DISSONANCE THEORY OF SOUND OBJECTS

Colby Nelson Leider

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDICACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF MUSIC

January 2007

ABSTRACT

With the central compositional role it places on timbre, electroacoustic music poses new questions and problems in existing definitions of musical dissonance. However, no comprehensive musical theory of dissonance as it relates to sound objects themselves or their syntactical constructs has yet been proposed. While the theory described in this dissertation is not comprehensive by any means, it attempts to paint a skeletal model of one through both musical and technical approaches.

This dissertation first proposes the concept that individual sound objects can be considered relatively consonant or dissonant, just as theorists have historically described dyads (and only more recently, chords) comprised of distinct pitches, by examining historical writings on dissonance and projecting linkages to the electroacoustic medium. Furthermore, it proposes a rudimentary musical model—not a mathematical or cognitivist one, of which several already do exist—of "timbral consonance" defined in terms of compositional aspects of electro-acoustic music. In a manner similar to how purely pitch-based notions of dissonance has informed the syntax of tension and release in tonal music, so has the timbral counterpoint of "consonant" and "dissonant" sound objects, if at only an intuitive level.

Next, the concept of quantifying dissonance—both in historical context and current theoretical practice—is addressed, along with the extent to which objectification of dissonance is musically informative for analysis of particularly non-notated music. Subsequently, the dissertation formally defines the term “sound object,” motivating the necessity of viewing musical dissonance as an inherent property of a sound object, and it argues that sound objects themselves naturally fall onto multidimensional axes of consonance and dissonance with respect to particular physical and perceptual properties. A battery of listening tests that were conducted to test the proposed theory are then described and analyzed to attempt to find physical properties of sound and psychoacoustic factors that may influence the perception of sound-object dissonance.

As a supplement to this dissertation, my composition *Tilt* for 7.1-channel computer playback with optional live electronics is included on an enclosed DVD. The work was commissioned in 2002 by the International Computer Music Association, and, while a separate project from this essay, is included here as a supplement. The DVD contains program notes and data files containing audio tracks, along with a two-channel mixdown.

ACKNOWLEDGMENTS

Thank you to Princeton University for their support of this work and providing an incredible, wonderful, and stimulating environment while I lived and studied there for four unforgettable years. Thank you to my advisors Perry Cook and Paul Lansky, who read and commented this essay along the way, allowing me to pester them at conferences and festivals over the years after I moved away from Princeton. And thank you to all of the great folks I met while a student there.

To my good friends Paul Koonce and Alan Shockley, thank you for extremely helpful observations, ideas, and many late-night phone calls.

Thank you to my parents, Dr. Fred Leider and Jaci Leider, and my grandparents, Ben and Peggy Alley, for everything.

I am also indebted to the University of Miami's Frost School of Music and our music engineering students for their support and for participating in the listening tests described in Chapter 6. In particular, thank you to my colleagues Ken Pohlmann and Joe Abatti for their help and insights.

To my wife, Dr. Kristine H. Burns, and children Liam and Livya, thank you for putting up with me for eight years, five years, and one year, respectively, while I wrote this essay. Thank you especially for your support and understanding while my mind occasionally (well, often) wandered elsewhere to finish this work.

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

1 (CONSONANCE – DISSONANCE) ≠ (DISSONANCE – CONSONANCE)

That which is, is good.

—Thomas Aquinas (1225–1274)
Summa Theologica I, Question V, Article 1

If it sounds good, it is good!

—Duke Ellington (1899–1974)

The ultimate rule ought to be: if it sounds good to you, it's bitchin'; and if it sounds bad to you, it's shitty.

—Frank Zappa (1940–1993)

The sheer number of historical writings that address issues related to the tuning of musical scales—and, by implicit or explicit inference, the concept of "dissonance" of some kind—underscores our omnipresent interest in the inner functionality of music and emphasizes the importance of these facets in our thinking about musical experience. Volumes of music theory treatises from the

Middle Ages extensively debate the relative "levels" of dissonance and consonance that should be properly ascribed to various intervals. Curiously, little progress was made in our fundamental approach to understanding musical dissonance for some time, at least until Hermann von Helmholtz's late-nineteenth century studies. Even today, our understanding of the quantitative and qualitative musical, neurological, and psychological aspects of the constitution of dissonance—and as importantly, the ways in which these aspects interact—is tenuous at best.

Part of the reason for the tenuousness of this understanding is the detailed levels of study in such seemingly diverse fields as psychology, music theory, music composition, cognitive science, and neurology that are required to formulate a comprehensive of the myriad of roles dissonance plays in music. And because musical dissonance in general is a psychacoustic correlate of many physical phenomena, working together in concert, there is no one "correct" answer—and thus comprehensive, quantitative analysis tool—as we will see shortly.

The issue is also further compounded by the multitude of levels of examination possible when analyzing a musical work. For example, one might examine the dissonance structure posed by a particular combination of chords in the string section of an orchestral piece. Alternatively, one may wish to examine a dissonance structure that is posed by one interval in particular when played by a flute and an oboe by using the many tools offered by digital signal processing. In the former example, we are looking at notes only, and this is the primary way that musical dissonance has been explored in the past. "A major triad is consonant," we might say. In the latter example, we are "zooming in" with the microscope afforded by analysis of digital audio signals to examine the time- and frequency-domain properties of sound samples. Here, we might instead say, "The resulting interval is consonant owing to the relative harmonicity of the overtone structures of its constituent timbres and the mutual congruence of the feature vectors of the individual tones." Equipped with algorithmic transcription capabilities (for example, the ability to detect pitches and rhythms), the latter approach forms a more robust—and until recently relatively unexplored—manner in which to examine musical dissonance. The latter approach is the subject of this essay.

❧

This discussion is based on the central premise that music, as with any time-based medium, can only do one of three things to a listener's/viewer's attentive state at any given moment: (1) focus it, or increase attention; (2) leave it unchanged; or (3) blur it, or decrease attention. The same could be said for other aspects of the receiver's mental state, for example, perception of tension and release, stability and instability, familiarity and unfamiliarity, causality and non-causality, suspense and resolution. Said another way, time-based art can do one of three things at any point in time:

And because a work can do any of these things at any point in time, it can also do the same at any scale in time—on a local level as well as a global level.

Theories of dissonance have historically formed a significant compositional tool in the creation and ordering of such trajectories.

❧

The electro-acoustic medium, with the central musical role it places on timbre, poses new questions and problems when discussing the concept of dissonance. Is a particular *sound* inherently more musically dissonant than another sound? Can sound itself be characterized as dissonant, rather than just combinations of representational objects (e.g., notes on a staff)? How do certain sound objects interact in electro-acoustic music to create a perceivable, syntactical gestalt? In this thesis, I offer a partial answer these questions through examinations of literature from several disciplines and analysis of selected musical compositions.

I will first discuss the ideas of musical consonance and dissonance in their various historical incarnations and subsequently apply their relevant details within the context of electro-acoustic music. I propose the notion that individual sound objects, or relatively brief sonic gestalts, can be considered relatively consonant or dissonant, just as theorists have historically described dyads (and only more recently, chords) comprised of distinct pitches. Furthermore, I propose a rudimentary musical model—not a mathematical or cognitivist one, of which several exist—of "timbral consonance" defined in terms of compositional aspects of electro-acoustic music. In a manner similar to how purely pitch-based notions of dissonance has informed the syntax of tension and release in tonal music, so has the timbral counterpoint of "consonant" and "dissonant" sound objects. Specific examples of such music are discussed.

Clearly, this premise and ensuing discussion could be subverted based on one's definition of the term "sound object." In the casual sense, I am using the term to denote the shortest perceivable yet coherent sonic gestalt in electro-acoustic music, independent of its source. (A more formal definition will be offered in the third chapter.) For now, let us think of a sound object in digital terms as a short sound file—the building block of texture, gesture, and soundscape in electro-acoustic music composition.

As a part of the musical model describing the dissonance of the sound object, I discuss the role that both measurable (i.e., quantifiable) and subjective (i.e., qualitative) features such as reverberation, transients, "recognizability," perceived causality, spatial velocity, signal quality, context, and other aspects play in the reception and perception of dissonance of electro-acoustic music. I offer that the unique interactions among the physical properties of sound objects in conjunction with their predominantly nonlinear psychoacoustic correlates can inform the groundwork for such a new theory of dissonance: a dissonance theory of sound objects.

The second chapter addresses important historical theories of musical dissonance, followed by non-musical theories of dissonance culled from the literature of anthropology and psychology. Rather than linearly presenting historical dissonance theory, which has already been surveyed by James Tenney in his important work *A History of 'Consonance' and 'Dissonance'* (1988) and elsewhere, I categorize and discuss the global evolution of different theories and trends. In parallel with our increased understanding of acoustics, psychology, and human neurology/physiology, theories of musical dissonance have, over time, assigned greater responsibility to the role of timbre over pitch in our perception and categorization of dissonance.

In Chapter Three, I present issues concerning the quantification of dissonance, that is, attempting to "objectively" measure levels of dissonance in music that potentially encapsulates both the physical properties of sound objects and their psychoacoustic correlates in some meaningful way. Subsequently, Chapter Four offers a discussion of the value—and alternatively, the potential futility—of such objective tasks, except as a supplementary tool when performed in conjunction with a broader analysis of a work or as an aid in the development of more complete psychoacoustic models.

In Chapter Five, "Dissonance of Sound Objects," I formally define the term "sound object," motivate the necessity of viewing musical dissonance as an inherent property of a sound object, and I argue that sound objects themselves naturally fall onto multidimensional axes of consonance and dissonance with respect to particular physical and perceptual properties. A battery of listening tests that were conducted are then described and analyzed. Here, I formally outline many of the physical properties of sound objects and their psychoacoustic correlates that I submit contribute to our perception of the musical dissonance of sound objects.

Can the sound of two hands clapping be thought of as more consonant than that of a squeaky door? Under what circumstances, musical or other, might characteristics of each sound change their consonance relative to each other? More importantly, which *recorded* hand claps are more or less musically consonant than which *recorded* squeaky doors? Clearly, cultural training plays an important role in our overwhelming percept of "beauty" in the octave and "ugliness" in the minor second, but what about *sounds* in general? As Huron (1997) writes:

> Why *do humans find certain sound stimuli "ugly" or "repulsive?" Why would some simple sine-tone combinations cause distress or discomfort?*

In attempting to answer these kinds of questions from a music-compositional perspective, I offer implicit and perceptual properties of sound objects that join to create dissonance continua. For example, factors that most certainly affect our compositional assessment of a sound object's relative dissonance include the sound's inherent "harmonicity," transient nature, reverberation, "recognizability," causality, volume, context, spatial location, spatial velocity, signal quality, and any visual cues that may accompany the sound's production, along with our perception of whether the sound poses a biological threat that should elicit a necessary physical response. I discuss each of these factors and present sound examples to illustrate each of them.

❧

In contrast to other existing ideas, I am primarily concerned with a compositional and aesthetic, rather than a mathematical or scientific, description of musical dissonance. My discussion is related to the practice of composing electro-acoustic music and anecdotally includes how composers of such music might intuitively approach the juxtaposition and intersection of "consonant" and "dissonant" sounds in their compositions. As such, my discussion is descriptive, rather than cognitivist, empirical, or mathematical.

In a larger context, I am attempting to understand ways dissonance of sound objects may inform compositional choices in electro-acoustic music. Arguably, so much traditional acoustic music has been effective largely because composers have discovered or even perfected various means for creating feelings of relative “tension” and “release.” The same thing happens in successful electro-acoustic music, though the “tension” and “release” are created through alternating timbres and textures, for example, rather than notationally representible pitch classes, which suggests the categorical importance of a dissonance theory of sound objects.

The nonlinear relationship among physical phenomena of sound production, psychoacoustics, and our perception of musical dissonance underscores the hysteretic nature of dissonance. It is because of this very hysteresis that quantitative and qualitative conceptions of dissonance by necessity lag each other in perpetuity. Ultimately, I am ultimately asking many more questions than I can answer, but the cathartic exercise of pondering certainly sounds good, and if it sounds good, it is good, I hope.

2 PRELUDE: A SURVEY OF MUSICAL CONSONANCE

> *It is clear…that the fundamental reason for the great divergence in the ranking by experts and the consequent disparagement of the ranking of consonance and dissonance has been due to the failure to take common ground in the definition of these terms.* (Malmberg 1918, p. 108)

Attempts to quantify aspects of musical experience date as far back as the Pythagoreans, who computed and tabulated ratios that represented the correspondence between vibrating lengths of a monochord and the perceived pitches and intervals. Since then, the terms *consonance* and *dissonance* have been ascribed—often quite casually and inconsistently—to various features of both linear and vertical harmony throughout the history of music. Consonance, like timbre, is not easily defined, nor is it necessarily simple (or even meaningful, as we will see) to parameterize dissonance as a perceptual construct.

Like timbre, only specific elements of consonance exhibit any meaningful ordinality at all: one can only say that one interval is more consonant than another within a single given parameter, just like one can only quantitatively compare two sounds in terms of one particular timbral feature at a time. To say one interval is inherently more consonant than another may be true in one sense and not in another. The interval of a perfect fifth, for example, may not always be perceptually consonant contrary to popular opinion (and music theory textbooks). The dissonance of intervals is greatly dependent on the timbres that play each pitch.

Consider, for example, the following thought experiment: two tuning forks are pitched a perfect fifth apart and struck simultaneously. Compare the dissonance (in the casual sense of the term) of the resulting sound with that of two large cymbals whose fundamentals are tuned precisely a 3:2 perfect fifth apart but whose spectra are entirely inharmonic. Most people would say the perfect fifth resulting from the tuning forms is much more "consonant" than that exhibited by the cymbals.

Many other anomalies in this world of consonance and dissonance exist. Perhaps one of the most interesting is their non-complementarity: the degree to which an interval is classified as "dissonant" by a listener is not necessarily the inverse of the degree to which the same listener classifies the same interval as "consonant."

In this chapter, I will outline a few different ways the terms "consonance" and "dissonance" have been defined and used. I will then turn our discussion to realms outside of pitch (and music, for that matter), where we will briefly address rhythmic dissonance and cognitive dissonance. I will then present a brief history of attempts to quantify dissonance levels: the empiricism of dissonance, as it were. Next, from James Tenney's writings on musical structure and form, the idea of using consonance and dissonance as organizing principles in a musical work is motivated, supplemented by a brief discussion of several works organized in such a way. The chapter concludes with my observations on the idea and process of quantifying dissonance, specifically, what "works" and what does not.

2.1 The Legacy of "Consonance" and "Dissonance"

> *Eight softening bars tell us unambiguously that we approach a love scene. But this motive built on the alternation of two dissonant chords sounds rather like scratching a glass plate with a sharp knife. Like a cold snake-skin runs this love bliss down the spine.*
>
> —Eduard Hanslick, *Neue Freie Presse*, Vienna, November 30, 1876, speaking of Tchaikovsky's *Romeo and Juliet* Suite (quoted in Slonimsky 1953, p. 206)

That Western philosophy tends toward descriptive categorization is an understatement. Consider Aristotle's writings on the origins and kinds of humor in the *Poetics*, Augustine's categorization of sins in the *Confessions*, Aquinas' encyclopedic *Summa Theologica*, and Wittengenstein's outline-exposition of symbolism and language in *Tractatus Logico-Philosophicus*, to name but a few famous examples. Most Western writings about music are no different; they generally attempt to dissect and categorize, and writings and theories about music that lie outside of this expository tradition are often dismissed.

To the extent that music before the invention of electronics was written, described, and analyzed in terms of pitch, the perceptual correlate of frequency, it was only natural to begin dissecting the implications—be they mathematical, musical, affective, or spiritual—of all possible (or "allowable") combinations of pitches. And this is exactly what happened. Most early writings on music discussed intervals and their resulting dissonance. Quite simply, the more perfect the "consonance," the more "beautiful" the interval, and therefore the more perfect expression it is of divine truth. Clearly suggestive of a relative scale of consonance, it was not long until people began actually assigning numerical values to intervals as a supposed indication of their relative dissonance levels. (We will return to this later.)

It is often taken for granted that practice precedes theory in the classical Western tradition: composers write; theorists categorize. Consider the number of music theory textbooks and theoretical studies that are written each decade on tonal music, well after the so-called compositional demise of the tonal system. But, as Jeppesen (1939) notes, theory has occasionally preceded practice, and when this happened, it was generally in the context of theoretical treatments of consonance and dissonance. As an example, he notes the Ars Antiqua as an example:

> *…[T]he "Franconian" law, setting forth a prohibition against dissonances upon accented portions of the measure, was formulated by the theorists some time before it was carried out in actual practice. Likewise, although the prohibition against parallel fifths was proclaimed in the thirteenth century and was made more stringent by the theorists of the fourteenth, one cannot regard it as having been fully observed until the appearance of the* a capella *composers of the Palestrina period.*

And so the concepts of consonance and dissonance occupy a central historical role in the theory-practice dialectic. Even still, it seems no music theory textbook is complete without ample, albeit grossly simplified, definitions and categorizations of various "consonance" and "dissonances."

❧

One of the major impediments to a unified theory of musical consonance is simply the lack of consensus in terminology. And virtually no twentieth-century examination of the concept of musical dissonance is complete without a statement similar to the preceding sentence. Yet, millennia after the concepts of consonance and dissonance were manifest in dialogue about music, we still do not firmly embrace the true multidimensionality of the concept.

Epistemologically speaking, terminology itself can play a major role in understanding of any entity, as discussed in the writings of Bertrand Russell and Ludwig Wittgenstein. But the definition of the purely ineffable is no mean feat. However, contrary to Wittgenstein's idea on the issue—"Whereof one cannot speak, thereof one must be silent" (*Tractatus Logico-Philosophicus*7)—an attempt to discuss that which one cannot fully describe or observe (musical dissonance, in this case) can yield new ideas and creative insights.

Many composers and theorists are justifiably concerned over the lack of coherent thinking about the constitution of dissonance. As Partch (1974), p. 154 wrote of modern composers:

> *But whether they are consophiles or consophobes, they are justified in objecting to the common terms "pleasant" for consonance and "unpleasant" for dissonance, terms which are indefinite if not actually misleading.*

He quickly adds the following in a footnote, alluding to the ineffability of consonance:

> *Nor are the terms of the psychologists very clarifying. The criteria, and associated terms, for consonance encountered in their writings include: mechanism of synergy, conscious fusion, fusion, smoothness, purity, blending, fractionation. So many terms confuse the issue. The*

> *word consonance evolved as it did to express the idea that it does express, and—even thought it is one of a homonym—spelled in this way it expresses nothing else.*

Thankfully, we more or less all recognize dissonance when we hear it. And history is rife with fanciful descriptions: as early as 1573, Gabriel Harvey wrote of "Dissonant and iarring dittyes" (*Letter-Book* 1573–80, p. 117), and

As Hutchinson and Knopoff (1979) relate, Mersenne wrote of the "trembling" of mistuned organ pipes as early as 1636, and William Holder noted the "Battel in the Ayr" that results from adjacent low-frequency organ pipes (Wever 1929). John Milton's 1634 *Comus* notes "The…roar…filled the air with barbarous dissonance," while Joseph Addison's *The Spectator* (1711–1714) astutely observes the subjective nature of dissonance and musical preference, writing "What is Harmony to one Ear, may be Dissonance to another."

And thus, consonance is what it is. Dissonance is what it is. Perhaps that is the best we can say about it, analogously to James Tenney's circular definition of timbre as that which pitch and rhythm are not—as well as some of what they are as well. As a testament to the generality that the notion of dissonance can encompass both in and outside of music, consider the writings from a broad array of disciplines— from computational musicology to anthropology and psychology—that address the issue both colloquially and theoretically. Considering the amassed field of knowledge from each of these disciplines, it is interesting to search for areas of overlap among the conceptions proffered by each field. For example, to what extent can gestalt psychology and Festinger's psychological theory of cognitive dissonance inform a musical definition of dissonance?

In such an endeavor, common terminology is essential, as previously noted. To address (and admittedly potentially confuse) this issue, let us begin by briefly dissecting ten different ways the term "dissonance" has been used. We will consider the notion of "Music of the Spheres" with a brief foray into the strange world of comma phenomena. Next, we examine the concepts of consonance as "pleasantness" and musical stability. We then address Helmholtz's idea of consonance and beat frequencies, followed by early-twentieth-century explorations of the relationship between consonance and tonal fusion of partials.

Next, we will touch on William Sethares' conjecture of "tonal consonance," perhaps the first comprehensive dissonance theory that fully takes timbre into account, followed by a brief discussion of recent neurological experiments in the medical research community on dissonance perception. Moving beyond the realm of pitch and timbre, we conclude the tour by presenting the notions of metrical dissonance, rhythmic dissonance, and contextual and cognitive dissonance.

Discussions of dissonance in the context of music have almost exclusively occurred—until surprisingly recently—within the context of the interval, and hence, tuning theory has played an important role in formulating many theories of dissonance. The notion of relative dissonance of intervals has generally assumed of the preexistence of scale (a safe bet for Western music!), and so theoretical dissections of scale have almost always been predicated on a search for maximal consonance of chords within the scale. Numerological quests for simplicity and order in scale to achieve maximal consonance undoubtedly yielded beautiful chords, like 4:5:6 major triads, but not without undesirable side effects, viz. “wolf fifths,” comma phenomena, etc. It was as if a giant optimization problem were controlling everything: the more beautiful one thing, the more ugly the other. Clearly, early writers postulated, a cosmic force must be in control.

2.2 The Music of the Spheres

At least seven different meanings and connotations of the term “consonance” may be traced, beginning with the mystical—and later overtly theological—idea that perfect consonances of simple numeric ratios best express the divine proportions of the universe (the *musica mundi*, reflected downward into the *musica humana* and the audible *musica instrumentalis*). Dissonance, on the other hand, was associated with evil, epitomized in the tritone’s reputation as *diabolus in musica*.

The vast majority of Western writings on and about music until the late Renaissance occupied a peculiarly speculative space embedded simultaneously in theology, philosophy, cosmology, and mathematics. The prevailing tone was one of order, “perfectness,” and beauty of the cosmos, reflected most visibly in the motions of vibrating strings here on earth. Long before we derived the wave equation, a partial differential equation that clearly explains how wave phenomena in ideal vibrating strings lead to simple harmonic motion and perfectly in-tune harmonic modes of vibration, Pythagoras, Boethius, Augustine, and others recognized the theosophical beauty in simplicity, the reflections of the divine cosmos, in sound.

The esoteric concept of "music of the spheres" is well-documented elsewhere (e.g., Haar 1998; James 1993; Voss 1998; 2000), and a wonderfully concise and elegant history is presented by Haar in Wiener's *Dictioinary of the History of Ideas: Studies of Selected Pivotal Ideas* (1973). The central theme is that music and the cosmos are inextricably intertwined: that the divine manifests itself not in the physical nature of sound, but in the harmonious relationship of all elements in the universe. This idea found its basis in the cult of the Pythagoreans, which flourished around 500 BC, who among other ideologies held that humankind's spiritual communion with the divine was possible through expressions of the mathematical realities of the universe. (See Iamblichus' *Life of Pythagoras*.) Aristotle even wrote of this cult in his *Metaphysica:* "They supposed the elements of numbers to be the elements of all things, and the whole heaven to be a scale and a number " (*Metaphysica* A 5 986a, trans. W. D. Ross). The Pythagoreans believed that certain numbers themselves were beautiful, but apparently not all of them: it is said that Pythagoras ordered the beheading of one of his followers who showed that $\sqrt{2}$ was an irrational number.

The theoretical study of music offered a natural framework within which Pythagoreans could assert and test the correspondences among numbers, scales, and sound. The Greek biographer Diogenes Laërtius (fl. early third century), for example, credited Pythagoras of Samos with defining the four principal consonances that can be formed within the sacred *tetraktys* (the numbers 1–4), namely the unision (a 1:1 of string lengths on two monochords), octave (2:1), perfect fifth (3:2), and perfect fourth (4:3). Thus we still denote musical scales tuned according to the highest prime number in the *tetraktys* (3) as *Pythagorean* scales, to which we will return later.

Pythagoreanism itself laid at the historical nexus of East and West—clearly a confluence of Jewish Kabbalistic and Chaldean (Babylonean) numerological traditions that largely commingled stories of numerical harmony with creationism, of mathematical beauty and cosmological exegesis. Consider, for example, Plato's play *Timaeus*, written circa 360 B.C., in which the protagonist proclaims:

> *Moreover, so much of music as is adapted to the sound of the voice and to the sense of hearing is granted to us for the sake of harmony; and harmony, which has motions akin to the revolutions of our souls, is not regarded by the intelligent votary of the Muses as given by them with a view to irrational pleasure, which is deemed to be the purpose of it in our day, but as meant to correct any discord which may have arisen in the courses of the soul, and to be our ally in bringing her into harmony and agreement with herself; and rhythm too was given by them for the same reason, on account of the irregular and graceless*

> *ways which prevail among mankind generally, and to help us against them.* (*Timaeus* XIV, trans. B. Jowett)

One of the greatest expressions of the music of the spheres, written about the same time as *Timaeus*, is found in the Myth of Er from Plato's *Republic*, the impetus for Iannis Xenakis' octaphonic tape-music classic *La legende d'Eer* (1977–78). As Haar summarizes in Weiner (1973):

> *Er the Pamphylian, a hero slain in battle, was given the privilege of seeing the next world and then returning to life to describe what he had seen. The vision of Er includes once again a model of the universe, a set of concentric rings or whorls—the planets—hung on the spindle of Necessity. The rims of these whorls are of different sizes and colors, and they revolve at different speeds—all theinner ones in opposition to the movement of the outer rim, the firmament. The Pythagorean proportions of the* Timaeus *are lacking here; but present is actual music, for as the spindle turns, "on the upper surface of each circle is a siren, who goes round with them, hymning a single tone or note. The eight together form one harmony "* (*Republic* X. 617, trans. B. Jowett).

Although a Christian take on the concept of Music of the Spheres was offered by Augustine in his classic *De Musica* (written c. 391), itself perhaps influenced by Clement of Alexandria's second-century *Exhortation to the Greeks*, the sixth century saw two important works on the subject. The first, Cassiodorus's *Insitutione*, echoed much of Pythagoreanism in its espousal of the efficacy of music to affect the soul and in the important relationships between numbers and music. However, the primary testament of Post-Classical thought on the subject came from another writer. The terms most often associated with the Music of the Spheres were given voice by Anicius Manlius Torquatus Severinus Boethius (c. 480–525) in his *De Musica* (c. 500), namely the lofty *musica mundana*(the music and harmony of the cosmos as a whole), *musica humana* (the human "music"—harmony that results from the proper relationships between humans and the divine, particularly in a moralistic sense), and *musica instrumentalis*(the only music that most can actually hear, i.e., the actual sounds of musical instruments). (Incidentally, this work of Boethius was apparently required at Oxford University as a standard music theory text until 1856.) Boethius' other primary contribution to music theory, *De Institutione Musica* (c. 505), specifically defines consonance of musical sounds, a definition to which we will return later.

Aside: Comma Phenomena, or "Why Did God Do This?"

The impetus behind much of the concept of the Music of the Spheres was surely owing not in the least to the existence of comma phenomena, a numerical anomaly inherent in just-intonation tuning systems, which were discovered remarkably early by many of the Pythagoreans. Commas are small intervals that result from the slight inequality of successive just-intonation intervals that should ultimately result in a simple just interval but do not. For example, three successive major thirds that are equivalent in intervallic ratio should yield an octave ratio, but this is impossible without resorting to irrational numbers, because the only number x that solves the relation $x^3 = 2:1$ is $\sqrt[3]{2}$, which of course is irrational. The use of irrational numbers to represent frequency ratios of musical intervals of course violates the very idea of just intonation by definition, leading to the concept of temperament. In fact, it could be said that the rise of equal temperament had less to do with allowing modulation than the solution to alleviate commas by tempering and distributing them throughout the scale. The ability to play in any key was a nice side effect. (See Barker 1989; Chalmers 1993; Forster 2005 for descriptions of historically important tunings and scales and additional discussions of comma phenomena.)

There exist a mathematically unbounded number of commas. However, a few generally receive special attention owing to their importance and primacy, as well as to the historical nature of their discoveries. We will briefly outline each of these in turn. The most famous of all commas is of course the so-called Pythagorean comma, discovered by the Pythagoreans, which is a result of the fact that twelve diapente (3:2 perfect fifths) do not equal seven diapason (2:1 perfect octaves). This comma is a number we can represent as

$$\frac{\left(\frac{3}{2}\right)^{12}}{2^7} = \frac{531441}{524288}$$

$$\approx 1.0136432647705078125$$

Expressed in units of cents, the ratio becomes

$$1200\log_2\left(\frac{\left(\frac{3}{2}\right)^{12}}{2^7}\right) \approx 23.4600103846490129338407\ \text{¢}$$

or almost one-quarter tone in twelve-tone equal temperament.

Other important commas include dieses, the syntonic comma, and the Schisma. The Great Diesis is simply the interval by which a perfect octave exceeds three 5:4 major thirds:

$$\frac{2}{\left(\frac{5}{4}\right)^3} = \frac{129}{125}$$

$$= 1.024$$

In general, a diesis is any sufficiently small interval ε by which m octaves exceeds n major thirds, or

$$\left|\frac{2^m}{\left(\frac{5}{4}\right)^n} - 1\right| < \varepsilon$$

of which an unbounded number exist.

The Syntonic Comma, or Comma of Didymus, represents the interval by which four perfect fifths exceeds two perfect octaves plus a major third:

$$\frac{\left(\frac{3}{2}\right)^4}{2^2\left(\frac{5}{4}\right)} = \frac{81}{80}$$

$$= 1.0125$$

Finally, the Schisma is the interval by which eight fifths plus a 5:4 major third exceed five octaves:

$$\frac{\left(\frac{3}{2}\right)^8\left(\frac{5}{4}\right)}{2^5} = \frac{32805}{32768}$$

$$\approx 1.001129150390625$$

A thorough accounting of many more important commas is provided in the Appendix. A more rigorous mathematical motivation and treatment of various comma phenomena is given in Haluska (2004).

Commas are interesting quite simply because they exist. Just as the discovery of the ratio of the circumference of a circle to its diameter, π, was an irrational number must have been a crushing blow to early mathematicians, so too did the existence of commas befuddle and intrigue early music theorists.

Neo-Pythagoreanism

After centuries of near-dormancy, the ideals espoused by Pythagoreanism found fresh voice in the "esoteric science" movement of the Renaissance, which included the study of alchemy and other fringe topics. The term "Neo-Pythagoreanism" historically applies most often to the so-called "first generation" post-Pythagorean writers (e.g., Boethius, Cassiodorus, and Proclus), we use it here to denote the Renaissance astronomers, philosophers, and mathematicians whose writings so often echo traces of Classical Pythagoreanism.

Many writings of the major astronomers—Galileo, Copernicus, and Newton, to name a few—directly address cosmic harmony in a neo-Pythagorean light. The "Music of the Spheres" concept had in particular served as a lifelong devotion of Johannes Kepler (1571–1630), who as a corollary to his heliocentric, qualitative laws of planetary motion noted the musical intervals formed by the ratios of the angular velocities exhibited by each planet at aphelion (the position furthest from the sun) and perihelion (the position nearest the sun). Thus was formed a quantitative and quite literal Music of the Spheres, directly within our solar system, perhaps predating modern ideas of sonification of inherently nonmusical data.

Consider the Fifth Book of Kepler's *Harmonices Mundi* (1619), which addresses "celestial harmonies" in terms of arithmetic proportions of distance among the planets as well as analogously to tuning ratios in music. (Particularly interesting is the eighth question, “In the Celestial Harmonies, Which Planet Sings Soprano, Which Alto, Which Tenor, and Which Bass?”) He writes:

> *Accordingly the movements of the heavens are nothing except a certain everlasting polyphony (intelligible, not audible) with dissonant tunings, like certain syncopations or cadences (wherewith men imitate these natural dissonances), which tends towards fixed and prescribed clauses—the single clauses having six terms (like voices)—and which markes out and distinguishes the immensity of time with those notes. Hence it is no longer a surprise that man, the ape of his Creator, should finally have discovered the art of singing polyphonically…which was unknown to the ancients, namely in order that he might play the everlastingness of all created time in some short part of an hour by means of an artistic concord of many voices and that he might to some extent taste the satisfaction of God the Workman with His own works, in that very sweet sense of delight elicited from this music which imitates God.* (Quoted in Hawking 2005, pp. 45–46.)

For Kepler and other astronomers, the mimetics of music in its parallels to and imitations of the cosmos clearly echo the sentiments of Pythagoreans. Such a thought was also promulgated in Mario Bettini’s *Apiaria* (1641–42), Giambattista

Riccioli's *Almagestum novum (*1651), and Marin Mersenne's *Harmonie universelle (*1636).

Neo-Pythagoreanism indeed flourished particularly in the 17th century in Europe, some principles of which were directly reflected notably in the famous *Musurgia Universalis* (1650) of Athanasius Kircher (1602–1680) along with many other of his prolific writings. Kircher's book, one of the most important musicological works of the 17th century, is replete with hermetic symbolism and is known to have been influential on J. S. Bach and other composers. The work, in its broad mixture of music theory, instrumentation, pedagogy, instrument building, and an overview of the human auditory system, particularly benefited from Kircher's membership in the 40,000-member Society of Jesus. Kircher frequently corresponded with international Jesuit priests, who sent information about musical instruments from distant lands. (See also the Athanasius Kircher Correspondence Project, available online at http://193.206.220.68/kircher/index.html, for more details.)

But the revitalization of the concepts surrounding the Music of the Spheres, particularly the more esoteric principles, was perhaps best summarized graphically in work by the English physician and erstwhile Hermetic philosopher Robert Fludd (1574–1637), a contemporary of Johannes Kepler. Often called the last Renaissance Man, Godwin (1979) remarks that "He lived at the very end of an era in which it was possible for one mind to encompass the whole of learning." Fludd's writings betray his obsession with unifying the microcosmic and macrocosmic, the divine and the mundane, the inner self of the individual within the totality of cosmic harmony. And music, of course, figured into his world view, the descendant of a decidedly unique lineage of mystical Christianity one could trace to Origen of Alexandria (A.D. 185–254) through Hildegard of Bingen (1098–1179) and Meister Eckhart (Eckhart con Hochheim, c. 1260–1327/8).

De Musica Mundana, a rather rambling and strange book (and somewhat rife with misprints), is contained within the first volume of Fludd's *Utriusque Cosmi Maioris scilicet et Minoris Metaphysica* (*History of the Macrocosm and Microcosm*, 1617/1618) and characterizes Fludd's Neo-Pythagorean approach to the place of music in the cosmos. Perhaps the most famous of the woodcuts contained therein is shown in Figure 2–1, "The Divine Monochord." Here, the interval separating the earth (Terra) and the highest of the heavens (at the top of the figure) is the *Disdiapason* (double octave, 4:1, listed on the left of the monochord string as *Proportio quadrupla*), while the interval from the earth to the sun is the *Diapason materialis* (2:1 octave, listed on the left of the monochord string as *Proportio dupla*). Note the upward progression from Earth (symbolized as the note *G* by the letter Γ), through Water (*Aqua*), Air (*Aer*), Fire (*Ignis*), the moon and other planets, to the heavens. And without any trace of subtlety, the very finger of God is itself serving as a cosmic tuning peg on the monochord, governing the precise frequency at which each element vibrates.

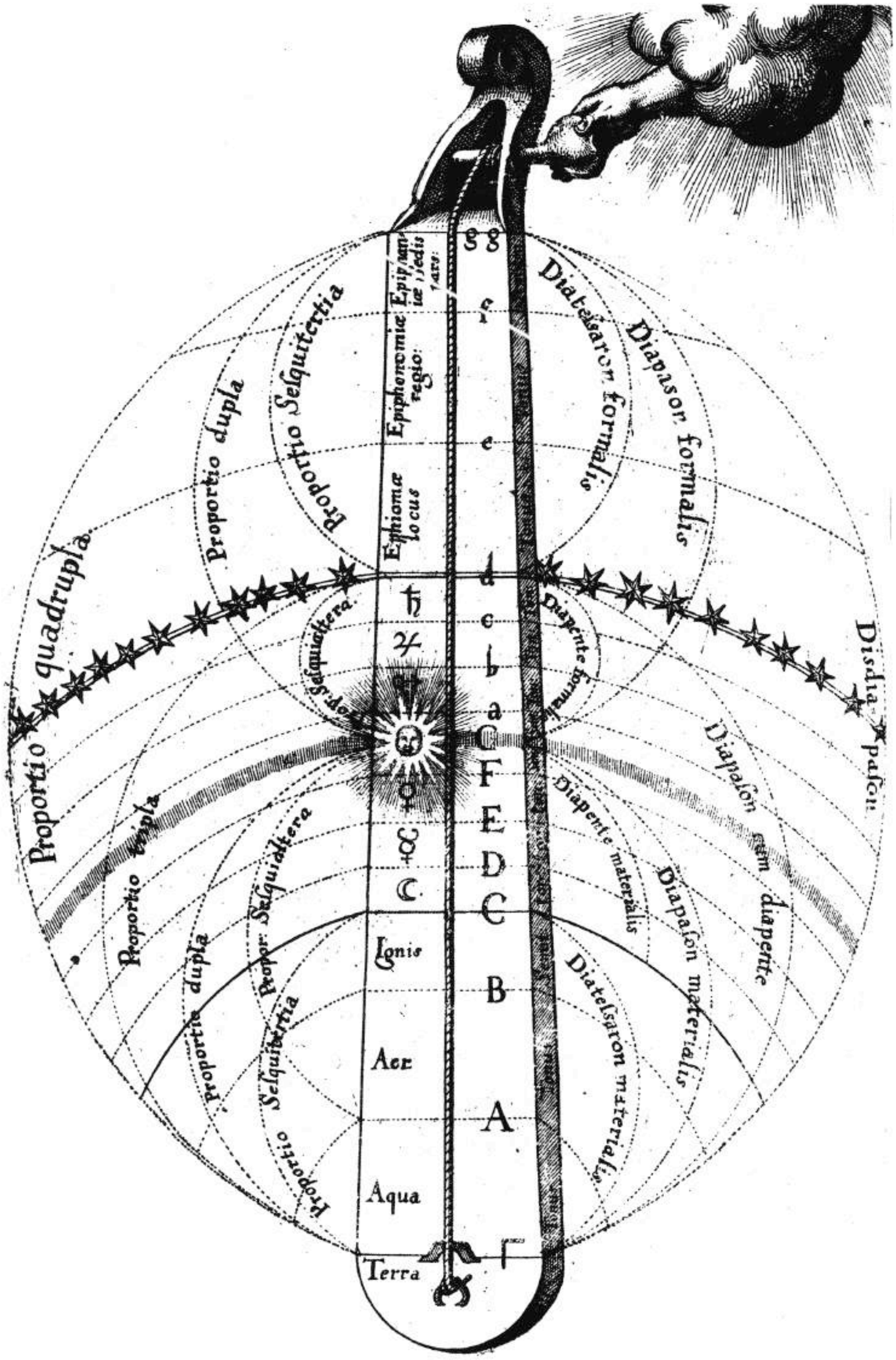


Figure 2–1. Robert Fludd, "The Divine Monochord" from *History of the Macrocosm and Microcosm* Volume I (1617), Tractate I, p. 90.

Fludd's "The Elemental Monochord" (Figure 2–2) similarly ascribes intervallic ratios to the relationships among the "three regions" (*Regio Primama, Secunda, and Tertia*) of each of the four elements (Earth, Water, Air, and Fire). Here, the "three regions"correspond to the "inferior" state of the element (in which case the element contains material from the immediately lower state), the "pure" state, and the "superior" state (in which case the element contains material from the immediately superior state). Curiously, as Godwin (1979) observes,

> *Perhaps more significant that this rather laboured system is the presence of the Sun at the monochord's peg, in the same position as the hand of God… [shown in "The Divine Monochord"]. Does this imply that as the Creator is to the universe, so is the Sun to the sublunary realm? Occult doctrine would certainly agree.*

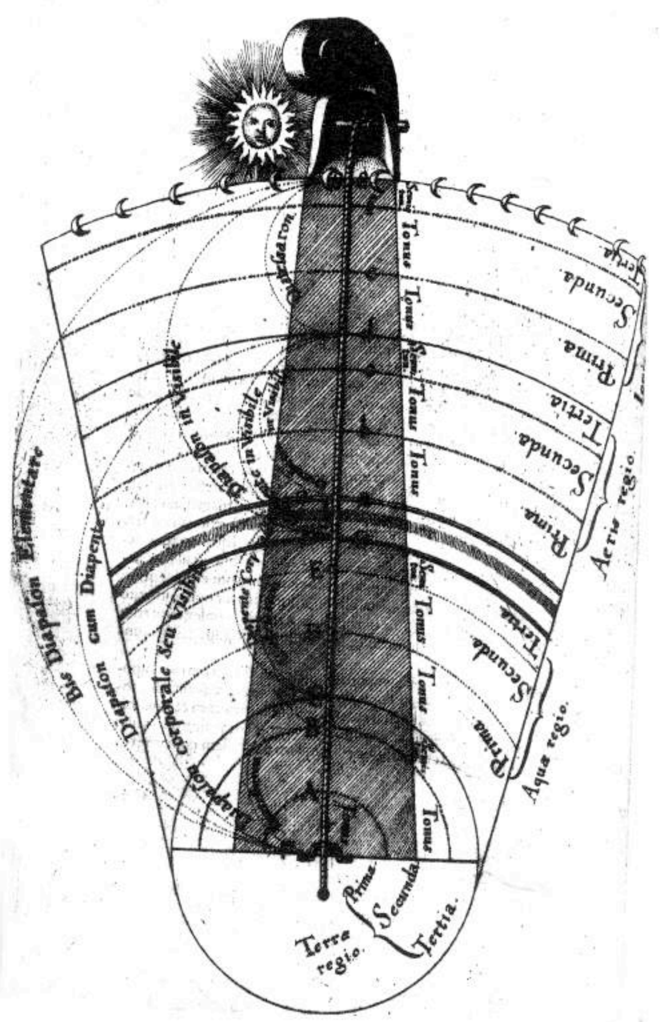


Figure 2–2. Fludd, "The Elemental Monochord" from *History of the Macrocosm and Microcosm* Volume I (1617), Tractate I, p. 100.

Among the most musically significant and symbol-rich Pythagorean iconography of Fluud is certainly *The Temple of Music*, shown in Figure 2–3, an amalgam of musical information. Of particular note here is the foundation on which the temple is built: (1) the lute, which Fludd honors in corresponding text as the most desirable of musical instruments; (2) the entry of Pythagoras into a blacksmithing shop, where he reportedly discovered correspondences between hammer weights and the intervals they produced; and (3) a “cheat sheet” of musical notation containing a scale on G, notated in the bass clef, along with successively shorter rhythmic values.

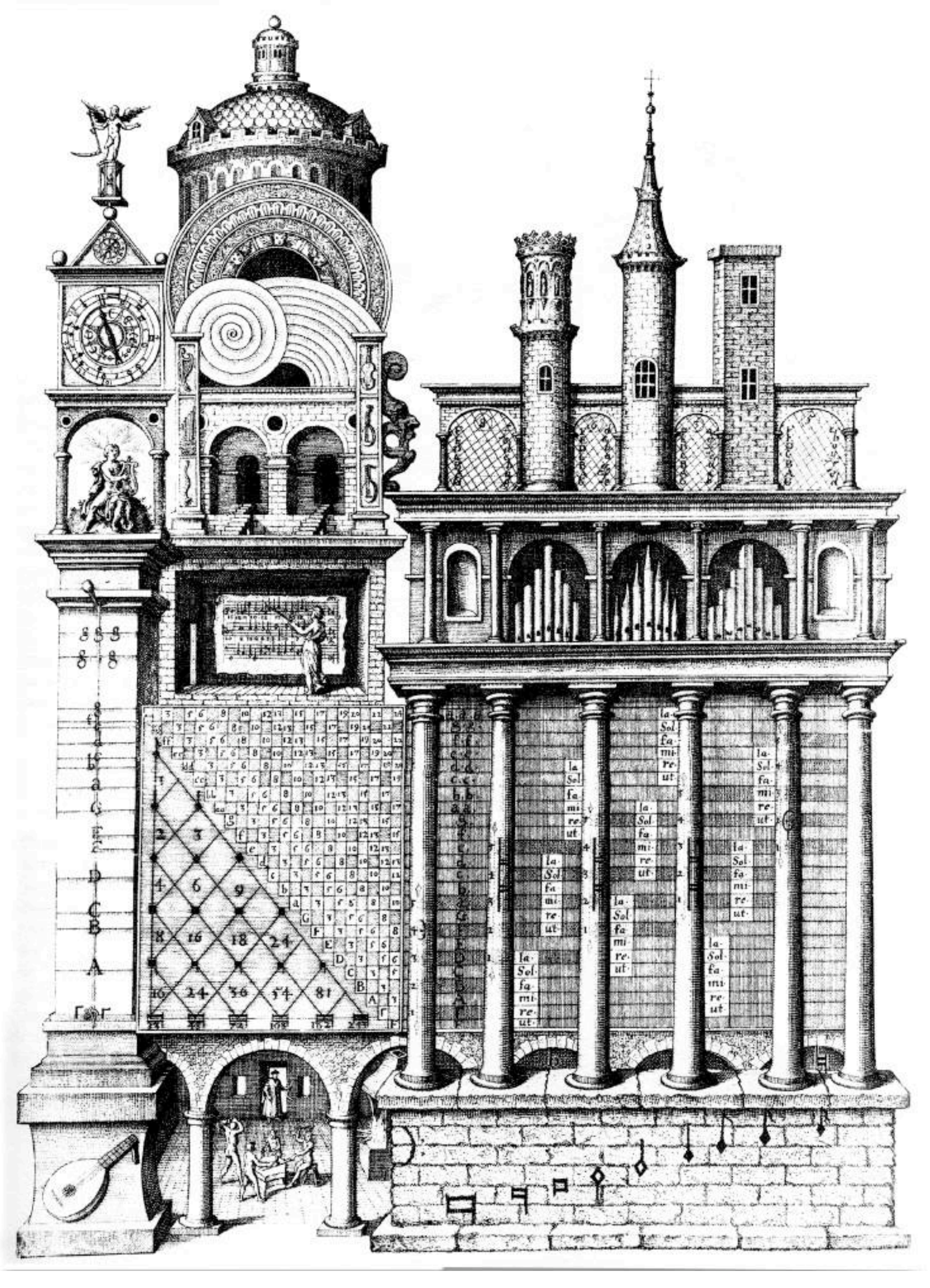


Figure 2–3. Fludd, "The Temple of Music" from *History of the Macrocosm and Microcosm* Volume I (1618), Tractate II, pp. 160–161.

Of the *Temple*, Fludd himself writes of the esoteric nature of its invention and description:

> *The poets, whose efforts are accustomed to be assiduously engaged with fables and images, would be singing about the buildings and wonderful site of this temple. Indeed, they may have pursued this subject with even greater acumen since, after all, music derived its name from their goddesses, the Muses, just as is evident from its etymology. I beseech, therefore, that I may ask pardon from them if I permit myself to be led very much without measure by the invention and stimulus of poetical madness in the description of this temple.* (Quoted in Barton 1978.)

Fludd addresses consonance as he continues to describe the Temple in an engaging and inventive narrative:

> *Thus, let us imagine this Temple of Music to be built on the top of Mt. Parnassus, the abode of the Muses, adorned in every part with eternally green and flowering woods and fields, and pleasantly surrounded by crystal fountains flowing here and there in different directions whose murmur often brings a peaceful sleep to passers by.*

Birds frequent theses parts and inhabit the woods pouring forth diverse consonances of sound in greater symphony. They seem diligently to lay the basis or foundation by means of their higher, more piercing song; through their melody the Nymphs themselves around the temple, the Satyrs led through the woods by Sylvanus and the shepherds led through the fields by Pan, are all moved to engage in choral dances.

Among these delights, therefore, that divine gift of Apollo is established, preserved and indeed worshipped by the adoration of all souls. All of its constituent parts are given up to peace and concord, in the mysteries of harmony and symphony, including the concords of heaven and the elements, so mutually bound to each other that it would be necessary for the whole world to perish and be reduced to nothing by the strifes of discord before these consonances would either disappear of be destroyed.

Therefore, the protectress or goddess of this temple is Concordia, ineffable Concord, great offspring of the Being of Beings, by whose adoration little things grow, and by whose contempt great things fall to pieces. Its guardian or priestess is Thalia, most delightful of the nine Muses, by the example of whose harmony the occult mysteries are explained to pilgrims who suppliantly seek her oracles.

Therefore, a man with a keen eye for knowledge will pay attention to any part of this structure and not disdain the smallest portion, *because it is moved by that harmonic soul of Apollo in each part as in its whole. That spirit of music, after the manner of a zephyr, is accustomed to blow through all the sinews of this building, soothing and gladdening the souls of living beings, carrying away with itself the lusts of man, and restraining the madness of evil daemons as if imbuing them with a certain humanity.*

You should eagerly contemplate the spiral revolution of the larger tower of the temple which denotes the motion of air, after it is caused to resound by sound or voice. the two doors represent the ears, the organs of hearing, without which the emitted sound cannot be perceived, nor may one enter this temple except by them. In the following place you will observe its three smaller towers representing the arrangements of notes, b rotundum b quadratum, and naturalis. And with observation of these, three rectangles must be carefully examined in order to determine the diverse natures, names and places of the aforementioned notes in the demonstrated system (anything placed under any tower is naturally related to that tower). The pipes or organs of these rectangles, distinct in their height, denote the difference of voices and sounds of any rectangle.

Indeed, the division of the column of this temple must not be disdained, since it will delineate the true proportions and diverse species of consonances. The clock must also be zealously pondered lest time waver unexpectedly or advance with too slow a pace, that is, one

> *which does not observe proportion or measure. And so, this clock is a sort of guardian of the regular times of the notes and a most ample mirror of their simple value.*
>
> *Why then will not the triangle of proportionate quantity have to be inspected, which probes into the diversity of the proportion of times in diminution as well as in augmentation and clearly shows the perfections and imperfection of the notes? Also the triangle of the system of harmonious intervals, as it were the end of all the remaining mysteries, ought to be looked into with no little care, since, through it and from it all the concords of music are produced, without which no harmony is made. Beyond this triangle is depicted the story in which the discovery of its consonances is told, namely the observations of Pythagoras, who passing by a certain blacksmith's shop by chance hearing an agreement from the striking of four hammers, ordered the hammers to be weighed, and from the difference of their weights he discovered the three musical proportions of consonances: diatesseron, diapente and diapason, which we have very plainly explained by the letters and connection of letters in the three windows of the temple, which are equally of use in composition of musical harmony and the harmonical triangle.*
>
> *Therefore, eager reader, if you keenly examined these parts of the temple, you will be a partaker of all of its mysteries and a great master of this excellent science.* (Quoted in Barton 1978.)

The image has much to say visually about dissonance. For example, the upper-triangular area immediately above the depiction of Pythagoras serves as a compositional aid by providing the note-to-note distances for consonant dyads. For example, the lowest note (F) against an A two blocks higher yields a "3," or third, a consonance. The intersection of the same low F against the B three blocks higher is empty, which indicates the fact that they are a dissonance. Note also the music notation above the checkerboard, in which "a Muse stands pointing at a phrase in three parts, the triumphant result of these compositional aids" (Godwin 1979). A transcription from Barton (1978) is shown in Figure 2–4.

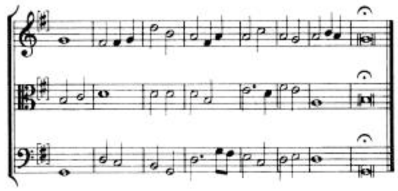



Figure 2–4. Transcription of the Muse's chorale from Fludd's *The Temple of Music* by Todd Barton.

Fludd's influence is clearly felt in the work of English composer and theorist Christopher Simpson (c. 1605–1669). Portions of his sprawling treatise on viol playing, *The Division-Viol* (1665), feature rather enigmatic descriptions of the power of musical numerology and the Music of the Spheres. Section thirteen, entitled "Reflections on the Concord of Musick," relates the author's conviction of divine numerology in sound:

> *And here I cannot but wonder, even to amazement, that from no more than Three Concords, (with some intervening Discords) there should arise such an infinite variety, as all the Musick that ever has been or ever shall be composed. And my wonder is increased by a consideration of the Seven Gradual Sounds of Tones, from whose various positions and Intermixtures those Concords and Discords do arise. These Gradual Sounds are distinguished In the Scale of Musick by the same seven Letters which in Kalender distinguish the seven dayes of the Week; to either of which, the adding of more is but a repetition of the former over again.*
>
> *This Mysterious number of seven, leads me into a contemplation of the Universe, whole Creation is deliverd unto our Capacity (not without some mystery) as begun and finished in seven dayes, which is thought to be figured long since by Orpheus his seven stringed Lyre. Within the Circumference of this great Universe, be seven Globes or Spherical Bodies in continual Motion, producing still new and various figures, according to their diverse positions one to another.*

Simpson continues as he draws connections to the twelve-membered Zodiac by noting the corresponding twelve-tone scale in common use.

About fifty years later, these concepts began to leave the fringe and crept into more formalized and widely accepted theories of music. Rameau's 1722 *Traite de l'Harmonie* and later writings by the mathematician Jean Le Rond d'Alambert developed a dissonance theory based on a musical memetics of nature whereby we perceive interval ratios between adjacent higher harmonics as increasingly dissonant. They argued that the intervals among lower harmonics were, in the analysis of Sir James Jeans (1937), "most consonant to the scheme of nature" because most sounds found in nature are harmonic and can be analyzed in terms of their harmonics relative to the fundamental bass.

The tradition of Music of the Spheres is remarkable in the continuity of writings about it, stretching forward to Leonhard Euler (1707–1783), whose 1739 work *Tentamen Novae Theoriae Musicae ex Certissismis Harmoniae Principiis Dilucide Expositae* ("Attempt at a New Theory of Music, Exposed in All Clearness from the Most Well-Founded Principles of Harmony") explores dissonance and attempts to make music a "part of mathematics and deduce in an orderly manner, from correct principles, everything which can make a fitting together and mingling of tones pleasing." (See Bailhache 1997 for a complete discussion.) For Euler, the perception of order and perfection was tantamount to consonance, and his *gradus suavitatis* (literally, "degree of pleasantness") espoused in the *Tentamen* attempts to quantify consonance through purely numerical means, in true Pythagorean fashion. As Leman (1995) notes, Euler was attempting to provide arithmetic logic behind Gottfried Wilhelm von Leibniz's (1646–1716) idea that the soul "secretly" calculated ratios of musical intervals. Even some modern theories of musical dissonance are based in part on modifications of Euler's writings or pay homage (e.g., Vogel 1993; Leman 1995). The basic idea is simple: the more "complicated" the numbers involved in an interval, the less "pleasing" the result. Leman (1995) summarizes from the *Tentamen* the basic principle, which is based on the fact that any number a can be decomposed into a product of n prime numbers $p_1,\ldots, p_n$, each raised to a corresponding exponent $e_1,\ldots, e_n$:

$$a = \prod_{k=1}^{n} p_k^{\,e_k}$$

Euler's *gradus suavitatis* measure Γ of an interval a is then given by

$$\Gamma(a) = 1 + \sum_{k=1}^{n} e_k (p_k - 1)$$

In the case of just intonation, in which a can be expressed as the ratio of two rational numbers p and q, we define

$$\Gamma\left(\frac{p}{q}\right) \equiv \Gamma(p \cdot q)$$

For example, consider the interval of a 5:4 major third. Here, $\Gamma(5/4) = \Gamma(20)$. But 20 can be decomposed as $5^1 2^2$. Thus we have

$$\begin{aligned}\Gamma(20) &= 1 + 1(5-1) + 2(2-1) \\ &= 7\end{aligned}$$

Plotting Euler's Γ operator as a function of various intervals provides what Leman calls the *tone profile* of the scale. We will return to this concept later.

Before continuing with our discussion of Neo-Pythagorism, it bears parenthetical mention here the extent to which Euler's measure of consonance as the numerical "simplicity" of an interval's ratio carried forward throughout history. The *Psychologische Studien* (second edition, 1905) of Theodore Lipps (1851–1914), in particular, documents the state of *fin de Siècle* approaches to the study of consonance from the viewpoint of leading German psychologists, including Wilhelm Wundt (1832–1920), often called the father of experimental psychology, Carl Stumpf (1848–1936), Max Meyer (1873–1967), and Felix Krüger (1874–1948). Lipps states that "[o]ne way or another, we can't help basing consonance on vibration ratios" (Lipps 1995, p. 91). He immediately continues:

> *So this is my theory. It may yet be possible to find a different basis for consonance from mine, one based on the simplicity of ratios. In any case, the fundamental idea of my theory remains, its foundation of consonance in ratio simplicities, dissonance in the opposite.*

Mathematics, physics, and psychology were not the only disciplines to address the numerical implications of consonance. Echoes of the more cosmic concepts associated with Pythagoreanism can be found also in the writings of philosopher Arthur Schopenhauer (1788–1860), whose *The World as Will and Representation* (1819/1844) refers to music as manifestation of the human will:

> *Music is a means of making rational and irrational relations of numbers comprehensible, not like arithmetic by the help of the concept, but by bringing them to a knowledge which is perfectly, directly, and simultaneously sensible. Consonances and dissonances, with their innumerable degrees of difference, portray the movements of the human will in its essential feelings of satisfaction and dissatisfaction.* (Quoted in Malmberg 1918, p. 96)

The tradition of Music of the Spheres was also carried forward through the 18th and 19th centuries in writings of Hassidic Jews, in particular the Kabbalah, with its emphasis on "sacred geometry" and the concept of the "Tree of Life." Indeed, the nexus of esotericism, symbolsm, and mysticism is strong in much Western music, including of course the Masonic-referential works of Mozart, including his *Masonic Funeral Music* (K.477), *Eine Kleine Freymaurer Kantate* (K. 623), and *Die Zauberflöte* (K. 620).

One of the even more esoteric, philosophical interpretation of the Music of the Spheres is found in the works of Hans Kayser (1891–1964) and Albert Freiherr von Thimus (1806–1878). Kayser's seminal work *Die Harmonikale Symbolik Des Alterthums* (1868–76) takes a simultaneously historical and cosmological view on the Harmony of the Spheres tradition. Kayser, who corresponded with Arnold Schönberg for a time, wrote a tome entitled *Lehrbuch der Harmonik* (1950) that explores Pythagorean harmonics at great length[i]. These works assert the cosmological significance of the unison (1/1) as God. Thimus in particular developed a certain harmonic diagram of Iamblichus (the Lambdoma) into what he referred to as the "Pythagorean Table" (Levary and Levy 1983; see Figure 2–5).

---

[i] An English translation is currently underway at the time of this writing.

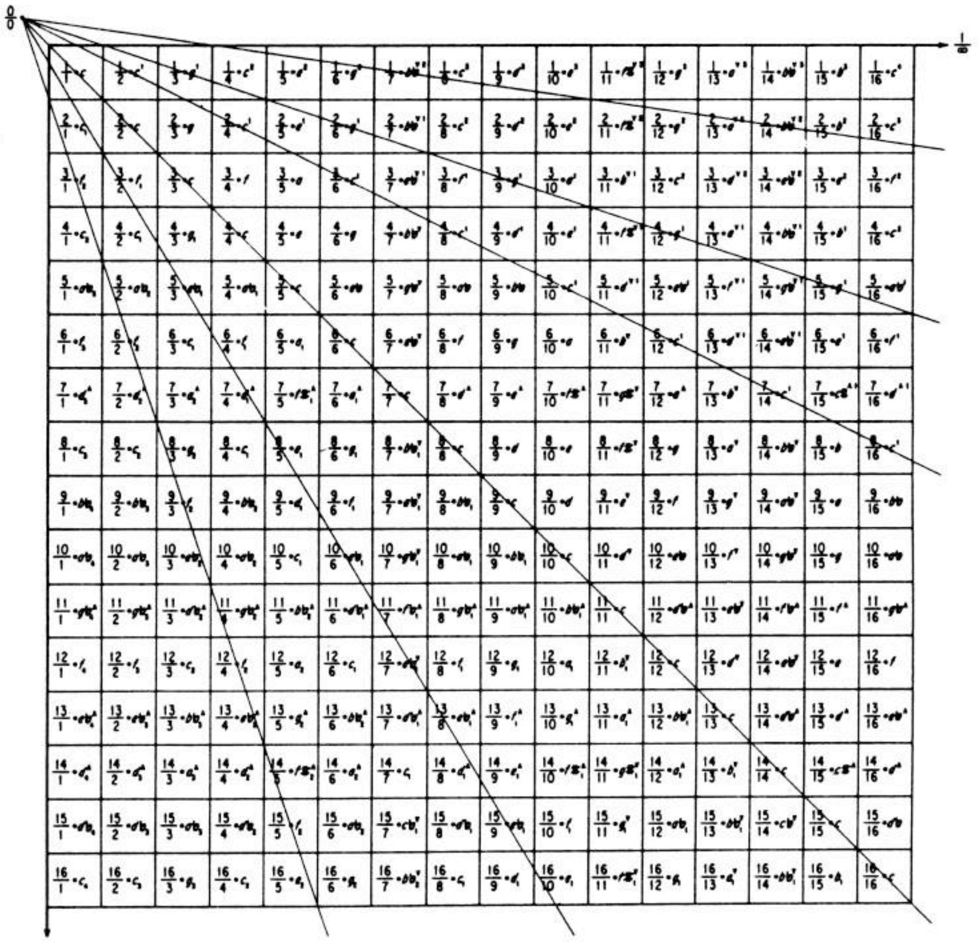


Figure 2–5. The "Pythagorean Table" of Albert Freiherr von Thimus (1806–1878), from Godwin (1995).

Thimus' Pythagorean Table is a simple matrix capable of expressing all possible combinations of integers, but with a curious property. As shown in Figure 2–5, connecting all possible numerically equivalent ratios (e.g., drawing a line connecting 6/2, 12/4, and 18/8) results in a set of lines that intersect graphically at a point that lies outside the matrix. Kayser called this ratio 0/0, representing the absence of being, of complete nothingness or Nirvana (and hence reflected physically in silence). Godwin (1987) quotes Kayser's comments on the duality of 1/1 and 0/0 in his work *Akroasis*:

> *[H]erein lies, in the symbolism of harmonics, a consoling certainty. In spite of being torn by the strife between light and dark…, in spite of consonance and dissonance, each single existence-value with its reincarnations is directed toward the divine, whence it receives its true innermost value.*

And so the philosophical quest for metaphysical interpretations of physical phenomena like simple harmonic motion and the overtone series continues.

❧

Clearly, the concept of Music of the Spheres is by definition tied to a view of consonance as certainty and order and dissonance as uncertainty and disorder in the grand scheme of the cosmos. This is a view that has manifest itself more recently in the writings of Paul Erlich, who, according to Monzo (2004), defines *harmonic entropy* as a measure of "the dissonance of an interval based on the uncertainty involved in interpreting that interval in terms of an integer ratio." He emphasizes that the notion of harmonic entropy "is intended to be a second component in measuring the sonance of an interval, alongside roughness." Thus, the modern notion of harmonic entropy as used in the music-tuning community is still rooted in Pythagorean concepts of order and the numerical beauty of simple integer ratios.

The words of Fludd, Simpson, Euler, and even later writers in many ways read just like those of philosophers well over a millennium earlier concerning dissonance and musical theosophy. The concept of Music of the Spheres has even informed and served as a focus for much twentieth-century music. In addition to Xenakis' *La legende d'Eer* mentioned earlier, important twentieth-century works following in this line include George Crumb's *Makrokosmos I*, and of course Holst's *The Planets Suite*, Op. 32 (1914–1916). In the words of Xenakis, "We are all Pythagoreans" (1977, p. 40).

2.2 All Pleasantries Aside

Other definitions of consonance tie the concept to the emotion of pleasantness. Again, Boethius was among the first to write of this connotation, in *De Institutione Musica* IV.1:

> *Consonae quidem sunt, quae simul pulsae suavem permixtumque inter se coniungunt sonum.*
>
> *Consonant pitches are those which when struck at the same time sound pleasant and intermingled with each other.* (Tr. Bower, p. 116)

But for Boethius, the sensation of pleasantness was not due to simple experiential pleasure per se, but rather an ineffable sense that one was communing directly with the grand order of the cosmos. As Umberto Eco notes, "Microcosm and macrocosm are tied by the same knot, simultaneously mathematical and aesthetic" (1986, p. 31). Here, the effect of consonance is a pleasing sensation caused by corporeal response to divine proportion. Pythagoreanism here meets emotion in a linear sense: the laws of the cosmos dictate the laws of music and the kinds of music we make; the perception (whether conscious or not—it does not matter) of divine proportions in music triggers a sensation of pleasantness; and we call this pleasant sensation "consonance." Such a schema is depicted in Figure 2–6.

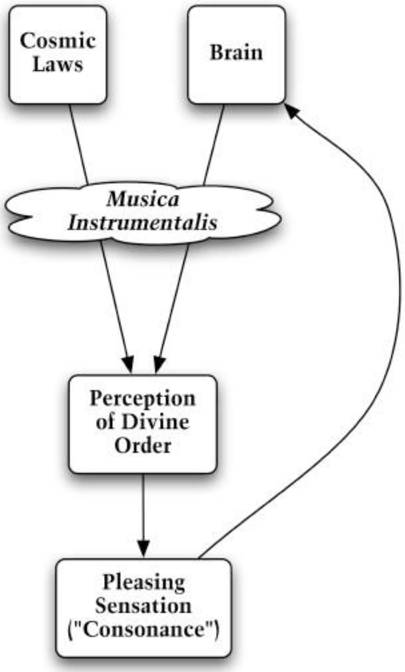


Figure 2–6: Boethius' conception of musical consonance.

Along this same line of thought (and over a millennium later), Leibnitz and Euler believed that an unconscious mechanism calculated the mathematical ratios present in intervals, and some unknown process conveyed this information to perceptions of agreeableness and pleasantness. Consonance and dissonance were not only about order and mathematical purity, but more accurately a function of pleasantness that resulted from cognition of these traits. In a sense, this notion grew from an infusion of Neo-Pythagoreanism with a primitive speculative psychology. Interestingly, the pleasure that results from musically "consonant" intervals was not ascribed to music-theoretic principles, but rather to perception (rather conscious or subconscious) of divine order, just as Boethius wrote much earlier. Colloquially, we might say that consonance "sounds good" not because it "sounds good," but because it conveys a fragment of divine order to humankind, which itself is arbiter and in turn bringer of pleasure.

Another approach to the equation of consonance with simple auditory pleasure is found in the writings of Johannes Tinctoris (c. 1430–c. 1511), whose 1477 treatise *Liber de Arte Contrapuncti* asserts that

> *Counterpoint is therefore a combination of tones. If this combination or mixture sounds pleasant, it is called consonance; if, on*

> *the other hand, it sounds harsh and unpleasant, it is called dissonance.*
> (Quoted in Jeppesen 1939.)

The return to a Pythagorean equation of pleasantness with perception of numerical order was instigated by Euler. For Euler, the consonance level of an interval was correlated to the numerical simplicity of the ratios involved in the interval. In general, the greater the sum of the numerator and denominator of the ratio, the more dissonant the interval, as shown in Table 2–1 after his 1739 *Tentamen Novae Theoriae Musicae*.

Numerator	**Denominator**	**Ratio**	**Dissonance**	**Notes**
1	1	1.00	2	Unison
9	8	1.13	17	3-limit major second
5	4	1.25	9	5-limit major third
4	3	1.33	7	Perfect fourth
3	2	1.50	5	Perfect fifth
5	3	1.67	8	3-limit major sixth
15	8	1.88	23	3-limit major seventh
2	1	2.00	3	Perfect octave

Table 2–1. Dissonance of selected intervals, according to the method of Euler's 1739 treatise *Tentamen Novae Theoriae Musicae.*

Euler's goal was to derive a correspondence between numerical complexity and unpleasantness: complexity of interval is inversely proportional to pleasantness. As Malmberg (1918, p. 103) observed:

> *Euler…agreeing essentially with Leibnitz' explanation, interpreted the feeling of agreeableness of the consonances as due to the ease of perceiving order or coherence in the simpler ratios. He divided the consonances into ten classes, ranking them according to the simplicity of their ratios. Euler was the first scientist to formulate the fundamental law of consonance that "the degree of consonance is in a direct ratio to the magnitude of the common divisor of the vibration frequencies.*

A more immediate interpretation of consonance as pleasantness was adopted by Malmberg (1918), whose study was perhaps the first comprehensive psychological approach to the study of dissonance. In short, Malmberg conducted listening tests in which he asked subjects which intervals they preferred, i.e., found more pleasant. The goal here was to establish a "standard order from the best consonance to the worst dissonances" (1918, p. 120) through experimental listening tests, a concept which formed the basis for many such "dissonance rating" tests since.

This connotation of consonance with pleasantness is still very much alive in modern times, as Plomp and Levelt (1965, p. 551) write:

> *For naive subjects…consonance and pleasantness are…similar concepts, as was demonstrated by the authors in an experiment in which 10 subjects had to judge a large number of intervals on 10*

> *different semantic scales. A high correlation between consonance and pleasantness scores was found. …[F]or the naive subject the notions consonance and pleasantness are nearly identical.*

The association of Plomp's and Levelt's definition of consonance—which they called "tonal consonance"—with pleasantness was reinforced by Kameoka and Kuriyagawa (1969) in their similar definition of "sensory consonance." The lineage of writings on consonance and its connotation with pleasantness can be traced to more recent writings on "auditory consonance" and "auditory disgust," which we treat separately later. Interestingly, recent neurological studies may indicate a fundamental link between the perception of dissonance and emotional states of pleasure and displeasure.

As a final aside on the subject, a unique approach to assessing ratings of auditory "likes" and "dislikes" in terms of subjective pleasentness is found in the work of R. Murray Schafer (1994) in the context of acoustic ecology. Schaeffer asked residents of various locations around the world to assess whether they liked or disliked particular sounds. The results, tabulated in his work *The Soundscape*, provide an interesting insight to one aspect of a final theory of sound-object dissonance. We will return in greater detail to Shafer in Chapter 6, where the results of his survey will be presented.

Research into the interplay of sensations of "pleasantness" and "likability" with the concepts of consonance and dissonance continues. For example, Ritossa and Rickard (2004) studied the use of these sensations in predicting emotional states induced in listeners on hearing a musical passage. Furthemore, related applications have evolved in the commercial world (e.g., MoodLogic; see http://www.moodlogic.com) in which pleasantness and emotional classifications are used to categorize large databases of music content, such as found in personal jukeboxes.

2.3 Dissonance and Instability

Another use refers to consonance as stability and dissonance as instability. Somehow, this view seems to permeate many reference works on music and for the general reader. The *Harvard Dictionary of Music*, for example, defines consonance and dissonance as "[t]he perceived stability or instability of a complex of two or more sounds" (Randel 1986). The idea that dissonances must be resolved to stable consonances—and methodologies for doing such—have occupied a great deal of the writings on tonal theory for two centuries. The equation of dissonance with instability has had far more to do with the classification of the perfect fourth as a dissonance by some tonal theorists than any acoustical definition of the term. (For example, the perfect fourth is generally unstable in two-voice counterpoint because it "should" resolve to the third.)

The rules of sixteenth-century counterpoint, for example, categorized dissonances according to the degree to which they created instability or "obtrusion." Jeppesen, in his classic 1939 study on the subject (p. 98), writes:

> *The same wariness against abrupt or unclear effects which is characteristic of the Palestrina style in the linear treatment is evident in the treatment of chords. Dissonances are used only in restricted forms and in places where they do not produce an obtrusive effect. Their use may be divided into three principal categories:*
>
> 1. *Passing dissonances.*
> 2. *Suspension dissonances.*
> 3. *Auxiliary dissonances (that is, dissonances which are introduced by step on weak beats and then return to the preceding tone.)*

Consonances are similarly categorized as perfect or imperfect, a tradition that has carried forward to the modern day.

The sense of dissonance as instability has informed much of the American music-theory textbook tradition. A classic example can be found in Walter Piston's 1941 text *Harmony*, in which the author defines a consonant interval as "one which sounds stable and complete" and characterizes dissonant intervals by their restlessness and…need for resolution to a consonant interval" (1978, p. 6). Whence did an interval *need* to do anything?

Piston, as well as most introductory texts that follow in its footsteps, conveniently outlines the "consonant" and "dissonant" intervals:

Consonant:	*the perfect intervals and the major and minor thirds and sixths;*
Dissonant:	*the augmented and diminished intervals and the major and minor seconds, sevenths, and ninths;*

However, an exception is noted that apparently results from the notion of dissonance as instability:

Exception: *the perfect fourth is dissonant when there is no tone below its lower tone. It is consonant when there is a third or perfect fifth below it.*

And thus the fourth in a I_4^6 chord is deemed dissonant, presumably owing to the tonal instability that results owing to the lack of its more stable grounding on the root of the chord.

Indeed, the formulation of dissonance as the perceptual correlate of musical stability seems to be the prevailing sentiment in modern music theory literature. This notion is also present in the work of Lerdahl and Jackendoff (1977), who assert, “Broadly, the relative stability of a pitch-event can be thought of in terms of its relative consonance or dissonance” (p. 117). One recent music theory textbook simply states:

> *Intervals are consonant if they produce a sense of stability. Dissonant intervals, on the other hand, create a sense of tension or instability, which we normally perceive as a clash that requires resolution to a consnance.* (Roig-Fancolí 2003, p. 15)

❧

The notion of stability introduces three important concepts into conceptions of dissonance, namely *motion*, *expectation* and *stability*. The musical relationship of stability and motion is again expressed in Piston (1978, p. 7):

> *Music without dissonant intervals is often lifeless and negative, since it is the dissonant element which furnishes much of the sense of movement and rhythmic energy…. It cannot be too strongly emphasized that the essential quality of dissonance is its sense of movement and not, as is sometimes erroneously assumed, its degree of unpleasantness to the ear.*

Paraphrasing this sentiment, it is often said that dissonance is the spice that "wakes up" an otherwise bland musical fabric. The relationship between dissonance and motion is also casually cited by many listeners in expressing their preference for equal temperament over just intonation, the idea being that the (perhaps learned) out-of-tuneness of equal temperament contributes to its apparent linear motion and potential forward progression. A counterargument lies in the truism that, for a given scale, just intonation can easily be constructed to “sweeten” particular consonances and “darken” particular dissonances.

Expectation involves two matters: culturally agreed-upon rules for manipulating musical expectation (e.g., a I–vi–ii–V^7 progression in tonal music, accompanied by a ritard, creates an expectation of resolution to the tonic); it also even perhaps involves acoustical principles of tension and release in that there do in fact exist some acoustical bases for certain rules of tonal music (Hutchinson and Knopoff 1979).

Stability, then, can be defined in musical contexts in either physical or cognitive ways as the result of both motion and expectation, specifically, the degree to which motion and expectation seem to agree. Physically, stability invokes basic principles of gestalt psychology that lie well beyond the scope of this essay as well as centuries of cultural indoctrination. Suffice it to say that this kind of stability is the kind invoked in textbook definitions of ending a piece of common-practice tonal music on a I chord. Similar “laws” dictate that a so-called “perfect authentic” V^7–I cadence should invoke a stronger sense of final stability at the end of a progression than should a ♭VI–I cadence, for example.

Physical components of the percept of stability are addressed elsewhere in the recent literature on consonance and dissonance and more general psychological studies. We might encapsulate the notion of physical stability of an isolated sound object in terms of a variety of acoustical factors, including perhaps some weighted combination of the quantifiable properties of harmonicity, spectral centroid, spectral flatness, and a variety of other features that can be extracted from a sound file using digital signal processing techniques. (We will return to this idea in Chapter 5.) One could also compute stability in terms of physical, acoustical features for a linear progression of sounds over time.

2.4 The Physics of Dissonance: Helmholtz' *Konsonanz*

As Hutchinson and Knopoff (1979) observe, the first account of the relationship between acoustical beating and dissonance was offered by Joseph Saveur in 1700. The idea was simple: the presence of beats accounts for dissonance, and the absence of beats accounts for consonance. Saveur's examination initiated a new wave of scientifically informed observations on the subject of dissonance, particularly in their consideration of the principles of physical principles.

The examination of dissonance from a physical—and soon purely acoustical—viewpoint culminated in the work of Hermann Ludwig Ferdinand von Helmholtz (1821–1894), whose 1862 book *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* ("On the Sensations of Tone as a Physiological Basis for the Theory of Music") eloquently summarizes late-nineteenth-century scientific thought on acoustics, music theory, and dissonance. The importance of this work of course cannot be overstated, leading as it did to important discussion, debate, and reference many years after its publication (e.g., Gurney 1880; Heffernan 1887; Jeans 1937).

Beginning with Helmholtz, the ranking of the relative dissonance of intervals led to the "modern" experimental quantification of so-called dissonance curves. In Figure 2–7, the *y*–axis represents an arbitrary dissonance level of dyads played on a violin, while the *x*–axis represents interval size (from 1:1 to 2:1). Helmholtz proposed a theory of dissonance based on the relative amount of acoustical beating that occurs among partials for a given interval and a given timbre:

> *When two musical tones are sounded at the same time, their united sound is generally disturbed by the beats of the upper partials, so that a greater or less part of the whole mass of sound in broken up into pulses of tone, and the joint effects is rough. This relation is called* Dissonance.
>
> *But there are certain determinate ratios between pitch numbers, for which this rule suffers an exception, and either no beats at all are formed, or at least only such as have so little intensity that they produce no unpleasant disturbance of the united sound. These exceptional cases are called* Consonances. (p. 194)

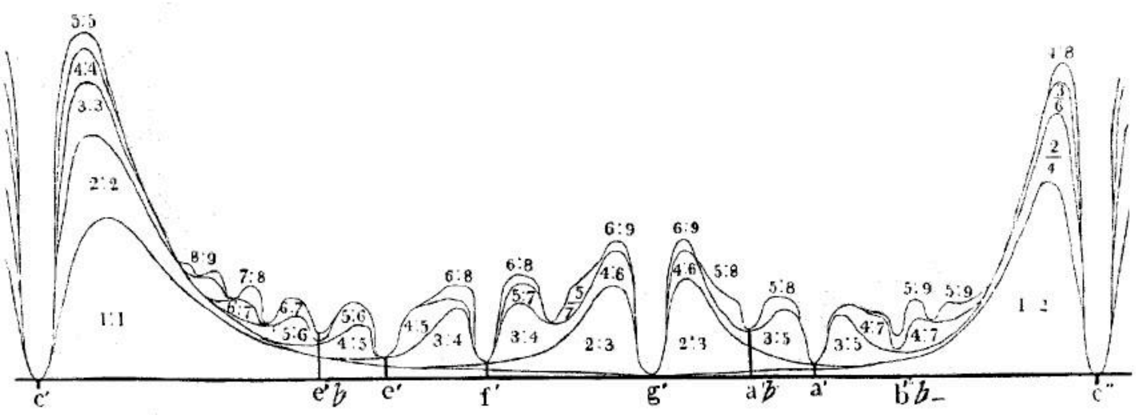


Figure 2–7. Helmholtz's plot of the relative consonance of harmonic-tone intervals (1885).

For example, the interval of the perfect octave is consonant because the majority of the overtones of each pitch are coincident, while the interval of a minor second is dissonant because the majority of the overtones of each pitch are slightly skew, causing beating among the overtones. Helmholtz's beat theory of consonance dominated the acoustical literature for almost a century after it was published, and it initiated a fifth use of consonance to refer specifically to *Konsonanz* (what was later called "sensory consonance")—as distinct from what he termed *Klangverwandtschaft* (context-dependent common harmonic practice).

Basing his definition of consonance on the absence of beats among partials, Helmholtz' beat theory in a sense circularly defines dissonance in terms of that which is not consonant. The beat theory has been faulted for three primary reasons. First, the presence of beats does not explain the marked preference for stretched octaves: intervals slightly larger than a perfect octave exhibit greater beating among partials in harmonic sounds, yet subjects generally hear these "stretched" octave ratios as more consonant than a true $\frac{2}{1}$ ratio. As Keislar (1991) notes, Western-trained musicians tend to prefer equal temperament to just intonation, presumably owing to cultural factors[ii]. Second, musicians tend to identify certain intervals as consonant and others as dissonant, even in the absence of any overtones in the stimuli. Again, Keislar (1991) studied this phenomenon in trained musicians, concluding that frequency ratio per se was a greater determinant of consonance judgments than beating among partials. Third, Vogel (1993) describes dichotic binaural studies in which pairs of tones classified as dissonant according to Helmholtz' beat theory were played, one tone in each ear, to test subjects. Presenting dissonant intervals in this manner, researchers found that the intervals were no longer perceived as dissonant.

That being said, the legacy of Helmholtz' scientific inquiry into the physical nature of musical dissonance was groundbreaking, and many aspects of it are not been disproven after more than a century. For example, Jacobsson and Jerkert (2000) found strong evidence to support Helmholtz' original beat theory, at least for trained musicians when rating inharmonic complex tones.

[ii] Other studies have examined cross-cultural ratings of consonance. Butler and Daston (1968) found similarity in interval ratings between Japanese and Americans, for example. On the other hand, Maher (1976) found marked dissimilarity in interval ratings among Indians and Canadians.

Helmhotz' legacy of the division of consonane into two distinct phenomena, one physical and the other psychacoustical and context-dependent, is perhaps the most important contribution to the study of musical dissonance in over a century, and this legacy carries onward. The components of sensory consonance and *Klangverwandtschaft* are treated extensively by modern writers, including Terhardt (1984, p. 278), for example, who defines consonance specifically within the confines of tonal music:

> *We consider the term musical consonance to be subsuming the principles that are regarded as governing tonal music. Those principles ordinarily are more or less loosely indicated by terms such as harmony, consonance, and dissonance. They can readily be verified by analysis of any piece of tonal music; thereby, typical and systematic tone relations (i.e., pitch and frequency relationships) will be revealed. The principle (whose nature to this point must be considered as unknown) that creates those specific relations is called musical consonance. …[I]t should be noted that it is this definition that establishes musical consonance as a link between music and psychoacoustics: On the one hand, musical consonance somehow represents certain essential features of tonal music; on the other hand, it can be reduced to established psychoacoustic phenomena such as pitch and roughness.*

Indeed, the twentieth-century concept of auditory roughness established one of the most important new bases for evaluation of musical dissonance in listening tests, and in particular led to more recent discussions of auditory masking as a primary contributor to the perception of dissonance. (We will return to this concept in section 2.8.) But this shade of meaning grows from a long-held notion of consonance as somehow reflective of unity and harmoniousness.

2.5 Harmoniousness, Purity, Roughness, Fusion

In its fourth meaning, consonance has also referred to properties of fit, belonging, harmoniousness, concord, or togetherness—both in aesthetic as well as cognitive senses of the term. Perhaps the earliest use of the English word "dissonance" to describe a lack of agreement or harmony in a general sense is found in William Caxton's *The boke yf Eneydos* (The Book of Eneydos vii, p. 32, 1490): "The maner of that countree…was all dissonaunt & dishoneste in regarde to that of Dydo." Caxton also uses "consonant" in as early as 1489 to denote harmony and agreement: "Thy raysons ben consonaunte" (*The book of fayttes of armes and of chyualrye* IV/xi, p. 260). Indeed, the equation of consonance with agreement and concord was probably the earliest use of the word in English, used in this way by monk and poet John Lidgate (c. 1370–c. 1450) in the *Chronicle of Troy* (1430).

In the aesthetic sense, this refers to the often ineffable way that we might say some pitches "work well together" or somehow fit together well; in electroacoustic music, we might make similar arguments for the conjoining of particular sounds on aesthetic grounds. In this light, Schoenberg's famous "emancipation of the dissonance" may be seen as an attempt to neutralize previously-held notions of the kinds of "togetherness" or "fit" that had constituted consonance for many centuries. In the cognitive sense, fit and harmoniousness have tended to invoke notions of "smoothness," "sweetness," and a lack of "roughness." Both meanings have existed for centuries.

Boethius, in Book I of *De Institutione Musica*, speaks to this very issue:

> *Quae sit natura consonantiarum.*
>
> *XXVIII. Consonantiam vero licet aurium quoque sensus diiudicet, tamen ratio perpendit. Quotiens enim duo nervi uno graviore intenduntur simulque pulsi reddunt permixtum quodammodo et suavem sonum, duaeque voces in unum quasi coniunctae coalescunt; tunc fit ea, quae dicitur consonantia. Cum vero simul pulsis sibi quisque ire cupit nec permiscent ad aurem suavem atque unum ex duobus compositum sonum, tunc est, quae dicitur dissonantia.*

> *28. What the nature of consonance is.*

> *Although the sense of hearing recognizes consonances, reason weighs their value. When two strings, one of which is lower, are stretched and struck at the same time, and they produce, so to speak, an intermingled and sweet sound, and the two pitches coalesce into one as if linked together, then that which is called "consonance" occurs. When, on the other hand, they are struck at the same time and each desires to go its own way, and they do not bring together a sweet sound in the ear, a single sound composed of two, then this is what is called "dissonance."* (Tr. C. M. Bower)

Boethius' comments were exceedingly long-legged, for they are echoed almost verbatim many centuries later. The fusion theory, the first modern explication of which was offered by Stumpf (1898), holds that the consonance of an interval is directly proportional to the degree to which the interval tends to fuse and provide a single sonic gestalt. Stumpf defines the term *Verschmelzung* as the primary factor in ranking the consonance and dissonance of intervals:

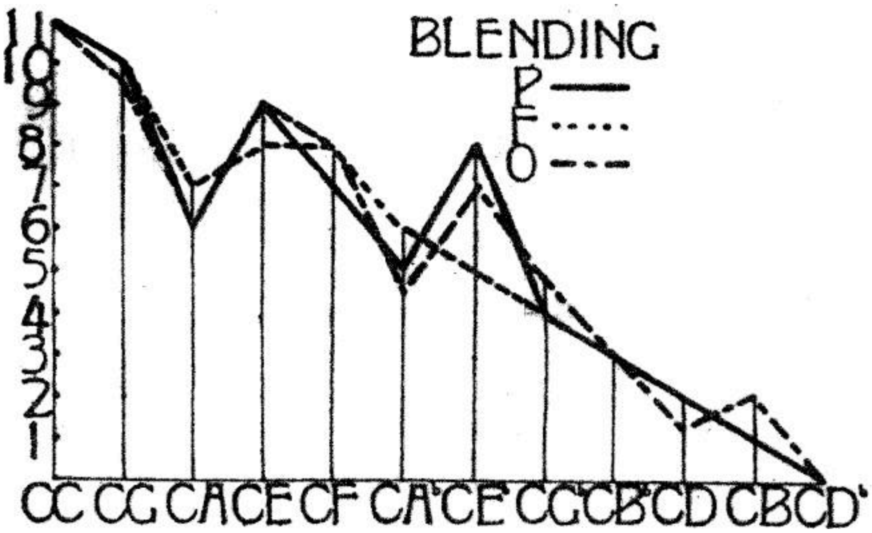
> *The sounding together of two tones approaches sometimes more, sometimes less, the impression of unity, and it is apparent that this is more the case, the more consonant the interval is. Even if we recognize the tones as two and separate from one another, yet they form a totality in the sensation, and this totality appears to us as possessing a greater or less degree of unity.* (Quoted in Malmberg 1918, p. 97)

Stumpf attempted to measure fusion in terms of subject reaction time, i.e., how long it took for subjects to perceive whether they heard one tone or two. Although the specifics of Stumpf's fusion theory were soon rejected (even by Stumpf himself) and aspects of it disproven in the psychological literature (due partially to his experimental methods and later examinations; see, for example, DeWitt and Crowder 1987; Huron 1991), the gist of consonance as sonic union was reiterated by Malmberg in his 1918 study, arguably the most complete survey of dissonance from a variety of disciplines at that point in time. Based on listening tests in which subjects were asked to rank the perceived dissonance of intervals played on tuning forks, a piano, and a pipe organ, Malmberg codifies four contributors of consonance: blending ("a seeming to belong together"), smoothness ("relative freedom from beats"), fusion ("a tendency to merge into a single tone, unanalyzable"), and purity ("resultant analogous to pure tone"). He defines dissonance as the "reciprocal" of consonance, identified also by four factors: disagreement ("incompatibility"), roughness ("harshness, unevenness or intermittence"), disparateness ("separateness or…'twoness'), and richness ("resultant analogous to rich tone"). He then offers the most complete synthesis of current musical and psychological theory on consonance in his definition:

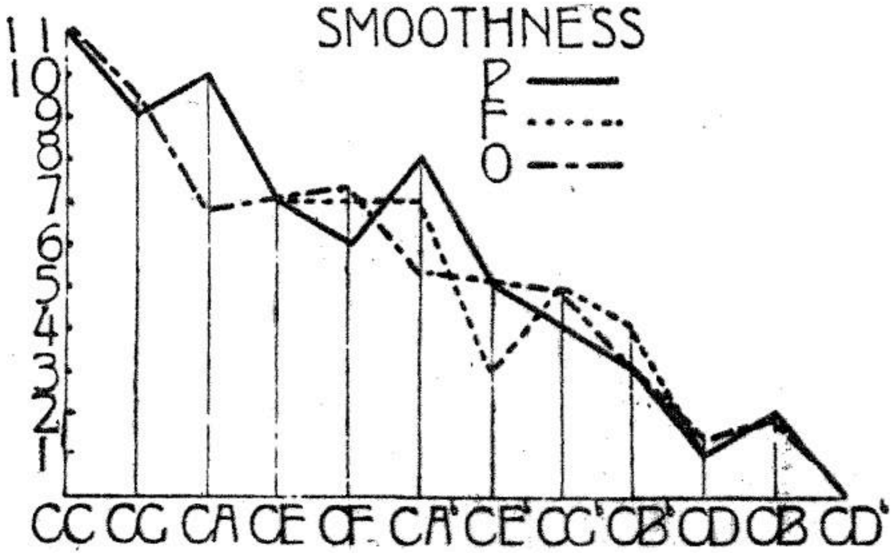
> *When the two tones of a two-clang tend to blend or fuse and produce a relatively smooth and pure resultant, they are said to be consonant. Dissonance is the reciprocal of this. "Agreeableness" which*

> *has played an important rôle in the popular conception and in the theory is here conspicuous by its absence. The perception of consonance as above defined therefore becomes a cognitive act of discrimination rather than a mere feeling of agreeableness. (Malmberg 1918, p. 108)*

After conducting listening tests of subjects' rankings of various dyads according to each of the four factors, Malmberg ranked each accordingly in decreasing order of blending, smoothness, purity, and fusion. For example, Figure 2–8 illustrates the results for two of the factors—(a) blending and (b) smoothness—of dyads played on the piano ("P"), tuning forks ("F"), and organ ("O"). Similar rankings were produced for purity and fusion.



(a)



(b)

Figure 2–8. Results from Malmberg's (1918) listening tests ranking (a) blending and (b) smoothness of dyads. Similar rankings were conducted for purity and fusion.

The results of the listening tests were synthesized into a "standard order from the best consonance to the worst dissonance" (p. 120) for dyads on a piano, as illustrated in Figure 2–9. The results differ from those of Helmholtz, especially with respect to the tritone and the minor seventh.

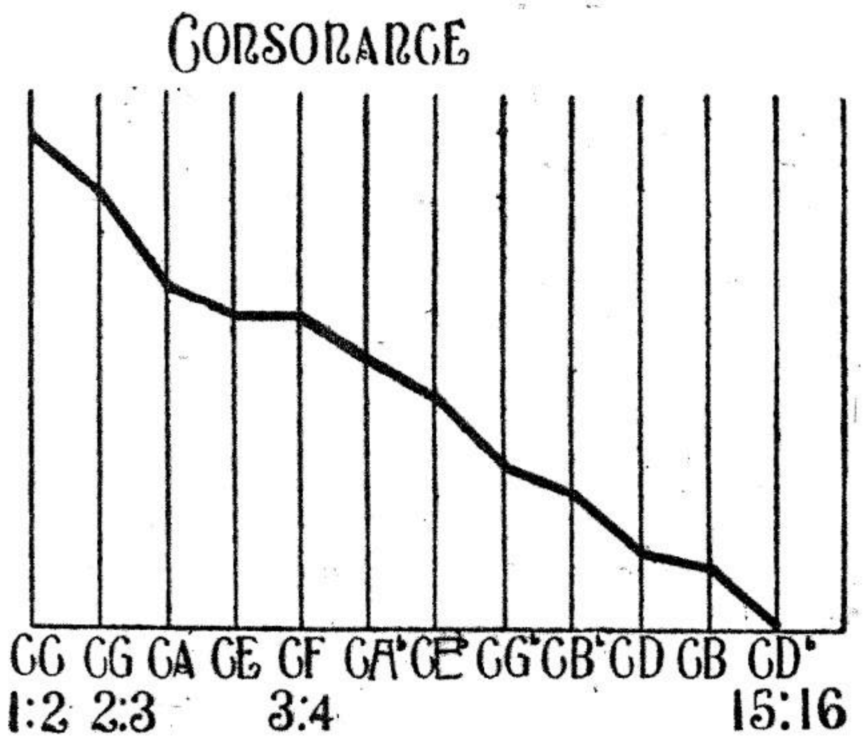


Figure 2–9. Malmberg's (1918) standardized consonance ratings for dyads on the piano.

Malmberg's codification of consonance as decomposable into factor of blending, smoothness, fusion, and purity was the prevailing sentiment for the first half of the twentieth century, echoed in Carl Seashore's seminal text *Psychology of Music* (1938). In Chapter 10, "Consonance," Seashore primarily summarizes Malmberg's results.

One of the fundamental problems in this approach to dissonance lies in the inherent conflation of fusion and "oneness." As Huron (2005) notes:

> *Bregman (1990) has pointed out that it is important not to conflating two different auditory experiences: "smooth sounding" versus "sounding as one."*

Clearly, the view of consonance as harmonic fusion and oneness provides only a partial glimpse into the entire world of dissonance.

Modern technical examination of the cognitive relationship between roughness and dissonance began with Plomp and Levelt (1965), who formally initiated a sixth use of "consonance" by defining *tonal consonance* in terms of the relationship of frequency ratio to the critical bandwidth (roughly a minor third for intervals over about 100 Hz)[iii]. The critical bandwidth is the interval around which a sensation of "roughness" occurs; smaller intervals tend to be heard as beating or chorusing of the fundamental, while larger intervals tend to be heard as two discrete tones[iv]. The existence of this phenomenon serves as a concrete example of the tangency between timbre, interval, and harmony.

[iii] Actually, the first study that investigated the relationship between critical bandwidth and perceptual dissonance is found in Greenwood (1961), although most of the more recent literature seems to defer to Plomp and Levelt's study.

[iv] The bark scale, named after German physicist and acoustician H. Barkhausen, attempts to model the percept of critical bandwidths; one bark is defined to be the width of one critical band. More recent studies (e.g., Smith and

Plomp and Levelt conducted listening tests and found that the perceived dissonance of simultaneous pairs of sinusoids was proportional to the critical bandwidth. Subjects in psychological tests chose the most dissonant dyads composed of pure tones to be those an interval one-quarter of the critical bandwidth apart, with a tapering off toward consonance for larger and smaller intervals to create a skewed inverted bell curve shape. (See Figure 2–10.) Plomp and Levelt's use of "consonance" thus refers to pairs of tones that lie at roughly the same frequency or at an interval greater than a critical bandwidth apart.

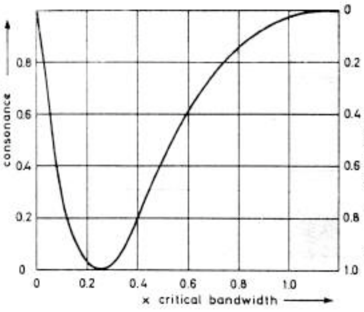


Figure 2–10. Tonal consonance is lowest at approximately 25% of a critical bandwidth between two pure tones (Plomp and Levelt 1965, p. 556).

Abel 1999) have found that the critical band scale is largely accurate when constructing cochlear filterbank models for machine-listening applications in the lower frequency range, but higher frequencies are better modeled using Equivalent Rectangular Bandwidth (ERB) filters (Moore and Glasberg 1996).

Extrapolating this characteristic to harmonic sounds by summing the contributions to dissonance of the interactions between each partial, the authors computed dissonance curves which greatly resembled Helmholtz' for a violin almost a century earlier. More recent evidence for the relationship between critical band and dissonance is provided by Simpson (1994), who applied cochlear models to analyze dissonance of chords. Considerations of roughness, critical bandwidth, and spectral components are tantamount to a full-blown analysis of the relationship between timbre and dissonance, to which our attention now turns.

2.6 The Contributions of Timbre

Although Helmholtz mentions the idea that timbre can directly affect musical dissonance, the tools available to him at the time certainly limited his scientific exploration of the possibilities. Malmberg (1918) experimented with a variety of sound sources and found different results for each source:

> *The order of the ranking of the intervals varies for different qualities of tone. The order has been established for tuning forks, piano, and pipe organ*. (p. 131)

But it was not until much later—and actually only recently—that systematic explorations of dissonance perception as a function of timbre and spectral content have been undertaken. Tools taken for granted in modern audio and music cognition research—spectrum analysis, sound synthesis, principle components analysis, and so on—were of course a major limiting factor.

Many did note the virtual absence of timbral considerations in dissonance curve generation and computation. One of the most blunt rejections of the many plots of consonance versus interval (of the kind produced by Helmholtz and Malmberg) owing to their overt generality is that of Partch (1949; 1974), who criticizes the idea of producing graphs and rankings of intervallic dissonance. In fact, the only such item in his encyclopedic *Genesis of a Music* is the classic "One-Footed Bride" graph, an almost-symmetrical plot of subjective consonance versus intervals alongside their complimentary ratios (e.g., the consonance of the augmented fourth 7:5 is plotted alongside that of the octave-normalized—but slightly different—augmented fourth 10:7). The graph is shown in Figure 2–11.

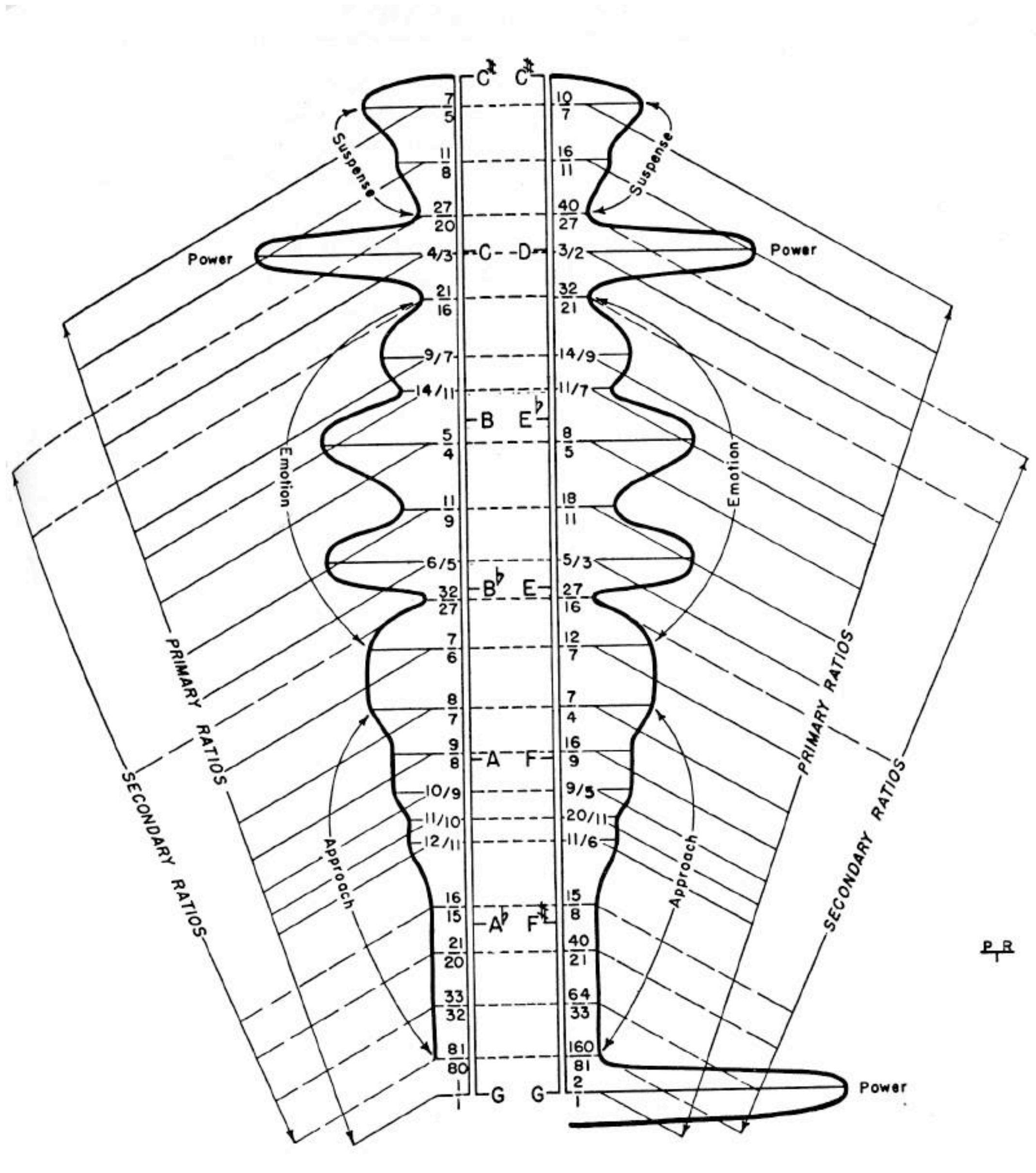


Figure 2–11. Partch's "One-Footed Bride" (1974, p. 155).

After explaining the graph's depiction of "Intervals of Power," "Intervals of Suspense," and "Intervals of Approach," Partch seemingly paradoxically renounces the idea outside of examinations of timbre and other relevant factors (p. 157):

> *It is fairly foolish to undertake to pin consonance to a graph less general than this unless it is predicated on specific range, specific quality of tone, specific relevance of combinational tones, and specific assurance that these qualitative and quantitative factors are invariable.*

However, he does acknowledge that "[s]hort of a lifetime of laboratory work which the composer cannot undertake, the general is the only practicable approach."

Early- and mid-century suggestions of the interactions between timbre and dissonance are found most notably in Carl Seashore (e.g., Seashore 1938), whose terminology for timbre ("sonance") even implicitly suggests its logical relationship with consonance and dissonance: that sounds of con-*sonance* literally "go together" timbrally, whereas sounds of dis-*sonance* do not.

I submit the following two observations. First, timbre, at some fundamental level, is an important, if not the primary, determinant of what most people consider musical dissonance. Consequently, it is entirely probable that we would arrive at the same dissonance judgment by listening to a high-quality recording of a dyad or chord as we would by listening to live instruments play the same pitch collection. Second, if we could devise a computer program to analyze the dissonance level of recorded dyads and chords with regard to a given set of features, it follows that such a program might in fact be useful in analyzing the relative dissonance of any kind of sound object, not just dyads or chords. This suggests that addressing the relative dissonance of sound objects, of clangs, might be useful in analyzing non-notated music, particularly computer music.

Given that the last century has seen a trend to thinking about dissonance in terms of spectral interactions between sounds—not just in terms of their frequency ratios—it is unfortunate that authors such as Hutchinson and Knopoff (1979), Danner (1985), and Marinis et al. (2005) ignore timbral contributions to perception of musical consonance. Contradictorily, Hutchinson and Knopoff reportedly take spectra into account by calculating dyadic consonance according a formula presented in Plomp and Levelt (1965). However, their discussion then proceeds without reference to timbre and spectra at all, ranking generic triads of unknown spectral composition according to consonance.

Similarly, Danner, basing his article on Hutchinson and Knopoff, ranks trichords according to their acoustic dissonance without regard to timbral constitution. Danner's graphical dissonance analysis of Elliot Carter's *Canon for 3* (1971) then has no meaning whatsoever, given that the work was written for unspecified instruments and that the analysis is not of a particular performance but of the score. The acoustical consonance—and, of course, the sensory consonance as well—will vary greatly depending on the chosen instrumentation. It is incredible that a century after the writings of Helmholtz that some theorists attempt to analyze musical dissonance without reference to instrumentation—or psychoacoustic modeling of any kind.

Euler’s measure of consonance, which we mentioned earlier, naturally grows out of the idea of harmonicity. For harmonic sounds at intervals that are closely related along the harmonic series (perfect fifths and major thirds, for example), the spectral components could be thought of as components of a missing or masked fundamental just as well as components of individual sounds. To a harmonicity algorithm, this would be of no consequence, because harmonicity computation does not need information on the kind or number of instruments involved in producing the sound. Whereas Euler’s measure in this case might incorrectly yield different consonance values for different intervals with the same sensory consonance, a harmonicity metric would quite possibly correctly identify both intervals as having the same consonance.

Recent research has begun addressing the interplay of dissonance and timbre (Bolger and Griffith 2003). However, if there is a recent movement toward thinking of consonance, dissonance, timbre, and tuning of musical scales as intertwined parts of a whole, the leader of the movement is William Sethares. Much of the most important recent examination of dissonance, however, and particularly the most technically informed one, is found in his work.

The notion that certain instruments sound “good” in certain tunings while others do not is not new; clearly, entire musical cultures, perhaps most famously in India and Southeast Asia, have responded to this observation by developing a myriad of scales intended for various instruments. Sethares has formalized this notion in great detail, and in doing so helped define how tuning, timbre, and consonance interact (Sethares 1993, 1998, 1999).

A starting point for Sethares is his parameterization of Plomp and Levelt's dissonance curves into a simple equation so that one can compute a numeric dissonance value for any interval between sine waves given their frequencies and amplitudes. He then extends the equation to complex sounds by using the formula to add the contribution of each pair of spectral components in the sounds. Sethares (1998), for example, summarizes his dissonance function $d(\cdot)$ (previously derived in Sethares 1993) that quantifies the dissonance between a pair of tones at frequencies $f_1$and f_2 at recpective amplitudes $v_1$and v_2:

$$d\left(f_{1,}f_2, v_1, v_2\right) = v_1 v_2 \left[e^{-as|f_2 - f_1|} - e^{-bs|f_2 - f_1|}\right]$$

where the scalar s is given by

$$s = \frac{d^\star}{s_1 \min(f_1, f_2) + s_2}$$

The parameters $a = 3.5$, $b = 5.75$, $d^* = 0.24$, $s_1 = 0.21$, and $s_2 = 19$ were deteremined by a least-squares fit of Plomp's and Levelt's data. The scalar d^* is the "interval at which maximum dissonance occurs," and the s function is employed to allow smooth interpolation among the various curves that Plompt and Levelt produced. (Recall that the actual bandidth of critical bands changes with respect to frequency.) Other example dissonance functions are provided in Haluska (2004).

Once an emperically determined dissonance function is accepted, the "total dissonance" D among all pairs of sine tones present in a complex tone comprised of Npure tone partials can be computed as

$$D = 0.5\sum_{i=1}^{N}\sum_{j=1}^{N} d(a_i f, a_j f, v_i, v_j)$$

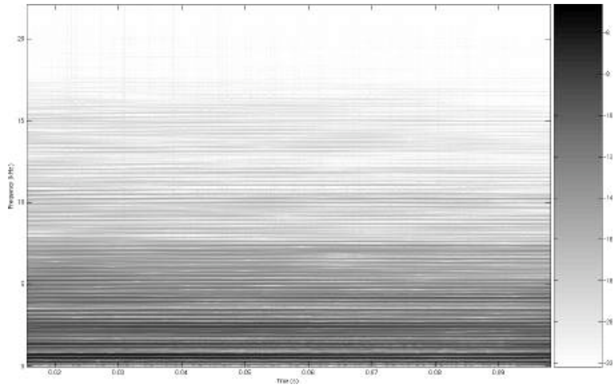
Here, we take the sound as decomposable in the Fourier sense into partials whose frequencies are given by the set $\{a_1 f, a_2 f, \ldots, a_N f\}$ with corresponding amplitudes of $\{v_1, v_2, \ldots, v_N\}$.

This computational process can be extended to real-world sounds and is illustrated graphically in Figure 2–12, which shows (a) the spectrogram of a sound file of a horn. In Figure 2–12(b), the eleven most-prominent spectral components have been threshholded to obtain the precise locations of the $\{a_1 f, a_2 f, \ldots, a_N f\}$ components.

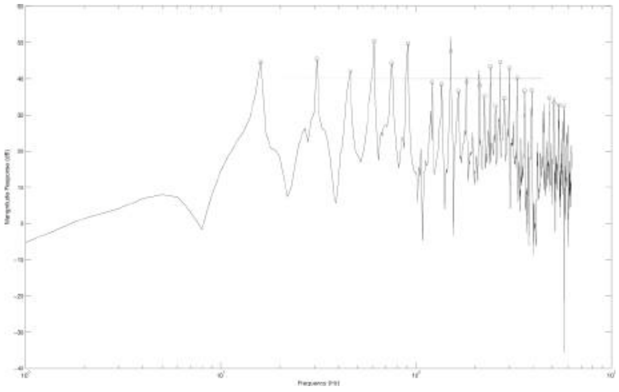
Here, the dissonance values between the lowest spectral component of the sound on the left and all components of the sound on the right are added. The process is then repeated for each component of the sound on the left. The sum total of these dissonance values yields the total dissonance of the interval.

Figure 2–12 (c) illustrates the first complete nested addition in the equation above; here, i is set to 1, and j is allowed to range from 1 to $N = 11$. Each arrow-tipped line denotes the next spectral line with which the dissonance function $d(\cdot)$

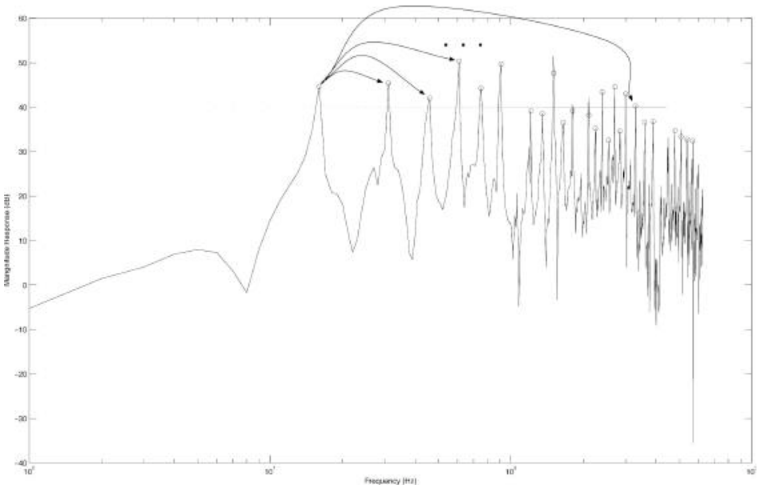
is calculated. Once all computations of dissonance values formed by pairing the fundamental with its harmonics have occured, all other possible pairs are then computed (i.e., $i = 1, 2, \ldots, 11$). Note that for time-varying sounds, the sound file must be windowed, and the total dissonance D must be computed separately for each frame.



(a)



(b)



(c)

Figure 2–12. Computing total dissonance of complex sounds, one partial at a time. (a) Spectrogram of a horn; (b) peak-picking; (c) calculating the sensory dissonance contributed by the lowest-frequency peak in a particular frame.

Remarkably, computing such dissonance curves for dyads of harmonic complex tones in which one sound's fundamental is fixed and other is allowed to continuously vary from a unison to an octave results in curves (tone profiles) extremely similar to those produced experimentally by listening tests. In particular, for a simple seven-partial harmonic spectrum, Sethares (1993) notes that nulls in the dissonance curves occur where we would expect, at 1:1 (unison), 7:6 (septimal minor third), 6:5 (five-limit minor third), 5:4 (five-limit major third), 4:3 (perfect fourth), 3:2 (perfect fifth), 5:3 (five-limit major sixth), 7:4 (harmonic minor seventh), and 2:1 (perfect octave). The greatest predicted consonances lie at the unison, perfect fifth, and octave.

An implication of his work is that a certain interval played on two instruments may have an identical dissonance as a different interval played on different instruments. For example, a chord played on a piano may have a near-identical sensory dissonance as a particular clarinet multiphonic. This raises the idea of developing an algorithm to compute the dissonance of a sound rather than an interval—without regard to its method of production or the number of instruments producing it, but rather the acoustical information alone present in the signal. What is needed is a method to measure the inherent dissonance of windowed audio signals.

One possible measure that would lend itself to easy implementation is a dissonance metric based on the inherent *harmonicity* of a sound. The more harmonic a sound is, the greater its tendency to be represented in terms of a fundamental with overtones arranged in an integer geometric series. Less harmonic sounds exhibit higher standard deviation of this geometric series.

2.7 The Neurology of Dissonance

A more recent addition to the many definitions of consonance and dissonance is found in neurological studies that directly measure the brain's response to musical stimuli. This forms the basis of a seventh, physiological, definition of consonance—quantifiably and directly measurable according to the brain's chemical and electrical reactions to musical stimuli. Although psychoacoustic correlates of aspects of music like frequency (pitch) and amplitude (loudness) have been well understood for decades, the neurological correlates of emotional aspects of music have not been thoroughly studied, and relatively little is known about them. The fundamental question here lies in finding the neurological components that contribute to the brain's assignment of intervals (and, by inference, sounds in general) as relatively consonant or dissonant.

Neurological aspects of dissonance perception were first hinted by Helmholtz and by his contemporary, the polymath Edmund Gurney (1847–1888). Gurney's 1880 *The Power of Sound*, a sprawling treatise on the philosophy of music, concludes with a remarkable appendix entitled, "On Discord." He speculates extensively on the nature of fatigue, wear, and repair of the auditory processing system and its relationship to dissonance, debating Helmholtz on several points. Gurney also addresses the importance of context in the cognition of dissonance:

> *A discord is always a discord wherever it occurs, and has the same wearing effect on the peripheral organs: but the action of the higher co-ordinating centres so overrides the natural character of the sensation as to convert it into an all-important feature of modern music, the simplest bit of which is often crammed with discord.*(Gurney 1880, p. 557)

A variety of recent hypotheses regarding the neurological basis of dissonance perception in music have been offered and are summarized in Huron (1997). Boomsliter and Creel (1961) suggest dissonance relates to the synchronization among neural firings; Terhardt (1974) proposes the activation of pattern-matching templates and the importance of virtual pitch; and Resnick (1981) suggests that the time delay inherent in pitch perception plays an important role.

Recent studies count the number of nerve fibers activated with the sounding of various intervals (e.g., Cariani, Delgutte, and Kiang 1992; Tramo, Cariani, and Delgutte 1992; Cariani and Delgutte 1996). Findings indicate that sounds historically called “consonant” yield few nerve fiber activations, while “dissonant” intervals yield a far greater number of activations.

In another approach, Blood et al. (1999) used positron emission tomography (PET) scans to continually monitor subjects’ neurological responses to a short harmonized melody. The melody was harmonized in various ways, from a “consonant” harmonization featuring major chords to a more “dissonant” harmonization featuring flat-13 triads. The authors reported a high correlation coefficient between subjects’ ratings of “pleasantness” with the more consonant harmonizations and “unpleasantness” with the more dissonant harmonizations. However, significantly less correlation was reported between subjects’ ratings of “happy” with the more consonant versions and “sad” with the more dissonant versions. Examinations of the resulting PET scans revealed various activation locations in the brain, primarily in the right hemisphere.

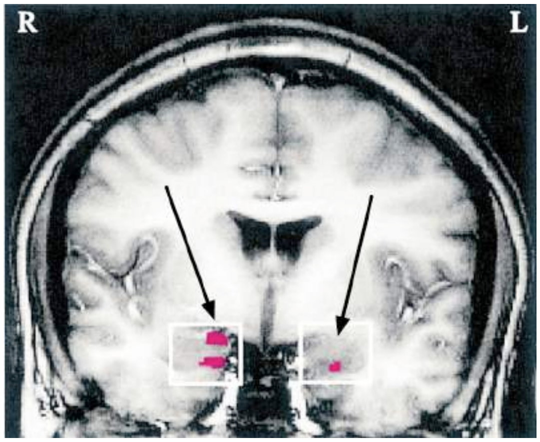
Interestingly, there seems to be some debate regarding the relationship between dissonance perception in music and the negative emotional states of fear and digust. In particular, fear tends to activate the amygdala (Adolphs et al. 1995; Hudgdahl et al. 1995; Morris et al. 1996; Rogan and LeDoux 1996). Amygdala activation has also been demonstrated in functional magnetic resonance imaging (fMRI) in response to listening to laughing and crying sounds (Sander and Scheich 2001; see Figure Figure 2–13). However, the study of Blood et al. (1999) notes that "amygdala activation was not detected," in disgust processing:

> *In summary, the findings in this study identify activity in paralimbic and neocortical regions correlated with degree of musical dissonance, and thus begin to characterize the neural basis for emotional responses to music. These regions have been previously shown to be associated with certain emotional processes. However, these regions differ from those that are active during perceptual aspects of music processing, as well as from those attributed to processing different emotions. The findings of this study not only begin to define a neural network associated specifically with emotional responses to music, but also demonstrate dissociations from other important cognitive processes.* (p. 386)

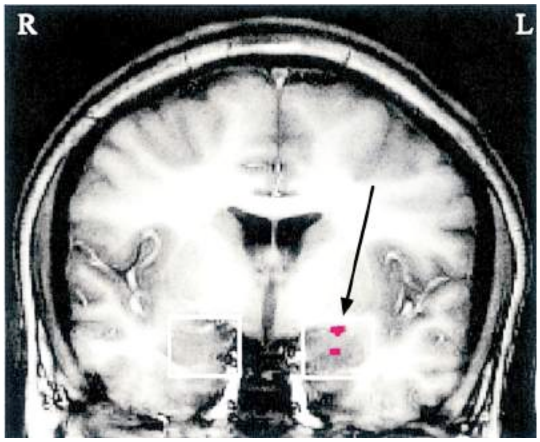
Other recent research addresses other neurological correlates of dissonance perception of chords in both monkeys and humans (Fishman and Steinschneider 2003).

In addition to studying the neurological mechanisms underlying sound processing with respect to dissonance, masking, and disgust in response to acoustic stimuli, a corresponding research problem has recently gained attention.

The question is in some ways the inverse of the above problem, namely, to synthesize audio given measured mental states (instead of measuring mental states in response to auditory input.) The work of Miranda et al. (2003) and Miranda, Roberts, and Stokes (2004), for example, attempts to synthesize musical structures in real time based on measured electroencephalogram (EEG) data. Work in the auditory display community (e.g., Guizatdinova and Guo 2003) addresses the sonification of facial features given an image-processing-based classifier capable of discerning disgust, fear, anger, surprise, happiness, sadness, and neutrality.



(a)



(b)

Figure 2–13. MRIs illustrating amygdala activation while listening to (a) laughing and (b) crying. Here “R” and “L” denote the right and left hemispheres of the brain. Perception of laughing activates both sides of the amygdala; crying only activates the left amygdala. (Sander and Scheich 2001; arrows added.)

The examination of cochleograms, PET scans, MRIs, and other neurological indicators provide an altogether new insight into the world of auditory dissonance. Studies of this sort have thus far concentrated on neurological reactions to relatively isolated sounds and short musical events. Further information will surely be gleaned by examinations of dissonance perception in longer musical passages, particularly as dissonance relates to memory, silence, and fatigue.

2.8 Masking and Auditory Disgust (and More Neurology)

A ninth conception of consonance has been advanced in recent years as involving the concepts of evolutionary development, auditory masking, and the basic emotions of digust and fear. The general idea is that because, as Huron (1997) notes, "Although musical consonance is known to be influenced by social, cultural, and other learned factors, response patterns continue to show transcultural similarities that suggest fundamental physiological concomitants," a fundamental evolutionary response mechanism must be involved in dissonance perception. More specifically, because most humans, regardless of cultural factors, tend (at least anecdotally) to classify certain sounds as "ugly" and others as "pretty," this raises the question of the involvement of innate, pre-programmed responses to auditory stimuli.

Some recent literature along these lines is predated in many ways by Gurney (1880), which was discussed in the last section. The first appendix of *The Power of Sound* is entitled "On Pleasure and Pain" and the speculative ideas it raises are entirely suggestive of more recent studies (although without experimental support at the time of its writing). Gurney distinguishes between *physical* and *psychical* phenomena of human experience, and echoes his contemporary Charles Darwin in arguing that physical responses to stimuli are innately based on the process of natural selection: quite simply, those who appropriately respond to environmental phenomena and/or develop helpful physical characteristics are more likely to survive and pass on these traits genetically. The psychical, or experiential, aspects of this process, which can in general be reduced to pleasure and pain, are useful biologically only to the extent that they invoke a physical response. Gurney comments at length on the relationship between these two forces, noting that when the physical and the psychical are not in agreement, natural selection will run its course to preserve appropriate linkages for future generations:

> *Perhaps the commonest form in which pleasure and pain are vaguely credited with objective powers, is not so much in the connection of natural selection with particular pleasures and pains as in a more general sort of argument, of this kind: that as it may be assumed that creatures in whom the agreeable did not predominate over the disagreeable sensations would perish off the face of the earth through a palpable want of the necessary adaptation to environment, therefore pleasurable sensation in general must be an advantageous thing. (p. 544)*

The biological relevance of hearing then, according to Gurney, is one of survival and adaptation, and hence there exists an evolutionary basis for our quest for pleasure, even in sound. But sound is of course an altogether different sensation than the other senses in that it is present intermittently, and we are able in some circumstances to "tune it out." (This has recently been shown possible with various visual optical illusions as well.) Gurney notes of stimuli that

> *the more striking and impressive they are the less agreeable do they appear, both as causing a physical shock or start and in primitive life probably suggesting danger. Perception and discrimination of sounds would doubtless be an advantage when attained; but the attainment, as in the case of the eye, would be come about through variations and differentiation of structure entailing new and unsought sensibilities.* (p. 546)

This concept forms the gist of modern explications of auditory dissonance in terms of evolutionary development and masking. In short, the idea is that sounds perceived as dissonant are those that most easily mask other sounds, thereby minimizing the effectiveness of our biological quest for information about our surroundings.

Huron (1997) summarizes recent work in this tradition by defining auditory dissonance as "a negative-valence emotion that arises in response to stimulus-engenedered degradation of the auditory system. In short, sounds that are recognized as reducing our capacity to hear other sounds tend to evoke an unpleasant phenomenal experience which in turn leads to stimulus-aversive behaviors." He notes that ethology informs us that behaviors that are likely to elicit strong sensations of pleasure or pain are thought strongly linked to evolutionary survival, noting food, fear, sex, and disgust as examples. A similar theory is advanced by Kamo and Iwasa (2000).

Owing to the primacy of the interval of a critical bandwidth in so many psychoacoustic discussions, particularly regarding its relationship to the phenomenon of masking (in which noise is imperceptible behind one or more pure tones, or vice versa), in conjunction with the proven relationship among the critical bandwidth, sensory dissonance, and musical practice (e.g., Plomp and Levelt 1962, 1965; Greenwood 1991; Huron 1991; Huron and Sellmer 1992; Huron 1994), Huron (1997) hypothesizes that masking must by inference play a crucial role in auditory dissonance.

Clearly, the relationship between masking and dissonance can take two forms. On the one hand, that a masker would be considered dissonant in general is predicted by the theory of natural selection. The tendency for one sound to obscure the perception of another by definition limits our ability to perceive new information about our environment (and hence survive), and so Huron argues that we are quite possibly innately programmed to produce an unavoidable adverse response (i.e., disgust) in response to the detection (whether conscious or not) of a masking sound. The disgust response, which is one of the six basic human emotions (Ekman 1992), is designed to inflict displeasure, thereby minimizing our tendency to place ourselves in environments in which significant masking would occur, thereby enhancing evolutionary survival. The importance of disgust in survival has been applied to other areas, such as moral codes, food, and sexual taboos (Looy 2001, 2004).

An interesting and as yet unexplored area of research is an examination of the processing mechanisms involved during audition of auditory paradoxes like those attributed Roger Shepard (1964) and Jean-Claude Risset (1997). In the case of the continuously descending Risset scale, for example, the traditional sound used to create the paradox (ten octaves of sine tones) itself would be judged quite consonant by most listeners, but the total experience of listening to the scale over time is jarring to most, and it seems plausible that the experience may also activate similar regions of the brain.

Looy (2001) also examines the neurobiology of disgust, noting the particular areas of the brain that are activated during the disgust response:

> *The neurobiological study of disgust has shown that facial expressions of disgust appear to involve activation of the basal ganglia, particularly the right anterior putamen and caudate nucleus, as well as the left anterior insular cortex (Sprengelmeyer et al., 1998; Phillips, Young, Scott et al., 1998; Phillips, Young, Senior, et al., 1997). These areas may also process responses to auditory disgust stimuli such as sounds of retching (Calder et al. 2000). The experience of disgust may involve similar regions (Sprengelmeyer et al. 1996; Calder et al. 2000), as well as the lateral cerebellum and the occipitotemporal cortex (Lane et al. 1997). These appear to be disgust-specific, instead of more generally processing perceptual abilities or basic emotions.*

The specific areas of the brain that are activated during disgust processing is still debated, however. For example, Schienle et al. (2002) offer support that the insular cortex is not used in processing of the disgust emotion. The work of Lane et al. (1997) suggests similarity of processing regions in the brain behind the basic emotions of happiness, sadness, and disgust. In their study, positron emission tomography illustrated activation of the thalamus and medial prefrontal cortex (Brodmann's area 9) of twelve healthy females in response to film and recall stimuli.

The basic emotion of disgust has also been linked to the experience of fear (Woody and Teachman 2000). In this light, Huron likens the phenomenon of auditory dissonance as akin to fear of the dark, which also inhibits the quest for information about one's environment. Surely, a comprehensive theory of musical dissonance will one day be greatly enhanced by the addition of recent insights into neurological and neurobiological contributors to dissonance perception.

2.9 Beyond the Realm of Pitch (and Music)

> *For thee, my gentlehearted Charles, to whom / No sound is dissonant which tells of life.*
>
> —Samuel Taylor Coleridge, This Lime Tree Bower My Prison(1797)

> *For God's sake (I was never more serious) don't make me ridiculous any more by terming me gentleharted in print….substitute drunken dog, ragged head, seld-shaven, odd-eyed, stuttering, or any*

> *other epithet which truly and properly belongs to the gentleman in question.*
>
> —Charles Lamb (1775–1834), Letter to Coleridge (August 1800)

The concept of dissonance is by no means limited to elements of pitch in music, or music at all, for that matter. Outside of music, the term “dissonance” was applied at least as early as 1597, when Bp. Joseph Hall (1574–1656) wrote of “The Tralation of one of Persius his Satyrs into English, the difficultie and dissonance wherof shall make good my assertion….” William Melmoth’s *Letters on Several Subjects, by the Late Sir T. Fitzosborne* (1763, p. 64) refers to the “harshness and dissonance of so unharmonious a sentence….” And Robert Southey’s *Joan of Arc: An Epic Poem* VI.180 notes the **“**dissonance of boisterous mirth.”

We can speak of dissonance as a form of “incongruency” with respect to various aspects of perception. In music, metrical dissonance can occur among rhythmic strains in a particular texture; structural dissonance can also exist among large-scale portions of a musical work. In poetry, dissonance of course has a slightly different meaning. Next, we trace an overview of past and recent scholarship concerning “visual dissonance.” This section concludes with a brief overview of perhaps the most well-known use of the term “dissonance” outside of music: cognitive dissonance.

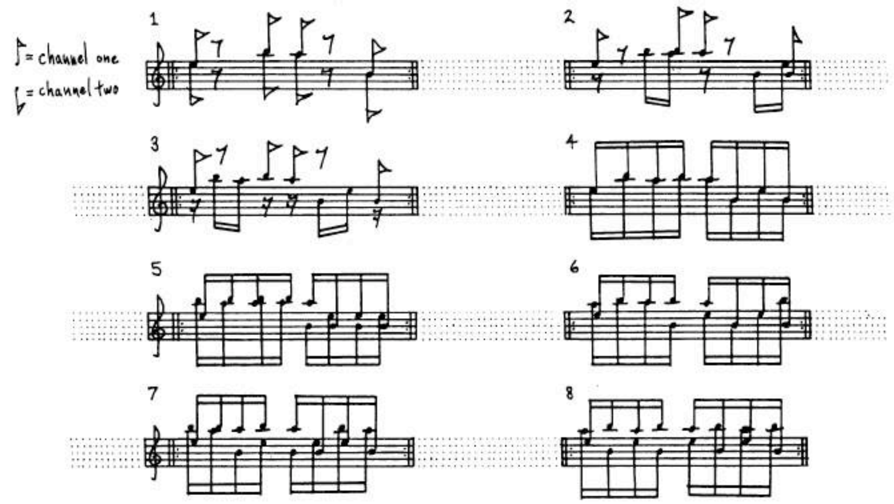
Rhythmic and Metrical Dissonance

Historically, discussions of musical dissonance have addressed the contributions of pitch to that percept exclusively. Several studies in musicological literature, however, were particularly forward-thinking in their adaptation of the concept of dissonance to notions of meter and rhythm.

The term *rhythmic dissonance* was coined by Yeston (1976) to describe the relationship between two or more rhythmic strata. Rhythmic consonance is said to occur when the basic pulse of one rhythmic layer is evenly divisible into that of the other (e.g., half notes against quarter-notes). Rhythmic dissonance occurs when this cannot happen (e.g., seven-against-four). Prime dissonant structures are those in which the divisors of the rhythmic pulse of each stratum are each prime (e.g., seven-against-five).

Of particular note also are studies by Krebs (1987, 1999), which define *metrical dissonance* as the disagreement among metrical layers in a work. The disagreement can be caused by two factors: the time division of pulses in each layer may not form an integer relationship, or there might be a constant and perceptible phase shift between metrical layers. A combination of these factors is present in the famous “phase pieces” of Steve Reich, for example *Piano Phase* (1967), in which a static melodic texture played on two pianos is displaced in time and then realigned periodically, and the lesser-known work from a year earlier entitled *Melodica* (Figure 2–14).

MELODICA



The music exists on magnetic tape. The only source recorded is a loop of the composer playing the original figure (at 1 above) on the Melodica. This loop is first recorded on channel one and is then recorded on channel two in unison with the first channel as shown at 1 above. The dotted lines indicate the gradual shift of phase as channel two begins to slowly move ahead of channel one. Thus at 2 above channel two has moved a sixteenth ahead of the first channel, and at 3, an eighth ahead. Between 3 and 4 there occurs the only splice in the tape as the combination of channels one and two (as they appear at 3) is looped and recorded on both channels. To begin with (at 4 above) both channels are in unison and thus there is no rhythmic change heard between 3 and 4. Then, as before, channel two begins to gradually move ahead and out of phase with channel one. By 5 it is a sixteenth ahead, at 6 an eighth ahead, at 7 a dotted eighth, and finally at 8 a quarter ahead. This last relationship is held steadily for more than 2½ minutes to permit the listener to examine the sound in detail without any phase shift to occupy his attention.

Melodica was conceived and realized in one day, May 22, 1966.

Steve Reich

Figure 2–14. Steve Reich, *Melodica* (1966).

Here, the metrical dissonance would theoretically lie at a minimum at the indicated points of phrase alignment (1–8), while it would lie at a maximum somewhere during the phasing process (where the staggered dots are shown in the score). Surely, it is easy to speculate that a metrical dissonance function during the phasing process must vary according to the temporal ratios of their start times, analogous to the mechanism by which Plomp and Levelt (1965) showed that dissonance of tones was a function of the ratio of their frequencies. For example, precisely halfway through the phasing advancement from point 1 to point 2, the sixteenth notes in each piano are maximally out of phase with respect to each other (i.e., their attacks are equidistant). The two, monophonic sixteenth-note textures can then be clearly fused (barring spatial or timbral cues) into one monophonic thirty-second note texture, thereby minimizing the metrical dissonance.

One could easily extend these simple concepts to envisage many other forms of metrical dissonance, at least in the sense of the word "dissonance" as rhythmic "incongruence" of some kind. For example, consider the result of effecting a triple-meter texture in a musical passage that is clearly notated in duple meter (i.e., hemiola), or vice versa, which of course occurs relatively commonly in various segments of music history, particularly Baroque dance music. Another example of what we might call metrical dissonance occurs more generally when rhythmic complexity is sufficient that the notion of meter becomes more or less irrelevant (i.e., any sensation of downbeat is continually thwarted or simply ignored).

These ideas suggest another form of dissonance in music—what we might term *notational dissonance*, in which the notation and the effective sonic result are seemingly at odds. Notational dissonance can be divided into four distinct but potentially overlapping categories: *rhythmic notational dissonance*, *pitch notational dissonance*, *timbral notational dissonance*, and *expressive notational dissonance*—each corresponding to a different variety of incongruence between notation and intended or achieved sonic result. As an example, a potential case of rhythmic notational dissonance is illustrated in Figure 2–15.

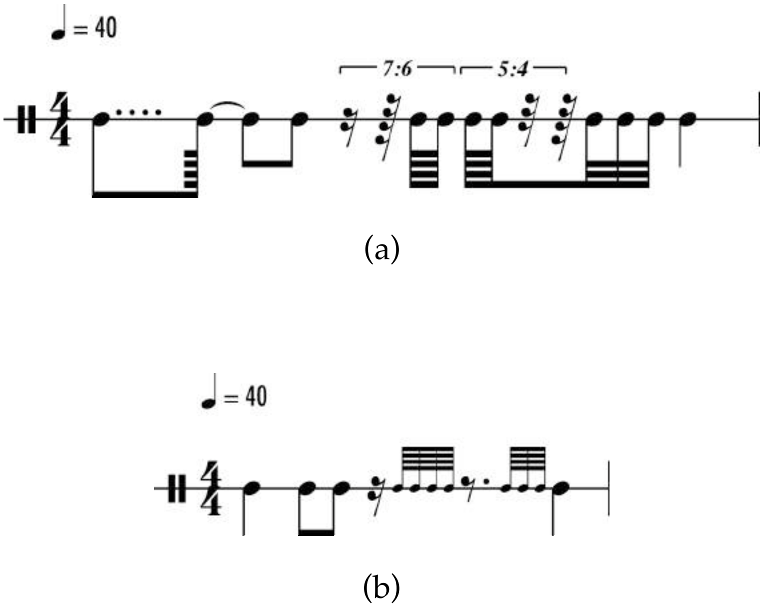


Figure 2–15. Rhythmic notational dissonance. (a) Densely notated rhythmic passage; (b) arguably, a perceptually congruent quantization.

In this case, for a particular range of tempi, any deviations from a resultant of simple, integer-based pulse percepts could be regarded as either (1) performance error (either mental or mechanical) or (2) expressive timing. The degree to which the perceptions of either of these cases are equally likely suggests the extent of the rhythmic notational dissonance.

Several recent examinations of metrical dissonance have addressed specific works and idioms, for example the music of Bartók (Roder 2001), Ravel (Bhogal 2000), Schönberg (Malin 2000), Elliot Carter (Koivisto 2000), electronic dance music (Butler 2002), and Led Zeppelin (Martens 2000). Risset (1997) describes work on the perceptual paradoxes that can result when tinkering with metrical dissonance.

Contextual and Structural Dissonance

Context clearly must also play a major role in a complete theory of musical dissonance. It has been noted that perhaps the first mention and theoretical treatment of context-dependent musical semantics informed by dissonance theory is Rameau's *Traité de l'Harmonie* (1722). Clearly, the language of common-practice tonal music developed under the premise that dissonances and consonances coexist in a mutually beneficial fabric, and that often one is compositionally preferred over the other depending on context.

It is often casually remarked that dissonance provides the "spice" in music—that without it, music can be boring. To the extent that consonance represents auditory/musical "simplicity" and dissonance represents corresponding "complexity," this may well be true, as psychologists tell us that occasional perception of complexity can be a key factor in maintaining attention spans. (This idea seems to contradict ideas of dissonance as comprised only of a neurobiological negative emotion intended to minimize discomfort; clearly, other factors are as—if not more—important in the realm of music.) As Parncutt (1989) notes, this is particularly compelling in tonal music: "the contrast between consonance and dissonance contributes to 'tension' and 'resolution' (Nielsen 1983) and thereby to a sense of 'forward motion' (Forte 1962, p. 15) in tonal music." When this "forward motion" is thwarted in some way, the sense of contextually implied dissonance structures is undermined, as Gestalt theory's principle of good continuation would imply.

Context-related factors can be reduced into three categories: familiarity and memory, cultural conditioning and stylistic cliché, and simultaneity (juxtaposition--Ives). The role familiarity plays in context-dependent dissonance perception has been examined by Valentine (1914), Cazden (1972), and others. In summary, the more familiar a musical passage, the less dissonant it is judged to be. Clearly relevant to the familiarity is the role of musical memory, a highly variable factor among listeners, which is directly proportional to familiarity.

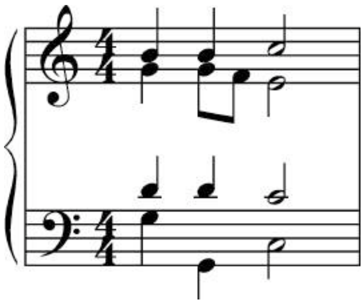
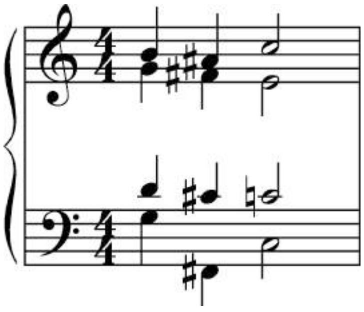
A related concept is the importance musical culture places on context-related dissonance perception. An example is that of compositional clichés (for example, the chord progression shown in Figure 2–16), which become culturally conditioned in a given compositional milieu. This cadence is of course quite familiar in tonal music contexts, and so the literature suggests that "appropriate" musical information presented in this harmonic context would tend to be judged as consonant.

(a)

(b)

Figure 2–16. (a) Tonal cadential cliché; (b) one possible corresponding example of contextual dissonance.

Contextual dissonance owing to cultural conditioning would result for most listeners in Figure 2–16(a) by simply replacing the penultimate chord with an F-sharp-major triad (perhaps retuned, for example, to A = 415 Hz instead of the A = 440 Hz tuning of the preceding and following chord). In isolation, the F-sharp-major triad tuned to A = 415 Hz would be judged as relatively consonant, but, owing to cultural conditioning and exposure to common-practice Western tonality, most listeners would find the new progression somewhat "jarring," to say the least.

Contextual dissonance perception was tested more recently by Bigand, Parncutt, and Lerdahl (1996). By examining four variables ("tonal hierarchies, sensory chordal consonance, horizontal motion, and musical training") while playing a variety of three-chord sequences (C major—[variable chord]—C major), the authors were able to draw conclusions regarding linear, contextual dissonance within the confines of tonal music. They note in conclusion that, as "a main outcome, it appears that judgments of tension arose from a convergence of several cognitive and psychoacoustics influences, whose relative importance varies, depending on musical training."

Another example of contextual dissonance is found in the infinitely ascending/descending scales of Shepard and Risset described previously. On a moment-to-moment basis, the traditional timbres employed would be judged as consonant, whereas the total experience can be difficult to parse contextually. This form of contextual dissonance results from the implausibility of the musical context (i.e., pitches that decrease or increase without limit in context).

A related term has been used casually by Jourdain (1997, p. 104), who refers to *structural dissonance* as occuring "when chords are combined in ways our brains have difficulty modeling" (Jourdain 1997, p.104). Clearly, he refers to the same phenomenon. Structural dissonance can also be taken on a much larger scale. Consider, for example, replacing certain culturally agreed-upon structural elements from, say, the sonata form, with altogether different constructs. The structural integrity of the expected sonata form is thwarted, presumably by intentional design.

Poetry

The concept of dissonance has also been employed explicitly outside the realm of music in poetry, in which dissonance is considered the avoidance of repeated vowel sounds or consonants. A distinction is often drawn between the related terms of assonance and poetic consonance: assonance is the deliberate repetition of vowel sounds (e.g., “geek week”), while consonance is the deliberate repetition of consonants (e.g., “Fee Fi Fo Fum”). Taken together, assonance and consonance are forms of alliteration.

Alliterative structures in English permeate much Old English verse, and much older examples can be found in other languages. As a poetic device, the concept seems most analogous to immediate repetition of neighboring notes or phrases in music.

Visual Dissonance

The term “visual music” is often applied to “time-based visual imagery that establishes a temporal architecture in a way similar to absolute music” (Evans 2005; see Mattis 2005 for a general introduction). Noting the common pedagogical *sine qua non* that an innate sensation of visual “rightness” or “correctness” is fundamental to two-dimensional composition in the visual arts, Evans declares that this “[v]isual ‘rightness’ is visual consonance,” and that this assertion “becomes an axiom from which we can build a grammar of visual music.” He correspondingly considers moments of visual tension or “wrongness” as “visual dissonance.”

Another form of visual dissonance is related to the musical structural dissonance of incongruity and implausibility; we cited infinitely ascending/descending scales as musical examples of this. In the visual realm, this experience takes the form of optical illusion, as exploited for example in the drawings of M. C. Escher (1898–1972). In recent years, a host of time-based visual paradoxes have been discovered as well. … new visual illusions

The concept of visual dissonance itself is not particularly new. Notably, animations and films by Oskar Fischinger (http://www.oskarfischinger.org), John Whitney, and Norman McLaren address the issue of visual dissonance as a function of time, with or without musical accompaniment.

Visual dissonance thus defined is then an individual’s perceptual correlate of an inherent property of an image, and a quality that potentially varies over time in the case of moving images. It thus resides exclusively on what many visual musicians call the *visual plane* of the work, or the collection of time-varying imagery. We have already defined musical and auditory dissonance as a property that correspondingly occurs on the *auditory/musical plane*. The simultaneous pairing of both then creates a new kind of dissonance field, which I call *audiovisual dissonance* (Figure 2–17). Audiovisual dissonance at its most fundamental level then represents the degree of correspondence, “rightness,” or simultaneity between auditory and visual planes.

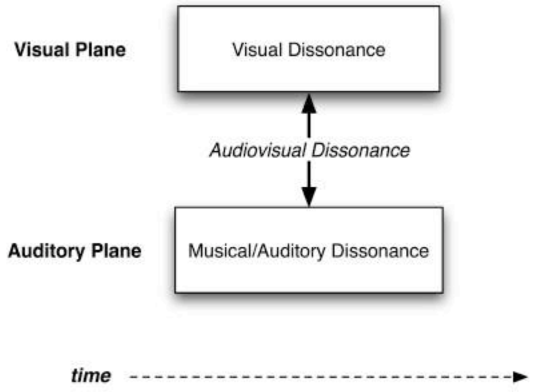


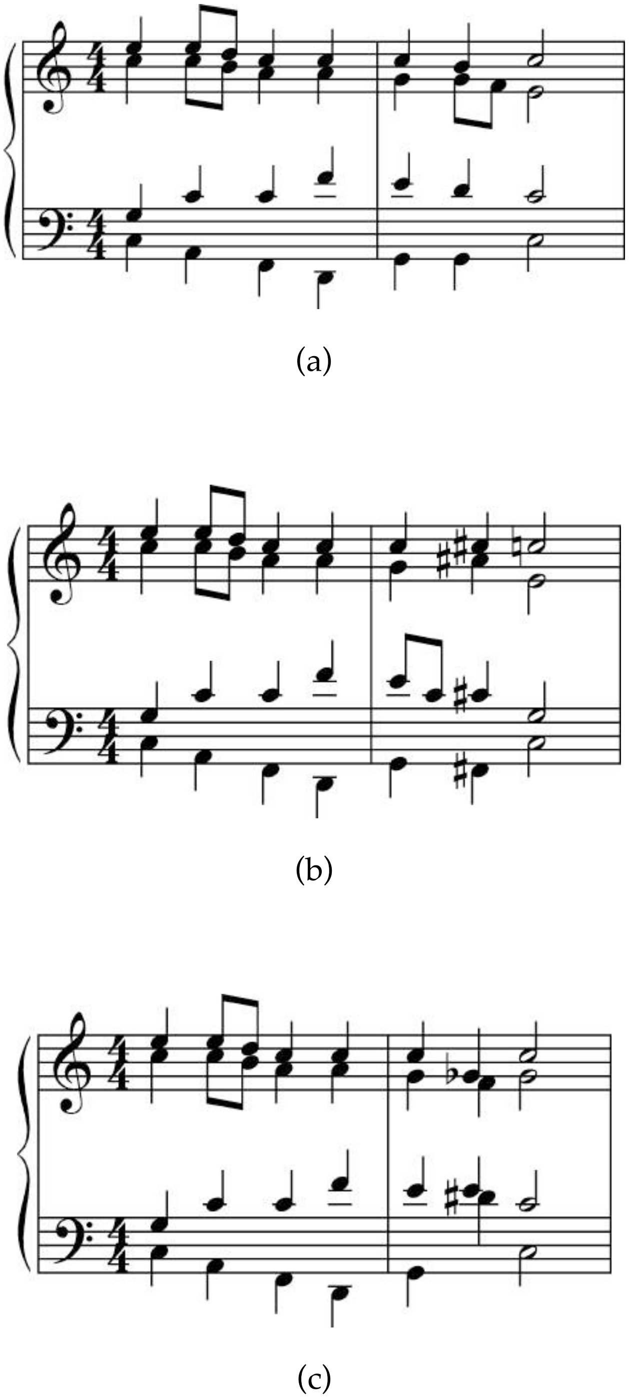
Figure 2–17. Visual, Audiovisual, and Musical/Auditory Dissonance planes.

As two examples of audiovisual dissonance, consider two extreme cases: computer audio visualization “skins,” and the collaborations of John Cage and Merce Cunningham. The former case exhibits a literal, one-to-one mapping from auditory to visual; furthermore, the interaction is one-sided, without feedback, in that the audio informs the video, but not vice-versa. Thus, the experience of audiovisual dissonance tends not to fluctuate significantly and is a function of the chosen mapping scheme. In the Cage-Cunningham works, however, the audiovisual dissonance is intentionally kept as random as possible; the dance and music are designed to occur independently. Therefore, moments of congruence and incongruence occur more or less happenstance, the spectrum between them occupying the fundamental aesthetic space.

Cognitive Dissonance

In his classic 1957 work *A Theory of Cognitive Dissonance*, social psychologist Leon Festinger introduced the term *cognitive dissonance* to denote a discrepancy or contradiction among one's tenets. It was also applied to a denote the mental condition that results in the case of a discrepancy between one's thoughts and actions, for example a mismatch between a moral conviction and a failure to act in accordance with that conviction. This use of "dissonance" as discrepancy and lack of cognitive unity or "oneness"—a kind of "out-of-placedness," as it were—lies surprisingly concomitant with the use of the term in much of the psychological literature up to that time in describing musical dissonance as lack of fusion of the harmonic components in a musical interval (e.g., Stumpf 1898).

For Festinger, dissonance is akin to the percept of hunger, in that it is a state of negative valence emotion that one attempts to remediate. He also classifies cognitive dissonance as a "post-decisional" state, in that, unlike cognitive conflict—a "pre-decisional" state—cognitive dissonance occurs after one has decided to act in a manner conflictive with one's convictions.

How could a comprehensive theory of musical dissonance incorporate this notion of cognitive dissonance? Consider the jarring effect, musically analogous to cognitive dissonance, that would result upon hearing the musical passage shown in Figure 2–18.

(a)

(b)

(c)

Figure 2–18. (a) Standard tonal progression; (b) unexpected substitution of one chord; (c) even less “expected” substitution of the penultimate chord.

Because we are clearly referencing a Western tonal language here, the F-sharp triad is out of place in what we would be led to expect was a straightforward I–vi–IV–ii–I^{6}_{4}–V^{7}–I progression. The cognitive dissonance would be heightened by replacing the F-sharp chord with, for example, a fortissimo tone cluster. The experience of cognitive dissonance can occur both in the listener (experienced as an out-of-placedness, a conflict of beliefs and actions), as well as in the post-decisional state of the composer, immediately upon writing such a passage. Such is the case with the music of Carlo Gesualdo, for example, who consciously broke the rules of music-making during his time.

But cognitive dissonance in music can also occur with sound itself, as in for example the case of electronic and computer music, in which seemingly jarring, out-of-context sounds suddenly appear. The “awkward” and surprising placement of a familiar sound—for example, a sudden burst of human speech—in the middle of an otherwise synthetic texture can indeed evoke a cognitive dissonance in the listener, perhaps as much as that experienced in the post-decisional state of the composer who placed the sound there. Such is the case it could be said with electronic works by composers such as Eric Lyon, Christopher Penrose, and others.

2.10 Summary of Musical Consonance

So what is dissonance? Jourdain (1997, p. 101) offers a succinct and quite accurate summary of our current understanding of and approach to dissonance. He summarizes in an accessible way the commentaries of other recent writings on the subject by noting that dissonance can be subsumed by factors arising from (1) acoustics, (2) neurology, and (3) structure. In this spirit, we here offer the Venn Diagram in Figure 2–19 as illustrative of the myriad of approaches to the study of dissonance.

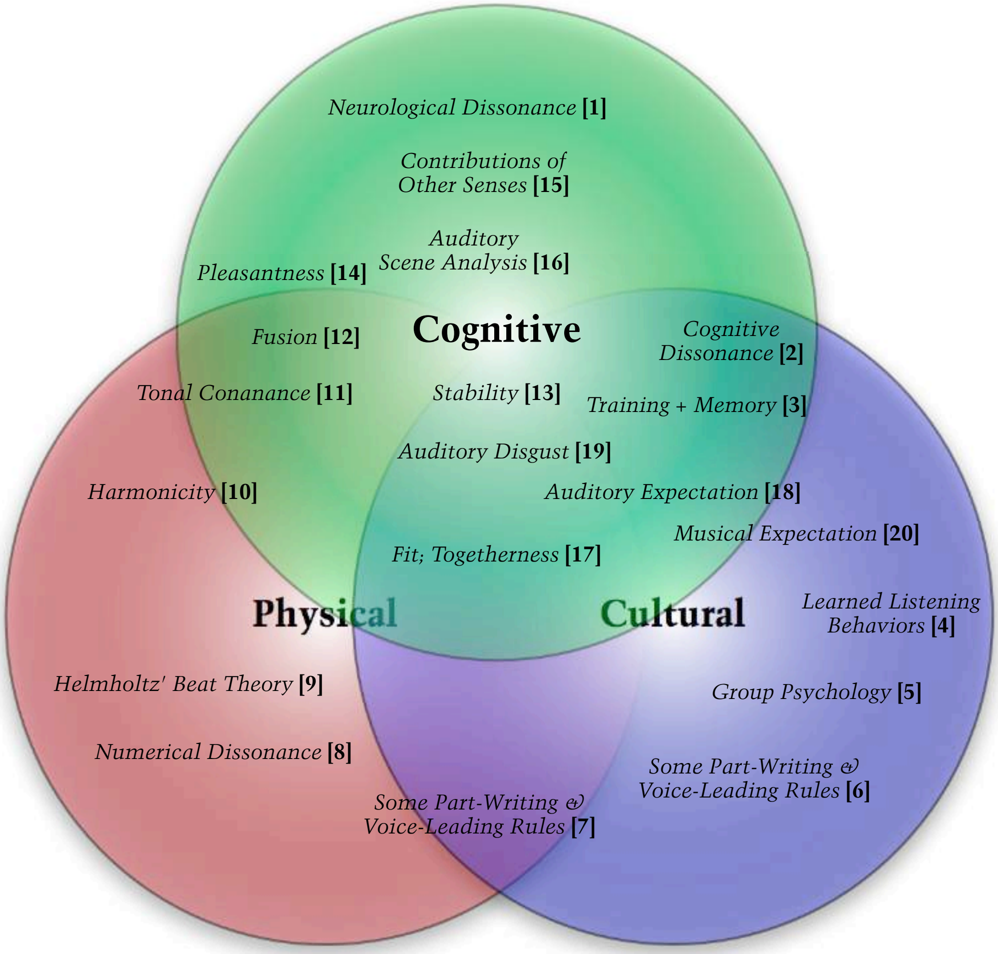


Figure 2–19. Summary of acoustical, psychoacoustic, and musical/contextual contributors suggested by primary theories of musical dissonance.

The confluence of the acoustical, psychoacoustic, and musical/contextual factors that contribute to modern theories of musical dissonance are summarized in Table 2–2. We can summarize at least twenty factors involved in dissonance perception, each of which could be classified under one or more of the rubrics of culture, cognition, and physics.

Reference Index	*Factor*	*Example*
1	Neurological Dissonance	Contributions of neural activity to pitch perception, tempo detection, etc.; place and template theories of hearing
2	Cognitive Dissonance	Festinger; can involve any combination of senses
3	Training + Memory	"I was taught that the minor second is called 'dissonant.'" "That sound reminds me of another that I previously categorized as 'dissonant.'"
4	Learned Listening Behaviors	"Yuck; I don't like the sound of tritones. They're the devil in music!"
5	Group Psychology	"My friends and I hate minor seconds, but we love tritones." "Most music theorists agree: the major seventh is more consonant than a minor second."
6	Some voice-leading and part-writing rules	The fourth should resolve to a third.
7	Some voice-	Parallel fifths are forbidden.

	leading and part-writing rules	
8	Numerical Dissonance	Numerical "complexity" of an interval; Euler's Γ function.
9	Helmholtz' Beat Theory	The minor second is more dissonant than the octave because more beat frequencies are created by the minor second.
10	Harmonicity	(1) The fifth is more consonant than the minor second because it is more "harmonic" or because it occurs "lower" in the harmonic series (e.g., Leibnitz, Rameau); (2) Harmonic sounds are more "consonant" than inharmonic sounds
11	Tonal Consonance	Sethares' Dissonance Theory
12	Tonal Fusion	The perfect fifth is more consonant than a tritone because the fifth fuses better owing to both physical and cognitive factors.
13	Stability	A major triad is more consonant ending for a composition than banging all notes of the piano simultaneously.
14	Pleasantness	"Mmmm…those pitches/sounds create a pleasing sensation."
15	Contributions of	Complementarity/Non-complementarity

	Other Senses	(i.e., agreement/disagreement) of all senses
16	Auditory Scene Analysis	Good continuation, grouping, etc.
17	Fit; togetherness	These pitches/sounds “work” well together; it sounds like they belong together.
18	Auditory Expectation	Auditory cognitive dissonance
19	Auditory Disgust	Negative-valence auditory emotional response.
20	Musical Expectation	Acquired/taught musical cognitive dissonance

Table 2–2. Summary of cognitive, physical, and cultural factors that contribute to dissonance. Reference indices correspond to those in Figure 2–19.

Once dissonance can be dismantled into its constituent parts, it becomes a candidate for quantification, perhaps even deserving of its own scale and units, like the Mel scale for pitch perception or the Phon/Sone scales for perceptual loudness. However, as previous attempts at defining numerical metrics for timbral descriptions of sound have proven problematic, so too is the multidimensional world of dissonance. Perhaps the greatest problem inherent in the study of musical dissonance is its refusal to be strictly quantified and measured, despite our best efforts.

3 INTERLUDE: THE EMPIRICISM OF DISSONANCE

> *When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science.*
>
> *—William Thomson, Lord Kelvin, Popular Lectures and Addresses (1891–1894)*

> *Measure what is measurable, and make measurable what is not so.*
>
> *—Gottlob Frege (1848–1925), Quoted in Weyl (1959)*

As has been demonstrated, Western writings on musical dissonance have historically approached dissonance from an empirical and quantitative perspective. The typical twentieth-century research paper on musical dissonance typically follows the following form: first, ask subjects to rate the relative dissonance of intervals; second, statistically analyze the results; finally, inductively generalize a graph, based on weighted averages of the subjects' responses, that illustrates the relative dissonance of all equal-tempered (or perhaps just-tuned) dyads within a given octave.

Before we concentrate on a few interesting sidebars to this quantitative study of dissonance—what we might call the attempted "empiricism of dissonance"—a curious alternative approach offered by Edmonds and Smith in 1923 is worth noting. These psychologists conducted an early study rooted in the supposition that intervals can be given an absolute dissonance rating, irrespective of context and relation to other intervals. They write that "[i]n this investigation…we have been concerned with the possibility of an absolute characterization, a phenomenological description of the bitonal process in and of itself" (p. 287). Starting from this decidedly quantitative approach, however, they conclude instead that "it is fairly easy to fit to bitonal fusions characteristic 'quantitative' names, derived from taste and touch, which make recognition of the fusions possible" (p. 291). Surprisingly, they offer in conclusion that the best way to characterize the dissonance of intervals might be to take a qualitative approach, rather than a quantitative one. Pairs of tones were played for the subjects, and they were asked to select from a list of adjectives that best described their experience of that interval or to freely come up with their own. The authors found that subjects used the following terminology to describe the "experience" of the octave: "smooth, like the surface of window glass"; "smooth, like polished steel"; and "smooth, a unitary experience like one note." Conversely, subjects described the major seventh as "astringent, like the taste of a persimmon [sic]";

"gritty, like the feel of small sharp granular objects"; "astringent, like strong vinegar or alum"; and "harsh, a nippy, biting effect like a strong astringent." The major second was "gritty, like the feel of small pebbles in the hand or grapenuts [sic] in the mouth"; "rough"; "gritty, like sand in one's teeth"; and "gritty, like sandpaper." The subjects' consistently similar choices of adjectives to describe intervals forms an interesting qualitative alternative to empirical study in the understanding of dissonance.

As noted, the vast majority of studies of musical dissonance attempt to discover a new empirical insight into the "proper" ranking of intervals in terms of their relative dissonances. We now concentrate on a brief study of the philosophy of empiricism from the writings of Gregory Bateson before motivating the potential usefulness of the empirical study of dissonance, particularly within the confines of musical analysis. Our journey continues with comments on the comparative ranking of intervals, followed by a summary of attempts to graphically display dissonance data (along with several proposed new displays). The chapter concludes with a discussion of dissonance as a musical control structure.

3.1 Introduction: Number versus Quantity

Before attempting to quantify an elusive concept like that of musical dissonance, we must consider various ways in which it is possible to assign numerical descriptors to objects. Clearly, one of the foremost lessons from twentieth-century science is that observation and measurement affect that which is observed and measured. Equally as important, however, is the distinction between number and quantity.

In a classic 1978 article, social scientist Gregory Bateson called for a fundamental recognition of the difference between numbers (“the product of counting”), and quantities (“the product of measurement”). Bateson distinguishes counting as a discrete, digital process, whereas measurement exhibits properties of analog systems. He furthermore notes a process that lies somewhere in between—a kind of gestalt pattern-recognition activity, in which we are capable of counting without counting, but rather simply glancing. Bateson concludes his brief argument by referencing Pythagoras and Augustine, observing that “we occidental humans get numbers by counting or pattern recognition, while we get quantities by measurement”—declaring this concept to be “some sort of universal truth.”

This argument forms an apt container for the discussion of analytical techniques whereby we listen to and make judgments regarding the dissonance of a sound. The two basic modes by which humans listen to sounds have been classified as analytic in nature (“analytical listening”) and holistic in nature, (“synthetic listening”). For related discussions, see Deutsch (1982); McAdams (1982); Parncutt (1989); and Doherty and Lutfi (1992). Because our psychoacoustic apparatus is clearly adept at both counting, measurement, and pattern recognition, it naturally follows that these entities must be tied together in some way, whether hard-wired or learned. A proposed arrangement is offered in Figure 3–1, in which an analytical listening state leads one to count (for example, the number of beats per second in an interval, as piano tuners do when tuning strings); on the other hand, a synthetic listening mode prepares one for a less-exacting, analogic form of measuring various aspects of dissonance perception (for example, measuring one's internal emotional disgust level upon hearing a dentist's drill). The gestalt perception of dissonance *in toto* then follows as a result of the fusion of the two modes of listening, whereby intervals and sounds are instantly classified according to their dissonance "level" using a combination of these techniques, akin to the manner in which one can "recognize" five apples by simply perceiving their proximity and recognizing "fiveness." We use the phrase "instantly classified" here to denote the auditory analog of the visual "glimpse"—the notion of the brief and the impermanent, coinciding with the sure and the certain.

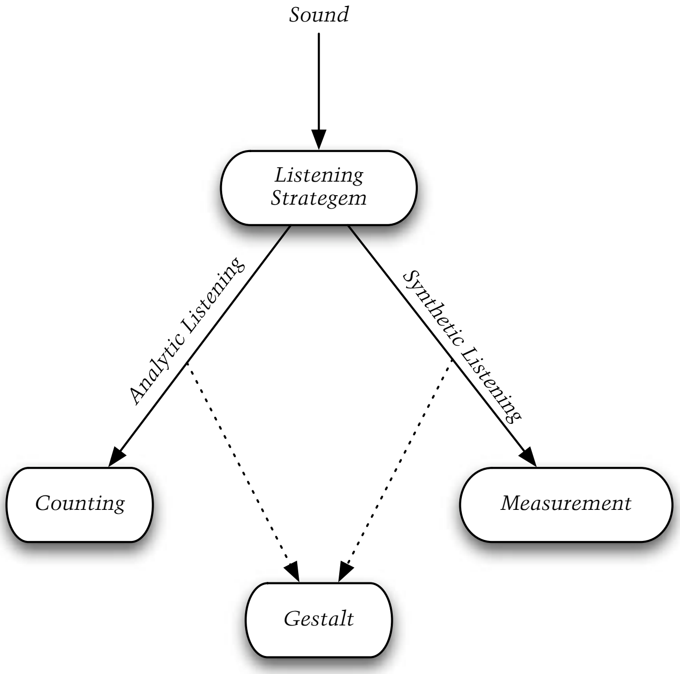


Figure 3–1. Correspondences among listening modes and dissonance analysis techniques.

The incongruence between number and quantity need not impede our attempts to quantify aspects of musical dissonance, but on the contrary, should inform the search. In particular, we should recognize those aspects of musical dissonance that are countable, as well as those that are merely measurable. Table 3-1 classifies various aspects of dissonance in this way.

Factor	*Category*

Neurological Dissonance	Measurable; some aspects, such as neural firing rates, can be countable, however
Cognitive Dissonance	Measurable
Training + Memory	Difficult to measure or count
Learned Listening Behaviors	Difficult to measure or count
Group Psychology	Difficult to measure or count
Some voice-leading and part-writing rules	Countable (in the sense that the rules are discrete in nature)
Numerical Dissonance	Countable
Helmholtz' Beat Theory	Countable
Harmonicity	Measurable
Tonal Consonance	Countable
Tonal Fusion	Measureable
Stability	Measurable
Pleasantness	Measurable

Contributions of Other Senses	Difficult to measure or count
Auditory Scene Analysis	Somewhat measurable and countable (viz. Computational auditory scene analysis)
Fit; togetherness	Measurable
Auditory Expectation	Measurable
Auditory Disgust	Measurable; gestalt
Musical Expectation	Measurable; gestalt

Table 3-1. Measurable and countable classifications of dissonance factors.

Classification of the factors affecting dissonance in general does not denote accuracy or meaningfulness; for example, one can accurately measure the auditory disgust response of a group of subjects using any of various means (facial expression, manual ranking by subjects, etc.). What is important here is awareness of the countability and/or measurability of that being observed.

3.2 Dissonance-Based Musical Analysis

Once dissonance levels have been quantified in some way, it naturally follows that the data they provide may yield insights into musical analysis. If we can assign a single number (through measurement) that relates in some way to the perceptual "dissonance" as a function of time that one experiences while listening to a piece of music, or on the other hand simply count a particular countable dissonance factor, we may be able to abstract useful information about aspects of the moment-to-moment experience of listening to the work. We may also then gain insights into the occurrences of tonal tension and release, for example. Several recent studies (e.g., Danner 1985) are rooted in this concept, which attempt to compute (i.e., count) the "dissonance level" at each quantized unit of time in the piece by adding the dissonance that results from each note-against-note combination. For example, in a four-voice canon, the total dissonance at a particular unit of time $\delta_{total}(t)$ might be calculated by adding the dissonance between each possible combination of voices. If we label the four voices *A*, *B*, *C*, and *D*, then we could compute the total dissonance as

$$\delta_{total}(t) = \delta_{AB}(t) + \delta_{BC}(t) + \delta_{CD}(t) + \delta_{AC}(t) + \delta_{BD}(t) + \delta_{AD}(t)$$

where, for example, the dissonance function $\delta_{AB}(t)$ computes the dissonance between voices *A* and *B* in isolation. However, dissonance curves in general analyze only one (or perhaps a small subset) of several factors that influence our perception of dissonance, generally irrespective of musical context. Calculating the beat-to-beat, note-against-note dissonance that results at each instant in a piece of music, irrespective of the timbres and other psychoacoustic factors involved, is essentially meaningless, although one can still find examples of such a practice occasionally even in modern analyses. Furthermore, it is clear that component intervals within chords cannot simply be isolated, quantified in terms of dissonance, and then added back together; this is a multidimensional problem, for the "dissonance" of each constituent interval affects the "dissonance" of each of the other component intervals, in addition to that of the entire chord. For example, in the equation above, a minor change in $\delta_{AB}(t)$—perhaps owing to a slight amplitude fluctuation in voice A, or a different articulation of voice B—could potentially impact the measured dissonance level of the total chord, and as such it should be able to impact the "countable" dissonance level as well. At the very least, we should attempt to compute the total dissonance as

```
$\delta_{total}(t) \leftarrow 0$                /* initialize total dissonance to 0 */
for i = 1 to numberOfVoices
      for j = 1 to numberOfVoices
            if i != j and we have not computed $\delta_{ij}(t)$ or $\delta_{ji}(t)$, then
                  compute $\alpha_{ij}(t)$
                  compute $\delta_{ij}(t)$
                  $\delta_{total}(t) \leftarrow \delta_{total}(t) + \alpha_{ij}(t)\delta_{ij}(t)$
            end
      end
end
```

where the $\alpha_{ij}(t)$ represent a psychoacoustic weighting function corresponding to the coexistence of voices *i* and *j* at time *t*.

Note that this is entirely different than Sethares' approach, which measures the "total dissonance" at a any point in time by computing the dissonance between each possible pair of spectral lines present in a recorded sound. This is one of the central arguments of this thesis: that strings of sounds are best taken *in toto* for musical dissonance analysis, irrespective of notation or their parametric descriptions.

Even so, in many ways, this brute-force approach to dissonance analysis is clearly only meaningful at a cursory level, as it neglects both timbral contributions, amplitude scaling of each note, psychoacoustic modeling, and musical context. Robert Jourdain, is his general-reader introduction to music psychology *Music, The Brain, and Ecstasy*, writes the following:

> *Dissonance is hard to tack down even in classical Western harmony. The overall dissonance of a piece cannot be measured simply by tallying the relative number of dissonant intervals formed by chords and melodic lines. When a dissonance falls at a point of harmonic arrival—a point often emphasized by rhythmic accentuation—the dissonance will clang in your ears. Yet the same dissonance will hardly register when it occurs at a less conspicuous position.* (Jourdain 1997, p. 104)

That being said, incorporation of quantified dissonance into the art of musical analysis can provide new insights, almost irrespective of the specific degree of accuracy of the measurement technique. For example, the analysis of Danner (1985), despite the criticisms outlined above, does provide some insight into the Carter *Canon*, although at a relatively cursory level that could be obtained through conventional direct harmonic analysis. The work of Huron (1991), however, has been particularly interesting in its computational dissonance analysis of the music of Bach. As Huron (2005) states, "Bach's polyphonic music is organized so as to minimize 'sounding as one' while maximizing 'sounding smooth.' Bach's musical organization is inconsistent with the theory that consonance is caused by tonal fusion."

In addition to distinguishing between number and quantity, tantamount to the success of dissonance-based musical analysis is an understanding of the ways in which intervals have been compared in Western musical theory and the various ways that dissonance curves have been constructed. Much of this exercise was undertaken in the preceding chapter; in the next section, we generalize about attempts to compare the dissonance of intervals in terms of number or quantity.

3.3 Comparative Rankings of Intervallic Dissonance

We have already examined several historical attempts (most of them quite recent) to quantify and compute dissonance curves; these are all indebted to Euler's *gradus suavitatis*function Γ, which represents the earliest major attempt to quantify dissonance. Plotting Γ for a particular scale is a computationally trivial method of achieving dissonance graphs or “tone profiles” for that scale. An example profile of the 11-limit scale {1:1, 16:15, 9:8, 6:5, 5:4, 4:3, 11:8, 3:2, 8:5, 5:3, 16:9, 15:8} is shown below in Figure 3–2.

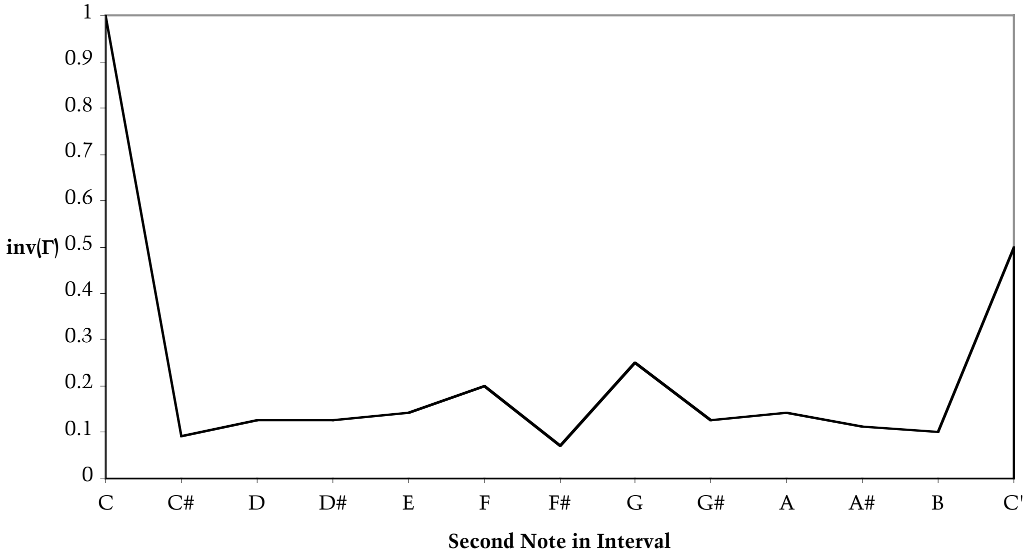


Figure 3–2. The inverse of Euler’s Γ-function for all dyads in a particular just scale. In each interval, the first pitch is C, and the second pitch is shown on the *x* axis.

Here, we plot the value of the Γ-function for the dyad formed when each member of the scale is played simultaneously with the root of the scale, assuming 1:1 = C (i.e., the scale is built on C). The graph more or less agrees with our perception of the dissonance of dyads, in the generic sense, removed from their sonic context.

A major flaw is easily uncovered with Euler's function, though. Consider two different augmented fourths: 11:8 and 45:32. Although listening tests would show a discrepancy in subjects' classification of the relative dissonance of these two intervals, Euler's method assigns each the same *gradus suavitatis*, since

$$\begin{aligned}\Gamma\left(\frac{11}{8}\right) &= \Gamma(11\cdot 8)\\ &= \Gamma\left(11^1\cdot 2^3\right)\\ &= 1+1(11-1)+3(2-1)\\ &= 14\end{aligned}$$

and

$$\begin{aligned}\Gamma\left(\frac{45}{32}\right) &= \Gamma(45\cdot 32)\\ &= \Gamma\left(5^1\cdot 3^2\cdot 2^5\right)\\ &= 1+1(5-1)+2(3-1)+5(2-1)\\ &= 14\end{aligned}$$

Another problem with this measure is that it neglects psychoacoustic models, which were of course not available to Euler. For example, listeners generally find stretched octaves more consonant that perfect 2:1 octaves, particularly on a piano whose string stiffness contributes to a slightly inharmonic tone. This finding stands in contrast to the prediction of Euler's model.

A third problem becomes apparent when one continually "zooms" in on the neighborhood of intervals immediately surrounding any given just interval. Number theory tells us that between any two just intervals exists another, *ad infinitem*; thus, between two relatively "consonant" intervals like the minor third and major third, an unbounded number of other just intervals exists. The problem is that we could easily construct a numerically "complicated" interval (in terms of the prime decompositions of its numerator and denominator) that lies squarely between the third and fourth, and thus Euler's model would predict a giant spike in the dissonance value. However, such a spike would contradict experimental evidence that all dyads in the neighborhood between a minor and major third are comparably consonant. In general, the greater level of detail with which one attempts to construct a tone profile using Euler's *gradus suavitatis*(i.e., the larger the set of dyads one examines), the less the resulting profile tends to agree with psychoacoustic evidence.

As we have already seen, following in this tradition, many other mathematical reductions of aspects of dissonance were proposed. Some of these models incorporated by modeling roughness and beating (Helmholtz 1877; Terhardt 1968, 1974; Plomp and Steeneken 1968), fusion (Stumpf 1898), critical bandwidths (Plomp and Levelt 1965; Sethares 1997), virtual pitch (Terhardt 1972), and more recently Equivalent Rectangular Bandwidths (Patterson 1976; Houtgast 1977; Weber 1977; Patterson et al. 1983; Shailer and Moore 1983; Fidell et al. 1983).

Another approach to quantifying dissonance lies in the method of mathematically modeling listener preference tests or sets of previously published data (Plompt and Levelt 1965; Kameoka and Kuriyagawa 1969; Sethares 1997; Haluska 2004; Manaris et al. 2005). This approach is also valuable, for it attempts to bridge the gap between well-understood acoustical models and more recent trends in cognition.

But just as flaws can be uncovered in Euler's *gradus*, so too can we point to counterexamples that may quickly dismiss these more recent dissonance curves, or indeed the very notation that such curves bear musical meaning or significance. But clearly, any psychoacoustic correlate is inherently flawed as a quantitative measure, but nonetheless useful as a starting point in developing models of machine listening. And that, perhaps, is the lesson to be learned—that mathematically empericizing dissonance from both physical and psychoacoustic models can ultimately meld to form a single, unified theory not only of dissonance, but pitch perception in particular and tone semantics in general.

3.4 Visualizing Quantified Dissonance

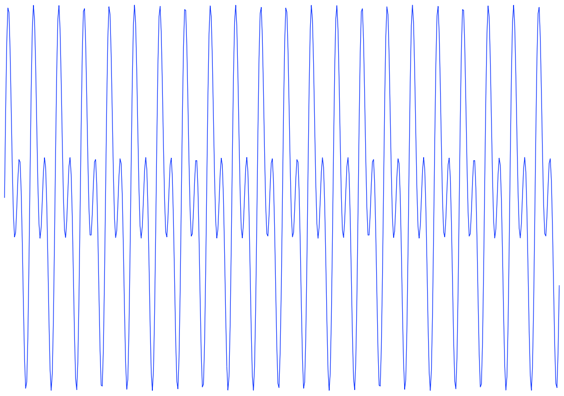
Creating graphical displays that illustrate in some way a psychoacoustic percept is quite challenging. To my knowledge, no such system for comprehensively displaying the percept of sensory dissonance has been developed or even attempted, largely owing to the lack of a clear and comprehensive model of musical dissonance. Hence, most methods of visualizing dissonance, as it were, attempt to show one or more physical contributors to the percept of dissonance (for example, roughness or beating).

The graphical representation of physical phenomena that contribute to sensory dissonance is quite easy, and several well-known techniques can be used for display of such data. We begin by addressing the obvious time-domain plots of amplitude for acoustic phenomena, which can clearly illustrate beating and other dissonance-related contributors. We then examine Lissajous Curves, Chladni Patterns, and lattice diagrams. This section concludes with my speculations on the applications of spherical harmonics to the display of Eulerian dissonance.

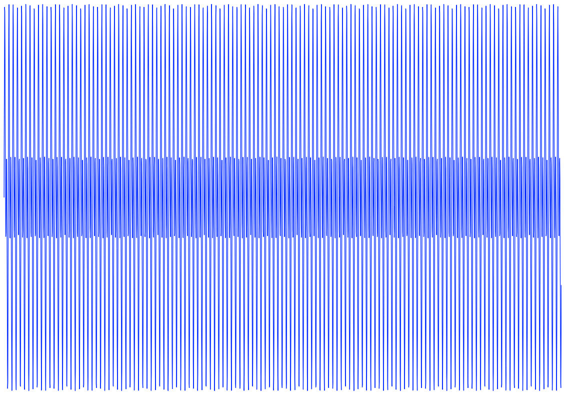
Amplitude Beating

As we have seen, beating and fusion lie at the core of previous theories of musical dissonance, arising from the tradition of Helmholtz. At the risk of stating the obvious, plotting the simultaneity of similar frequencies and their corresponding amplitude modulation patterns that result quite simply illustrates the beating phenomon. Consider Figure 3–3, which shows two perspectives on a

440 Hz tone together with an 880 Hz tone. The Eulerean metric predicts a highly consonant result, which the lack of beats illustrates. (Of course, perfect 2:1 octaves are often cited in the literature as less pleasing to human listeners. Again, simple graphs like these only display physical phenomena, not perceptual correlates.)



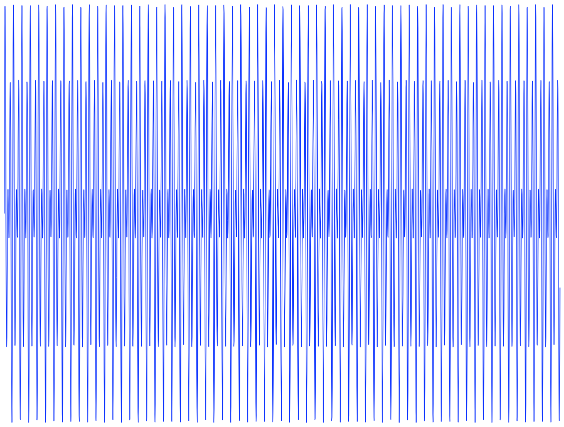
(a)



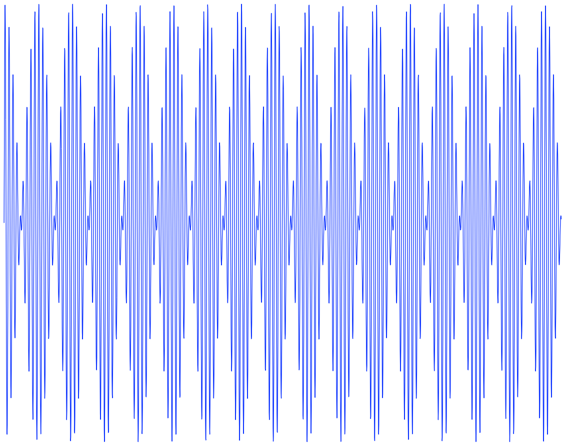
(b)

Figure 3–3. (a) Plot of $y = \sin(2\pi 440t) + \sin(2\pi 880t)$; (b) zoomed out. No amplitude beating results from the combination of 440 Hz and 880 Hz.

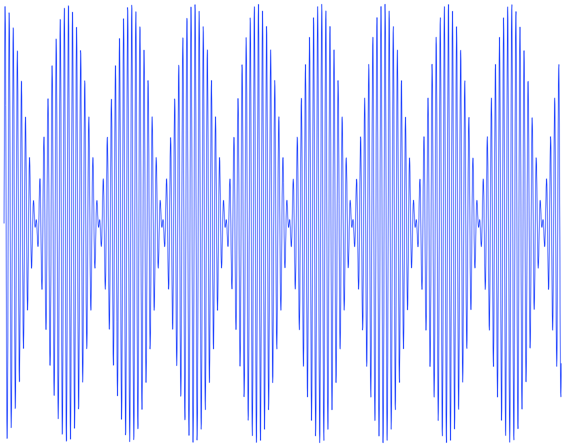
Beating patterns eloquently correspond to Euler’s *gradus suavitatis* dissonance measure. Consider, for example, the increased frequency of beating patterns corresponding to increasing Eulerean dissonance measures in Figure 3–4, which shows (a) a perfect fifth, (b) a 9:8 major second, a 16:15 minor second, and a syntonic comma.



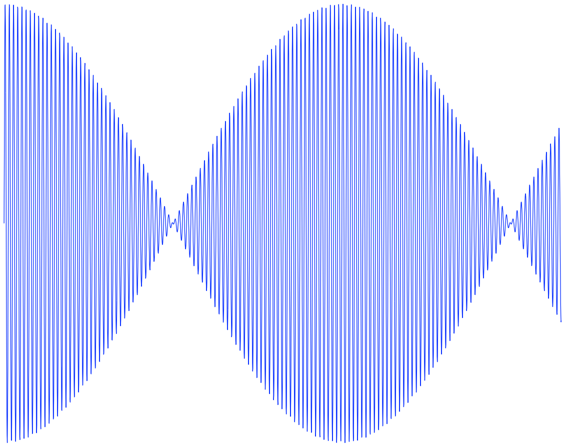
(a)



(b)



(c)



(d)

Figure 3–4. Graphical display of increasing Eulerean dissonance: (a) 440 Hz plus a 3:2 perfect fifth; (b) 440 Hz plus a 9:8 major second; (c) 440

Hz plus a 16:15 minor second; (d) 440 Hz plus a syntonic comma (81:80).

The corresponding beatings in each of the above scenarios are of course visible in time-frequency graphs (e.g., sonograms and waterfall plots) as well, provided the appropriate parameters are chosen.

Unfortunately, simple plots like these tell virtually nothing about acoustical dissonance present in complex sounds composed of many partials. They tell even less about the acoustical dissonance inherent in complex sound objects.

Lissajous Curves and The Harmonograph

Lissajous curves (also known as Lissajous figures or Bowditch curves) were studied first by Nathaniel Bowditch in the early nineteenth century, and later exploration was carried out independently about forty years later by Jules Antoine Lissajous. A Lissajous curve describes the system of parametric equations given by

$$\begin{Bmatrix} x(t) = A\cos(\omega_x t - \delta_x) \\ \\ y(t) = B\cos(\omega_y t - \delta_y) \end{Bmatrix}$$

which is sometimes written in a slightly different form:

$$\begin{Bmatrix} x(t) = A\sin(\frac{\omega_x}{\omega_y} t - \delta_r) \\ \\ y(t) = B\cos(t) \\ \end{Bmatrix}$$

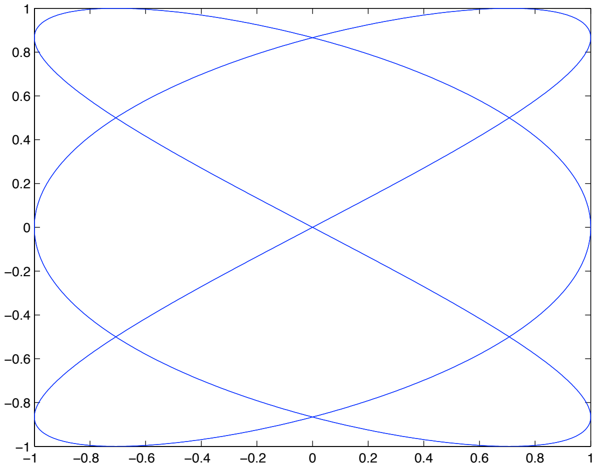
where δ_r is the relative phase difference between the two signals in rad/sec. In expressing the relationship between two audio frequencies f_1 and f_2 (in units of Hz) with a phase difference of δ_d δ_r equations is then simply

$$\begin{Bmatrix} x(t) = A\sin(\frac{f_x}{f_y} t - \delta_d) \\ \\ y(t) = B\cos(t) \\ \end{Bmatrix}$$

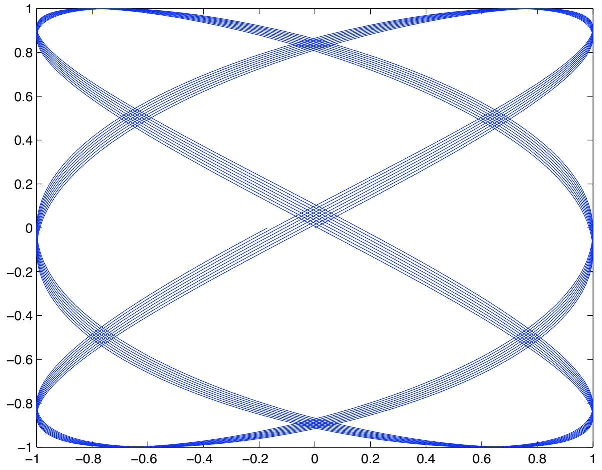
since $\omega = 2\pi f$.

The resulting curve succinctly illustrates several dimensions of data, namely the constants A, B, δ_r, f_x, f_y, δ_d, and the independent variable of time, t. Neglecting the phase difference between the two signals (δ_d) and fixing $A = B = 1$, Lissajous figures uniquely represent the ratio f_x: f_y, which is useful for visualizing musical intervals. Interestingly, these curves are closed only in the case where the ratio f_x: f_y is a rational number; that is, they will form a closed loop for just intervals but not for tempered ratios. Lissajous figures highlight periodicity and coincidence; thus, because irrational numbers can never be periodic with respect to one another (i.e., one cannot be evenly divided into another), the resulting curves are a bit messy.

Consider Figure 3–5, which compares the Lissajous figures for a 3:2 perfect fifth (Figure 3–5a) and a twelve-tone equal-tempered fifth (Figure 3–5b). Note that the frequency ratio for an equal-tempered fifth is given by $2^{7/12}:1$. Also, recall that, because the ratio representing an equal-tempered fifth is irrational, the curve will never close. For convenience, Figure 3–5b shows the first 16 cycles of the curve (i.e., plotted from 0 to 32π radians).



(a)



(b)

Figure 3–5. (a) Lissajous curve corresponding to a 3:2 perfect fifth; (b) Lissajous curve corresponding to the first 16 cycles of an equal-tempered fifth.

The above figures are easily generated with modern computers. But the nineteenth-century trade of ornamental turning, the art of etching geometric patterns onto wood and metal, provided the technology to realize such figures long before computers. The Geometric Chuck was used by metalsmiths to produce geometric designs that varied uniquely according to the frequency ratios of two gears. Examples are shown in Figure 3–6. (Note that the Lissajous figures are plotted in polar coordinates rather than in rectangular coordinates.) The Harmonograph, shown in Figure 3–7, was such a similar instrument made for drawing these curves on paper. Two kinds were available: the Lateral

Harmonograph and the Rotary Harmonograph, which produced Lissajous curves plotted along Cartesian (rectangular) and polar coordinate systems, respectively.

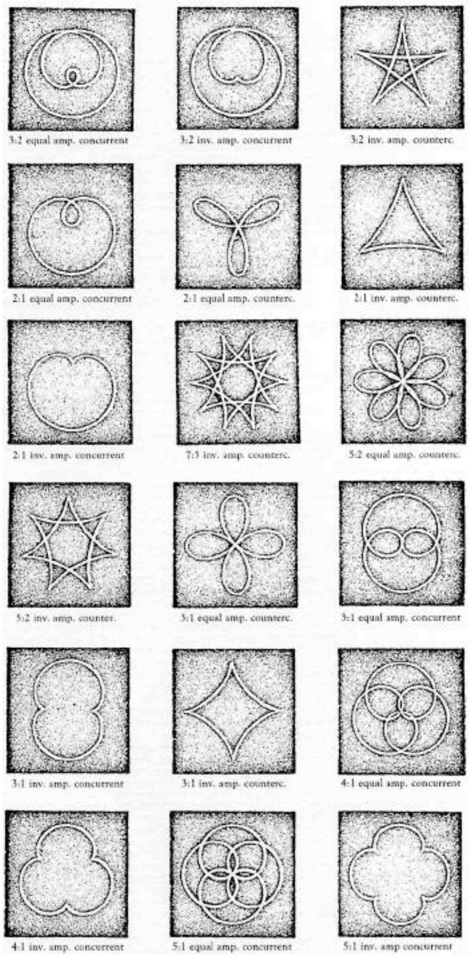


Figure 3–6. Output of various gear ratios from Sir Thomas Bazley's *Index to the Geometric Chuck: A Treatise upon the Description, in the Lathe, of Simple and Compound Epitrochoidal or "Geometric" Curves* (1875).

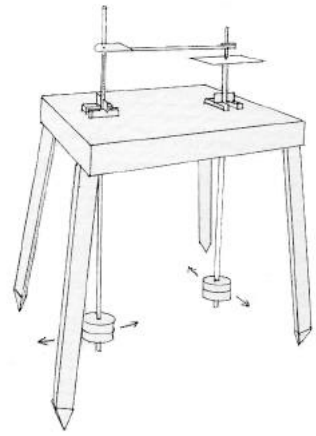
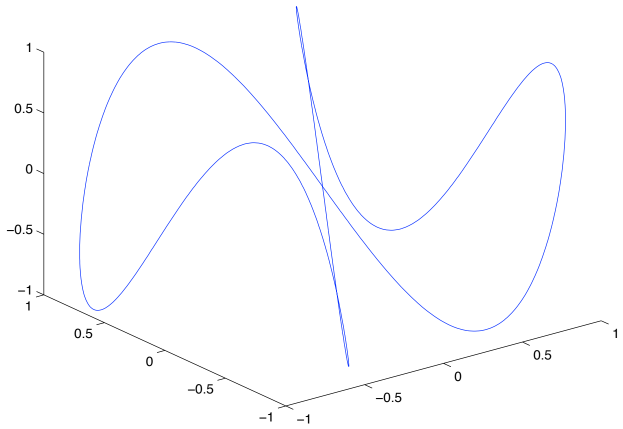
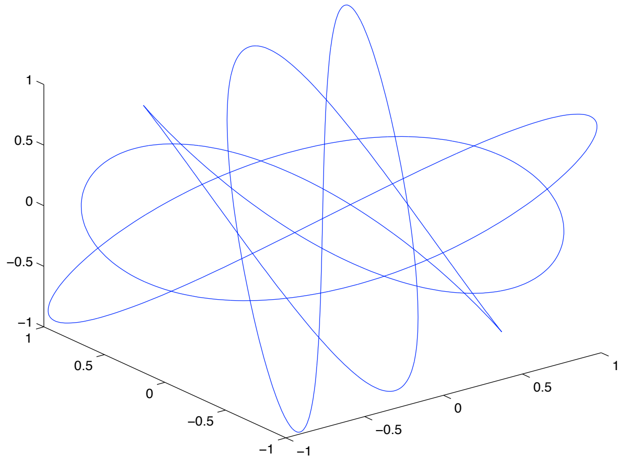


Figure 3–7. Lateral Harmonograph (Ashton 2003).

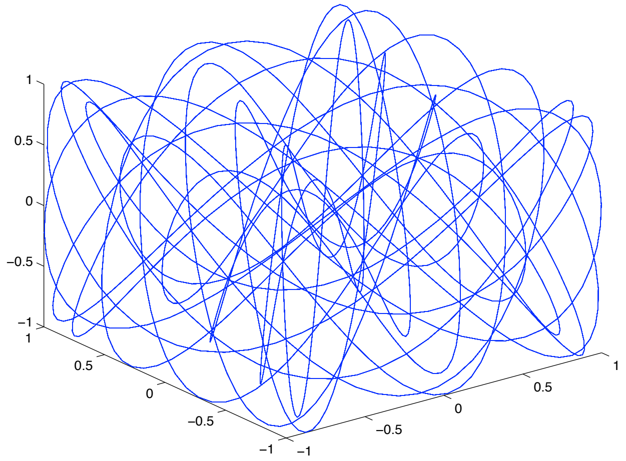
Classical Lissajous figures can easily be extended into three dimensions, and thus patterns that correspond to three frequencies of pure tones, for example, can be observed. Several example three-dimensional Lissajous figures are shown in Figure 3–8. Figure 3–8a illustrates a 1:2:4 chord (two octaves), Figure 3–8b shows a 4:5:6 major triad, and Figure 3–8c illustrates a 15:16:29 triad. It is apparent from the graphs that the "complexity" of each curve indicates a sense of Eurlerean dissonance.



(a)



(b)



(c)

Figure 3–8. Three-dimensional Lissajous curves representing (a) a 1:2:4 triad (i.e., two octaves); (b) a 4:5:6 major triad; and (c) a 15:16:29 triad.

Plotting time-varying frequency trios would illustrate a simple yet effective display of, for example, the fundamental frequencies in three-voice polyphony. Other extensions using colormaps and surfaces instead of simple curves might yield other insights. Lissajous curves have been replaced, so to speak, in modern signal-processing applications by functions like the autocorrelation and cross-correlation measures, which provide a number corresponding to the similarity of a signal with respect to itself (autocorrelation) or another signal (cross-correlation).

The Kaleidophone

Inspired by the then-recent invention of the kaleidoscope, Sir Charles Wheatstone (1802–1875) invented another means for visualizing numerical intervals. The device, called the kaleidophone (Figure 3–9), was particularly useful for graphically displaying patterns that correspond to musical intervals. Webster's 1913 Unabridged Dictionary offers a definition of the device:

> *An instrument invented by Professor Wheatstone, consisting of a reflecting knob at the end of a vibrating rod or thin plate, for making visible, in the motion of a point of light reflected from the knob, the paths or curves corresponding with the musical notes produced by the vibrations.*

The instrument could emit surprisingly interesting images depending on how the kaleidophone was struck. The generated images could subtly change by then bowing the thin rod with a violin bow. Ashton (2003) relates that Wheatstone referred to the instrument as a "philosophical toy."

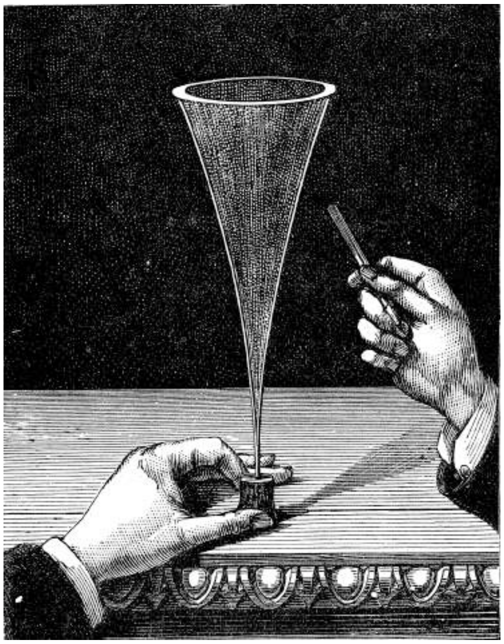


Figure 3–9. Sir Charles Wheatstone's Kaleidophone (1827) (reprinted in Ashton 2003).

Chladni Patterns

By sprinkling sand or salt on a plate and forcing the plate to vibrate, the sand will naturally relocate to the locations of the nodes that result on the surface. The pattern that results, called a Chladni Pattern and discovered by scientist and amateur musician Ernst Florens Friedrich Chladni (1756–1827), graphically illustrates the outline of the eigenmodes of the surface and is directly related to the frequency at which the surface vibrates[i]. The surface can be made to vibrate in several ways: (1) by a high sound-pressure-level sound wave; (2) by friction (e.g., bowing a surface with a violin bow); or (3) by direct stimulation with a mass, for example via direct coupling to an eletromechanical transducer. Chladni Patterns have been used for centuries to tune violin plates and soundboards of various other musical instruments; by adjusting the mass and/or geometry of the soundboard, it can be tuned to resonate in a desired way. (See Levin 2003 for an interesting discussion.) An example pattern is shown in Figure 3–10.

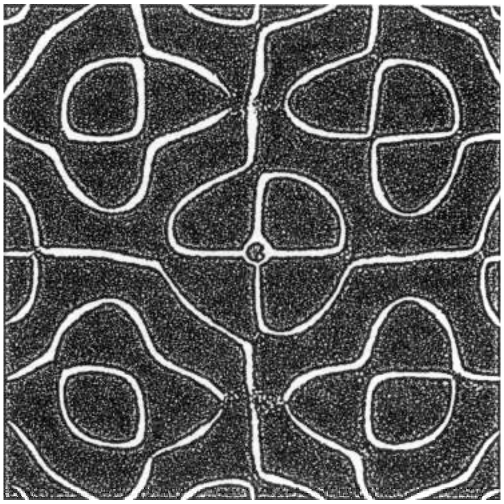


Figure 3–10. Chladni pattern from Jenny (1967/1972, reprinted in Ashton 2003) for a particular frequency of plate vibration.

---

[i] Legend has it that Napoleon gave Chladni 6000 Francs for his demonstration, and he offered a reward of 3000 Francs to anyone who could explain it. The reward was given to Sophie Germain in 1816.

Chladni patterns can be extended in several interesting ways. By forcing a surface to vibrate at two or more frequencies simultaneously, characteristic patterns will form. Thus, for example, different musical intervals can form identifying patterns, allowing us to visualize and compare them graphically.

Chladni patterns have also been extended beyond two-dimensional surfaces to illustrate the eigenmodes of N-dimensional shapes[ii]. This is predicted entirely by the Helmholtz wave equation, which describes the propagation of waves through a function p of N dimensions:

$$\nabla^2 p + k^2 p = 0$$

where $\nabla^2 p$ is the gradient of the gradient of the function p, and k is the so-called wave number. (In one dimension, the gradient of the gradient can be thought of as the "curvature" of the string; in two dimensions, it can represent the curvature of a surface, such as a drum head; in three dimensions, we can think of it as the acceleration of particle flow per unit volume of air.) The wave number is defined by the ratio ω / c, where ω is the radian frequency of oscillation of the wave, and c describes the speed of the wave. In the simple case of a one-dimensional p, say

---

[ii] Chladni patterns in one dimension should be familiar: vibrating strings naturally show their nodes and antinodes when struck, bowed, plucked, or otherwise excited.

p(z), the gradient-squared reduces to a second derivative, and thus the equation reduces to

$$\frac{d^2p}{dz^2} + k^2 p = 0$$

The solution to this second-order differential equation is of course simply the complex exponential $p(z) = Ke^{-jkz}$, which, according to Euler's identity, can be reduced to

$$p(z) = K\big(\cos(kz) + j\sin(kz)\big),$$

where *K* is an arbitrary constant.

Interesting patterns were produced on real-world objects by Hans Jenny (1904–1972), a Swiss doctor (and perhaps a kind of twentieth-century polymath Robert Fludd), who coined the term *cymatics* to describe the study of wave phenomena from a Neo-Pythagorean stance. His book *Cymatics: The Structure and Dynamics of Waves and Vibrations* (Volume 1, 1967; Volume 2, 1972) documents his many experiments and results, and it highlights some of his inventions, including the tonoscope, a non-electronic vocal-sound visualization apparatus based on Chladni-type phenomena. Jenny's ideas reportedly influenced American composer Alvin Lucier, particularly in his work *Queen of the South* (1972).

Multidimensional Chladni surfaces illustrate nodes of vibrating surfaces along orthogonal axes; said another way, they graphically show the places where vibrational harmonics cancel each other. A three-dimensional Chladni surface can be described by any linear combination of sinusoids, for example

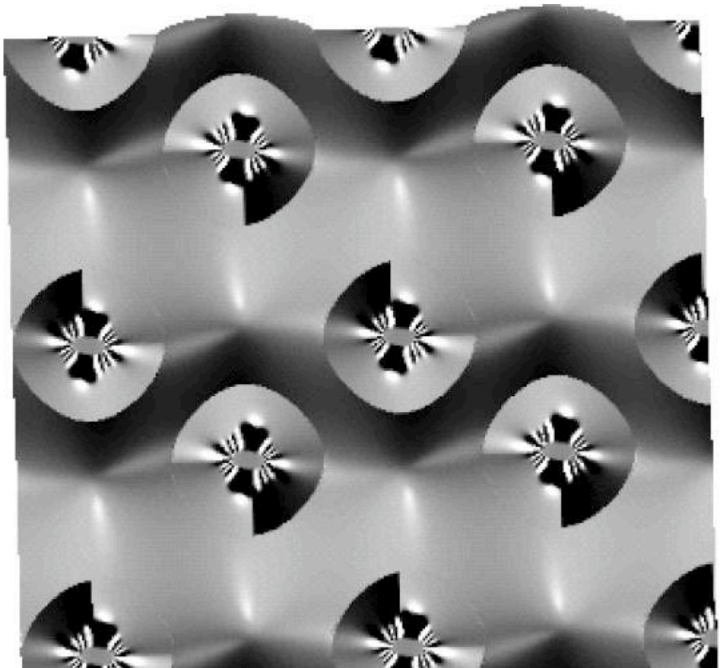
$$\cos(ax) + \cos(by) + \cos(cz) = 0$$

Solving for *z*, we obtain

$$z = \frac{1}{-c}\cos^{-1}\left(\cos(ax) + \cos(by)\right)$$

If $0 \leq x < \pi$ and $0 \leq y < \pi$, it follows that the three-dimensional surface defined by *z* is periodic outside these intervals owing to the periodicity intrinsic in the cosine function. Thus, three-dimensional Chladni patterns are easy to synthetically generate. An example surface for the case $a = b = c = 0.5$ is shown in Figure 3–11. By making the parameters *a*, *b*, and *c* represent audio frequencies, we can visualize three-dimensional Chladni vibration patterns that are characteristic for trichords of pure tones.

(a)

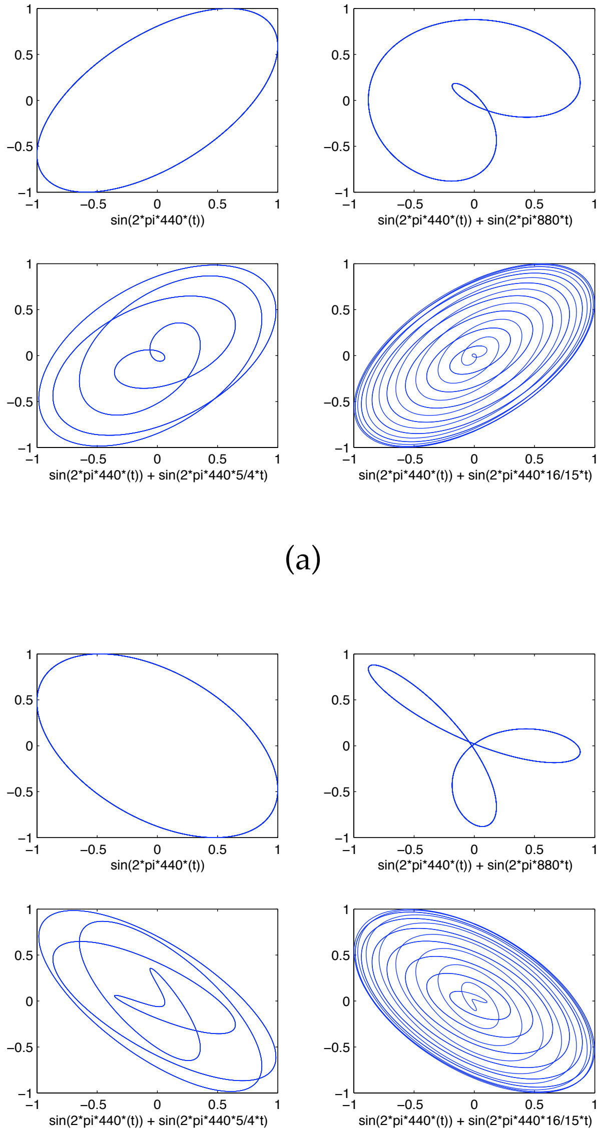


(b)

Figure 3–11. (a) Mesh surface for $\cos(0.5x)+\cos(0.5y)+\cos(0.5z)=0$; (b) another perspective. (Axes have been removed for clarity.) Note the spatial periodicity.

Phase-Space Diagrams

Phase-space diagrams, also called state-space or lag-space plots, have been a part of the standard digital signal-processing arsenal for many years, and they are widely used in many other fields as well, from nonlinear dynamics to biology and neurophysiology. In its two-dimensional formulation, the phase-space diagram represents the value of a signal as a function of a delayed version of itself. For example, Figure 3–12 illustrates lag-space diagrams of pure-tone signals. Figure 3–12a shows plots for intervals of (a) a unison (a 440 Hz sine wave); (b) octave (440 Hz and 880 Hz); (c) 5/4 major third (440 Hz and 550 Hz); and (d) 16/15 minor second (440 Hz and $469\frac{1}{3}$ Hz). Each combined signal is plotted here against itself delayed by 15 samples at a sampling rate of $F_s = 44.1$ kHz. Figure 3–12b shows the same set of intervals, but with a lag of 33 samples at $F_s = 44.1$ kHz. Note the closed loop that results, which is indicative of periodic, non-chaotic signals.



(a)

(b)

Figure 3–12. Lag-space plots of a pure-tone unison, octave, 5/4 major third, and 16/15 minor second for a lag time of (a) 15 samples and (b) 33 samples.

Whereas the autocorrelation of a signal is a function that describes lags of maximal self-similarity, phase-space diagrams graphically illustrate the effect of a single given lag on a signal. Such diagrams have been used in various aspects of audio signal analysis and machine listening applications, for example, auditory display (Gerhard 1999), sound synthesis (Di Scipio 1999), and timbre classification (e.g., Stone 1998). Parenthetically, autocorrelation-based methods have found similar use in tempo tracking (e.g., Ellis and Arroyo 2004), pitch detection (Rabiner, Dubnowski, and Schafer 1976; Rabiner 1977), and other tasks.

Lattice Graphs

The beauty of the aforementioned displays is their generality: we can quickly generate a graphical display that corresponds in a predictable way to the *acoustic* dissonance (not its many psychological correlates) of any sound—not just sine tones. Thus, acoustic phenomena like beats and harmoncity of intervals and even sound objects can generate unique and identifying patterns.

Other, more utilitarian depictions of musical intervals have been devised that are particularly useful for visualizing tunings, scales, and intervals. While many lattice-based displays of scales exist in antiquity, their modern refinement lives in Harry Partch's tonality diamond and expanded tonality diamond (Figure 3–13).

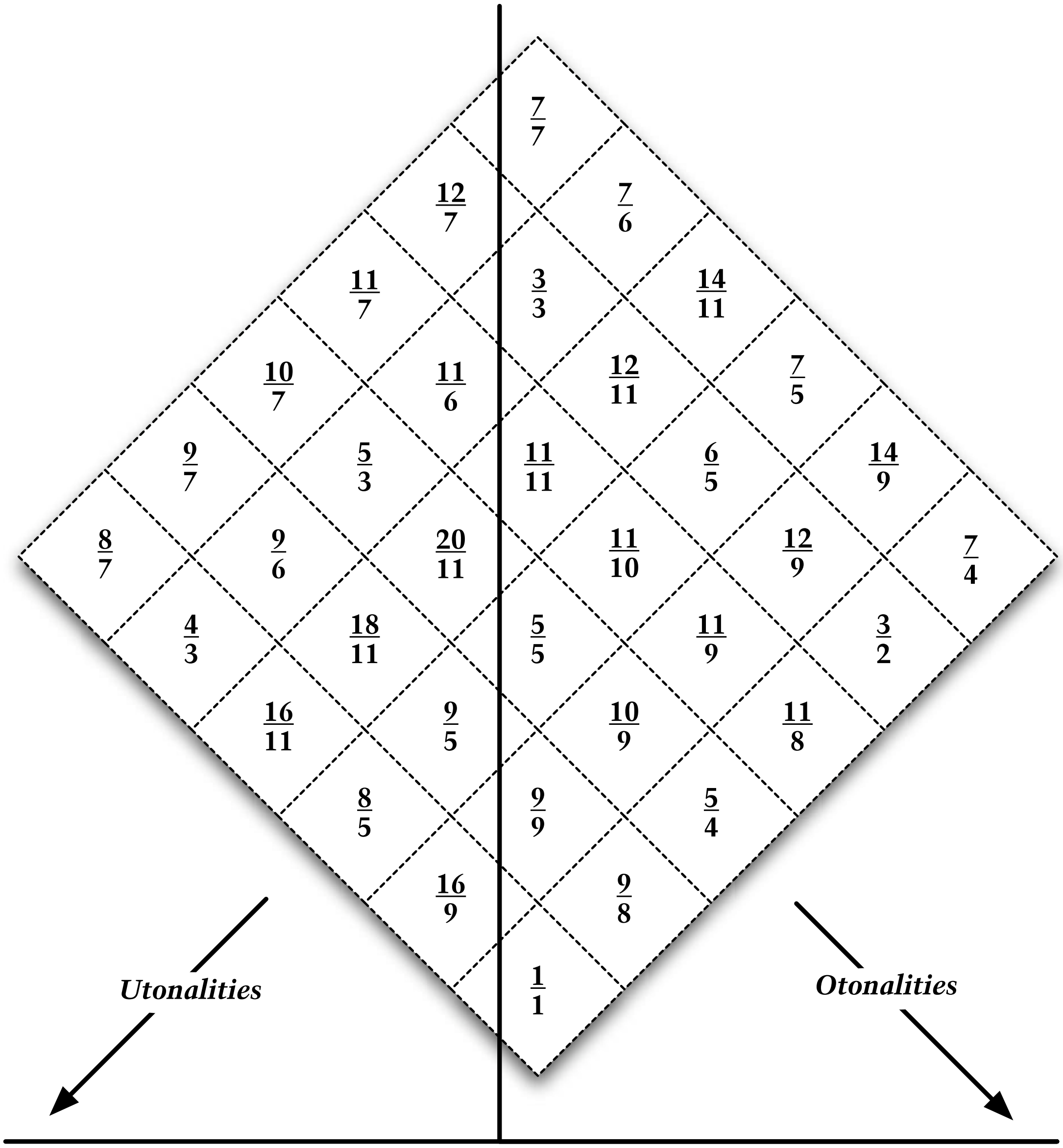


Figure 3–13. An 11-Limit Expanded Tonality Diamond (after Partch 1974).

The 11-Limit Diamond expresses the possible tonalities present using what Partch calls "identities" up to and including 11 (i.e., 1, 3, 5, 7, 9, 11). The diamond unfolds from the horizontal center outward by placing each unisons comprised of each identity along the central axis (1:1, 9:9, 5:5, and so on). Each row-column at which a given unison intersects is then populated with the corresponding identity. Ascending (i.e., traversing south-east, from left to right), the identity populates the numerator of the ratios in that row. Descending (i.e., traversing southwest), the identity populates the denominator. The set of ascending intervals derived from a pair of identities Partch calls *Otonalities*, and the set of descending intervals derived from a pair of identities he calls *Utonalities*.

The idea of using "identities" to generate a tuning lattice can be easily modified to illustrate a component of dissonance quite easily in which distance from the origin in the lattice yields an approximation of a given interval's "dissonance." Consider a simple lattice comprised of intervals produced only from the prime numbers 3 and 5. We call the result a 5-limit lattice, because 5 is the largest prime number contained in the graph. The dimensions of the graph are prime-orthogonal vectors. For example, in its two-dimensional form, the x axis may represent increasingly complex intervals (relative to the origin, a 1:1 unison) on 3: 3^0, 3^1, 3^2, …. The $y$axis may represent the same but on a different prime, say 5^0, 5^1, 5^2, …. The units are then octave-rectified (i.e., divided or multiplied by 2 to ensure each member R satisfies $1 \leq R < 2$, yielding in our example 1:1, 3:2, 9:8, … on the x axis and 1:1, 5:4, 25:16, … on the y axis. In the negative x and y directions, the exponents are simply inverted. (See Figure 3–14.)

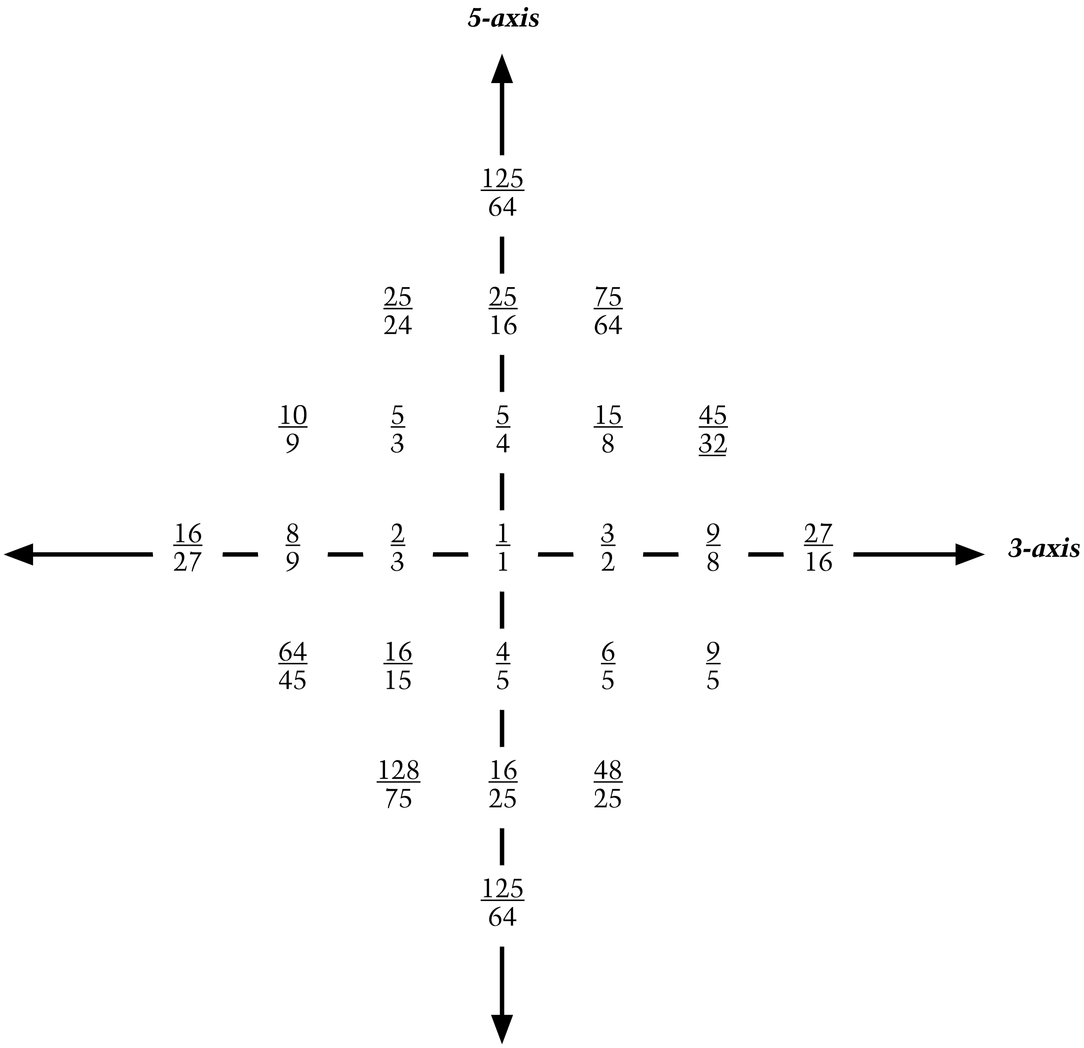


Figure 3–14. A simple 5-limit lattice.

In this formulation, the value of the interval $R(x,y)$ is simply the product of the corresponding x and y coordinate values. For example,

$$R(1,1) = \frac{3}{2} \cdot \frac{5}{4} = \frac{15}{8}$$

In prose, this equation simply states that a 5:4 major third plus a 3:2 perfect fifth yields a 15:8 major seventh. As another example, consider

$$R(-1,-2) = \frac{2}{3} \cdot \frac{16}{25} = \frac{32}{75}$$

which reduces to a downward step within the lower octave of 64:75 (or 128:75 when rectified and octave-reduced to the upper octave). In prose, this equation states that walking down a perfect fifth (2:3) and walking down another augmented sixth (16:25), we land on a combined interval of an augmented second (128:75) when octave-reduced with respect to unison (i.e., when the augmented twelfth is reduced to an augmented second).

Among other things, tonality diamonds illustrate a path by which the distance from the origin $R(0,0)$ to the interval $R(x,y)$ is directly proportional to Eulerean dissonance; that is, the further we stray from the origin, the more numerically complex the intervals become.

This can be extended to an arbitrary number of dimensions, allowing visualization of numerical dissonance along multiple prime-orthogonal bases. This concept will be addressed in musical terms later.

Provided the basis vectors are orthogonal, tonality diamonds and lattices could be extended to display computational dissonance of objects other than musical intervals. For example, a three-dimensional lattice could be formed to graphically represent the “dissonance” of successive transformations of a sound object. In this case, the origin would represent the original, unprocessed sound object. The *x* axis could represent increasing discrete tonal fusion metrics, the *y* axis could represent increasing harmonic richness (i.e., decreasing purity), and the *z* axis could represent decreasing harmonicity (i.e., increasing inharmonicity). Such a depiction would be difficult in practice, however, for two reasons: the search for and proof of orthogonal psychological correlates is difficult, if not impossible (i.e., how can one prove conclusively that one correlate is always exclusive of another?), and the formulation of sound-processing algorithms that could yield results that exhibited predictable feature vectors in multiple dimensions would be difficult.

One of the primary values of lattice diagrams is the primacy they give prime numbers. Indeed, personal experience suggests that "primeness" plays a fundamental role at some level in our perception of combinations of tones. Additively synthesizing a complex tone from pure sine waves sounds altogether different when prime harmonics are emphasized or omitted. Primeness is clearly related to the fusion of complex tones. The confluence of prime numbers, dissonance, and lattice diagrams are summarized by Joe Monzo (2004):

> *My own theory of sonance actually holds that there are two separate continua of sensation, one determined by the values of the prime-factors of the ratios interpreted by the listener as being that of the two tones in the interval, and the other determined by the values of the exponents of those factors. Dissonance increases (and consonance*

> *simultaneously decreases) as both the prime-factors and the values of the exponents of those factors become larger. This idea was expressed earlier by Ben Johnston and others; the earliest reference to it which I have seen is in* The true character of modern music*, written in 1764 by the mathematician Leonhard Euler. Harmonic lattice diagrams are a graphical representation of this theory of sonance*[iii].

Spherical Harmonics

Spherical harmonics (Ferrers 1877) provide a beautiful, elegant, and concise way to illustrate four or more dimensions of data in a simple three-dimensional projection. They are used quite often in chemistry and quantum mechanics to describe the possible surfaces of travel of electrons in an atom, for example. They have also been used in acoustics recently to study multichannel recording techniques (e.g., Moorer 2001) and acoustic propagation of sound waves (Giron 1996).

Spherical harmonics are defined as the angular portion of the solution p to Laplace's Equation, which is given by

---

[iii] Some recent tuning theorists distinguish *concord* as purely a psychoacoustic percept divorced from musical context, whereas *sonance* or *sensory context* takes into account both psychoacoustic modeling as well as musical context.

$$\nabla^2 p = 0$$

Laplace's Equation itself is a special case of the Helmholtz form of the wave equation in which $k = 0$:

$$\nabla^2 p + k^2 p = 0$$

where k is the wave number, defined as the ratio of the radian frequency of oscillation of the wave ω to the velocity of propagation of the wave c. A concise derivation of the solution to the angular portion of this equation is given in Weisstein (2004), and the result for three dimensions in terms of polar angles θ and ϕ is

$$Y_l^m(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} \cdot P_l^m(\cos\theta)e^{jm\phi}$$

where l and m are constants (with $-l \le m \le l$), $P_l^m(\bullet)$ is the lth Legendre Polynomial raised to the m power, and $j = \sqrt{-1}$.

Spherical harmonics can also be illustrated in a simpler manner by plotting linear sums of sinusoids in a spherical coordinate system, for example

$$r(\theta,\phi)\left\{=\sum_{k=0}^{N-1}\sin(a_k\phi)^{a_k'}+\cos(b_k\phi)^{b_k'}\right\}+\left\{\sum_{k=0}^{N-1}\sin(c_k\theta)^{c_k'}+\cos(d_k\theta)^{d_k'}\right\}$$

In this case, the radius r that defines the three-dimensional surface is a weighted combination of spatial periodicities of azimuth and elevation (θ and ϕ). In the case of four sinusoids, we could rewrite this in the form

$$r(\theta,\phi)=\sin(a_0\phi)^{a_1}+\cos(b_0\phi)^{b_1}+\sin(c_0\theta)^{c_1}+\cos(d_0\theta)^{d_1}$$

Provided that the constants $\{a_0, a_1, b_0, b_1, c_0, c_1, d_0, d_1\}$ are real integers, the resulting surface described by $r(\theta,\phi)$ is a closed surface.

One of the beauties of spherical harmonics lies in their intuitive interpretation; they illustrate the spherical nodes and antinodes of phenomena that obey the wave equation—that is, the angles and magnitudes at which harmonics cancel and reinforce. And quite simply, the more numerically "complex" the relationships among the constants $\{a_0, a_1, b_0, b_1, c_0, c_1, d_0, d_1\}$ (or l and m in the Legendre form), the more "pointy" and "rough" the resulting surface becomes. Hence, spherical harmonics offer an elegant and concise depiction of Eulerean consonance by simply mapping the frequency content of a signal (say, for example, the frequencies present in a triad composed of pure sine tones) to the set of constants above.

Consider, for example, Figure 3–15, which illustrates spherical harmonics of a perfect unison. Note that the exponents a_1, b_1, c_1, and d_1 each effectively turn "on" or "off" the corresponding frequency constant by setting each exponent to 0 (off) or 1 (on). Thus, to display a 1:1 unison, we can set $a_0 = 1$, $b_0 = 1$, $c_0 \in \Re$, $d_0 \in \Re$, $a_1 = 1$, $b_1 = 1$, $c_1 = 0$, and $d_1 = 0$. (Here, $\Re$ represents any real number.) Thus, the display shown in Figure 3–15plots a smooth, spherical "apple".

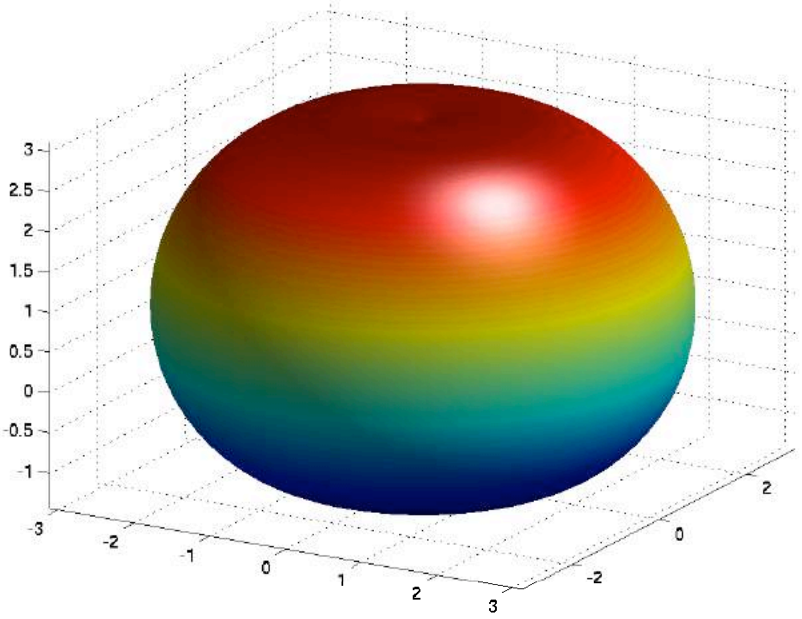


Figure 3–15. Spherical harmonics of a perfect unison.

To display a 3:2 frequency ratio, we set $a_0 = 3$, $b_0 = 2$, $c_0 \in \Re$, $d_0 \in \Re$, $a_1 = 1$, $b_1 = 1$, $c_1 = 0$, and $d_1 = 0$. The result is shown in Figure 3–16.

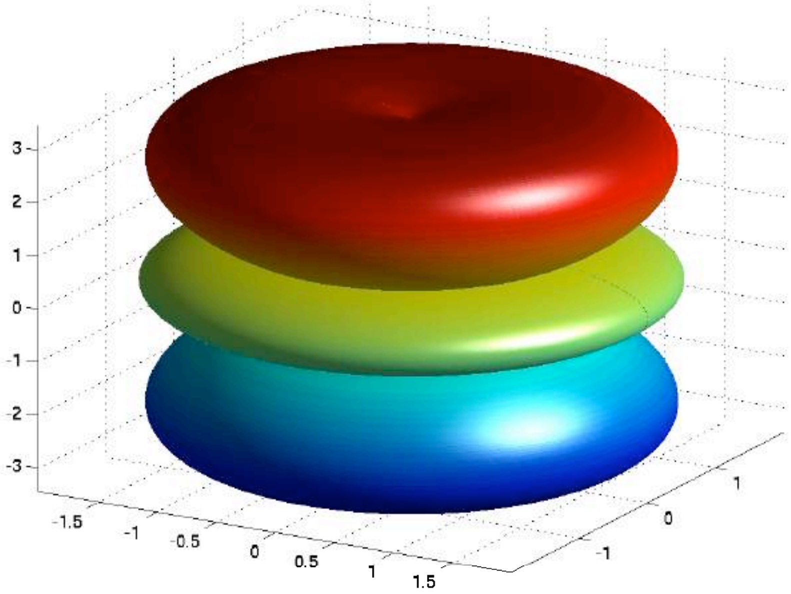


Figure 3–16. Spherical harmonics of a 3:2 perfect fifth.

Compare the surface illustrating the 3:2 perfect fifth with the spherical harmonics of a particular mistuned (“wolf”) fifth given by 55:36, shown in Figure 3–17. The 55:32 exhibits a much higher Eulerean dissonance rating, and this is reflected in the corresponding loss of smoothness in the surface.

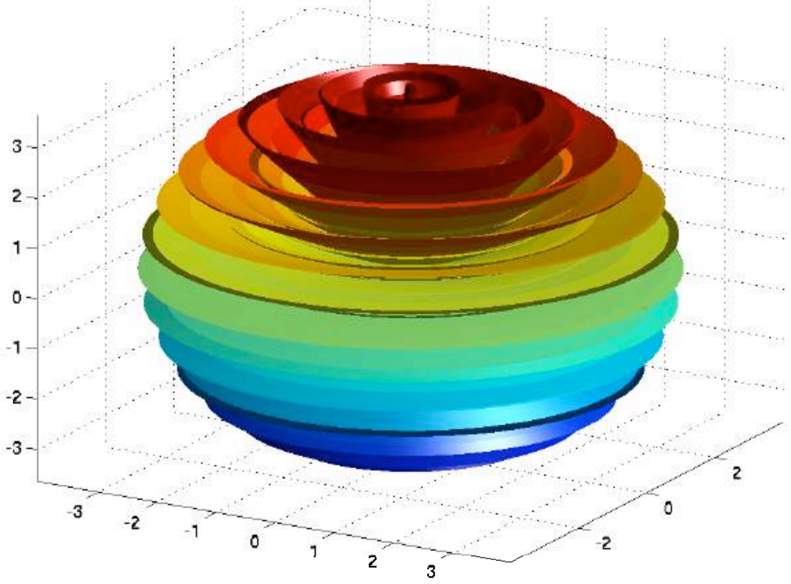
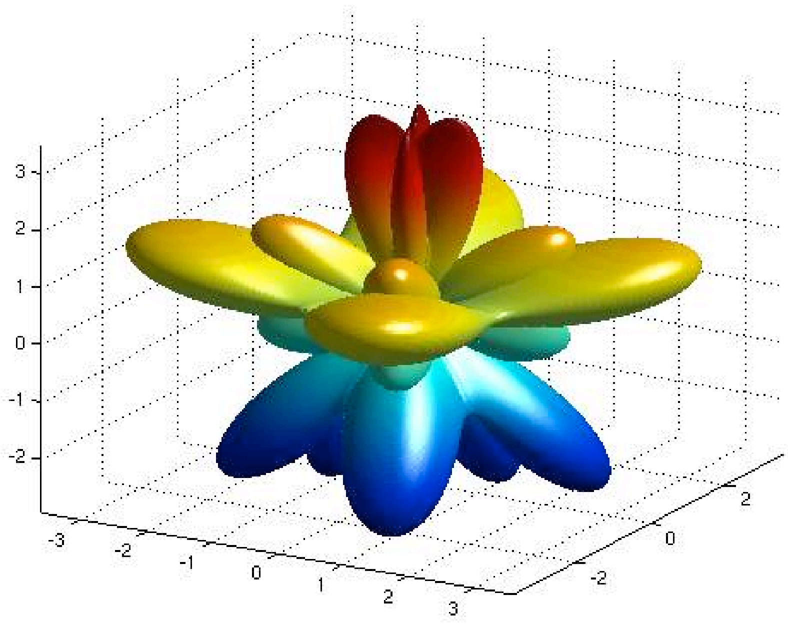
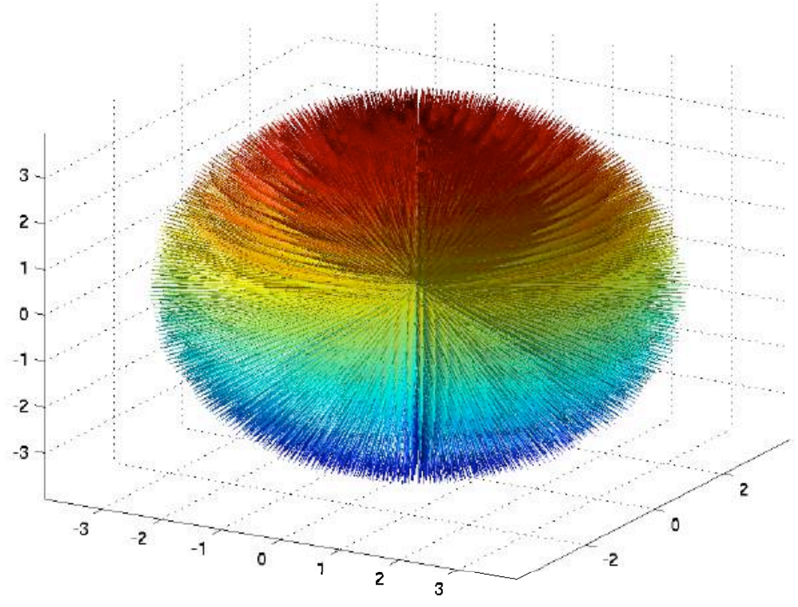


Figure 3–17. Spherical harmonics of a 55:36 wolf fifth.

Spherical harmonics can be used to display interactions of up to four normalized frequencies by setting a_0, b_0, c_0, and d_0 appropriately. As a final example, the spherical harmonics of a major triad tuned to $\frac{1}{1}:\frac{5}{4}:\frac{3}{2}$ (i.e., 4:5:6) and a more "gritty" $\frac{1}{1}:\frac{16}{15}:\frac{11}{8}$ (120:128:165) triad are shown in Figure 3–18. Note the relative smoothness in the surface contour of the major triad and the "jaggedness" of the triad in (b). We could quantify the "grittiness" of the spherical harmonics in various ways: for example, we might calculate the number of spherical peaks for a given neighborhood size, or we might instead simply compute the gradient ($\nabla r(\theta,\phi) = \frac{\partial r}{\partial \theta} \cdot \hat{\theta} + \frac{\partial r}{\partial \phi} \cdot \hat{\phi}$ in polar coordinates, or $\nabla r(x, y, z) = \frac{\partial r}{\partial x} \cdot \hat{x} + \frac{\partial r}{\partial y} \cdot \hat{y} + \frac{\partial r}{\partial z} \cdot \hat{z}$ in Cartesian coordinates) of each surface.



(a)



(b)

Figure 3–18. Spherical harmonics of (a) $\frac{1}{1}:\frac{5}{4}:\frac{3}{2}$ major triad and (b) $\frac{1}{1}:\frac{16}{15}:\frac{11}{8}$ triad.e

One might argue that the $\frac{1}{1}:\frac{16}{15}:\frac{11}{8}$ set produces a more "soft" or "fuzzy" plot than does the $\frac{1}{1}:\frac{5}{4}:\frac{3}{2}$ set. As others have noted, any attempt to map data from one sense to another (e.g., audition to vision) runs the risk of interference with other senses (e.g., touch) (Cook 2005). Clearly, the study of "reverse sonification," or visualization—that is, the graphical display of auditory phenomena—is relatively young, and many problems must still be solved. A comprehensive study in this area would also need to incorporate psychoacoustic models that inform the physical data, because as we know, perceptual correlates of vision and audition are entirely nonlinear with respect to physical phenomena. For example, by infusing graphical displays like those provided by spherical harmonics with psychoacoustic modeling of dissonance perception, one might be able to match the auditory “smoothness” to linearly complement the visual “smoothness” of the display.

Dissonance Displays: Comments and Future Work

The list of potential dissonance displays presented here is far from complete. Other novel methods for visualizing audio signals, such as information derived from the wavelet transform (Cheng 1996; Tzanetakis, Essl, and Cook 2001) and Wigner Distributions (Preis and Georgopoulos 1998; DeMeo 2002) may yield many interesting insights when applied to the examination of audio signal dissonance.

However, potential problems abound when one attempts to map data from one domain into another domain of perception. Furthermore, the aforementioned displays are limited in that they only consider physical phenomena and not their perceptual correlates. Development of a robust auralization that corresponds to the complete experience of musical/auditory dissonance must certainly include accurate models of both physical attributes and perceptual contributors.

Because consonance and dissonance subsume the interaction of a finite but unbounded number of dimensions of data (both acoustic and perceptual), all of the aforementioned displays are inherently limited. They are useful only to the extent that we realize the limited aspects of consonance and dissonance that they display. As such, the simplest kinds of dissonance measures to display are numerical in nature, as with Euler's dissonance metric. That being said, the attempt to quantify and display the confluence of physical and perceptual data, however limited in execution, can be extremely helpful. Consider, for example, the Fletcher-Munson equal-loudness contours and the Mel scale of pitch perception, both of which try to display graphically the psychoacoustic correlate of a physical phenomenon. Both scales, many decades after their introduction, are used in virtually all modern applications that incorporate a psychoacoustic model of some sort, for example as found in recent MPEG standards. The continued exploration of the graphical display of dissonance data, even of a limited number of dimensions, may prove fruitful in developing appropriate feature vectors for signal-processing analysis of audio streams.

3.5 Dissonance as a Musical Control Structure

Although consonance and dissonance are difficult to define, several contemporary composers have selected a particular definition of the terms that suits a musical idea and used it to structure an entire composition. Inspiration for employing consonance and dissonance theory to form the basis of compositional macrostructure comes particularly from James Tenney, who wrote in his master's thesis in 1961 that "all parameters may be involved in the determination of structure in a musical configuration" (Tenney 1992, p. 63). He continued: "any parameter may function as the primary determinant of form in a clang—if only because it is possible to reduce to zero the degree of articulation of every other parameter within the clang." (Borrowing from gestalt psychology, he defines *clang* as "a sound or sound-configuration which is perceived as a primary musical unit or aural *Gestalt*.")

If, as Tenney wrote, "the form of a musical configuration is primarily determined by the effective differences between its successive parts," and we have any conceivable parameter available for creating musical differences, consonance theory may potentially serve to tightly organize an entire work. Wessel (1979) and others have written about timbre as a structuring principle in musical compositions, and the ability to sculpt timbre at the micro-level with a computer (and hence explicitly organize a work according to some timbral property of properties) has allowed entire cultures and sub-cultures of music to flourish recently. Given this, the extrapolation to dissonance—an entity somewhere between pitch and timbre—as an organizing principle in a work is not difficult.

Indeed, theories of consonance and dissonance have for centuries been involved in structuring music. The difference here is the explicit, detailed level to which consonance theory may be applied, and the primacy to which its focus is given. Furthermore, the extent to which measures of consonance and dissonance, both acoustically and perceptually, are quantified is much greater in certain recent compositions than in the past.

As an example, a work structured with regard to a particular composer's conception of consonance and dissonance may take an arch-like shape, as shown in Figure 3–19.

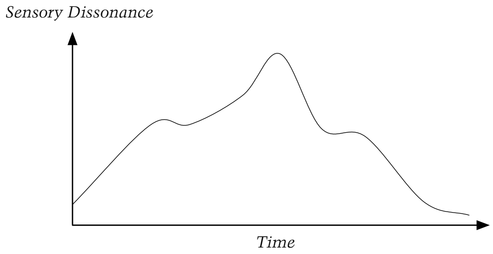


Figure 3–19. Arch-shaped dissonance form.

When used in conjunction with other formal procedures (repetition, variation, etc.), consonance and dissonance can serve as effective vehicles for musical form. *Critical Band* (Tenney 1988), *Psaltery* (Polansky 1978), and my own compositions *Nexus* (1997) and *Motus Lentus* (1997–) exemplify the interweaving of process and dissonance theory as formal mechanisms.

3.6 Composition within Dissonance Space

James Tenney's *Critical Band* (1988) for unspecified sustaining instruments examines, as its name implies, the phenomenon of the critical bandwidth. After establishing a unison on A (440 Hz), the instruments begin to deviate geometrically in frequency on either side of the A. For the first half of the piece, the sonic effect changes from mistuning to chorusing, finally reaching a level of maximum roughness at the critical bandwidth. After the intervals among the instruments have exceeded the critical bandwidth (about a minor third), the resultant texture begins to force a harmonic, rather than purely timbral, interpretation.

Larry Polansky's tape composition *Psaltery* was written in 1978 and is dedicated to Lou Harrison. The 51 pitches in the work are tape loops recorded from a bowed psaltery and were all derived from the first seventeen harmonics of each member of an E major triad whose members are in a 4:5:6 intervallic ratio. The first seventeen harmonics of the root of the triad are slowly presented, clearly establishing a "tonic" or harmonic mode. Polansky (1989) describes the ensuing process:

> *[P]itches from the next series…begin to replace their closest neighbors until the series on 5 is complete. This process happens twice more, moving to the perfect fifth…and then back to the fundamental. finally, the series on the fundamental drops out.*
>
> *Harmonics entire according to their "prime complexity" in this order: 17, 13, 11, 14, 7, 5, 10, 5, 9, 12, 6, 3, 16, 8, 4, 2, 1.*

The replacement of neighboring harmonics in order of prime complexity creates a subtle modulation—a kind of harmonic "cross-fade" to the next series. This can be thought of as replacing the more dissonant intervals first (which the ear is perhaps ready to replace after hearing them for several minutes), and replacing the more consonant intervals last.

For reference, a sonogram of the opening 3 sec of the composition is given in Figure 3–20. (For this and future sonograms, the left audio channel alone was analyzed; the sound file was also normalized to 50% amplitude.) Note the continuous presence of the fundamental and the majority of the harmonics in roughly constant amplitude levels.

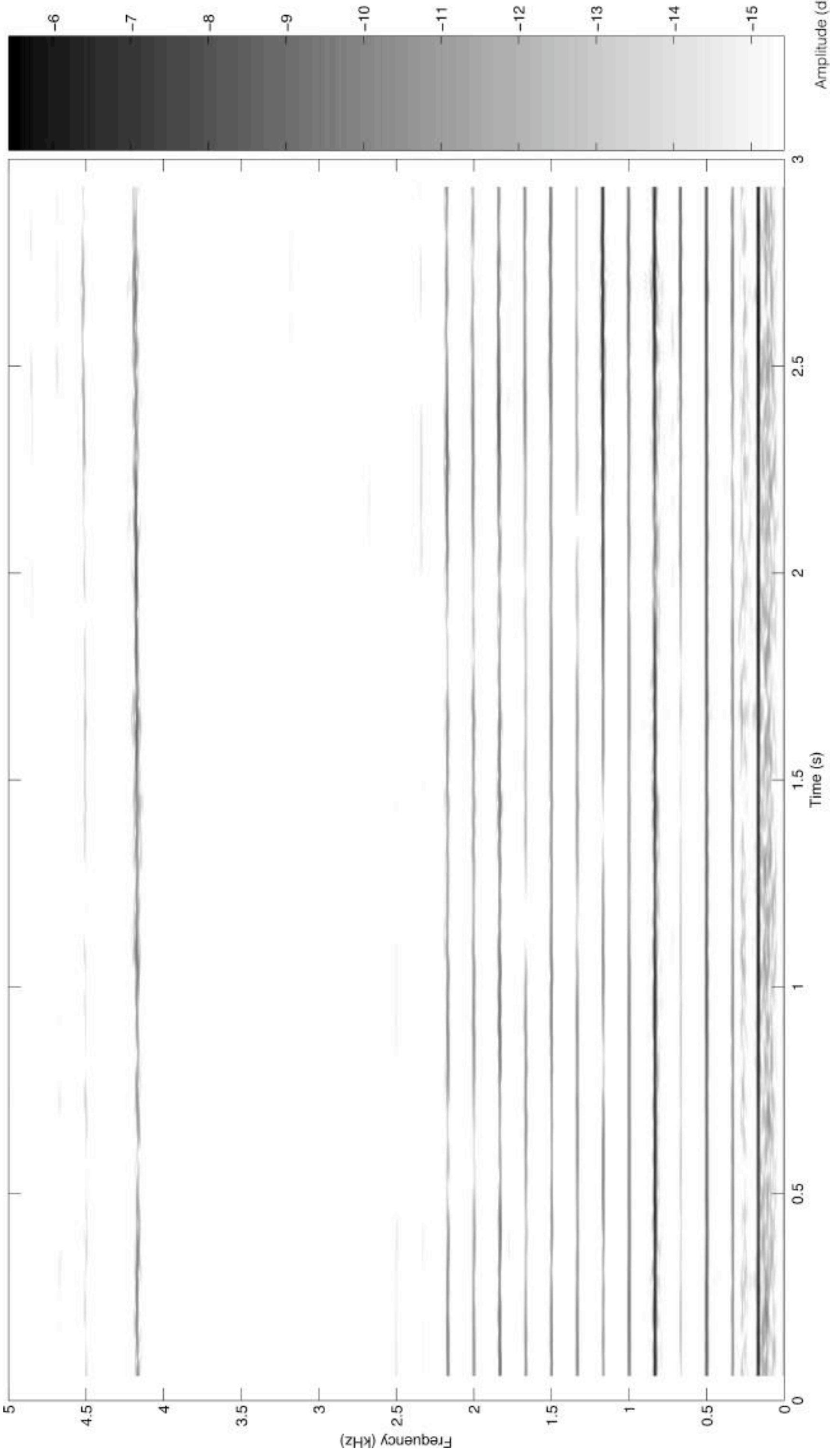


Figure 3–20. Sonogram of Larry Polansky's *Psaltery* (1978), opening 3 sec.

Figure 3–21 illustrates a sonogram of the composition from approximately 0:39 (zero minutes, thirty-nine seconds) to 0:42. Note the intermittent shading of the upper partials: this represents phasing that occurs as a result of the introduction of a member of the G-sharp series, which lies at a slightly different frequency from harmonics on the E series. The texture is static, but the introduction of partials from a different harmonic series creates the "blotching" of the sonogram's lines. In many ways, the compositional process may be seen as playing with the boundary between interval and timbre.

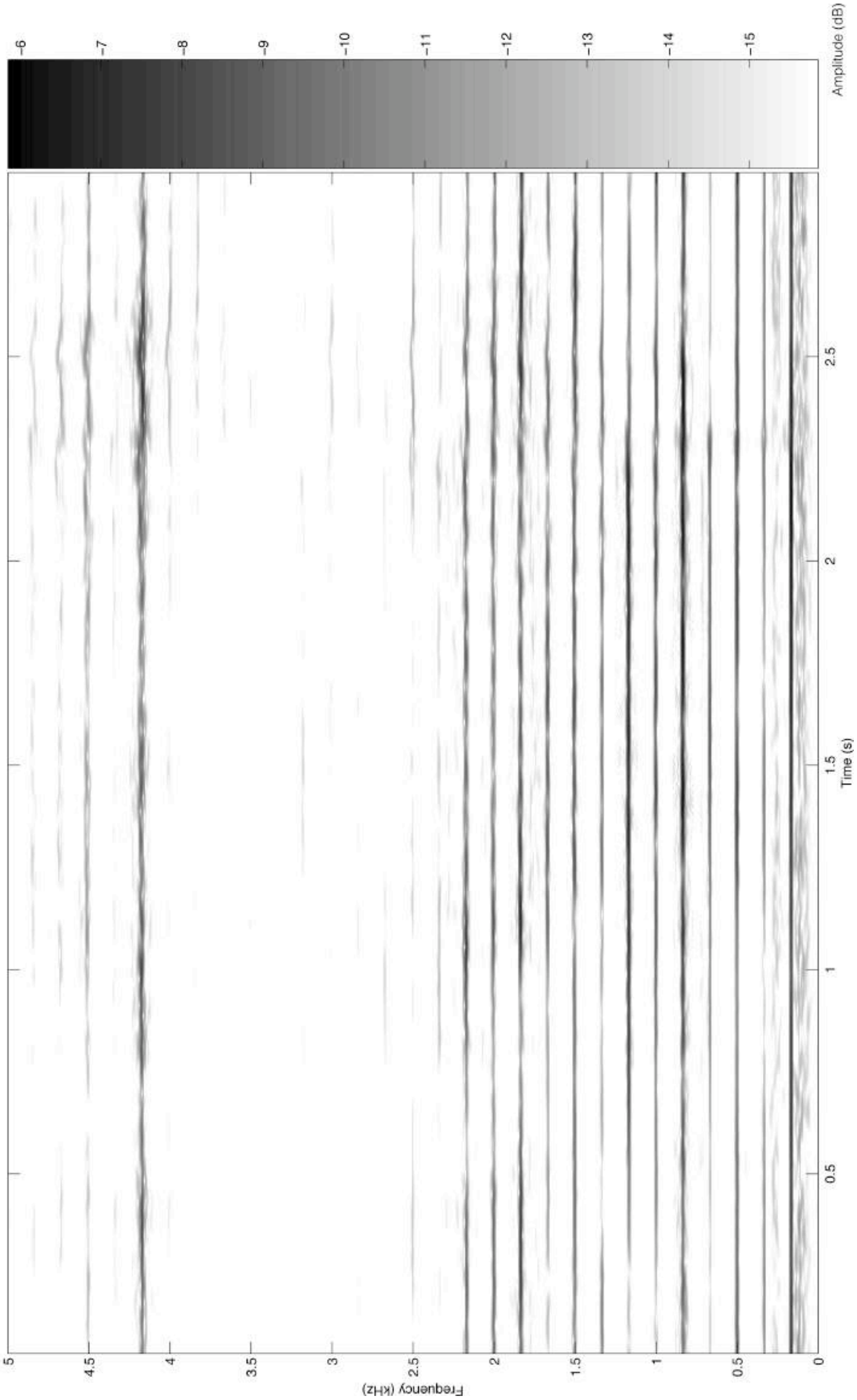


Figure 3–21. Larry Polansky, *Psaltery* (1978), 0:39 to 0:42.

This chorusing effect caused by the introduction of new pitches from the next harmonic series is illustrated again in Figure 3–22, which presents 0:46 to 0:49 of *Psaltery*. Note the increasing chorusing, which appears as “fuzziness” in the sonogram, of the harmonics present in the signal.

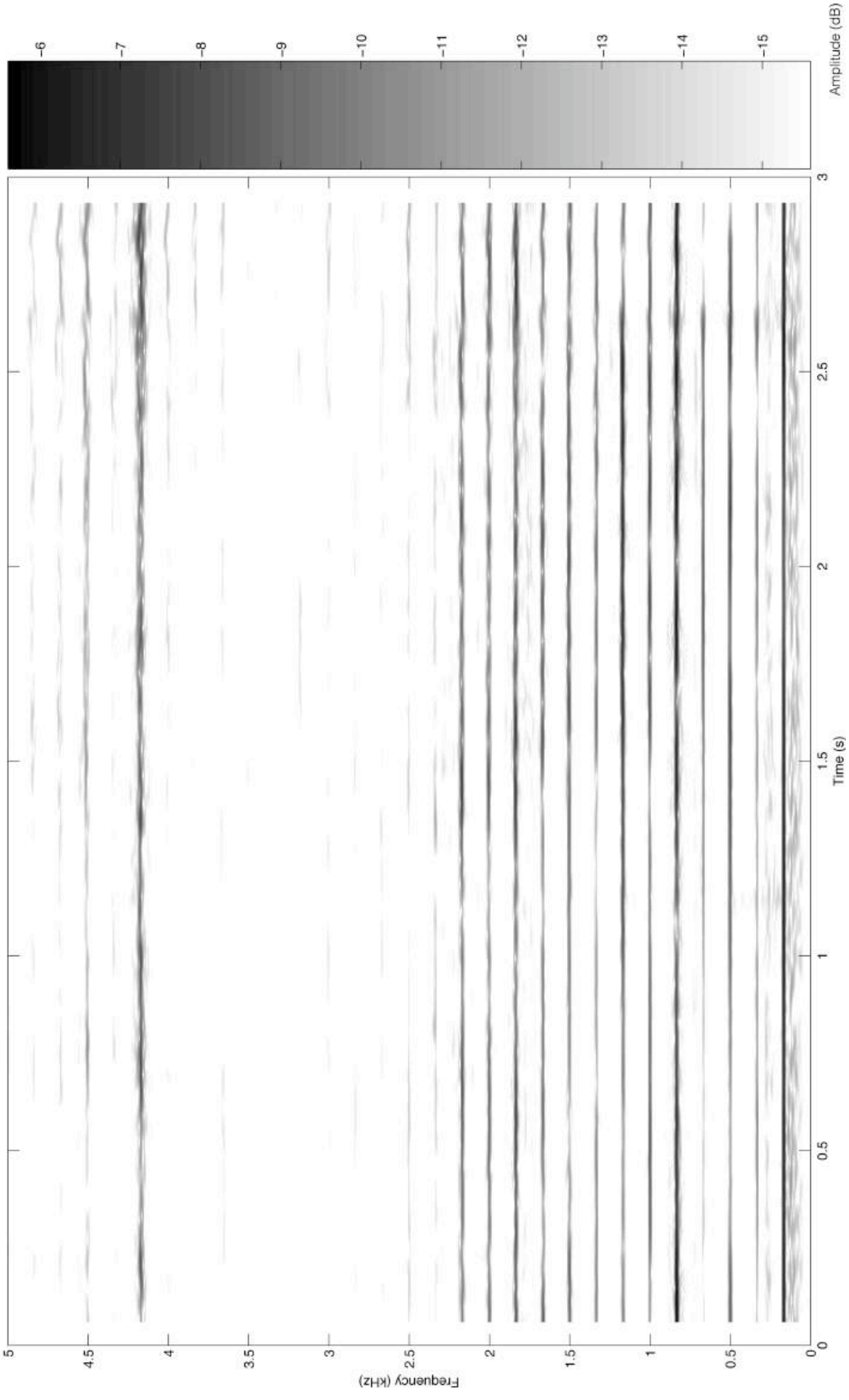


Figure 3–22. Larry Polansky, *Psaltery* (1978), 0:46 to 0:49.

Polansky's composition *Choir* (1978), which exists in a version for chorus singing in just intonation (tuning via headphones) and in a version for computer-synthesized stereo tape and solo voice, uses a similar process. Calculated navigation through a dissonance space by defining a metric for the measure of dissonance can serve as an effective organizing methodology for a composition.

My own composition *Nexus* (1997) for contrabass trio in just intonation is based on a catalog of all possible trichords from the first eleven members of the harmonic series. All three bassists play only harmonics on the first string: G for basses I and II, and a retuned F-sharp for bass III. (The G and F-sharp are tuned 6:5 and 9:8, respectively, above a "phantom" low E fundamental.) The piece is organized roughly along the lines of a logarithmic measure of dissonance into short phrases, so that more consonant trichords are presented initially and give way to more dissonant ones by the end of the work.

Figure 3–23 shows measures 4–6 of the work. The number below each note represents the harmonic number, while the number above each note gives the deviation in cents from equal temperament for the notated pitch. (This notation is based on that used by Polansky and others in works for instruments in just intonation.) Note that the resultant trichords often constitute relatively consonant minor triads (E minor and B minor in this example).

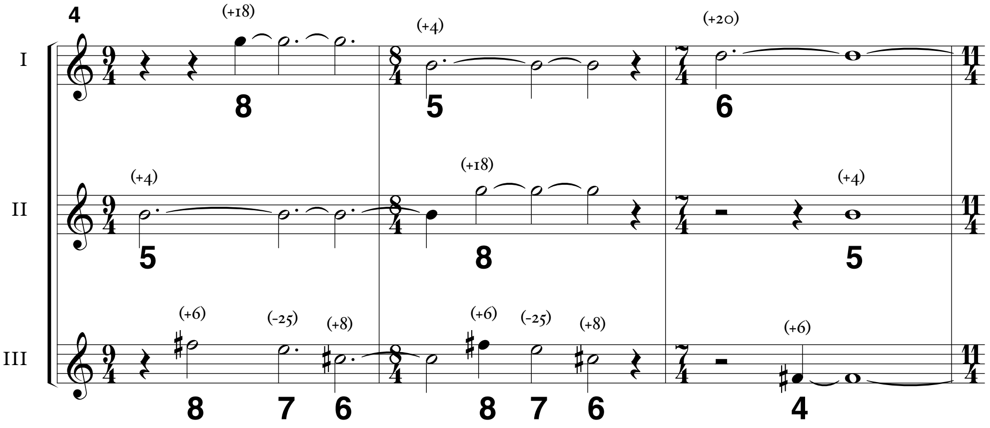


Figure 3–23. Colby Leider, *Nexus* (1997), mm. 4–6.

Figure 3–24 shows the final three measures of the piece. Clearly, the trichords are spaced much more closely—in fact, within a critical bandwidth. The structure of the work lies in its process—process based on thinking about consonance and dissonance.

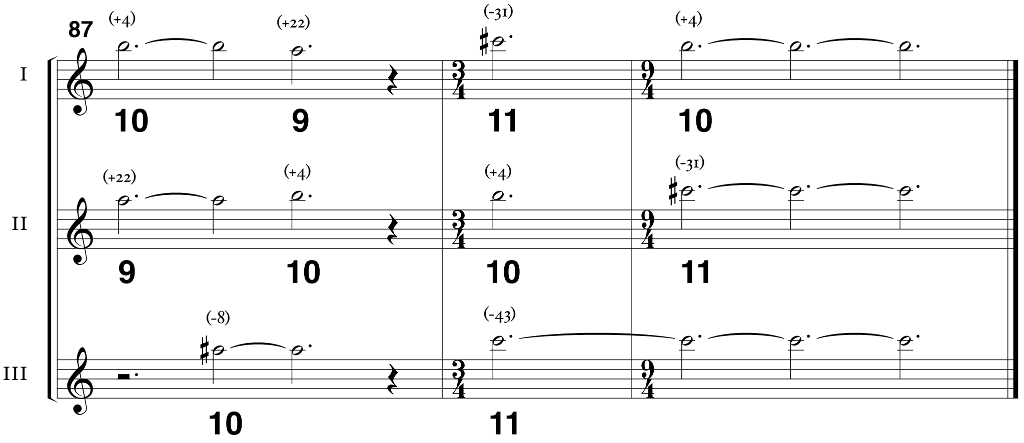


Figure 3–24. Colby Leider, *Nexus*, mm. 87–89.

Another composition of mine that directly employs dissonance calculation as a musical control structure is a synthesized work of unspecified duration for an unspecified number of channels, *Motus Lentus* (1997–). Inspired by Harry Patch's tonality diamond and a composition by composer Carter Scholz entitled *Lattice*, I constructed three three-dimensional just-intonation lattices to create three different pitch spaces. The first lattice I used is shown in Figure 3–25.

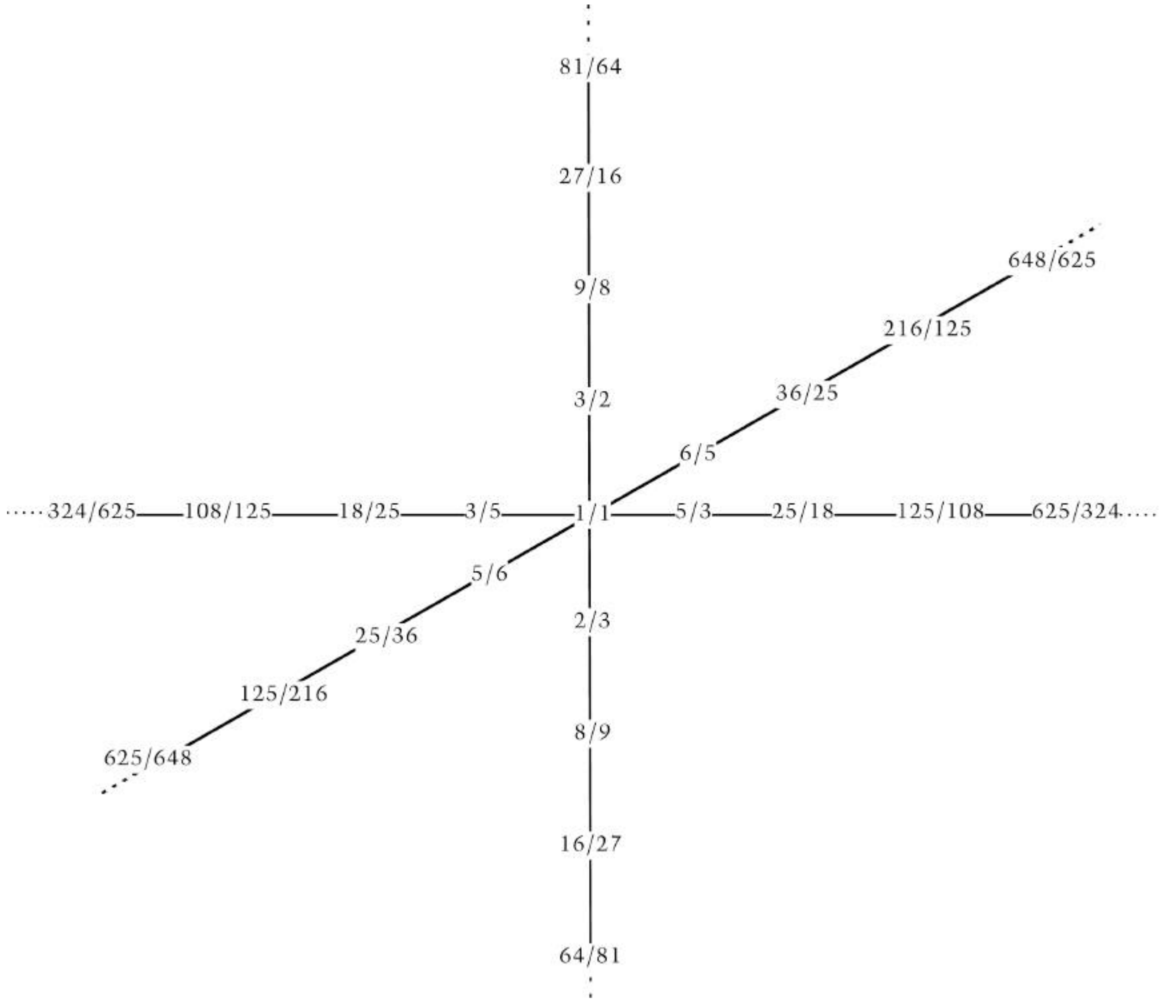


Figure 3–25. The first of three three-dimensional just-intonation lattices used in the composition of *Motus Lentus* (1997–).

Each lattice contains a different base interval on each of its x, y, and z axes. Axes are geometric in nature rather than arithmetic; for example, the members of the 9/8 major-second axis are

$$\left\{\left(\frac{9}{8}\right)^0, \left(\frac{9}{8}\right)^1, \left(\frac{9}{8}\right)^2, \ldots\right\}$$

Each result is octave-rectified to generate a potentially unbounded number of pitches within the octave (i.e., its pitch chroma is determined). Pitches at a given coordinate are given by the product of each corresponding base coordinate. For example, the interval R at the location (2, 2, 3) is given by the expression

$$R(2,2,3) = R(2,0,0) \cdot R(0,2,0) \cdot R(0,0,3)$$

In the previous figure, $R(2, 2, 2)$ is computed by

$$R(2,2,3) = \frac{25}{18} \cdot \frac{9}{8} \cdot \frac{216}{125}$$

$$= \frac{48600}{18000}$$

$$\equiv 81/60 \text{ (after octave - reduction)}$$

In simple prose, this process illustrates that stepping up two 5/3 major sixths (i.e., $R(2,0,0)$), followed by two 3/2 perfect fifths (i.e., $R(0, 2, 0)$), and finally adding three 6/5 minor thirds (i.e., $R(0, 0, 3)$) yields an 81/60 augmented fourth.

Motus Lentus begins by stochastically choosing pitches in a small radius around the origin of the first lattice. The orbit of possible choices gradually expands, resulting in overlapping chords of greater dissonance. The pitch-space gradually "morphs" into the second lattice via statistical replacement of pitches from the new lattice using the same radius of choices. Eventually, only pitches from the second lattice are chosen, and the radius of choices now shrinks in size until it reaches the origin of the second sphere, landing on a new unison. The process repeats as the pitch-space travels to the third lattice, and then the entire process unfolds in the reverse direction.

This process of statistical replacement of pitches between lattices is somewhat analogous to a modulation to a new tonal center and seems to have a similar musical effect. The transition is subtle, however, as most of the pitch-space replacement occurs at large radii (i.e., large prime-multiple ratios) of the lattices.

A signal-processing approach to dissonance calculation, based on some kind of parameterization of existing timbre- and dissonance-analysis models, would offer a distinct advantage: it would allow a new kind of analysis of non-notated music. Consonance measurement as an analysis tool has restricted itself to notated music; using a signal-processing approach, new insights could be gained into music of oral and non-Western traditions as well as electro-acoustic and computer music.

3.7 Dissonance and Compositional Microstructure

A digital signal processing approach to examining consonance and dissonance through harmonicity and other parameters would offer new ways of organizing compositional microstructure. At least two foreseeable means would be possible. The first, described by Sethares (1998), is consonance-based modulation. A second means would lie in extending Trevor Wishart's concept of the *timbre* tree (Wishart in preparation) to the *harmonicity tree*. Each of these possibilities is now explained.

By "consonance-based modulation" within a sound, Sethares suggests the possibility of altering a sound's inherent consonance over time. For example, a harmonic, highly consonant guitar strum may become dissonant in the course of the strum by gradually altering its spectrum. Alternatively, the tuning system in which a passage is played may change gradually, thereby affecting the perceived consonance level (Sethares 1998). A variety of means of constructing tuning system morphologies are explored and categorized in Leider (1996).

Wishart describes his compositional process as the construction, navigation, and mixing of trees of timbrally related sounds. In Wishart's timbre tree, a particular source sound is placed at the root of the tree, and all nodes below the root are processed versions of it. The tree is constructed according to the perceptual or statistical similarity of the sounds and the number of levels of processing each node has undergone, rather than according to the methods used to produce the sounds.

The notion of the timbre tree can easily be extended to incorporate the inherent consonance and dissonance of sounds, as shown in Figure 6. At each parent node, the choice of which child node to receive the processed sound is based on the harmonicity, for example: sounds with lower harmonicity are placed in the left child node, while sounds exhibiting higher harmonicity are placed in the right child node. Constructing a harmonicity tree in this way would provide a ready index of processed sounds, sorted according to harmonicity, for creating consonance-based musical microstructures.

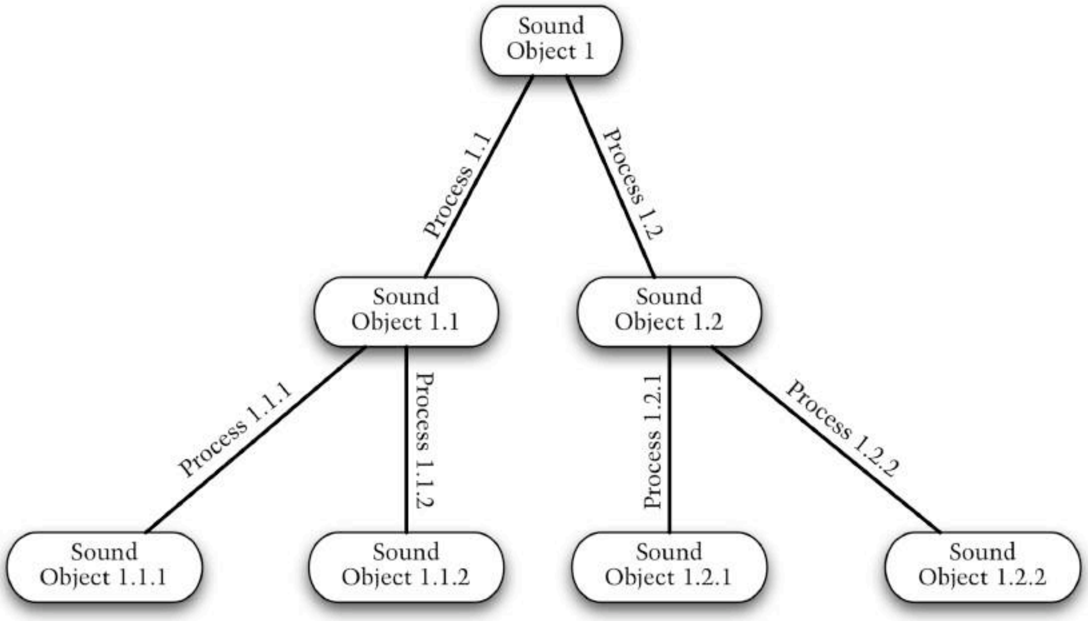


Figure 3–26. Harmonicity tree. The children of each node are arranged in decreasing harmonicity, left to right.

Is one linear sequence of abstract sound objects more consonant than another? Is one superimposition of sound objects more pleasing than another? Many factors, including harmonicity, may play a role in answering these questions.

Consonance and dissonance are difficult to define and equally difficult to discuss. The concepts of musical consonance, inherent consonance, sensory consonance, and tonal consonance all have different shades of often-contradictory meaning. Whatever their exact meanings, quantified models of the concepts have proven capable vehicles of musical form. The ideas presented in this section, specifically of the harmonicity tree and its logical extension—the consonance tree—offer new ways of thinking about the terms and possibilities for structuring music according to consonance theory.

4 POSTLUDE: THE DISSONANCE OF EMPIRICISM

> *Humor can be dissected as a frog can, but the thing dies in the process and the innards are discouraging to any but the pure scientific mind.*
>
> *—E. B. White (1899–1985), "Some Remarks on Humor"*

Clearly, many techniques are available to assist in the analysis of dissonance in music, from exacting neurological and cognitive measures to more general aesthetic criteria. But what is the point of assigning numbers to consonance levels? What is a consonance "level" anyway? Roederer (1973), in his introductory text on the subject on the psychophysics of music, writes

> *Consonance and dissonance are subjective feelings associated with two (or more) simultaneously sounding tones, of a nature much less well defined than the psychophysical variables of pitch and loudness, and even quality* [i.e., timbre]. *Whereas there is a relatively small variance among individual judgments regarding the latter, there is a much wider disparity when a given group of subjects judges the "consonance" or "dissonance."*

This sentiment summarizes the apparent futility in ranking consonance and dissonance "levels." Quantification of a subjective feeling may not seem like a worthwhile venture, and without proper context, it may not be. To the extent that empiricizing dissonance does not aid understanding, analysis, or composition on some level of specific musical works, it is of course a futile venture. Vague theoretical studies, such as those of Danner (1985), offer little insight to development of a post-Helmholtzian understanding of the musical continuum between consonance and dissonance.

History offers many examples of individuals who attempted to quantize aspects of music into discrete representations. Consider, for example, the most basic of these—the notion of scale—a notion that virtually all musical cultures of the world share. The scale, in its role as a building block of compositional structure, has been extraordinarily useful over centuries of development, perhaps largely because in it exists a kind of measurable, specifiable, certainty: the fundamental frequency of A4 is 440 Hz, and that of A-sharp4 is 440 Hz multiplied by the twelfth root of two, and so on, for example. Once we have defined a scale in which to compose, it functions as a collection of basis elements onto which notes are projected to produce musical sound. The same can be said for rhythm.

At one time, quantifying dissonance levels of dyads and chords served a useful compositional and pedagogical purpose, particularly in training new composers to follow the rules of the establishment. The rules of Western counterpoint rely heavily on predefined dissonance levels of all possible dyads (and, by inference, chords of arbitrary constitution). As such, these dissonance levels served as a kind of "rule book" that informed many compositional choices (q.v. Fludd's *Temple of Music*).

Of course, modernism threw convention out the compositional window, and post-modernism restructured convention, effectively "emancipating dissonance"—or at least turning our conceptions of it upside down. When one is permuting tone rows, the quantification of consonance and dissonance is perhaps the last thing on one's mind. So why do we still care about dissonance, and in particular, ranking it in some meaningful way?

I submit that the modern-day study of empirical dissonance is useful in three areas: as a compositional control structure, analogous to Wessel's treatment of timbre space as a musical control structure (1979); as an insight into an automated music analysis tool, particularly of non-notatable music; and as part of a broader music-classifier and representational system. In this first capacity, a well-defined approach to dissonance has informed the unique compositional ideologies of composers like James Tenney. A wealth of interesting music has been written that responds in some way, either directly or indirectly, to a dissonance metric.

The usefulness of automated musical analysis is often debated, since the results of listening to a piece of music are experienced uniquely by each individual. However, preliminary results of other forms of automated musical analysis have proven fruitful in solving many problems associated with automated music transcription and harmonic analysis, for example. Automated dissonance analysis may form a small part in the future of such systems. his

The concept of computational dissonance analysis may inform part of a larger automated musical analysis system, for example using the scheme shown in

Figure 4–1. Here, on the left part of the figure, the input audio file is computationally dissected to produce some form of symbolic representation (still a very difficult task, admittedly). This symbolic representation is then analyzed harmonically to inform the rule-based, symbolic musical analysis of the target dissonance model of the audio file. On the right part of the figure, the input audio file is analyzed computationally to extract feature vectors deemed important in the construction of a computational dissonance model. These data inform the computational aspect of the target dissonance model. The combination of the strictly symbolic and the strictly computational would indeed result in a robust analysis model.

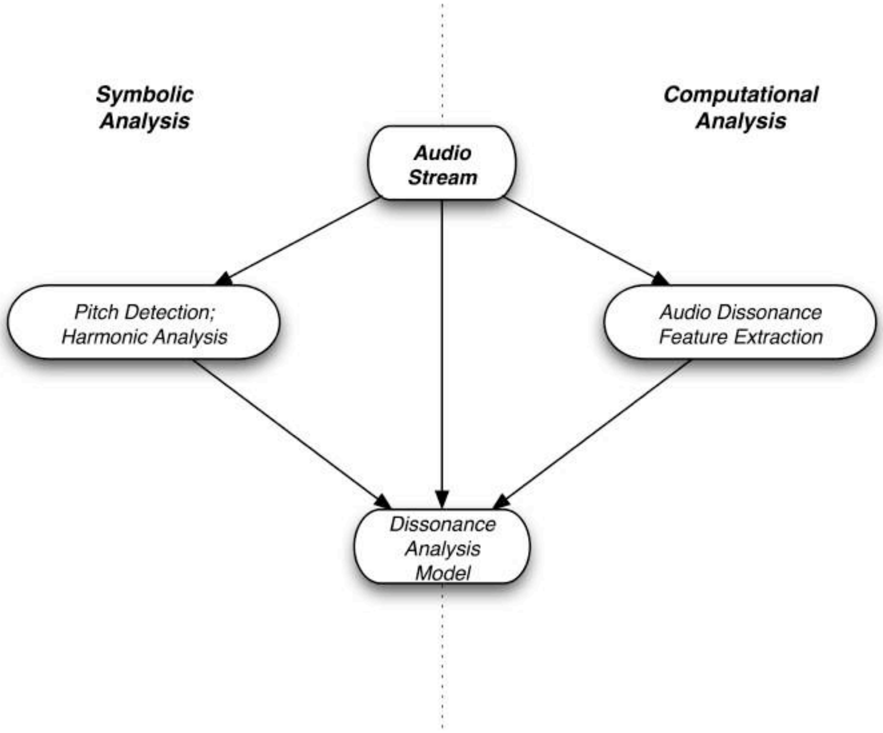


Figure 4–1. Symbolic+computational dissonance analysis model.

Finally, quantizing aspects of dissonance perception in music can lead to several relevant feature vectors for music classification. For example, genre identification in music could be improved by incorporating aspects of dissonance analysis into machine-listening algorithms. Further study will surely lead to new insights in these directions.

Consonance and dissonance are but two related aspects of musical experience, clearly difficult to define, let alone quantify. Therein lies the dissonance of empiricism.

❧

And so dissonance is what it is. Consonance is what it is.

5 DISSONANCE OF SOUND OBJECTS

> *Today, music, as it becomes continually more complicated, strives to amalgamate the most dissonant, strange, and harsh sounds. In this way, we come ever closer to* noise-sound.
>
> *—Luigi Russolo, "The Art of Noises" (1913)*

Before formulating a prototype theory of sound-object dissonance, it is necessary to first define the notion of the sound object formally while considering a variety of writings on the subject. To this end, we here discuss four prominent theoretical treatments of sound objects, drawing on the work on Pierre Schaeffer, Denis Smalley, R. Murray Schafer, and Trevor Wishart. The chapter continues by hypothesizing potential contributing acoustical and perceptual factors that may characterize sound-object dissonance. We then lay the groundwork for conducting specific listening tests in the next chapter to judge subjects' categorization of the relative dissonance of sound objects, from which the accuracy of the proposed theory of sound-object dissonance can be judged.

5.1 L'Objet Sonore

The idea of the sound object—*l'objet sonore*—is crucial to any complete compositional or analytical system that addresses electroacoustic music, because the amalgamation, juxtaposition, and linear unfolding in time of sound objects is central to so much of the literature. To the extent that sound and timbre have served and continue to serve as central organizing principles and control structures for electroacoustic music, an examination of both the nature of the sound object and the contributing factors to acoustical and psychoacoustic dissonance as a means of creating tension and release is important.

Definitions of "sound object" abound; each one tends to define the concept relative to one or more physical and/or cognitive-correlative features. Aesthetic inquiry of the concept, however, traces roots to the Italian Futurist painter Luigi Russolo (1885–1947) and French radio engineer/composer Pierre Schaeffer (1910–1995). Russolo's *L'Arte dei Rumori* ("The Art of Noises," written in March 1913 and published in Russolo 1916) calls for the creation of a new kind of music, composed entirely of sounds for their own sake; to do so, we "must break out of this limited circle of sounds and conquer the infinite variety of noise-sounds." For Russolo, this music was necessitated by both the acoustical impositions of modern, industrialized society upon the silence of nature, as well as by the corresponding attempt to satiate the increasing desire of culture toward dissonance:

> *At first, the art of music sought purity, limpidity, and sweetness of sound. Then, different sounds were amalgamated, care being taken,*

> *however, to caress the ear with gentle harmonies. Today, music, as it becomes continually more complicated, strives to amalgamate the most dissonant, strange, and harsh sounds. In this way, we come ever closer to* noise-sound.

Although he did not explicitly discuss the notion of the sound object, Russolo classified "families of noises"—irrespective to some extent of their means of production. This foreshadowed a more specific theory on sound objects offered several decades later by Pierre Schaeffer.

For Schaeffer, the sound object is a philosophical construct, one that is definable only within the confines the phenomenological philosophy espoused by Edmund Husserl (1859–1938). Indeed, Schaeffer's *objet sonore*, a concept examined at length in two books (Schaeffer 1966, 1967) and a pedagogical set of recordings made with Guy Reibel, Beatriz Ferreyra, and Henri Chiarucci (1967), is a phenomenological object itself, an abstract entity divorced from any hint of (or at least attention to) reference, utility, or identification. It is thus distinct from the *sound event*, a phrase more recently used by Truax (1999) and others in the context of acoustic ecology to describe sound not in terms of its abstract characteristics but rather explicitly in terms of its signification and semantic importance. We could say that a "whoosh" is a sound object, irrespective of its means of production, but the sound of seven rifles firing simultaneously three times is a sound event produced by firearms with a non-musical meaning.

The sound object is itself a transcendental-phenomenological object (Kane 2005), perceived only through the course of *écoute réduite* ("reduced listening," one of Schaeffer's four basic modes of listening). Thus, for Schaeffer, the sound object is inexorably married to its means of receptivity, perceivable only in the course of focused listening to sound for its own sake, divorced from its means of production. Kane (2005) effectively summarizes Schaeffer's definition from his 1966 *Traité des objets musicaux* by stating, "Sound, holding itself at the threshold of the transcendental-phenomenological reduction, asserting no claim about the exterior world, and maintaining its stubborn integrity in the face of occultation by signification, is *l'objet sonore*."

Furthermore, central to Schaeffer's treatise is the understanding of sound objects as they relate to perception. To be recognized phenomenologically as an object per se, a sound must by definition be perceived as a sonic gestalt. And as such, it must retain its ability to be perceived as a gestalt upon repeating listenings, regardless of context or method of production.

The sound object, like James Tenney's classic definition of timbre, is also definable in terms of what it is not. Landy (2005) paraphrases Chion (1983) by noting the following:

> *Schaeffer suggests that there is some confusion concerning the notion whilst adding: a) The sound object is not the sound body, b) The sound object is not the physical signal, c) The sound object is not a recorded fragment, d) The sound object is not a notated symbol on a score, e) The sound object is not a state of mind (it remains the same across different listening modes).*

More recent definitions of the sound object involve the time frame in which a sound event occurs (e.g., Roads 2001), an idea that reinforces the connectivity between phenomenology and cognition. Such definitions may be particularly apropos, because the mechanisms of cognition themselves are a function of the time scale on which attention is focused. (Recall Morton Feldman's famous observation often quoted from his Universal Edition brochure: "Up to one hour you think about form, but after an hour and a half it's scale.") Furthermore, perceptual (subjective) time is different than clock time (see Kramer 1988, for example). Furthermore, one recent neurological study found that sequences of tones played within a 240 msec timeframe are much more likely to be "bound together into a single acoustic event " (Atienza et al. 2003). In this light, Roads (2001) codifies nine time scales on which audio events can unfold, from the infinite to the infinitesimal in duration. He defines the sound object as sound that occurs between the *meso* and *micro* time scales, occupying from a small fraction of a second to several seconds in duration. For Roads, the meso scale refers to "Divisions of form," such as sections and phrases, while the micro scale is occupied by "sound particles" that approach the lower limits of human temporal perception.

Several objections to this basic theory of Schaeffer have been raised. Perhaps the most powerful of these is raised again by Kane (2005), who notes that "it still remains unclear what exactly it means, experientially, to perceive a sound-as-such." Whether we are cognitively able to disassociate any and all sounds from past experience, cultural signification, and semiotic representation requires a leap of faith, particularly from the standpoint of evolutionary biology, which suggests that we dissect all incoming sounds first for fight-or-flight response.

Kane also cites the objection to phenomenological reduction as a basis for electroacoustic music as postulated by Trevor Wishart. For example, Wishart (1985) argues that, to be engaging to listeners, electroacoustic music must consider the primacy of gesture and the difficulty of its effective dissociation from sonic result. Furthermore, it is impossible across cultures to escape the psychological dominance of the vocal utterance as a sound, irrespective of cultural affiliation. That is, most people find it difficult to hear the human voice as sound for sound's sake; we are physiologically and psychologically programmed otherwise. (One need only look to the Fletcher-Munson equal-loudness contours or their more recent modifications, for example, to witness this, as we are most sensitive to frequencies in the frequency range of the human voice.)

For the purposes of the present discussion, let us define the sound object as a relatively short basic musical entity, divorced from any surrounding physical or musical context, unambiguously perceived as a single unit upon multiple hearings by a variety of listeners. Central to this definition is the consideration of the sound object's perceptual "objectness." That is, apart from its consideration as a single perceptual unit (object), it transforms from "sound object" to more generally "sound." Just as the traditional musical note is unique in its "objectness" or "noteness," so too is the more general concept of the sound object.

A sound object may comprise multiple sounds from a variety of natural and/or synthetic sources, but the object must be generally taken as a single sonic gestalt. It is the perceptual encapsulation of timbre—a kind of timbral unit. In other words, the concepts of note and local gesture within the context of traditionally notated music are reflected in notion of the sound object in electronic music (or any kind of abstract structuring of sounds). The sound object is created from the intersection of contributions from pitch, rhythm, and frequency content (loosely, “timbre”), as shown in Figure 5–1. A diagram of this sort may seem too simplistic, but note how it naturally incorporates a variety of musical and psychoacoustic phenomena that fall into the intersections among pitch, rhythm, and timbre. For example, consider the linkage between tuning, pitch, and timbre (Sethares 1993a, 1993b, 1998, 1999). Playing a particular chord tuned in, say, twelve-tone equal temperament might allow or even encourage trained listeners to separate the constituent pitches of the chord, while playing the same chord in a certain just-intonation scale may encourage some listeners to perceive the entire chord as a single timbre (i.e., enhance spectral fusion), owing to the potential overlap of harmonics of each of the constituent pitches of the chord (Sethares 1999).

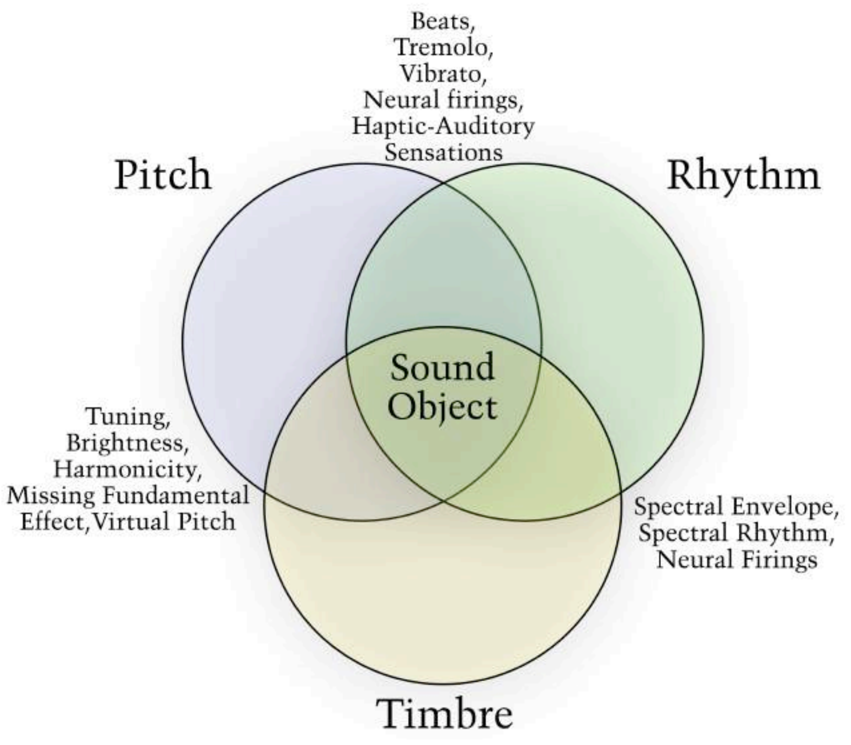


Figure 5–1. The abstract concept of the sound object is tied to the intersection of pitch, rhythm, and timbre, in which the totality of each is taken as a single perceptual unit.

In the context of computer-based analysis, the abstract *objet sonore* is often considered to be sound file containing digital audio samples: to a computer, a sound file is a data object, divorced from its method of acoustical production[i]. However, unlike other theories of sound objects, I contend that the consideration of a sound divorced from the space in which it occurs—whether synthetic or

---

[i] (Later, we distinguish among the abstract sound object, the "real" sound object, and this, the "stored" sound object.)

actual—fails to consider the totality of the object. This is because our definition here of the sound object is inexorably linked to human perception, and the perception of gestalt can significantly change based on the location and space in which the sound occurs. Consider, for example, that extensive reverberation may tend to blur sound objects together into a single perceivable unit. The same sound objects played with no reverberation added might be perceived as separate objects. Thus, consideration of sound objects within the confines of the space in which they actually occur (or are artificially made to occur) necessarily elevates their status to "real" (in the Lacanian sense) sound objects.

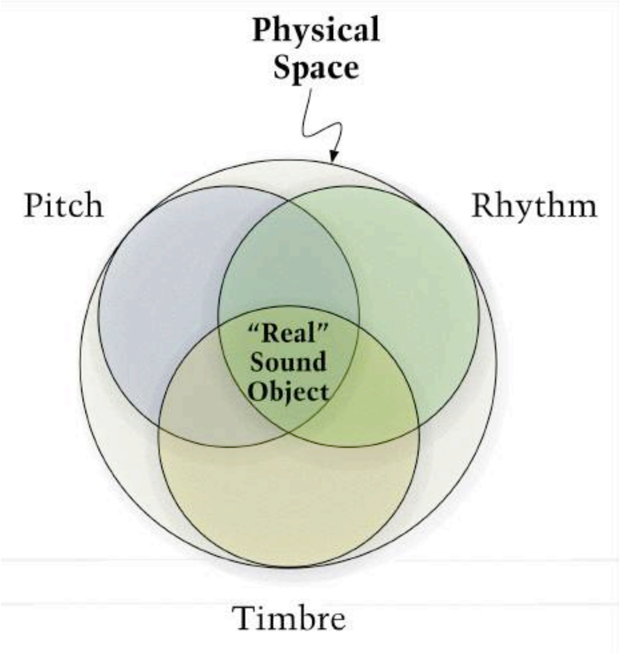


Figure 5–2. Placing a sound within an acoustical environment forces an abstract sound image into a concrete, "real" sound object.

A basic thesis of this dissertation is that a specific examination of computer and human approaches to the objectification of sound-object dissonance may at some point provide insights into a more general analytical tool for music that exists apart from any concrete representation, such as standard musical notation or graphical performance instruction. By leveraging knowledge gained from subjective dissonance ratings of sound objects, new analytical insights may be gained into electro-acoustic music.

Corollary to the variety of aesthetic premises on which considerations of the sound object are based, multiple categorizations of sound objects have been proposed, reflecting the bias of the one who proposed it. We will now consider several of these before settling on one to use as a rough framework for listening tests in the next chapter.

5.2 Taxonomies of Sound (Objects)

The twentieth century saw several attempts to classify sound. These attempts took one of two forms: aesthetic, musically oriented taxonomies, or scientific, quantifiable taxonomies. We will examine each of these major taxonomies in roughly chronological order.

Perhaps the earliest important classification of sounds, at least from a musical perspective, is that of Luigi Russolo. In his 1913 manifesto, he outlines "six families of noises of the Futurist orchestra," shown in Table 5–1. (These are taken from Barclay Brown's 1986 translation, with the exception that Brown's translation of the word *Ronzii* as "Humming" has been changed to "Buzzing" (literally, "buzzes") to avoid overlap with human sounds in the sixth column.)

1	2	3	4	5	6
Roars Thunderings Explosions Hissing roars Bangs Booms	Whistling Hissing Puffing	Whispers Murmurs Mumbling Muttering Gurgling	Screeching Creaking Rustling Buzzing Crackling Rubbing	Noises obtained by beating on metals woods skins stones pottery etc.	Voices of animals and people Shouts Screams Shrieks Wails Hoots Howls Death rattles Sobs

Table 5–1. Russolo's "six families of noises."

Russolo's sound families are organized according to a mix of their sonic properties as well as their means of production. Casually, we could describe Group 1 as "big sounds," Group 2 as "bright, soft sounds," Group 3 as "dark, soft sounds," and Group 4 as "bright, loud sounds." On the other hand, the last two groups are organized according to their method of production. These sounds were grouped to enable the development of the "noise orchestra" by which a variety of nontraditional timbres that embraced the sounds of modern society could be explored.

About two decades after Russolo's manifesto, Pierre Schaeffer experimented with composing a music made from the grouping of sound objects in different ways. From the beginning, he acknowledged the primacy of the human voice, dividing sound objects for his collaborative work with Pierre Henry *Symphonie pour un Homme Seul* (1950) into two groups: human sounds and non-human sounds. But ultimately, Schaeffer's philosophical bent toward phenomenology led him to dissociate the method of sound production from the acoustical result, writing the following in 1957:

> *All call into question the notion of the instrument. Sound can no longer be characterized by its causal element, it has to be characterized by the effect only. Hence it must be classed according to its particular morphology, rather than according to instrumental provenance. It must be considered in itself. The best proof of this: once the most interesting sonorities produced by the new techniques have been recorded on tape, it is impossible to say how, and by what ensemble of procedures or instruments, they have been produced.*[ii]

An important aspect of Schaefferian theory lies in a comprehensive approach to the examination, classification, and musical use of sound objects—sounds cut and isolated from their surrounding musical context, acousmatically shrouding their means of production—which he divides into four broad areas: typology, morphology, characterology, and analysis/synthesis. Of these, the first two stages are concerned with the classification (taxonomy) of sounds; the final two are concerned with their musical use.

[ii] The quotation and translation here is taken from Palombini (1993).

Schaeffer's typology concerns the broad classification of a sound's "type" according to "mass/facture," "duration/variation," and "equilibrium/originality," which we can briefly summarize as the global aspects of the sound's spectral features and duration. Typological classification organizes sounds according to basic, primordial types (of which Schaeffer defines around thirty.) Sounds can be grouped typologically into "Balanced Objects," "Redundant Sounds," "Eccentric Sounds," and "Varying Sounds."

Second, morphology examines in greater detail a sound's evolution over time. Morphologically, sound objects can be described in terms of "Matter," "Shape," and "Variation." Schaeffer decomposes matter into three elements: (1) mass (representing the object's location along the continuum of "pitchedness" to "noiseiness"); (2) harmonic timbre (loosely, the "brightness" or "darkness" of an object); and (3) grain (or micro-level sonic properties). Next, the shape of a sound object is described by its dynamics (amplitude envelope) and its "allure" (frequency modulation). Finally, Schaeffer describes the variation of a sound object in terms of its "melodic profile" and "mass profile," which refer to the object's pitch evolution and position along the "pitchedness"-"noisiness" continuum. Taken together with typological descriptors, the morphological properties of matter, shape, and variation properties are used to construct a working solfège of the sound object, a system that has been expanded in Dennis Smalley's more recent writings on spectromorphology (e.g., Smalley 1986).

The transition from the examination of typology and morphology to that of characterology and analysis/synthesis crosses the schism from decomposition to re-composition—from dissection and taxonomy to grouping and music composition. Characterology attempts to group sounds into "genres" or "families," for example "bell-like" sounds or "metallic sounds," a research problem explored much later using computers to model psychoacoustic timbral similarity (Wessel 1973; Grey 1977; Toiviainen, Kaipainen, and Louhivuori 1995; Aucouturier and Pachet 2004a, 2004b). And because this process can lead to the grouping of timbrally similar sounds produced by a variety of physical methods (e.g., the sustained portion of both a long flute note and that of a piano, which many listeners confound), Schaeffer's original phenomenological premise that sounds should be taken naïvely as objects apart from reference to their production or signification.

Finally, the analysis/synthesis stage addresses the assemblage of classified and grouped sound objects and the creation of an intelligent musical fabric. Through the analysis process, sounds belonging in the same family can be analyzed for their suitability to create "scales" for later musical use (synthesis) according to variation in a primary psychoacoustically perceivable features. This allows the potential to create a musical syntax based on timbre in general instead of pitch and rhythm, which was of course Schaeffer's intended goal.

Still other and more modern theories and taxonomies of the sound object exist, including those of Dennis Smalley (1986) and Trevor Wishart (1985). (We will discuss these in some detail in the next section.) The relatively young field of acoustic ecology, spearheaded in many respects by composer R. Murray Schafer, asserts a broader organization of sounds according to a variety of means. Schafer's taxonomy, set forth in his 1977 book *The Tuning of the World*, groups sounds variously according to the method of production; source of production (human, animal, nature, etc.); geographical location of occurrence; cultural signification and semiotic value; intended use; and perceptual factors such as loudness. For reference, Schafer’s taxonomy is outlined below in Table 5–2.

I. Natural Sounds
  a. Sounds of creation
  b. Sounds of apocalypse
  c. Sounds of water
    1. Oceans, seas, and lakes
    2. Rivers and brooks
    3. Rain
    4. Ice and snow
    5. Steam
    6. Fountains, etc.
  d. Sounds of air
    1. Wind
    2. Storms and hurricanes
    3. Breezes
    4. Thunder and lightning, etc.
  e. Sounds of earth
    1. Earthquakes
    2. Landslides and avalanches
    3. Mines
    4. Caves and tunnels
    5. Rocks and stones
    6. Other subterranean vibrations

    7. Trees
    8. Other vegetation

f. Sounds of fire
    1. Large conflagrations
    2. Volcanoes
    3. Hearth and camp fires
    4. Matches and lighters
    5. Candles
    6. Gas lamps
    7. Oil lamps
    8. Torches
    9. Festival or ritual fires

g. Sounds of birds
    1. Sparrow
    2. Pigeon
    3. Killdeer
    4. Hen
    5. Owl
    6. Lark, etc.

h. Sounds of animals
    1. Horses
    2. Cattle
    3. Sheep
    4. Dogs
    5. Cats

        6. Wolves
        7. Gophers, etc.
    i. Sounds of insects
        1. Flies
        2. Mosquitoes
        3. Bees
        4. Crickets
        5. Cicadas, etc.
    j. Sounds of fish and sea creatures
        1. Whales
        2. Purpoises
        3. Turtles, etc.
    k. Sounds of seasons
        1. Spring
        2. Summer
        3. Fall
        4. Winter

II. Human Sounds
    a. Sounds of the voice
        1. Speaking
        2. Calling
        3. Whispering
        4. Crying
        5. Screaming
        6. Singing

        7. Humming
        8. Laughing
        9. Coughing
        10. Grunting
        11. Groaning, etc.
    - b. Sounds of the body
        1. Heartbeat
        2. Breathing
        3. Footsteps
        4. Hands (Clapping, Scratching, etc.)
        5. Eating
        6. Drinking
        7. Evacuating
        8. Lovemaking
        9. Nervous System
        10. Dream Sounds, etc.
    - c. Sounds of clothing
        1. Clothing
        2. Pipe
        3. Jewelry, etc.

- III. Sounds and Society
    - a. General descriptions of rural soundscapes
        1. Britain and Europe
        2. North America
        3. Latin and South America

    4. Middle East
    5. Africa
    6. Central Asia
    7. Far East

b. Town soundscapes
    1. Britain and Europe, etc.

c. City soundscapes
    1. Britain and Europe, etc.

d. Maritime soundscapes
    1. Ships
    2. Boats
    3. Ports
    4. Shoreline, etc.

e. Domestic soundscapes
    1. Kitchen
    2. Living room and hearth
    3. Dining room
    4. Bedroom
    5. Toilets
    6. Doors
    7. Windows and Shutters, etc.

f. Sounds of trades, professions, and livelihoods
    1. Blacksmith
    2. Miller
    3. Carpenter

    4. Tinsmith, etc.

g. Sounds of factories and offices

    1. Shipyard
    2. Sawmill
    3. Bank
    4. Newspaper

h. Sounds of entertainments

    1. Sports events
    2. Radio and television
    3. Theater
    4. Opera, etc.

i. Music

    1. Musical instruments
    2. Street music
    3. House music
    4. Bands and orchestras, etc.

j. Ceremonies and festivals

    1. Music
    2. Fireworks
    3. Parades, etc.

k. Parks and gardens

    1. Fountains
    2. Concerts
    3. Birds, etc.

l. Religious festivals

        1. Ancient Greek
        2. Byzantine
        3. Roman Catholic
        4. Tibetan, etc.

IV. Mechanical Sounds

- a. Machines (general descriptions)
- b. Industrial and factory equipment (general descriptions)
- c. Transportation machines (general descriptions)
- d. Warfare machines (general descriptions)
- e. Trains and trollies
    1. Steam locomotives
    2. Electric locomotives
    3. Diesel locomotives
    4. Shunting and yard sounds
    5. Coach sounds
    6. Street cars, etc.
- f. Internal combustion engines
    1. Automobiles
    2. Trucks
    3. Motorcycles, etc.
- g. Aircraft
    1. Propeller aircraft
    2. Helicopters
    3. Jets
    4. Rockets, etc.

- h. Construction and demolition equipment
    - 1. Compressors
    - 2. Jackhammers
    - 3. Drills
    - 4. Bulldozers
    - 5. Pile drivers, etc.
- i. Mechanical tools
    - 1. Saws
    - 2. Planes
    - 3. Sanders, etc.
- j. Ventilators and air conditioners
- k. Instruments of war and destruction
- l. Farm machinery
    - 1. Threshing machines
    - 2. Binders
    - 3. Tractors
    - 4. Combines, etc.

V. Quiet and Silence

VI. Sounds as indicators

- a. Bells and gongs
    - 1. Church
    - 2. Clock
    - 3. Animal, etc.
- b. Horns and whistles
    - 1. Traffic

    2. Boats
    3. Trains
    4. Factory, etc.
- c. Sounds of time
    1. Clocks
    2. Watches
    3. Curfew
    4. Watchmen, etc.
- d. Telephones
- e. (Other) warning systems
- f. (Other) signals of pleasure
- g. Indicators of future occurrences

Table 5–2. R. Murray Schafer's sound-object taxonomy.

As thorough as this list seems to be, objections are easily raised. A primary objection is that the taxonomy seems to exhibit redundancy and overlap in its classifications. For example, sounds of the season Spring, mentioned in Section I.k.1, might coincide with the classification of a human voice speaking during Spring, outlined in Season II.a.1. Or perhaps "Opera," mentioned in III.h.4, could just as well have been listed under "Music" in III.i. The criticism is answered, however, by noting that sounds should be classified under this taxonomy according to the environmental frame of focus by which they are intended to be classified. For example, a human voice speaking during Spring should be classified under II.a.1 (human speech) rather than I.k.1 (sounds of the Spring season) if the sounds “speechness” is more important than its “Springness.” Said another way, the classification of sounds is in some ways a rather personal matter, and the availability of overlapping classification categories represents simply a means by which the classifier can emphasize a single desired property of a sound.

The problem of classifying sounds has been reinvigorated in recent years owing to the field of multimedia content retrieval and the power of modern computers. The categorization of sound objects according to the methodology of Schaeffer’s *Solfège*, for example, has been automated in the work of Dack (1999). Various systems for classifying sounds from the standpoint of music information retrieval that rely on signal-processing techniques have been proposed (e.g., Foote 1997, 1999; Zhang and Kuo 1998; Tzanetakis and Cook 2000; Downie 2003). New systems for indexing sounds have been suggested that use a variety of techniques, for example MPEG-7 content descriptors (Herrera , Serra, and Peeters 1999; Philippe 2000; Casey 2001; Kostek and Czyzewski 2001; Herre 2003; Gómez et al. 2003; Cano et al. 2004), onomatopoeia or more general descriptive adjectives (von Bismarck 1974a; von Bismarck 1974b), semantic tags that represent information about the creation of the sound (e.g., where and when it was recorded or synthesized, what kind of microphone was used, etc.). However, some have criticized the concept of semantic descriptors, as noted by Cano et al. (2004), because descriptions about the recording or production of a sound do not necessarily reflect anything important or intrinsic about the sound itself.

Although the present essay addresses the dissonance of sound objects in isolation, it is here worth pausing to consider in some detail the variety of theoretical and aesthetic treatments of the sound object in the context of the electroacoustic music composition. In so doing, we provide a context for the proposed dissonance theory of the sound object.

5.3 Four Theories of Objects in Electroacoustic Composition

Among the most prominent musical developments of the twentieth century was most certainly the elevation of the role of timbre in the traditional Western hierarchy of musical parameters, made possible in part by the advent of recorded sound. This has allowed (and perhaps required) the development of new theories of music based almost exclusively on timbre and of the sound object itself, for the sound object by definition represents the very encapsulation of a perceptual timbral unit. The compositions and theories of Pierre Schaeffer, R. Murray Schafer, Trevor Wishart, and Denis Smalley question our previously-held notions of what constitutes music, the act of composition, the composer-performer-audience relationship, and the role technology should assume in music and musical institutions in the twentieth century. I briefly summarize and discuss the theories of Schaeffer, Schafer, Wishart, and Smalley, and then compare and contrast them with one another in this section, before continuing our discussion of quantitative characteristics of sound objects in the next section.

In 1951, the Radiodiffusion-Télévision Française (RTF) chartered the Groupe de Recherches de Musique Concrète, a group consisting of Pierre Schaeffer, the composer Pierre Henry, and the sound engineer Jaques Poullin. Schaeffer had already been studying and experimenting with what was known as *musique concrète* as early as 1949, and the Groupe de Recherches de Musique Concrète, the first designed electronic music studio (Palombini 1993a, p. 542)—later known as the Groupe de Recherches Musicales—provided an environment where further work in this area could be carried out under the auspices of the RTF.

With his earliest piece of musique conrète, *Cinq études de bruits* (1948), Schaeffer had begun to formulate a primitive theory of *conrète* composition in which the main goal was source decontextualization—that is, the disembodiment of sound through dissection, separating the cause of the sound from its sonic effect. He achieved this goal from the onset through two main methods, as he himself notes in *À la recherche d'une musique concrete*:

> To distinguish *an element (to hear it in itself, for the sake of its texture, its matter, its colour).*
>
> To repeat it. *Repeat the same sonic fragment: there is not an event any more, there is music.*

In making such a radical break with musical tradition, Schaeffer, a self-described musical "anarchist" (Hodgkinson 1987), found himself swimming in a sea of infinite possibilities, and he began to espouse the value of experimentation above all else. This fact is clearly evidenced in the following dialogue:

> *M. Pierret:— Can we go as far as saying that, if you wrote today either* la Coquille à planètes *or* Orphée, *then you'd show more care*

> *for the œuvre, you'd no longer be the victim of your own experiments…*
>
> *P. Schaeffer:— Certainly not! We always commit the same mistakes again, and 'je ne regrette rien'! I tell you: I prefer an experiment, even aborted, to a successful œuvre.* (Pierret 1969, p. 105, translated in Palombini 1993a, p. 542)

Schaeffer further advocates experimentation for its intrinsic educational value. He seems to be purporting (and rightly so) that if we are going to create a music based solely on timbre, we need to study sound and timbre as much as possible. Clearly, compositional techniques traditionally applied to pitch-based material are not generally applicable to non-pitch-based materials, and thus we are forced to confront this new music with new ears. As Palombini (1993a) writes, "[N]ot only a new instrumental apprenticeship is necessary: the apprenticeship of sonority itself is imposed. The choice is therefore between using concrete material to create œuvres and doing research into sonority to discover musicality" (pp. 547–548). We must also listen with new ears, "whence the idea of a sol-fa of the sound object to train the ear to listen in a new way; this requires that the conventional listening habits imparted by education first be unlearned" (Schaeffer and Reibel 1967).

In promoting experimentation over prolificacy, Schaeffer raises the question of what constitutes a musical composition. In an interview in 1986 (Hodgkinson 1987), he remarked that in retrospect he does not think of himself, or anyone else who wrote musique concrète, as a composer, because

> *Musique Concrete in its work of assembling sound, produces sound-works, sound-structures, but not music. We have to not call music things which are simply sound structures.*

He continues:

> *There are many people working with sound. It's often boring, but not necessarily ugly. It contains dynamic and kinaesthetic impressions. But it's not music.*

An irony in the thinking of Schaeffer becomes apparent as his revolutionary ideas about the sound object and the notion of musique concrète clash with a traditionalist view of the constitution of music, musicality, and composition.

For Schaeffer, musical value is inexorably tied to its use in the context of a system and the idea of *reproducibility*—the notion that music can be performed and realized by different people at different times in different venues in different ways. When asked what he thought constituted musical value, Schaeffer responded:

> *The best analogy is with language—since we talk of musical languages. People who share the same language, French or Chinese or whatever, have the same vocal chords and emit sounds which are basically the same, as they come from the same throats and lungs. So*

> *this is a sound world. But the same sounds have linguistic values and this makes them different. These linguistic values derive from their role within a system. In the same way, musical value is inseparable from the idea of system.*

This seems to validate the aesthetic legitimacy of musique concrète provided the treatment is in some way systematic. For that matter, it seems to validate *any* music provided the treatment operates within a system or musical language. But not so; even atonality and serialism per se are suspect:

> *In so far as atonality for instance presented only a destructive face, pretending to organize the twelve tones in ignorance of their degree quality, and considering them solely as terms of an algebraic permutation, one could be shocked by so premature a denial of a tradition that I shall call—no pun intended—dominant.* (Schaeffer 1957, trans. by and quoted in Palombini 1993a, p. 544)

Palombini (1993b) summarizes Schaeffer's view on serialism thus:

> *In principle, but not in practice, it is unacceptable to apply serialism to traditional musical material. In principle, but not in practice, it is acceptable to apply serialism to concrete material.*

The distinction seems to be that the application of serial procedures to traditionally notated music results in an abstract music in which the row and system are generally not heard anyway. Concrete music does not need the application of serialism to achieve this level of abstraction, although "it was perhaps useful to put the straitjacket on these new materials for a year or two, so as to demonstrate at least the possibility of submitting them to construction" (quoted in Palombini 1993a). (He added: "Why twelve notes when electronic music has introduced so many more?")

Schaeffer's traditionalist tendencies are evidenced in his pessimistic outlook on the future of music. To him, the best music has already been composed, and we can create a new music only when we "realize that there's no way out of traditional music" and "get down to a baroque music for the 21st century" (Hodgkinson 1987). He observes:

> *each time I was to experience the disappointment of not arriving at music. I couldn't get to music—what I call music. I think of myself as an explorer struggling to find a way through in the far north, but I wasn't finding a way through.... There is no way through. The way through is behind us.*

Reflecting on his life in 1986 in the same interview (Hodgkinson 1992), he wrote:

> *So these were the three circumstances that compelled me to experiment in music: I was involved in music; I was working with turntables (then with tape-recorders); I was horrified by modern 12-tone music. I said to myself, 'Maybe I can find something*

> *different…maybe salvation, liberation, is possible.' Seeing that no-one knew what to do anymore with DoReMi, maybe we had to look outside that… Unfortunately it took me forty years to conclude that nothing is possible outside DoReMi… In other words, I wasted my life.*

❧

Like Pierre Schaeffer, the Canadian composer R. Murray Schafer admonishes us to "clean our ears" to appreciate the totality of the world of sound. Schafer challenges our notions of instrument, performer, composer, and audience in both his music and his numerous books and articles, many of which are didactic in nature.

One of Schafer's primary concerns lies in reestablishing in our consciousness the idea of the musician as a "full creative human being and not merely a technician and a virtuoso repeater of past practices and received interpreted 'truths'" (Coleman 1994). As an author and pedagogue, he emphasizes learning by doing—learning about music by making it, much in the manner that Zoltán Kodály (1882–1967) and Carl Orff (1895–1982) did earlier in this century. He adamantly attempts to break the myth of the composer as a member of the musical elect by involving all of his students in the process of composition, for example, by asking them to extemporize a setting of a text simply by using sounds from their own voices or available instruments or by having students make as many sounds as possible with a piece of paper (Schafer 1967). His use of aleatory and nonstandard graphic notation in works such as

Epitaph for Moonlight (1968) for youth choir and optional unspecified bells blurs the divisions among composer, conductor, and performer.

For Schafer, however, the distinction between "music" and "sonic construction" is not only unclear, but, more importantly, irrelevant. In 1972, he founded the World Soundscape Project "for the purpose of exploring the relationship between people and their acoustic world" (Slonimsky 1997). He notes that certain other cultures do not even posses the word "music," and that the "origins of this concept of music owes much to the transition from outdoor to indoor living" (Schafer 1992). The Western notion of music as an abstract entity requiring intense concentration and focus has necessitated vast indoor concert halls and rooms strictly for the presentation of musical compositions, and, in the process, taken music away from the masses and placed it in the hands of an elite few.

Schafer expounds a theory of musical mimetics in which music reacts *to* nature and our surroundings. For example, when discussing an 1864 bill that had been passed to ban street music in London, he notes

> *The street had now become the home of non-music, where it mixed with other kinds of sound-swill and sewage. From now on chamber music and street noise would develop obversely: the more intricate the one became, the cruder the other seemed.* (Schafer 1992)

The ubiquity of noise in our environment has had other several notable effects. Nostalgia itself has developed as a musical tool, and the concert hall is a "virtual space" in which we can hear the forgotten sounds of nature in

Beethoven's Sixth Symphony or Respighi's *Pines of Rome*. There has been a yet more directly observable effect, however:

> *The frequency range of the music is another unconscious initiation of the external soundscape. Mozart's music is made up of mid- and high-frequency sounds as was his world, whereas the heavy infrasound of the modern city is reproduced in the guitars of the modern rock group.* (Schafer 1992)

The overall volume level of both music and the soundscape has also increased, as he notes in *The Tuning of the World* (1977, p. 116).

Another element at the heart of Schafer's theory is a reconciliation of modernist and post-modernist tendencies which have bitterly divided composers throughout the twentieth century (Coleman 1994). He is concerned with new musical notation, aleatory, and avant-garde musical theater while often employing traditional musical forces and texts in dead languages (which tends to highlight their phonetic nature rather than any implicit meaning (*The New Grove Dictionary of Music and Musicians* 588). Coleman (1994, p. 1189) notes:

> *The multiplicity of levels and potential meanings, the diminishment of hierarchy, the use of crosscultural and transcultural elements, the profusion of information, and the concomitant acceptance of ambiguity attend upon the resulting complexities make these works a fascinating blend of the modern and postmodern, or, as*

> *one might surmise, just about the location that Schafer would no doubt favor for his efforts.*

Schafer ultimately advocates a confluence of elements to create a work, one in which "music, non-music, and silence are woven together artistically and therapeutically to bring about a new consciousness where art and life touch, merge, and are lost in one another" (1992, p. 45). The arbitrary line between a musical composition and soundscape is lost when one considers the beauty of an individual sonic element on its own terms without regard to the artificiality of its construction, appreciating simply the immediacy and physicality of the sonic experience.

❧

Denis Smalley is exclusively an electroacoustic composer, and the corpus of his theoretical writing may be divided into three broad categories: a theory of listening, a theory of musical fields, and a general theory of musical values and the role of technology in music. Central to Smalley's theory of listening, itself a synthesis of other theories, is the idea of a hierarchy of listening modes. We may listen to a work in any of the various modes, which may or may not overlap, and modal shifting occurs during the course of listening to a piece of music "depending on our attention and focus, and on our competence and experience as listeners" (Smalley 1992, p. 517). Smalley's listening theory begins with an overview of the four modes of Pierre Schaeffer, outlined in his 1966 *Traité des Objets Musicaux* and summarized in "The Listening Imagination: Listening in the

Electroacoustic Era" (pp. 515–517). These four modes include information-gathering, in which "we are occupied with the provenance of the sound and the 'message' it carries"; passive reception, in which the listener has no choice or intent in listening; appreciating and responding to attributes of sounds themselves based on "spectro-morphological criteria"; and abstraction of pertinent values.

Smalley also incorporates Schachtel's (1984) psychological theories of autocentricity and allocentricity into his theory of listening:

> *The* autocentric *or subject-centered senses focus on basic responses and feelings of pleasure and displeasure. The emphasis is on subjective reaction to something....*
>
> *The* allocentric *perceptual mode is object-centered in that it involves perceiving something independent of the perceiver's needs.... It is a process of active and selective focusing on an object, being able to discern distinguishing features in a non-partisan way.* (Smalley 1992)

Again, a hierarchy is established in which allocentricity represents a higher level of musical comprehension and concentration than does autocentricity. The dynamic between these two modes of listening, however, is not only inevitable, but it forms a "fundamental part of the listening process" (p. 519).

Smalley's listening theory also codifies the possible relationships between subject and object. Indicative relationships convey information about environmental events, e.g., the sound of gunshot indicates that a gun has just been fired. Reflexive relationships concern autocentric emotional response to a sound and may be active or passive. Finally, interactive relationships exist whenever the subject and object enter a dialogue and the subject continuously explores "the qualities and structure of the object" (p. 520).

In light of his theory of listening, Smalley proposes a theory of indicative musical fields that, once established, attempts to promote the supremacy of electroacoustic music in terms of its ability to explore indicative relationships.

According to Smalley, nine indicative fields exist: gesture, utterance, behavior, energy, motion, object/substance, environment, vision, and space. For Smalley, these fields by definition cannot include compositional models based on "scientific, mathematical, statistical or other theories, regardless of any universal validity…because these models cannot be 'understood' without explanation" (p. 522). Nevermind that a mathematically based compositional object may, through its details, indicate a general and easily-perceived morphological property such as growth, division, multiplicity, or stasis; examples of this abound in the electroacoustic literature, in distinction to Smalley's contention. Smalley apparently maintains that a comprehension of the compositional model in all detail is necessary for it to be part of an indicative field.

Smalley notes that electroacoustic music, unlike instrumental music, exhibits various degrees of surrogacy. A first order surrogate consists of a sample of an instrument in which the sound source is identifiable, while a second order surrogate includes sampled sound spectrally altered in such a way that "vestiges of human gestural activity…are surmised from the sound" (p. 524) but are not easily explained physically. The final stage of abstraction, remote surrogacy, consists of "a state where neither gesture-type nor source can be surmised" (p. 524).

Utterance constitutes an important reflexive element in electroacoustic music, as the human voice "makes utterance intimate and emotionally charged" (p. 525). In Smalley's indicative field theory, utterance always announces a human presence and focuses attention on the musical object that contains the utterance.

Behaviour encompasses the function of sounds in a musical context and is concerned with three dynamics: dominance/subordination, conflict/coexistence, and causality. Causality is of particular importance in electroacoustic music, for causal relationships are not often immediately apparent:

> *This type of causality is surmised rather than known: visual or experiential knowledge cannot verify the relationship or test it by recreating the temporal sequence.* (p. 527)

Causality, when linked to the "fields of gesture, energy, motion, and object/substance…tends to add impetus to the forward motion of musical structure" (p. 527) and may be perceived as an independent indicative musical field when appropriately employed in the context of these other related fields.

Energy and motion fields may be created by various means, including spatialization, diffusion, and spectromorphology. Smalley notes that this field is in constant flux between "compaction and dispersal" (p. 528), and its existence is necessary to create sonic trajectories and spectral textures.

The substance/object field relates to the perception of a sound object as a discrete thing, which may be suggested by any of three means. Substance may be created by reference in some way (directly or indirectly) to other objects that have substance, such as physically perceivable gesture sources (for example, bowing, hitting, or scraping). It may also be created by "types of motion that suggest analogies with the motion of objects", or from any substance for that matter, as long as a "semblance of a plausible gestural origin" (p. 529) is maintained.

The environment field consists of the incorporation of environmental sounds into music and is closely related to the space field. The space field may be discussed in terms of (1) articulation of structure, (2) articulation of composed spatial content via live diffusion, and (3) the listener's experience. In all three areas, ideas regarding composed versus listening space are implied, and the composer must be aware of them.

Smalley's notion of the visual field, perhaps the least well-defined indicative network in his theory, implies a sense of synæsthesia on the part of the listener. He proposes that "music, and electroacoustic music in particular, is not a purely auditory art but more integrated, audio-visual art, albeit that the visual aspect is frequently invisible" (p. 530).

Many musical stalwarts have traditionally objected to concerts of tape music because a performer is not visible. We like to *see* music being made right before our eyes, because the establishment of indicative fields on the part of the composer and their perception on the part of the listener is not as necessary, or at least not as difficult. When I hear an utterance in a concert of acoustic music, I may witness its physical production, and thus a more self-evident sense of causality exists. The establishment of virtual correlates of physical entities is not necessary in acoustic music, for they exist by virtue of their means of production. Not to say that musical metaphor is impossible, but that the notion of surrogacy does not exist.

In light of these nine musical fields, Smalley (1992) frequently asserts that electroacoustic music is in fact particularly well suited to the establishment of indicative fields:

> *Electroacoustic music, through its extensive sounding repertory drawn from the entire sound-field, reveals the richness and depth of indicative relationships more clearly and comprehensively than is possible with other musics.* (p. 521)

And later:

> *The widest possible repertory of motions is possible in electroacoustic music because of the spectro-morphological freedom of the medium which allows both an extensive variety of attitudes towards (dis)continuity and conjunct/disjunct motion, and an unrivaled elasticity of temporal flux.* (p. 528)

In outlining his theories of listening and indicative fields, Smalley at several points indicates his own musical value system and thoughts on the role of technology in music. In order to create a theory of electroacoustic music based on indicative musical fields, we must redefine the relationship between ourselves and our musical implements and proceed with open minds:

> *In the most interesting and original electroacoustic music, the traditional notion of the instrument, so fundamentally linked to causal identity and so intimately tied to human agency via gestural activity, no longer provides such a dominating and fundamental indicative link. It is often not so much a case of stretching (or contracting) the notion of 'instrument' but of discarding it altogether.* (p. 540)

He also notes that

> *The history of contemporary instrumental music (and I would not like to say where that history starts) is bound up with the blurring of the distinctions between harmony and timbre and between pitch and "noise", and the consequent potential for the creation of the continua between pitch and noise, and between pitch and timbre. Once composers started to explore more fully the noise and timbral poles of those continua, the transformation of instrumental spectromorphologies (particularly the spectral aspect) could become a central feature of musical discourse rather than a peripheral feature.* (Smalley 1993, pp. 291–292)

In summary, Smalley astutely observes that “[r]egrettably there is too much electroacoustic music that demonstrates a disdain for listeners’ indicative needs and the spectro-morphological means of achieving them. (1992, p. 551).

❧

Trevor Wishart’s theories of music concern the sonic continuum, the sonic landscape, utterance and extended vocal technique, and the potential of technology to transform musical ideology. His musical manifesto *On Sonic Art* (1985) defines much of this theory. Wishart charts this history of music from the perspective of the “lattice”—the organizational framework upon which its primary elements (pitch, rhythm, and timbre) unfold and are quantized into discrete values. The keyboard, so central to Western musical thought, “represents the ultimate rationalization of a lattice-based view of music” (Wishart 1985, p. 17). Only gesture, which is “essentially an articulation of the continuum” (p. 12), can save us: “[In] music which attempts to deal with the continuum (rather than the lattice), gestural structure becomes the primary focus of organisational effort.” (p. 13)

Throughout the course of the book, Wishart attempts to establish “criteria for composing music with non-lattice materials which ‘work’ in some experientially verifiable sense that is not merely circular” (p. 25). Wishart, like Lerdahl (1997, p. 118–120), seems to be ardently concerned with finding the Elysian fountain of musical youth: an objective phenomonology of what constitutes “good” music.

Wishart discusses timbre and sound objects at length and attempts to establish a fundamental vocabulary. Because timbre eludes a physical metric and is multidimensional in nature, subjective descriptions must generally suffice for musical purposes. But we may create a music whose structure is derived from timbral organization by using methods analogous to traditional instrumental music:

> *Modulation (in the sense of clear progression from one field of sound-objects to another field) can be clearly demonstrated and utilized in timbre-space. Modulation between different timbral sets could clearly be used as a basis for the large-scale architecture of a work….* (p. 48)

Wishart defines a landscape as "the imagined source of the perceived sounds." We may compose with landscape by creating metaphorical relationships between sound objects and transforming those relationships during the course of the composition. For the remainder of *On Sonic Art*, Wishart presents a catalogue of virtually every possible spatial motion of a sound with respect to a listener, followed by a catalogue of many extended vocal techniques (which is also summarized in Wishart 1990, pp. 313–314). Curiously, the catalogue omits vocal techniques of musics from Eastern cultures that many would classify as "extended" in various ways, such as Tibetan chant and Tuvan throat singing.

Elsewhere, Wishart (1992) outlines the potential of technology to affect our music-making. He notes first the impact of recordings:

> *At the most obvious level, the advent of musical recording has made more kinds of music available to more people than was ever before possible. The ability to hear (and re-hear) any specific piece of music at any time has immense repercussions in the field of music learning.* (p. 566)

He later points out the pedagogical implications, particularly in ear training and computer-assisted instruction:

> *Accessible technology also has much to offer in the more conventional areas of music education. Simple aural training, which can be a drudge for both students and teachers alike, can be transformed using interactive computer programs, in the manner of computer games, which the student can adjust to his or her level/needs, leaving the teacher free to deal with higher level concerns.* (p. 566)

Furthermore, the use of synthesizers with timbres that approximate that of the instruments we are employing in our composition would allow us “to adopt a heuristic approach to building musically effective structures” so that we could “[b]uild, test by listening, rebuild” (p. 568).

The remainder of Wishart’s commentary addresses the impact of computers on self-publication of musical scores, musical instruments, and new performance paradigms. One statement deserves particular attention:

> *We should not, however, be fixated by the goal of real-time operation. Certainly, for most music-performance operations, a system that operates in real time is essential. But musical composition has never been a real-time occupation, and we should not be surprised if producing a complex sound or sound-sequence, through synthesis or analysis and spectral shaping, takes a certain time! Computers can certainly take some of the drudgery out of the compositional process. But the development of new ideas always takes time.* (Wishart 1992, p. 574)

Fortunately, many other composers and software designers do not share this view; recent advances in real-time computer-music environments have engendered the development of new forms of expression, from improvisation to interactive video art. Wishart's statement here seems to contradict a notion he expressed earlier: the "[b]uild, test by listening, rebuild" approach (1985, p. 568). I am not implying that real-time execution provides necessary and sufficient conditions for the creation of a work, but it certainly does not hinder the composer. And thanks to the ever-increasing speed at which microprocessors operate and new real-time software synthesis programs, the difficulty of real time is rapidly becoming a moot issue for most musical situations.

Wishart concludes his discussion on the impact of technology on music in another article, "From Architecture to Chemistry" (1993), in which he begins, "In the late twentieth century out principal metaphor for musical composition must change from one or [*sic*] architecture to one of chemistry." (p. 301)

A staunch modernist, Wishart observes that because "the possibilities of conventional instruments have been explored to their limit" (p. 301), we must summon the aid of the computer to provide us with new sounds and help us "take apart what were once the raw materials of music, reconstitute them, or transform them into new and undreamt of musical materials" (p. 302). Thus, the very nature of music and composition is transformed:

> *In this context, sound materials become like clay in the hands of a potter, and music becomes a plastic art, where sonic objects of any origin can be moulded to the particular shape required by the composer.… Again the computer provides the means to produce arbitrary transformations of the material, and compositional skill lies in both an understanding of musical acoustics–giving an insight into what kind of transformations will lead to what type of results—and aural judgement of those results.* (1993, p. 575)

His basic view of the impact of the computer on composition may be summarized thus:

> *The computer opens up areas of compositional exploration that were previously inaccessible. The precision with which sound materials can be specified implies two things: (1) Given an understanding of acoustics, sounds can be transferred directly from the composer's imagination to the performance situation; (2) Areas of sonic organization previously inaccessible to composers through the*

> *existing media of notation can be explored, opening up a new world of dreamed of, but unsung possibilities.* (Wishart 1988, p. 27)

In his other book and accompanying CD-ROM, *Audible Design* (1994), Wishart furthers his metaphor of the composer as chemist by providing a slightly more technical framework on manipulation and creation of sonic materials and how best to use them in composition.

Finally, the musical theories of Trevor Wishart include apocalyptic discussions of "Populism" and "Scholasticism" in music. Scholasticism refers to the evolution of musical language is controlled by an elite few, while Populism "assumes that not only must the language of music be recognisable to a large public, but that the discourse of music must be popular with a large group" (Wishart 1983, p. 106). Wishart argues that Scholasticism has dominated twentieth century musical discourse, and continuing "in this direction it can become only the handmaiden of an autocratic and elitist culture" (p. 106). Fortunately, he once again offers an ontological catalogue of factors that contribute to good music—"possible socio-musical 'givens' that might be acceptable as the basic roots of any musical language":

> *(1) Rhythm in the sense of felt, danced, human movement, speech rhythm, but not the psychologically arbitrary arrangement of "duration-structures."*

> *(2) Melody in the commonly accepted sense of recognisable tune-like gestalts (I have yet to hear someone whistle Schoenberg's tunes as he predicted).*
>
> *(3) Language, and all human utterances, and all extensions of these, and the articulations and timbral patterns arising from them.*
>
> *(4) Landscape, in the sense of recognisable real-world sounds and sound-environments and sound-constructs deriving from these in various ways.*
>
> *(5) Music-Theatre… combination of musical-organization with theatrical gesture and situation, visual props and effects… preferably pointing outside the confined world of professional musical performance itself, and its idiosyncracies* [sic]. (p. 106 et seq.)

Followed by: "I am not suggesting that all of these are necessary features of an accessible musical language. But perhaps at least ONE of them is." (p. 107)

These "givens" remind one of Lerdahl's two aesthetic claims that the "best music utilizes the full potential of our cognitive resources" (Lerdahl 1997, p. 118) and the "best music arises from an alliance of a compositional grammar with the listening grammar" (p. 119).

It is of course an onerous and perhaps impossible task to attempt to outline the basis for a music which is universally acceptable, and certainly futile to make aesthetic claims as Lerdahl does which are nothing more than value judgments regarding the constitution of "good music," but Wishart's call for an end to unmusical, intellectual "Scholasticism" is perhaps both timely and laudable.

❧

The theories of Schaeffer, Schafer, Smalley, and Wishart posses a common focus: providing new perspectives on the acts of composition, performance, and listening. With the exception of Schaeffer, who did not consider his output music, they all note that, in order for music to utilize a non-lattice-based construction and organization, our ears must be open to appreciating sounds at all levels, both in our environment and in the concert hall. Non-lattice-based music may be composed using techniques analogous to lattice-based notated music, but it must be based on archetypal notions of gesture, utterance, behavior, and their morphologies in order to provide a sense of structural coherence.

All four similarly object to the application of serial principles to non-lattice-based composition. Wishart gives one reason why when discussing his use of extended vocal techniques in *Vox-1*:

> *Each of the second order morphologies may be associated with the first-order morphologies which exhibit change. This gives us twenty-two perceptible different morphologies and each of these may be associated with the three magnitudes, giving a total of sixty-six gestural archetypes. If we now remember that these articulations may be applied to both the frequency width and the rate of iteration of the*

> *vibrato, we now have 3,756 ways of articulating vibrato! We may now apply the same gestural criteria to the overall dynamic envelope of the note and the tremolo characteristics. We can describe 14,106,536 possible articulations for a standard musical note. If we now enter the field of the true continuum and consider portamento motions of the pitch and timbral transformation of the pitch through time, we discover 50,000,000,000 perceptibly distinguishable sound-objects. At this point, serial methodology loses its charm.* (Wishart 1985, p. 67)

Simply put, now that we are composing at the level of sound itself rather than the higher, abstracted level of musical notation, there are simply too many data to serialize.

Wishart registers another critique of serialism applied specifically to loudness:

> *The formalistic assignment of a series of different dynamic levels to musical objects, which was experimented with in the Total Serial aesthetic leaves a sense of arbitrariness or agitation (neither of which is usually intended) because it ignores the landscape basis of our perception of loudness.* (1985, p. 99)

The arbitrariness to which he refers in the context of his landscape theory of music is that of causality. The laws of causality rule the musical landscape, and, simply put, the total serialist aesthetic forms the basis of a non-causal system. Wishart, in addressing the “arbitrariness” of serialism and advocating the use of psychologically causal compositional grammars, joins Lerdahl ("serial (or 12-tone) organizations are cognitively opaque"; see Lerdahl 1997, p. 97) and Reich:

> *John Cage has used processes and has certainly accepted their results, but the processes he used were compositional ones that could not be heard when the piece was performed. The process of using the* I Ching *or imperfections in a sheet of paper to determine musical parameters can't be heard when listening to music composed that way. The compositional processes and the sounding music have no audible connection. Similarly in serial music, the series itself is seldom audible.* (1974, p. 10)

As mentioned previously, Schaeffer admitted the usefulness of serially organizing sounds during the early years of musique concrète for exemplary purposes, but such a system was ultimately unacceptable in his opinion as far as concrète was concerned "insofar as it displays the rigidity of a method" (Palombini 1993b, p. 19).

Wishart and Schafer both address aleatory either directly or indirectly. Schafer frequently employs aleatory to involve the performers at a very direct level in the realization of the composition, a notion which, as mentioned, is fundamental to Schafer's theory. Scores such as the aforementioned *Epitaph for Moonlight* (1968) feature approximate timelines and instructions such as "A medium high note ad lib.," "Solo Sopranos and Altos ad lib.—free pitch and rhythm," and "All Instruments soft glissandi." Suggestive pitch contours are frequently hand-drawn in works such as the *First String Quartet* (1973).

Wishart, on the other hand, tends to employ maximally descriptive non-standard notation and more rigidly defined instructions to the performers in compositions such as *Vox-I* and *Anticredos*.

Wishart and Schafer both discuss the role that technology should play in music education. As mentioned earlier, Wishart advocates the use of computers and game-like software for ear-training. Schafer, who founded the electronic music studio at Simon Fraser University in 1965 (Truax 1990), is also "notable for its emphasis on electronic music for schools and colleges," as Manning notes (1993, p. 186). In addition, Schafer's many books emphasize an experimental, hands-on, low-tech method of music education in which students make music with ordinary objects. Wishart, too, has employed similar pedagogical techniques in group games such as "pass the sound."

A final and very fundamental distinction may be drawn among theories in this group regarding the "musicality" of sound objects. Palombini discusses Schaeffer's view:

> *The* Étude aux chemins de fer *posed the problem of musically organizing sounds produced by six locomotives at the Batignolles station. Schaeffer recorded the stokers' improvisation. Rhythmic leitmotives were then isolated. Montage (mixing) attempts led to both dramatic and musical sequences. Dramatic sequences, referring the listener back to events (departure, stopping, etc.), were considered unmusical by Schaeffer.…Dramatic sequences were not eliminated, but the discerning listener was expected to prefer the musical ones. (1993b, p. 15)*

In fact, the very focus of the *Solfége de l'Objet Sonore* is the quest for "musical" sound objects. Wishart voices the Cagean opposition in the introduction to *On Sonic Art*:

> *Also, one further important point, in contradistinction to what is implied in the "Solfege de l'Objet Sonore". This book assumes that there is no such thing as an unmusical sound-object.* (Wishart 1985, p. 6)

Schaeffer, Schafer, Smalley, and Wishart have all clearly provided major contributions to the development of a theory of non-lattice-based composition at the level of sound itself. Perhaps, as a result, Schaeffer's dream of a "baroque music for the 21st century" (Hodgkinson 1987) will come to fruition as the lack of proliferation of a dominant language continues to roam the musical landscape.

5.4 Implicit Characteristics of Sound Objects

It is clear from historical experiments with dissonance perception of intervals that various factors contribute to the perception of auditory dissonance; in previous chapters, concepts such as fusion, harmonicity, beating, and roughness were mentioned. The same factors, and arguably more, must be at play when a listener consciously analyzes the perceptual dissonance of an isolated sound object, for these reasons. First, a classical dissonance listening test could easily be constructed in which intervals are played from a recording through loudspeakers instead of by using two tuning forks, for example. The recording of such a sound in isolation is itself a sound object as I have defined it, and thus the listening test that used actual tuning forks must itself have been testing listeners' judgment of a subset of possible sound objects (in this case, pure-tone dyads). Because a recorded tuning-fork dyad and an actual tuning-fork dyad can be perceptually indistinguishable given the proper sound-reproduction equipment, the actual method of presentation of the sound object is irrelevant. And so, because we know of particular acoustical and psychoacoustical factors that contribute to listeners' perception of dissonance of pitch-based phenomena, the same factors must be at play in the perception of dissonance of sound objects in general, for pitch-centric sounds form a subset of all possible sound objects.

Second, more factors must be considered when analyzing sound-object dissonance than those used when examining interval alone. This is true because sound objects occupy an entire spectrum of "mass," to use Schaeffer's terminology. Many sound objects invoke no sensation of pitch perception, and therefore other factors besides the historically examined elements of smoothness, purity, and so on, must contribute to dissonance perception. For example, I hypothesize that most listeners would classify a short, loud burst of white noise as more dissonant than a longer, soft, low-pass filtered noise with a gentle attack and decay envelope. None of the spectral components of either sound tend to fuse (other than temporally, in terms of the common fate of their amplitude envelopes); they certainly are not "smooth" or "pure," either.

While it is impossible to accurately guess all specific factors that contribute to one's perception of sound-object dissonance, I hypothesize the primacy of factors that fall naturally into five categories: spectral features of the sound object, temporal features, spatial features, other acoustical features, and perceptual (psychoacoustic) features. I will briefly outline each of these below. Each of these groups is discussed in terms of its potential to affect a listener's judgment of auditory dissonance. After actual listening tests have been conducted, their respective contributions will be evaluated, as much as feasible, and compared to the experimental results.

Spectral Features

The first group of hypothesized contributors to dissonance of the musical sound object involves the spectrum of sound in terms of its static (global, or averaged) features. The properties that are most suggestive perhaps from previous dissonance studies include *harmonicity*, *spectral overlap, spectral centroid*, *spectral flatness*, and *spectral fluxoid*. Of course, a sound object's time-varying spectral properties are relevant as well, and so each of the following primary features should be examined for temporal properties as well.

Harmonicity

The harmonicity of an audio segment is directly related to its periodicity. Fourier theory tells us that periodic sounds are by nature harmonic, and non-periodic sounds are inharmonic. To the extent that a sound is periodic and therefore harmonic, the frequencies of its spectral lines exhibit simple harmonic number ratios. At its simplest form, this relates to Schaeffer's notion of "mass," for the more harmonic a sound, the greater our sense of "pitchedness" upon hearing it. Various mechanisms for quantifying harmonicity have been proposed (most notably the harmonic product spectrum and harmonic sum spectrum; see Noll 1969), and these have found use in pitch-determination algorithms (Cuadra, Master, and Sapp 2001).

Extensions of simple harmonicity calculations have been more recently applied to the perceptual coding of audio signals. One of the primary stages in such systems is the computation of the "tonal" and "noise" components of an audio signal (e.g., Johnston 1998); because tonal components tend to possess different auditory masking properties than do noisy components (e.g., noise masks tones in general much better than vice versa), the tonality index of each bin (or bandwidth in a cochlear filter model) can be used to determine the psychacoustic salience of that component in context.

Other methods for measuring the periodicity or noisiness of a signal are available, including time-domain methods such as linear prediction, as well as other frequency-domain methods, such as spectral compactness and sharpness (von Bismarck 1974a, 1974b) and spectral flatness.

Spectral Centroid

The spectral centroid, a metric applied to audio signals by Beauchamp (1982) and many others since, is simply a weighted average of the spectrum's frequency components. Casually, it is said that the centroid indicates the relative "brightness" or "darkness" of the sound by computing the spectral center of gravity. Its computation has been used for many years in a variety of signal-processing contexts.

Spectral Flatness

Spectral flatness is related to the concept of harmonicity, in that flatness is generally often used to compute the tonality index of a sound in perceptual coding systems. It is defined the ratio of the geometric mean frequency to arithmetic mean frequency of the power spectral density (PSD) exhibited within each critical bandwidth. This metric has found recent use in the automatic segmenting of audio streams by noting the high correlation between changes in spectral flatness and desired segmentation tasks (Izmirli 2000). It would make sense that rapid changes in the spectral flatness measure (SFM) of a sound indicates a transient, constantly changing sound object, which should clearly in turn have a bearing on listeners' judgments of dissonance of that sound object.

Spectral Smoothness and Spectral Fluxoid

The spectral smoothness measure, proposed by McAdams (1999) indicates the envelope of a sound's spectrum for a given frame by quantifying the difference in amplitude between adjacent bins. Spectral smoothness has found applications in several areas, including fundamental frequency-estimation algorithms (Klapuri 2003). Similarly, the spectral fluxoid (or spectral flux) attempts to quantify spectral changes over time by computing a difference function between the spectra of adjacent audio frames. In a sense, it quantifies the temporal stability or "constancy" of the spectrum.

Signal Quality and Data-Reduction Artifacts

At the risk of stating the obvious, audio signal quality must play a role in the perception of consonance and dissonance of sound objects. Inasmuch as modern electroacoustic music is created, stored, and played back from digital audio samples, the precise method used to create, store, and playback those bits of data can impact a listener's perception of the intended sound objects.

But first, an important philosophical point must be reiterated regarding the distinction between the idealized sound object and real sound object. We noted earlier that an abstract sound object exists only in the mind's eye, and that its existence in acoustical reality is tied not only to its pitch-oriented, timbral, and temporal features, but in particular its physical existence as an acoustic phenomenon in space. In the case of sound objects that are recorded or synthesized and then stored digitally, however, the signal quality of the storage-playback system must be added as a requirement for its "realness." We now have three categories of sound object as a result: the *abstract* or *idealized sound object* (a theoretical construction), the *real sound object* (an acoustical product), and the *stored sound object* (a digitally encoded representation of the real sound object).

Analysis of the dissonance of abstract sound objects, while interesting, is a more esoteric task of perhaps limited use, because we all remember things differently. Asking subjects to assess the dissonance of abstract sound objects described on paper (e.g., "waterfall" or "car engine"), as Schafer (1977) does, may provide useful information from an acoustic-ecology point of view, but it fails to provide any kind of meaningful acoustic or psychoacoustic insights. Similarly, measuring "real" sound objects (those that acoustically exist but are not recorded; e.g., asking listeners to assess the dissonance of hearing an actual hand clap) introduces an unwelcome variable into the test, namely, repeatability. The most logical choice, then, which mitigates these problems, is to assess dissonance of sound objects using sound objects that are digitally stored and equivalently reproduced for a variety of subjects. The primary layer of complexity in the progression from "real" sound object to stored object lies in the realm of the digital codec used to store and reproduce the sound object.

We can subdivide quantifiable dissonance contributors of this type into three broad categories: raw audio signal quality, hardware encoder/decoder (codec) distortion, and software codec distortion. The precise means by which the "signal quality" of a sound stored digitally should be described is of course open to interpretation, but several obvious features of signals lend themselves to such a global description. For example, the MPEG-7 standard includes an audio signal quality descriptor (`AudioSignalQualityDS`) that reports features such as digital clips, clicks, cross-channel correlation of multichannel sound files, signal bandwidth, background noise level, DC offset level, and other measures.

Hardware interfaces can potentially audibly distort audio during recording or playback, particularly if basic tenets of digital signal processing are not observed (for example, the Nyquist theorem). Quantization noise and other artifacts can result if too few bits are used to store digital audio samples, particularly when recording sounds at relatively low sound pressure levels. Furthermore, timing jitter, or even simple lack of a stable word clock reference, can easily lead to audible artifacts that can be readily quantified and measured, particularly when recording a known signal. The analog world should not be discounted here either: microphone self-noise as well as amplifier harmonic distortion and other factors should be classified into this category as well, as in a sense they are the analog form of digital codecs (i.e., transducers).

The contributions of software codecs to audio signal quality—and to human assessment of signal quality—are perhaps even more prevalent, particularly now in an ever-expanding age of low-bit-rate audio coding. Despite their best intentions, modern data-reduction codecs, such as Apple Computer's Advanced Audio Coding (AAC), Dolby Laboratory's AC3 multichannel codec, Microsoft's WMA codec, and the ubiquitous MP3 specification, lead to measurable loss of signal quality by definition. Even though codecs that attempt to incorporate models of human auditory perception in order to eliminate psychoacoustically redundant data can lead to significant data reduction, their artifacts of course are easily quantifiable. They are thus worth noting, if only briefly, in the context of outlining measurable characteristics of recorded sound objects.

Aside from bit rate and data reduction rate, other signal quality parameters can be precisely measured. These include simple measures such as the codec residual (the difference between original and encoded files, found simply by piecewise subtracting the samples of the encoded file from the original, unencoded file), as well as traditional measures like the signal-to-noise ratio/dynamic range of the codec. However, no widely adopted and systematic method of precisely assessing and comparing audio codecs currently exists. Most often, perceptual success of audio codecs is generally measured via listening tests in which individuals blindly select their preferred codec. Furthermore, as Pohlmann (2005) notes, traditional quantitative measures generally fail to provide any meaningful insight, adding that the specific commercial pedigree of the codec, taken in conjunction with bit rate, represent perhaps the two most important determinants of signal quality.

(Note that signal quality can be measured in terms of spectral, temporal, or combined spectro-temporal measures. For example, while most studies concentrate on spectral distortion introduced by codecs, temporal artifacts—particularly smearing of transients and pre-masking—are manifest as well. However, signal quality is discussed here alongside spectral features of sound objects for convenience.)

Other Spectral Measures

Research into timbre analysis and the sound segregation problem has yielded many other measures of the spectral features of a sound. (See Park 2004 for an overview.) These include the log spectrum spread, spectral cutoff (or rolloff), spectral shimmer, spectral jitter, harmonic slope (Pollard and Jansson 1982), cepstral analysis measures such as MFCCs (Davis and Mermelstein 1980), and features derived via wavelet analysis (e.g., Tzanetakis, Essl, and Cook 2001).

Temporal Features

Clearly, the shape of a sound object’s amplitude evolution over time affects dissonance judgments. It is a reasonable assumption that loud sounds with fast attack times could tend to be perceived as “threatening” by many listeners, thereby hastening the classification of “dissonant” by listeners according to the extent that biology, auditory masking, and evolution contribute to our sense of dissonance.

Aside from extracting the amplitude envelope of a sound object in terms of attack, sustain, release, and decay times—which can be quite difficult for unknown audio signals—more general measures can be extracted for the sound object as a whole. These include the amplitude peak, root-mean-square (RMS) amplitude, and temporal centroid.

Spatial Features

Aside from spectral and temporal features of a sound, the spatial properties of a sound object may contribute to dissonance judgments. Examinations of spatial feature vectors seem to have been neglected in favor of spectral and temporal features, partially owing to the exclusive considerations of monaural sound files in the literature and the increased number of testing variables that spatial considerations suggest. However, with the increasing availability and even ubiquity of surround audio recording and playback systems, spatial feature vectors may indeed prove an important area of future investigation.

Reverberation

It has often been casually remarked that applying reverberation to an audio signal is akin to sprinkling food with sugar…it makes anything taste (or sound) pleasing. (John Chowning reportedly said that reverberation is the "ketchup" of computer music.) To my knowledge, no studies have conclusively measured the relationship between dissonance perception and reverberation, but anecdotal evidence seems to point toward a highly correlative relationship.

Spatial Location in a Surround Field

Few metrics attempt to quantify the average spatial location of sound (relative to the listener), spatial smoothness, spatial fluxoid, or any kind of similar measure, apart from a few “fuzzy” auditory displays of spatial location (e.g., Digidesign’s SurroundScope, shown in Figure 5–3) and quantitative measurements of individual head-related transfer functions (HRTFs). It would indeed be useful to quantify spatial features of a sound field to be able to say, for example, that a multichannel sound object is “75% front-center” and “25% rear-left.” Such metrics may prove useful in investigating dissonance perception in surround audio fields.

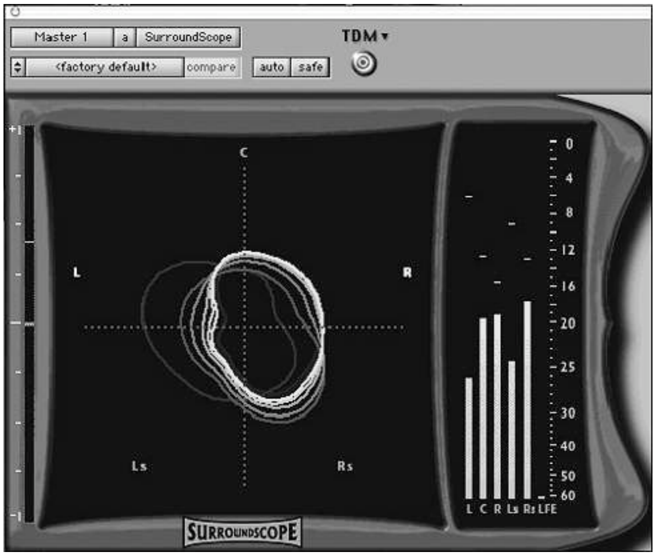


Figure 5–3. Digidesign’s SurroundScope surround-audio monitoring plug-in.

Headphones

Listening to music over headphones creates a far different spatial effect than does listening to the same music over loudspeakers at a distance. With the exception of binaural recordings, most recorded music was not mixed and mastered for headphone presentation, and as such, certain spatial cues may be exaggerated or even incorrect over headphones, which can certainly affect perceptual dissonance of the sound. For example, headphones tend to exaggerate the interaural intensity difference (IID) cue, which could result in a false spatial image and an increased perception of acoustic threat. Furthermore, even inexpensive headphones are also capable of producing extremely high sound pressure levels, even at high frequencies, because they are almost directly coupled to the ear canal. Such high levels, even if caused by an accidental turn of the volume knob on an MP3 player, could certainly induce annoyance and threat, thereby affecting judgments of relative perceptual dissonance.

Other Acoustic Features

It is easy to hypothesize other important contributors to dissonance perception of sound objects. For instance, we may consider the autocorrelation or self-similarity of a sound, the fractal dimensionality, or the signal quality itself, examined from a communications engineering perspective. For example, we may consider the signal-to-noise ratio, noise-floor level, bit rate, or data compression or codec artifacts, as potential contributors to dissonance.

Many of these and the measures discussed above are conveniently available when working with sound within the MPEG-7 framework as descriptors, simplifying their computation. Some features, however, are not possible currently to compute directly. For example, the blind computation of absolute attack time and reverberation level is still an active area of research.

Perceptual Qualities

The features above deal exclusively with physical, quantifiable, acoustical properties of sounds. The computation of their psychoacoustic correlates is much more difficult and is an active area of research as well, thanks largely to ongoing psychoacoustic and neurological studies. Regarding perceptual correlates of the acoustical features listed above, perhaps the easiest to directly compute, at least to an approximation, are (1) the correlate of the fundamental frequency (whether actual or virtual), which we generally call pitch or "viritual pitch" (Terhardt 1972); (2) the correlate of spectral flatness and harmonicity taken together, which we might call "pitchedness"; and (3) the correlate of amplitude, which is traditionally referred to as "loudness." Furthermore, the first and last of these psychoacoustic correlates has a well-defined and experimentally supported measurement scale: pitch can be measured and quanitified to some extent in units of mels (Stevens, Volkmann, and Newmann 1937), while loudness can be measured and quantified in units of phones or sones (Stevens 1955).

Other perceptual features of sounds have not yet been fully examined. One of these is the correlate of tempo, which can be confounded owing to tempo "octave" errors; that is, one listener may tap internally to quarter notes in a passage, which the other may tap to eighth notes. Various psychoacoustic models of tempo perception have been proposed (e.g., Desain 1992; Parncutt 1994; McAuley 1995); however, no complete model that attempts to quantify the psychoacoustic correlate of tempo has yet been produced. Were such a model to exist, though, it could certainly assist beat- and tempo-tracking algorithms by incorporating tempo-related psychoacoustic phenomena that have not yet been fully incorporated into any existing model.

Several perceptually important features of sounds, however, do not lend themselves to quantization on a scale or are somewhat more ineffable in nature. Moreover, because they are psychoacoustic features, their perception can vary greatly among listeners. Features of this type include (1) the "recognizability" of a sound; (2) perceptual correlates of spatial dissonance; (3) causal dissonance; (4) context and "acoustic cognitive dissonance"; (5) the perception of auditory threat or danger; and (6) potential interplay affected by visual cues accompanying the sound. The first of these, recognizability, goes to Schaeffer's idea of reduced listening—that is, the listening of a sound for its properties exclusively as a sound, not for ancillary properties such as physical causation or recognition of the sound. Anecdotally, I can assert that the degree to which a sound object is recognized by a listener can greatly impact that listener's perception of its dissonance level owing to memory, cultural factors, and prior experiences with that sound (or other sounds related in some way). Lack of recognizability, or "unidentifiability," on the other hand, naturally encourages Schaeffer's *écoute réduite* by operating on an auditory "blank canvas" within the listener's mind.

The next kind of dissonance in this category involves the perceptual correlates that accompany acoustical spatial dissonance. An example of acoustical spatial dissonance might be a multichannel sound object in which, owing to the spatial mixing technique used in the sound's production, certain frequency components simply mask each other. (For example, consider a multichannel object in which all of the sound is mixed into the front-center loudspeaker, and some of the components within each individual channel are masked as a result.) Their acoustical spatial dissonance could perhaps be lessened by widening the surround image, thereby potentially minimizing the destructive interference and phase cancellations and limiting the masking that would otherwise occur. Psychoacoustic spatial dissonance, then, is simply the perceptual correlate of this type of phenomenon. If we define acoustical spatial dissonance as the degree to which auditory masking and phase cancellations are exhibited in a mix, then the correspondence between the quantifiable reduction in masking and the perception of such a reduction defines the notion of psychoacoustic spatial dissonance.

But the definition of spatial dissonance need not be limited to auditory masking caused by spatial location; other scenarios may promote spatial dissonance. It could consist of, for example, spatial orientations of sound objects that defy convention in some way, or perhaps contradict previous experience with that particular sound. A simple example of this might be reversing the left and right channels of a stereo sound file containing a stereo recording of a piano; the listener may expect to hear the lowest pitches of the piano on the left channel, with the pitches increasing toward the right channel. This is of course our spatial experience when playing a piano. Flipping the channels of the recording such that the lowest pitch is on the right might cause spatial dissonance to be perceived, albeit perhaps only for trained listeners.

As another example, I recall from conversations with composer Paul Lansky the spatial dissonance he experienced while recording samples of passing cars for his composition *Night Traffic*. He had apparently swapped stereo channels when plugging a stereo microphone into a portable digital audio recorder while standing in front of a highway; the experience of seeing a car passing from right to left while hearing it pass from left to right must surely have been disconcerting, if not spatially dissonant. This scenario alludes to simulacra of visual and auditory dissonance in general. In this particular example of a passing car, whether the experience is spatially dissonant or visually dissonant depends on one's frame of reference and acceptance of reality. That is, are the ears correct, or are the eyes correct?

The third kind of this perceptual dissonance could be termed *causal dissonance*, whereby a listener's psychological expectations of cause and effect are thwarted in some way. While abstract sound objects may or may not lend themselves to a natural spectromorphology, or "ideal" spectral continuation—an idea that has is debated in Wishart (1985) and elsewhere—recognizable sounds exhibiting an experientially familiar cause-and-effect relationship are especially subject to causal dissonance. An example of a pair of sound objects that exhibit a high degree of causal consonance is the sound of a chainsaw ripping through a tree trunk, immediately followed by the sound of the tree's fall and crashing into the ground. As another example, consider popping the cork on a bottle of champagne, followed by the overflow of bubbles outside the bottle—a cause-and-effect relationship of actions that is also accompanied by characteristic sounds.

A related category is found in sound objects that do not necessarily exhibit a cause-and-effect relationship, but nevertheless tend to follow a consistent order in our previous experience. Think here of the sound of cracking an egg followed by the sizzle of the egg cooking in a pan, or the sound of someone whistling a continuously descending tone followed by a vocalized “explosion” sound. An even better example is found in backward reverberation, a common effect in much popular music, in which a reverberant wash of sound precedes early reflections, which in turn precede direct sound.

A fourth variety of sonic-perceptual dissonance could be termed *contextual dissonance*. In this case, a listener’s expectation of one or more elements of a spectromorphology or sonic continuation are usurped. This usurpation can be accomplished in one of two ways: usurpation primarily via repetition, stability, pattern, or good continuation; and usurpation of expectation primarily based on a listener’s prior experience. For example, a sound object that exhibits a steady rhythmic pulse for a time (perhaps the sound of a woodpecker) and then suddenly discontinues the established, expected pattern, exhibits contextual dissonance owing to usurpation of repetition. On the other hand, a sound object that begins with the sound of a child’s singing “Row, Row, Row, Your Boat,” in which the word “Boat” were replaced by the sound of an explosion, creates contextual dissonance based on prior experience. A musical example of this is found in Pierre Schaeffer’s late tape work *Bilude* (1979), in which notes from the first prelude of Book I of J. S. Bach’s *Well-Tempered Clavier* are replaced by out-of-context short recordings. To the extent that a listener’s prior experience coincides with physical causality, then both contextual dissonance and causal dissonance of the sound object are invoked.

Dissonance of sound objects is also created when a listener consciously or subconsciously perceives threat, danger, annoyance, or disgust from a sound object. This notion has been explored elsewhere (e.g., Huron 1997), and the general idea is that sounds that tend to invoke “fight-or-flight” responses (e.g., gunfire, screaming, and other sounds of violence and destruction) tend to be considered negatively by listeners. To the extent that consonance is in general considered a positive trait and dissonance a negative trait (at least psychologically speaking), then negative-emotion-inducing sounds should be considered as relatively dissonant vis-à-vis positive-emotion-inducing sounds.

Lastly, a sixth variety of perceptual dissonance can also be created from codecs (both hardware and software). We noted earlier some of the quantifiable measures used to rate signal quality, and the impact of codecs on the potential perceptual dissonance of sound objects must not be discounted, either. Most studies that address psychoacoustic assessment of codecs tend to report simple binary measures (i.e., “Which codec sounds ‘better’?”) after conducting a variety of listening tests (e.g., ITU Recommendation BS. 1116–1, “Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems”). More specific listening tests, perhaps containing more pointed questions of the participants (e.g., “Which codec makes the sound file sound ‘brighter’?” or “Which codec makes the sound file sound more ‘present’?”), will lead to more specific answers regarding the impact of codecs on dissonance ratings of sound objects by listeners.

5.5 A Prototype Dissonance Theory of Sound Objects

With the forgoing concepts in mind, then, I propose the following six fundamental components of a theory regarding the acoustical and auditory dissonance of sound objects. Again, I only refer here to sound objects considered in isolation, apart from their environmental and/or musical contexts.

1. **All else being equal, sound objects that mask desired auditory information (i.e., information the listener desires to hear) will tend to be classified by listeners as more dissonant than those that exhibit no such masking.**

This contention is supported by recent research (Huron 1997; Fishman et al. 2001; Bolger and Griffith 2003). A sound object that contains two or more simultaneous streams of musical data (for example, the sound of a loud waterfall and the sound of a human voice speaking) will be more dissonant because one stream may mask another. In this example, to the extent that the sound of the waterfall masks the human speech (assuming the listener desires to hear and understand the speech), the combined sound object is more dissonant than, say, the sound of the same human speech without the waterfall. Furthermore, the degree to which desired sonic information is masked must relate proportionately to what we might call *auditory frustration*. This first assertion is a natural extension to generalized sound objects of Plomp's and Levelt's (1965) classic theory of music consonance as it relates to critical bandwidth.

2. **All else being equal, sounds that evoke negative-valence sensations of annoyance, dislike, fear, disgust, threat, anger, boredom, or hate will tend to be classified by listeners as dissonant.**

Historically, consonance per se has been treated as a musical positive, and dissonance as a musical negative. The claim that they correspond respectively to high preference and low preference is supported in a variety of studies, including studies that indicate infants' preference for consonance over dissonance (Zentner and Kagan 1998) and even animals' preference for consonance (Borchgrevink 1975; Hulse, Bernard, and Braaten 1995). Although extrapolating the listening preferences of albino rats (Borchgrevink 1975) and European Starlings (Hulse, Bernard, and Braaten 1995) to human audition may be problematic, humans do seem to possess some kind of biological auditory pre-programming. But the case of infants' preference is indeed clear; Zentner and Kagan (1998) write that

> *Infants looked significantly longer at the source of sound and were less motorically active to consonant compared with dissonant versions of each melody. Further, fretting and turning away from the music source occurred more frequently during the dissonant than the consonant versions. The results suggest that infants are biologically prepared to treat consonance as perceptually more pleasing than dissonance.*

However, it is fair to say that far more studies have been conducted on listening preferences of "consonant" and "dissonant" versions of intervals and melodies than of sound itself. The latter is the subject of listening tests discussed in the next chapter of this thesis. At the time of this writing, I am aware of several recent studies (Marquis-Favre, Premat, Aubrée, and Vallet 2005; Marquis-Favre, Premat, and Aubrée 2005; Lee et al. 2005) that address the latter by measuring listeners' annoyance with various sounds.

The concept of sound inducing a sensation of threat or danger makes sense from a biological standpoint, in that certain "real-life" sounds can of course cause panic or fear. But how can recordings of sounds do the same? From my own experience, highly reverberated sound objects tend to sound more consonant than dry sounds; it is often remarked that adding reverberation to a sound not only masks technical and musical deficiencies in the recording, but that just the right amount of reverberation makes the sound sound better. (I recall a teacher of mine at one point who had just given an organ recital in Westminster Cathedral remarked that "even a train wreck would sound good in that place because of all the reverb!") I submit that this phenomenon may relate in some way to the biologically innate perception of threat and danger from sound. Sounds objects that exhibit an higher ratio of reverberant to dry sound will tend to be classified as more consonant than those with a lower ratio of reverberant to dry sound, and one of the reasons for this is that highly reverberant sounds seem to pose no danger because the sound source is so far away from us. Following this line of reasoning, the more smooth a sound object's amplitude envelope and the more low-pass-filtered it sounds, the more likely it will be to perceived as consonant because of its interpreted physical distance will generally be greater than transient, bright, “in-your-face” sound objects.

3. **Sound objects that exhibit periodicity, predictability, and regularity of acoustic phenomena (e.g., regular sinusoidal beating, consistent tempo, and high degree of harmonicity) will tend to be classified as more consonant by listeners than those sound objects exhibiting lower corresponding periodicity, predictability, and regularity.**

Even though our subconscious does not enumerate the ratios of musical intervals that we hear (as Leibniz and others of the 18th century thought), centuries of writings about the dissonance of intervals has shown that we tend to find “simpler” interval ratios more consonant. This is due at least partly, as Helmholtz (and of course Pythagoras and most people in between) thought, to the regularity and certainty with which the periods of the intervals constituent frequencies align. (The combined waveform for a 3:2 perfect fifth, for example, repeats after every two periods of the lower frequency and three periods of the higher frequency[iii]. On the other hand, the pattern for a 16:15 minor second repeats much less frequently—in fact, 40 times less frequently per unit time.)

[iii] More precisely, the air molecules themselves that are directly in front of two ideal tuning forks tuned 3:2 reach equilibrium and repeat following this pattern.

Clearly, far more factors contribute to the perception of simple intervallic dissonance than the combined periodicity and harmonicity of the resultant waveform, particularly thanks to 20th-century analyses of listener preferences regarding intervals played in equal temperament versus just intonation (see for example Borden 2003).

However, psychologists do assert a natural human preference for both auditory and visual regularity and consistency, particularly with perceptual grouping tasks (Bregman 1990), possibly motivated by an innate tendency toward data reduction (Barlow 1959; Smaragdis 2001). In addition, recent work in examining Mismatch Negativity—an electrical response in the brain that results when stimuli fed to a human subject are suddenly changed (Näätänen 1995; Näätänen, Jacobsen, and Winkler 2005)—continue to shed more light on the subject. Other work by Katz (2004) asserts that music and sounds are preferred "to the extent that [they induce] synchrony in those brain structures that are responsible for processing the passage."

This being said, it is a small jump to assert that listeners will tend to classify sound objects that exhibit a high degree of regularity and consistency (in terms of their harmonic, rhythmic, and/or timbral components) as more consonant than those that exhibit lower degrees of regularity and consistency. The precise means with which regularity and consistency can be quantified and rated are varied (just as the concept of musical similarity is; see Berenzweig, Logan, Ellis, and Whitman 2004), but the spirit of this assertion should be clear when comparing sounds qualitatively.

This assertion is potentially complicated by the preference of many listeners, at least from my own experience, for unpredictability and slight variation often referred to as “warmth” (or colloquially as “phatness”) in a sound object, although (again anecdotally) this seems to apply almost exclusively to pitched sound objects. (The terms are often used as descriptors of a musical instrument’s tone rather than of sounds in general.) This is of course a valid criticism of claim 3, along with the claim by many musicians that equal temperament sounds “better” than just-intoned scales for this very reason; the same is often said for the preference of vibrato in solo instruments and the singing voice. However, some may argue that cultural factors are more responsible for these phenomena than simple first-response reactions to sound objects played in isolation.

4. **All else being equal, sounds objects that exhibit good continuation of spatial trajectory, amplitude envelope, and frequency content will tend to be classified as more consonant than those that exhibit lower degrees of good continuation of the same properties.**

Bregman (1990) notes our preference for good continuation of sound contours, and that we tend to use the gestalt principle of good continuation as a primary means of grouping sounds. In this context of the present essay, sounds that are grouped together in the gestalt sense over a relatively short time span are referred to as a sound object. It is therefore a small jump to assert that sound objects that by definition exhibit high degrees of “gestaltness” or auditory unity do so potentially in part due to the high degree of good continuation of their constituent parameters (i.e., rhythmic, harmonic, timbral, and spatial.

This assertion raises the idea that some sound objects are innately "better" at being sound objects than other sound objects. (Said another way, this means some potentially some sound objects exhibit higher degrees of "sound-objectness" than do other sound objects.) This assertion carries over from basic Gestalt psychology, which admits that perception of "gestaltness" varies among subjects.

5. **All else being equal, digital audio signal quality will exhibit a strong correlation to listeners' perception of the dissonance of sound objects.**

This concept is also related to contention 1, inasmuch as signal quality can lead to undesirable masking effects and therefore impede biologically important auditory sensations such as recognizability, familiarity, intelligibility, memory, and so on. To the extent that (1) consonance is linked to "auditory preference" and dissonance is linked to "auditory dislike" and (2) auditory fidelity is viewed as a preferred quality in the electronic reproduction of sound objects, it follows that signal quality will directly relate to listeners' perception of sound-object dissonance.

❧

In closing, the consonance-dissonance continuum occupied by sound objects represents a fundamental parameter in composing, listening to, and analyzing electroacoustic music. Surprisingly little has been written about the relationship between timbre and dissonance, particularly in the context of electroacoustic music. It naturally follows, however, from writings that explicitly ascribe the creative power of the medium to the unprecedented potential of composition within a continuum rather than the confines of discrete notation systems[iv] (e.g., Xenakis 1971; Estrada 1994) and from modern work that address the compositional-organizational potential of musical parameters other than pitch and rhythm (e.g., Wessel 1979), that the consonance-dissonance continuum represents yet another potentially valuable resource.

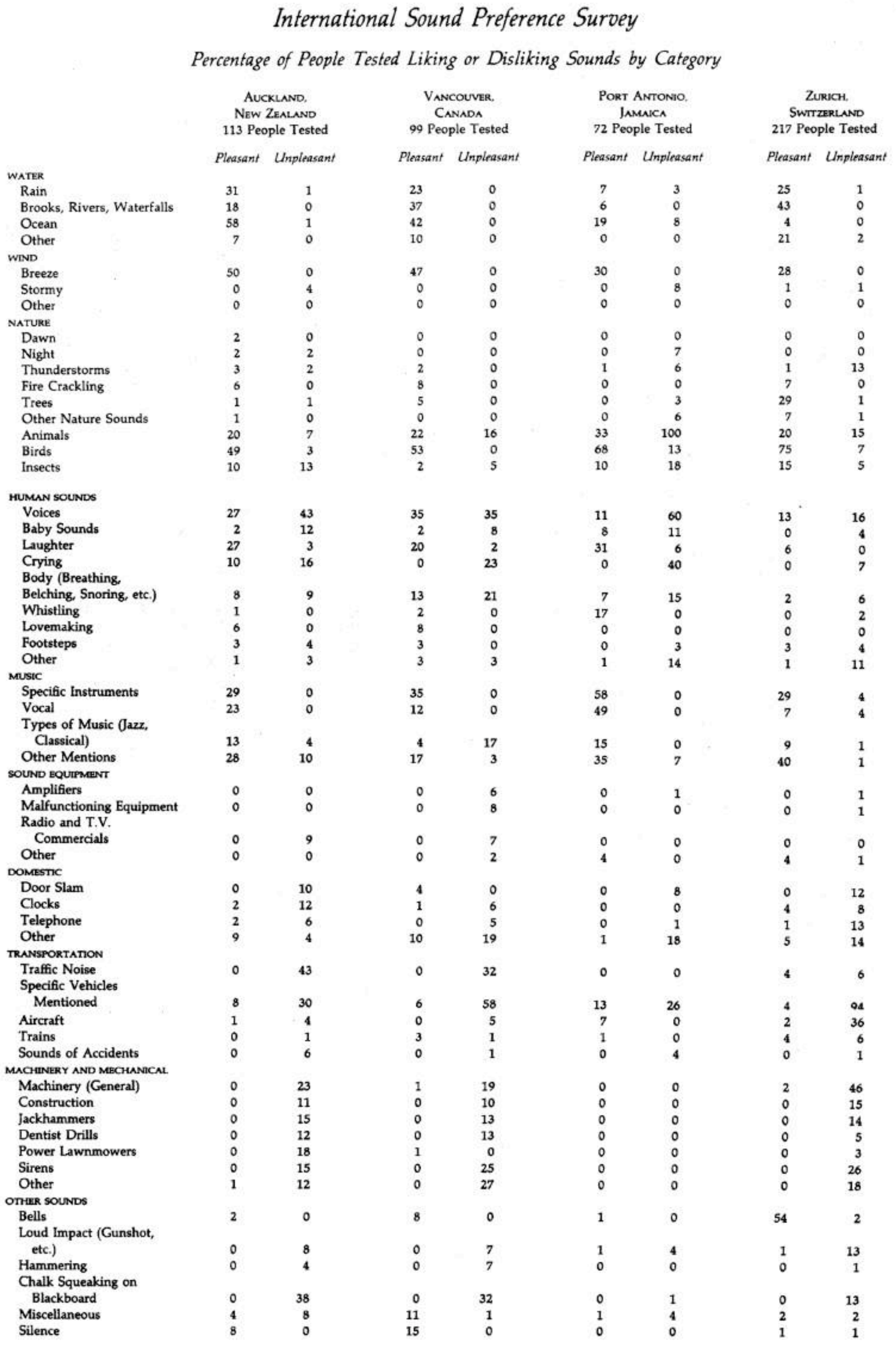
To the extent that the creation of psychoacoustic tension and release in music is important; and to the extent that electroacoustic music can operate outside the realm of pitch by focusing aesthetic attention on timbre, rhythm, and space; and to the extent that electroacoustic music is music, then, the consonance-dissonance continuum represents a fundamental area of inquiry regarding electroacoustic music.

[iv] or, as Julio Estrada calls it, the "discontinuum"

6 LISTENING TESTS

To test listeners' subjective responses to various sound objects in terms of perceptual dissonance, a battery of listening tests was constructed. As noted, dissonance-oriented listening tests in the past have historically asked listeners to rate the "dissonance"—or more often a specific and supposedly contributing factor, like "fusion," "roughness," or "purity"—of intervals played on one or more particular musical instruments in isolation (i.e., outside of a larger musical context). Based on statistical analysis of listeners' judgments, inductive conclusions were then typically drawn regarding the relationship between a specific contributing factor and the larger issue of musical dissonance.

Comprehensive tests of a similar nature have not yet been conducted for recordings of sound objects in general, at least to my knowledge. The closest related test of which I am aware was reported by composer and acoustic ecologist R. Murray Schafer in his book *The Soundscape* (1977, reprinted 1994), in which the author asked subjects in various world cities to respond by mailed questionnaire regarding their like or dislike of particular sounds. The results of the questionnaire are summarized in Appendix II of his book, and they are shown here in Table 6-1.



International Sound Preference Survey

Percentage of People Tested Liking or Disliking Sounds by Category

	AUCKLAND, NEW ZEALAND 113 People Tested		VANCOUVER, CANADA 99 People Tested		PORT ANTONIO, JAMAICA 72 People Tested		ZURICH, SWITZERLAND 217 People Tested	
	Pleasant	*Unpleasant*	*Pleasant*	*Unpleasant*	*Pleasant*	*Unpleasant*	*Pleasant*	*Unpleasant*
WATER								
Rain	31	1	23	0	7	3	25	1
Brooks, Rivers, Waterfalls	18	0	37	0	6	0	43	0
Ocean	58	1	42	0	19	8	4	0
Other	7	0	10	0	0	0	21	2
WIND								
Breeze	50	0	47	0	30	0	28	0
Stormy	0	4	0	0	0	8	1	1
Other	0	0	0	0	0	0	0	0
NATURE								
Dawn	2	0	0	0	0	0	0	0
Night	2	2	0	0	0	7	0	0
Thunderstorms	3	2	2	0	1	6	1	13
Fire Crackling	6	0	8	0	0	0	7	0
Trees	1	1	5	0	0	3	29	1
Other Nature Sounds	1	0	0	0	0	6	7	1
Animals	20	7	22	16	33	100	20	15
Birds	49	3	53	0	68	13	75	7
Insects	10	13	2	5	10	18	15	5
HUMAN SOUNDS								
Voices	27	43	35	35	11	60	13	16
Baby Sounds	2	12	2	8	8	11	0	4
Laughter	27	3	20	2	31	6	6	0
Crying	10	16	0	23	0	40	0	7
Body (Breathing, Belching, Snoring, etc.)	8	9	13	21	7	15	2	6
Whistling	1	0	2	0	17	0	0	2
Lovemaking	6	0	8	0	0	0	0	0
Footsteps	3	4	3	0	0	3	3	4
Other	1	3	3	3	1	14	1	11
MUSIC								
Specific Instruments	29	0	35	0	58	0	29	4
Vocal	23	0	12	0	49	0	7	4
Types of Music (Jazz, Classical)	13	4	4	17	15	0	9	1
Other Mentions	28	10	17	3	35	7	40	1
SOUND EQUIPMENT								
Amplifiers	0	0	0	6	0	1	0	1
Malfunctioning Equipment	0	0	0	8	0	0	0	1
Radio and T.V. Commercials	0	9	0	7	0	0	0	0
Other	0	0	0	2	4	0	4	1
DOMESTIC								
Door Slam	0	10	4	0	0	8	0	12
Clocks	2	12	1	6	0	0	4	8
Telephone	2	6	0	5	0	1	1	13
Other	9	4	10	19	1	18	5	14
TRANSPORTATION								
Traffic Noise	0	43	0	32	0	0	4	6
Specific Vehicles Mentioned	8	30	6	58	13	26	4	94
Aircraft	1	4	0	5	7	0	2	36
Trains	0	1	3	1	1	0	4	6
Sounds of Accidents	0	6	0	1	0	4	0	1
MACHINERY AND MECHANICAL								
Machinery (General)	0	23	1	19	0	0	2	46
Construction	0	11	0	10	0	0	0	15
Jackhammers	0	15	0	13	0	0	0	14
Dentist Drills	0	12	0	13	0	0	0	5
Power Lawnmowers	0	18	1	0	0	0	0	3
Sirens	0	15	0	25	0	0	0	26
Other	1	12	0	27	0	0	0	18
OTHER SOUNDS								
Bells	2	0	8	0	1	0	54	2
Loud Impact (Gunshot, etc.)	0	8	0	7	1	4	1	13
Hammering	0	4	0	7	0	0	0	1
Chalk Squeaking on Blackboard	0	38	0	32	0	1	0	13
Miscellaneous	4	8	11	1	1	4	2	2
Silence	8	0	15	0	0	0	1	1

Table 6-1. Results from R. Murray Schafer's International Sound Preference Survey (1977/1994).

Similar results were reported more recently by Yang and Kang (2005), in which urban-residing respondents typically indicated preference for “natural sounds” (i.e., sounds emanating from nature exclusively) over “urban sounds” (i.e., sounds emanating from typical urban soundscapes).

Problems naturally arise in any attempt to measure perceptually salient features of recorded sound objects. The most obvious lies in the mechanism for selecting and ordering sounds for the test. Many taxonomies of sound objects exist, as noted in Chapter 5. For the listening tests conducted here, the spirit of Schafer’s taxonomy was chosen as a rough basis, with several recorded sounds chosen from each broad category to cover a diverse range of sound objects. Because it is only possible to measure listeners’ subjective responses to a finite number of sounds, any taxonomy of sounds should ideally suffice; that of Schafer’s was chosen in particular because it is not concerned overtly with musical composition or analysis, and as such, offers a distinctly blank canvas from which to work.

6.1 Description of Tests

To test listeners’ judgments of sound-object dissonance, a protocol for listening tests was designed with the following goals in mind:

(1) The subjects should represent a consistent background in terms of musical training to eliminate a potentially prominent testing dimension;

(2) Subjects should be given identical time frames within which to complete the listening test, to eliminate another testing dimension; and

(3) The subject pool should be large enough to yield acceptable statistical confidence measures so that appropriate generalizations can be drawn.

A fundamental deviation from other protocols here is that the term “dissonance” was not defined *a priori* for test subjects; instead, subjects were free to assert whatever meaning(s) they deemed appropriate based on their musical experiences. This decision was based on the hypothesis that auditory processing and classification mechanisms might be different (perhaps significantly) for different categories of sound objects. For example, when listening to a simple major triad played on a piano, we know from the literature that the classical factors of tonal fusion, purity, roughness, and so forth must be involved. When unpitched, highly transient, unfamiliar, or unclassifiable sounds are played, it makes sense that listeners may either adjust their internal definition of “dissonance” and “consonance,” or perhaps recognize the existence of a broader spectrum of these perceptions than previously assumed. As such, the results of the test should ideally point to an underlying semantic meaning of the term "dissonance" as it relates to different musical contexts and varieties of sound objects.

Sixty-four sound objects were chosen at random from a variety of audio recordings, including synthesized sounds, commercial recordings, and sound-effects libraries, and these were grouped into thirty-two pairs to construct a forced-choice comparison test. Sound files ranged in duration from 2 sec to 19 sec. Sounds of similar durations were grouped in pairs for comparison to eliminate sound-object duration as a significant testing dimension.

The thirty-two pairs of sound objects were further subdivided randomly into four groups. For each group of eight pairs, subjects were asked to compare the corresponding sound objects by answering a series of related binary questions. Furthermore, after hearing each pair, subjects were asked which sound object was perceived as more dissonant.

The overall structure of the test is shown in Table 6-2, and the study questionnaire is reprinted in its entirety Appendix A. Each sound was followed by a short silence, and each pair was played twice in succession.

[Sound 1a] — [3 sec silence] — [Sound 1b] — [3 sec silence]
[Sound 1a] — [3 sec silence] — [Sound 1b] — [10 sec silence]

[Sound 2a] — [3 sec silence] — [Sound 2b] — [3 sec silence]
[Sound 2a] — [3 sec silence] — [Sound 2b] — [10 sec silence]
•
•

-

[Sound 32a] — [3 sec silence] — [Sound 32b] — [3 sec silence]
[Sound 32a] — [3 sec silence] — [Sound 32b]

Table 6-2. Structure of the listening test protocol.

At the completion of the listening tests, the subjective data were gathered and entered into a table. At the same time, each sound object was analyzed with a feature-extraction algorithm in the MATLAB environment with the goal of extracting objective qualities of each sound. Each sound file was then rated in terms of its peak amplitude in decibels, root-mean-square (RMS) amplitude in decibels, spectral centroid in Hz, spectral rolloff in Hz, hamonicity as a percentage, temporal centroid in sec, and temporal centroid as a percentage.

The first two of these objective measures attempts to provide a rough estimate of perceptual loudness. Of course, the measures used here are quite simplistic given the more advanced and accurate computational loudness models available, yet for the test here, they were thought to provide adequate resolution and accuracy. The spectral centroid, or spectral "center of gravity," is a widely accepted measure of perceptual brightness. Spectral rolloff is a measure of the frequency below which 85% of the spectral energy occurs; as such, it provides an additional computational measure of brightness and noisiness. The harmonicity measure used here approximates the extent to which a sound is perceived as harmonic by taking the second peak of the normalized autocorrelation (Arfib, Keiler, and Zölzer 2002). Finally, temporal centroid in sec and temporal centroid as a percentage define the transient nature of the sound object. Objects with relatively low values of these measures suggest a sound with a sharp attack, while low values suggest a subtler or relatively imperceptible attack.

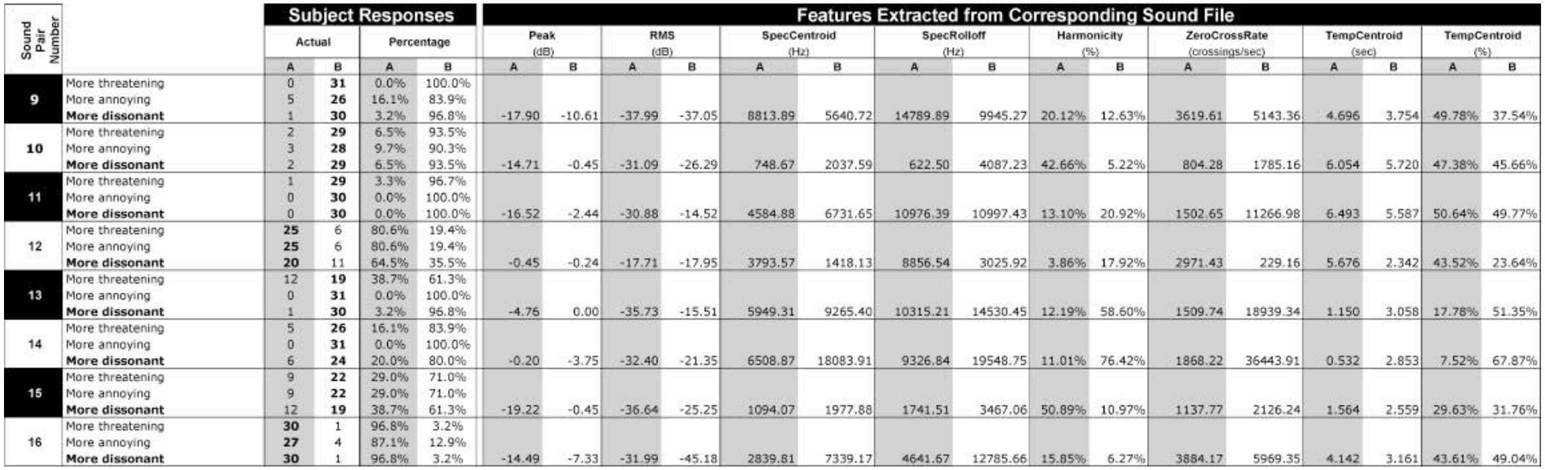
Once subjective and objective data were tabulated, conclusions could be drawn regarding their respective relevance to ratings of sound-object dissonance. Next, the results of the listening test are presented.

6.2 Results

Raw results for each set of sound-object listening tests are reported in Tables 6-3, 6-4, 6-5, and 6-6. Note that each table demarcates subjective human responses to each questionnaire (“Subject Responses”) from objectively computed features (“Features Extracted from Corresponding Sound File”).

Sound Pair Number		Subject Responses				Features Extracted from Corresponding Sound File															
		Actual		Percentage		Peak (dB)		RMS (dB)		SpecCentroid (Hz)		SpecRolloff (Hz)		Harmonicity (%)		ZeroCrossRate (crossings/sec)		TempCentroid (sec)		TempCentroid (%)	
		A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
1	More frustating	**13**	18	41.9%	58.1%																
	More difficult to understand	**25**	6	80.6%	19.4%																
	More intelligible	**10**	21	32.3%	67.7%																
	More dissonant	**22**	9	71.0%	29.0%	-1.34	-11.80	-21.10	-29.44	2403.70	1582.66	4086.31	3152.25	23.04%	14.81%	2606.71	1194.96	4.206	4.655	47.70%	42.98%
2	More frustating	**25**	6	80.6%	19.4%																
	More difficult to understand	**27**	4	87.1%	12.9%																
	More intelligible	**6**	25	19.4%	80.6%																
	More dissonant	**29**	2	93.5%	6.5%	-5.72	-22.29	-19.09	-38.55	1855.57	1259.89	3308.04	2224.39	29.48%	23.91%	1340.07	1269.75	7.458	5.121	48.26%	35.50%
3	More frustating	5	**26**	16.1%	83.9%																
	More difficult to understand	8	**23**	25.8%	74.2%																
	More intelligible	18	**13**	58.1%	41.9%																
	More dissonant	2	**29**	6.5%	93.5%	-12.45	-2.18	-31.64	-24.48	1769.23	8063.18	3009.55	15744.71	6.71%	11.85%	3607.56	11564.30	3.533	3.963	47.78%	55.08%
4	More frustating	**13**	18	41.9%	58.1%																
	More difficult to understand	**30**	1	96.8%	3.2%																
	More intelligible	**6**	25	19.4%	80.6%																
	More dissonant	**21**	10	67.7%	32.3%	-11.71	-1.35	-25.11	-28.03	3960.39	3988.91	7804.18	8562.45	1.44%	9.08%	4940.81	5179.90	3.034	2.507	49.83%	45.12%
5	More frustating	**14**	17	45.2%	54.8%																
	More difficult to understand	**22**	9	71.0%	29.0%																
	More intelligible	**9**	21	30.0%	70.0%																
	More dissonant	**21**	9	70.0%	30.0%	-5.51	-20.63	-26.57	-36.10	8195.68	1255.15	21783.57	1233.67	48.38%	25.13%	15721.26	1215.84	5.793	5.420	67.23%	55.22%
6	More frustating	4	**27**	12.9%	87.1%																
	More difficult to understand	27	**4**	87.1%	12.9%																
	More intelligible	9	**22**	29.0%	71.0%																
	More dissonant	3	**28**	9.7%	90.3%	-14.27	-15.55	-33.13	-25.90	3281.04	4487.28	6601.17	6919.89	6.17%	77.36%	3265.61	5410.62	7.356	6.087	57.36%	49.84%
7	More frustating	5	**26**	16.1%	83.9%																
	More difficult to understand	8	**23**	25.8%	74.2%																
	More intelligible	24	**7**	77.4%	22.6%																
	More dissonant	12	**19**	38.7%	61.3%	-0.45	-4.91	-23.67	-22.91	1609.17	2959.88	2004.66	4387.89	21.32%	15.50%	885.04	4154.51	4.813	4.801	50.68%	48.29%
8	More frustating	17	**14**	54.8%	45.2%																
	More difficult to understand	8	**23**	25.8%	74.2%																
	More intelligible	23	**8**	74.2%	25.8%																
	More dissonant	7	**24**	22.6%	77.4%	-7.29	-15.12	-25.32	-40.03	2064.02	3908.10	3164.02	5620.95	7.59%	15.96%	1731.41	4503.00	5.134	2.994	50.93%	21.63%

Table 6-3. Subject responses and features for each sound file, pairs 1–8.



Sound Pair Number		Subject Responses				Features Extracted from Corresponding Sound File															
		Actual		Percentage		Peak (dB)		RMS (dB)		SpecCentroid (Hz)		SpecRolloff (Hz)		Harmonicity (%)		ZeroCrossRate (crossings/sec)		TempCentroid (sec)		TempCentroid (%)	
		A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
9	More threatening	0	**31**	0.0%	100.0%																
	More annoying	5	**26**	16.1%	83.9%																
	More dissonant	1	**30**	3.2%	96.8%	-17.90	-10.61	-37.99	-37.05	8813.89	5640.72	14789.89	9945.27	20.12%	12.63%	3619.61	5143.36	4.696	3.754	49.78%	37.54%
10	More threatening	2	**29**	6.5%	93.5%																
	More annoying	3	**28**	9.7%	90.3%																
	More dissonant	2	**29**	6.5%	93.5%	-14.71	-0.45	-31.09	-26.29	748.67	2037.59	622.50	4087.23	42.66%	5.22%	804.28	1785.16	6.054	5.720	47.38%	45.66%
11	More threatening	1	**29**	3.3%	96.7%																
	More annoying	0	**30**	0.0%	100.0%																
	More dissonant	0	**30**	0.0%	100.0%	-16.52	-2.44	-30.88	-14.52	4584.88	6731.65	10976.39	10997.43	13.10%	20.92%	1502.65	11266.98	6.493	5.587	50.64%	49.77%
12	More threatening	**25**	6	80.6%	19.4%																
	More annoying	**25**	6	80.6%	19.4%																
	More dissonant	**20**	11	64.5%	35.5%	-0.45	-0.24	-17.71	-17.95	3793.57	1418.13	8856.54	3025.92	3.86%	17.92%	2971.43	229.16	5.676	2.342	43.52%	23.64%
13	More threatening	12	**19**	38.7%	61.3%																
	More annoying	0	**31**	0.0%	100.0%																
	More dissonant	1	**30**	3.2%	96.8%	-4.76	0.00	-35.73	-15.51	5949.31	9265.40	10315.21	14530.45	12.19%	58.60%	1509.74	18939.34	1.150	3.058	17.78%	51.35%
14	More threatening	5	**26**	16.1%	83.9%																
	More annoying	0	**31**	0.0%	100.0%																
	More dissonant	6	**24**	20.0%	80.0%	-0.20	-3.75	-32.40	-21.35	6508.87	18083.91	9326.84	19548.75	11.01%	76.42%	1868.22	36443.91	0.532	2.853	7.52%	67.87%
15	More threatening	9	**22**	29.0%	71.0%																
	More annoying	9	**22**	29.0%	71.0%																
	More dissonant	12	**19**	38.7%	61.3%	-19.22	-0.45	-36.64	-25.25	1094.07	1977.88	1741.51	3467.06	50.89%	10.97%	1137.77	2126.24	1.564	2.559	29.63%	31.76%
16	More threatening	**30**	1	96.8%	3.2%																
	More annoying	**27**	4	87.1%	12.9%																
	More dissonant	**30**	1	96.8%	3.2%	-14.49	-7.33	-31.99	-45.18	2839.81	7339.17	4641.67	12785.66	15.85%	6.27%	3884.17	5969.35	4.142	3.161	43.61%	49.04%

Table 6-4. Subject responses and features for each sound file, pairs 9–16.

Sound Pair Number		Subject Responses				Features Extracted from Corresponding Sound File															
		Actual		Percentage		Peak (dB)		RMS (dB)		SpecCentroid (Hz)		SpecRolloff (Hz)		Harmonicity (%)		ZeroCrossRate (crossings/sec)		TempCentroid (sec)		TempCentroid (%)	
		A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
17	More easily recognizable	29	**2**	93.5%	6.5%	-1.18	-0.47	-26.94	-26.56	1699.42	3044.67	3402.36	5505.11	7.26%	5.56%	1768.02	2140.68	0.913	3.212	10.67%	32.98%
	More predictable	27	**4**	87.1%	12.9%																
	More consistent	19	**12**	61.3%	38.7%																
	More dissonant	8	**28**	22.2%	77.8%																
18	More easily recognizable	**0**	31	0.0%	100.0%	-3.69	-2.66	-16.60	-25.18	6625.99	6371.49	13797.00	12447.91	69.54%	1.24%	2376.50	7221.51	2.031	1.427	50.77%	28.71%
	More predictable	**6**	25	19.4%	80.6%																
	More consistent	**21**	10	67.7%	32.3%																
	More dissonant	**25**	6	80.6%	19.4%																
19	More easily recognizable	5	**26**	16.1%	83.9%	-8.77	-6.58	-24.65	-31.20	4986.93	6372.20	10478.20	12447.90	6.04%	1.25%	5412.18	4432.05	4.307	1.434	51.87%	17.75%
	More predictable	10	**21**	32.3%	67.7%																
	More consistent	16	**15**	51.6%	48.4%																
	More dissonant	13	**18**	41.9%	58.1%																
20	More easily recognizable	**1**	30	3.2%	96.8%	-9.43	-3.14	-38.30	-25.44	5983.41	3706.43	10193.44	9069.07	17.20%	3.09%	4678.39	3926.37	2.807	5.558	20.28%	40.11%
	More predictable	**7**	24	22.6%	77.4%																
	More consistent	**13**	18	41.9%	58.1%																
	More dissonant	**30**	1	96.8%	3.2%																
21	More easily recognizable	**30**	1	96.8%	3.2%	-4.18	-11.83	-31.41	-32.47	2180.11	5465.08	4007.96	14887.67	11.68%	15.10%	1499.14	3492.28	5.412	3.849	55.51%	42.75%
	More predictable	**20**	11	64.5%	35.5%																
	More consistent	**8**	23	25.8%	74.2%																
	More dissonant	**18**	12	60.0%	40.0%																
22	More easily recognizable	**21**	10	67.7%	32.3%	-2.55	-2.13	-18.76	-18.59	2374.44	4388.97	5438.23	16999.79	1.94%	22.71%	1539.90	5602.57	5.126	4.211	46.77%	46.81%
	More predictable	**30**	1	96.8%	3.2%																
	More consistent	**27**	4	87.1%	12.9%																
	More dissonant	**22**	9	71.0%	29.0%																
23	More easily recognizable	12	**19**	38.7%	61.3%	-1.47	-2.58	-29.99	-20.08	7499.99	3108.61	13824.79	4434.19	8.02%	8.52%	10951.61	2323.20	7.145	1.885	79.40%	33.71%
	More predictable	19	**12**	61.3%	38.7%																
	More consistent	19	**12**	61.3%	38.7%																
	More dissonant	13	**18**	41.9%	58.1%																
24	More easily recognizable	**1**	30	3.2%	96.8%	-0.45	-3.88	-14.48	-9.67	1096.61	2096.56	1880.96	4016.10	29.08%	31.16%	530.79	1150.69	5.887	4.811	51.86%	38.07%
	More predictable	**3**	28	9.7%	90.3%																
	More consistent	**2**	29	6.5%	93.5%																
	More dissonant	**18**	13	58.1%	41.9%																

Table 6-5. Subject responses and features for each sound file, pairs 17–24.

Sound Pair Number		Subject Responses				Features Extracted from Corresponding Sound File															
		Actual		Percentage		Peak (dB)		RMS (dB)		SpecCentroid (Hz)		SpecRolloff (Hz)		Harmonicity (%)		ZeroCrossRate (crossings/sec)		TempCentroid (sec)		TempCentroid (%)	
		A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
25	More pitched	**13**	18	41.9%	58.1%																
	Smoother	**24**	7	77.4%	22.6%																
	More regular	**18**	13	58.1%	41.9%																
	More dissonant	**26**	5	83.9%	16.1%	-10.86	-2.60	-25.11	-29.92	3991.86	6759.79	7769.05	9742.89	1.49%	15.70%	5091.71	13683.54	4.586	3.552	50.01%	37.25%
26	More pitched	**12**	19	38.7%	61.3%																
	Smoother	**12**	19	38.7%	61.3%																
	More regular	**9**	21	30.0%	70.0%																
	More dissonant	**25**	6	80.6%	19.4%	-9.63	-9.75	-27.47	-28.89	4217.47	8318.37	8421.37	15712.89	4.91%	48.78%	5232.19	4428.76	5.340	5.321	53.79%	57.06%
27	More pitched	6	**25**	19.4%	80.6%																
	Smoother	21	**10**	67.7%	32.3%																
	More regular	20	**11**	64.5%	35.5%																
	More dissonant	11	**20**	35.5%	64.5%	-4.35	-14.23	-29.72	-37.60	5238.82	2956.69	10025.79	5503.08	4.80%	3.61%	5936.40	3033.04	5.699	1.568	53.61%	15.56%
28	More pitched	5	**26**	16.1%	83.9%																
	Smoother	22	**9**	71.0%	29.0%																
	More regular	14	**17**	45.2%	54.8%																
	More dissonant	1	**30**	3.2%	96.8%	-14.07	-3.00	-39.12	-18.53	8393.39	6050.67	14559.25	10895.57	17.43%	79.25%	7614.94	12937.93	1.659	2.288	29.22%	48.32%
29	More pitched	22	**9**	71.0%	29.0%																
	Smoother	29	**2**	93.5%	6.5%																
	More regular	18	**13**	58.1%	41.9%																
	More dissonant	3	**28**	9.7%	90.3%	-5.33	-13.40	-20.22	-20.48	509.90	2095.71	795.42	4013.39	82.08%	31.20%	239.88	880.35	5.315	4.812	32.38%	28.30%
30	More pitched	6	**25**	19.4%	80.6%																
	Smoother	14	**17**	45.2%	54.8%																
	More regular	24	**7**	77.4%	22.6%																
	More dissonant	8	**23**	25.8%	74.2%	-6.18	-1.72	-26.03	-23.80	1887.02	4401.11	4039.95	9606.09	6.88%	3.44%	1443.89	3291.68	6.668	8.194	53.65%	58.61%
31	More pitched	27	**4**	87.1%	12.9%																
	Smoother	30	**1**	96.8%	3.2%																
	More regular	23	**8**	74.2%	25.8%																
	More dissonant	0	**31**	0.0%	100.0%	-18.78	0.00	-30.19	-24.85	1129.72	11060.85	1543.83	15828.89	44.55%	35.89%	936.61	7946.80	4.167	5.310	32.96%	90.19%
32	More pitched	**7**	24	22.6%	77.4%																
	Smoother	**2**	29	6.5%	93.5%																
	More regular	**5**	26	16.1%	83.9%																
	More dissonant	**26**	5	83.9%	16.1%	0.00	-3.20	0.00	-18.86	6564.19	530.86	14502.79	819.73	7.54%	31.97%	2455.83	573.21	2.806	4.396	50.00%	32.94%

Table 6-6. Subject responses and features for each sound file, pairs 25–32.

Subjective Data

Once both subjective and objective data were gathered, a preliminary analysis could be divided into two phases. In the first phase, attempts were made to discern the extent to which the (subjective) answers of each question in the questionnaire possibly affected the relative rating of “dissonant” by the subjects. For example, I wanted to understand to what extent a rating of “more frustrating to hear” would correspond to a rating of “more dissonant” for a given sound objects. In the second phase, attempts were made to correlate objectively computed values (e.g., spectral centroid and spectral rolloff) to subjective ratings of sound-object dissonance.

The next figures graphically illustrate the results of the objective-ratings aspect of the listening test. Figure 6-1 shows the cumulative ratings of sound pairs 1–8, sorted according to increasing dissonance rating. Along with the cumulative dissonance rating of each sound object, the cumulative rating of the answers to teach of the questions “Which sound is more frustrating to hear?”, “Which sound is more difficult to understand?”, and “Which sound is more intelligible?” are shown.

In each figure, the *y*-axis shows the raw rating for each series of questions, i.e., the raw number of subjects that answered in the affirmative the question corresponding to the sound object shown along the *x* axis. To produce the curves, the eight sounds judged as “more dissonant” in each pair were sorted in increasing order of their raw scores (i.e., the number of respondents that rated that sound object as more dissonant). (Note that, as a result, the raw score for the dissonance curve then by definition is always greater than one-half the total number of respondents). Next, the corresponding raw ratings for each question were plotted overlaid on the dissonance curve. In the legend of Figure 6-1, “Frust” refers to the question “Which sound is more frustrating to hear?”, “Diff” refers to the question “Which sound is more difficult to understand?”, and “Intell” corresponds to the question “Which sound is more intelligible?”.

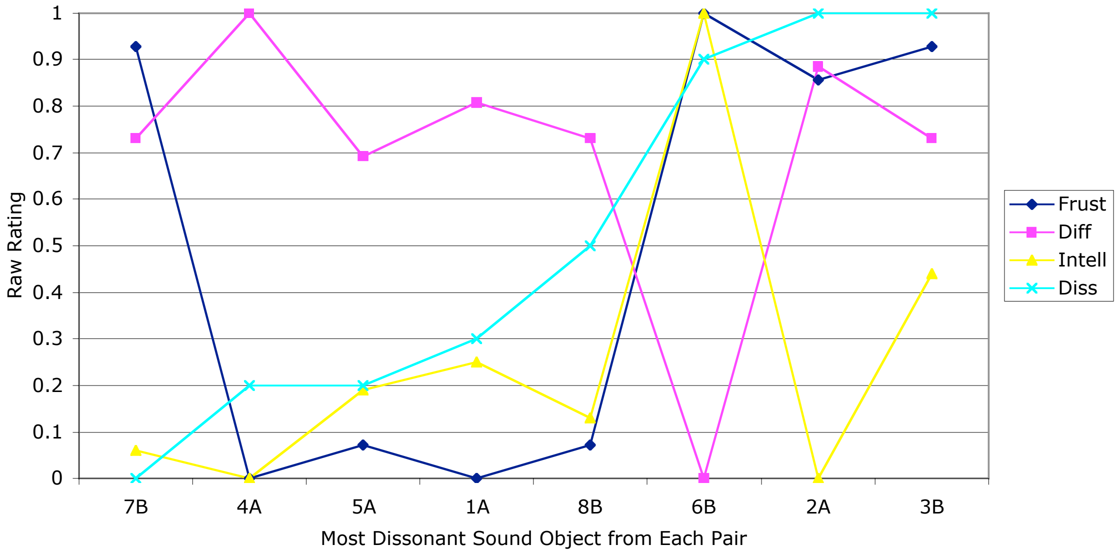


Figure 6–1. Normalized subjective ratings of sound pairs 1–8.

Figure 6-2 shows the same results with a cubic-spline interpolation of the subjectively reported data. Comparing affirmative answers to a variety of questions is of course akin to comparing psychological apples and oranges, so to speak, and thus perhaps moreso than any precise statistical comparisons, it was judged that qualitative, graphical comparison of subjective-response curves might provide as much or more insight into the problem of sound-object dissonance assessment and its psychological contributors than direct statistical computations on the reported data.

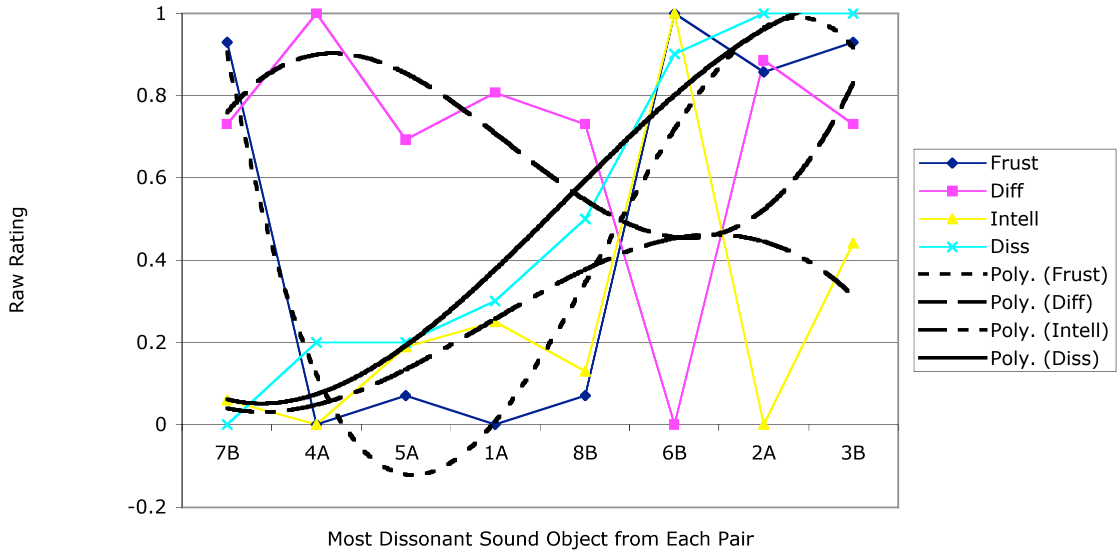


Figure 6–2. Cubic-spline interpolation of data from Figure 6-1.

Upon examining Figure 6-2, one could argue that the intelligibility curve most closely models the dissonance-rating curve. Once again, comparing these essentially unit-less curves in a meaningful manner is difficult; however, we can examine the global trend of each curve to subjectively point to relative similarities/correspondences.

Figure 6-3 reports responses to the next sequence of sound-object pairs, labeled 9–16. The raw data reported here correspond to the questions “Which sound is more threatening?” (“Threat”) and “Which sound is more annoying?”.

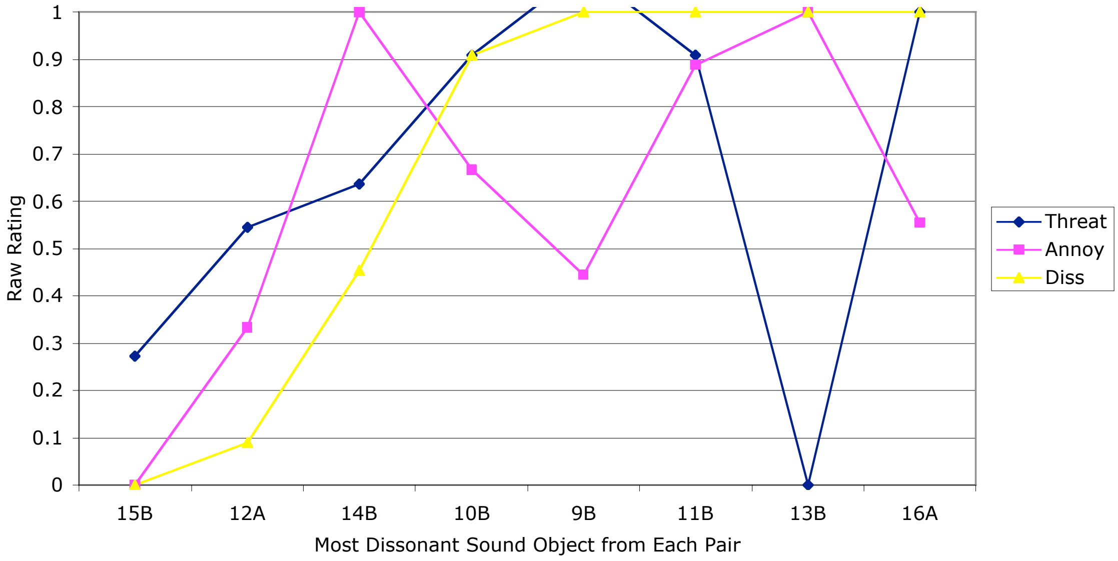


Figure 6–3. Subject ratings of sound pairs 9–16.

Figure 6-4 shows a cubic-spline interpolation of the data in Figure 6-3, indicating a relatively strong general trend, particularly if the "Threat" score for sound object 13B is omitted.

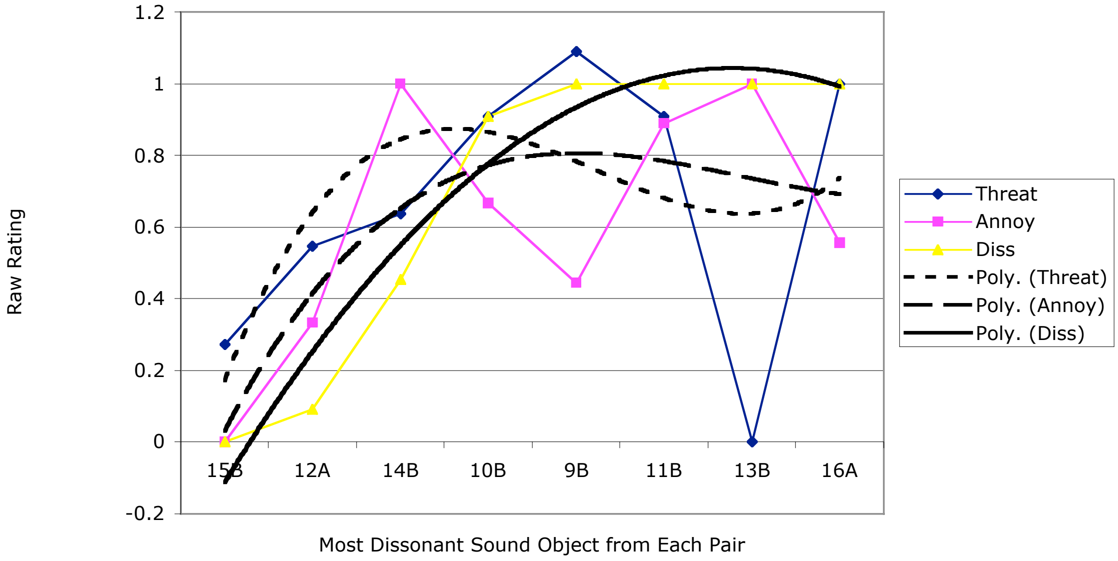


Figure 6–4. Cubic-spline interpolation of data from Figure 6-3.

Figure 6-5 illustrates subject ratings of sound objects 17–24, in which relatively little correspondence among curves can be observed. The legend refers to raw ratings for the questions “Which sound is more easily recognizable?” (“Recog”), “Which sound is more predictable?” (“Predict”), and “Which sound is more consistent?” (“Constnt”).

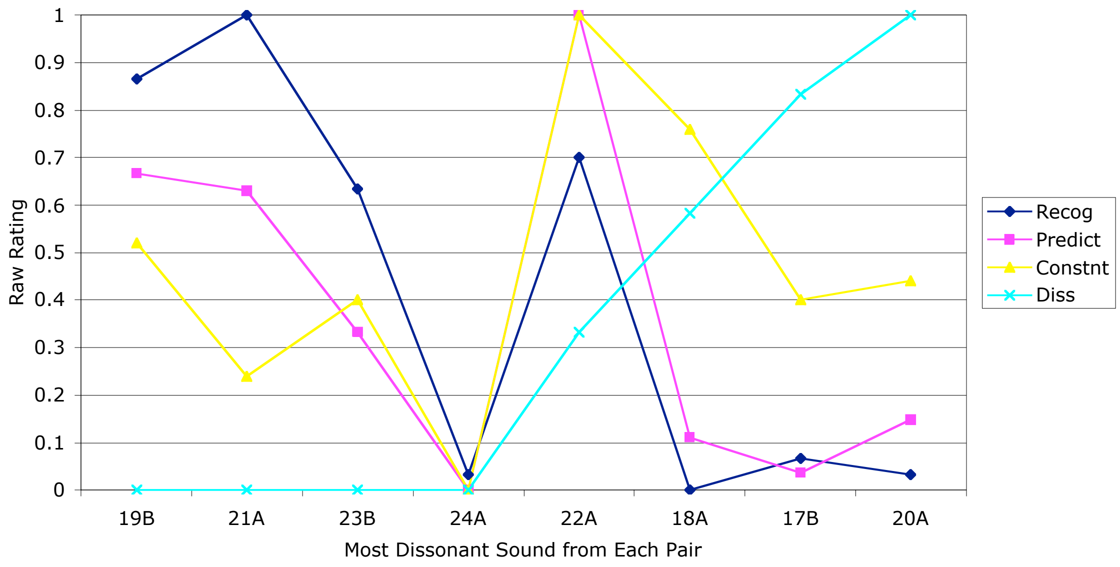


Figure 6–5. Subject ratings of sound pairs 17–24.

Figure 6-6 shows a cubic-spline interpolation of the data, which once again highlights the low degree of visual correspondence among the curves.

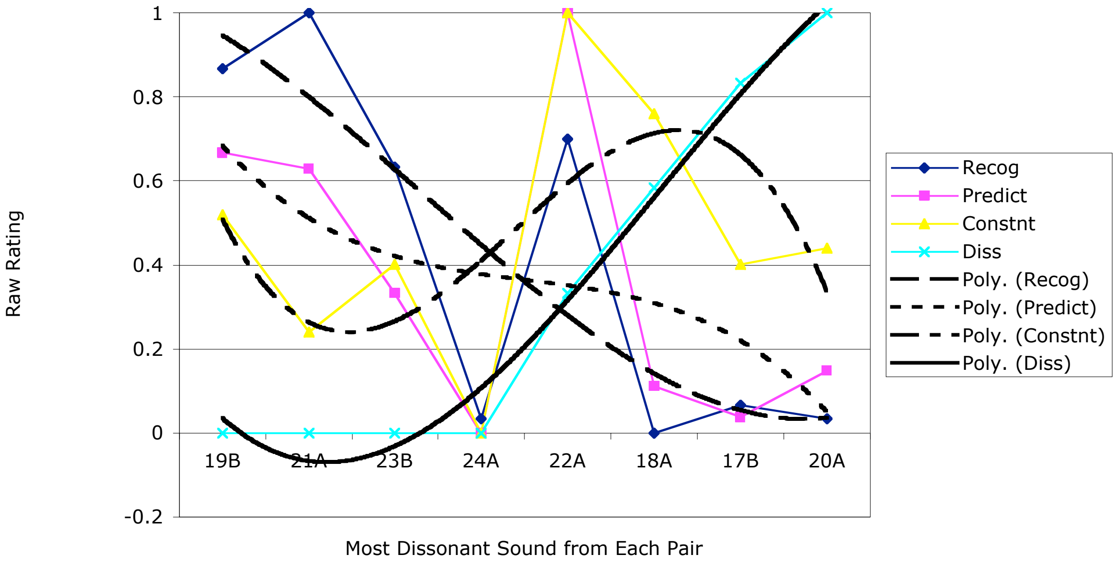


Figure 6–6. Cubic-spline interpolation of data from Figure 6-5.

Next, Figure 6-7 illustrates subject responses for the final group of sound objects, pairs 25–32. The legend refers to the questions “Which sound is more pitched?” (“Pitched”), “Which sound is smoother?” (“Smoother”), and “Which sound is more regular?” (“Regular”).

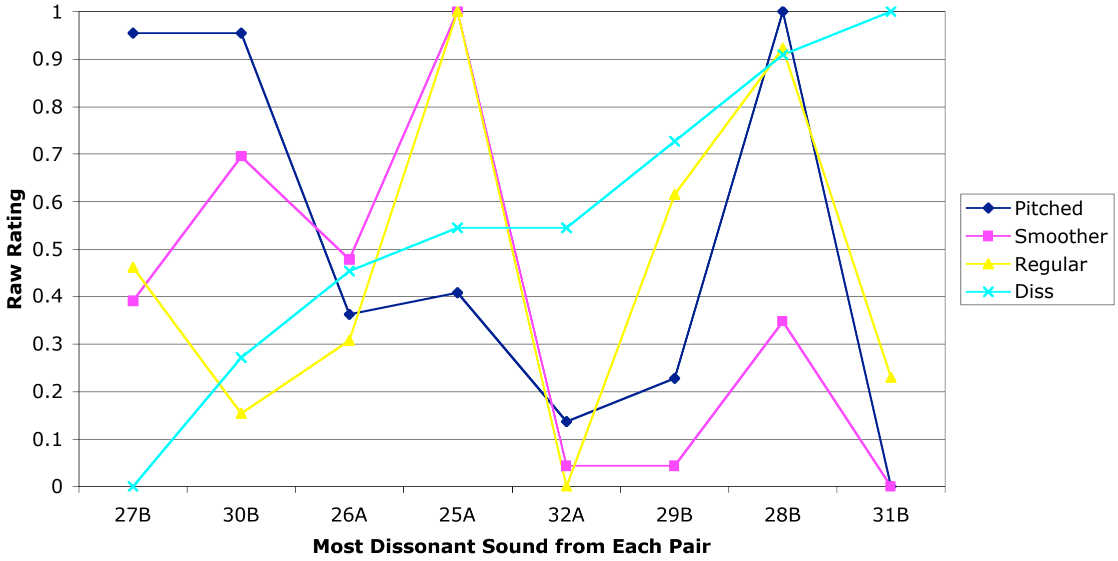


Figure 6–7. Subject ratings of sound pairs 25–32.

Figure 6-8 shows the data of Figure 6-7 with overlaid cubic-spline interpolations. The curves show a somewhat surprising lack of correspondence, except in the middle region, for which the slope of the dissonance curve seems well matched to the slope of the “Pitched” and “Regular” curves.

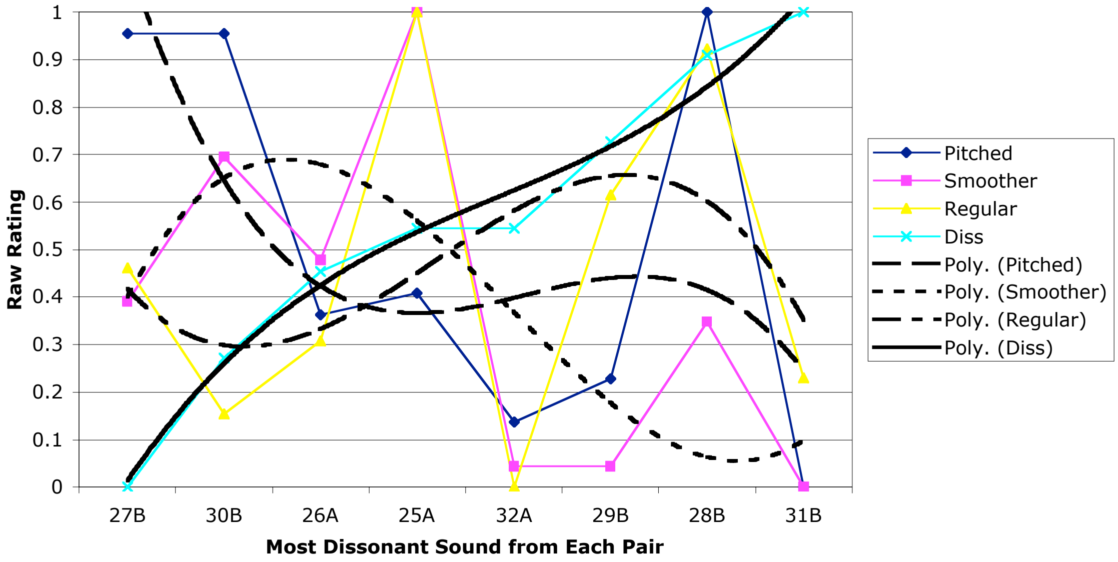


Figure 6–8. Cubic-spline interpolation of data from Figure 6-7.

Objective Data

In the second phase of results-gathering and preliminary data analysis, the goal was to determine correspondences among objectively computed values for the sound file in each pair judged as "more dissonant" to subjective ratings of sound-object dissonance. A simple function called `extractFeatures` was written in MATLAB to compute and return values of the peak sample value, root-mean-square (RMS) sample level, spectral centroid, spectral rolloff, harmonicity, zero-crossing rate, temporal centroid in Hz, and temporal centroid as a percentage of the sound-object duration. The results are shown in Table 6-8.

Sound #	Diss	Peak (dB)	RMS (dB)	SpecCentroid (Hz)	SpecRolloff (Hz)	Harmonicity (%)	ZeroCrossRate (crossings/sec)	TempCentroid (sec)	TempCentroid% (%)
1A	22	-1.34	-21.10	2403.70	4086.31	23.04%	2606.71	4.206	47.70%
2A	29	-5.72	-19.09	1855.57	3308.04	29.48%	1340.07	7.458	48.26%
3B	29	-2.18	-24.48	8063.18	15744.71	11.85%	11564.30	3.963	55.08%
4A	21	-11.71	-25.11	3960.39	7804.18	1.44%	4940.81	3.034	49.83%
5A	21	-5.51	-26.57	8195.68	21783.57	48.38%	15721.26	5.793	67.23%
6B	28	-15.55	-25.90	4487.28	6919.89	77.36%	5410.62	6.087	49.84%
7B	19	-4.91	-22.91	2959.88	4387.89	15.50%	4154.51	4.801	48.29%
8B	24	-15.12	-40.03	3908.10	5620.95	15.96%	4503.00	2.994	21.63%
9B	30	-10.61	-37.05	5640.72	9945.27	12.63%	5143.36	3.754	37.54%
10B	29	-0.45	-26.29	2037.59	4087.23	5.22%	1785.16	5.720	45.66%
11B	30	-2.44	-14.52	6731.65	10997.43	20.92%	11266.98	5.587	49.77%
12A	20	-0.45	-17.71	3793.57	8856.54	3.86%	2971.43	5.676	43.52%
13B	30	0.00	-15.51	9265.40	14530.45	58.60%	18939.34	3.058	51.35%
14B	24	-3.75	-21.35	18083.91	19548.75	76.42%	36443.91	2.853	67.87%
15B	19	-0.45	-25.25	1977.88	3467.06	10.97%	2126.24	2.559	31.76%
16A	30	-14.49	-31.99	2839.81	4641.67	15.85%	3884.17	4.142	43.61%
17B	28	-0.47	-26.56	3044.67	5505.11	5.56%	2140.68	3.212	32.98%
18A	25	-3.69	-16.60	6625.99	13797.00	69.54%	2376.50	2.031	50.77%
19B	18	-6.58	-31.20	6372.20	12447.90	1.25%	4432.05	1.434	17.75%
20A	30	-9.43	-38.30	5983.41	10193.44	17.20%	4678.39	2.807	20.28%
21A	18	-4.18	-31.41	2180.11	4007.96	11.68%	1499.14	5.412	55.51%
22A	22	-2.55	-18.76	2374.44	5438.23	1.94%	1539.90	5.126	46.77%
23B	18	-2.58	-20.08	3108.61	4434.19	8.52%	2323.20	1.885	33.71%
24A	18	-0.45	-14.48	1096.61	1880.96	29.08%	530.79	5.887	51.86%
25A	26	-10.86	-25.11	3991.86	7769.05	1.49%	5091.71	4.586	50.01%
26A	25	-9.63	-27.47	4217.47	8421.37	4.91%	5232.19	5.340	53.79%
27B	20	-14.23	-37.60	2956.69	5503.08	3.61%	3033.04	1.568	15.56%
28B	30	-3.00	-18.53	6050.67	10895.57	79.25%	12937.93	2.288	48.32%
29B	28	-13.40	-20.48	2095.71	4013.39	31.20%	880.35	4.812	28.30%
30B	23	-1.72	-23.80	4401.11	9606.09	3.44%	3291.68	8.194	58.61%
31B	31	0.00	-24.85	11060.85	15828.89	35.89%	7946.80	5.310	90.19%
32A	26	0.00	0.00	6564.19	14502.79	7.54%	2455.83	2.806	50.00%

Table 6-7. Feature evaluation of sounds judged as the "most dissonant" from each pair.

However, a problem arises upon examination of the data in Table 6-7, namely, that the feature values have different units, eliminating the value of direct numerical comparison. Thus, the dissonance ratings and feature values were normalized to unity, facilitating the comparison of features with unlike units. The results of this normalization are shown in Table 6-8.

Finally, to prepare the data for the curve-comparison technique previously used for the subjective portion of the data, the data were sorted according to increasing dissonance rating of each sound object. The results of this normalization and sorting operation are shown in Table 6-9.

Sound #	Diss	Peak	RMS	SpecCentroid	SpecRolloff	Harmonicity	ZeroCrossRate	TempCentroid	TempCentroid%
1A	**0.3077**	0.8285	0.0789	0.0769	0.1108	0.2794	0.0578	0.4101	0.4307
2A	**0.8462**	0.4210	0.1021	0.0447	0.0717	0.3619	0.0225	0.8911	0.4382
3B	**0.8462**	0.7337	0.0502	0.4101	0.6966	0.1359	0.3072	0.3741	0.5296
4A	**0.2308**	0.1115	0.0460	0.1686	0.2976	0.0025	0.1228	0.2366	0.4592
5A	**0.2308**	0.4363	0.0373	0.4179	1.0000	0.6042	0.4230	0.6448	0.6924
6B	**0.7692**	0.0000	0.0411	0.1996	0.2532	0.9758	0.1359	0.6884	0.4594
7B	**0.0769**	0.4819	0.0622	0.1097	0.1260	0.1828	0.1009	0.4980	0.4386
8B	**0.4615**	0.0102	0.0000	0.1655	0.1879	0.1885	0.1106	0.2308	0.0813
9B	**0.9231**	0.1532	0.0041	0.2675	0.4052	0.1458	0.1284	0.3432	0.2946
10B	**0.8462**	0.9399	0.0389	0.0554	0.1109	0.0509	0.0349	0.6340	0.4033
11B	**0.9231**	0.7057	0.1797	0.3317	0.4581	0.2521	0.2989	0.6144	0.4584
12A	**0.1538**	0.9398	0.1214	0.1588	0.3505	0.0335	0.0680	0.6275	0.3746
13B	**0.9231**	1.0000	0.1593	0.4809	0.6356	0.7353	0.5126	0.2402	0.4795
14B	**0.4615**	0.5792	0.0764	1.0000	0.8877	0.9637	1.0000	0.2100	0.7009
15B	**0.0769**	0.9399	0.0451	0.0519	0.0797	0.1246	0.0444	0.1663	0.2170
16A	**0.9231**	0.0259	0.0153	0.1026	0.1387	0.1872	0.0934	0.4005	0.3759
17B	**0.7692**	0.9363	0.0374	0.1147	0.1821	0.0553	0.0448	0.2631	0.2335
18A	**0.5385**	0.5848	0.1393	0.3255	0.5987	0.8756	0.0514	0.0882	0.4717
19B	**0.0000**	0.3625	0.0177	0.3106	0.5309	0.0000	0.1086	0.0000	0.0293
20A	**0.9231**	0.2049	0.0022	0.2877	0.4177	0.2045	0.1155	0.2031	0.0632
21A	**0.0000**	0.5415	0.0171	0.0638	0.1069	0.1338	0.0270	0.5884	0.5353
22A	**0.3077**	0.6951	0.1064	0.0752	0.1787	0.0088	0.0281	0.5462	0.4182
23B	**0.0000**	0.6911	0.0900	0.1184	0.1283	0.0932	0.0499	0.0668	0.2432
24A	**0.0000**	0.9399	0.1806	0.0000	0.0000	0.3568	0.0000	0.6588	0.4865
25A	**0.6154**	0.1434	0.0460	0.1704	0.2958	0.0030	0.1270	0.4662	0.4617
26A	**0.5385**	0.1957	0.0327	0.1837	0.3286	0.0469	0.1309	0.5778	0.5123
27B	**0.1538**	0.0328	0.0032	0.1095	0.1820	0.0303	0.0697	0.0198	0.0000
28B	**0.9231**	0.6496	0.1096	0.2916	0.4529	1.0000	0.3455	0.1263	0.4390
29B	**0.7692**	0.0562	0.0855	0.0588	0.1071	0.3841	0.0097	0.4997	0.1707
30B	**0.3846**	0.7838	0.0551	0.1945	0.3881	0.0280	0.0769	1.0000	0.5768
31B	**1.0000**	0.9999	0.0477	0.5866	0.7008	0.4441	0.2065	0.5735	1.0000
32A	**0.6154**	1.0000	1.0000	0.3219	0.6342	0.0806	0.0536	0.2029	0.4615

Table 6-8. Feature evaluation of sounds judged as the "most dissonant" from each pair. Dissonance ratings and computed features are here normalized to a maximum of unity to facilitate comparison across units.

Sound #	Diss	Peak	RMS	SpecCentroid	SpecRolloff	Harmonicity	ZeroCrossRate	TempCentroid	TempCentroid%
19B	**0.0000**	0.3625	0.0177	0.3106	0.5309	0.0000	0.1086	0.0000	0.0293
21A	**0.0000**	0.5415	0.0171	0.0638	0.1069	0.1338	0.0270	0.5884	0.5353
23B	**0.0000**	0.6911	0.0900	0.1184	0.1283	0.0932	0.0499	0.0668	0.2432
24A	**0.0000**	0.9399	0.1806	0.0000	0.0000	0.3568	0.0000	0.6588	0.4865
7B	**0.0769**	0.4819	0.0622	0.1097	0.1260	0.1828	0.1009	0.4980	0.4386
15B	**0.0769**	0.9399	0.0451	0.0519	0.0797	0.1246	0.0444	0.1663	0.2170
12A	**0.1538**	0.9398	0.1214	0.1588	0.3505	0.0335	0.0680	0.6275	0.3746
27B	**0.1538**	0.0328	0.0032	0.1095	0.1820	0.0303	0.0697	0.0198	0.0000
4A	**0.2308**	0.1115	0.0460	0.1686	0.2976	0.0025	0.1228	0.2366	0.4592
5A	**0.2308**	0.4363	0.0373	0.4179	1.0000	0.6042	0.4230	0.6448	0.6924
1A	**0.3077**	0.8285	0.0789	0.0769	0.1108	0.2794	0.0578	0.4101	0.4307
22A	**0.3077**	0.6951	0.1064	0.0752	0.1787	0.0088	0.0281	0.5462	0.4182
30B	**0.3846**	0.7838	0.0551	0.1945	0.3881	0.0280	0.0769	1.0000	0.5768
8B	**0.4615**	0.0102	0.0000	0.1655	0.1879	0.1885	0.1106	0.2308	0.0813
14B	**0.4615**	0.5792	0.0764	1.0000	0.8877	0.9637	1.0000	0.2100	0.7009
18A	**0.5385**	0.5848	0.1393	0.3255	0.5987	0.8756	0.0514	0.0882	0.4717
26A	**0.5385**	0.1957	0.0327	0.1837	0.3286	0.0469	0.1309	0.5778	0.5123
25A	**0.6154**	0.1434	0.0460	0.1704	0.2958	0.0030	0.1270	0.4662	0.4617
32A	**0.6154**	1.0000	1.0000	0.3219	0.6342	0.0806	0.0536	0.2029	0.4615
6B	**0.7692**	0.0000	0.0411	0.1996	0.2532	0.9758	0.1359	0.6884	0.4594
17B	**0.7692**	0.9363	0.0374	0.1147	0.1821	0.0553	0.0448	0.2631	0.2335
29B	**0.7692**	0.0562	0.0855	0.0588	0.1071	0.3841	0.0097	0.4997	0.1707
2A	**0.8462**	0.4210	0.1021	0.0447	0.0717	0.3619	0.0225	0.8911	0.4382
3B	**0.8462**	0.7337	0.0502	0.4101	0.6966	0.1359	0.3072	0.3741	0.5296
10B	**0.8462**	0.9399	0.0389	0.0554	0.1109	0.0509	0.0349	0.6340	0.4033
9B	**0.9231**	0.1532	0.0041	0.2675	0.4052	0.1458	0.1284	0.3432	0.2946
11B	**0.9231**	0.7057	0.1797	0.3317	0.4581	0.2521	0.2989	0.6144	0.4584
13B	**0.9231**	1.0000	0.1593	0.4809	0.6356	0.7353	0.5126	0.2402	0.4795
16A	**0.9231**	0.0259	0.0153	0.1026	0.1387	0.1872	0.0934	0.4005	0.3759
20A	**0.9231**	0.2049	0.0022	0.2877	0.4177	0.2045	0.1155	0.2031	0.0632
28B	**0.9231**	0.6496	0.1096	0.2916	0.4529	1.0000	0.3455	0.1263	0.4390
31B	**1.0000**	0.9999	0.0477	0.5866	0.7008	0.4441	0.2065	0.5735	1.0000

Table 6-9. Feature evaluation of sounds judged as the "most dissonant" from each pair, normalized and sorted by increasing dissonance rating.

Figure 6–9 shows a plot of normalized dissonance scores and normalized peak sample value for the sound judged as most dissonant from each pair in the entire data set. The data are presented sorted according to increasing dissonance rating. Also shown in the figure are cubic-spline interpolations of the data. For clarity, not all sound-object labels are shown on the *x* axis, but only every other one. The complete data set is shown in the previous two tables.)

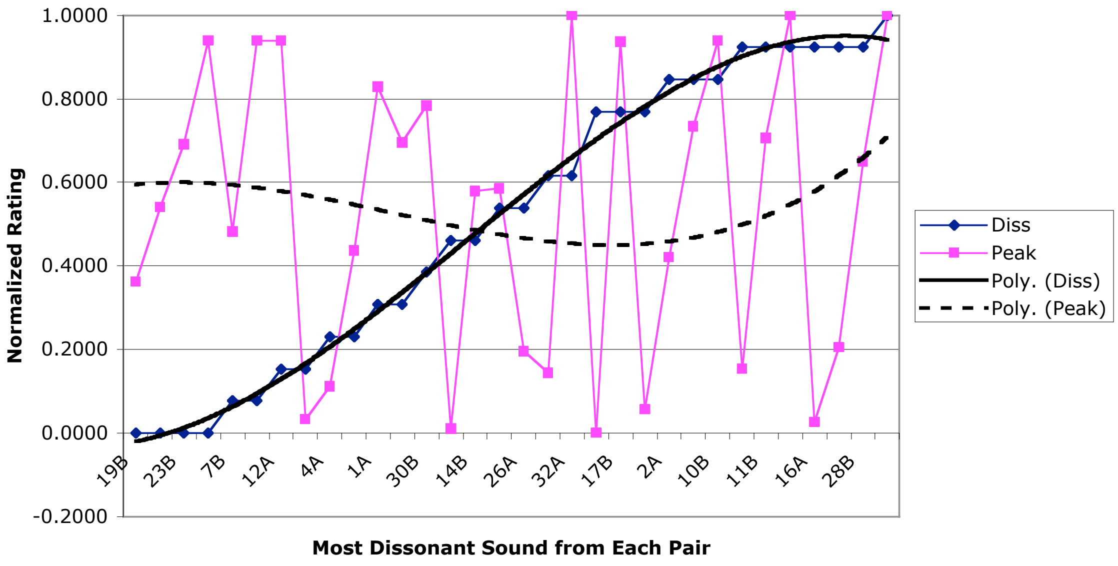


Figure 6–9. Normalized dissonance score and normalized peak sample value for the sound object rated as most dissonant from each pair, sorted by increasing dissonance rating.

Similarly, **Figure 6–10** through **Figure 6–16** illustrate RMS sample level, spectral centroid, spectral rolloff, harmonicity, zero-crossing rate, temporal centroid, and temporal centroid as a percentage of sound-object duration, respectively, versus the normalized dissonance rating of the sound object judged as most dissonant from each pair. Cubic-spline interpolations of the data are also provided in each figure.

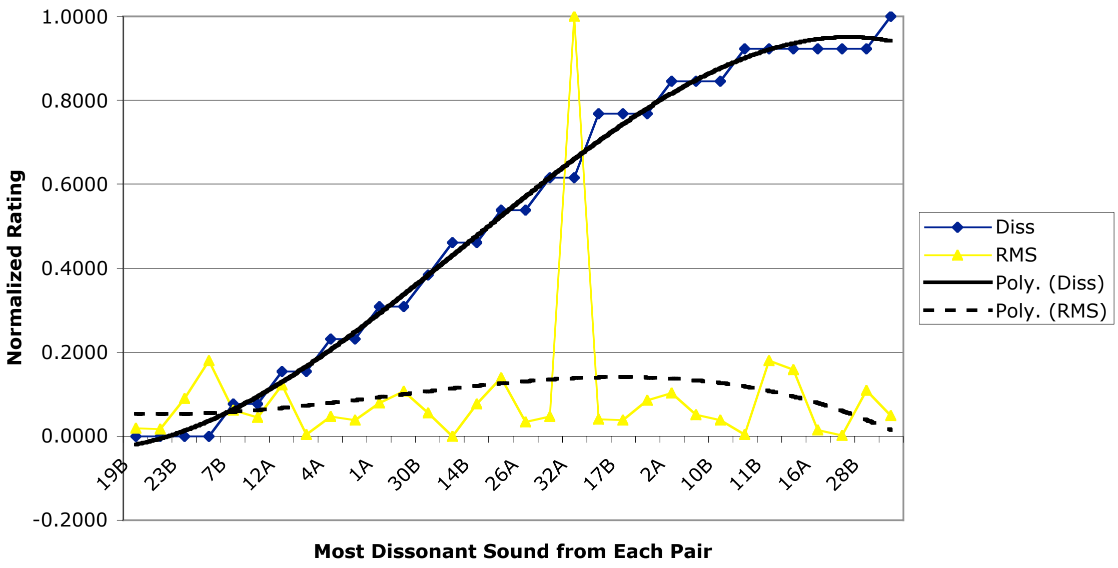


Figure 6–10. Normalized dissonance score and normalized root-mean-square (RMS) sample value for the sound object rated as most dissonant from each pair, sorted by increasing dissonance rating.

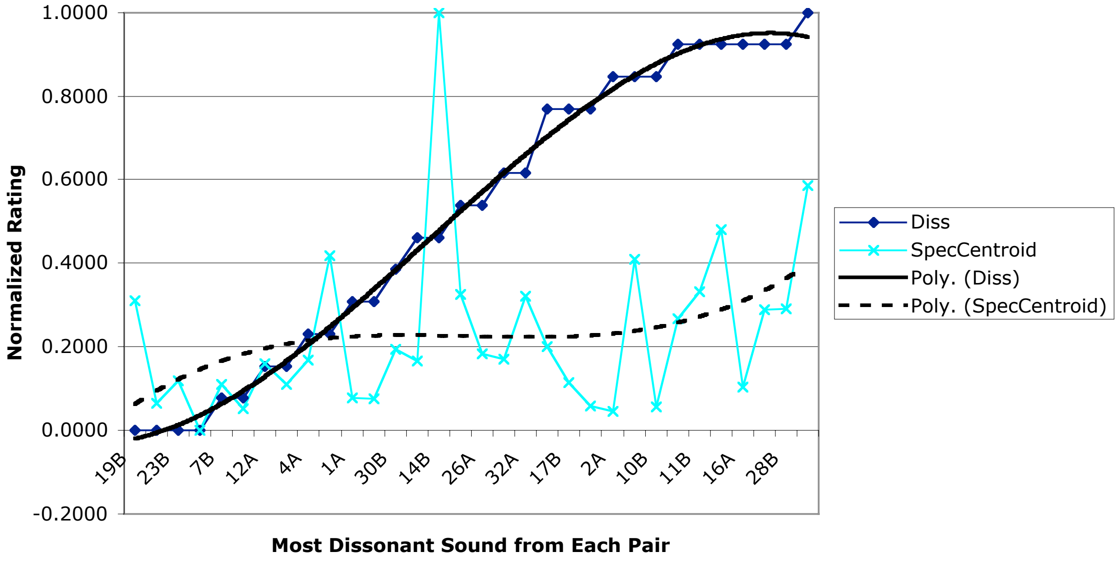


Figure 6–11. Normalized dissonance score and normalized spectral centroid value for the sound object rated as most dissonant from each pair, sorted by increasing dissonance rating.

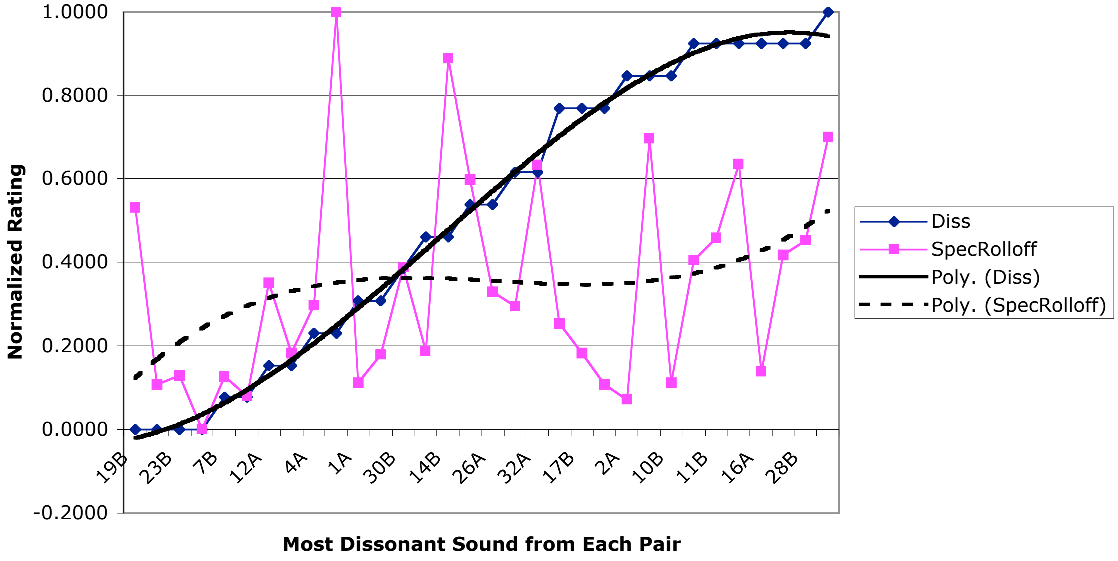


Figure 6–12. Normalized dissonance score and normalized spectral rolloff value for the sound object rated as most dissonant from each pair, sorted by increasing dissonance rating.

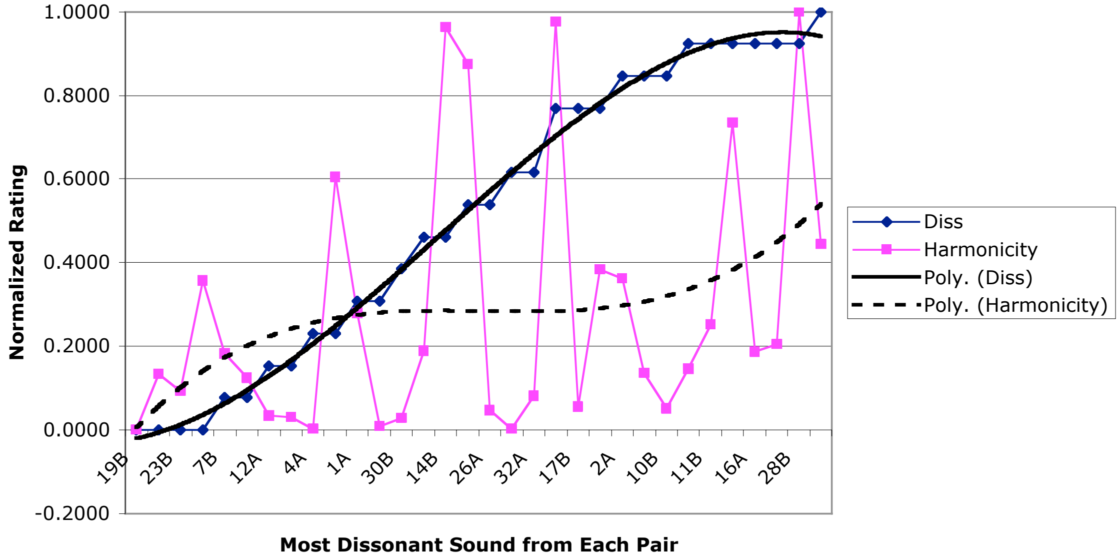


Figure 6–13. Normalized dissonance score and normalized harmonicity value for the sound object rated as most dissonant from each pair, sorted by increasing dissonance rating.

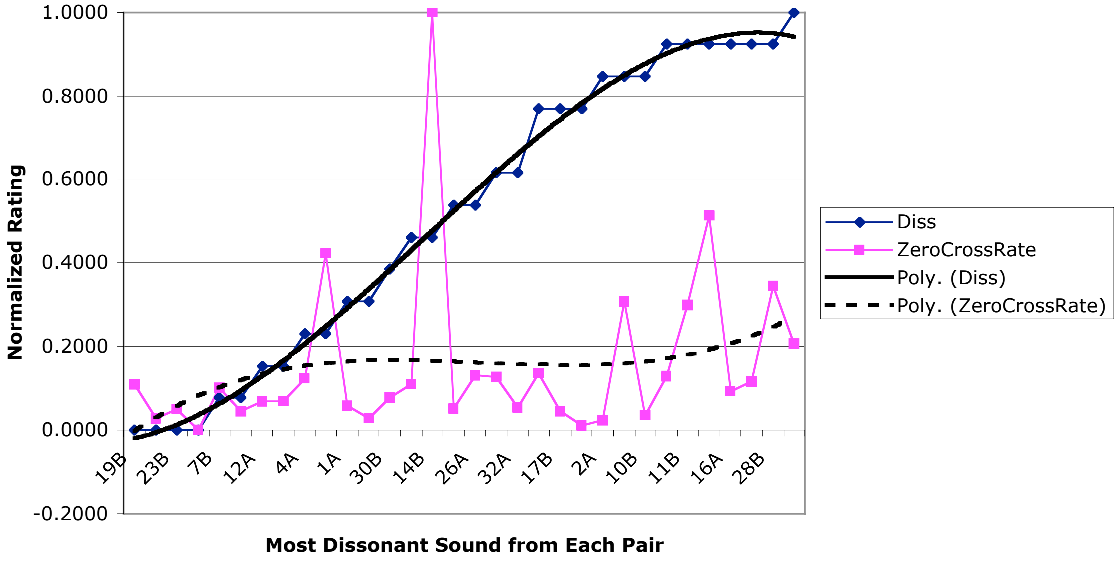


Figure 6–14. Normalized dissonance score and normalized zero-crossing rate for the sound object rated as most dissonant from each pair, sorted by increasing dissonance rating.

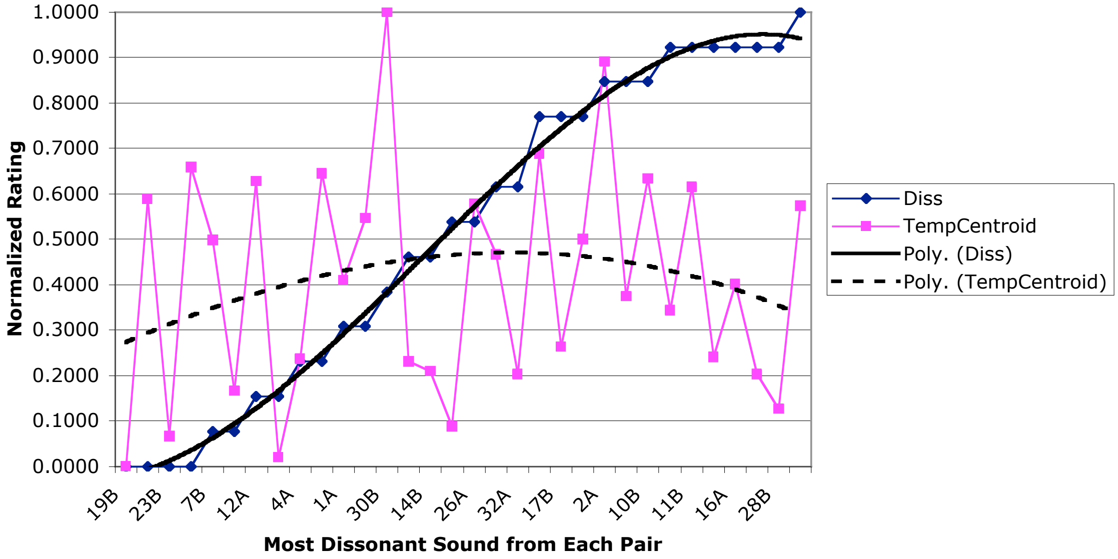


Figure 6–15. Normalized dissonance score and normalized temporal centroid for the sound object rated as most dissonant from each pair, sorted by increasing dissonance rating.

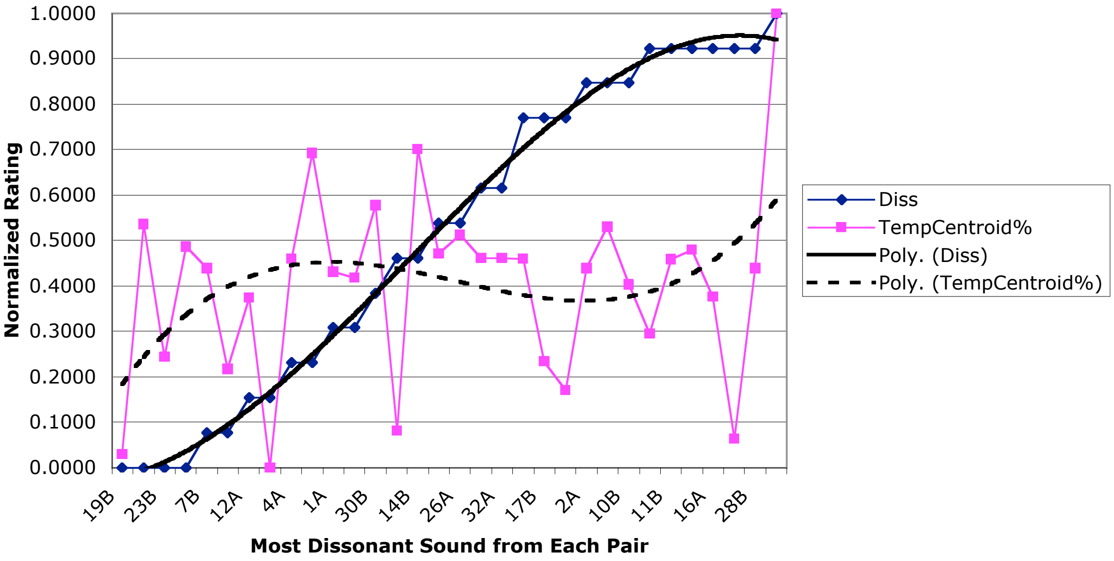


Figure 6–16. Normalized dissonance score and normalized temporal centroid (as a percentage of the sound-object duration) for the sound object rated as most dissonant from each pair, sorted by increasing dissonance rating.

Now that results and preliminary observations on the collected subjective data and computed objective data have been presented, the next section presents analysis of the test results.

6.3 Analysis and Discussion

As will be demonstrated, the results partially confirm the assertions of Chapter 5. However, they also fail to provide support for several of the assertions, as we will see.

Analysis of the test data was divided into two parts: (1) analysis of the subjective data (listener ratings) from the listening test, and (2) analysis of the computed feature data on the sound objects used in the listening test. Analysis of each type of data is now presented in the following sections.

Subjective Data

Subjective evaluation of the cubic-spline curves presented above reveals hints of correspondences among listener ratings to their rating of dissonance levels of the sound objects. As a means of quantifying these relationships further, the correlation coefficient (i.e., Pearson moment, or *r* value) was computed between normalized raw dissonance ratings and normalized responses to each of the listening-test questions. The results of these calculations are shown in **Table 6-10**.

	Correlation
Frust-Diss	0.5630
Diff-Diss	–0.3522
Intell-Diss	0.4881
Threat-Diss	0.4368
Annoy-Diss	0.6047
Recog-Diss	–0.6697
Predict-Diss	–0.4084

Constnt-Diss	0.2911
Pitched-Diss	–0.4904
Smoother-Diss	–0.4296
Regular-Diss	0.2282

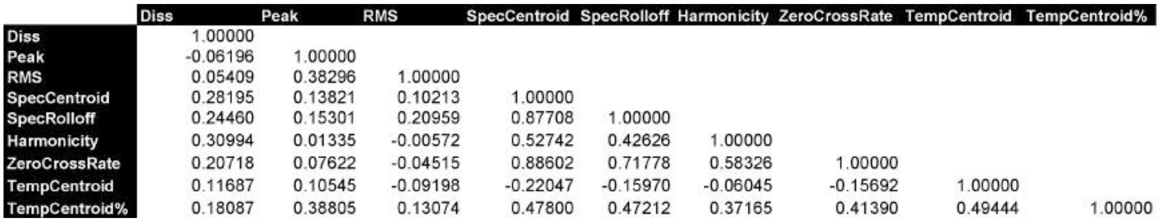
Table 6-10. Correlation coefficients for the normalized listening-test results.

Note that frustration-dissonance, intelligibility-dissonance, threat-dissonance, annoyance-dissonance, consistent-dissonance, and regular-dissonance exhibit positive correlation coefficients, indicating some degree of relationship. In particular, frustration and annoyance correlate strongly ($r > 0.5$), as does perception of threat ($r = 0.4368$) to dissonance rating, as proposed in Chapter 5, supporting assertion #2. Suprisingly, the positive correlation among intelligibility, consistency, and regularity per se to dissonance contradict somewhat assertion #3; however, the correlation coefficients involving consistency and regularity were at least small ($r < 0.3$).

On the other hand, negative correlation coefficients are displayed by difficult-dissonance, recognizable-dissonance, predictable-dissonance, pitched-dissonance, and smoother-dissonance. The strong negative correlation ($r < –0.5$) of recognizable-dissonance supports several of the assertions of Chapter 5, as does the negative correlation coefficients of predictable-dissonance, pitched-dissonance, and smoother-dissonance, supporting assertion #3. The results for difficult-dissonance, however, are somewhat surprising, indicating that subjects tended to rate sounds that were more difficult to understand as less dissonant.

Objective Data

Next, features computed for each sound object were compared, and the corresponding correlation coefficient between each feature and the normalized dissonance rating was computed. The results are presented as a correlation matrix in **Table 6-11**.



	Diss	Peak	RMS	SpecCentroid	SpecRolloff	Harmonicity	ZeroCrossRate	TempCentroid	TempCentroid%
Diss	1.00000								
Peak	-0.06196	1.00000							
RMS	0.05409	0.38296	1.00000						
SpecCentroid	0.28195	0.13821	0.10213	1.00000					
SpecRolloff	0.24460	0.15301	0.20959	0.87708	1.00000				
Harmonicity	0.30994	0.01335	-0.00572	0.52742	0.42626	1.00000			
ZeroCrossRate	0.20718	0.07622	-0.04515	0.88602	0.71778	0.58326	1.00000		
TempCentroid	0.11687	0.10545	-0.09198	-0.22047	-0.15970	-0.06045	-0.15692	1.00000	
TempCentroid%	0.18087	0.38805	0.13074	0.47800	0.47212	0.37165	0.41390	0.49444	1.00000

Table 6-11. Correlation matrix of objective data.

With the exception of normalized peak sample value, all other computed features correlate positively to normalized dissonance rating, albeit most of them weakly. The strongest correlation to dissonance involves spectral centroid ($r = 0.28195$), spectral rolloff ($r = 0.24460$), harmonicity ($r = 0.30994$), and zero-crossing rate ($r = 0.20718$). It can be argued that the positive correlations involving spectral centroid, spectral rolloff, and zero-crossing rate support the spirit of assertions #1–3, inasmuch that we assume that bright, broadband, and noisy sounds would tend to be less-liked (and therefore potentially evoke a negative-valence emotion and/or indicate low levels of predictability and regularity owing to its noisiness).

While analysis of both subjective and objective data generally supports at least portions of the theory proposed in Chapter 5, several contradictions exist.

For example, why do sounds that exhibit greater harmonicity tend to be classified as more dissonant by listeners? Also, why does RMS sample level not correlate strongly either positively or negatively to rated dissonance? The only explanation is that specific combinations of these and quite likely other subjective and objective features of each sound objects must combine somehow to encourage listeners to consider the musical dissonance of sound objects. For example, a sound object with a high RMS sample level may lead to a low dissonance rating, provided the harmonicity level is high ("Turn it up to eleven; the louder, the better!"), while a high RMS sample level may lead to a high dissonance rating when the harmonicity value is low ("Turn off that loud racket!").

Cluster Analysis and Visualization of Objective Data

To explore this, the multidimensional data must be reduced or clustered in some way to attempt to provide an insight. Many such algorithms are known; for suitability to this specific task, the k-means clustering algorithm (MacQueen 1967) was chosen to group the data set into meaningful clusters. This algorithm is a classic unsupervised learning algorithm that assumes data exhibit the potential to be meaningfully grouped—or clustered—in some way. The number of groups k to find is given as an input to the algorithm, and unfortunately, no general method of computing the optimum k for a given data set is available. The algorithm works by finding clusters through iteratively minimizing the mean-squared error J between data points and each proposed cluster center. The error J is given by

$$J = \sum_{j=1}^{k} \sum_{n} \left| x_n - \mu_j \right|^2$$

where x_n represents the nth point of data vector x, μ_j is the geometric centroid of the data points in the entire data collection, and k represents the number of clusters to be found.

To group the subjective and objective data, clustering of sizes $k \in \{2,3,4,5,6,7\}$ was attempted. The results of the k-means clustering are shown in **Table 6-12** ($k = 2$), **Table 6-13** ($k = 3$), **Table 6-14** ($k = 4$), **Table 6-15** ($k = 5$), **Table 6-16** ($k = 6$), and **Table 6-17** ($k = 7$).

Diss	Peak	RMS	SpecCentroid	SpecRolloff	Harmonicity	ZeroCrossRate	**TempCentroid**	**TempCentroid%**	Cluster
0.2308	0.4363	0.0373	0.4179	1.0000	0.6042	0.4230	0.6448	0.6924	1
0.4615	0.5792	0.0764	1.0000	0.8877	0.9637	1.0000	0.2100	0.7009	1
0.5385	0.5848	0.1393	0.3255	0.5987	0.8756	0.0514	0.0882	0.4717	1
0.6154	1.0000	1.0000	0.3219	0.6342	0.0806	0.0536	0.2029	0.4615	1
0.7692	0.0000	0.0411	0.1996	0.2532	0.9758	0.1359	0.6884	0.4594	1
0.8462	0.7337	0.0502	0.4101	0.6966	0.1359	0.3072	0.3741	0.5296	1
0.9231	0.7057	0.1797	0.3317	0.4581	0.2521	0.2989	0.6144	0.4584	1
0.9231	1.0000	0.1593	0.4809	0.6356	0.7353	0.5126	0.2402	0.4795	1
0.9231	0.6496	0.1096	0.2916	0.4529	1.0000	0.3455	0.1263	0.4390	1
1.0000	0.9999	0.0477	0.5866	0.7008	0.4441	0.2065	0.5735	1.0000	1
0.0000	0.3625	0.0177	0.3106	0.5309	0.0000	0.1086	0.0000	0.0293	2
0.0000	0.5415	0.0171	0.0638	0.1069	0.1338	0.0270	0.5884	0.5353	2
0.0000	0.6911	0.0900	0.1184	0.1283	0.0932	0.0499	0.0668	0.2432	2
0.0000	0.9399	0.1806	0.0000	0.0000	0.3568	0.0000	0.6588	0.4865	2
0.0769	0.4819	0.0622	0.1097	0.1260	0.1828	0.1009	0.4980	0.4386	2
0.0769	0.9399	0.0451	0.0519	0.0797	0.1246	0.0444	0.1663	0.2170	2
0.1538	0.9398	0.1214	0.1588	0.3505	0.0335	0.0680	0.6275	0.3746	2
0.1538	0.0328	0.0032	0.1095	0.1820	0.0303	0.0697	0.0198	0.0000	2
0.2308	0.1115	0.0460	0.1686	0.2976	0.0025	0.1228	0.2366	0.4592	2
0.3077	0.8285	0.0789	0.0769	0.1108	0.2794	0.0578	0.4101	0.4307	2
0.3077	0.6951	0.1064	0.0752	0.1787	0.0088	0.0281	0.5462	0.4182	2
0.3846	0.7838	0.0551	0.1945	0.3881	0.0280	0.0769	1.0000	0.5768	2
0.4615	0.0102	0.0000	0.1655	0.1879	0.1885	0.1106	0.2308	0.0813	2
0.5385	0.1957	0.0327	0.1837	0.3286	0.0469	0.1309	0.5778	0.5123	2
0.6154	0.1434	0.0460	0.1704	0.2958	0.0030	0.1270	0.4662	0.4617	2
0.7692	0.9363	0.0374	0.1147	0.1821	0.0553	0.0448	0.2631	0.2335	2
0.7692	0.0562	0.0855	0.0588	0.1071	0.3841	0.0097	0.4997	0.1707	2
0.8462	0.4210	0.1021	0.0447	0.0717	0.3619	0.0225	0.8911	0.4382	2
0.8462	0.9399	0.0389	0.0554	0.1109	0.0509	0.0349	0.6340	0.4033	2
0.9231	0.1532	0.0041	0.2675	0.4052	0.1458	0.1284	0.3432	0.2946	2
0.9231	0.0259	0.0153	0.1026	0.1387	0.1872	0.0934	0.4005	0.3759	2
0.9231	0.2049	0.0022	0.2877	0.4177	0.2045	0.1155	0.2031	0.0632	2

Table 6-12. K-means clustering of objective data, $k = 2$.

Diss	Peak	RMS	SpecCentroid	SpecRolloff	Harmonicity	ZeroCrossRate	TempCentroid	TempCentroid%	Cluster
0.0000	0.3625	0.0177	0.3106	0.5309	0.0000	0.1086	0.0000	0.0293	1
0.1538	0.0328	0.0032	0.1095	0.1820	0.0303	0.0697	0.0198	0.0000	1
0.2308	0.1115	0.0460	0.1686	0.2976	0.0025	0.1228	0.2366	0.4592	1
0.4615	0.0102	0.0000	0.1655	0.1879	0.1885	0.1106	0.2308	0.0813	1
0.5385	0.1957	0.0327	0.1837	0.3286	0.0469	0.1309	0.5778	0.5123	1
0.6154	0.1434	0.0460	0.1704	0.2958	0.0030	0.1270	0.4662	0.4617	1
0.7692	0.0000	0.0411	0.1996	0.2532	0.9758	0.1359	0.6884	0.4594	1
0.7692	0.0562	0.0855	0.0588	0.1071	0.3841	0.0097	0.4997	0.1707	1
0.9231	0.1532	0.0041	0.2675	0.4052	0.1458	0.1284	0.3432	0.2946	1
0.9231	0.0259	0.0153	0.1026	0.1387	0.1872	0.0934	0.4005	0.3759	1
0.9231	0.2049	0.0022	0.2877	0.4177	0.2045	0.1155	0.2031	0.0632	1
0.0000	0.5415	0.0171	0.0638	0.1069	0.1338	0.0270	0.5884	0.5353	2
0.0000	0.6911	0.0900	0.1184	0.1283	0.0932	0.0499	0.0668	0.2432	2
0.0000	0.9399	0.1806	0.0000	0.0000	0.3568	0.0000	0.6588	0.4865	2
0.0769	0.4819	0.0622	0.1097	0.1260	0.1828	0.1009	0.4980	0.4386	2
0.0769	0.9399	0.0451	0.0519	0.0797	0.1246	0.0444	0.1663	0.2170	2
0.1538	0.9398	0.1214	0.1588	0.3505	0.0335	0.0680	0.6275	0.3746	2
0.3077	0.8285	0.0789	0.0769	0.1108	0.2794	0.0578	0.4101	0.4307	2
0.3077	0.6951	0.1064	0.0752	0.1787	0.0088	0.0281	0.5462	0.4182	2
0.3846	0.7838	0.0551	0.1945	0.3881	0.0280	0.0769	1.0000	0.5768	2
0.7692	0.9363	0.0374	0.1147	0.1821	0.0553	0.0448	0.2631	0.2335	2
0.8462	0.4210	0.1021	0.0447	0.0717	0.3619	0.0225	0.8911	0.4382	2
0.8462	0.9399	0.0389	0.0554	0.1109	0.0509	0.0349	0.6340	0.4033	2
0.2308	0.4363	0.0373	0.4179	1.0000	0.6042	0.4230	0.6448	0.6924	3
0.4615	0.5792	0.0764	1.0000	0.8877	0.9637	1.0000	0.2100	0.7009	3
0.5385	0.5848	0.1393	0.3255	0.5987	0.8756	0.0514	0.0882	0.4717	3
0.6154	1.0000	1.0000	0.3219	0.6342	0.0806	0.0536	0.2029	0.4615	3
0.8462	0.7337	0.0502	0.4101	0.6966	0.1359	0.3072	0.3741	0.5296	3
0.9231	0.7057	0.1797	0.3317	0.4581	0.2521	0.2989	0.6144	0.4584	3
0.9231	1.0000	0.1593	0.4809	0.6356	0.7353	0.5126	0.2402	0.4795	3
0.9231	0.6496	0.1096	0.2916	0.4529	1.0000	0.3455	0.1263	0.4390	3
1.0000	0.9999	0.0477	0.5866	0.7008	0.4441	0.2065	0.5735	1.0000	3

Table 6-13. K-means clustering of objective data, $k = 3$.

Diss	Peak	RMS	SpecCentroid	SpecRolloff	Harmonicity	ZeroCrossRate	TempCentroid	**TempCentroid%**	Cluster
0.0000	0.3625	0.0177	0.3106	0.5309	0.0000	0.1086	0.0000	0.0293	1
0.1538	0.0328	0.0032	0.1095	0.1820	0.0303	0.0697	0.0198	0.0000	1
0.2308	0.1115	0.0460	0.1686	0.2976	0.0025	0.1228	0.2366	0.4592	1
0.4615	0.0102	0.0000	0.1655	0.1879	0.1885	0.1106	0.2308	0.0813	1
0.5385	0.1957	0.0327	0.1837	0.3286	0.0469	0.1309	0.5778	0.5123	2
0.6154	0.1434	0.0460	0.1704	0.2958	0.0030	0.1270	0.4662	0.4617	2
0.7692	0.0000	0.0411	0.1996	0.2532	0.9758	0.1359	0.6884	0.4594	2
0.7692	0.0562	0.0855	0.0588	0.1071	0.3841	0.0097	0.4997	0.1707	2
0.8462	0.4210	0.1021	0.0447	0.0717	0.3619	0.0225	0.8911	0.4382	2
0.9231	0.1532	0.0041	0.2675	0.4052	0.1458	0.1284	0.3432	0.2946	2
0.9231	0.0259	0.0153	0.1026	0.1387	0.1872	0.0934	0.4005	0.3759	2
0.9231	0.2049	0.0022	0.2877	0.4177	0.2045	0.1155	0.2031	0.0632	2
0.2308	0.4363	0.0373	0.4179	1.0000	0.6042	0.4230	0.6448	0.6924	3
0.4615	0.5792	0.0764	1.0000	0.8877	0.9637	1.0000	0.2100	0.7009	3
0.5385	0.5848	0.1393	0.3255	0.5987	0.8756	0.0514	0.0882	0.4717	3
0.6154	1.0000	1.0000	0.3219	0.6342	0.0806	0.0536	0.2029	0.4615	3
0.8462	0.7337	0.0502	0.4101	0.6966	0.1359	0.3072	0.3741	0.5296	3
0.9231	0.7057	0.1797	0.3317	0.4581	0.2521	0.2989	0.6144	0.4584	3
0.9231	1.0000	0.1593	0.4809	0.6356	0.7353	0.5126	0.2402	0.4795	3
0.9231	0.6496	0.1096	0.2916	0.4529	1.0000	0.3455	0.1263	0.4390	3
1.0000	0.9999	0.0477	0.5866	0.7008	0.4441	0.2065	0.5735	1.0000	3
0.0000	0.5415	0.0171	0.0638	0.1069	0.1338	0.0270	0.5884	0.5353	4
0.0000	0.6911	0.0900	0.1184	0.1283	0.0932	0.0499	0.0668	0.2432	4
0.0000	0.9399	0.1806	0.0000	0.0000	0.3568	0.0000	0.6588	0.4865	4
0.0769	0.4819	0.0622	0.1097	0.1260	0.1828	0.1009	0.4980	0.4386	4
0.0769	0.9399	0.0451	0.0519	0.0797	0.1246	0.0444	0.1663	0.2170	4
0.1538	0.9398	0.1214	0.1588	0.3505	0.0335	0.0680	0.6275	0.3746	4
0.3077	0.8285	0.0789	0.0769	0.1108	0.2794	0.0578	0.4101	0.4307	4
0.3077	0.6951	0.1064	0.0752	0.1787	0.0088	0.0281	0.5462	0.4182	4
0.3846	0.7838	0.0551	0.1945	0.3881	0.0280	0.0769	1.0000	0.5768	4
0.7692	0.9363	0.0374	0.1147	0.1821	0.0553	0.0448	0.2631	0.2335	4
0.8462	0.9399	0.0389	0.0554	0.1109	0.0509	0.0349	0.6340	0.4033	4

Table 6-14. K-means clustering of objective data, $k = 4$.

Diss	Peak	RMS	SpecCentroid	SpecRolloff	Harmonicity	ZeroCrossRate	TempCentroid	TempCentroid%	Cluster
0.0000	0.3625	0.0177	0.3106	0.5309	0.0000	0.1086	0.0000	0.0293	1
0.1538	0.0328	0.0032	0.1095	0.1820	0.0303	0.0697	0.0198	0.0000	1
0.2308	0.1115	0.0460	0.1686	0.2976	0.0025	0.1228	0.2366	0.4592	1
0.4615	0.0102	0.0000	0.1655	0.1879	0.1885	0.1106	0.2308	0.0813	1
0.6154	0.1434	0.0460	0.1704	0.2958	0.0030	0.1270	0.4662	0.4617	1
0.7692	0.0562	0.0855	0.0588	0.1071	0.3841	0.0097	0.4997	0.1707	1
0.9231	0.1532	0.0041	0.2675	0.4052	0.1458	0.1284	0.3432	0.2946	1
0.9231	0.0259	0.0153	0.1026	0.1387	0.1872	0.0934	0.4005	0.3759	1
0.9231	0.2049	0.0022	0.2877	0.4177	0.2045	0.1155	0.2031	0.0632	1
0.0000	0.5415	0.0171	0.0638	0.1069	0.1338	0.0270	0.5884	0.5353	2
0.0769	0.4819	0.0622	0.1097	0.1260	0.1828	0.1009	0.4980	0.4386	2
0.3077	0.6951	0.1064	0.0752	0.1787	0.0088	0.0281	0.5462	0.4182	2
0.3846	0.7838	0.0551	0.1945	0.3881	0.0280	0.0769	1.0000	0.5768	2
0.5385	0.1957	0.0327	0.1837	0.3286	0.0469	0.1309	0.5778	0.5123	2
0.8462	0.4210	0.1021	0.0447	0.0717	0.3619	0.0225	0.8911	0.4382	2
0.0000	0.6911	0.0900	0.1184	0.1283	0.0932	0.0499	0.0668	0.2432	3
0.0769	0.9399	0.0451	0.0519	0.0797	0.1246	0.0444	0.1663	0.2170	3
0.6154	1.0000	1.0000	0.3219	0.6342	0.0806	0.0536	0.2029	0.4615	3
0.7692	0.9363	0.0374	0.1147	0.1821	0.0553	0.0448	0.2631	0.2335	3
0.0000	0.9399	0.1806	0.0000	0.0000	0.3568	0.0000	0.6588	0.4865	4
0.1538	0.9398	0.1214	0.1588	0.3505	0.0335	0.0680	0.6275	0.3746	4
0.3077	0.8285	0.0789	0.0769	0.1108	0.2794	0.0578	0.4101	0.4307	4
0.8462	0.9399	0.0389	0.0554	0.1109	0.0509	0.0349	0.6340	0.4033	4
0.2308	0.4363	0.0373	0.4179	1.0000	0.6042	0.4230	0.6448	0.6924	5
0.4615	0.5792	0.0764	1.0000	0.8877	0.9637	1.0000	0.2100	0.7009	5
0.5385	0.5848	0.1393	0.3255	0.5987	0.8756	0.0514	0.0882	0.4717	5
0.7692	0.0000	0.0411	0.1996	0.2532	0.9758	0.1359	0.6884	0.4594	5
0.8462	0.7337	0.0502	0.4101	0.6966	0.1359	0.3072	0.3741	0.5296	5
0.9231	0.7057	0.1797	0.3317	0.4581	0.2521	0.2989	0.6144	0.4584	5
0.9231	1.0000	0.1593	0.4809	0.6356	0.7353	0.5126	0.2402	0.4795	5
0.9231	0.6496	0.1096	0.2916	0.4529	1.0000	0.3455	0.1263	0.4390	5
1.0000	0.9999	0.0477	0.5866	0.7008	0.4441	0.2065	0.5735	1.0000	5

Table 6-15. K-means clustering of objective data, $k = 5$.

Diss	Peak	RMS	SpecCentroid	SpecRolloff	Harmonicity	ZeroCrossRate	TempCentroid	TempCentroid%	Cluster
0.1538	0.0328	0.0032	0.1095	0.1820	0.0303	0.0697	0.0198	0.0000	1
0.2308	0.1115	0.0460	0.1686	0.2976	0.0025	0.1228	0.2366	0.4592	1
0.4615	0.0102	0.0000	0.1655	0.1879	0.1885	0.1106	0.2308	0.0813	1
0.5385	0.1957	0.0327	0.1837	0.3286	0.0469	0.1309	0.5778	0.5123	1
0.6154	0.1434	0.0460	0.1704	0.2958	0.0030	0.1270	0.4662	0.4617	1
0.7692	0.0562	0.0855	0.0588	0.1071	0.3841	0.0097	0.4997	0.1707	1
0.9231	0.1532	0.0041	0.2675	0.4052	0.1458	0.1284	0.3432	0.2946	1
0.9231	0.0259	0.0153	0.1026	0.1387	0.1872	0.0934	0.4005	0.3759	1
0.9231	0.2049	0.0022	0.2877	0.4177	0.2045	0.1155	0.2031	0.0632	1
0.0000	0.5415	0.0171	0.0638	0.1069	0.1338	0.0270	0.5884	0.5353	2
0.0769	0.4819	0.0622	0.1097	0.1260	0.1828	0.1009	0.4980	0.4386	2
0.3077	0.6951	0.1064	0.0752	0.1787	0.0088	0.0281	0.5462	0.4182	2
0.3846	0.7838	0.0551	0.1945	0.3881	0.0280	0.0769	1.0000	0.5768	2
0.8462	0.4210	0.1021	0.0447	0.0717	0.3619	0.0225	0.8911	0.4382	2
0.0000	0.3625	0.0177	0.3106	0.5309	0.0000	0.1086	0.0000	0.0293	3
0.0000	0.6911	0.0900	0.1184	0.1283	0.0932	0.0499	0.0668	0.2432	3
0.0769	0.9399	0.0451	0.0519	0.0797	0.1246	0.0444	0.1663	0.2170	3
0.0000	0.9399	0.1806	0.0000	0.0000	0.3568	0.0000	0.6588	0.4865	4
0.1538	0.9398	0.1214	0.1588	0.3505	0.0335	0.0680	0.6275	0.3746	4
0.3077	0.8285	0.0789	0.0769	0.1108	0.2794	0.0578	0.4101	0.4307	4
0.2308	0.4363	0.0373	0.4179	1.0000	0.6042	0.4230	0.6448	0.6924	5
0.4615	0.5792	0.0764	1.0000	0.8877	0.9637	1.0000	0.2100	0.7009	5
0.5385	0.5848	0.1393	0.3255	0.5987	0.8756	0.0514	0.0882	0.4717	5
0.7692	0.0000	0.0411	0.1996	0.2532	0.9758	0.1359	0.6884	0.4594	5
0.9231	0.6496	0.1096	0.2916	0.4529	1.0000	0.3455	0.1263	0.4390	5
0.6154	1.0000	1.0000	0.3219	0.6342	0.0806	0.0536	0.2029	0.4615	6
0.7692	0.9363	0.0374	0.1147	0.1821	0.0553	0.0448	0.2631	0.2335	6
0.8462	0.7337	0.0502	0.4101	0.6966	0.1359	0.3072	0.3741	0.5296	6
0.8462	0.9399	0.0389	0.0554	0.1109	0.0509	0.0349	0.6340	0.4033	6
0.9231	0.7057	0.1797	0.3317	0.4581	0.2521	0.2989	0.6144	0.4584	6
0.9231	1.0000	0.1593	0.4809	0.6356	0.7353	0.5126	0.2402	0.4795	6
1.0000	0.9999	0.0477	0.5866	0.7008	0.4441	0.2065	0.5735	1.0000	6

Table 6-16. K-means clustering of objective data, $k = 6$.

Diss	Peak	RMS	SpecCentroid	SpecRolloff	Harmonicity	ZeroCrossRate	TempCentroid	TempCentroid%	Cluster
0.0000	0.3625	0.0177	0.3106	0.5309	0.0000	0.1086	0.0000	0.0293	1
0.1538	0.0328	0.0032	0.1095	0.1820	0.0303	0.0697	0.0198	0.0000	1
0.2308	0.1115	0.0460	0.1686	0.2976	0.0025	0.1228	0.2366	0.4592	1
0.4615	0.0102	0.0000	0.1655	0.1879	0.1885	0.1106	0.2308	0.0813	1
0.0000	0.5415	0.0171	0.0638	0.1069	0.1338	0.0270	0.5884	0.5353	2
0.0769	0.4819	0.0622	0.1097	0.1260	0.1828	0.1009	0.4980	0.4386	2
0.3077	0.6951	0.1064	0.0752	0.1787	0.0088	0.0281	0.5462	0.4182	2
0.0000	0.6911	0.0900	0.1184	0.1283	0.0932	0.0499	0.0668	0.2432	3
0.0769	0.9399	0.0451	0.0519	0.0797	0.1246	0.0444	0.1663	0.2170	3
0.0000	0.9399	0.1806	0.0000	0.0000	0.3568	0.0000	0.6588	0.4865	4
0.1538	0.9398	0.1214	0.1588	0.3505	0.0335	0.0680	0.6275	0.3746	4
0.3846	0.7838	0.0551	0.1945	0.3881	0.0280	0.0769	1.0000	0.5768	4
0.5385	0.1957	0.0327	0.1837	0.3286	0.0469	0.1309	0.5778	0.5123	5
0.6154	0.1434	0.0460	0.1704	0.2958	0.0030	0.1270	0.4662	0.4617	5
0.7692	0.0000	0.0411	0.1996	0.2532	0.9758	0.1359	0.6884	0.4594	5
0.7692	0.0562	0.0855	0.0588	0.1071	0.3841	0.0097	0.4997	0.1707	5
0.8462	0.4210	0.1021	0.0447	0.0717	0.3619	0.0225	0.8911	0.4382	5
0.9231	0.1532	0.0041	0.2675	0.4052	0.1458	0.1284	0.3432	0.2946	5
0.9231	0.0259	0.0153	0.1026	0.1387	0.1872	0.0934	0.4005	0.3759	5
0.9231	0.2049	0.0022	0.2877	0.4177	0.2045	0.1155	0.2031	0.0632	5
0.3077	0.8285	0.0789	0.0769	0.1108	0.2794	0.0578	0.4101	0.4307	6
0.7692	0.9363	0.0374	0.1147	0.1821	0.0553	0.0448	0.2631	0.2335	6
0.8462	0.9399	0.0389	0.0554	0.1109	0.0509	0.0349	0.6340	0.4033	6
0.2308	0.4363	0.0373	0.4179	1.0000	0.6042	0.4230	0.6448	0.6924	7
0.4615	0.5792	0.0764	1.0000	0.8877	0.9637	1.0000	0.2100	0.7009	7
0.5385	0.5848	0.1393	0.3255	0.5987	0.8756	0.0514	0.0882	0.4717	7
0.6154	1.0000	1.0000	0.3219	0.6342	0.0806	0.0536	0.2029	0.4615	7
0.8462	0.7337	0.0502	0.4101	0.6966	0.1359	0.3072	0.3741	0.5296	7
0.9231	0.7057	0.1797	0.3317	0.4581	0.2521	0.2989	0.6144	0.4584	7
0.9231	1.0000	0.1593	0.4809	0.6356	0.7353	0.5126	0.2402	0.4795	7
0.9231	0.6496	0.1096	0.2916	0.4529	1.0000	0.3455	0.1263	0.4390	7
1.0000	0.9999	0.0477	0.5866	0.7008	0.4441	0.2065	0.5735	1.0000	7

Table 6-17. K-means clustering of objective data, $k = 7$.

Next, the data sets were interactively visualized in the YALE machine-learning and data-mining environment (Mierswa et al. 2006). Owing to the relatively small number of test trials (32) by data-mining standards, as well as the desire to minimize the difficulty of achieving meaningful and easily readable visualizations, the case of binary clustering (i.e., $k = 2$) was interactively explored and probed for meaning. Furthermore, binary clustering was chosen to examine the extent to which one cluster would represent the "more consonant" sound objects and the other the "more dissonant" sound objects. Note that in the binary case, cluster #1 exhibits a higher mean dissonance rating.

First, the validity of the clustering was tested using a self-organizing map (SOM; Kohonen 1988, 1997), which attempts to map large-dimensional data sets into fewer dimensions (akin to the purpose of classical multidimensional scaling), allowing for visualization. Dimensionality reduction is achieved here primarily by iteratively training a neural network on the input data to produce graphical "islands" and "shores" onto which data points can be drawn. A 40-pixel by 30-pixel, two-dimensional SOM of the binary-clustered data is shown in Figure 6–17. Here, 25 neural-network training rounds were used with an adaptation radius of 15 pixels.

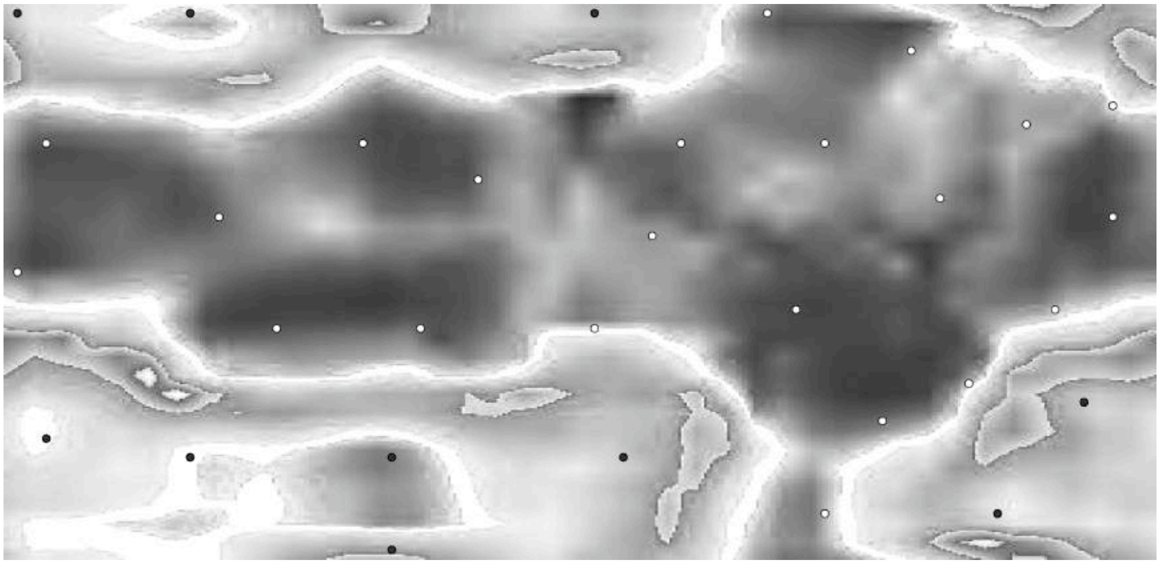


Figure 6–17. SOM for binary-clustered sound-object feature data. The dark circles indicate cluster #1, and the white circles indicate cluster #2.

Examination of Figure 6–17 clearly indicates the success of the binary-clustering operation previously performed. The first cluster of data points (shown on the map as dark circles) tend to form around the outside boundary of the map, and with only one exception, that cluster appears only on the "shore", or light-colored area of the map. On the other hand, the second cluster of data points (shown on the map in white circles) tend to form directly in the "ocean"—in the middle of the map.

Further support of the binary clustering is provided by examination of an Andrews Curve (Andrews 1972) of the data set. In this visualization, multidimensional data sets are mapped to Fourier coefficients to produce continuous curves. Data vectors of high correspondence will tend to be displayed as synchronized curves, while vectors of low correspondence will exhibit little curvilinear similarity. An Andrews Curve of the binary-clustered feature data is shown in Figure 6–18. Curve color indicates the cluster number (darker = cluster #1, lighter = cluster #2). The curves clearly form two rough groups as expected, particularly around their local extrema; however, the groupings are not particularly pronounced, at least relative to typical Andrews Curve models, suggesting a potentially subtle transition between cluster #1 and cluster #2.

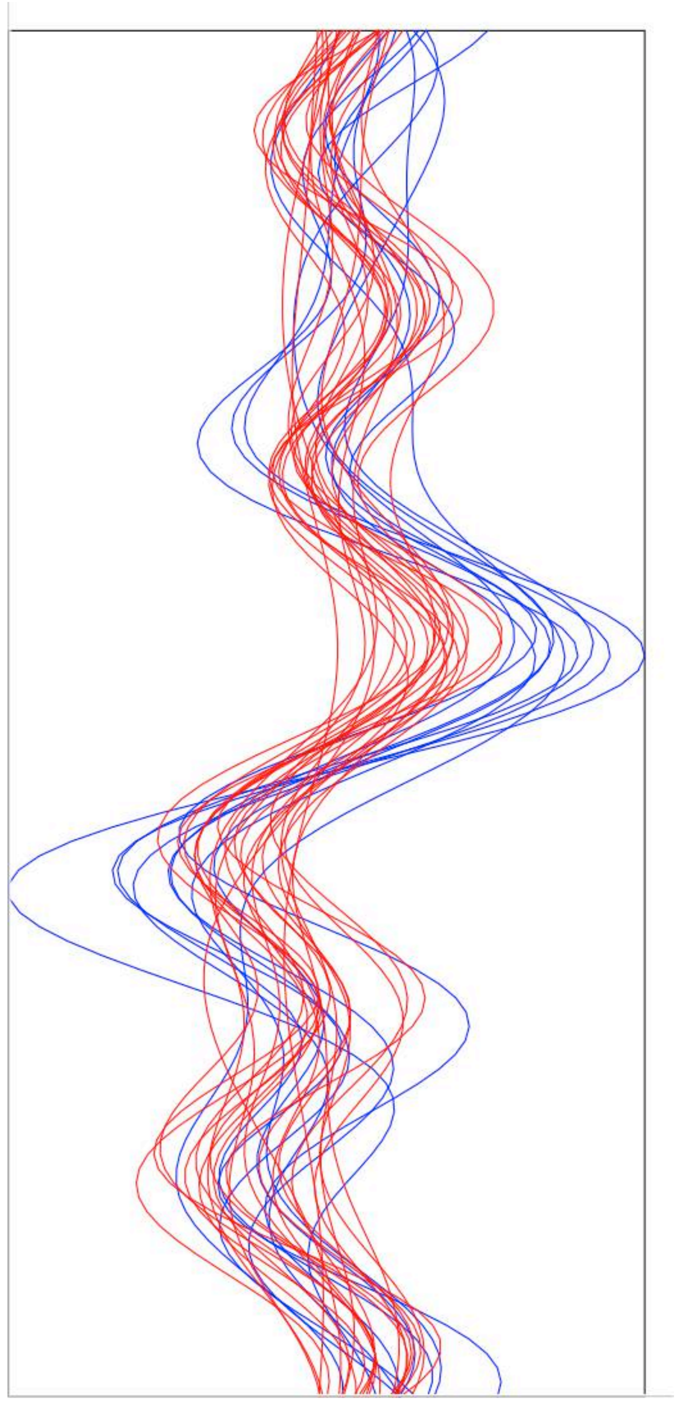


Figure 6–18. Andrews Curve of binary-clustered feature data.

Next, attempts to discern the relative contributions of each feature to the normalized dissonance rating and cluster grouping were made. Several visualizations were used to this end. The first, a RadViz plot (Ankerst, Keim, and Kriegel 1996), is shown in Figure 6–19, and indicates a strong clustering of high dissonance rating with high temporal centroid, peak sample value, and spectral centroid. A mild clustering of weak dissonance, low spectral rolloff, low zero-crossing rate, and low harmonicity is also exhibited.

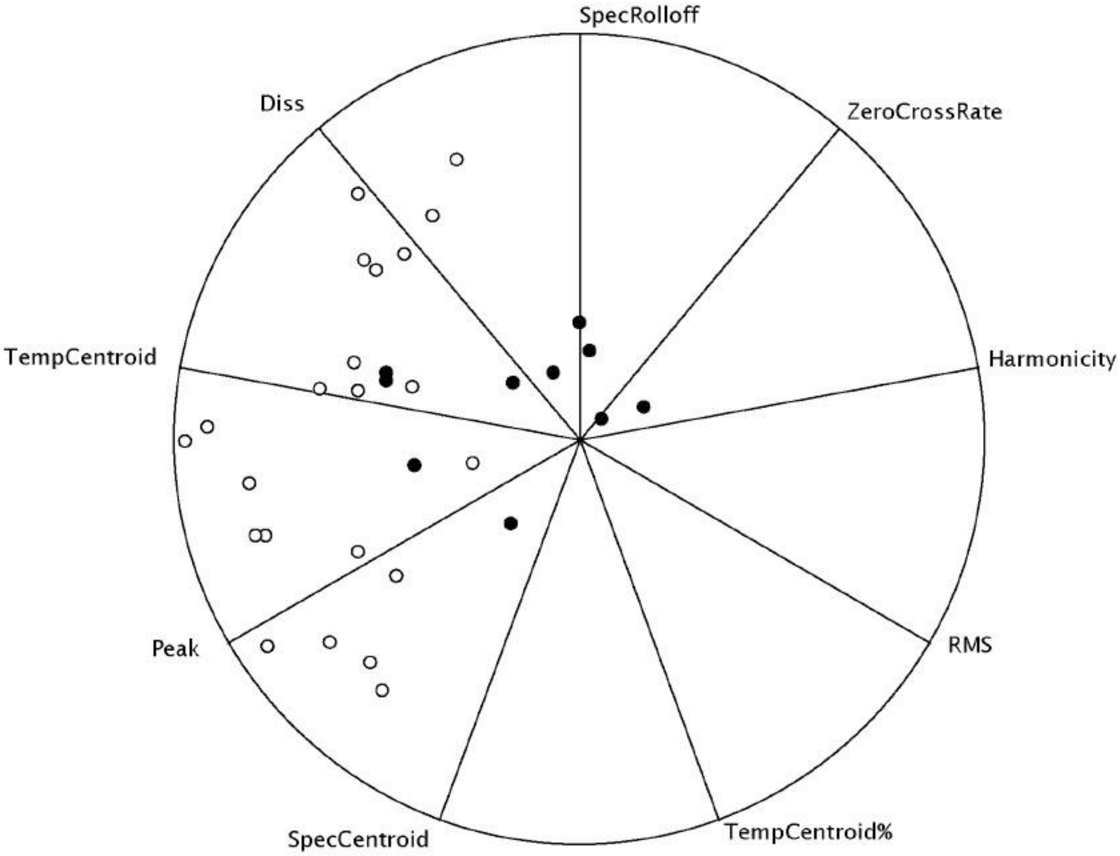


Figure 6–19. RadViz plot showing cluster #1 (white-filled circles) and cluster #2 (dark-filled circles).

Additional insights are gained upon examination of a quartile-matrix visualization of the binary-clustered normalized features. Such a visualization is depicted in **Figure 6–20**.

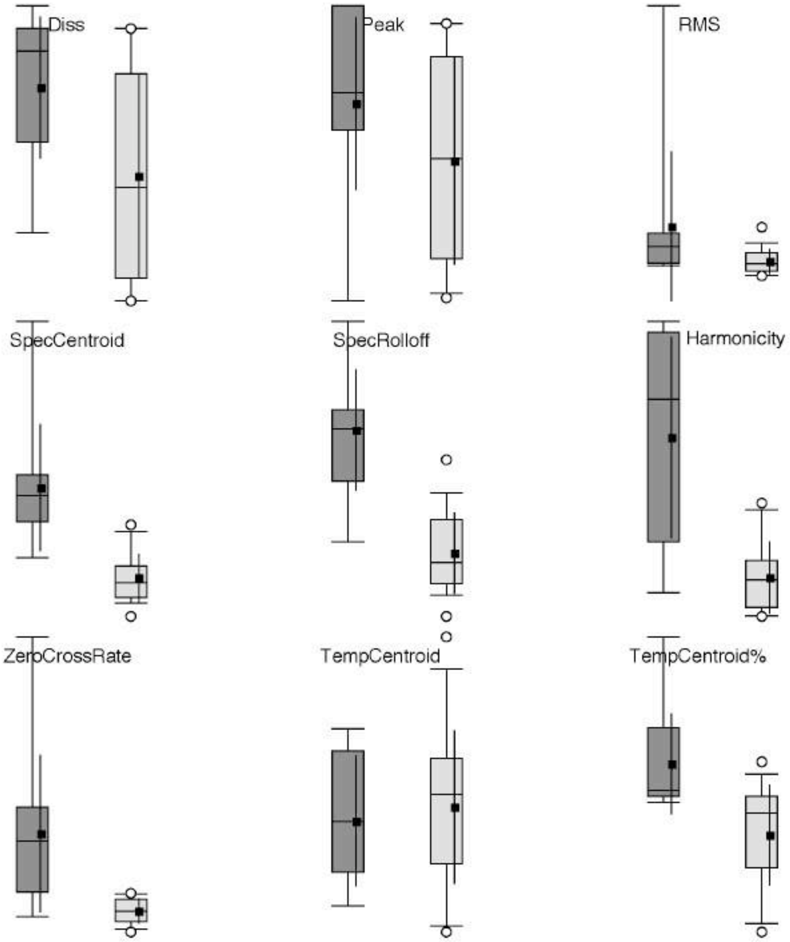


Figure 6–20. Quartile-matrix visualization of $k = 2$ clustered data. Cluster #1 is shaded dark-gray, and cluster #2 is shaded light-gray.

Note that the quartile-matrix visualization indicates the strongest groupings along the dimensions of normalized dissonance rating (as we would expect), peak sample value, spectral centroid and rolloff, harmonicity, zero-crossing rate, and temporal centroid as a percentage of the sound-object duration. The most pronounced groupings were apparently related to spectral centroid and spectral rolloff. Weak clustering was apparently evoked along the dimensions of RMS sample level and temporal centroid.

Further evidence for these claims is supported by the visualization of Figure 6–21, a so-called Parallel Plot, which indicates that spectral centroid, spectral rolloff, harmonicity, and zero-crossing rate provide potentially informative criteria along which data can be clustered into two groups as to relative dissonance. A Histogram Color Matrix representation of the multidimensional data set, shown in Figure 6-22, also supports the assertion that the data cluster fairly well according to spectral rolloff, spectral centroid, and temporal centroid as a percentage of sound-object duration. Clearly, other features such as peak sample value and RMS sample value provide relatively little information regarding partitioning and classification of the data set according to dissonance.

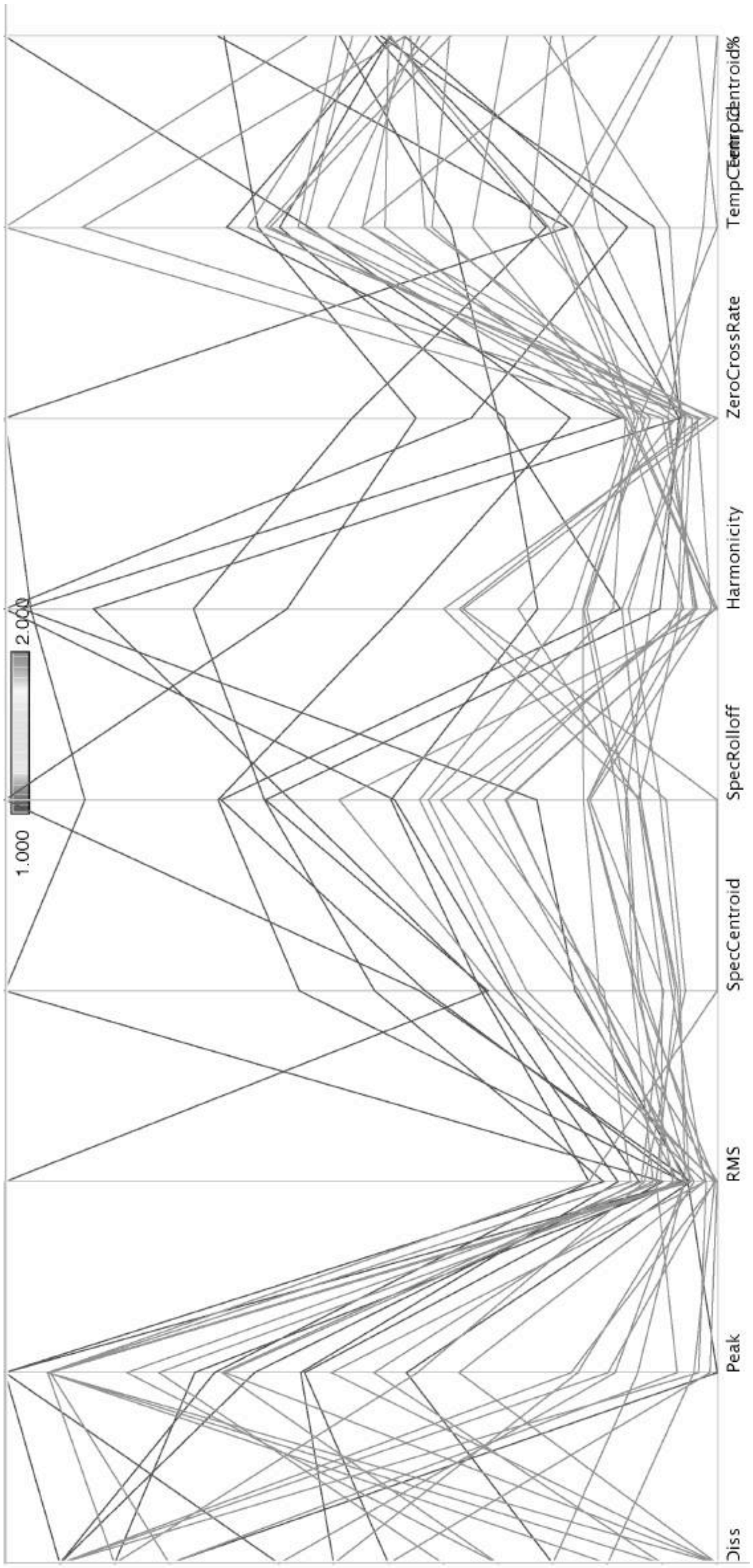


Figure 6–21. Parallel-plot visualization of each binary cluster.

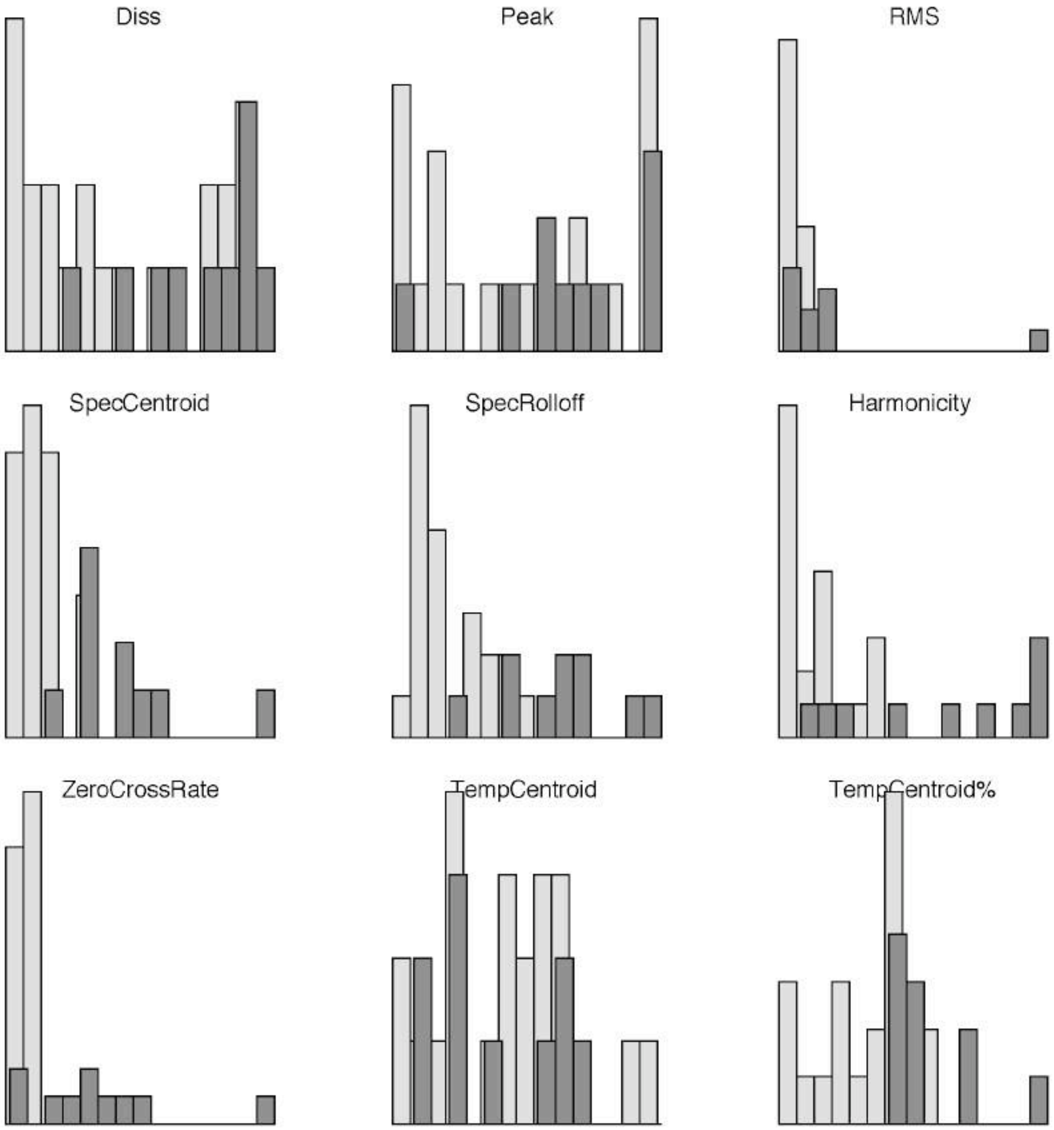


Figure 6–22. Histogram color matrix of the nine computed features of sound object.

Each of these visualizations of the multivariate data set provides a glimpse into the perceptual and physical factors that contribute to listener perception of sound-object dissonance. Casually speaking, brightness (as denoted by spectral centroid), brightness/noisiness (as denoted by spectral rolloff), and to a lesser extent harmonicity and zero-crossing rate, seem to correspond most strikingly to the classification of a sound object as relatively dissonant by trained listeners. In other words, brighter and noisier sounds tend to be classified as more dissonant. The results also contradict the claim of Chapter 5 that perceptually louder sounds (as approximately rated here by peak and RMS sample level) will be rated as more dissonant. It can be extrapolated that a favorable ("consonant") sound that is played louder will, up to a degree, be deemed as more consonant than its quieter counterpart.

These conclusions are supported by analytical examination of numerical data (e.g., via *r* values and k-means clustering), as well as through a variety of visualization schemes aimed at understanding multivariate data sets. Furthermore, these results support the majority of assertions from Chapter 5.

6.4 Future Work

To continue this work on sound-object dissonance analysis and classification, several avenues of exploration must be pursued. These can be divided into two broad areas: (1) enhanced testing methodologies based on lessons learned from this exercise, and (2) the employ of more robust analysis techniques of such tests. Both areas of improvement will lead to a better technical—and more importantly, musical—understanding of what is meant by sound-object dissonance in the context of the composition and performance of electroacoustic music.

First, future work should address larger-scale testing modalities. For example, greater confidence measures could be achieved by employing larger test populations. Furthermore, the same subjective questions should be asked in the listening test questionnaire of each sound object, allowing for correspondences among subjective ratings to be drawn. For example, had the listening test questionnaire used in this thesis asked each listener the same set of questions (e.g., “Which sound is more threatening?” and “Which sound is more intelligible?”) rather than dividing the questionnaire into isolated sets, then global correlation matrices and conditional probabilities could have been reported. As an example, it would be informative to know the degree to which listener assessment of a sound as being “more threatening” might also predict or correspond to their assessment of that sound as “more intelligible” for a given pair of sound objects.

Future tests should also employ larger feature sets. Many other audio feature measures are commonly used in the literature, and those reported here represent only a representative subset. In addition, time-varying feature extraction on a frame-by-frame basis from each sound object would produce true feature vectors, allowing for more subtle correspondences between time-varying objective data and listener classification of dissonance to be discovered.

Second, the work presented here can only form at best a portion of our musical understanding of sound-object dissonance in the context of composition, because more robust analysis techniques should be employed. First, the “best” weighting of the features extracted from each sound object as this weighting corresponds to correct prediction of sound-object dissonance should be explored. In addition, "best" weighting of features before clustering (i.e., a preprocessing stage) should be explored. Several machine-learning techniques for such an exploration currently exist and might prove useful for more robust analysis. Next, analysis of the testing results could be improved by employing additional multidimensional data visualization algorithms. One particular area of interest to me is to explore Chernoff faces (Chernoff 1973), which display multiple dimensions of data as features such as smile/frown and eyebrow height on a synthetically generated representation of the human face. It would be interesting to explore such a visualization as a means of meaningfully representing the multivariate data set gathered from the form of dissonance-rating listening test proposed in this thesis, particularly in such a way that the resulting Chernoff faces accurately model typical human facial response upon hearing sound objects.

7 POSTLUDE: BRIEF REMARKS ON DISSONANCE AND ELECTROACOUSTIC MUSIC

The results of the listening test presented in this thesis illustrate but a small portion of the role that psychological and physical assessment of sound-object features play in the broader realm of musical dissonance. The relatively recent acceptance (and return to) in the Western "art music" tradition of non-notated music—after a hiatus of over a millennium—in the form of improvisation, new media, musical interactivity, and music intended for playback solely over loudspeakers, necessitates new ways of considering musical dissonance and consonance, of tension and release. Music that exists as instructions on a page to a performer engenders a certain facility of analysis in this realm; for example, the minor second in this measure will sound more dissonant on the piano than the perfect fifth in the subsequent measure. That is, notation provides a starting point for discussion and analysis of musical architecture. Of course, much more is at play in the analysis of musical tension and release in notated music, as the isolated intervals present from moment to moment in a composition reflect but one dimension of musical experience. But at least the notation provides a fertile ground for discussion, and to some extent, understanding.

Electroacoustic music currently knows no such analytical starting point, for better or worse. We hear moments of harmonicity and inharmonicity, of brightness and darkness, of density or sparseness, of spatial trajectories that are fluid or disjunct—and we are left with but our ears and memories to interpret and analyze the experience. That being said, electroacoustic music, as with all music, is ultimately interesting and successful to the extent that it engages its listeners in some meaningful way, apart from any analytical framework. With remarkably few exceptions, fruitful analysis and dissection does not lead to a set of rules that enable the successful building of new works; it seems only the combination of intuition and creativity can do that.

But perhaps a subjective and objective analysis of sound objects, which form the basis of the composition of much electroacoustic music, such as that presented here will inform an enhanced understanding and acceptance of such music. Clearly, sound-object dissonance will provide only a small part of such an understanding, together with theories of gesture, texture, and space, but it may serve as a stepping-stone in some small way.

APPENDIX A: LISTENING TEST PROTOCOL

Musical Dissonance Ratings of Sound Objects by Trained Listeners

OFFICE USE ONLY Subject ID:

Thank you for participating in this listening study. Before we begin, please complete the informational questionnaire below.

Today's Date:

Month Day Year

University of Miami Major (if applicable)

Age: **Sex:** M F **Years of Musical Training:**

Primary Musical Instrument:

Please briefly describe any documented hearing problems:

You will now hear a series of audio sound samples of approximately 10 sec duration each, played one at a time. There are 32 pairs of sounds in the entire study; these pairs of sounds are not necessarily related in any way. The study takes approximately 60 minutes to complete.

Several times during the study, I will announce a one-minute break in which a randomly chosen musical passage is played to "clean" your ears.

The sounds will be presented as follows:

[**Sound 1a**] — [click + 2 sec silence] — [**Sound 1b**] — [click + 2 sec silence]
[**Sound 1a**] — [click + 2 sec silence] — [**Sound 1b**] — [triple-click + 15 sec silence]

[**Sound 2a**] — [click + 2 sec silence] — [**Sound 2b**] — [click + 2 sec silence]
[**Sound 2a**] — [click + 2 sec silence] — [**Sound 2b**] — [triple-click + 15 sec silence]

•
•
•

[**Sound 32a**] — [click + 2 sec silence] — [**Sound 32b**] — [click + 2 sec silence]
[**Sound 32a**] — [click + 2 sec silence] — [**Sound 32b**] —

For each pair of sounds, please quickly write a word in the spaces provided to describe your response to each of the sounds in that pair. Please also circle your response to each question. There are no correct or incorrect answers.

You are free to define the questions on your own terms. For example, you are free to consider sounds as "consonant" or "dissonant" however you wish, provided you attempt to be consistent.

Please make every attempt to separate the context in which you hear each sound from your assessment of it. In other words, please consider each sound on its own terms, irrespective of neighboring sound pairs.

Example

Sound 0a:__________________ Sound 0b: __________________

Which sound is more frustrating to hear?	Sound 0a	Sound 0b
Which sound is more difficult to understand?	Sound 0a	Sound 0b
Which sound is more intelligible?	Sound 0a	Sound 0b
Which sound is more dissonant?	Sound 0a	Sound 0b

Sound Pairs 1–8

Sound 1a:____________________ Sound 1b: ____________________

Which sound is more frustrating to hear?	Sound 1a	Sound 1b
Which sound is more difficult to understand?	Sound 1a	Sound 1b
Which sound is more intelligible?	Sound 1a	Sound 1b
Which sound is more dissonant?	Sound 1a	Sound 1b

Sound 2a:____________________ Sound 2b: ____________________

Which sound is more frustrating to hear?	Sound 2a	Sound 2b
Which sound is more difficult to understand?	Sound 2a	Sound 2b
Which sound is more intelligible?	Sound 2a	Sound 2b
Which sound is more dissonant?	Sound 2a	Sound 2b

Sound 3a:____________________ Sound 3b: ____________________

Which sound is more frustrating to hear?	Sound 3a	Sound 3b
Which sound is more difficult to understand?	Sound 3a	Sound 3b
Which sound is more intelligible?	Sound 3a	Sound 3b
Which sound is more dissonant?	Sound 3a	Sound 3b

Sound 4a:____________________ Sound 4b: ____________________

Which sound is more frustrating to hear?	Sound 4a	Sound 4b
Which sound is more difficult to understand?	Sound 4a	Sound 4b
Which sound is more intelligible?	Sound 4a	Sound 4b
Which sound is more dissonant?	Sound 4a	Sound 4b

Sound 5a:____________________ Sound 5b: ____________________

Which sound is more frustrating to hear?	Sound 5a	Sound 5b
Which sound is more difficult to understand?	Sound 5a	Sound 5b
Which sound is more intelligible?	Sound 5a	Sound 5b
Which sound is more dissonant?	Sound 5a	Sound 5b

Sound 6a:____________________ Sound 6b: ____________________

Which sound is more frustrating to hear?	Sound 6a	Sound 6b
Which sound is more difficult to understand?	Sound 6a	Sound 6b
Which sound is more intelligible?	Sound 6a	Sound 6b
Which sound is more dissonant?	Sound 6a	Sound 6b

Sound 7a:__________________ Sound 7b: __________________

Which sound is more frustrating to hear?	Sound 7a	Sound 7b
Which sound is more difficult to understand?	Sound 7a	Sound 7b
Which sound is more intelligible?	Sound 7a	Sound 7b
Which sound is more dissonant?	Sound 7a	Sound 7b

Sound 8a:__________________ Sound 8b: __________________

Which sound is more frustrating to hear?	Sound 8a	Sound 8b
Which sound is more difficult to understand?	Sound 8a	Sound 8b
Which sound is more intelligible?	Sound 8a	Sound 8b
Which sound is more dissonant?	Sound 8a	Sound 8b

[A short musical passage will now be played to clean your ears.]

Sound Pairs 9–16

Sound 9a:__________________ Sound 9b: __________________

Which sound is more threatening?	Sound 9a	Sound 9b
Which sound is more annoying?	Sound 9a	Sound 9b
Which sound is more dissonant?	Sound 9a	Sound 9b

Sound 10a:__________________ Sound 10b: __________________

Which sound is more threatening?	Sound 10a	Sound 10b
Which sound is more annoying?	Sound 10a	Sound 10b
Which sound is more dissonant?	Sound 10a	Sound 10b

Sound 11a:__________________ Sound 11b: __________________

Which sound is more threatening?	Sound 11a	Sound 11b
Which sound is more annoying?	Sound 11a	Sound 11b
Which sound is more dissonant?	Sound 11a	Sound 11b

Sound 12a:__________________ Sound 12b: __________________

Which sound is more threatening?	Sound 12a	Sound 12b
Which sound is more annoying?	Sound 12a	Sound 12b
Which sound is more dissonant?	Sound 12a	Sound 12b

Sound 13a:__________________ Sound 13b: __________________

Which sound is more threatening?	Sound 13a	Sound 13b
Which sound is more annoying?	Sound 13a	Sound 13b
Which sound is more dissonant?	Sound 13a	Sound 13b

Sound 14a:__________________ Sound 14b: __________________

Which sound is more threatening?	Sound 14a	Sound 14b
Which sound is more annoying?	Sound 14a	Sound 14b
Which sound is more dissonant?	Sound 14a	Sound 14b

Sound 15a:__________________ Sound 15b: __________________

Which sound is more threatening?	Sound 15a	Sound 15b
Which sound is more annoying?	Sound 15a	Sound 15b
Which sound is more dissonant?	Sound 15a	Sound 15b

Sound 16a:__________________ Sound 16b: __________________

Which sound is more threatening?	Sound 16a	Sound 16b
Which sound is more annoying?	Sound 16a	Sound 16b
Which sound is more dissonant?	Sound 16a	Sound 16b

[A short musical passage will now be played to clean your ears.]

Sound Pairs 17–24

Sound 17a:__________________ Sound 17b: __________________

Which sound is more easily recognizable?	Sound 17a	Sound 17b
Which sound is more predictable?	Sound 17a	Sound 17b
Which sound is more consistent?	Sound 17a	Sound 17b
Which sound is more dissonant?	Sound 17a	Sound 17b

Sound 18a:__________________ Sound 18b: __________________

Which sound is more easily recognizable?	Sound 18a	Sound 18b
Which sound is more predictable?	Sound 18a	Sound 18b
Which sound is more consistent?	Sound 18a	Sound 18b
Which sound is more dissonant?	Sound 18a	Sound 18b

Sound 19a:__________________ Sound 19b: __________________

Which sound is more easily recognizable?	Sound 19a	Sound 19b
Which sound is more predictable?	Sound 19a	Sound 19b
Which sound is more consistent?	Sound 19a	Sound 19b
Which sound is more dissonant?	Sound 19a	Sound 19b

Sound 20a:__________________ Sound 20b: __________________

Which sound is more easily recognizable?	Sound 20a	Sound 20b
Which sound is more predictable?	Sound 20a	Sound 20b
Which sound is more consistent?	Sound 20a	Sound 20b
Which sound is more dissonant?	Sound 20a	Sound 20b

Sound 21a:__________________ Sound 21b: __________________

Which sound is more easily recognizable?	Sound 21a	Sound 21b
Which sound is more predictable?	Sound 21a	Sound 21b
Which sound is more consistent?	Sound 21a	Sound 21b
Which sound is more dissonant?	Sound 21a	Sound 21b

Sound 22a:__________________ Sound 22b: __________________

Which sound is more easily recognizable?	Sound 22a	Sound 22b
Which sound is more predictable?	Sound 22a	Sound 22b
Which sound is more consistent?	Sound 22a	Sound 22b
Which sound is more dissonant?	Sound 22a	Sound 22b

Sound 23a:__________________ Sound 23b: __________________

Which sound is more easily recognizable?	Sound 23a	Sound 23b
Which sound is more predictable?	Sound 23a	Sound 23b
Which sound is more consistent?	Sound 23a	Sound 23b
Which sound is more dissonant?	Sound 23a	Sound 23b

Sound 24a:____________________ Sound 24b: ____________________

Which sound is more easily recognizable?	Sound 24a	Sound 24b
Which sound is more predictable?	Sound 24a	Sound 24b
Which sound is more consistent?	Sound 24a	Sound 24b
Which sound is more dissonant?	Sound 24a	Sound 24b

[A short musical passage will now be played to clean your ears.]

Sound Pairs 25–32

Sound 25a:____________________ Sound 25b: ____________________

Which sound is more pitched?	Sound 25a	Sound 25b
Which sound is smoother?	Sound 25a	Sound 25b
Which sound is more regular?	Sound 25a	Sound 25b
Which sound is more dissonant?	Sound 25a	Sound 25b

Sound 26a:____________________ Sound 26b: ____________________

Which sound is more pitched?	Sound 26a	Sound 26b
Which sound is smoother?	Sound 26a	Sound 26b
Which sound is more regular?	Sound 26a	Sound 26b
Which sound is more dissonant?	Sound 26a	Sound 26b

Sound 27a:____________________ Sound 27b: ____________________

Which sound is more pitched?	Sound 27a	Sound 27b
Which sound is smoother?	Sound 27a	Sound 27b
Which sound is more regular?	Sound 27a	Sound 27b
Which sound is more dissonant?	Sound 27a	Sound 27b

Sound 28a:____________________ Sound 28b: ____________________

Which sound is more pitched?	Sound 28a	Sound 28b
Which sound is smoother?	Sound 28a	Sound 28b
Which sound is more regular?	Sound 28a	Sound 28b
Which sound is more dissonant?	Sound 28a	Sound 28b

Sound 29a:__________________ Sound 29b: __________________

Which sound is more pitched?	Sound 29a	Sound 29b
Which sound is smoother?	Sound 29a	Sound 29b
Which sound is more regular?	Sound 29a	Sound 29b
Which sound is more dissonant?	Sound 29a	Sound 29b

Sound 30a:__________________ Sound 30b: __________________

Which sound is more pitched?	Sound 30a	Sound 30b
Which sound is smoother?	Sound 30a	Sound 30b
Which sound is more regular?	Sound 30a	Sound 30b
Which sound is more dissonant?	Sound 30a	Sound 30b

Sound 31a:__________________ Sound 31b: __________________

Which sound is more pitched?	Sound 31a	Sound 31b
Which sound is smoother?	Sound 31a	Sound 31b
Which sound is more regular?	Sound 31a	Sound 31b
Which sound is more dissonant?	Sound 31a	Sound 31b

Sound 32a:__________________ Sound 32b: __________________

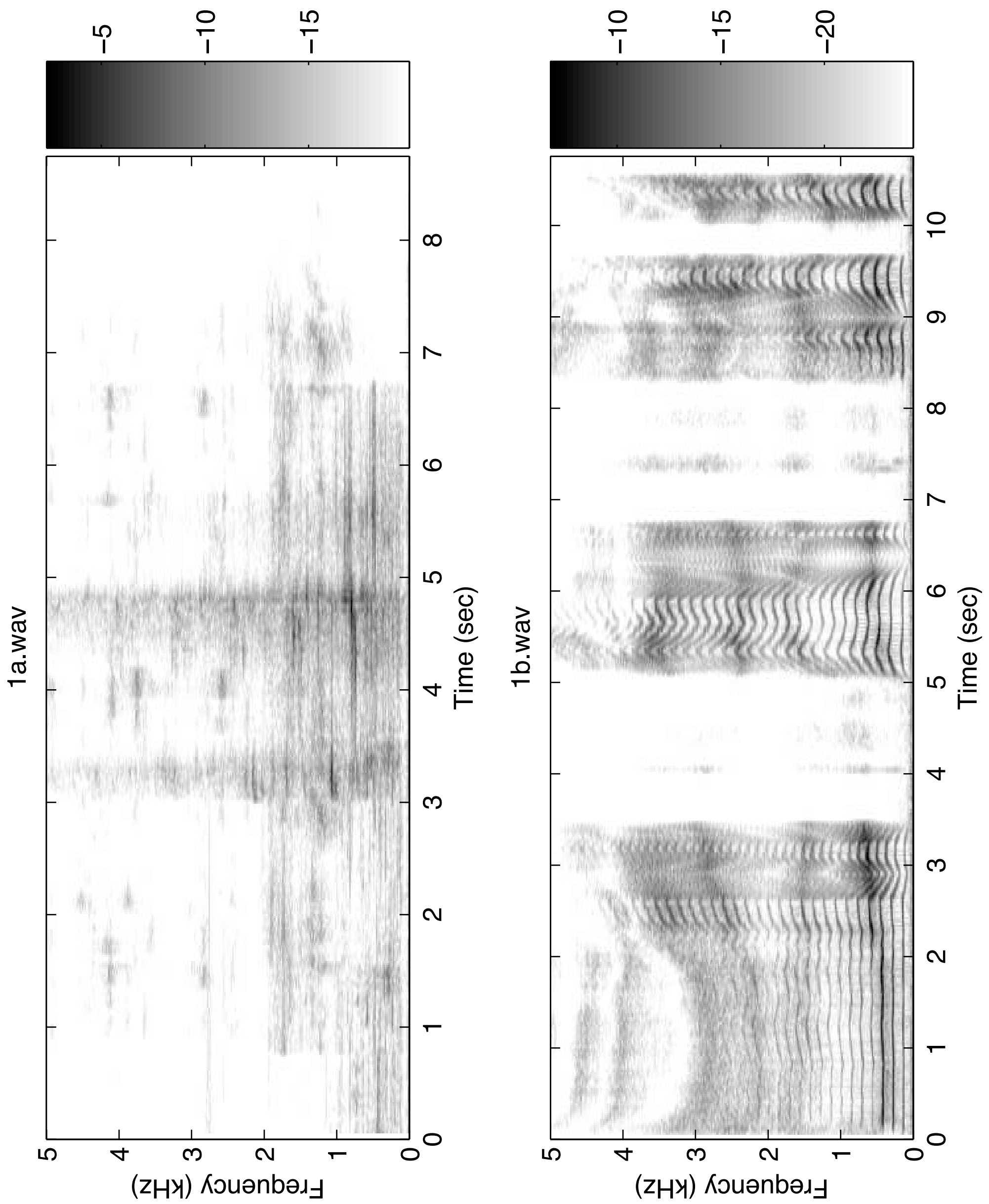
Which sound is more pitched?	Sound 32a	Sound 32b
Which sound is smoother?	Sound 32a	Sound 32b
Which sound is more regular?	Sound 32a	Sound 32b
Which sound is more dissonant?	Sound 32a	Sound 32b

Thank you for participating in the listening test!

APPENDIX B: SPECTROGRAMS OF TEST SOUNDS

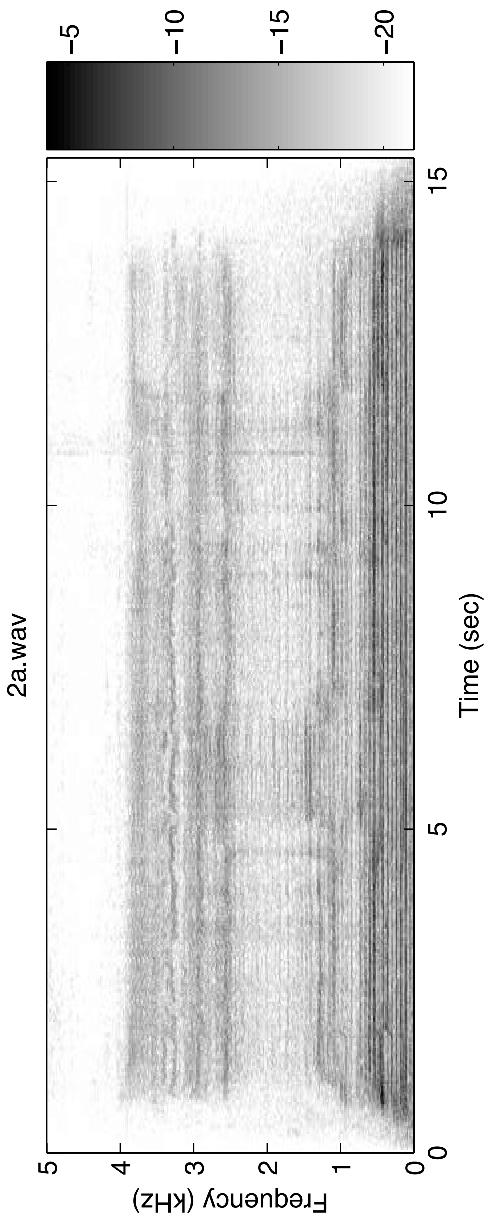
As a visual display aid for subjective analysis, spectrograms were computed and graphed in MATLAB for each pair of sound files used in the listening test. Each graph shown in the following pages plots frequency in kilohertz versus time in seconds for each sound; the grayscale legend on the right of each graph indicates the relative amplitude of each frequency in decibels (dB) in the graph according to pixel brightness along the colormap. Underneath each graph, the corresponding questionnaire statistics are listed.

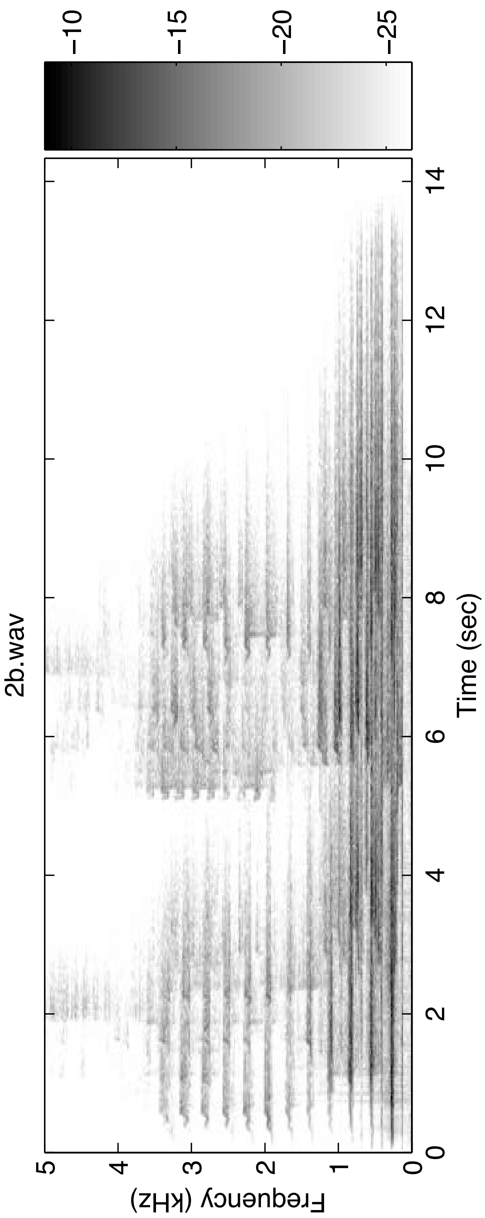
Pair #1



	A	B
More frustating	41.9%	**58.1%**
More difficult to understand	**80.6%**	19.4%
More intelligible	32.3%	**67.7%**
More dissonant	**71.0%**	29.0%

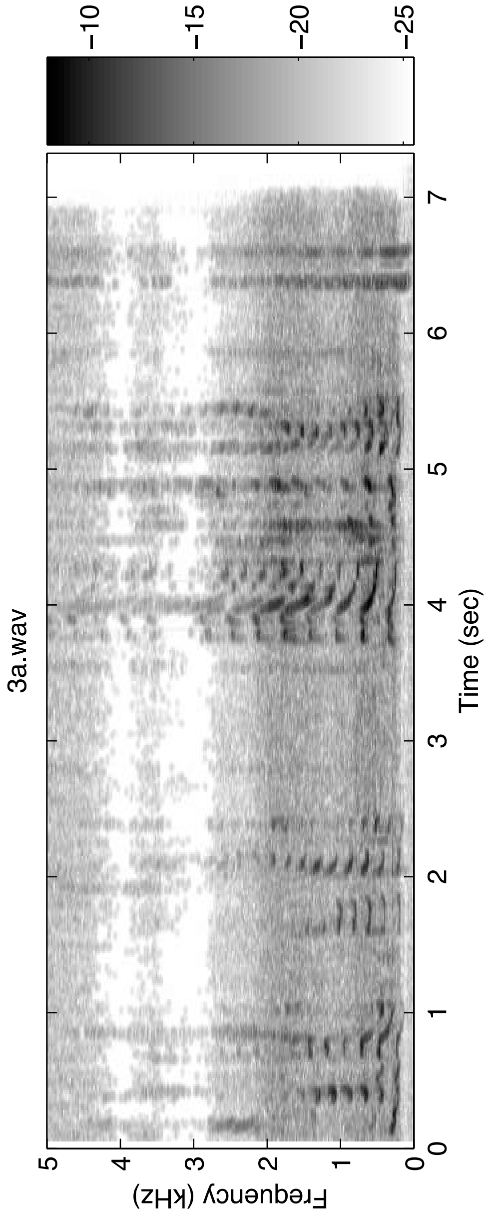
Pair #2

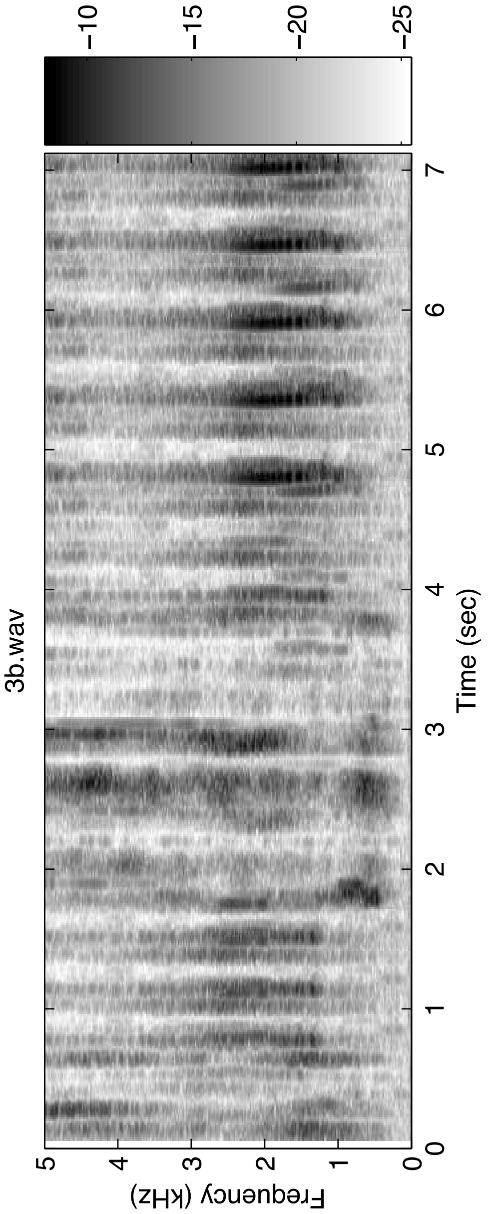




	A	B
More frustating	**80.6%**	19.4%
More difficult to understand	**87.1%**	12.9%
More intelligible	19.4%	**80.6%**
More dissonant	**93.5%**	6.5%

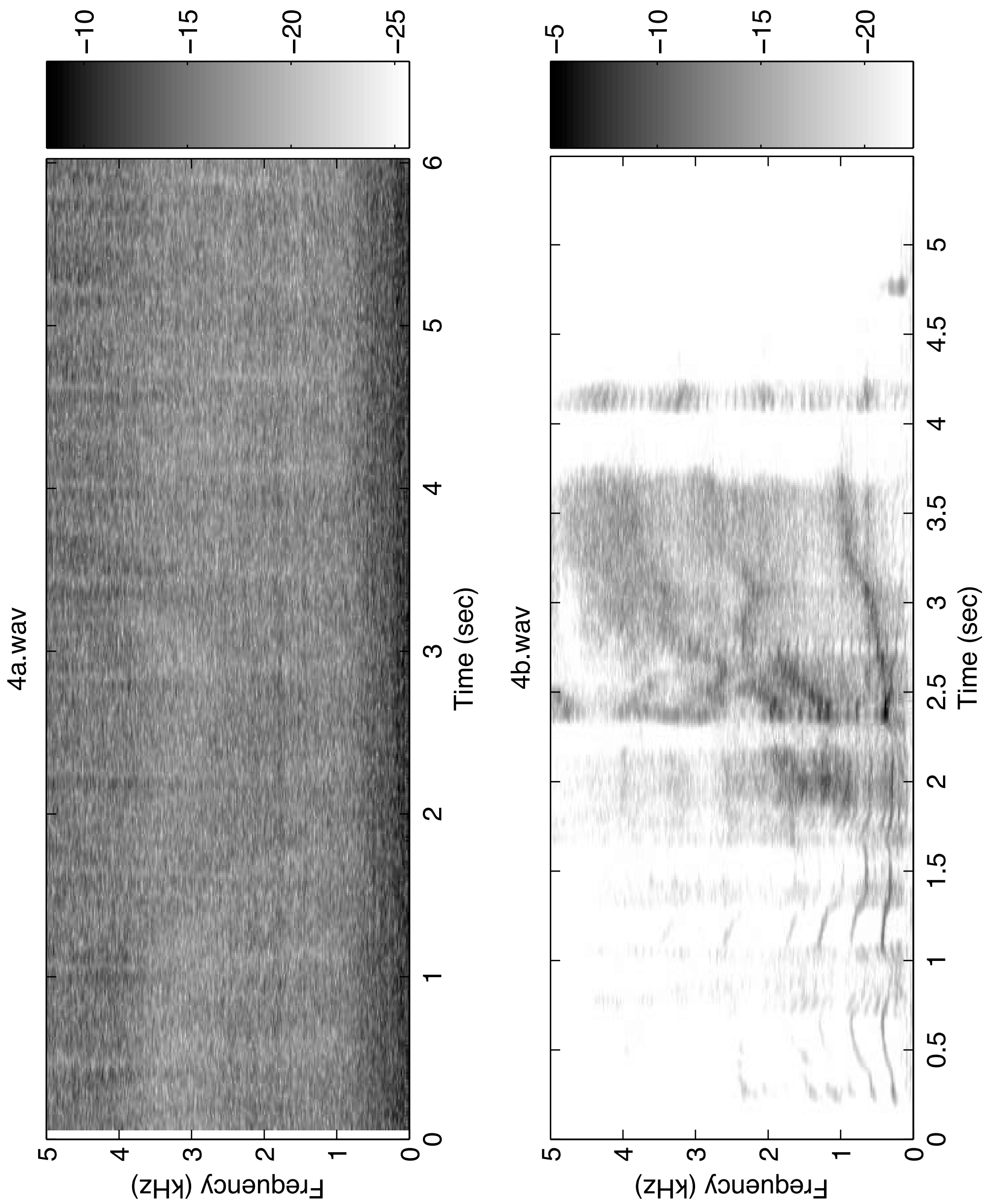
Pair #3





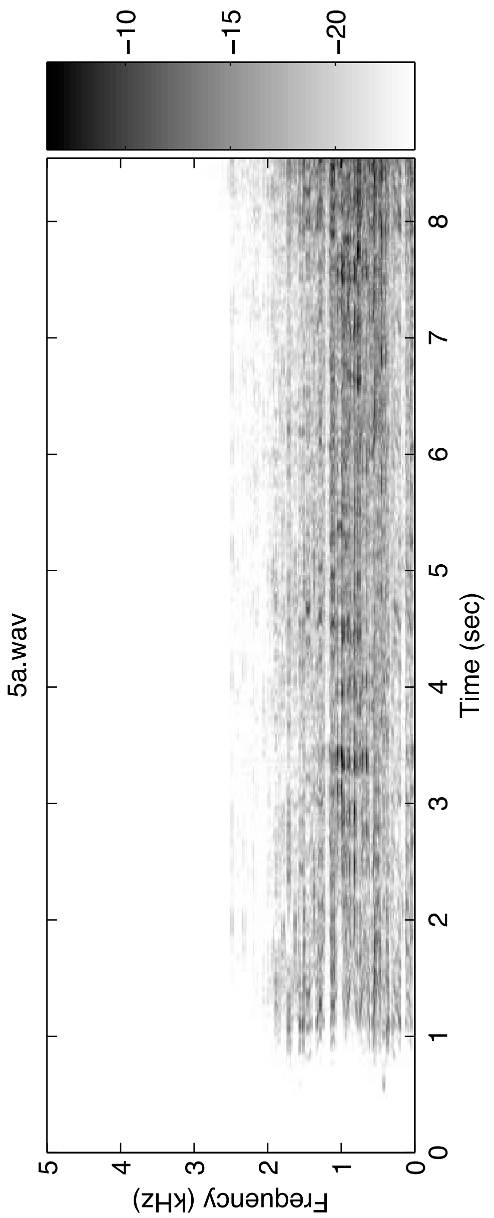
	A	B
More frustating	16.1%	**83.9%**
More difficult to understand	25.8%	**74.2%**
More intelligible	**58.1%**	41.9%
More dissonant	6.5%	**93.5%**

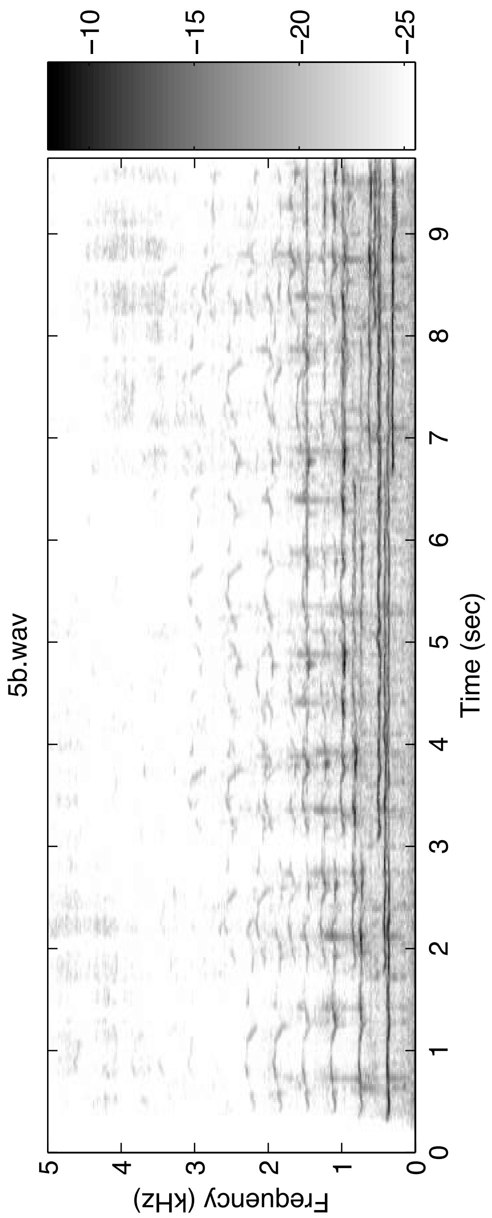
Pair #4



	A	B
More frustating	41.9%	**58.1%**
More difficult to understand	**96.8%**	3.2%
More intelligible	19.4%	**80.6%**
More dissonant	**67.7%**	32.3%

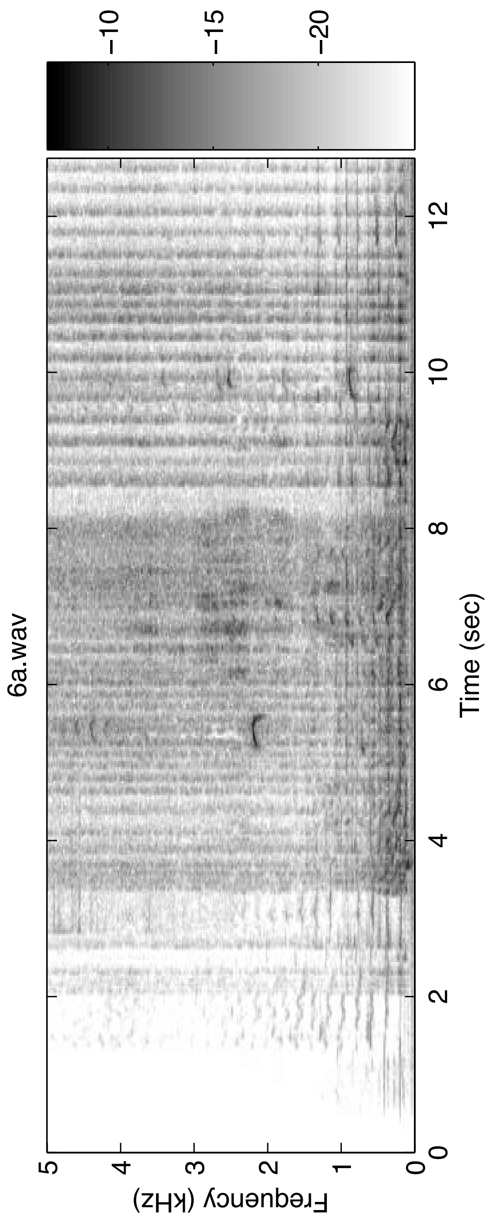
Pair #5

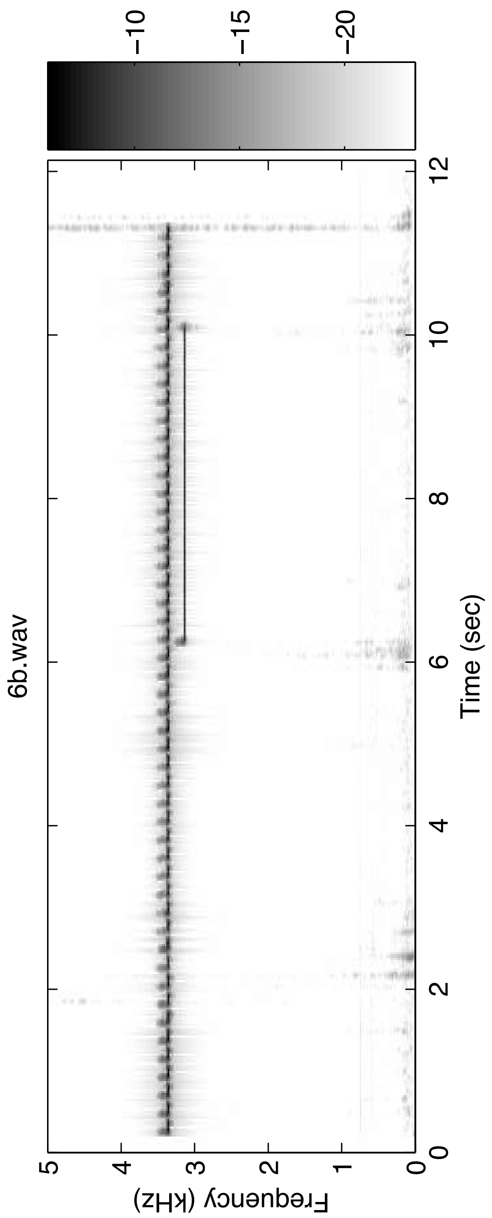




	A	B
More frustating	45.2%	**54.8%**
More difficult to understand	**71.0%**	29.0%
More intelligible	30.0%	**70.0%**
More dissonant	70.0%	**30.0%**

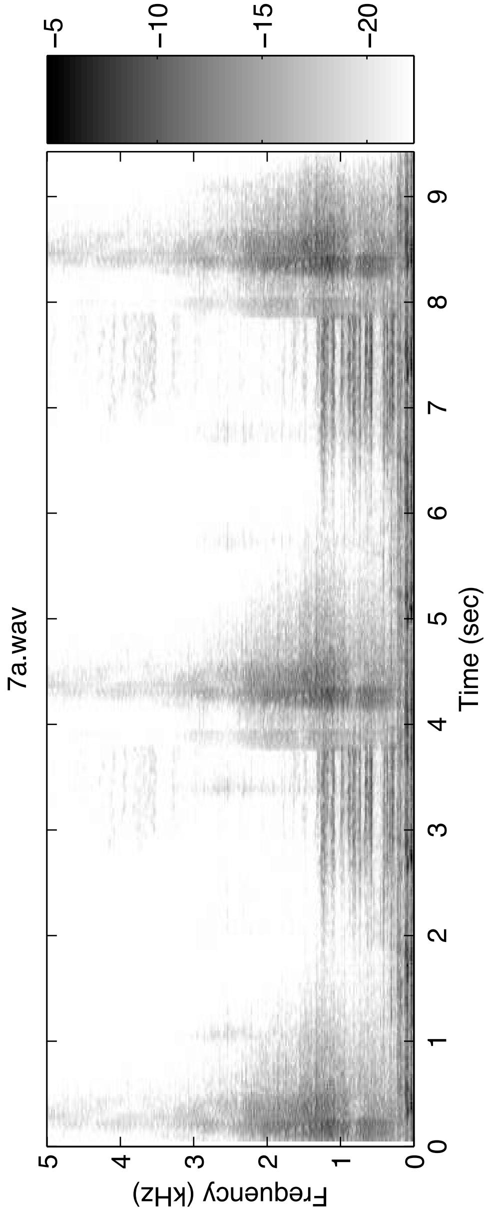
Pair #6

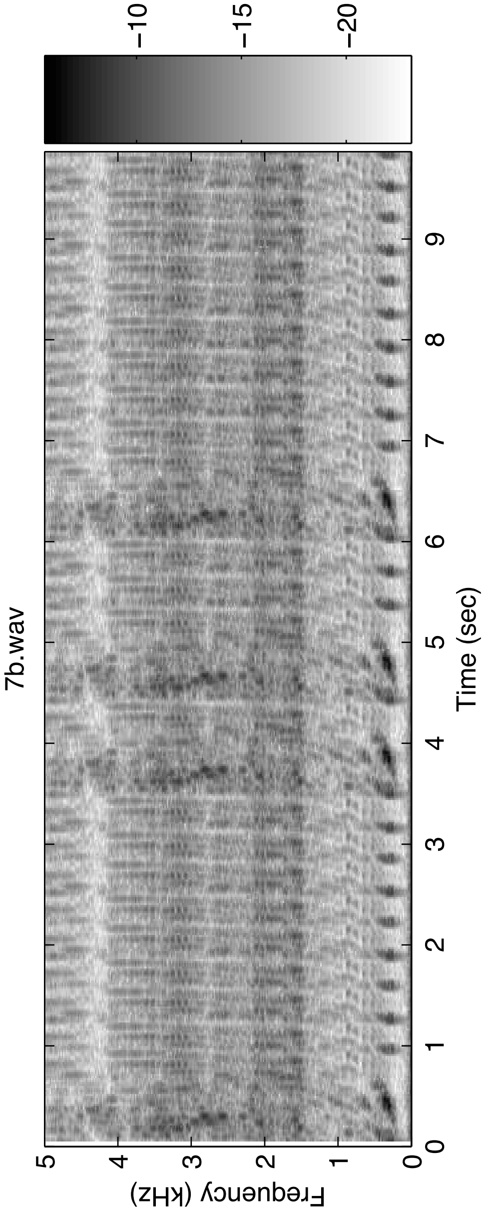




	A	B
More frustating	12.9%	**87.1%**
More difficult to understand	**87.1%**	12.9%
More intelligible	29.0%	**71.0%**
More dissonant	9.7%	**90.3%**

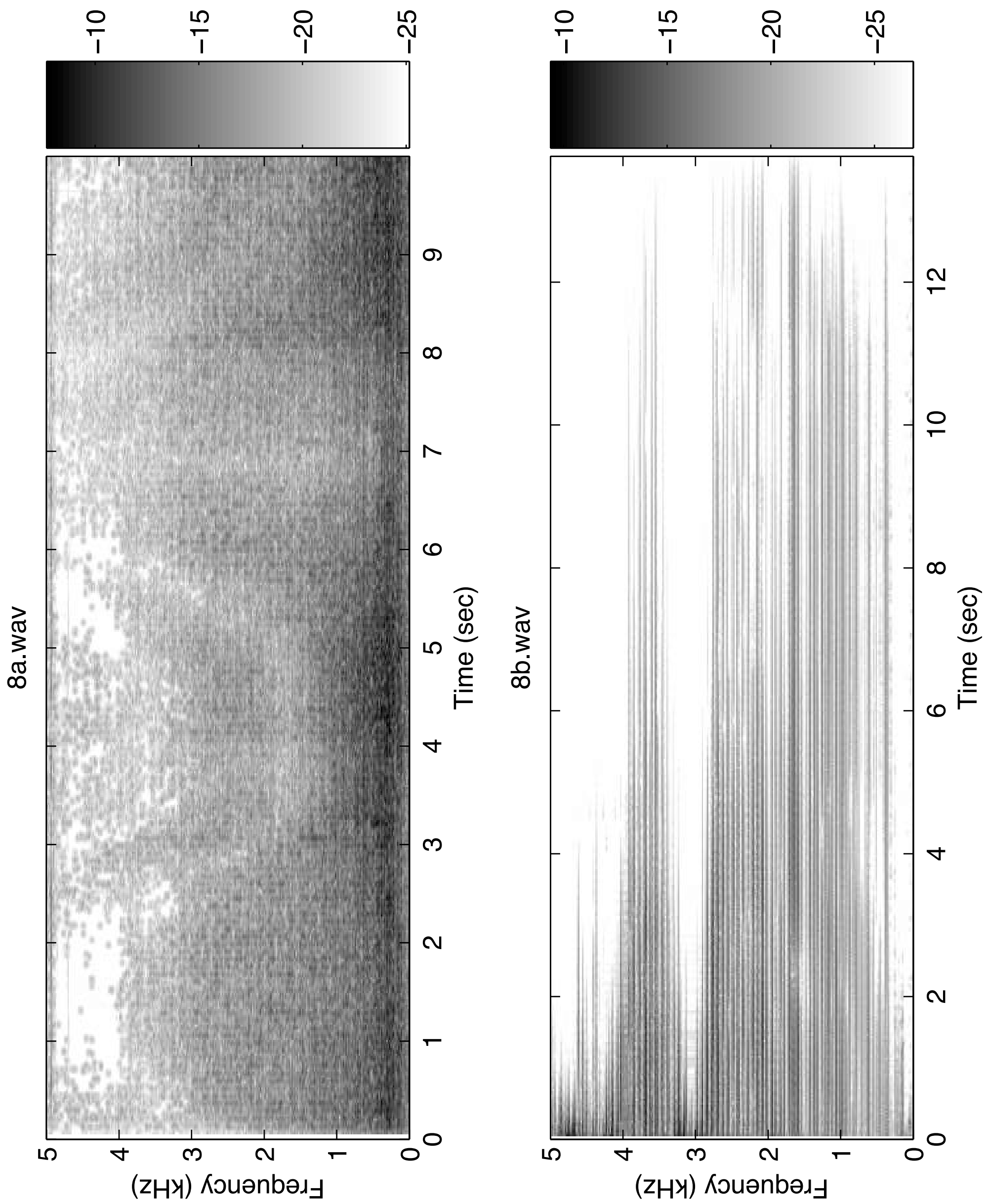
Pair #7





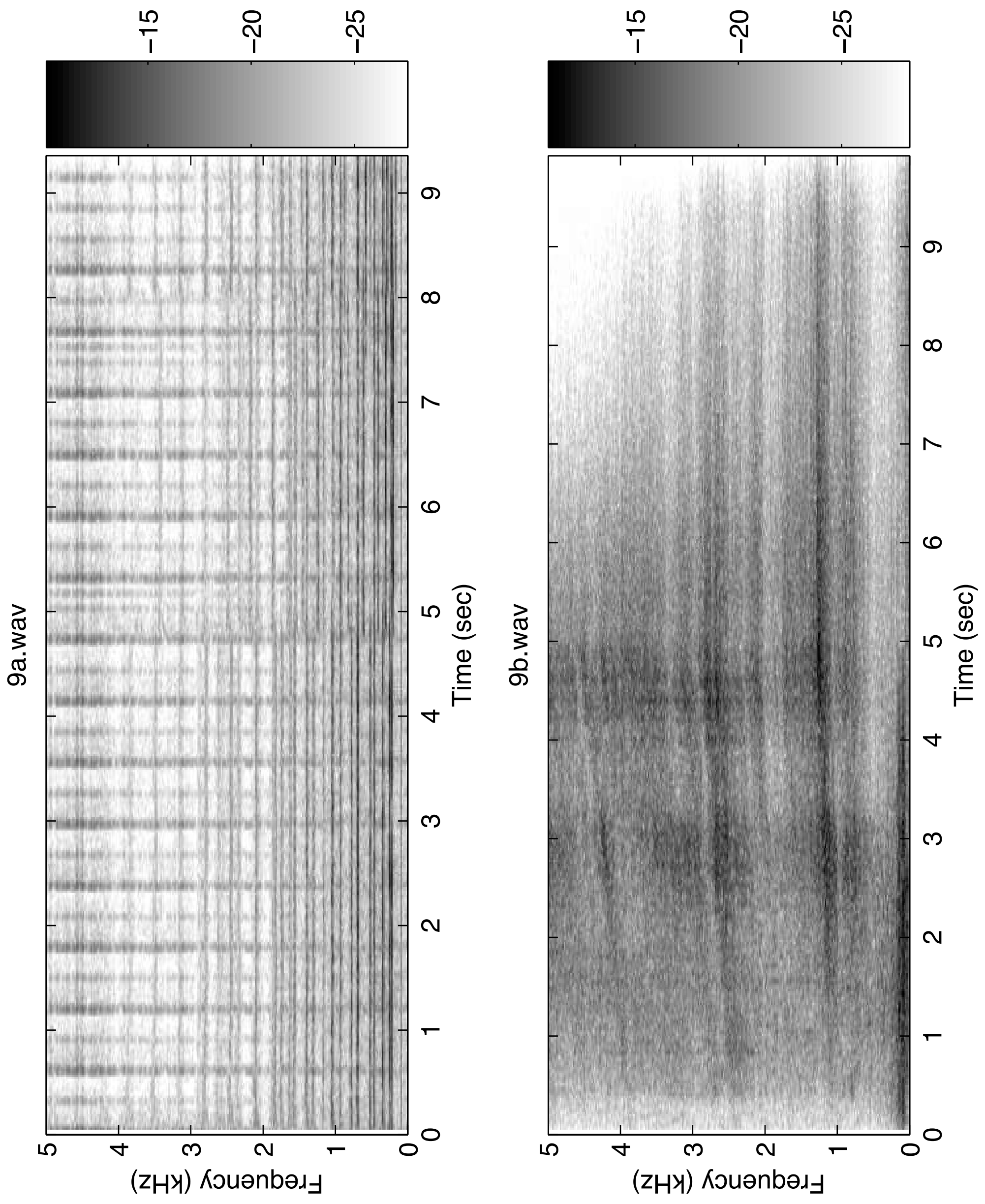
	A	B
More frustating	16.1%	**83.9%**
More difficult to understand	25.8%	**74.2%**
More intelligible	**77.4%**	22.6%
More dissonant	38.7%	**61.3%**

Pair #8



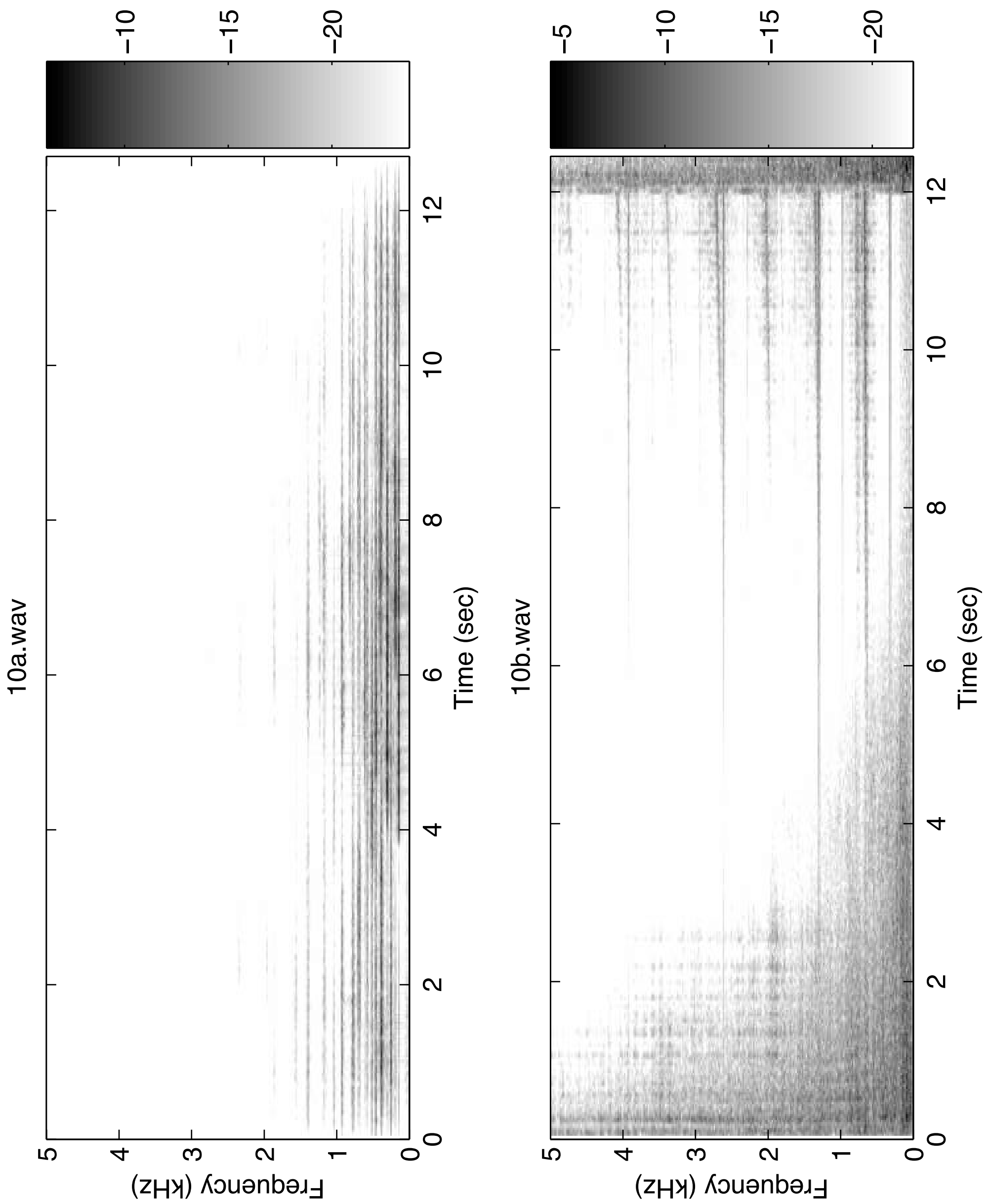
	A	B
More frustating	**54.8%**	45.2%
More difficult to understand	25.8%	**74.2%**
More intelligible	**74.2%**	25.8%
More dissonant	22.6%	**77.4%**

Pair #9



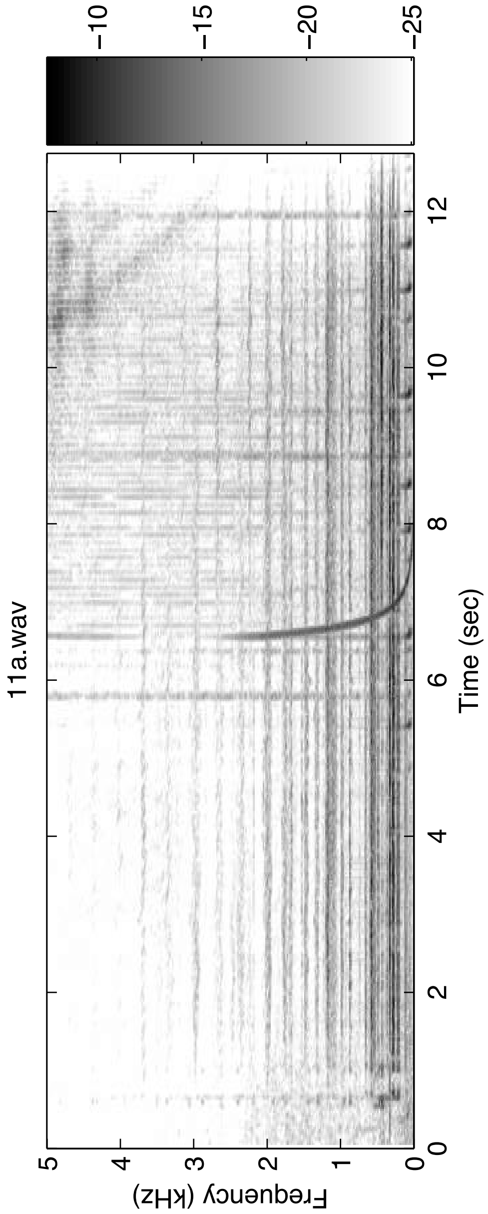
	A	B
More threatening	0.0%	100.0%
More annoying	16.1%	83.9%
More dissonant	3.2%	96.8%

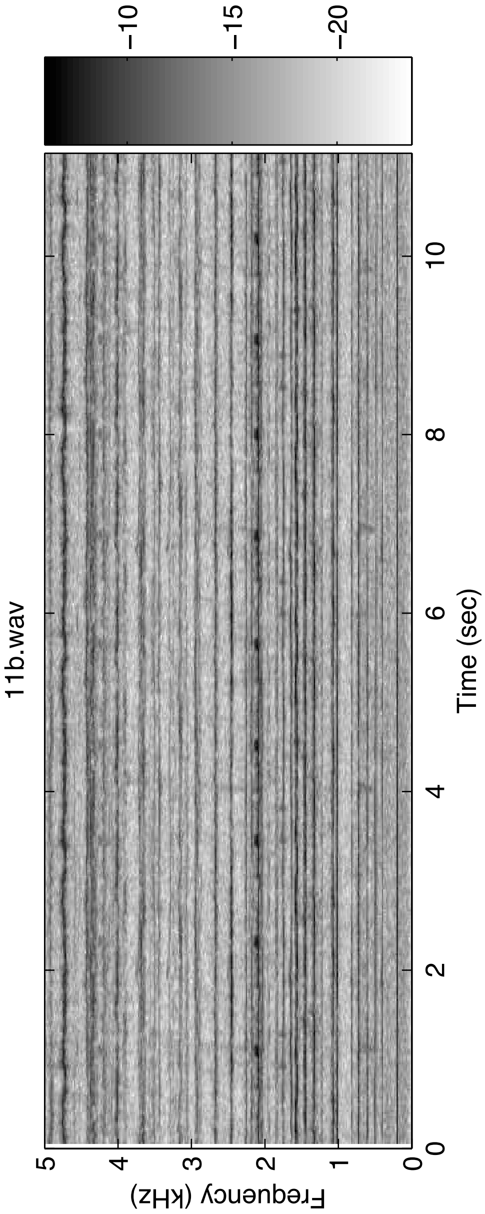
Pair #10



	A	B
More threatening	6.5%	93.5%
More annoying	9.7%	90.3%
More dissonant	6.5%	93.5%

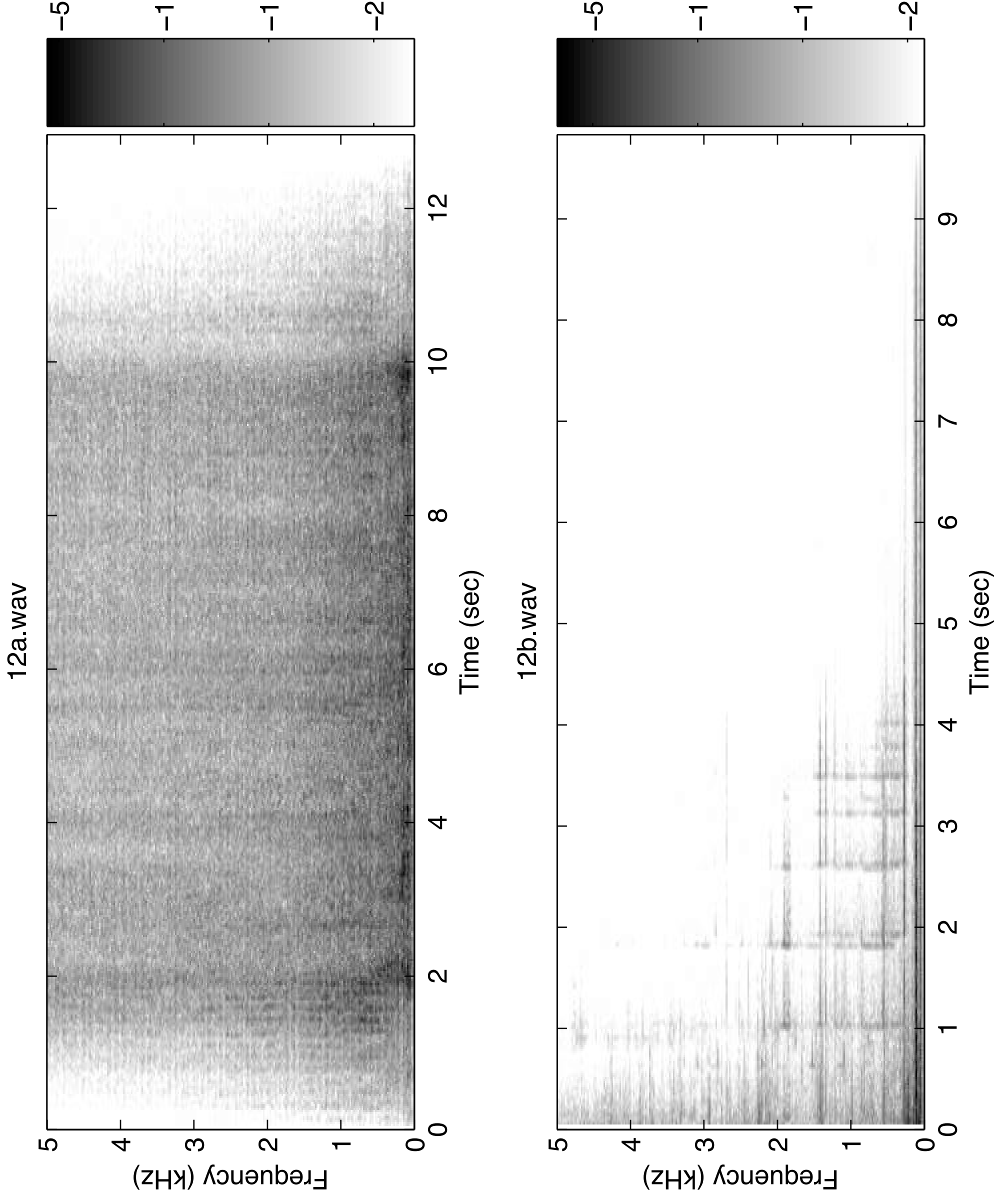
Pair #11





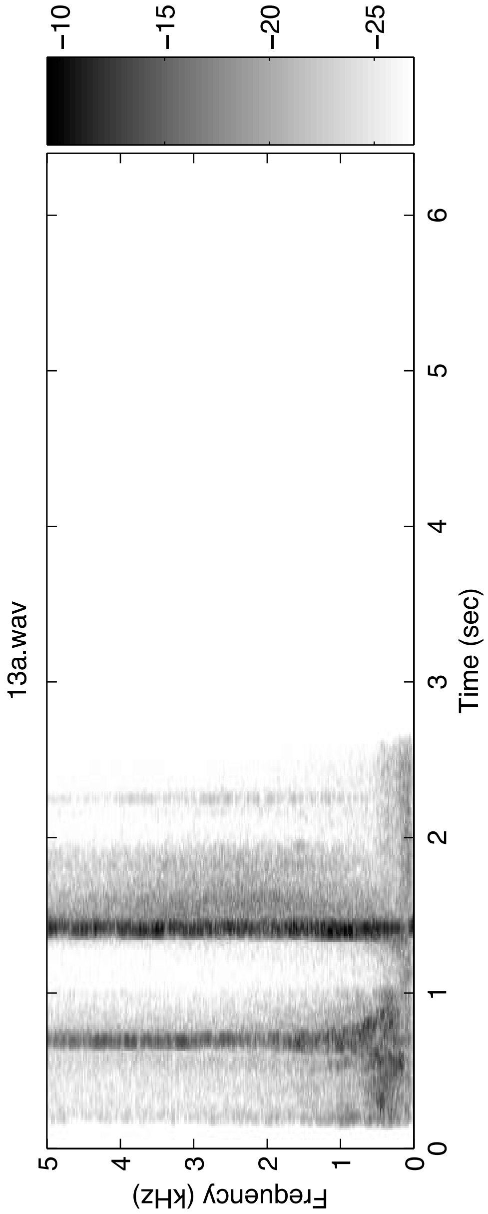
	A	B
More threatening	3.3%	96.7%
More annoying	0.0%	100.0%
More dissonant	0.0%	100.0%

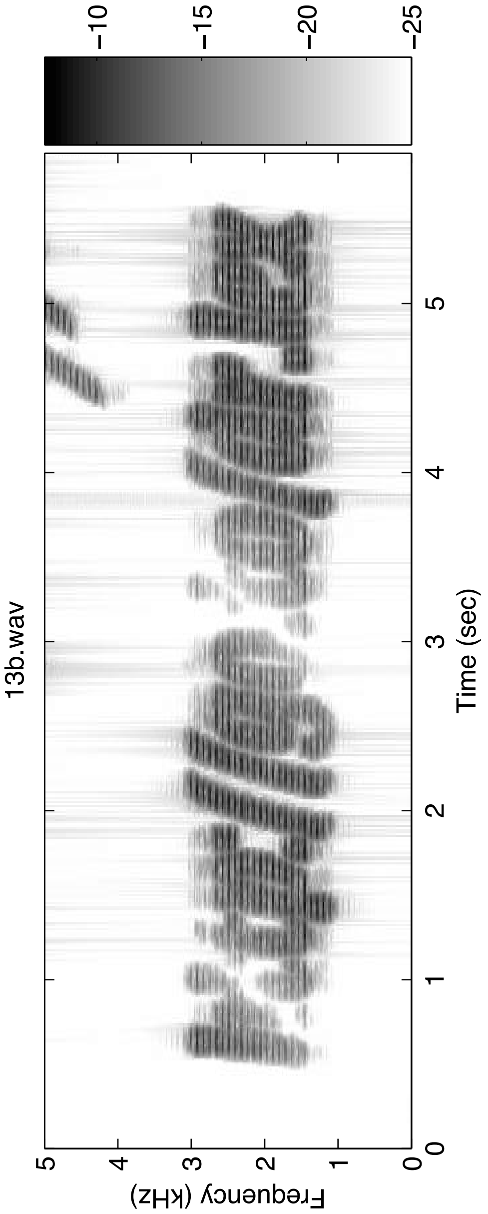
Pair #12



	A	B
More threatening	80.6%	19.4%
More annoying	80.6%	19.4%
More dissonant	64.5%	35.5%

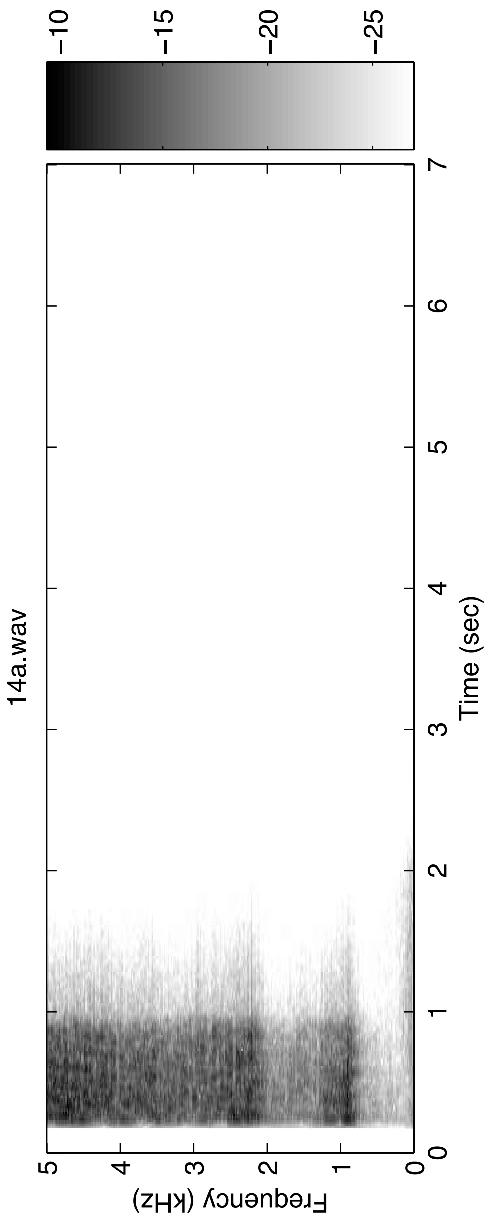
Pair #13

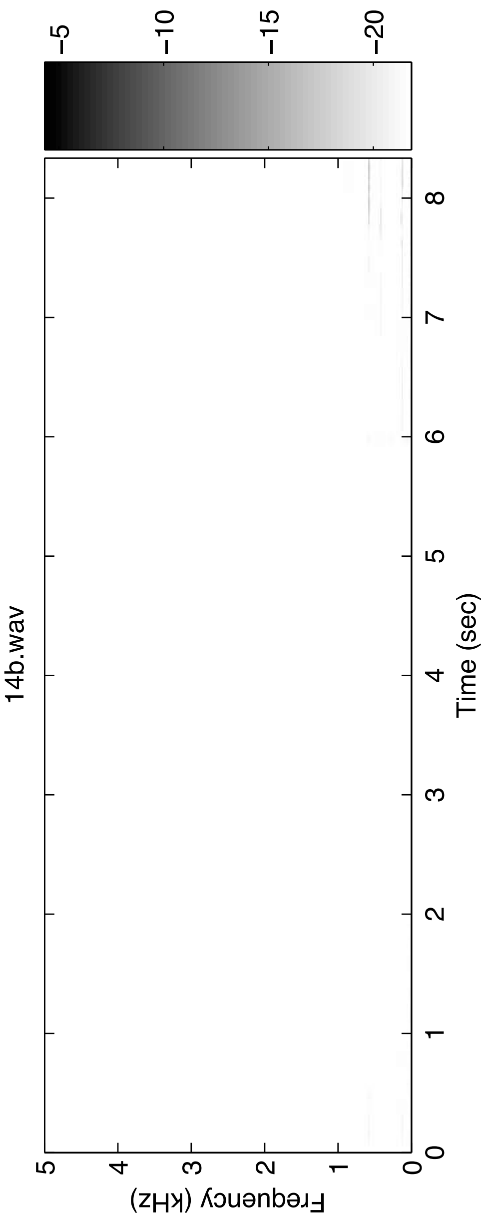




	A	B
More threatening	38.7%	61.3%
More annoying	0.0%	100.0%
More dissonant	3.2%	96.8%

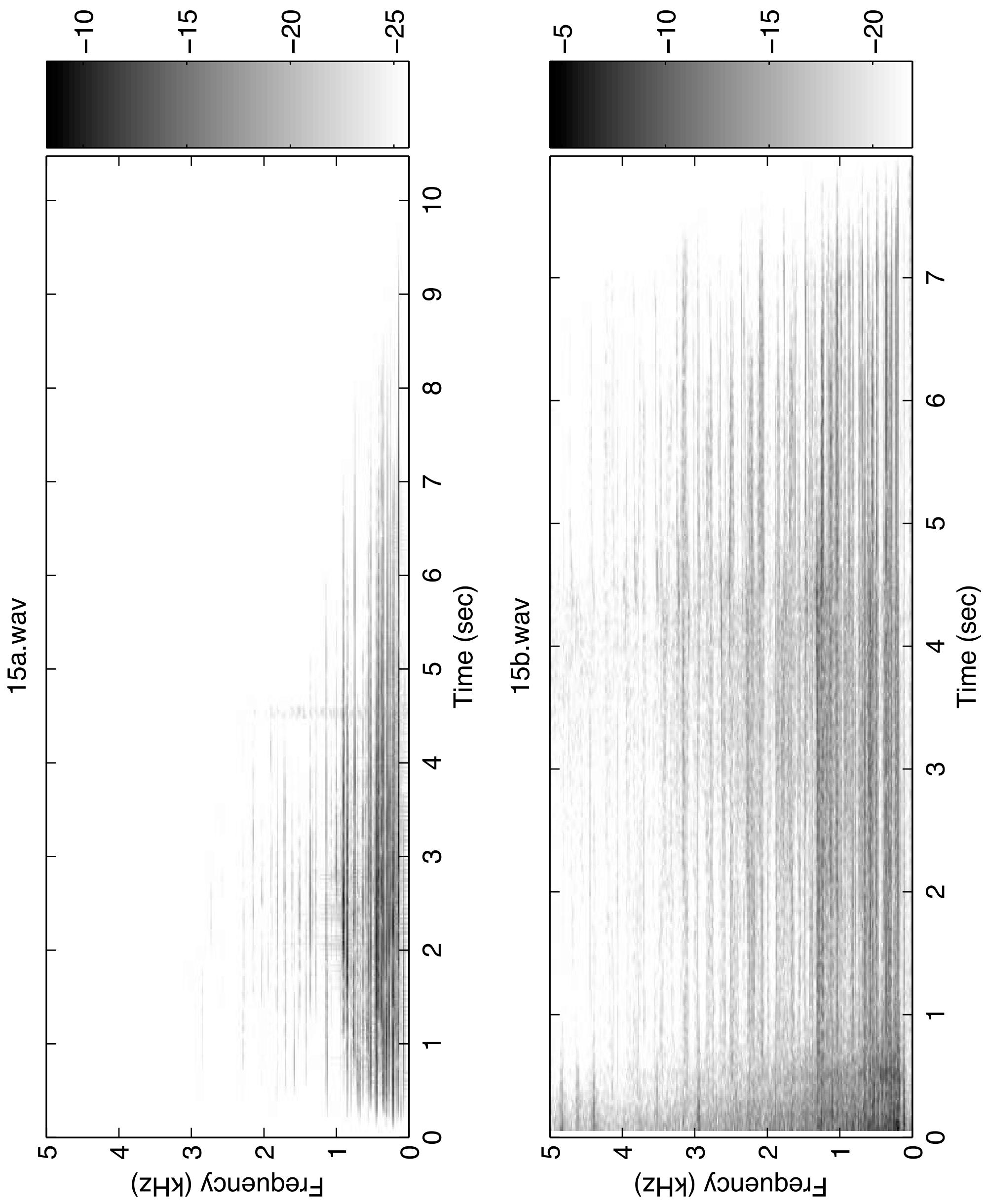
Pair #14





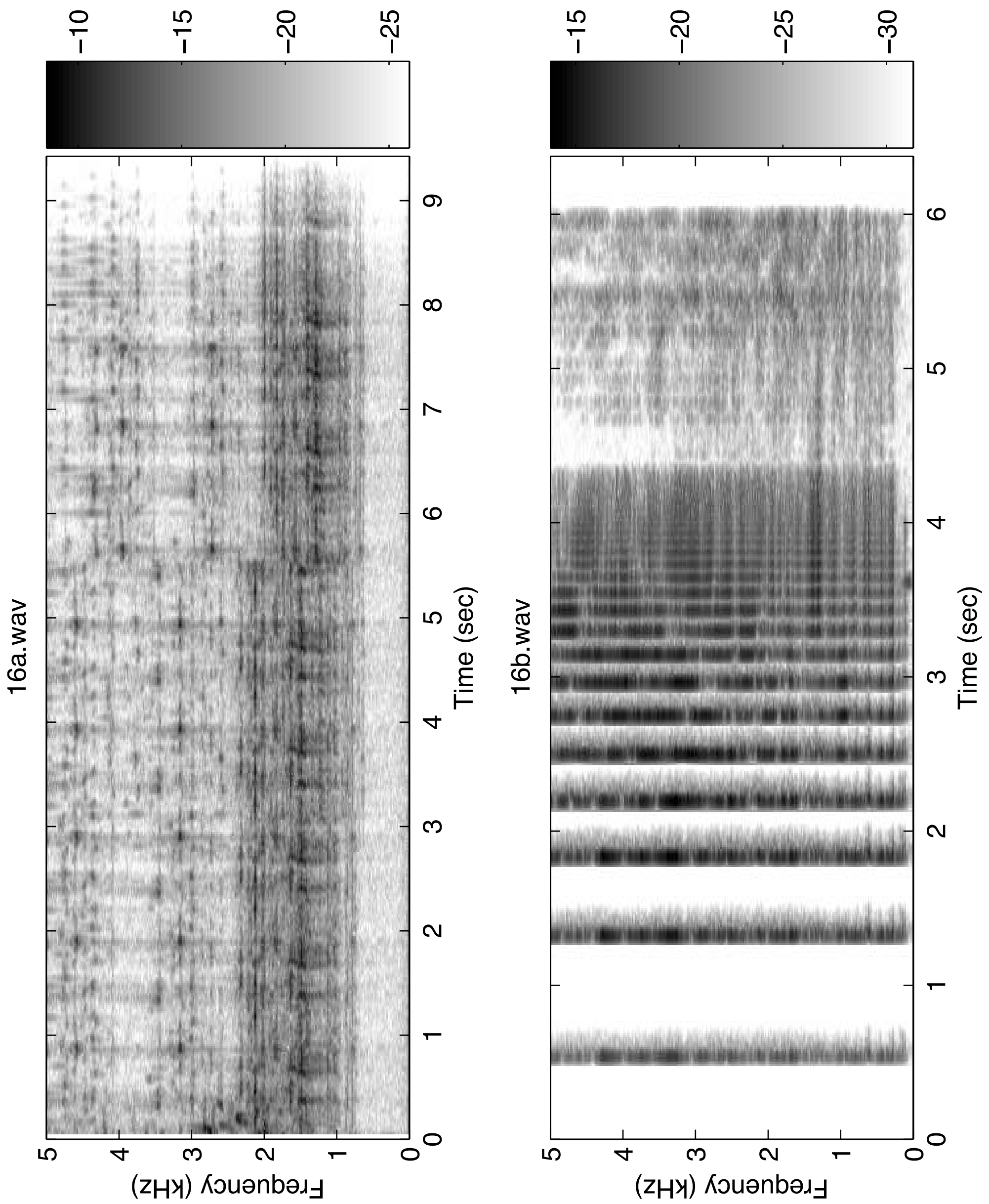
	A	B
More threatening	16.1%	83.9%
More annoying	0.0%	100.0%
More dissonant	20.0%	80.0%

Pair #15



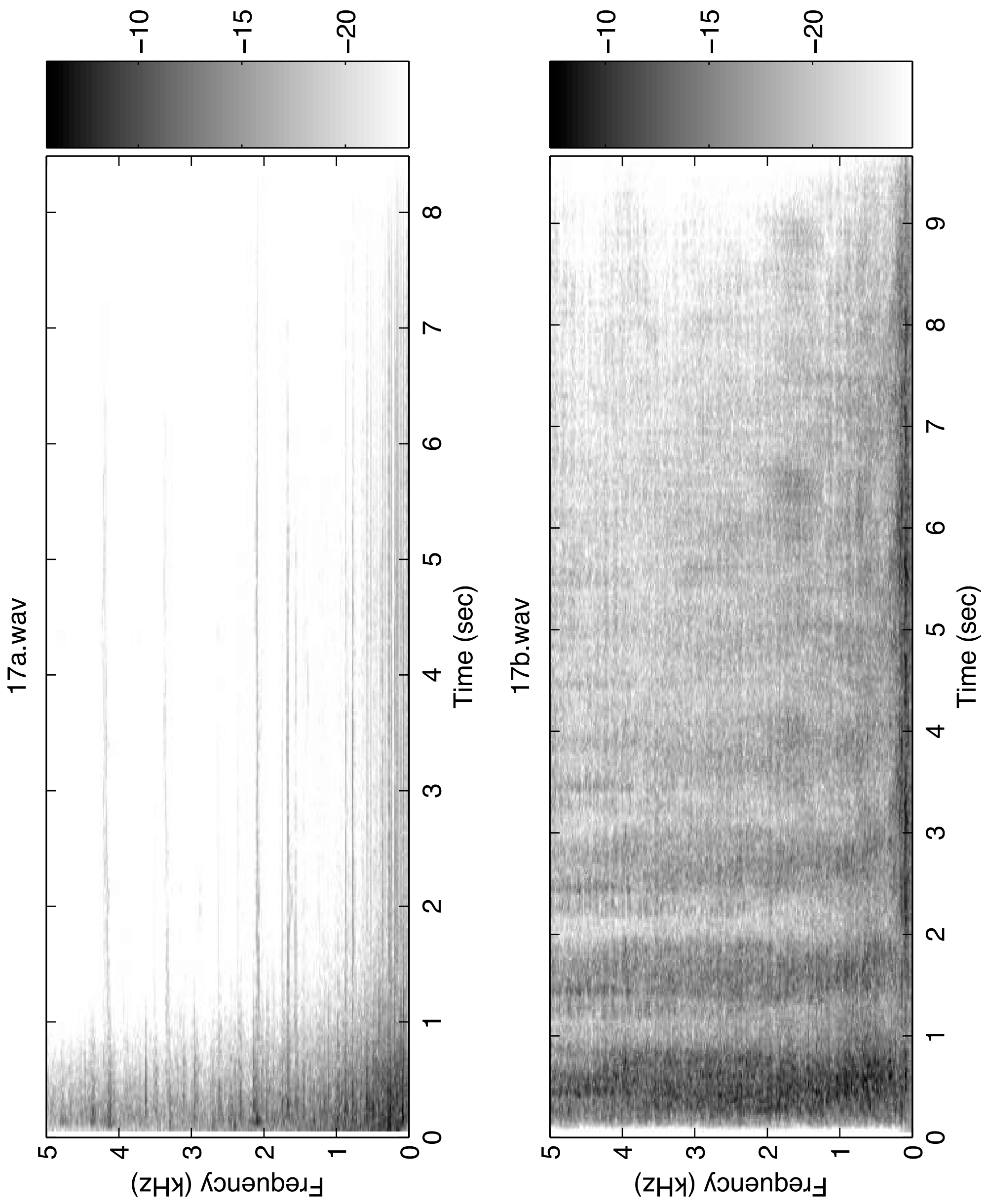
	A	B
More threatening	29.0%	71.0%
More annoying	29.0%	71.0%
More dissonant	38.7%	61.3%

Pair #16



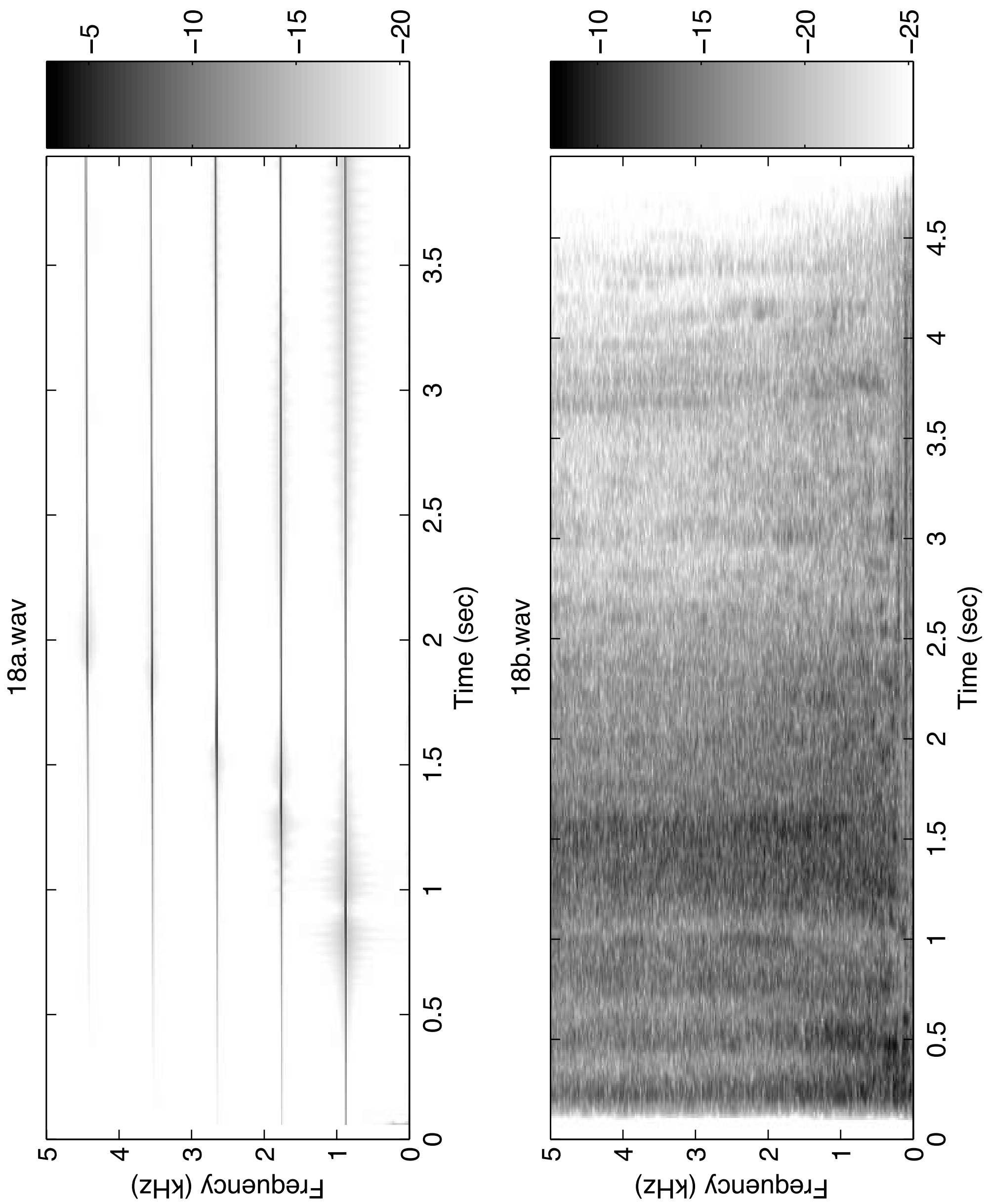
	A	B
More threatening	96.8%	3.2%
More annoying	87.1%	12.9%
More dissonant	96.8%	3.2%

Pair #17



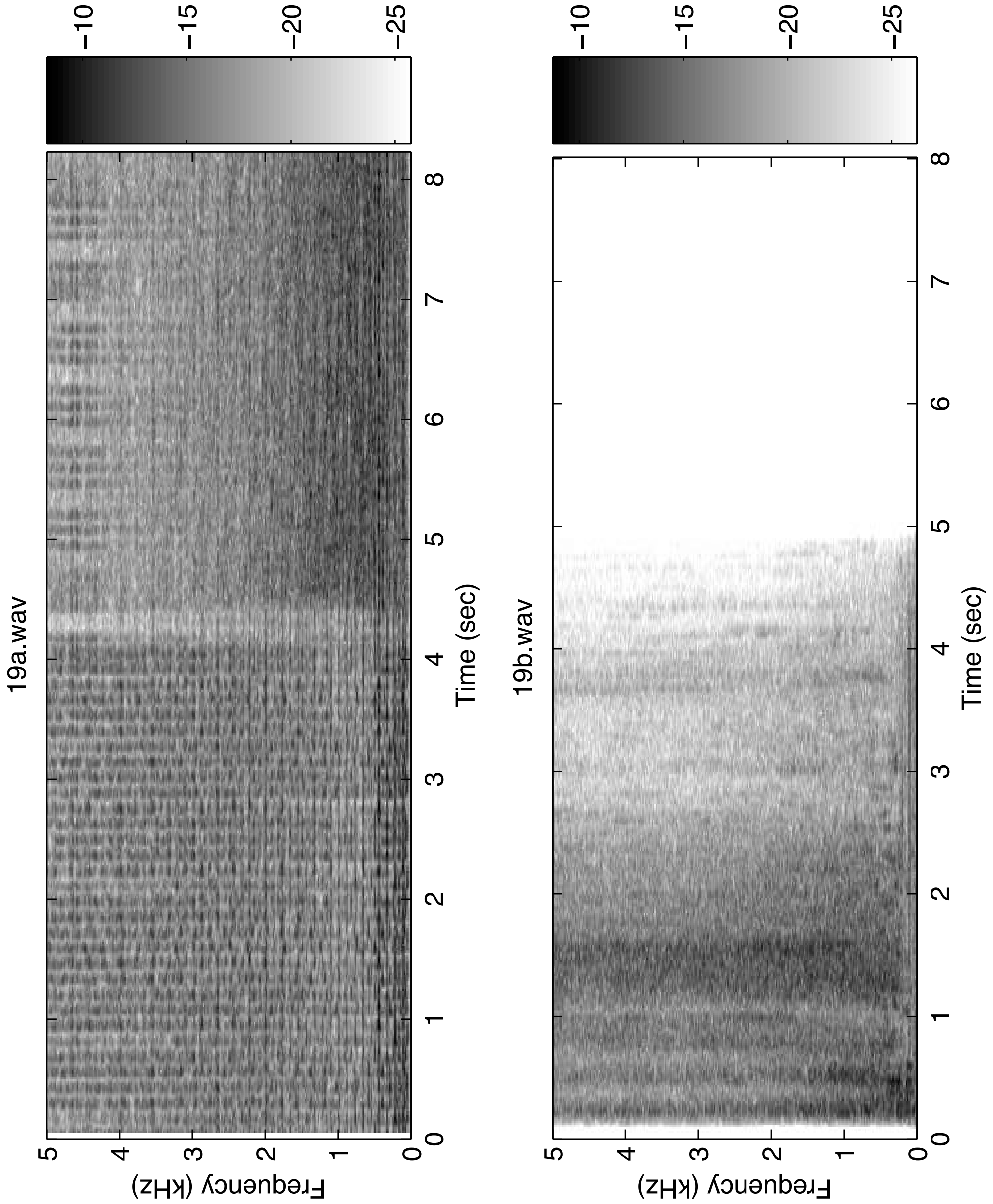
	A	B
More easilty recognizable	93.5%	6.5%
More predictable	87.1%	12.9%
More consistent	61.3%	38.7%
More dissonant	22.2%	77.8%

Pair #18



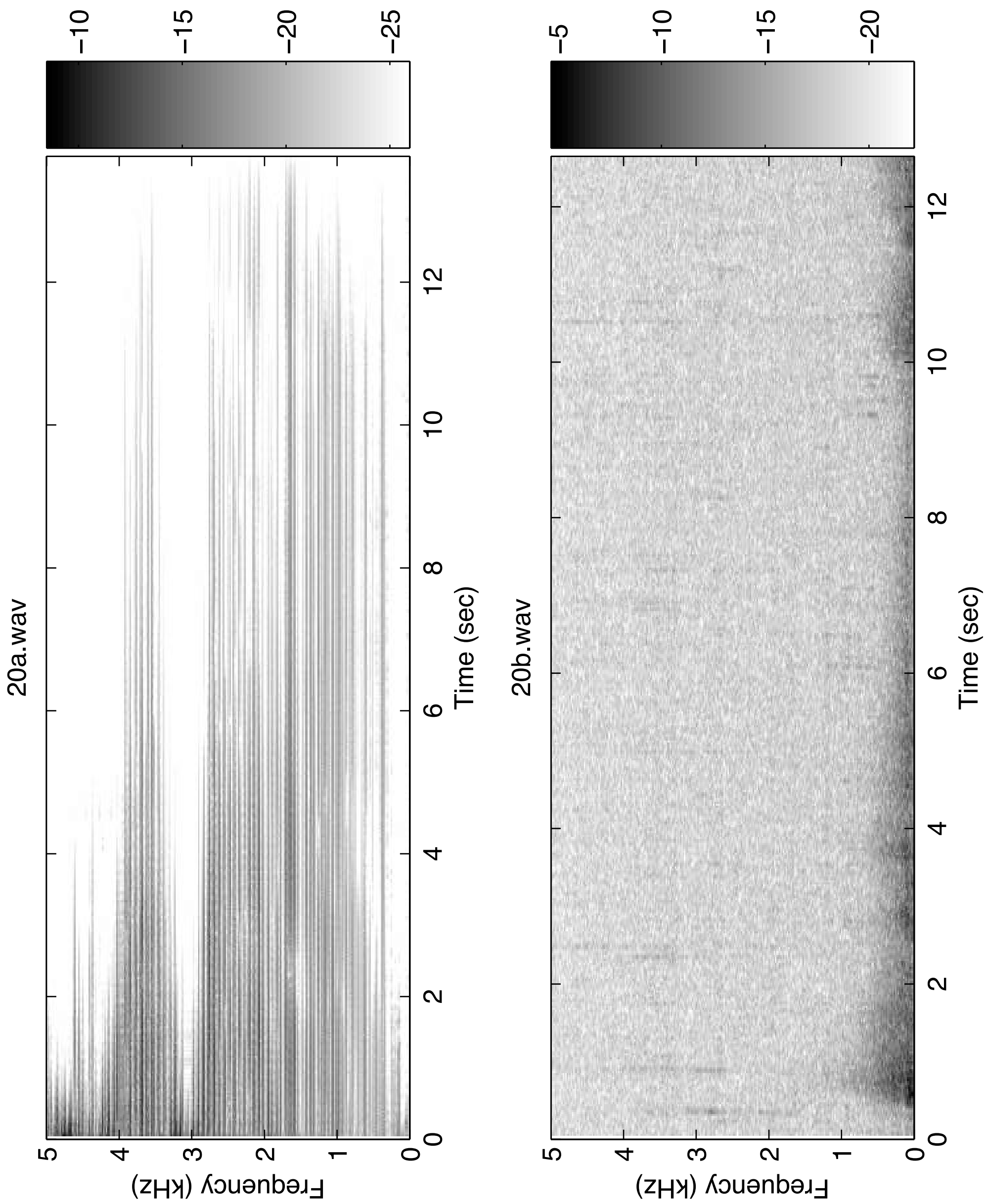
	A	B
More easilty recognizable	0.0%	100.0%
More predictable	19.4%	80.6%
More consistent	67.7%	32.3%
More dissonant	80.6%	19.4%

Pair #19



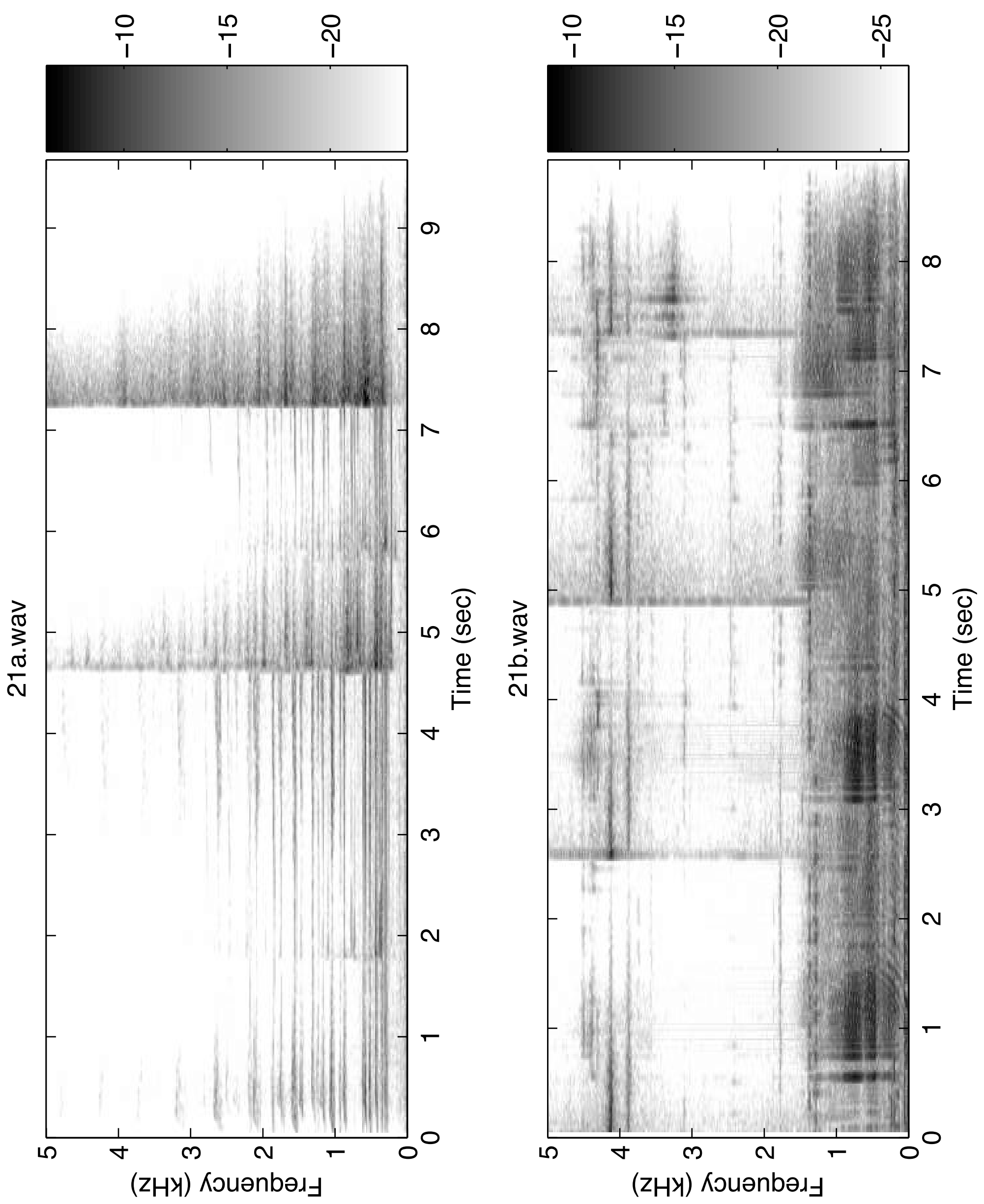
	A	B
More easilty recognizable	16.1%	83.9%
More predictable	32.3%	67.7%
More consistent	51.6%	48.4%
More dissonant	41.9%	58.1%

Pair #20



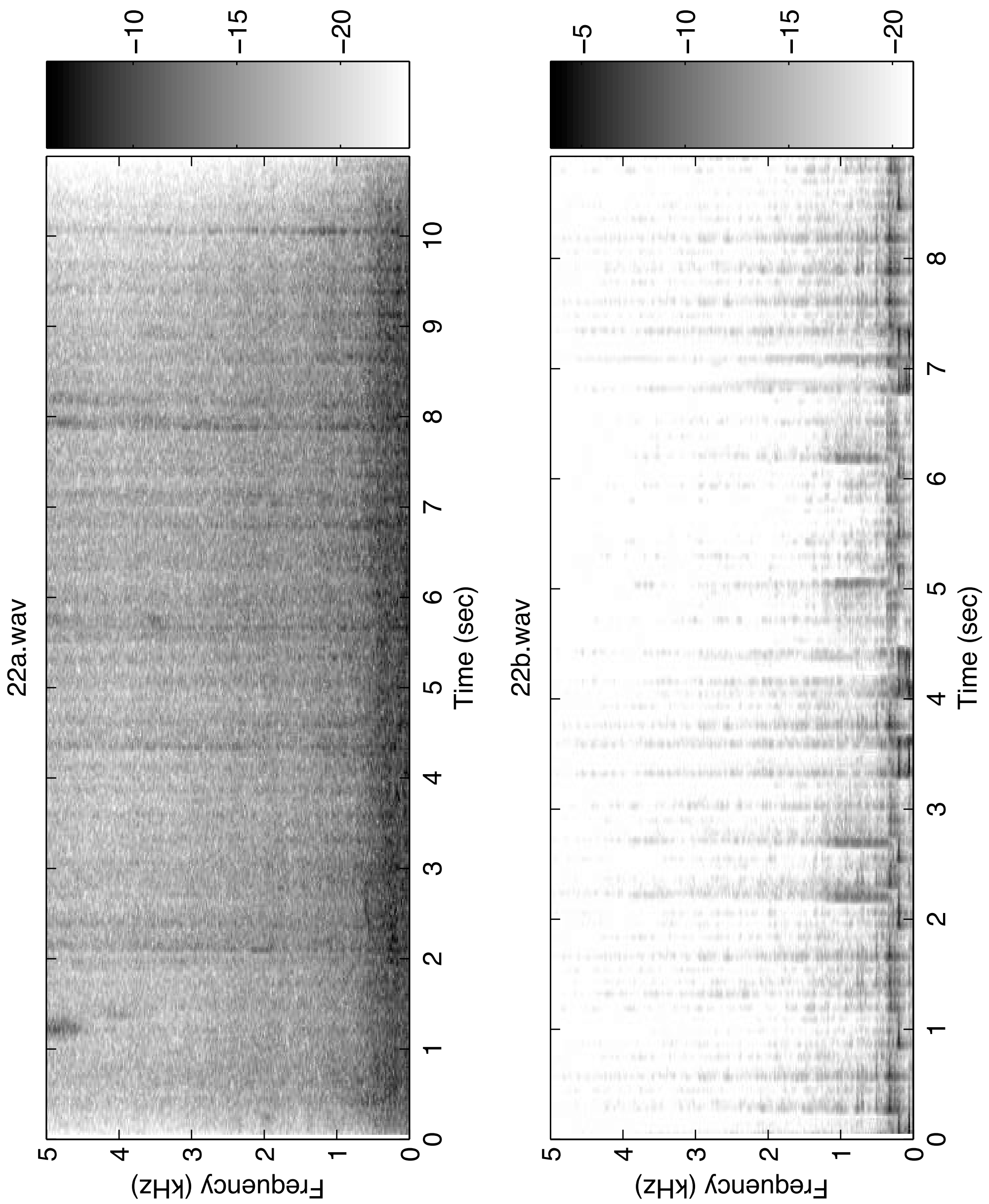
	A	B
More easilty recognizable	3.2%	96.8%
More predictable	22.6%	77.4%
More consistent	41.9%	58.1%
More dissonant	96.8%	3.2%

Pair #21



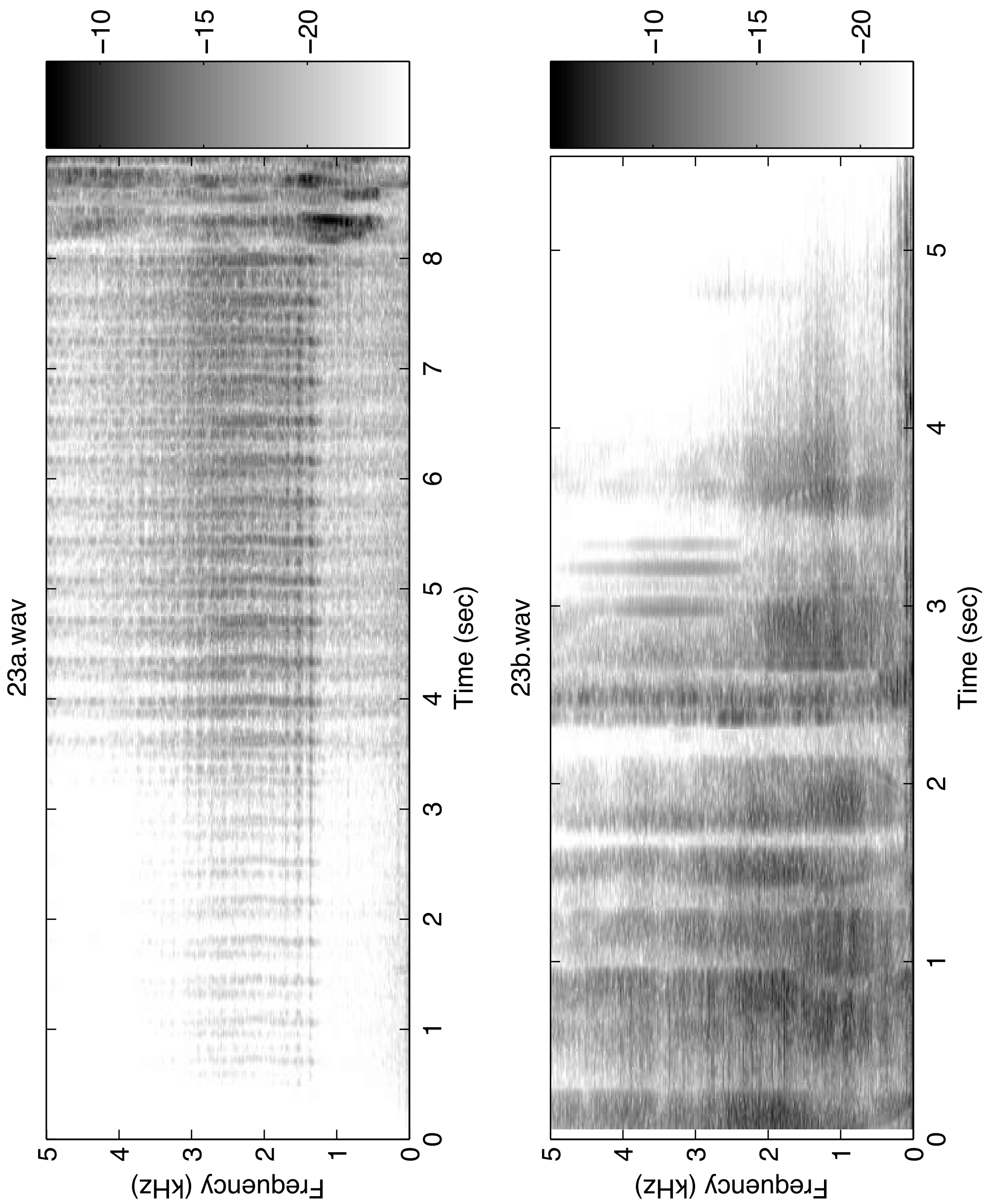
	A	B
More easilty recognizable	96.8%	3.2%
More predictable	64.5%	35.5%
More consistent	25.8%	74.2%
More dissonant	60.0%	40.0%

Pair #22



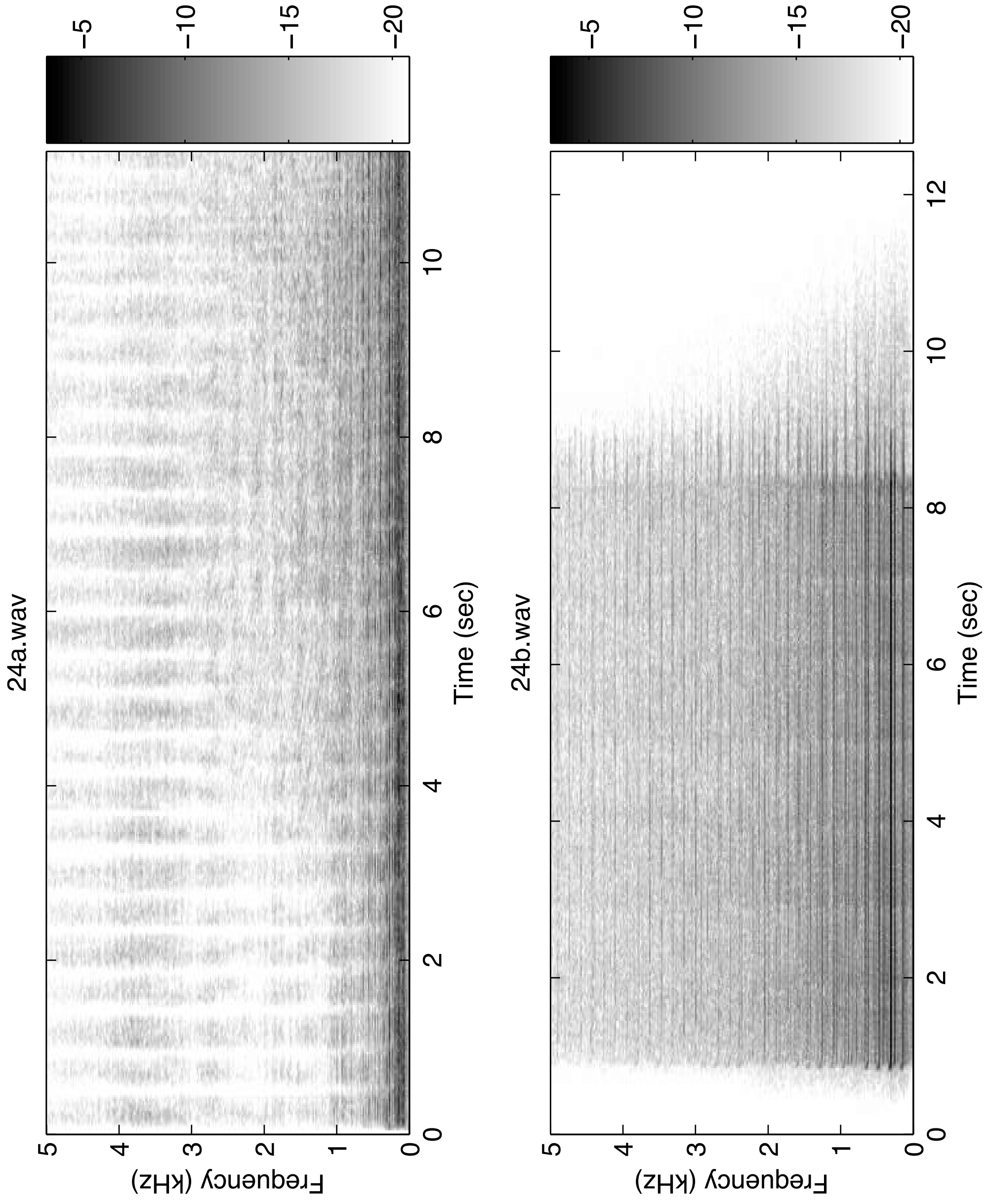
	A	B
More easilty recognizable	67.7%	32.3%
More predictable	96.8%	3.2%
More consistent	87.1%	12.9%
More dissonant	71.0%	29.0%

Pair #23



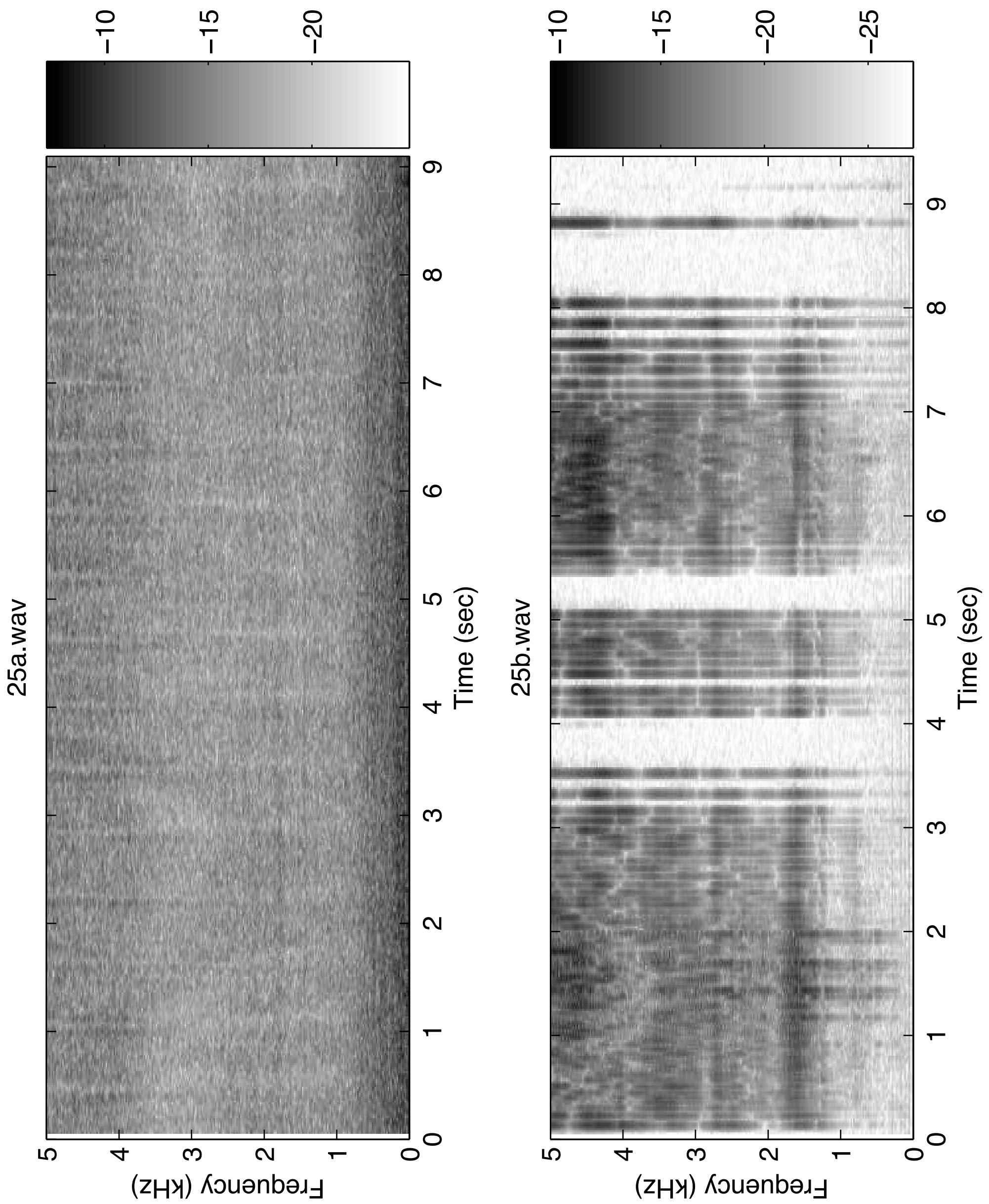
	A	B
More easilty recognizable	38.7%	61.3%
More predictable	61.3%	38.7%
More consistent	61.3%	38.7%
More dissonant	41.9%	58.1%

Pair #24



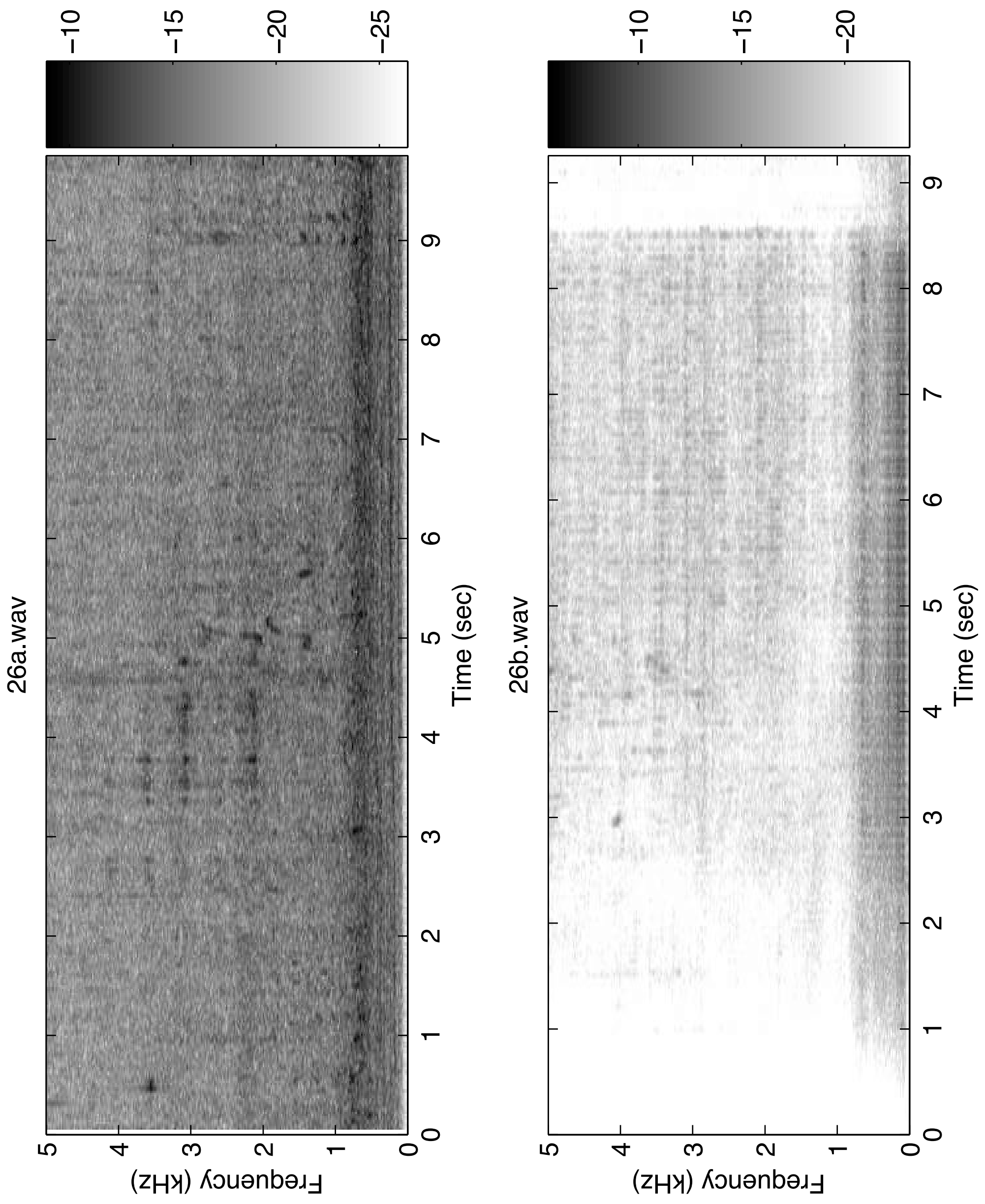
	A	B
More easilty recognizable	3.2%	96.8%
More predictable	9.7%	90.3%
More consistent	6.5%	93.5%
More dissonant	58.1%	41.9%

Pair #25



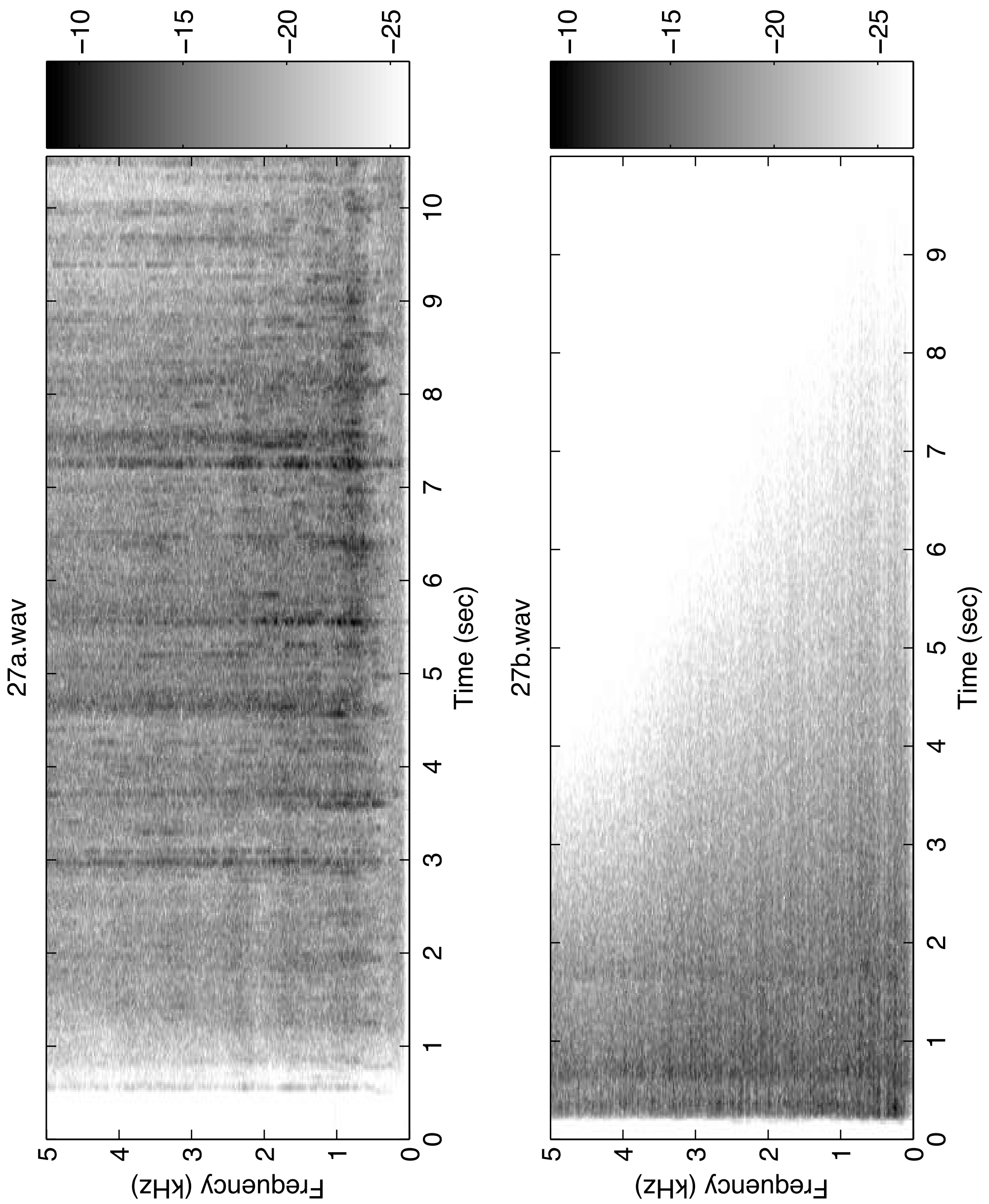
	A	B
More pitched	41.9%	58.1%
Smoother	77.4%	22.6%
More regular	58.1%	41.9%
More dissonant	83.9%	16.1%

Pair #26



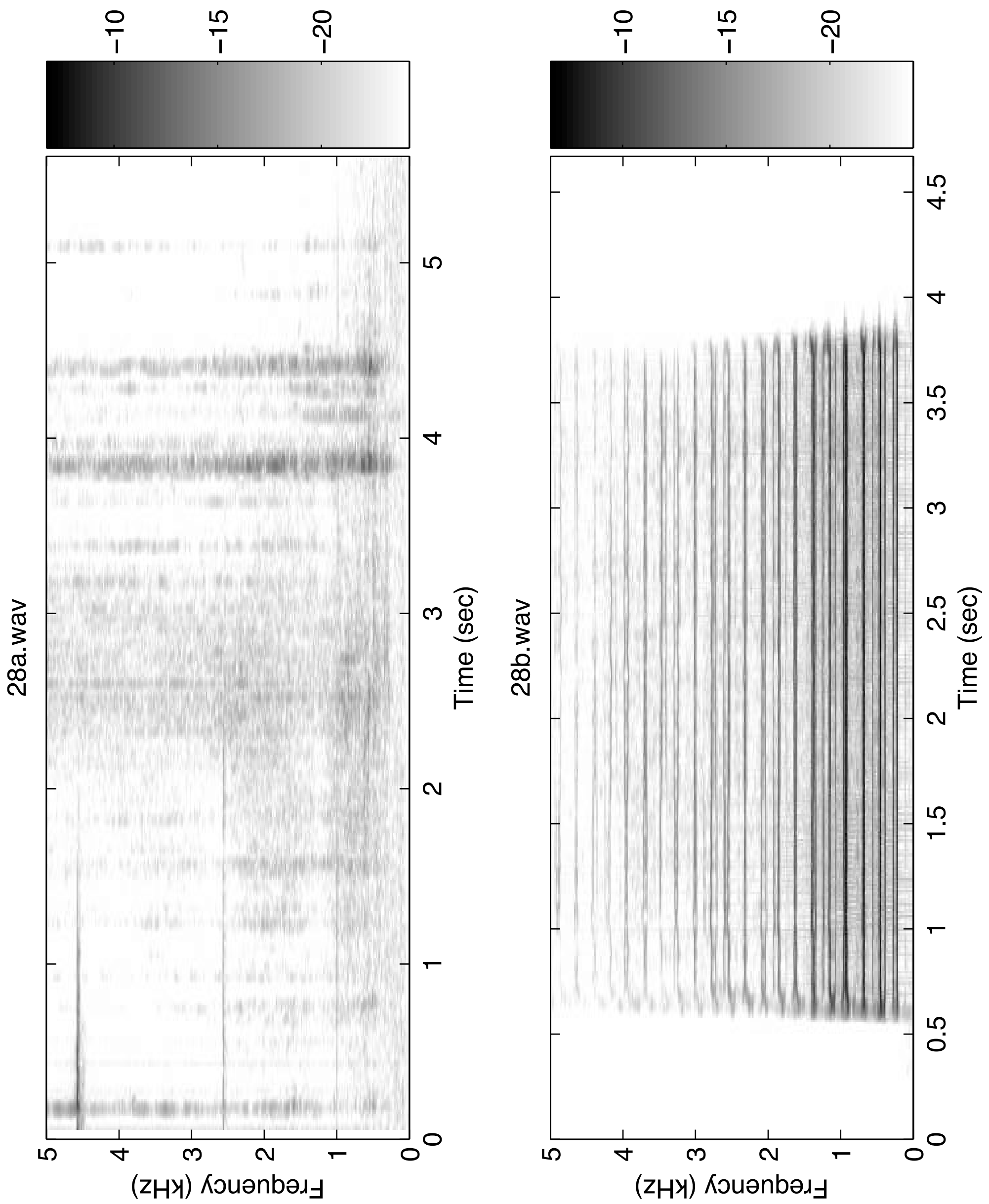
	A	B
More pitched	38.7%	61.3%
Smoother	38.7%	61.3%
More regular	30.0%	70.0%
More dissonant	80.6%	19.4%

Pair #27



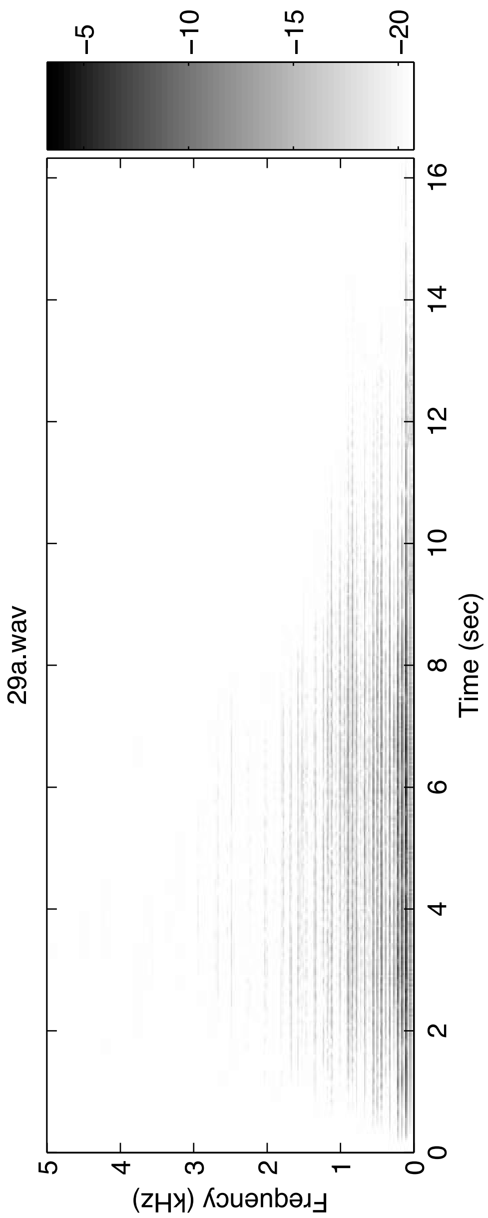
	A	B
More pitched	19.4%	80.6%
Smoother	67.7%	32.3%
More regular	64.5%	35.5%
More dissonant	35.5%	64.5%

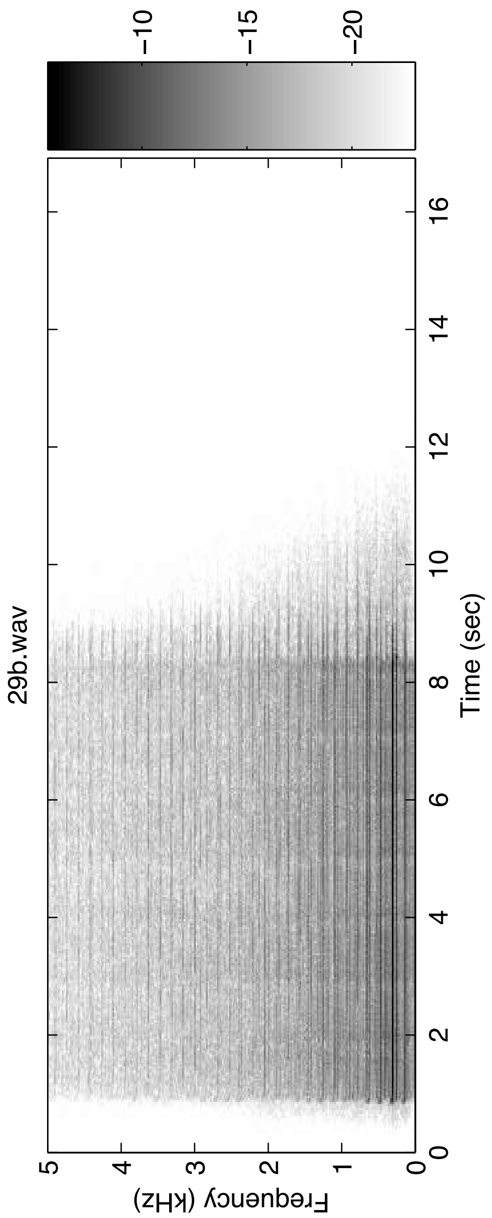
Pair #28



	A	B
More pitched	16.1%	83.9%
Smoother	71.0%	29.0%
More regular	45.2%	54.8%
More dissonant	3.2%	96.8%

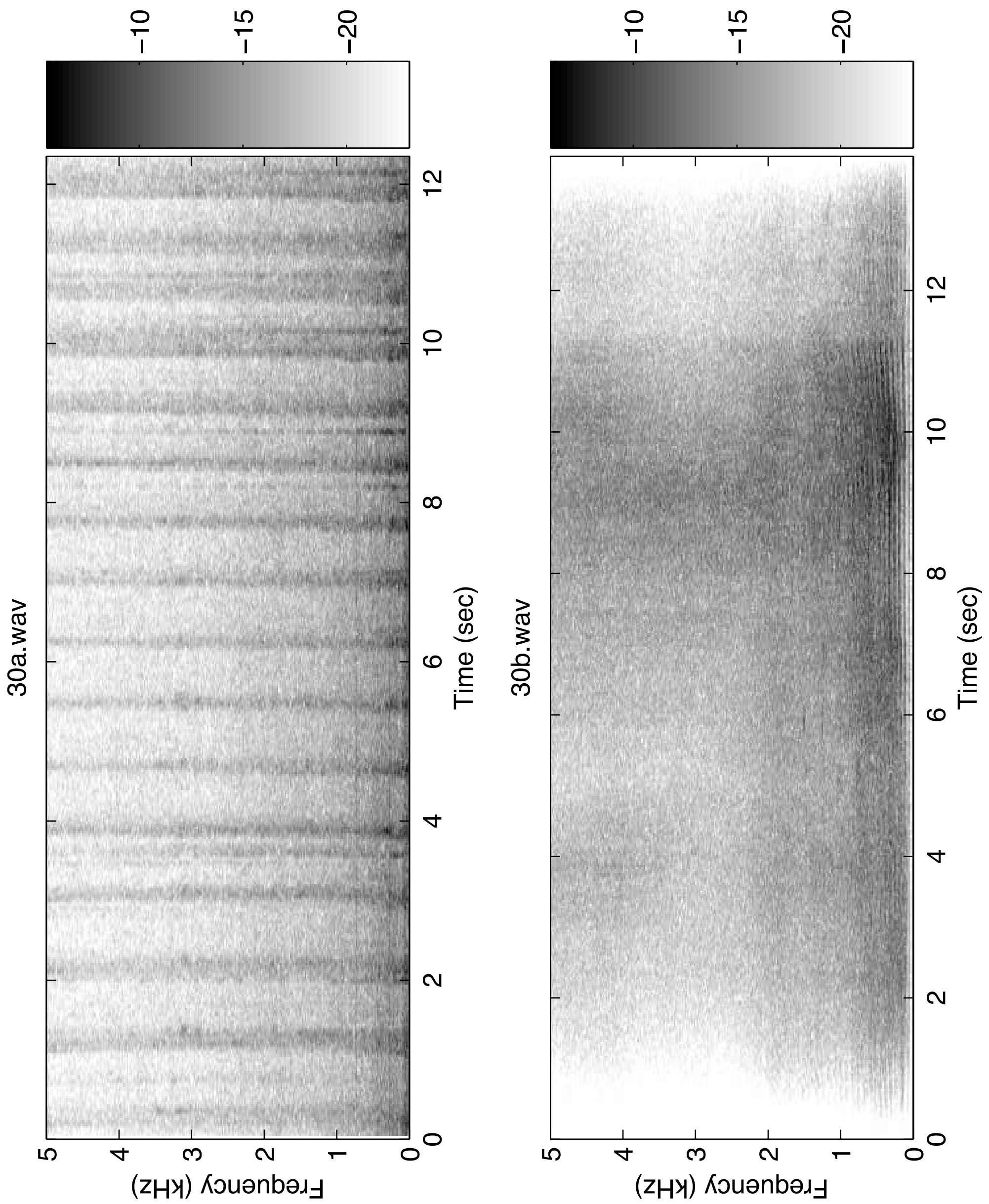
Pair #29





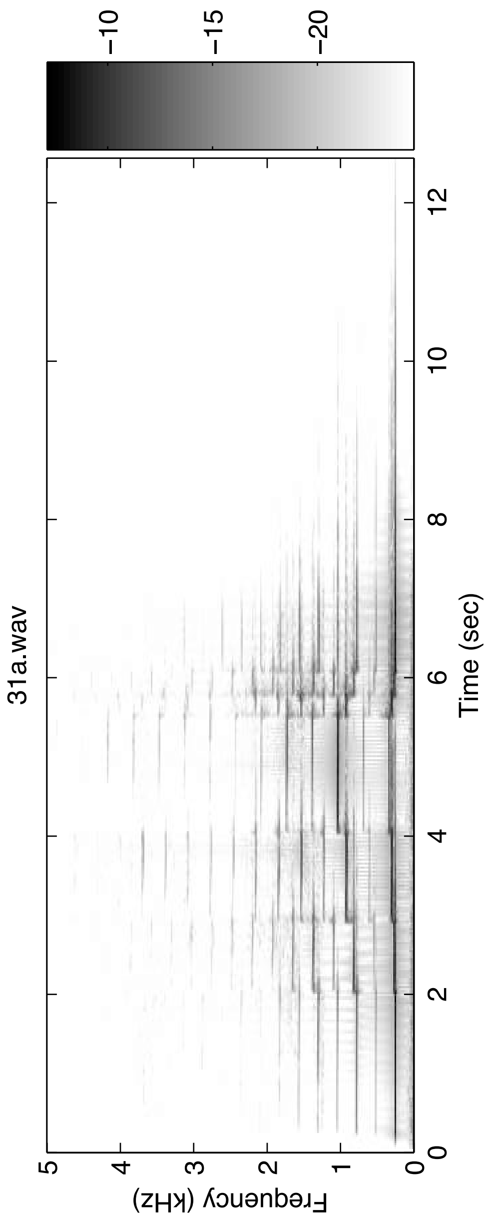
	A	B
More pitched	71.0%	29.0%
Smoother	93.5%	6.5%
More regular	58.1%	41.9%
More dissonant	9.7%	90.3%

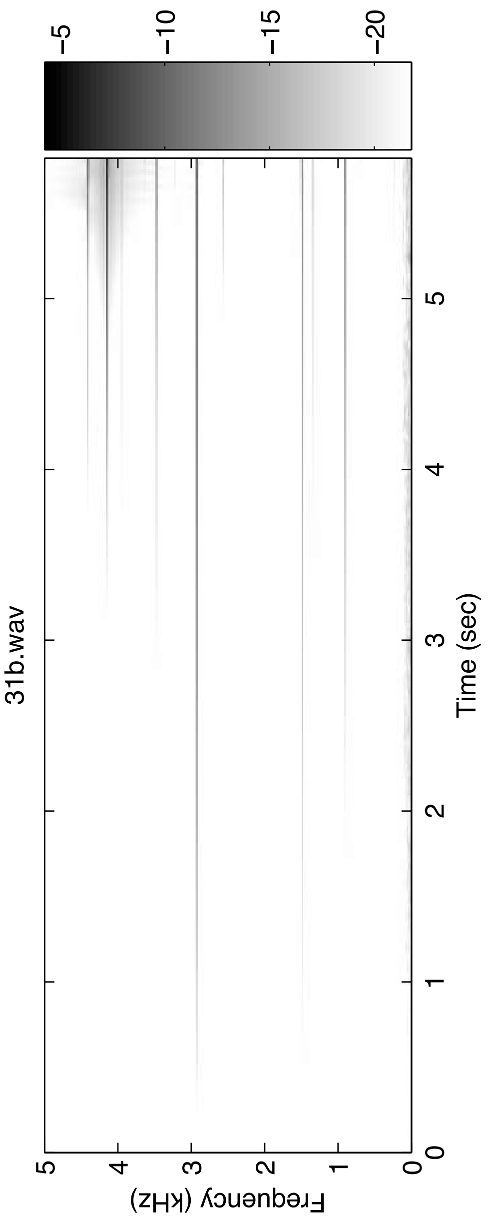
Pair #30



	A	B
More pitched	19.4%	80.6%
Smoother	45.2%	54.8%
More regular	77.4%	22.6%
More dissonant	25.8%	74.2%

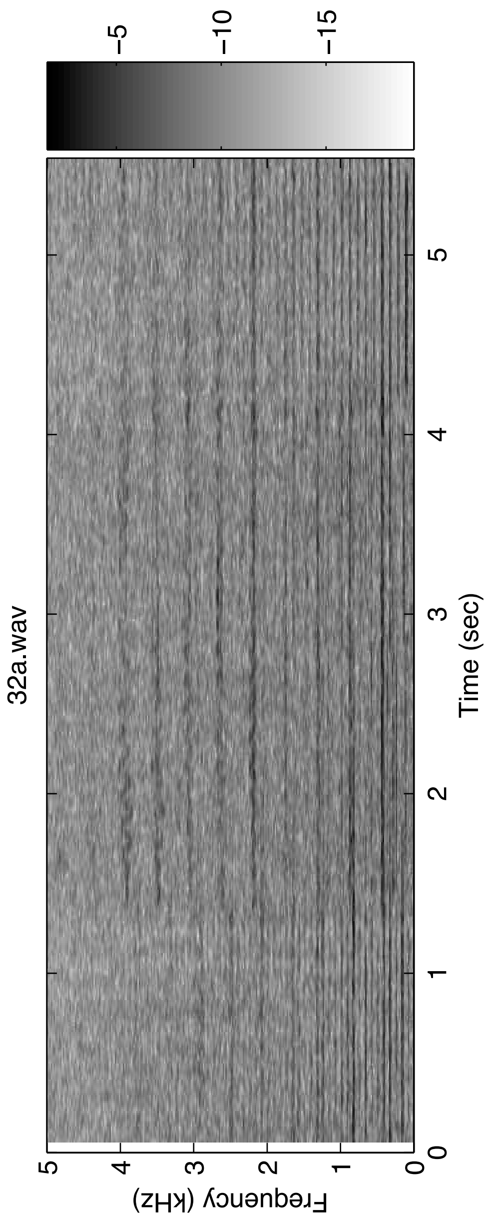
Pair #31

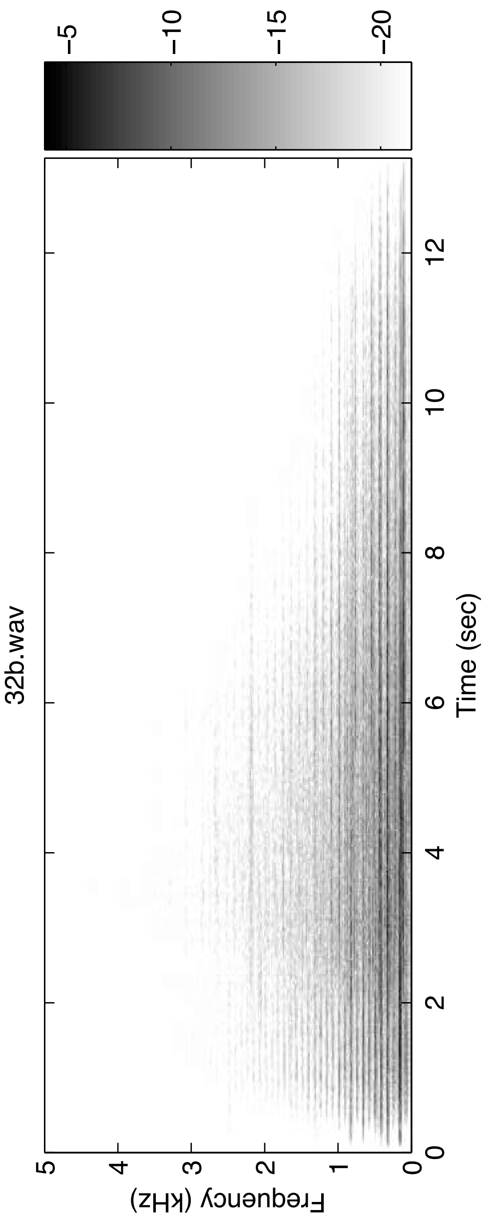




	A	B
More pitched	87.1%	12.9%
Smoother	96.8%	3.2%
More regular	74.2%	25.8%
More dissonant	0.0%	100.0%

Pair #32





	A	B
More pitched	22.6%	77.4%
Smoother	6.5%	93.5%
More regular	16.1%	83.9%
More dissonant	83.9%	16.1%

BIBLIOGRAPHY

Adolphs, R., et al. 1995. "Fear and the Human Amygdala." *J.Neurosci*.15:5879–5891.

Andrews, D. 1972. "Plots of High-Dimensional Data." *Biometrics* 28:125–136.

Ankerst, M., D. A. Keim, and H.-P. Kriegel. 1996. "Circle Segments: A Technique for Visually Exploring Large Dimensional Data Sets." In *Proceedings of the IEEE Visualization Conference*, San Francisco.

Arfib, D., F. Keiler, and U. Zölzer. 2002. "Time-Frequency Processing." In U. Zölzer, ed. *DAFX—Digital Audio Effects*. West Sussex, UK: Wiley, pp. 237–298.

Atienza, M., J. L. Cantero, C. Grau, C. Gomez, E. Dominguez-Marin, and C. Escera. 2003. "Effects of Temporal Encoding on Auditory Object Formation: A Mismatch Negativity Study." *Cognitive Brain Research* 16(3):359–371.

Aucouturier, J., and F. Pachet. 2004a. "Improving Timbral Similarity: How High's the Sky?" *Journal of Negative Results in Speech and Audio Sciences* 1(1).

Aucouturier, J., and F. Pachet. 2004b. "Tools and Architecture for the Evaluation of Similarity Measures: Case Study of Timbral Similarity." *Proceedings of the 2004*

International Conference on Music Information Retrieval. Barcelona, Spain: Audiovisual Institute, Universitat Pompeu Fabra.

Bailhache P. 1994. "Mersenne, Marin 'Harmonie Universelle': Situating Music Between Mathematics, Physics, Metaphysics and Religion." *Etudes Philosophiques* (1–2):13–24

Bailhache, P. 1986. "Valeur actuelle de l'acoustique musicale de Helmholtz." *Revue d'histoire des sciences* 39(4):301–324.

Bailhache, P. 1997. "Music Translated into Mathematics: Leonhard Euler." Paper presented at the conference Problems of Translation in the Eighteenth Century, Nantes, France, January 17. Trans. by Joe Monzo, available online at <http://sonic-arts.org/monzo/euler/euler-en.htm>.

Bailhache, P. N.D. "D'Alembert théoricien de la musique: empirisme et nature."Available online at <http://bailhache.humana.univ-nantes.fr/thmusique/d'alembert.html>.

Bailhahce, P. 1995. "Deux mathématiciens musicians: Euler et d'Alembert." *Physis, Rivista internazionale di storia della scienza* 32:1–35 .

Barbieri P. 1987. "Caramuel, Juan (1606–1682) and His Essay on Musical Logarithms and Musical Tuning." *Muskitheorie* 2 (2):145–168 .

Barbieri P. 1992. "Tartini 3rd tone and Euler harmonic Exponents, With a Previously Unpublished Manuscript by Tartini." *Musiktheorie* 7(3):219–234.

Barker, A. D. 1989 *Greek Musical Writings, Vol. II: Harmonic and Acoustic Theory*. Cambridge: Cambridge University Press.

Barlow, H. B. 1959. “Sensory Mechanisms, the Reduction of Redundancy, and Intelligence.” In *National Physical Laboratory Symposium No. 10*, The Mechanization of Thought Processes. London: Her Majesty’s Stationery Office, pp. 535–539.

Barton, T. D. 1978. *Robert Fludd’s Temple of Music: A Description and Commentary*. M.A. Thesis, University of Oregon School of Music.

Beauchamp, J. W. 1982. “Synthesis by Spectral Amplitude and ‘Brightness’ Matching Analyzed Musical Sounds.” *Journal of Audio Engineering Society* 30(6):396–406.

Berenzweig, A., B. Logan, D. P. W. Ellis, and B. Whitman. 2004. A Large-Scale Evaluation of Acoustic and Subjective Music-Similarity Measures. *Computer Music Journal* 28(2):63–76.

Bhogal, G. K. 2000. “Disappearing into the Ether: Metric Stability in Ravel’s ‘Noctuelles’ (from *Miroirs*, 1905).” Paper presented at Music Theory Midwest 2000 Eleventh Annual Conference, Lawrence University, 19 May.

Bigand, E., R. Parncutt, and F. Lerdahl. 1996. “Perception of Musical Tension in Short Chord Sequences: The Influence of Harmonic Function, Sensory Dissonance, Horizontal Motion, and Musical Training.” *Perceptual Psychophysics* 58(1):124–141.

Bilotta, E., and P. Pantano. 2000. “In Search of Musical Fitness on Consonance.” Paper presented at VII Simpósio Brasileiro de Computação e Música, Curitiba, PR, 17–20 July.

Blood, A., R. J. Zatorre, P. Bermudez, and A. C. Evans. 1999. “Emotional Responses to Pleasant and Unpleasant Music Correlate with Activity in Paralimbic Brain Regions.” *Nature Neuroscience* 2(4):382–387.

Boethius, Anicius Manlius Severinus. c. 524. *De Institutione Musica*. Trans. by Calvin M. Bower, ed. by C. V. Palsica, 1989. New Haven: Yale University Press.

Bolger, D., and N. Griffith. 2003. "Dissonance and Tonality: The Role of Timbre in Melodic Relations." Proceedings of the 5th Triennial ESCOM Conference, pp. 523–526.

Borchgrevink, H. M. 1975. "Musikalske akkordpreferanser hos mennesket belyst ved dyreforsok." [Musical chord preferences in humans as demonstrated through animal experiments]. *Tidskrift for den Norske Laegeforening* 95:356–358.

Borden, L. 2003. "Preferences for Major Triads Tuned in Just Intonation and Equal Temperament." M.S. Thesis, Vanderbilt University.

Bregman, A. 1990. *Auditory Scene Analysis*. Cambridge, Massachusetts: MIT Press.

Brues, A.M. 1927. "The Fusion of Non-Musical Intervals." *American Journal of Psychology* 38: 624–638.

Butler, J. W., and P. G. Daston. 1968. "Musical Consonance as Musical Preference: A Cross-Cultural Study." *Journal of General Psychology* 79:129–142.

Butler, M. J. 2001. "Turning the Beat Around: Reinterpretation, Metrical Dissonance, and Asymmetry in Electronic Dance Music." *Music Theory Online* 7(6).

Buttram, J. B. 1969. "Perception of Musical Intervals." *Perceptual and Motor Skills* 28:391–394.

Calinger R. 1966. "Leonhard Euler: The First St. Petersburg years (1727–1741)." *Historia Mathematica* 23(2):121–166 .

Cano, P., et al. 2004. "Sound Effects Taxonomy Management in Production Environments." Paper presented at the Audio Engineering Society 25th International Conference, London, June 17–19. Paper available online at <http://audioclas.iua.upf.edu/pdf/AES25-pcano.pdf>.

Cariani P. A., and B. Delgutte. 1996. "Neural Correlates of the Pitch of Complex Tones I: Pitch and Pitch Salience." *Journal of Neurophysiology* 76(3):1698–1716.

Cariani, P. A., B. Delgutte, and N. Y. S. Kiang. 1992. "The Pitch of Complex Sounds Is Simply Coded in Interspike Interval Distributions of Auditory Nerve Fibers." *Society for Neuroscience Abstracts* 18:383.

Casey, M. A. 2001. "General Sound Classification and Similarity in MPEG-7." *Organized Sound* 6:153–164

Cazden N. 1980. "The Definition of Consonance and Dissonance." *International Review of the Aesthetics and Sociology of Music* 11(2):123–168.

Chalmers, J. 1993. *Divisions of the Tetrachord.* Lebanon, New Hampshire: Frog Peak Music.

Cheng, C. I. 1996. "Wavelet Signal Processing of Digital Audio with Applications in Electro-Acoustic Music." A.M. Thesis, Electro-Acoustic Music, Dartmouth College.

Chernoff, H. 1973. "The Use of Faces to Represent Points in k-Dimensional Space Graphically." *Journal of the American Statistical Association* 68(342):361–368.

Chion, M. 1983. *Guide des Objets Sonores*. Paris: Editions Buchet/Chastel.

Clement of Alexandria. [late second century]. *Exhortation to the Greeks*. Trans. G. W. Butterworth, 1919. Cambridge, Massachusetts: Loeb Classical Library.

Cogan, R. 1984. *New Images of Musical Sound*. Cambridge, Massachusetts: Harvard University Press.

Coleman, R. 1994. "—". *Notes* 50(3):1186–1190.

Cook, P. R. 2005. Personal communication. 12 May 2005.

Cuadra, P., A. Master, and C. Sapp. 2001. "Efficient Pitch Detection Techniques for Interactive Music." *Proceedings of the 2001 International Computer Music Conference*. San Francisco: International Computer Music Association.

Dack, J. 1999. "Ear Training Using the Computer and PROGREMU." Paper presented at the Arts Symposium of the International Conference for Advanced Studies in Systems Research and Cybernetics, Baden-Baden, Germany, August.

Daniélou, A.1958. *Tableau Comparatif des Intervalles Musicaux*. Pondichéry, France: Intitut Français D'Indologie.

Danner, G. 1985. "The Use of Acoustic Measures of Dissonance to Characterize Western Pitch-Class Sets." *Music Perception* 3(1):103–122.

Davis, S.B., and P. Mermelstein. 1980. "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences." *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28(4):357–366.

de la Cuadra, P., A. Master, and C. Sapp. "Efficient Pitch Detection Techniques for Interactive Music." *Proceedings of the 2001 International Computer Music Conference* San Francisco: International Computer Music Association, pp. 403–406.

DeMeo, W. J. 2002. “Characterizing Musical Signals with Wigner-Ville Interferences.” *Proceedings of the 2002 International Computer Music Conference*. San Francisco: International Computer Music Association.

Desain, P. 1992. “A (De)Composable Theory of Rhythm Perception.” *Music Perception* 9(4):439–454.

Deutsch, D., ed. 1982. *The Psychology of Music*. New York: Academic Press.

DeWitt, L.A., and R. G. Crowder. 1987. “Tonal Fusion of Consonant Musical Intervals: The Oomph in Stumpf.” *Perception & Psychophysics* 41(1):73–84.

Di Scipio, A. 1999. “Synthesis of Environmental Sound Textures by Iterated Nonlinear Functions.” *Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects (DAFx99).* December 9–11, 1999.

Doherty, K. A., and R. A. Lutfi. 1992. “Efficiency of Analytic versus Synthetic Listening in Spectral Pattern Discrimination.” Paper presented at the 124th meeting of the Acoustical Society of America, New Orleans, October.

Downie, J. S. “Music Information Retrieval.” In B. Cronin, ed. *Annual Review of Information Science and Technology* 37:295–340. Medford, New Jersey: Information Today. Available online at <http://music-ir.org/downie_mir_arist37.pdf>.

Eco, U. 1986. *Art and Beauty in the Middle Ages*. H. Bredin, trans. New Haven: Yale University Press.

Edmonds, E. M., and M. E. Smith. 1923. "The Phenomenological Description of Musical Intervals." *American Journal of Psychology* 34:287–291.

Ellis, D. P., and J. Arroyo. 2004. "Eigenrhythms: Drum Pattern Basis Sets for Classification and Generation." Paper presented at the 2004 International Symposium on Music Information Retrieval, Universitat Pompeu Fabra, Barcelona, Spain.

Estrada, J. 1994. "Transkriptionsmethoden in einem Kontinuum von Rhythmus und Klang. *MusikTexte 55*.

Ethington, R., and B. Punch. "SeaWave: A System for Musical Timbre Description." *Computer Music Journal* 18(1):30–39.

Euler, L. 1739. *Tentamen novae theoriae musicae ex certissismis harmoniae principiis dilucide expositae.* Saint Petersburg Academy. Republished in *Musique mathématique*, Paris 1865, pp. I–VII, 1-215 [E33Aa]

Evans, B. 2005. "Foundations of a Visual Music." *Computer Music Journal* 29(4).

Faist, A. 1900. *Über Consonanz und Dissonanz*. Dissertation, Graz.

Ferrers, N. M. 1877. *An Elementary Treatise on Spherical Harmonics and Subjects Connected with Them*. London: Macmillan.

Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Stanford, California: Stanford University Press.

Fishman, Y. I., and M. Steinschneider. 2001. "Consonance and Dissonance of Musical Chords: Neural Correlates in Auditory Cortex of Monkeys and Humans." *The Journal of Neurophysiology* 86(6):2761–2788.

Fludd, R. 1617. *Utriusque Cosmi Maioris Scilicet et Minoris Metaphysica Physica Atque Technica Historia in Duo Volumnia Secundum Cosmi Differentiam Divisa—Tomus Primus De Macrocosmi Historia.* Oppenheim: Johan Theodori de Bry.

Fludd, R. 1618. *Utriusque Cosmi Maioris Scilicet et Minoris Metaphysica Physica Atque Technica Historia in Duo Volumnia Secundum Cosmi Differentiam Divisa—Tractatus Secundus De Naturaæ Simia Seu Technica Macrocosmi Historia in Partes Undecim Divisa.* Oppenheim: Johan Theodori de Bry.

Fokker, A. D. 1951. "Les gammes et le temperament egal." *Acustica* 1(1):29–35.

Fokker, A. D. 1975. *New Music with 31 Notes*, Trans. Leigh Jardine. Bonn: Verlag für Systematische Musikwissenschaft GmbH.

Foote, J. 1997. "Content-Based Retrieval of Music and Audio," in C.-C. J. Kuo et al., eds. *Multimedia Storage and Archiving Systems II—Proceedings of SPIE* 3229:138–147

Foote, J. 1999. "An Overview of Audio Information Retrieval." *Multimedia Systems* 7(1):2–11.

Forster, C. M. L. 2005. *Musical Mathematics: A Practice in the Mathematics of Tuning Instruments and Analyzing Scales.* Unpublished manuscript, available online at <http://www.chrysalis-foundation.org>.

Fyk, J. 1982. "Tolerance of Intonation Deviation in Melodic Intervals in Listeners of Different Musical Training." *Archives of Acoustics* 7(1):13–28.

Gaver, W. 1993. "What in the World Do We Hear? An Ecological Approach to Auditory Event Perception." *Ecological Psychology* 5(1):1–29.

Gell-Mann, M. 1995. “What Is Complexity?” *Complexity* 1(1).

Gerhard, D. 1999. “Audio Visualization in Phase Space.” *Bridges: Mathematical Connections in Art, Music, and Science*. Winfield, Kansas, pp. 37–48.

Giron, F. 1996. *Untersuchungen zur Richtcharakteristik von Schallquellen.* [Investigations About the Directivity of Sound Sources.] Ph.D. Dissertation, Rurh-Universität Bochum, Fakultät für Elektrotechnik und Informationstechnik.

Godwin, J. 1986. *Music, Mysticism and Magic: A Source Book*. New York: Routledge and Kegan Paul.

Godwin, J. 1987. *Harmonies of Heaven and Earth: From Antiquity to the Avant-Garde.* Rochester, Vermont: Inner Traditions.

Godwin, Joscelyn. *Robert Fludd: Hermetic Philosopher and Surveyor of Two Worlds.* Phanes Press.

Gómez, E., et al. “Using and Enhancing the Current MPEG-7 Standard for a Music Content Processing Tool.” Paper presented at the 114th Convention of the Audio Engineering Society, Amsterdam, March 22–25.

Greenwood, D. D. 1991. “Critical Bandwidth and Consonance in Relation to Cochlear Frequency-Position Coordinates.” *Hearing Research* 54(2):164–208.

Guernsey, M. 1928. “The Role of Consonance and Dissonance in Music.” *American Journal of Psychology* 40:173–204.

Guizatdinova, I., and Z. Guo. 2003. "Sonification of Facial Expressions." Paper presented at *New Interaction Techniques '03*. Tampere, Finland: University of Tempere.

Gurney, E. 1880. *The Power of Sound*. London: Smith, Elder, and Company.

Haar, J. 1998. *The Science and Art of Renaissance Music*. Princeton, NJ: Princeton University Press.

Hall, D. E., and J. T. Hess. 1984. "Perception of Musical Interval Tuning." *Music Perception* 2:166–195.

Hall, M. P. 1928. *The Secret Teachings of All Ages*. (Self-published.) Available online at <http://www.sacred-texts.com/eso/sta>.

Haluska, J. 2004. *The Mathematical Theory of Tone Systems.* New York: Marcel Dekker.

Heffernan, J. 1887. "On Musical Beats and Their Relationship to Consonance and Dissonance." *Proceedings of the Musical Association* 14th session (1887–1888), pp. 1–21.

Helmholtz, H. 1858. "The Physical Cause of Harmony and Disharmony." *Naturforschervers.* In *Karlsruhe*, Septmber.

Helmholtz, H. 1877. *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik.* 1877, 6th ed., Braunschweig: Vieweg, 1913; trans. by A.J. Ellis as *On theSensations of Tone as a Physiological Basis for the Theory of Music*. (1885). Reprinted New York: Dover, 1954.

Herre, J. 2003. "MPEG-7 Audio: Tools for Semantic Audio Description and Processing." Keynote address given at the 6th International Conference on Digital Audio Effects, DAFx-03. 8-11 September,
Queen Mary, University of London. Available online at <http://www.elec.qmul.ac.uk/dafx03/pdfs/MPEG-7_Tools_3.pdf>.

Herrera, P., X. Serra, and G. Peeters, 1999. "Audio Descriptors and Descriptor Schemes in the Context of MPEG-7." *Proceedings of the 1999 International Computer Music Conference*. San Francisco: International Computer Music Association.

Hodgkinson, T. 1987. "Pierre Schaeffer interview." *Recommended Records Quarterly* 2:1. Available online at <http://www.cs.nwu.edu/%Etisue/schaeffer.html>.

Hugdahl, K., et al. 1995. "Brain Mechanisms in Human Classical Conditioning: A PET Blood Flow Study." *Neuroreport 6*:1723–1728.

Hulse, H. S., D. J. Bernard, and R. F. Braaten. 1995. "Auditory Discrimination of Chord-Based Spectral Structures by European Starlings (*Stumus vulgaris*)." *Journal of Experimental Psychology: General* 124:409–423.

Huron, D. 1997. "A Possible Evolutionary Origin for Auditory Dissonance." Paper presented at the CCRMA Hearding Seminar, January 23.

Huron, D. 2005. "Course notes for Music 829B." Available online at <http://dactyl.som.ohio-state.edu/Music829B/music829B.html>. Accessed 6 May, 2005.

Huron, D., and P. Sellmer. 1992. “Critical Bands and the Spelling of Vertical Sonorities.” *Music Perception* 10(2):129–149.

Hutchinson, W., and L. Knopoff. 1979. “The Significance of the Acoustic Component of Consonance in Western Triads.” *Journal of Musicological Research* 3:5–22.

Iamblichus. N.d. *Life of Pythagoras*. Trans. T. Taylor. London: J. M. Watkins, 1818. Reprinted 1986, Rochester, Vermont: Inner Traditions.

Izmirli, O. 2000. “Using a Spectral Flatness Based Feature for Audio Segmentation and Retrieval.” *Proceedings of the International Symposium on Music Information Retrieval (ISMIR2000)*. October 23–25, 2000, Plymouth, Massachusetts, USA.

Jacobsson, B., and J. Jerkert. 2000. “Consonance of Non-Harmonic Complex Tones: Testing the Limits of the Theory of Beats.” Available online at <http://www.f.kth.se/~f97-jje/musikteori/consonance.html>.

James, J. 1993. *Music of the Spheres: Music, Science, and the Natural Order of the Universe*. New York: Grove Press.

Jeans, J. 1937. *Science and Music*. Reprinted New York: Dover Publications, 1968.

Jenny, H. *Cymatics: The Structure and Dynamics of Waves and Vibrations* (Volume 1, 1967; Volume 2, 1972).

Jeppesen, K. 1939. *Counterpoint: The Polyphonic Vocal Style of the Sixteenth Century*, trans. Glen Hardon. Englewood Cliffs, New Jersey: Prentice-Hall.

Johnston, J. D. 1988. “Transform Coding of Audio Signals Using Perceptual Noise Criteria.” *IEEE Journal of Selected Areas in Communication* 6:314–323.

Jourdain, R. 1997. *Music, The Brain, and Ecstasy.* New York: Avon Books.

Kameoka, A., and M. Kuriyagawa. 1969a. "Consonance Theory, Part I: Consonance of Dyads." *Journal of the Acoustical Society of America* 45(6):1451–1459.

Kameoka, A., and M. Kuriyagawa. 1969b. "Consonance Theory, Part II: Consonance of Complex Tones and Its Computation Method." *Journal of the Acoustical Society of America*45(6):1460–1469.

Kamo, M., and Y. Iwasa. 2000. "Evolution of Preference for Consonances as a By-Product." Evolutionary Ecology Research 2:375–383.

Kane, B. 2005. "*L'Objet Sonore Maintenant*: Reflections on the Philosophical Origins of Musique Concrète." *Proceedings of the Spark 2005 Festival of Electronic Music and Art*. Minneapolis: University of Minnesota School of Music, pp. 58–61.

Katz, B. 2004. "A Measure of Musical Preference." *Journal of Consciousness Studies* 11(3–4):28–57.

Keislar, D. F. 1991. "The Relevance of Beating Partials for Musical Intonation." *Proceedings of the International Computer Music Conference 1991*. San Francisco: International Computer Music Association, pp. 202–205.

Keistner, G. 1909. "Untersuchungen über den Gefühlseindruck unanalysierter Zweiklänge." *Psychologische Studien* 4:473–504.

Kepler, J. 1619. *Harmonies of the World, Book Five. [Harmonices Mundi.*] Stephen Hawking, ed. Trans. Andrew Motte, 2005. Philadelphia: Running Press Book Publishers.

Kircher, A. 1650. *Musurgia Universalis*. Rome.

Klapuri, P. 2003. "Multiple Fundamental Frequency Estimation Based on Harmonicity and Spectral Smoothness." *IEEE Transactions on Speech and Audio Processing* 11:804–816.

Klyve, D., et al. 2006. "The Euler Archive." Available online at <http://math.dartmouth.edu/~euler>.

Kohonen, T. 1988. *Self-Organization and Associative Memory.* New York : Springer.

Kohonen, T. 1997. *Self-Organizating Maps*. New York : Springer.

Koivisto, T. 2000. "Rhythm, Meter, and Notated Meter: Structure in Elliott Carter's Second String Quartet." Paper presented at Music Theory Midwest 2000 Eleventh Annual Conference, Lawrence University, 19 May.

Kostek, B.,and A. Czyzewski. 2001. "Representing Musical Instrument Sounds for Their Automatic Classification." *Journal of the Audio Engineering Society* 768–785.

Krebs, H. 1987. "Some Extensions of the Concepts of Metrical Consonance and Dissonance." *Journal of Music Theory* 31:99–120.

Krebs, H. 1999. *Fantasy Pieces: Metrical Dissonance in the Music of Robert Schumann.* New York: Oxford University Press.

Kuhn, T. S. 1976. "Mathematical vs. Experimental Traditions in Development of Physical Science." *Journal of Interdisciplinary History* 7(1):1–31 .

L. R. Rabiner, J. J. Dubnowski, and R. W. Schafer. 1976. "Real-Time Digital Hardware Pitch Detector." *IEEE Transactions on Acoustics, Speech, and Signal Processing* ASSP-24(1):2–8.

L. R. Rabiner. 1977. "On the Use of Autocorrelation for Pitch Detection." *IEEE Transactions on Acoustics, Speech, and Signal Processing* ASSP-25(1):24–33.

Landy, L., ed. 2005. *Electro-Acoustic Resource Site*. Available online at <http://www.ears.dmu.ac.uk>.

Lane, R. D., et al. 1997. "Neuroanatomical Correlates of Happiness, Sadness, and Disgust." *American Journal of Psychiatry* 154:926–933.

Lee, R. E., K. M. Booth, J. Y. Reese-Smith, G. Regan, and H. H. Howard. 2005. "The Physical Activity Resource Assessment (PARA) Instrument: Evaluating Features, Amenitiese, and Incivilities of Physical Activity Resources in Urban Neighborhoods." *International Journal of Behavioral Nutrition and Physical Activity* 2(13).

Leider, C. 1996. "Morphological Metrics and Tuning Spaces." Unpublished manuscript.

Leider, C. 1997–. *Motus Lentus*. Audio compact disc. Homestead, Florida: Everglade.

Leider, C. 1997. *Nexus*. Homestead, Florida: Everglade.

Leider, C. 2000. "Dissonance as a Musical Control Structure." Unpublished manuscript.

Leman, M. 1995. *Music and Schema Theory*. Berlin: Springer, pp. 3–19.

Lerdahl, F. 1997. "Cognitive Constraints on Compositional Systems." *Contemporary Music Review* 6(2):97–121.

Lerdahl, F., and R. Jackendoff. 1977. "Toward a Formal Theory of Tonal Music." *Journal of Music Theory* 21:111–171.

Levelt, W., J. Van de Geer, and R. Plomp. 1966. "Triadic Comparisons of Musical Intervals." *British Journal of Mathematical and Statistical Psychology* 19:163–179

Levin, T. Y. 2003. "'Tones from out of Nowhere': Rudolph Pfenninger and the Archaeology of Synthetic Sound ." *Grey Room* 12:32–79.

Lipps, T. 1905. *Psychologische Studien*, 2nd ed. Leipzig: Verlag der Dürr'Schen Bachhandlung.

Lipps, T. 1995. *Consonance and Dissonance in Music*. Trans. W. Thompson. San Marino, California: Everett Books.

Looy, H. 2001. "Disgust, Morality, and Human Identity." Philadelphia, PA: Metanexus Institute. Available online at <http://www.metanexus.net/metanexus_online/show_article.asp?4978>)

Looy, H. 2004. "Embodied and Embedded Morality: Divinity, Identity, and Disgust." *Zygon* 39(1):219–235.

Lundin, R. W. 1947. "Toward a Cultural Theory of Consonance." *Journal of Psychology* 23:45–49.

M. M. Sohndi. 1968. "New Methods of Pitch Extraction." *IEEE Transactions on Audio and Electroacoustics* AU-16(2):262–266.

MacQueen, J. B. 1967. Some Methods for Classification and Analysis of Multivariate Observations." *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability.* Berkeley: University of California Press, 281–297.

Maher, T. (1982). "Verbal and Exploratory Responses to Melodic Musical Intervals." *Psychology of Music* 10(1):11–27.

Maher, T. 1980. "A Rigorous Test of the Proportion That Musical Intervals Have Different Psychological Effects." *American Journal of Psychology* 93(2), 309-327.

Maher, T. F. 1976. "'Need for Resolution' Ratings for Harmonic Musical Intervals: A Comparison Between Indians and Canadians." *Journal of Cross Cultural Psychology* 7:259–276.

Maher, T. F. 1980. "A Rigorous Test of the Proposition That Musical Intervals Have Different Psychological Effects." *American Journal of Psychology* 93(2):309–327.

Makeig, S. 1979/1980. "The Affect of Musical Intervals." *Interval* 2(1):18–19, 24–26; 2(2–3): 112–116; 2(4): 11–13.

Malin, J. 2000. "Metrical Dissonance, Energetics and Music-Text Relations in Schoenberg's 'Valse de Chopin'." Paper presented at Music Theory Midwest 2000 Eleventh Annual Conference, Lawrence University, 19 May.

Malmberg, C. F. 1918. "The Perception of Consonance and Dissonance." *Psychological Monographs* 25(2):93–133.

Manaris, B., et al. 2005. "Zipf's Law, Music Classification, and Aesthetics." *Computer Music Journal* 29(1).

Manning, P. 1993. *Electronic & Computer Music*. Oxford: Clarendon Press.

Marquis-Favre, C., E. Premat, and D. Aubrée. 2005. "Noise and its Effects—A Review on Qualitative Aspects of Sound. Part II: Noise and Annoyance." *Acta Acustica united with Acustica* 91(4):626–642.

Marquis-Favre, C., E. Premat, D. Aubrée, and M. Vallet. 2005. "Noise and its Effects—A Review on Qualitative Aspects of Sound. Part I: Notions and Acoustic Ratings." *Acta Acustica united with Acustica* 91(4):613–625.

Martens, P. 2000. "Non-Binary Strategies in Led Zeppelin's 'The Ocean'." Paper presented at Music Theory Midwest 2000 Eleventh Annual Conference, Lawrence University, 19 May.

Mathews, M. V., and J. R. Pierce. 1980. "Harmony and Nonharnomic Partials." *Journal of the Acoustical Society of America* 68:1252–1257.

Mattis, O. 2005. *Visual Music*. London: Thames and Hudson.

Mayer, A. M. 1894. "Researches in Acoustics, No. IX." *Philosophical Magazine* 37:259–288.

McAdams, S. 1982. "Spectral Fusion and the Creation of Auditory Images." In M. Clynes, ed. *Music, Mind and Brain*. New York: Plenum.

McAdams, S. 1999. "Perspectives on the Contribution of Timbre to Musical Structure." *Computer Music Journal* 23(3):85–102.

McAuley, J. D. 1995. "Perception of Time as Phase: Toward and Adaptive Oscillator Model of Rhythmic Pattern Processing." Ph.D. thesis, Computer Science and Cognitive Science, Indiana University.

Meyer, M. 1903. "Experimental Studies in the Psychology of Music." *American Journal of Psychology* 14:456–478.

Mierswa, I., et al. 2006. "YALE: Rapid Prototyping for Complex Data Mining Tasks." In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-06).*

Miranda, E. R., K. Sharman, K. Kilborn, and A. Duncan. 2003. "On Harnessing the Electroencephalogram for the Musical Braincap." *Computer Music Journal* 27(2):80–102.

Miranda, E. R., Roberts, S. and Stokes, M. 2004. "On Generating EEG for Controlling Musical Systems." *Biomedical Engineering* 49(1):75–76.

Montani, A. 1947. "Outline of a Physiological Theory of Musical Consonance." *Riv. Musicale Ital.* 49:168–176.

Monzo, J. N.d. *Encyclopedia of Tuning*. Available online at <http://tonalsoft.com/enc/fr_index.html>. Last accessed 14 November, 2005.

Moore, B. C. J. 1989. *An Introduction into the Psychology of Hearing*, 3rd ed. London: Academic Press.

Moore, B. C. J., and B. R. Glasberg. 1996. "A Revision of Zwicker's Loudness Model." *Actua Acustica* 82:335–345.

Moorer, J. A. 2001. "Extension of the Method of Spatial Harmonics to Three Dimensions." Unpublished manuscript. Available online at <http://www.jamminpower.com/main/unpublished.jsp>.

Moran, H., and C. C. Pratt. 1926. "Variability of Judgments of Musical Intervals." *Journal of Experimental Psychology* 9:492–500.

Morris, J. S., et al. 1996. "A Differential Neural Response in the Human Amygdala to Fearful and Happy Facial Expressions." *Nature* 383:812–815.

Moylan, W. D. 1983. *An Analytical System for Electronic Music*. D. A. Dissertation, Ball State University.

Muzzulini D. 1994. "Euler, Leonhard Theory of Consonance." *Musiktheorie* 9(2):135–146.

Näätänen, R. 1995. "The Mismatch Negativity: A Powerful Tool for Cognitive Neuroscience." *Ear Hear* 16(1):6–18.

Näätänen, R., T. Jacobsen, and I. Winkler. 2005. "Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence." *Psychophysiology* 42(1):25.

Nesse, R. M. 1991. "What Good Is Feeling Bad? The Evolutionary Benefits of Psychic Pain." *The Sciences* November/December:30–37.

Noll, A. M. 1969. "Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate." Proceedings of the Symposium on Computer Processes in Communication, pp. 779–798.

Nordmark, J., and L. E. Fahlén. 1988. "Beat Theories of Musical Consonance." *Quarterly Progress and Status Report* 1:111-112. Stockholm: Royal Institute of Technology, Speech Transmission Laboratory.

O'Keeffe, V. 1975. "Psychophysical Preference for Harmonized Musical Passages in the Just and Equal Tempered Systems." *Perceptual and Motor Skills* 40:192–194.

Palombini, C. 1993b. "Machine Songs V: Pierre Schaeffer—From Research into Noises to Experimental Music." *Computer Music Journal* 17(3):14–19.

Palombini, C. 1993. "Pierre Schaeffer, 1953: Towards and Experimental Music." *Music and Letters* 74:542–557.

Palombini, C. 1993a. "Pierre Schaeffer, 1953: Towards and Experimental Music." *Music and Letters* 74:542–557.

Palombini, C. V. de Lima. 1993. *Pierre Schaeffer's Typo-Morphology of Sonic Objects*. Ph.D. Dissertation. University of Durham, UK.

Parker, D. B. 1986. "A More Natural Octavian Scale Derived from Euler." *Journal of Musicological Research* 6(4):353–355.

Parncutt, R. 1989. *Harmony: A Psychoacoustical Approach*. Berlin: Springer-Verlag.

Parncutt, R. 1994. "A Perceptual Model of Pulse Salience and Metrical Accent in Musical Rhythms." *Music Perception* 11(4):409–464.

Partch, H. 1979. *Genesis of a Music*. New York: Da Capo Press.

Peinaar, E. A. 1997. "Number Symbolism in Music: George Crumb's *Makrokosmos I* and the *Music of the Spheres*." M. Mus. thesis. University of Cape Town.

Philippe, P. "Low-Level Musical Descriptors for MPEG-7." *Signal Processing: Image Communication* DIH/HDM Department, CNET-France Telecom, BP59, 35512 Cesson Sévigné Cédex, France, pp. 181–191.

Plato. Ca. 360 B.C. *Timaeus*. Trans. B. Jowett (1817–1893). Available online at <http://www.gutenberg.org/etext/1572>.

Plomp, R. 1967. "Beats of Mistuned Consonances." *Journal of the Acoustical Society of America* 42:462–474.

Plomp, R. 1970. "Old and New Data on Tone Perception." Instituut Voor Zintuigfysiologie, RVO-TNO, Report No. IZF1970-14.

Plomp, R. 1976. *Aspects of Tone Sensation: A Psychophysical Study.* London: Academic Press.

Plomp, R., and A. M. Mimpen. 1968. "The Ear As a Frequency Analyzer." *Journal of the Acoustical Society of America* 36:1628–1636.

Plomp, R., and H. J. M. Steeneken. 1968. "Interference Between Two Simple Tones." *Journal of the Acoustical Society of America* 43:883–884.

Plomp, R., and W. J. M. Levelt. 1962. "Musical Consonance and Critical Bandwidth." *Proceedings of the Fourth International Congress on Acoustics,* Copenhagen.

Plomp, R., and W. J. M. Levelt. 1965. "Tonal Consonance and Critical Bandwidth." *Journal of the Acoustical Society of America* 38:548–560.

Pohlmann, K. 2005. Personal communication. 15 August.

Polansky, L. 1978. *Psaltery*. Recorded on *The Theory of Impossible Melody*. Audio compact disc. Berkeley, California: Artifact Recordings, 1989.

Polansky, L. 1989. Liner notes to accompany *The Theory of Impossible Melody*. Berkeley, California: Artifact Recordings.

Pollard, H. F., and E. V. Janson. 1982. "A Tristimulus Method for the Specification of Musical Timbre." *Acustica* 51: PAGES.

Preis, D., and V. C. Georgopoulos. 1998. "Wigner Distribution Representation and Analysis of Audio Signals: An Illustrated Tutorial Review." *Proceedings of the 104th Audio Engineering Society Convention*, Preprint 4682.

Rakowski, A. 1985. "The Perception of Musical Intervals by Music Students." *Council for Research in Music Education Bulletin* 85:175–186.

Rakowski, A., and A. Miskiewicz. 1985. "Deviations from Equal Temperament in Tuning Isolated Musical Intervals." *Archives of Acoustics* 10(2):95–104.

Randel, D. M., ed. *The New Harvard Dictionary of Music*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press, p. 197.

Rasch, R. 1984. "Theory of Helmholtz-Beat Frequencies." *Music Perception* 1:308–322.

Rasch, R., and R. Plomp. 1982. "The Perception of Musical Tones." In D. Deutsch, ed. *The Psychology of Music.* San Diego: Academic Press, pp. 89–112.

Reich, S. 1974. *Writings About Music*. New York: New York University Press.

Reilly, R. R. 1999. "Is Music Sacred?" *Crisis Magazine* September:27–31.

Risset, J.-C. 1969. "Pitch Control and Pitch Paradoxes Demonstrated with Computer-Synthesized Sounds." *Journal of the Acoustical Society of America* 46(1A):88.

Risset, J.-C. 1997. "Rhythmic Paradoxes and Illusions: A Musical Illustration." *Proceedings of the 1997 International Computer Music Conference*. San Francisco: International Computer Music Association, pp. 7–10.

Risset, J.-C., and D. Wessel. "Exploration of Timbre by Analysis and Synthesis." In D. Deutsch, ed. *The Psychology of Music*, Second Edition. San Diego: Academic Press, pp. 113–???

Ritossa, D. A., and N. S. Rickard. 2004. "The Relative Utility of 'Pleasantness' and 'Liking' Dimensions in Predicting the Emotions Expressed in Music." *Psychology of Music*. 32(1):5–22.

Roberts, L. A. 1986. "Consonance Judgments of Musical Chords by Musicians and Untrained Listeners." *Acustica* 62(2):163–171.

Roeder, J. 2001. "Pulse Streams and Problems of Grouping and Metrical Dissonance in Bartók's 'With Drums and Pipes.'" *Music Theory Online* 7(1).

Rogan, M. T., and J. E. LeDoux. 1996. "Emotion: Systems, Cells, Synaptic Plasticity." *Cell* 85:469–475.

Roig-Francolí, M. A. 2003. *Harmony in Context*. New York: McGraw-Hill.

Russolo, L. 1916. *The Art of Noises*, trans. B. Brown (1986). New York: Pendragon Press.

Samarotto, F. P. 2000. "'The Body that Beats': Review of Harald Krebs, *Fantasy Pieces: Metrical Dissonance in the Music of Robert Schumann*." *Music Theory Online* 6(4).

Sander, K., and H. Scheich. 2001. "Auditory Perception of Laughing and Crying Activates Human Amygdala Regardless of Attentional State." *Cognitive Brain Research* 12:181–198 .

Schachtel, E. G. 1984. *Metamorphosis: On the Development of Affect, Perception, Attention, and Memory*. New York: Da Capo Press.

Schaeffer, P. 1966. *Traité des objets musicaux.* Paris: Éditions du Seuil.

Schaeffer, P. 1967. *La Musique Concréte*. Paris: Presses Universitares de France.

Schaeffer, P., and G. Reibel. 1967. *Solfège de l'Objet Sonore*. Paris: Éditions du Seuil.

Schaeffer, P., ed. 1957. "Vers une musique expérimentale." *La revue musicale* 236. Paris: Richard-Masse, pp. 11–27.

Schafer, R. M. 1968. *Epitaph for Moonlight*. Don Mills, Ontario: BMI Canada Limited.

Schafer, R. M. 1973. *First String Quartet*. Toronto: Universal Edition (Canada) Ltd.

Schafer, R. M. 1992. "Music, Non-Music and the Soundscape." In J. Paynter, T. Howell, R. Orton, and P. Seymour, eds. *Companion to Contemporary Musical Thought*, Vol 1. London: Routledge, pp. 34–45.

Schafer, R. M. 1967. *Ear Cleaning*. Don Mills, Ontairo: Berandol Music Limited.

Schafer, R. M. 1994. *The Soundscape: Our Sonic Environment and the Tuning of the World*. Rochester, Vermont: Destiny Books. (First edition, New York: Random House, 1977.)

Schienle, A., et al. 2002. "The Insula Is Not Specifically Involved in Disgust Processing: An fMRI Study." *Neuroreport* 13(16):2023–2026.

Schulze W. 1987. "A Harmonical Analog: Leibniz's Genealogical Scheme in *Dissertatio de Arte Combinatoria*." *Studia Leibnitiana* 19(1):98–116.

Scriba, C. J. 1990. "Mathematics and Music." *Normat* 38(1):3–17, 52.

Seashore, C. E. 1967. *Psychology of Music*. Peter Smith Publishers.

Sethares, W. 1993. "Local Consonance and the Relationship Between Timbre and Scale." *Journal of the Acoustical Society of America* 94(3):1218–1228.

Sethares, W. 1993. "Relating Timbre and Tuning." *Experimental Musical Instruments* 12:22–29.

Sethares, W. 1998. "Consonance-Based Spectral Mappings." *Computer Music Journal* 22(1):56–72.

Sethares, W. 1999. *Tuning, Timbre, Spectrum, Scale*. New York: Springer.

Shepard, R. N. 1964. "Circularity in Judgments of Relative Pitch." *Journal of the Acoustical Society of America* 36:23–46.

Simpson, J. 1994. *Cochlear Modeling of Sensory Dissonance and Chord Roots*. Master of Applied Science thesis, University of Waterloo.

Simpson, R. 1665. *The Division-Viol.* London. (Facsimile Edition published by Musikverlag Emil Katzbichler, Munchen-Salzburg, 1983.)

Slonimsky, N. 1997. "R. Murray Schafer." *Baker's Biographical Dictionary of Twentieth-Century Musicians*. Ed. Laura Kuhn. New York: Schirmer Books, 1997. 1186.

Smalley, D. 1992. "The Listening Imagination: Listening in the Electroacoustic Era." *Companion to Contemporary Musical Thought*, Vol 1. In J. Paynter, T. Howell, R. Orton, and P. Seymour, eds. *Companion to Contemporary Musical Thought*, Vol 1. London: Routledge, pp. 514–554.

Smalley, D. 1993. "Defining Transformations." *Interface* 22:279–300.

Smalley, D. 1986. "Spectro-Morphology and Structuring Processes." In S. Emmerson ed. *The Language of Electroacoustic Music* London: Macmillan.

Smaragdis, P. 2001. "Redundancy Reduction for Computational Audition: A Unifying Approach." Ph.D. Thesis, Perceptual Computing, Massachusetts Institute of Technology.

Steedman, M. 1994. "The Well-Tempered Computer." *Philosophical Transactions of the Royal Society of London, Series A*, 349: 115–131.

Steedman, M. 2002. "Helmholtz' and Longuet-Higgins' Theories of Consonance and Harmony." Unpublished manuscript.

Stevens, S. S. 1955. "The Measurement of Loudness." *Journal of the Acoustical Society of America* 27(5):815–829.

Stevens, S. S., J. Volkman, and E. B. Newman. 1937. “A Scale for the Measurement of the Psychological Magnitude of Pitch.” *Journal of the Acoustical Society of America* 8(3):185–190.

Stone, L. 1998. “The Correlation Dimension as a Metric for Timbre.” A. M. Thesis, Electro-Acoustic Music, Dartmouth College.

Stumpf, C. 1890. *Tonpsychologie*. Leipzig: S. Hirzel.

Stumpf, C. 1898. *Konsonanz und Dissonanz*. Beiträge zur Akustischen Musikwissenschaft, Vol. 1, pp. 1–108.

Szeider, S. H. 1990. “The Semantics of Musical Scores: An Approach to the Mechanical Composition of Microtonal Music.” *Interface* 19(2-3):175–184.

Tame, D. 1984. *The Secret Power of Music*. Rochester, New York: Destiny Books.

Tenney, J. 1965. *The Physical Correlates of Timbre*. *Gravesaner Blaetter* 26:106–109.

Tenney, J. 1988. *A History of ‘Consonance’ and ‘Dissonance.’* New York: Excelsior.

Tenney, J. 1988. *Critical Band*. Recorded on Relâche, *On the Edge*. Audio compact disc. Key Gardens, New York: Mode Records, 1990.

Tenney, J. 1992. *META-HODOS and META Meta-Hodos*, 2nd ed. Hanover, New Hampshire: Frog Peak Music.

Terhardt, E. 1972. VIRTUAL PITCH ARTICLE.

Terhardt, E. 1974a. “Pitch, Consonance, and Harmony.” *Journal of the Acoustical Society of America* 55(5):1061–1069.

Terhardt, E. 1974b. ‹‹Über akustische Rauhigkeit und Scwankungsstärke›› (“On the Perception of Period Sound Fluctuations (Roughness)”). *Acustica* 20:215–224.

Terhardt, E. 1976. “Concept and Psychoacoustic Foundations of Musical Consonance.” *Acustica* 36(3):121–137.

Terhardt, E. 1984. “The Concept of Musical Consonance: A Link Between Music and Psychoacoustics.” *Music Perception* 1(3):276–295.

Toiviainen, P., M. Kaipainen, and J. Louhivuori. 1995. “Musical Timbre: Similarity Ratings Correlate with Computational Feature Space Distances.” *Journal of New Music Research* 24:282–298.

Tramo, M. J., et al. 2001. “Neurobiological Foundations for the Theory of Harmony in Western Tonal Music.” *Ann. NY Acad. Sci.* 930:92–116.

Tramo, M. J., P. C. Cariani, and B. Delgutte. 1992. “Representation of Tonal Consonance in the Temporal Fring Patterns of Auditory-Nerve Fibers: Responses to Musical Intervals Composed of Pure Tones vs. Harmonic Complex Tones.” *Society for Neuroscience Abstracts* 18:382.

Trehub, S., G. Schellenberg, and D. Hill. 1997. “The Origins of Music Perception and Cognition: A Developmental Perspective.” In I. Deliège and J. Sloboda, eds. *Perception and Cognition of Music*. East Sussex, UK: Psychology Press Ltd., pp. 103–128.

Truax, B. 1990. “Electro-Environment.” *EAR Magazine* 15(7):28–29.

Truax, B. 1999. *Handbook for Acoustic Ecology*, 2nd ed. Vancouver: Cambridge Street Publishing.

Tzanetakis, G., and P. Cook. 2000. "Audio Information Retrieval (AIR) Tools." *Proceedings of the International Symposium on Music Information Retrieval.*

Tzanetakis, G., G. Essl, and P. Cook. 2001. "Audio Analysis Using the Discrete Wavelet Transform." *Proceedings of the Conference on Acoustics and Music Theory Applications*, September.

Ungeheuer G. 1983. "On the Acoustics of Vowel Sounds in the 18th Century: The Euler-Lambert Correspondence and Kratzenstein." *Phonetica* 40(2):145–171.

Vogel, M. 1993. *On the Relations of Tone*. Trans. V. J. Kisselbach, ed C. A. Poldy. Bonn: Verlag für Systematische Musikwissenschaft GmbH.

von Bismarck, G. 1974a. "Timbre of Steady Sounds: A Factorial Investigation of its Verbal Attributes." *Acustica* 30:146–159.

von Bismarck, G. 1974b. "Sharpness as an Attribute of the Timbre of Steady Sounds." *Acustica,* 30:159–172.

Vos, J. 1986. "Purity Ratings of Tempered Fifths and Major Thirds." *Music Perception* 3(2):221–258.

Voss, A. 1988. "The Music of the Spheres, Marsilio Ficino and Renaissance Harmonia." *Culture and Cosmos* 2(2):16–38.

Voss, A. 2000. "The Astrology of Marsilio Ficino: Divination or Science?" *Culture and Cosmos* 4(2):29–46.

Waller, M. D. 1961. *Chladni Figures: A Study in Symmetry*. London: G. Bell and Sons.

Wapnick, J., G. Bourassa, and J. Sampson. 1982. "The Perception of Tonal Intervals in Isolation and in Musical Context." *Psychomusicology* 2(1):21–37.

Watkins, A. J., and M. C. Dyson. 1985. "On the Perceptual Organization of Tone Sequences and Melodies." In P. Howell, I. Cross, and R. West, eds. *Musical Structure and Cognition*. London: Academic Press.

Watkins, A., and M. Dyson. 1985. "On the Perceptual Organization of Tone Sequences and Melodies." In *Musical Structure and Cognition*. London: Academic Press, pp. 71–119.

Weinberger, N. 1999. "Music and the Auditory System." In D. Deutsch, ed. *The Psychology of* Music, Second Edition. San Diego: Academic Press, pp. 47–87.

Weinberger, N. M. 1997. "The Musical Infant and the Roots of Consonance." *MuSICA* 4(1). Available online at <http://www.musica.uci.edu/mrn/V4I1S97.html#consonance>.

Weisstein, E. W. 2004. "Spherical Harmonic." From *MathWorld*—A Wolfram Web Resource. Availanle online at <http://mathworld.wolfram.com/SphericalHarmonic.html>.

Wessel, D. 1973. "Psychoacoustics and Music: A Report from Michigan State University." *PACE: Bulletin of the Computer Arts Society* 30:1–2.

Wessel, D. 1979. "Timbre Space as a Musical Control Structure." *Computer Music Journal* 3(2):45–52.

Weyl, H. 1959. "Mathematics and the Laws of Nature." In I. Gordon and S. Sorkin, eds. *The Armchair Science Reader*. New York: Simon and Schuster.

Wiener, P. P. 1973. *Dictioinary of the History of Ideas: Studies of Selected Pivotal Ideas*. New York: Charles Scribner's Sons.

Wilson C. 1985. "The Great Inequality of Jupiter and Saturn: From Kepler to Laplace." *Archive for History of Exact Sciences* 33(1–3):15–290 .

Wishart, T. 1988. "The Composition of *Vox-5*." *Computer Musical Journal* 12(4):21–27.

Wishart, T. 1993. "From Architecture to Chemistry." *Interface* 22:301–315.

Wishart, T. 1994. *Audible Design: A Plain and Easy Introduction to Practical Sound Composition*. York, UK: Orpheus the Pantomime.

Wishart, T. 1983. "Music/Change." *Interface* 12:103–109.

Wishart, T. 1985. *On Sonic Art*. York, UK: Imagineering Press.

Wishart, T. 1990. "Extended Vocal Technique." *The Musical Times* 121:1647, May 1990: 313-314.

Wishart, T. 1992. "Music and Technology: Problems and Possibilities." *Companion to Contemporary Musical Thought*, Vol 1. In J. Paynter, T. Howell, R. Orton, and P. Seymour, eds. *Companion to Contemporary Musical Thought*, Vol 1. London: Routledge, pp. 565–582.

Wishart, T. 2000. "Sonic Composition in *Tongues of Fire*." *Computer Music Journal* 24(2).

Woody, S. R., and B. A. Teachman. 2000. "Intersection of Disgust and Fear: Normative and Pathological Views." *Clinical Psychology* 7(3):291–311.

Xenakis, I. 1966. “Towards a Philosophy of Music.” *Gravesäner Blätter* 29:39–52.

Xenakis, I. 1971. *Formalized Music: Thought and Mathematics in Composition*, 2nd ed. Bloomington, Indiana: Indiana University Press.

Xenakis, I. 1978. *The Legend of Er*.Recorded on *Iannis Xenakis—Volume 5*. Audio compact disc and DVD. Mode Records 148.

Y. I. Fishman, et al. 2001. “Consonance and Dissonance of Musical Chords: Neural Correlates in Auditory Cortex of Monkeys and Humans.” *Journal of Neurophysiology* 86:2761–2788.

Yang, W., and J. Kang. 2005. “Soundscape and Sound Preferences in Urban Squares: A Case Study in Sheffield.” Journal of Urban Design 10(1):61–80.

Yeary, M. 2003. “Einheitlichkeit: The Legacy of Carl Stumpf’s Theory of Tonal Fusion.” Unpublished manuscript. Available online at <http://home.uchicago.edu/~yeary/papers.html>.

Yeston, M. 1976. *The Stratification of Musical Rhythm.* New Haven: Yale University Press.

Zentner, M., and J. Kagan. “Infant’s Perception of Consonance and Dissonance in Music.” *Infant Behavior and Development* 21(3):483–492.