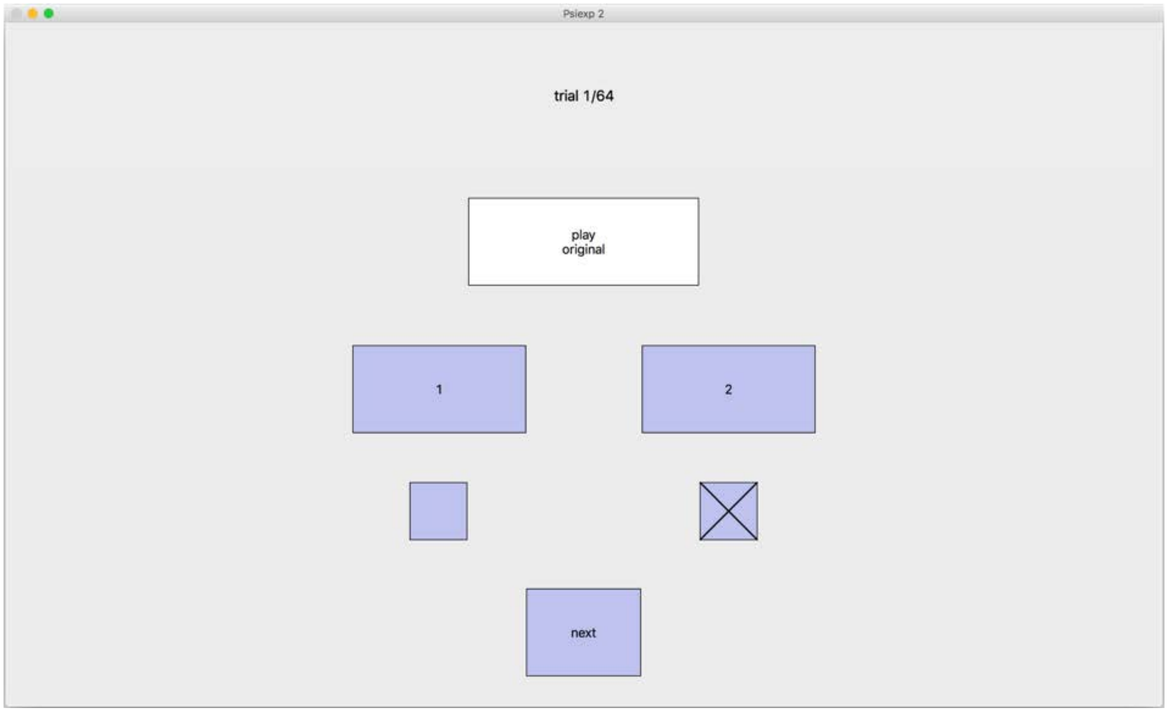

Fig. 4.1 The graphical user interface used for each experimental trial.

Participants first listened to the original excerpt, which began playing automatically at the beginning of each trial. They then had the option to freely click either button 1, button 2, or the "play original" button. If one of the three buttons were pressed during the playback of another button, the playback would stop, and the stimulus corresponding to the pressed button would play from its beginning. If a stimulus was returned to after having been interrupted, it would be played again from its beginning. As the stimuli were short, there was no limit to the number of times a participant could listen to the sounds associated with each button. However, the square boxes beneath buttons 1 and 2 were not activated and therefore could not be checked by the participant until both the true *shouga* and the retrograde foil had been listened to in their entirety at least once. After participants made their selection, the "next" button would activate, allowing them to proceed to the next trial. After the practice trials, participants were instructed to contact the experimenter for clarification regarding any questions they might have before continuing to the experiment proper.

The experiment proper consisted of 64 trials, one for each of the 64 stimulus families. Stimulus families of the experiment proper were presented in random order for each participant. The experimental procedure for each trial was identical to that described for the practice trials. After completing the experiment proper, participants were linked to the demographic questionnaire.

Data analysis

I. Sign test of the proportion of participants detecting true shouga *sequences at rates above chance*

If participants correctly identified the true *shouga* sequence on a given trial the response was registered as a "hit" and scored as 1. Otherwise it was registered as a "miss" and scored as 0. For each participant (N = 40), we averaged the responses scores across each of the 64 trials, resulting in the proportion of all trials for which the participant correctly identified the true *shouga* stimulus. Because the resultant proportions were not normally distributed across all 40 participants, we conducted the non-parametric *sign test* to determine whether the observed frequency of participants (S) identifying the true *shouga* stimulus at rates greater than the expected median of 0.5 was greater than chance. According to the null hypothesis, S should be N/2 = 20. P-values of the sign test correspond with the binomial probability of S statistics at least as extreme as the observed S statistic (Maxwell & Delaney, 2004).

II. Proportion of participants correctly identifying a given true shouga *sequences*

For each stimulus (N = 64), we averaged the response scores provided by each of the 40 participants, resulting in the proportion of all participants who correctly distinguished the true *shouga* stimulus from its retrograde foil. Because the resultant proportions were non-normally distributed across all 64 participants, we calculated a second S statistic for the number of true *shouga* stimuli correctly identified by more than the median proportion of participants (0.5). According to the null hypothesis of chance discrimination, S should be N/2 = 32.

To further investigate participant sensitivity to each of the *true shouga* sequences we ran a binomial test on the number of participants correctly identifying each sequence. We report the total number of *true shouga* sequences identified by more participants than predicted by chance according to the binomial distribution and present a histogram of the number of participants (N = 40) correctly identifying each *true shouga* sequence.

Results

I. Sign test of the proportion of participants detecting true shouga *sequences at rates above chance*

Thirty-seven (37, 92.5%) out of 40 participants correctly identified a majority of *true shouga* sequences (S = 37, p = 8.99E-9). The average number of *true shouga* sequences identified by a given participant was 43.0/64 (67.2%).

II. Proportion of participants correctly identifying a given true shouga *sequences*

Fifty-eight (58, 90.6%) out of 64 stimuli were correctly identified by the majority of participants (S = 58, p = 4.06E-12). The average number of participants identifying a given *true* *shouga* sequence was 27 (67.5%). On the level of individual differences, 42 (65.6%) of 64 true *shouga* sequences were identified by participants at rates better than predicted by chance according to the binomial distribution. The results for all *true shouga* sequences are presented in Figure 4.2.

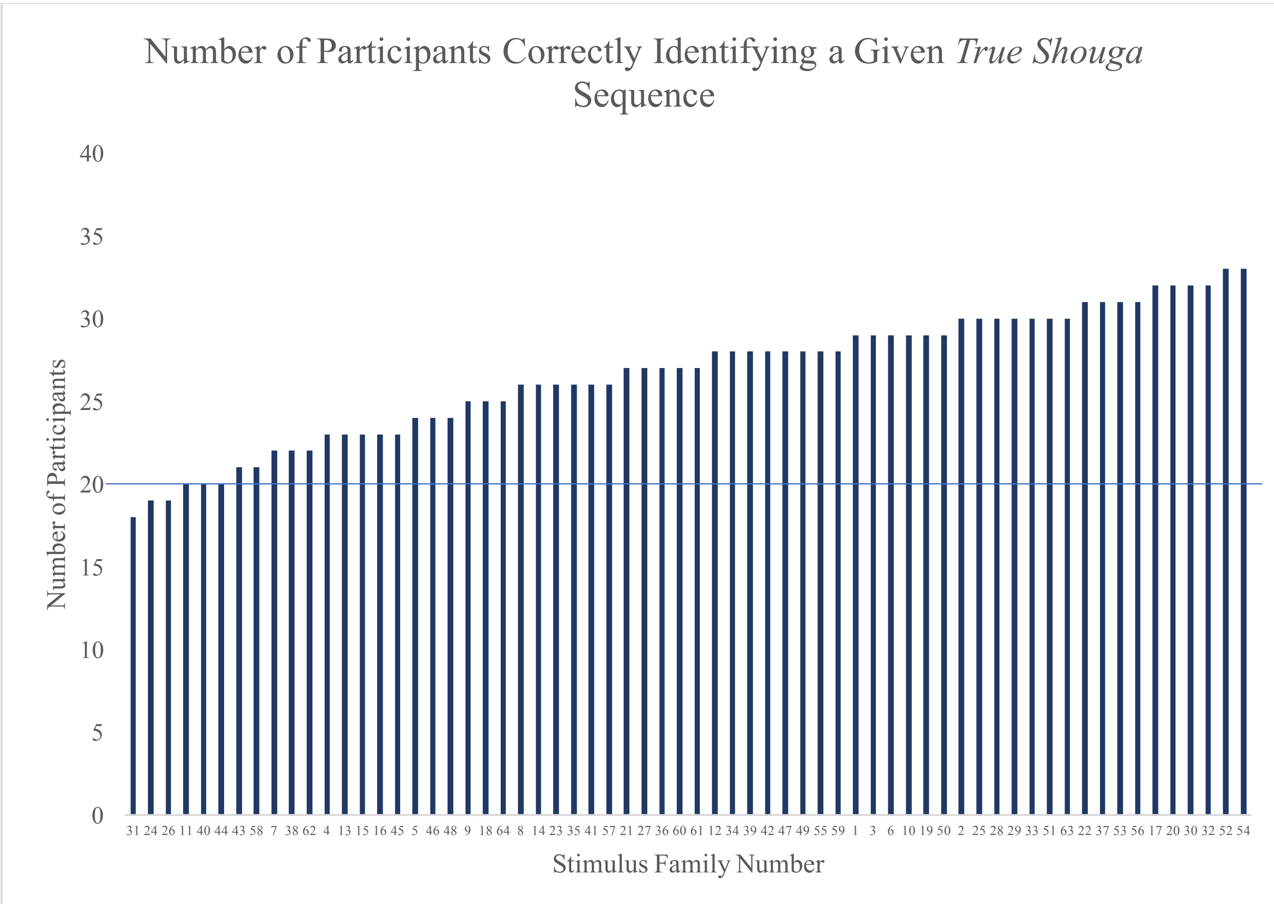

Fig. 4.2 The number of participants (N = 40) correctly identifying each *true shouga* sequence. The horizontal line indicates the median of the binomial distribution.

As a series of follow-up analyses attempting to glean a cursory insight into the empirical structures of the *shouga* associated with the subjective impressions, we conducted three post-hoc correlations between the number of participants correctly identifying a given *true shouga* sequence and 1) the length of the sequence, $r(38) = 0.22$, $p = 0.08$, 2) the number of unique morae within the sequence, $r(38) = 0.14$, $p = 0.11$, and 3) the ratio between the number of unique morae within the sequence and its length, $r(38) = –0.15$, $p = 0.22$.

Summary and Discussion

The present results provide a decisive and almost unanimous indication that untrained listeners—naïve to the pedagogical and musical origins of the *shouga* and with no previous training in Japanese music or the Japanese language—can discern true *shouga* sequences from arbitrary combinations of the same speech sounds in their association with traditional Japanese melodic phrases. As such the results provide strong evidence in support of the amodal *shape cognition* model proposed by Godøy (2019), and reveal a new and potentially rich avenue for the exploration of the perception of musical form.

The results are also consistent with the findings of Mungan et al. (2017) that the effects of culture-specific expertise are secondary and complementary to the basic ability of listeners to detect musical forms. Almost all of the Western listeners in the present sample were not only able to process the musical form of traditional Japanese music, but to associate this form with an analogous perceptual form resulting from the concatenation of foreign speech sounds. This result is reminiscent of the bouba-kiki effect (Köhler, 1947/1970), where non-linguistic strings of speech sounds are readily associated with particular *visual* shapes (see also Liew et al., 2018); in the case of the *shouga,* the form of the sequence of speech sounds is readily associated with the form of the melodic sequence to which it corresponds. The additional and considerable effect here is the fact that the 42 of 64 distinct *shouga* sequences each comprise a unique sequence variation of these speech sounds, indicating that the Gestalt sensitivity to perceptual forms is highly nuanced in music and readily transposable onto analogous perceptual arrays. This perceptual sensitivity suggests that the simplistic definition of musical contour as "the patterns of ups and downs in the musical melody," (cf., Dowling, 1978) may not be sufficiently detailed as an operational parameter for the description of a subject's experience of musical form: All *true shouga* and *retrograde foil* pairs shared identical melodic "contours" as operationally defined in the current music perception literature, and yet the musical morphology as directly perceived by the subject—*the melodic contour per se*—was readily and systematically identified by participants. What exactly this system is we cannot explicitly say; it is encoded within the internal logic of the *shouga,* which we do not explicitly understand at present. But the fact that participants can readily distinguish *true shouga* sound shapes from retrograde foils indicates that the perceptible detail of the musical form present in the mind of the listener is more sophisticated than we are able to capture with our current operational definitions of musical form.

To this end we address our post-hoc analyses. It is worth emphasizing that we did not experimentally control for the length or number of unique morae within the *shouga* sequences presented in this study. Nevertheless, the correlations between the number of participants correctly identifying a given *true shouga* sequence and 1) the length of the sequence and 2) the number of unique morae within the sequence did trend toward significance, indicating that these parameters may serve as useful objective vantage points for future investigations of the perceptual analogies between the *shouga* and the musical forms they represent. A larger stimulus set will also help to further specify these analyses. In particular, as the length and variability of a *shouga* sequence increase, the morphological specificity of the sonic image increases, ostensibly allowing its form to be more readily distinguished from dissimilar forms. These considerations do not, of course, address the question of which objective parameter states correspond with perceptually meaningful image contours, but a systematic investigation of these parameters in terms of robust *shouga* idioms (see Chap. 3, Study 3) may reveal additional insights.

Concerning the direct experience of our participants, very few indicated a high level of confidence in their responses during the experiment when asked informally about their experience after the fact. Most participants, however, were curious about the experiment and reported—as ethnomusicologists and expert musicians have before them—being intrigued that they could somehow intuitively "feel" what the right answer was. Many participants mentioned the idiom *ko-ro-ri,* which has received specific treatment by both Ando (1986) and Mananoi and Akira (1978) in the description of important *soukyoku* idioms. The idiom *ko-ro-ri* is fairly common within the corpus of *Rokudan no Shirabe* (it happens 69 times in total), so one is initially compelled to argue that the idiom was identified as "correct" because of its relative statistical prevalence. But it must be re-emphasized that every instance of *ko-ro-ri* was paired with and randomly sequenced alongside its retrograde counterpart *ri-ro-ko* within this experiment. So, mere exposure alone cannot explain why participants were intuitively sensitive to this idiom. A few participants indicated that the sound *ro* in *ko, ro, ri* was a sort of soft and rolling pattern which they came to recognize. This is particularly interesting when one considers the fact that *ko-ro-ri* always accompanies a descending melodic line from one of the primary pillar tones (scale degree $\hat{1}$ or $\hat{5}$) to a relatively less stable scale degree, and is, with only one exception, associated with a "long–short–short" rhythm, most often characterized by a dotted quarter note–eighth note–quarter note pattern (Bonin & McAdams, in preparation; see Chap. 3). Though anecdotal in the context of the present experiment, these results indicate several additional avenues for future research.

In particular, future research should investigate whether proficiency with these speech sounds in the context of formal language helps or hinders performance relative to non-Japanese speaking listeners, and whether a deliberately selected sample of listeners with no formal musical training differ from a sample including trained and untrained listeners.[103] It could also investigate the extent to which certain idioms are associated with metaphoric image descriptors (see e.g., Noble, 2018b; Marks, 2000; Eitan and Granot, 2006), whether and to what extent the "form" of the *shouga* is associated with onomatopoetic or acoustic-iconic factors (e.g., Hughes, 2000) vs embodied, motoric, and proprioceptive factors (e.g., Giolai, 2019), and whether particular idioms and phrases are associated with particular objective stimulus parameters of the musical melody (Bonin & McAdams, in preparation; see Chap. 3). Research into the perceptual basis of the functioning of the *shouga* may also open doors to the investigation of the unstructured and spontaneous formal analogies that occur in the working discourse among creative professionals. Bonin, Degtyareva and Hahn (2019) found a remarkable correspondence between the content and form of the free associations spontaneously used by performers, composers, and recording engineers in order to describe excerpts of Mahler's fifth symphony. Interviews with professional musicians are often replete with imagistic language, and the holistic experiential integration of the perceptual forms elicited by art objects with extra-artistic aesthetics and psychological tendencies is the basis of the *rasa* tradition in India (Chaudhury, 1965).[104]

[103] Two pilot studies under way with Japanese speakers (N = 12) and untrained listeners (N = 6) do not find any preliminary evidence of clear group differences.
[104] *Rasa* theory extends back to the earliest Vedic texts, where it literally means "essence," "juice," or "taste." In the *Natya Shastra* (~200 BC), an ancient Hindu treatise on artistic expression, *rasa* describes the "aesthetic flavour" of a piece of work which, the Hindus assert, cannot be expressed in words. Says the text, "The primary goal is to create *rasa* so as to lift and transport the spectators, unto the expression of ultimate reality and transcendent values." This *rasa* is produced from a combination of determinants (*vibhava*), or inputs to the conscious receptions, consequents (*anubhava*), or outputs of the conscious response and transitory states (*vyabhicaribhava*) between the various dimensions of the strand of consciousness. The *rasas* can be thought of as distilled strains of consciousness which are uniquely elaborated upon in each performance by the interplay of the distinct *vibhava, anubhava,* and *vyabhi-caribhava* that the audience brings to bear on the experience (Chaudhury, 1965). There are eight primary *rasas* listed in the *Natya Shastra*: Enamoured and erotic (S*ryngaram*), humorous and comical (*Hasyam*), pathetic and disgusting (*Bibhatsam*), infuriated and angry (*Raudram*), compassionate and sympathetic (*Karunyam*), heroic (*Viram*), terrible and horrifying (*Bhayanakam*), and marvellous and amazing (*Abhutam*).

As music perception research continues to develop into the 21st century, it may well be discovered that what is needed for advancement of contemporary perceptual science is not merely the development of new, abstract, symbolic representations of subjective experience (stronger, faster, linear computers), but rather a renewed interest and commitment to the mastery of the techniques cultivated within the musical traditions themselves. The present research indicates that such a possibility is not only relevant to the practicing musical artist but can also contribute to purely scientific and intellectual discoveries and descriptions of the arts.

Author Note:

This research was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Alexander Graham Bell Canada Graduate Scholarship (CGS-D) and a Japan Society for the Promotion of Sciences (JSPS) Doctoral fellowship to TB, and FRQSC (2014-SE-171434) and NSERC (RGPIN 2015-05280) grants and a Canada Research Chair (950-223484) to SM.

Thank you to Bennett K Smith for programming the experiment interface.

References

Addessi, A.R., & Caterina, R. (2000) On segmentation of post-tonal music. In Woods, C., Luck, G.B., Brochard, R., O'Neill, S. A., and Sloboda, J. A. (Eds.) *Proceedings of the Sixth International Conference on Music Perception and Cognition.* Keele, Staffordshire, UK: Department of Psychology. CD-ROM.

Ando, M. (1986). 生田流の箏曲 (*Koto of the Ikuta School*). H. Yoshikawa (Ed.). Tokyo, JP: Kodansha.

Ayari, M., & McAdams, S. (2003). Aural analysis of Arabic improvised instrumental music (taqsīm). *Music Perception, 21*(2), 159-216.

Bonin, T., Degtyareva, K., Hahn, E. (2019). Investigating the content and form of musical imagery evoked by excerpts of Mahler's 5th symphony. Paper presented at the CIRMMT-OICRM-BRAMS Symposium. May 9–10, 2019. Montreal, CA.

Bonin, T. & McAdams, S. (in preparation). 箏の唱歌. Structural relationships between the *shouga* and melody of *Rokudan no Shirabe* (六段の調).

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: The MIT Press.

Chaudhury, P. J. (1965). The theory of rasa. *The Journal of Aesthetics and Art Criticism*, *24*(1), 145-149.

Deliège, I. (1989). A perceptual approach to contemporary musical forms. *Contemporary Music Review*, *4*(1), 213-230.

Douglas, C., Noble, J., & McAdams, S. (2016). Auditory scene analysis and the perception of sound mass in Ligeti's continuum. *Music Perception*, *33*(3), 287-305.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*(4), 341-354.

Drake, C. & El Heni, J. B. (2003). Synchronizing with music: Intercultural differences. *Annals of the New York Academy of Sciences*, *999*(1), 429-437.

Eitan, Z. & Granot, R. Y. (2006). How music moves: Musical parameters and listeners images of motion. *Music Perception*, *23*(3), 221-248.

Giolai, A. (2019). Sensing the music: oral mnemonics as a technique of affective sensitization in Japanese “court music”. *Asian Anthropology*, 1-19. doi: 10.1080/1683478X.2019.1629472

Godøy, R. I. (2019). Musical shape cognition. In Grimshaw-Aagaard, M. Walther-Hansen, M. Knakkergaard, M. (Eds.) *The Oxford handbook of sound and imagination* (Vol. 2), (237-258). New York, NY: Oxford University Press.

Hughes, D. W. (2000). No nonsense: The logic and power of acoustic-iconic mnemonic systems. *British Journal of Ethnomusicology*, *9*(2), 93-120.

Köhler, W. (1947/1970). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York, NY: W.W. Norton & Company.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York, NY: Oxford University Press.

Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception*, *13*(3), 401-432.

Lerdahl, F., & Jackendoff, R. (1983). An overview of hierarchical structure in music. *Music Perception*, *1*(2), 229-252.

Liew, K., Lindborg, P., Rodrigues, R., & Styles, S. J. (2018). Cross-modal perception of noise-in-music: Audiences generate spiky shapes in response to auditory roughness in a novel electroacoustic concert setting. *Frontiers in Psychology*, 9:178. doi: 10.3389/fpsyg.2018.00178

Malm, W. (1983). Some of Japan's musics and musical principles. In E. May (Ed.), *Musics of many cultures: An introduction* (pp. 46-62). Los Angeles, CA: University of California Press.

Mananoi, Y. & Akira, S. (1978). 口唱歌大系,日本 の楽器のソルミゼーション (*Kuchishouga taikei: Solmization of Japanese instruments*). Tokyo, JP: CBS-Sony.

Marks, L. E. (2000). Synesthesia. In E. A. Cardeña, S. J. Lynn, & S. C. Krippner (Eds.), *Varieties of anomalous experience: Phenomenological and scientific foundations*. Washington, DC: American Psychological Association.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data* (2nd ed.). New York: Routledge.

McAdams, S. (1989). Psychological constraints on form-bearing dimensions in music. *Contemporary Music Review*, *4*(1), 181-198.

Meyer, L. B. (1996). *Style and music: Theory, history, and ideology*. Chicago, IL: University of Chicago Press.

Mungan, E., Yazıcı, F. & Kaya, M. U. (2017). Perceiving boundaries in unfamiliar Turkish makam music: Evidence for gestalt universals? *Music Perception*, *34*(3), 267-290.

Nelson, S. G. (2008a). Court and religious music (1): History of gagaku and shōmyō. In A. M. Tokita and D. W. Hughes (Eds.), *The Ashgate research companion to Japanese music* (pp. 35–48). Farnham, UK: Ashgate.

Nelson, S. G. (2008b). Court and religious music (2): Music of gagaku and shōmyō. In A. M. Tokita and D. W. Hughes (Eds.), *The Ashgate research companion to Japanese music* (pp. 49–76). Farnham, UK: Ashgate.

Noble, J. D. (2018a). What can the temporal structure of auditory perception tell us about musical “timelessness”? *Music Theory Online*, *24*(3).

Noble, J. (2018b). *Perceptual and Semantic Dimensions of Sound Mass* (Doctoral dissertation, McGill University, Montreal, QC).

Smith, B. (1995). PsiExp: An environment for psychoacoustic experimentation using the IRCAM musical workstation. In D. Wessel (Ed.), Society for music perception and cognition. Berkley, CA: University of California.

Stevens, C. J. (2012). Music perception and cognition: A review of recent cross-cultural research. *Topics in Cognitive Science*, *4*(4), 653-667.

Tenney, J., & Polansky, L. (1980). Temporal gestalt perception in music. *Journal of Music Theory*, *24*(2), 205-241.

Appendix 4A – *True shouga* and *retrograde foil* sequences for each stimulus family

Stimulus Family	True *Shouga* sequence	Retrograde Foil sequence
1	*te e n to n shya n*	*shya a n to n te n*
2	*shya shya ko ro ri chi i to n*	*to chi ri ro ko shya a shya n*
3	*chi n te tsu ko ro ri chi i to n*	*to n chi ri ro ko tsu te e chi n*
4	*shya n te e n shya shya chi i n*	*chi n shya a n shya te shya a n*
5	*to te tsu u u n*	*tsu te to o o n*
6	*shya shya te tsu shya shya te tsu*	*tsu te shya shya tsu te shya shya*
7	*to n ko ro ri n shya n*	*shya n ri ro ko n to n*
8	*tsu n te tsu ko ro ri chi i*	*chi n ri ro ko tsu te tsu u*
9	*tsu n te n chi n chi i n*	*chi n chi n te n tsu u n*
10	*te tsu ko ro ri n*	*ri ro ko tsu te n*
11	*to n te n to n te n tsu n te n*	*te n tsu n te n to n ten to n*
12	*chi tsu te chi tsu n te n*	*te tsu chi te tsu n chi n*
13	*tsu n te tsu shya shya te chi i tsu n te n*	*te n tsu chi te shya shya tsu u te n tsu n*
14	*chi i to n*	*to o chi n*
15	*tsu te tsu te*	*te tsu te tsu*
16	*tsu n ko ro ri chi i to n*	*to n chi ri ro ko o tsu n*
17	*shya n te e n shya shya chi i n to te tsu u u n*	*tsu n te e n to chi shya a n shya te shya a a n*
18	*to n ka ra te to te to te tsu to n*	*to n tsu te to te to te ra ka to n*
19	*tsu n te tsu chi tsu te tsu chi tsu ko*	*ko n tsu chi tsu te tsu chi tsu te tsu*
20	*chi n to te tsu n*	*tsu n te to chi n*
21	*to te chi tsu te tsu chi tsu u*	*tsu chi tsu te tsu chi te to o*
22	*ko ro ri n te n to n te n*	*te to te n ri n ro n ko n*
23	*chi te chi tsu ru tsu u to te to n*	*to te to tsu ru tsu u chi te chi n*
24	*ko ro ri n ko ro ri n*	*ri ro ko n ri ro ko n*
25	*chi tsu te chi i ko ro ri chi i to n*	*to chi ri ro o ko chi te tsu u chi n*
26	*sa ra ri n*	*ri ra sa n*
27	*to n te n to n shya n*	*shya n to n te n to n*
28	*shya n tsu n chi tsu te tsu ko ro ri n shya n*	*shya n ri n ro ko tsu te tsu chi tsu n shya n*
29	*shya n te e n shya shya chi i n to te tsu u u n*	*tsu n te e n to chi shya a n shya te shya a a n*
30	*shya shya te n te n*	*te te shya n shya n*
31	*to n ko ro ri n shya n*	*shya n ri ro ko n to n*
32	*shya n chi n tsu n te n chi i tsu n te n*	*te n tsu n chi n te n tsu u chi n shya n*
33	*chi n to n te n, tsu n tsu n te n*	*te n tsu n tsu n, te n to n chi n*
34	*tsu n tsu u to n ko ro ri n te n*	*te n ri i ro n ko to tsu n tsu n*
35	*chi tsu ko ro ri n chi te tsu n te n*	*te tsu te chi ri n ro ko tsu n chi n*
36	*tsu n te n to ka ra te chi i tsu n te n to n te n*	*te n to n te tsu chi te ra a ka n to n te n tsu n*
37	*te n ko ro ri n te n to n te n to n te n*	*te n to te to n te n ri n ro n ko n te n*
38	*tsu n to n te n chi n tsu n te n*	*ten n tsu n chi n te n to n tsu n*
39	*ko ro ri n te n tsu n te n ko ro ri n ko ro ri n*	*ri ro ko n ri n ro n ko n te tsu te n ri ro ko n*
40	*te chi i ko ro ri n te e n to n shya n*	*shya to o te ri ro n ko o n chi n te n*
41	*shya n te n shya n te n shya n chi i n shya shya te n*	*te n shya n shya n chi n shya n te e n shya te shya n*
42	*shya n chi n te n chi n tsu n te n ko ro ri n shya n*	*shya n ri n ro n ko n te n tsu n chi te chi n shya n*
43	*te n te n to n te n chi n tsu n te n*	*te n tsu n chi n te n to n te n te n*
44	*ko ro ri n te n tsu n ko ro ri n*	*ri ro ko n tsu n te n ri ro ko n*
45	*te to to te to te tsu n to n*	*to tsu te to te to to n te n*
46	*te to te to n ko ro ri n tsu n*	*tsu ri ro ko n to te to n te n*
47	*te to chi to te tsu n ko ro ri n chi n tsu n chi n tsu n shyu*	*shyu tsu chi tsu chi ri n ro ko tsu n te n to n chi n to n te*
48	*ko ro ri n to n*	*to ri ro n ko n*
49	*te n te n tsu to n te n*	*te n to n tsu te n te n*
50	*shya n te e n shya n chi i n shya n tsu u n*	*tsu n shya a n chi n shya a n te n shya a n*
51	*ko ro ri n shya shya te n shya shya ko ro ri n te n*	*te ri ro n ko shya shya n te shya shya ri ro n ko n*
52	*chi n chi n te n chi n ko ro ri n*	*ri n ro n ko n chi n te chi chi n*
53	*shya n chi n tsu n te n tsu n ko ro ri n*	*ri n ro n ko n tsu n te n tsu chi shya n*
54	*shya n te n shya n te n shya n tsu n ko ro ri n*	*ri n ro n ko n tsu n shya n te n shya te shya n*
55	*to to te to te tsu n to n te n*	*te to tsu te to te n to n to n*
56	*to n to n te n to n te n*	*te n to n te n to n to n*
57	*chi n te n chi n tsu n tsu n ko ro ri n tsu n*	*tsu n ri n ro n ko n tsu n tsu chi te n chi n*
58	*ko ro ri n chi i tsu te n to n shya n*	*shya to te n tsu u chi ri n ro n ko n*
59	*shya n chi n te n chi n tsu n tsu n ko ro ri n te n*	*te n ri n ro n ko n tsu n tsu n chi te chi n shya n*
60	*ko ro ri n te n tsu n te n ko ro ri n*	*ri ro ko n te n tsu n te n ri ro ko n*
61	*to n te n to n te n tsu n ko ro ri n tsu n*	*tsu n ri n ro n ko n tsu n te to te n to n*
62	*tsu n te tsu u n te e n sa ra ri n shya n*	*shya n ri ra n sa a n te n tsu te n tsu n*
63	*tsu n tsu u to n ko ro ri n te n*	*tsu n te e to n chi te tsu n ko n*
64	*to n ka ra te to te to te tsu to n*	*to n tsu te to te to te ra ka to n*

Chapter 5 – Conclusion

The present psychological dissertation has concerned the perception of musical *form.* The musical *form* appearing in the mind of the listening subject is amenable to identification, abstract manipulation, and comparison with other more or less similar *forms.* As such this *form* was conceptualized as a property of a *psychological entity,* situated within an intelligible *psychological domain.* The *psychological entity* regarded in the general sense was termed the *idea*; regarded in the specific case of the perception of musical *form*, it was termed the *musical image.* The *psychological domain* was assumed to be in some sense systematically associated with the *physical domain,* in which physical entities are created, measured, and reproduced. The physical entity regarded in the general sense was termed the *material*; regarded in the specific case of the perception of musical *form* it was termed the *musical signal.* According to this conceptualization, systematic changes in *ideas* correspond with systematic changes in *materials,* and vice versa, although *ideas* cannot be reduced to *material* states, nor can *materials* be reduced to *ideal* states.

In Chapter 1, I outlined various theoretical considerations necessary for a scientific investigation of the perception of musical *form.* In particular, this chapter focused on the systematic relations between musical *ideas* (*the musical image*) and musical *materials* (the *musical signal*). To this end I introduced the notions of the *generative* and *representative* functions of the *musical image.* The *representative function* of the *musical image* is familiar to the perceptual sciences in that it describes the capacity of the *musical image* to serve as a *psychological representation* of a *physical object.* I suggest that this function is counterbalanced by a complementary *generative function,* which is the capacity of the *musical image* to serve as a psychological point of reference for the creation of a physical object. That an *idea* can serve as a point of reference indicates that perceiving subjects are capable of differentially orienting themselves to psychological entities, in the same way a subject differentially orients toward a physical entity serving as a point of reference. The *generative function* of the *musical image* is central to the *creative process.* A complete understanding of the perception of musical form requires the integration of both perspectives.

While this notion of *ideas* as generative of *material* states is relatively foreign to the perceptual sciences, it has been highly refined by experts of the creative arts (including the musical arts). Posing a difficulty for the scientist, however, is the fact that the artist is typically less interested in a formal conversation regarding the physical dimensions of art objects; to the artist, the *ideas* and subjective experience associated with these objects possess merit in and of themselves, and little is gained, according to this perspective, by painstakingly associating the precise nuances of a given *idea* with specific physical parameter states. Thus, the artistic understanding of the *generative function* of *musical ideas* is ill-suited to descriptions employing current scientific language, which tends to reduce ideas to *representations* of physical events.

A complete understanding of the perception of musical *form* to be effectively implemented within a scientific discipline, a generalized empirical model—applying equally well to the description of the perceiving subject's interaction with *ideas* and *materials*—must be developed. To this end I developed the notion of *perceptual geometry,* which illustrates the analogous processes (or, arguably, singular process) by which the perceiving subject comes to an intelligible comprehension of *ideal* and *material* forms. The utility of this model is that it allows the scientist to regard the *creative process* as a fundamentally rational activity. Appendix 1A of Chapter 1 provides a neurophysiological exposition of the artist's creative process, while

Appendix 1B couches this notion of "the objective idea" within the larger trajectory of the 20th century empirical sciences.

Such a shift in perspective allows for new lines of empirical research. For example, the physical structure of art objects can be understood to reflect psychological constraints on the perception of *form* (according to the *generative function* of the *idea*). In turn, the subject's perception of *form* when presented with a physical stimulus (according to the *representative function* of the *idea*) can be regarded as an intelligible process, however enigmatic it might be. This proposed interplay between the *generative* and *representative* functions of the *musical image* is explored through the empirical investigations of Chapters 2, 3, and 4.

Chapter 2 provided evidence of Japanese, Western, and non-Japanese non-Western listeners' perceptual sensitivity to the *contour profiles* of traditional Japanese and Western styles. To design the experiment, I consulted with artists, who arrived at a general consensus that one of the salient perceptual differences between Japanese and Western musical forms is that the latter style is "straightened out" as compared with the former. This distinction was operationalized through modifications of the phase relations among the local pitch height maxima and minima, the strong and weak tones of the implied harmony, and the strong and weak beats of the implied meter within monophonic isochronous melodies: Japanese melodies were said to possess a *polyphasic* contour profile, while Western melodies were defined as having a *monophasic* contour profile. Since the Japanese and Western musical traditions also exhibit perceptually salient differences in their respective musical tonalities, each contour profile was presented in both a traditional Japanese scale and a traditional Western scale, producing four stimulus variants [Japanese tonality x Japanese contour profile (JJ), Japanese tonality x Western contour profile (JW), Western tonality x Japanese contour profile (WJ), and Western tonality x Western contour profile (WW)]. Participants of all three listener demographics were randomly presented with these stimulus variants and asked to report their perception of musical style on a 7-point Likert scale, with one extreme corresponding to “Very Japanese,” the other extreme corresponding to “Very Western,” and the intermediate position corresponding with cultural ambivalence. Analyses revealed that participants were sensitive to both the tonality and contour profile manipulations, with all four stimulus variants being rated as perceptually distinct from one another.

These results indicate first and foremost that listeners across cultures and regardless of prior musical training can perceive subtle nuances in the temporal structure of musical sounds.[105] Furthermore, they indicate that an artistic insight, communicated in lay language such as “straightened out,” can be operationally defined and systematically implemented within a scientific investigation of musical form.[106] Finally, these artistic insights are predictive of the musical forms perceived by non-musicians and across cultures—the respective stimuli anticipated to best reflect the Japanese and Western musical styles were in fact perceived as such by naive listeners exposed to these stimuli in random order.[107]

Chapters 3 and 4 investigated a traditional Japanese musical transliteration known as the *shouga.* The *shouga* operates by evoking in the subject an *idea* whose form is analogous to that elicited by the corresponding musical art object.[108] Thus, the psychological form of the *shouga* is experienced in terms of its perceived likeness with the music it represents. The physical structure of the *shouga* can be described as a concatenation of various Japanese speech sounds (*morae*),

105 cf., the *representative function* of the *musical image*
106 cf., the *generative function* of the *musical image*
107 cf., the interplay between the *generative* and *representative functions* of the *musical image*
108 cf., the interplay between the *generative* and *representative* functions of the *musical image*; see description of the creative act presented in the *Perceptual Geometry* section of Chapter 1.

known as *idioms.* Particular *idioms* correspond with particular motifs, each with their respective *ideal forms,* as perceived by the subject. In this way, the temporal sequencing of the idioms within the *shouga* corresponds with the temporal sequencing of physical parameter states within the music, each experienced by the subject as exhibiting a common *form.* As such, this device can be understood as a protolinguistic representation of the *perceptual geometry of ideal forms* experienced by the perceiving subject, thereby providing researchers with an unprecedented opportunity to study the perception of musical form. The *shouga* should prove as an invaluable resource as Western music perception research turns to dynamic and spatialized investigations of "musical shape cognition" (Godøy, 2019).

To this end, Chapter 3 presented three corpus-analytic investigations of the structural correspondence between the *shouga* and melody of the seminal *soukyoku* (koto music) work, *Rokudan no Shirabe* by YATSUHASHI Kengyo. Study 1 presented a replication and extension of the pioneering work by David Hughes (2000), demonstrating a structural correspondence between the vowel components of the *morae* within the *shouga* and the rising and falling pitch contours of the musical melody. Study 2 presented a series of exploratory analyses in search of additional indices of structural correspondence between the *shouga* and the melody of *Rokudan no Shirabe.* Critically, these results indicated: 1) that the use of Japanese speech sounds within the *shouga* is highly constrained—both in terms of the *morae* used and the way in which they are sequenced—, 2) that the *idioms* of the *shouga* correspond with constrained relative interval sequences within the melody, which can be transposed throughout the prevailing pitch space of the music, 3) that the consonant components of the *morae* within the *shouga* play a critical role in the articulation of musical form, and 4) that movement from one *mora* to the next within the *shouga* is highly constrained in terms of its association with the direction and magnitude of the pitch shifts (intervals) within the melody. Study 3 extended this exploratory analysis by providing a series of parametric analyses of the pitch, interval, scale degree, rhythm, and metric position states associated with *shouga* idioms identified by expert practitioners. The parametric distributions of each idiom were clearly distinct, although the ways in which the parameter states of a given idiom were distinguished from others were highly dynamic and contingent on case-by-case comparisons. This result indicates that the structural relations between *shouga* and melody are dynamic as a function of the common *form* they represent.

Together the studies of Chapter 3 indicate a remarkable structural correspondence between the *shouga* and melody of *Rokudan no Shirabe.* The full extent of this correspondence can only be intimated on the basis of the results produced by the limited conceptual and empirical tools employed within the current research. Extensive research, and the corresponding development of new empirical concepts and techniques, will be required to grapple with this highly refined system of musical transliteration. Nevertheless, the present studies provide an initial proof of concept that the *shouga* is a systematic and intelligible articulation of musical *form*. As a final step of this proof of concept, Chapter 4 investigated the extent to which this articulation corresponds with the perception of musical *form* experienced by naïve listeners.

On each trial of the experiment reported in Chapter 4, naïve listeners[109] were presented with three stimuli. The first was the reference stimulus, an original excerpt of a koto performance of *Rokudan no Shirabe*. The other two were the test stimuli: One was a vocal rendition of the *true shouga* sequence taken from the corpus of *Rokudan no Shirabe.* The other was a vocal rendition of a *retrograde foil,* which maintained but re-sequenced the speech sounds of the *true shouga* sequence. On each trial, the participant was asked to identify which of the two vocal

[109] unable to speak and with minimal exposure to the Japanese language and with no prior training in traditional Japanese music

renditions exhibited a “sound shape” more like that of the *original excerpt.* If, according to the null hypothesis, the *shouga* is an arbitrary or culturally conditioned articulation of musical form, then naive listeners should have identified the *true shouga* and *retrograde foil* sequences equally as often, performing at chance rates of discrimination. In fact, 37 of 40 participants (92.5%) identified the *true shouga* sequence more often than its *retrograde foil* counterpart, with the average rate of correct identification across all participants being > 67% (i.e., the average naïve listener correctly detected the *true shouga* sequence on two out of every three trials). Likewise, 58 of the 64 *true shouga* stimuli (90.6%) were correctly associated with their respective *original excerpt* by the majority of participants, with the average number of participants correctly identifying a given *true shouga* sequence being > 67% (i.e., the average *true shouga* sequence was correctly identified by 2 out of every 3 naïve listeners). These proportions were all highly significant according to the binomial distribution of chance performance predicted by the null hypothesis.

The results provide an unequivocal indication that the *shouga* is an intelligible representation of *musical form.* As noted in the discussion of Chapter 4, most participants were intrigued by the experiment because they felt they could intuit an association between the form of the *shouga* and the form of the original music, although they could not articulate why this was so, and very few generated an explicit set of guiding principles to make their judgments. This experience is highly reminiscent of the difficulty artists have in expressing their *ideas* or creative process in explicit language. And yet the results of this experiment indicate that regardless of this difficulty, the expert artist is capable of cultivating physical objects (e.g., the music and the *shouga* of *Rokudan no Shirabe*) that possess perceptually common *forms* as directly experienced by the perceiving subject, and that this experience is accessible to untrained listeners who have not themselves cultivated a particular creative technique.[110]

Taken together the present work emphasizes the need to consider both the creative and receptive processes that underlie the perception of musical *form*. This consideration entails a recognition in the perceptual sciences of what I have termed the *generative function* of ideas, such as *musical images*. Such a recognition results in the conceptualization of the *percept* as more than a mere *representation* of physical events: The *ideas* perceived by the conscious subject are both representative and generative of *material* objects. Indeed, the empirical results presented in Chapters 2–4 cannot be reasonably interpreted without this theoretical concession. In this sense, the present work is all but preliminary. Ideally, it will contribute to the materialization of a greater psychological theory of the creative process and the perception of musical *form*.

[110] cf., the interplay between the *generative* and *representative functions* of the *musical image*

References

Addessi, A.R., & Caterina, R. (2000) On segmentation of post-tonal music. In Woods, C., Luck, G.B., Brochard, R., O'Neill, S. A., and Sloboda, J. A. (Eds.) *Proceedings of the Sixth International Conference on Music Perception and Cognition.* Keele, Staffordshire, UK: Department of Psychology. CD-ROM.

Adriaansz, W. (1973). *The Kumiuta and Danmono traditions of Japanese koto music*. Los Angeles, CA: University of California Press.

Alías, F., Socoró, J., & Sevillano, X. (2016). A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds. *Applied Sciences*, *6*(5), 1–44.

Ando, M. (1986). 生田流の箏曲 (*Koto of the Ikuta School*). H. Yoshikawa (Ed.). Tokyo, JP: Kodansha.

Ando, M. (1989). Koto scales and tuning. *Journal of the Acoustical Society of Japan (E)*, *10*(5), 279-287.

Ando, M. (2017). Personal communication. Jun–Dec, 2017. Setagaya, JP.

Ayari, M., & McAdams, S. (2003). Aural analysis of Arabic improvised instrumental music (taqsīm). *Music Perception, 21*(2), 159-216.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211-227.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617-645.

Bergson (1907/1911). *Creative evolution* (A. Mitchell, Trans.). New York, NY: Henry Holt and Company.

Bigand, E. (1990). Abstraction of two forms of underlying structure in a tonal melody. *Psychology of Music*, *18*(1), 45-59.

Bernstein (1976). Music. In Ratcliffe, S. (Ed.), *Oxford Essential Quotations*. Retrieved 13 Aug. 2018, from https://www.oxfordreference.com/view/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00007743.

Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nature Neuroscience*, *2*(4), 382–387.

Boasen, J., Takeshita, Y., Kuriki, S., & Yokosawa, K. (2018). Spectral-spatial differentiation of brain activity during mental imagery of improvisational music performance using MEG. *Frontiers in Human Neuroscience*, *12*(156), 1–12.

Bohr, N. (1960). *Atomic physics and human knowledge.* New York, NY: John Wiley & Sons.

Bonin, T., Degtyareva, K., Hahn, E. (2019). Investigating the content and form of musical imagery evoked by excerpts of Mahler's 5th symphony. Paper presented at the CIRMMT-OICRM-BRAMS Symposium. May 9–10, 2019. Montreal, CA.

Bonin, T., Matsunaga, R. & McAdams, S. (in preparation). Cross-cultural perception of Japanese- and Western-typical melodic contour profiles.

Bonin, T. & McAdams, S. (in preparation). 箏の唱歌. Structural relationships between the *shouga* and melody of *Rokudan no Shirabe* (六段の調).

Bonin, T., & Smilek, D. (2016). Inharmonic music elicits more negative affect and interferes more with a concurrent cognitive task than does harmonic music. *Attention, Perception, & Psychophysics*, *78*(3), 946-959.

Bonin, T., Trainor, L. J., Belyk, M., & Andrews, P. W. (2016). The source dilemma hypothesis: Perceptual uncertainty contributes to musical emotion. *Cognition*, *154*, 174-181.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT press.

Brown, D. (2015). Richard Wagner. *The Esoteric Quarterly, 10*(4), 55–83.

Butler, D. (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), 219-241.

Butler, D. (1990a). A study of event hierarchies in tonal and post-tonal music. *Psychology of Music*, *18*(1), 4-17.

Butler, D. (1990b). Response to Carol Krumhansl. *Music Perception*, *7*(3), 325-338.

Cabanac, M. (1996). On the origin of consciousness, a postulate and its corollary. *Neuroscience & Biobehavioral Reviews*, *20*(1), 33-40.

Cabanac, M. (1999). Emotion and phylogeny. *Journal of Consciousness Studies*, *6*(6-7), 176-190.

Campbell, J. (1949). *The hero with a thousand faces.* New York, NY: Pantheon Books.

Chaudhury, P. J. (1965). The theory of rasa. *The Journal of Aesthetics and Art Criticism*, *24*(1), 145-149.

Cheyne, J. A., Bonin, T., Wright, C., Carriere, J. S., Danckert, J., & Smilek, D. (2016). "You're on ten, where can you go from there?" Tufnel problems in repeated experiential judgments. *Consciousness and Cognition*, *42*, 311-324.

Colman, A. M. (2009). *A dictionary of psychology*. Oxford, UK: Oxford University Press.

Condo, G. (2017). *The way I think.* Interview for Louisiana Museum of Modern Art. Soho, New York.

Cosmides, L., & Tooby, J. (1994). Better than rational: Evolutionary psychology and the invisible hand. *The American Economic Review*, *84*(2), 327-332.

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. *Handbook of Emotions*, *2*(2), 91-115.

Cuddy, L. L., Cohen, A. J., & Mewhort, D. J. K. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(4), 869-883.

Davis, S. (2006). Implied polyphony in the solo string works of JS Bach: A case for the perceptual relevance of structural expression. *Music Perception*, *23*(5), 423-446.

Deliège, I. (1989). A perceptual approach to contemporary musical forms. *Contemporary Music Review*, *4*(1), 213-230.

Descartes, René (1644/1983). *Principia philosophiae* (V. Rodger & R.P. Miller., Trans.). Dordrecht, NL: Reidel.

Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review*, *3*(4), 357-370.

Douglas, C., Noble, J., & McAdams, S. (2016). Auditory scene analysis and the perception of sound mass in Ligeti's continuum. *Music Perception*, *33*(3), 287-305.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*(4), 341-354.

Dowling, W. J. (1999). The development of music perception and cognition. In D. Deutsch (Ed.), *The psychology of music* (2nd Ed.) (pp. 603-625). San Diego, CA: Academic Press.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *The Journal of the Acoustical Society of America*, *49*(2B), 524-531.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition*. San Diego, CA: Academic Press.

Dowling, W. J., Tillman, B., & Ayers, D. F. (2001). Memory and the experience of hearing music. *Music Perception, 19*(2), 249-276.

Drake, C. & El Heni, J. B. (2003). Synchronizing with music: Intercultural differences. *Annals of the New York Academy of Sciences*, *999*(1), 429-437.

Eitan, Z. & Granot, R. Y. (2006). How music moves: Musical parameters and listeners images of motion. *Music Perception*, *23*(3), 221-248.

Eiting, M. H. (1984). Perceptual similarities between musical motifs. *Music Perception*, *2*(1), 78-94.

Farbood, M. M. (2012). A parametric, temporal model of musical tension. *Music Perception*, *29*(4), 387-428.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*(2), 181-195.

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, *100*(1), 173-215.

Galileo (1623/1957) *The assayer.* In S. Drake (Ed.), *Discoveries and opinions of Galileo.* New York, NY: Doubleday.

Gendlin, E.T. (1970). The significance of felt meaning. In R. Cormier, E. Chinn & R.H. Lineback (Eds.), *Encounter: An introduction to philosophy* (pp. 561–566). Glenview, IL: Scott, Foresman & Co.

Gibson, J. J. (1960). The concept of the stimulus in psychology. *American Psychologist*, *15*(11), 694-703.

Gibson, J. J. (1975). Events are perceivable but time is not. In J. T. Fraser & N. Lawrence (Eds.), *The study of time II* (pp. 295-301). New York, NY: Springer-Verlag.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin Harcourt.

Gill, S. P. (2015). *Tacit engagement: Beyond interaction.* New York, NY: Springer Cham.

Giolai, A. (2019). Sensing the music: oral mnemonics as a technique of affective sensitization in Japanese "court music". *Asian Anthropology*, 1-19. doi: 10.1080/1683478X.2019.1629472

Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für mathematik und physik*, *38*(1), 173-198.

Gödel (1961/1995). The modern development of the foundations of mathematics in light of philosophy. In S. Feferman, J. Dawson, S. Kleene, G. Moore, R. Solovay & J. van Heihenoort (Eds.), *The collected papers of Kurt Gödel. III: Unpublished essays and lectures.* Oxford, UK: Oxford University Press.

Godøy, R. I. (2019). Musical shape cognition. In M. Grimshaw-Aagaard, M. Walther-Hansen & M. Knakkergaard (Eds.), *The Oxford Handbook of Sound and Imagination II* (pp. 237–258). New York, NY: Oxford.

Good, M. (2013). Interview for CJXY FM radio, June 27, 2013. Hamilton, ON.

Gray, J. A. (1982). Précis of The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system. *Behavioral and Brain Sciences*, *5*(3), 469-484.

Granot, R. Y., & Eitan, Z. (2011). Musical tension and the interaction of dynamic auditory parameters. *Music Perception*, *28*(3), 219-246.

Halliwell, P. (1994). Learning the koto. *Canadian University Music Review/Revue de musique des universités canadiennes*, *14*, 18-48.

Hauser, M. D., & McDermott, J. (2003). The evolution of the music faculty: A comparative perspective. *Nature Neuroscience*, *6*(7), 663-668.

Hofstadter, D., & Sander, E. (2013). *Surfaces and essences: Analogy as the fuel and fire of thinking*. New York, NY: Basic Books.

Hughes, D. W. (1989). The historical uses of nonsense: Vowel-pitch solfège from Scotland to Japan. In M Philipp (Ed.), *Ethnomusicology and the historical dimension* (pp. 3–18). Ludwigsburg, Germany: Philipp Verlag.

Hughes, D. W. (1991). Oral mnemonics in Korean music: Data, interpretation, and a musicological application. *Bulletin of the School of Oriental and African Studies*, *54*(2), 307-335.

Hughes, D. W. (2000). No nonsense: The logic and power of acoustic-iconic mnemonic systems. *British Journal of Ethnomusicology*, *9*(2), 93-120.

Huron, D. (1994). Interval-class content in equally tempered pitch-class sets: Common scales exhibit optimum tonal consonance. *Music Perception*, *11*(3), 289-305.

Huron, D. (2001). Is music an evolutionary adaptation?. *Annals of the New York Academy of Sciences*, *930*(1), 43-61.

Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *The Journal of the Acoustical Society of America*, *124*(4), 2263-2271.

James, W. (1909). *The meaning of truth*. Mineola, NY: Dover Publications.

Jung, C. G. (1922/1923). *Psychological types: Or the psychology of individuation.* (H.G. Baynes, Trans.). New York, NY: Harcourt, Brace & Company.

Jung, C. G. (1944/1968). *Psychology and alchemy* (R.F.C. Hull, Trans). Princeton, NJ: Princeton University Press.

Jung, C. G. (1951/1959). *Aion: Researches into the phenomenology of the self.* (R.F.C. Hull, Trans.). Princeton, NJ: Princeton University Press.

Jung, C. G. (1956/1970). *Mysterium coniunctionis: An inquiry into the separation and synthesis of psychic opposites in alchemy*. (R.F.C. Hull, Trans.). Princeton, NJ: Princeton University Press.

Kant, I. (1783/1902). *Prolegomena to any future metaphysics that will be able to present itself as a science*. (P. Carus, Trans.). Chicago, IL: Open Court Publishing Company.

Koelsch, S. (2009). Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. *Psychophysiology*, *46*(1), 179-190.

Köhler, W. (1947/1970). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York, NY: W.W. Norton & Company.

Krumhansl, C. L. (1990). Tonal hierarchies and rare intervals in music cognition. *Music Perception, 7*(3), 309-324.

Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception*, *13*(3), 401-432.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*(4), 334-368.

Laban, R. (1960). *The mastery of movement.* London, UK: MacDonald & Evans.

Lakoff, G. (1987). *Women, fire, and dangerous things.* Chicago, IL: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119-159.

Lecoq, J. (2001). *The moving body (Le corps poétique): Teaching creative theatre.* New York, NY: Routledge

Lerdahl, F., & Jackendoff, R. (1983). An overview of hierarchical structure in music. *Music Perception*, *1*(2), 229-252.

Levenson, R. W. (1999). The intrapersonal functions of emotion. *Cognition & Emotion*, *13*(5), 481-504.

Lévy-Bruhl, L. (1938). *L'expérience mystique et les symboles chez les primitifs*. Paris, FR: Alcan.

Liew, K., Lindborg, P., Rodrigues, R., & Styles, S. J. (2018). Cross-modal perception of noise-in-music: Audiences generate spiky shapes in response to auditory roughness in a novel electroacoustic concert setting. *Frontiers in Psychology*, 9:178. doi: 10.3389/fpsyg.2018.00178

Liu, C. (2013). Just noticeable difference of tone pitch contour change for English-and Chinese-native listeners. *The Journal of the Acoustical Society of America, 134*(4), 3011-3020.

Locke, J. (1690). *An essay concerning human understanding. Book II: Ideas.* Retrieved from https://www.earlymoderntexts.com/assets/pdfs/locke1690book2.pdf. Accessed July 1, 2019.

Mach, E. (1906/2011). *Space and geometry: In the light of physiological, psychological, and physical inquiry* (T.J. McCormack, Trans.). Ocean Shores, WA: Watchmaker Publishing.

Malm, W. (1983). Some of Japan's musics and musical principles. In E. May (Ed.), *Musics of many cultures: An introduction* (pp. 46-62). Los Angeles, CA: University of California Press.

Malm, W.P. (2015). Japanese music. In *Encyclopedia Britannica.* Retrieved from https://www.britannica.com/art/Japanese-music. Accessed July 15, 2019.

Mananoi, Y. & Akira, S. (1978). 口唱歌大系,日本 の楽器のソルミゼーション (*Kuchishouga taikei: Solmization of Japanese instruments*). Tokyo, JP: CBS-Sony.

Margolis, E. & Laurence, S. (2018). Concepts, *The Stanford Encyclopedia of Philosophy,* Edward N. Zalta (ed.), Retrieved from https://plato.stanford.edu/archives/sum2019/entries/concepts/. Accessed May 1, 2018.

Margolis, E. & Laurence, S. (2015). *The conceptual mind: New directions in the study of concepts*, Cambridge, MA: MIT Press.

Marks, L. E. (2000). Synesthesia. In E. A. Cardeña, S. J. Lynn, & S. C. Krippner (Eds.), *Varieties of anomalous experience: Phenomenological and scientific foundations*. Washington, DC: American Psychological Association.

Matsunaga, R., Yokosawa, K., & Abe, J. I. (2012). Magnetoencephalography evidence for different brain subregions serving two musical cultures. *Neuropsychologia*, *50*(14), 3218-3227.

Matsunaga, R., Yokosawa, K., & Abe, J. I. (2014). Functional modulations in brain activity for the first and second music: a comparison of high-and low-proficiency bimusicals. *Neuropsychologia*, *54*, 1-10.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data* (2nd ed.). New York: Routledge.

Mayer, J. (2016, November). *JoJo Mayer - PASIC16.* Workshop presented at PASIC 2016. Indianapolis, IN.

McAdams, S. (1984). The auditory image: A metaphor for musical and psychological research on auditory organization. In W. R. Crozier & A. J. Chapman (Eds.), *Cognitive processes in the perception of art* (pp. 289-323). Amsterdam, NL: North-Holland.

McAdams, S. (1987). Music: A science of the mind? *Contemporary Music Review*, *2*(1), 1-61.

McAdams, S. (1989). Psychological constraints on form-bearing dimensions in music. *Contemporary Music Review*, *4*(1), 181-198.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*(3), 177-192.

McGilchrist, Iain (2009). *The master and his emissary.* New Haven, CT: Yale University Press.

McGuire, P. K., Silbersweig, D. A., Murray, R. M., David, A. S., Frackowiak, R. S. J., & Frith, C. D. (1996). Functional anatomy of inner speech and auditory verbal imagery. *Psychological medicine*, *26*(1), 29-38.

Meier, C. A. (Ed.). (2001). *Atom and archetype: The Pauli Jung letters, 1932-1958* (Vol. 631). Princeton, NJ: Princeton University Press.

Meyer, L. B. (1996). *Style and music: Theory, history, and ideology*. Chicago, IL: University of Chicago Press.

Mittledorf, J. (2003). *Probability of duplicate pairs.* Retrieved from http://mathforum.org/library/drmath/view/62941.html. Accessed June 1, 2019.

Mungan, E., Yazıcı, F. & Kaya, M. U. (2017). Perceiving boundaries in unfamiliar Turkish makam music: Evidence for gestalt universals? *Music Perception*, *34*(3), 267-290.

Näätänen, R. & Gaillard, A.W.K. (1983). The orienting reflex and the N2 deflection of the event-related potential (ERP). In A.W.K. Gaillard & W. Ritter (Eds.), Tutorials in event related potential research: Endogenous components (pp. 119-141). Amsterdam: North Holland.

Nelson, S. G. (2008a). Court and religious music (1): History of gagaku and shōmyō. In A. M. Tokita and D. W. Hughes (Eds.), *The Ashgate research companion to Japanese music* (pp. 35–48). Farnham, UK: Ashgate.

Nelson, S. G. (2008b). Court and religious music (2): Music of gagaku and shōmyō. In A. M. Tokita and D. W. Hughes (Eds.), *The Ashgate research companion to Japanese music* (pp. 49–76). Farnham, UK: Ashgate.

Neumann, E. (1949/1970). *The origins and history of consciousness* (R.F.C. Hull, Trans.). Princeton, NJ: Princeton University Press (Bollingen)

Neumann, E. (1955/1972). *The great mother* (R. Mannheim, Trans.). Princeton, NJ: Princeton University Press.

Neumann, E. (1959). *Art and the creative unconscious.* London, UK: Routledge & Kegan Paul.

Noble, J. D. (2018a). What can the temporal structure of auditory perception tell us about musical "timelessness"? *Music Theory Online*, *24*(3).

Noble, J. (2018b). *Perceptual and semantic dimensions of sound mass* (Doctoral dissertation, McGill University, Montreal, QC).

Paraskeva, S., & McAdams, S. (1997, September). Influence of timbre, presence/absence of tonal hierarchy and musical training on the perception of musical tension and relaxation schemas. In T. Rikakis (Ed.), *Proceedings of the 1997 International Computer Music Conference* [CD-ROM]. Thessaloniki, Greece: Aristotle University.

Peterson, J. B. (1999). *Maps of meaning: The architecture of belief.* London, UK: Routledge.

Peterson, J. B. (2000). *The pragmatics of meaning.* Retrieved from https://semioticon.com/frontline/jordan_b.htm. Accessed June 1, 2018.

Peterson, J. B., & DeYoung, C. G. (2000). Metaphoric threat is more real than real threat. *Behavioral and Brain Sciences*, *23*(6), 992-993.

Piaget, J. (1971a). *Biology and knowledge: An essay on the relations between organic regulations and cognitive processes* (B. Walsh, Trans.). Chicago, IL: University of Chicago Press.

Piaget, J. (1971b). *The theory of stages in cognitive development*. In D.R. Green, M.P. Ford, & G.B. Flamer (Eds.), *Measurement and Piaget* (pp. 1–11). New York, NY: McGraw-Hill.

Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures* (A. Rosin, Trans.). New York, NY: Viking.

Peng, G., Zheng, H. Y., Gong, T., Yang, R. X., Kong, J. P., & Wang, W. S. Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics,* 38(4), 616-624.

Polanyi, M. (1966). *The tacit dimension.* Chicago, IL: The University of Chicago Press.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3-25.

Razran, G. (1961). The observable and the inferable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychological Review*, *68*(2), 81-147.

Rao, P., Ross, J. C., Ganguli, K. K., Pandit, V., Ishwar, V., Bellur, A., & Murthy, H. A. (2014). Classification of melodic motifs in raga music with time-series matching. *Journal of New Music Research*, *43*(1), 115-131.

Rose, F. & Kapuscinski, J. (2013). Orchestration in gagaku music. Retrieved from https://ccrma.stanford.edu/groups/gagaku/theory/pitch-en.html. Accessed May 1, 2019.

Rose, F. & Kapuscinski, J. (2014). Japanese traditional orchestral music: The correlation between time and timbre. Retrieved from http://www.jaroslawkapuscinski.com/pdf/japanese-traditional-orchestral-music.pdf. Accessed May 1, 2019.

Rubia, K., Smith, A. B., Brammer, M. J., & Taylor, E. (2003). Right inferior prefrontal cortex mediates response inhibition while mesial prefrontal cortex is responsible for error detection. *Neuroimage*, *20*(1), 351-358.

Rubia, K., Smith, A. B., Brammer, M. J., Toone, B., & Taylor, E. (2005). Abnormal brain activation during inhibition and error detection in medication-naive adolescents with ADHD. *American Journal of Psychiatry*, *162*(6), 1067-1075.

Rumsfeld, D. H. (2002). *DoD news briefing – Secretary Rumsfeld and Gen. Myers.* February 12, 2002, 11:30 AM EDT. Retrieved from https://archive.defense.gov/Transcripts/Transcript.aspx?TranscriptID=2636. Accessed Aug 13, 2019.

Russell, B. (1967). *Autobiography*. Oxford, UK: Routledge.

Russo, F. A., & Thompson, W. F. (2005). An interval size illusion: The influence of timbre on the perceived size of melodic intervals. *Perception & Psychophysics*, *67*(4), 559-568.

Schaeffer, P. (1966/2017). *Treatise on musical objects: An essay across disciplines.* (J. Dack & C. North, Trans.). Oakland, CA: University of California Press.

Scheirer, E. D. (2000). *Music-listening systems* (Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA).

Schmuckler, M. A. (2010). Melodic contour similarity using folk melodies. *Music Perception*, *28*(2), 169-194.

Schubert, E., Murari, M., Rodà, A., Canazza, S., Da Pos, O., & De Poli, G. (2019). Verbal and cross-modal ratings of music: Validation and application of an icon-based rating scale. *i-Perception*, *10*(3), 1–21. doi: 10.1177/2041669519852643

Sears, David R. W. 2008. Processing surprise tension in tonal melodies. *Inquiry Research Journal, 9,* 16–23.

Seligman, P. (1974). *Being and not-being: An introduction to Plato's Sophist*. The Hague, NL: Nijhoff.

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America,* 1964, *36,* 2346-2353.

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89, 305–333.

Shove, P., & Repp, B. H. (1995). Musical motion and performance: Theoretical and empirical perspectives. In J. Rink (Ed.) *The practice of performance: Studies in musical interpretation* (pp. 55-83). Cambridge, UK: Cambridge University Press.

Smith, B. (1995). PsiExp: An environment for psychoacoustic experimentation using the IRCAM musical workstation. In Society for music perception and cognition (D. Wessel, Ed.). Berkley, CA: University of California Press.

Spencer-Brown, G. (1969). *Laws of form*. London, UK: Allen & Unwin.

Stetson, C., Fiesta, M. P., & Eagleman, D. M. (2007). Does time really slow down during a frightening event? *PloS One*, *2*(12), e1295.

Stevens, C. J. (2012). Music perception and cognition: A review of recent cross-cultural research. *Topics in Cognitive Science*, *4*(4), 653-667.

Tenney, J., & Polansky, L. (1980). Temporal gestalt perception in music. *Journal of Music Theory*, *24*(2), 205-241.

Tokita, A. (2014). Bi-musicality in modern Japanese culture. *International Journal of Bilingualism*, *18*(2), 159-174.

Tokita, A., & Hughes, D. W. (2008). *The Ashgate research companion to Japanese music*. Farnham, UK: Ashgate.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J.H. Barkow, L. Cosmides, J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (19–136). New York, NY: Oxford University Press.

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development*, 821-830.

Trehub, S. E., & Hannon, E. E. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition*, *100*(1), 73-99.

van Kranenburg, P., Volk, A., & Wiering, F. (2013). A comparison between global and local features for computational classification of folk song melodies. *Journal of New Music Research*, *42*(1), 1-18.

Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception & Psychophysics*, *64*(2), 198-207.

Whitaker, A. K. (Ed.). (1996). *Plato's Parmenides*. Indianapolis, IN: Hackett.

Whitehead, A. N. & Russell, B. (1910), *Principia mathematica,* Volume 1. Cambridge, UK: Cambridge University Press.

Whitehead, A. N. & Russell, B. (1912), *Principia mathematica,* Volume 2. Cambridge, UK: Cambridge University Press.

Whitehead, A. N. & Russell, B. (1913), *Principia mathematica,* Volume 3. Cambridge, UK: Cambridge University Press.

Whitehead, A. N. (1925). *Science and the modern world*. Cambridge, UK: Cambridge University Press.

Whitehead, A.N. (1929). *Process and Reality. An Essay in Cosmology.* Cambridge, UK: Cambridge University Press.

Wigner, E. P. (1961), Remarks on the mind-body question. In: I. J. Good (Ed.), *The Scientist Speculates* (pp. 171–184). London, UK: Heinemann

Wittgenstein, L. (1921/1922). *Tractatus logico-philosophicus* (C.K. Ogden, Trans.). London, UK: Routledge & Kegan Paul.

Zbikowski, L. M. (1997). Conceptual models and cross-domain mapping: New perspectives on theories of music and hierarchy. *Journal of Music Theory*, *41*(2), 193-225.

Zhang, Y., Chen, G., Wen, H., Lu, K. H., & Liu, Z. (2017). Musical imagery involves Wernicke's area in bilateral and anti-correlated network interactions in musicians. *Scientific Reports*, *7*(1), 7:17066. doi: 10.1038/s41598-017-17178-4

Zuckerkandl, V. (1965/1973). *Sound and symbol: Music and the external world.* Trans. W. R. Trask. Princeton, NJ: Princeton University Press.