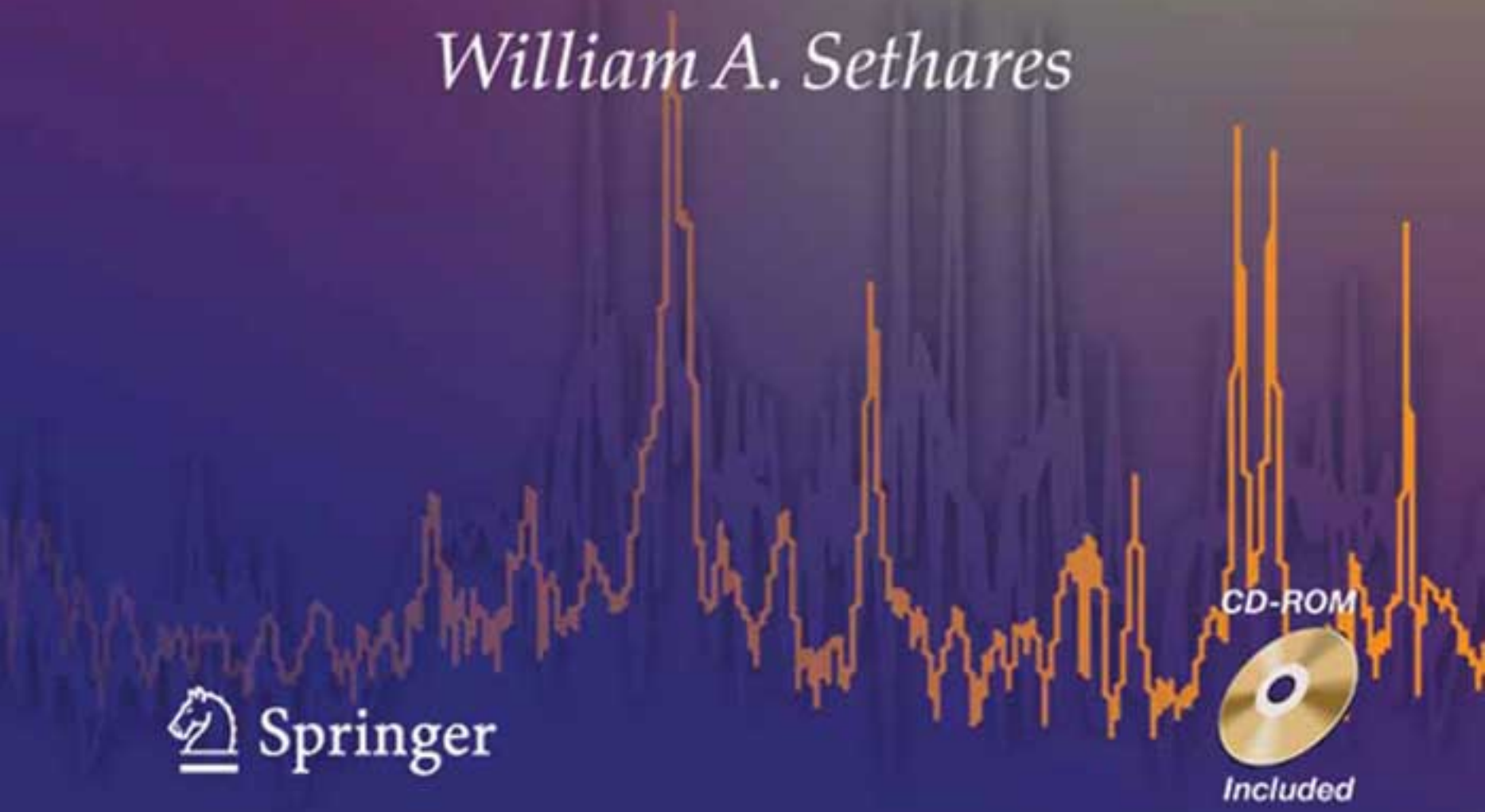
Tuning, Timbre, Spectrum, Scale

Second Edition

William A. Sethares



Springer

Tuning, Timbre, Spectrum, Scale

William A. Sethares

Tuning, Timbre, Spectrum, Scale

Second Edition

With 149 Figures

Springer

William A. Sethares, Ph.D.
Department of Electrical and Computer Engineering
University of Wisconsin–Madison
1415 Johnson Drive
Madison, WI 53706-1691
USA

British Library Cataloguing in Publication Data
Sethares, William A., 1955–
Tuning, timbre, spectrum, scale.—2nd ed.
1. Sound 2. Tuning 3. Tone color (Music) 4. Musical intervals and scales 5. Psychoacoustics 6. Music—Acoustics and physics
I. Title
781.2′3
ISBN 1852337974

Library of Congress Cataloging-in-Publication Data
Sethares, William A., 1955–
Tuning, timbre, spectrum, scale / William A. Sethares.
p. cm.
Includes bibliographical references and index.
ISBN 1-85233-797-4 (alk. paper)
1. Sound. 2. Tuning. 3. Tone color (Music) 4. Musical intervals and scales.
5. Psychoacoustics. 6. Music—Acoustics and physics. I. Title.
QC225.7.S48 2004
534—dc22 2004049190

ISBN 1-85233-797-4 2nd edition Springer-Verlag London Berlin Heidelberg
ISBN 3-540-76173-X 1st edition Springer-Verlag Berlin Heidelberg New York
Springer Science+Business Media
springeronline.com

Printed in the United States of America

First published 1999
Second edition 2005

Typesetting: Camera-ready by author
69/3830-543210 Printed on acid-free paper SPIN 10937952

Prelude

> *The chords sounded smooth and nondissonant but strange and somewhat eerie. The effect was so different from the tempered scale that there was no tendency to judge in-tuneness or out-of-tuneness. It seemed like a peek into a new and unfamiliar musical world, in which none of the old rules applied, and the new ones, if any, were undiscovered.* F. H. Slaymaker [B: 176]

To seek out new tonalities, new timbres...
To boldly listen to what no one has heard before.

Several years ago I purchased a musical synthesizer with an intriguing feature—each note of the keyboard could be assigned to any desired pitch. This freedom to arbitrarily specify the tuning removed a constraint from my music that I had never noticed or questioned—playing in 12-tone equal temperament.[1] Suddenly, new musical worlds opened, and I eagerly explored some of the possibilities: unequal divisions of the octave, n equal divisions, and even some tunings not based on the octave at all.

Curiously, it was much easier to play in some tunings than in others. For instance, 19-tone equal temperament (*19-tet*) with its 19 equal divisions of the octave is easy. Almost any kind of sampled or synthesized instrument plays well: piano sounds, horn samples, and synthesized flutes all mesh and flow. 16-tet is harder, but still feasible. I had to audition hundreds of sounds, but finally found a few good sounds for my 16-tet chords. In 10-tet, though, none of the tones in the synthesizers seemed right on sustained harmonic passages. It was hard to find pairs of notes that sounded reasonable together, and triads were nearly impossible. Everything appeared somewhat out-of-tune, even though the tuning was precisely ten tones per octave. Somehow the timbre, or tone quality of the sounds, seemed to be interfering.

The more I experimented with alternative tunings, the more it appeared that certain kinds of scales sound good with some timbres and not with others. Certain kinds of timbres sound good in some scales and not in others. This raised a host of questions: What is the relationship between the timbre of a

[1] This is the way modern pianos are tuned. The seven white keys form the major scale, and the five black keys fill in the missing tones so that the perceived distance between adjacent notes is (roughly) equal.

sound and the intervals, scale, or tuning in which the sound appears "in tune?" Can this relationship be expressed in precise terms? Is there an underlying pattern?

This book answers these questions by drawing on recent results in psychoacoustics, which allow the relationship between timbre and tuning to be explored in a clear and unambiguous way. Think of these answers as a model of musical perception that makes predictions about what you hear: about what kinds of timbres are appropriate in a given musical context, and what kind of musical context is suitable for a given timbre.

Tuning, Timbre, Spectrum, Scale begins by explaining the relevant terms from the psychoacoustic literature. For instance, the perception of "timbre" is closely related to (but also distinct from) the physical notion of the *spectrum* of a sound. Similarly, the perception of "in-tuneness" parallels the measurable idea of *sensory consonance*. The key idea is that consonance and dissonance are not inherent qualities of intervals, but they are dependent on the spectrum, timbre, or tonal quality of the sound. To demonstrate this, the first sound example on the accompanying CD plays a short phrase where the octave has been made dissonant by devious choice of timbre, even though other, nonoctave intervals remain consonant. In fact, almost any interval can be made dissonant or consonant by proper sculpting of the timbre.

Dissonance curves provide a straightforward way to predict the most consonant intervals for a given sound, and the set of most-consonant intervals defines a scale *related* to the specified spectrum. These allow musicians and composers to design sounds according to the needs of their music, rather than having to create music around the sounds of a few common instruments. The spectrum/scale relationship provides a map for the exploration of inharmonic musical worlds.

To the extent that the spectrum/scale connection is based on properties of the human auditory system, it is relevant to other musical cultures. Two important independent musical traditions are the gamelan ensembles of Indonesia (known for their metallophones and unusual five and seven-note scales) and the percussion orchestras of classical Thai music (known for their xylophone-like idiophones and seven-tone equal-tempered scale). In the same way that instrumental sounds with harmonic partials (for instance, those caused by vibrating strings and air columns) are closely related to the scales of the West, so the scales of the gamelans are related to the spectrum, or tonal quality, of the instruments used in the gamelan. Similarly, the unusual scales of Thai classical music are related to the spectrum of the xylophone-like *renat*.

But there's more. The ability to measure sensory consonance in a reliable and perceptually relevant manner has several implications for the design of audio signal processing devices and for musical theory and analysis. Perhaps the most exciting of these is a new method of *adaptive tuning* that can automatically adjust the tuning of a piece based on the timbral character of the music so as to minimize dissonance. Of course, one might cunningly seek to

maximize dissonance; the point is that the composer or performer can now directly control this perceptually relevant parameter.

The first several chapters present the key ideas in a nonmathematical way. The later chapters deal with the nitty-gritty issues of sound generation and manipulation, and the text becomes denser. For readers without the background to read these sections, I would counsel the pragmatic approach of skipping the details and focusing on the text and illustrations.

Fortunately, given current synthesizer technology, it is not necessary to rely only on exposition and mathematical analysis. You can actually listen to the sounds and the tunings, and verify for yourself that the predictions of the model correspond to what you hear. This is the purpose of the accompanying CD. Some tracks are designed to fulfill the predictions of the model, and some are designed to violate them; it is not hard to tell the difference. The effects are not subtle.

Madison, Wisconsin, USA *William A. Sethares*
August 2004

Acknowledgments

This book owes a lot to many people.

The author would like to thank *Tom Staley* for extensive discussions about tuning. Tom also helped write and perform *Glass Lake*. *Brian McLaren* was amazing. He continued to feed me references, photocopies, and cartoons long after I thought I was satiated. Fortunately, he knew better. The names *Paul Erlich*, *wallyesterpaulrus*, *paul-stretch-music*, *Jon Szanto*, and *Gary Morrison* have been appearing daily in my e-mail inbox for so long that I keep thinking I know who they are. Someday, the galactic oversight of our never having met will be remedied. This book would be very different without *Larry Polansky*, who recorded the first gamelan "data" that encouraged me to go to Indonesia and gather data at its source. When *Basuki Rachmanto* and *Gunawen Widiyanto* became interested in the gamelan recording project, it became feasible. Thanks to both for work that was clearly above and beyond my hopes, and to the generous gamelan masters, tuners, and performers throughout Eastern Java who allowed me to interview and record. *Ian Dobson* has always been encouraging. He motivated and inspired me at a very crucial moment, exactly when it was most needed. Since he probably doesn't realize this, please don't tell him – he's uppity enough as it is. *John Sankey* and I co-authored the technical paper that makes up a large part of the chapter on musicological analysis without ever having met face to face. Thanks, bf250. *David Reiley* and *Mary Lucking* were the best guinea pigs a scientist could hope for: squeak, squeak. *Steve Curtin* was helpful despite personal turmoil, and *Fred Spaeth* was patently helpful. The hard work of *Jean-Marc Fraïssé* appears clearly on the CD interface. Thanks also to my editors *Nicholas Pinfield*, *Christopher Greenwell*, and *Michael Jones* for saying "yes" to a book that otherwise might have fallen into the cracks between disciplines, and to *Deirdre Bowden* and *Alison Jackson* for putting up with mutant eps files and an unexpected appendectomy. *Anthony Doyle* and *Michael Koy* were instrumental in the preparation of this second edition. Several of the key ideas in this book first appeared in academic papers in the *Journal of the Acoustical Society of America*, and *Bill Strong* did far more than coordinate the review

process. Bill and his corps of anonymous reviewers were enlightening, thought provoking, and frustrating, sometimes all at once. *Marc Leman's* insights into psychoacoustics and *Gerard Pape's* thoughts on composition helped me refine many of the ideas. Thanks to everyone on the "Alternate Tuning Mailing List" and the "MakeMicroMusic list" at [W: 1] and [W: 18] who helped keep me from feeling isolated and provided me with challenge and controversy at every step.

The very greatest thanks go to *Ann Bell* and the *Bunnisattva*.

Contents

Variables, Abbreviations, Definitions

a_i	Amplitudes of sine waves or partials.
attack	The beginning portion of a signal.
[B:]	Reference to the bibliography, see p. 381.
cent	An octave is divided into 1200 equal sounding parts called cents. See Appendix B.
-cet	Abbreviation for cent-equal-temperament. In n-cet, there are n cents between each scale step; thus, 12-tet is the same as 100-cet.
CDC	Consonance-Dissonance Concept, see [B: 192].
$d(f_i, f_j, a_i, a_j)$	Dissonance between the partials at frequencies f_i and f_j with corresponding amplitudes a_i and a_j.
[D:]	Reference to the discography, see p. 395.
DFT	Discrete Fourier Transform. The DFT of a waveform (sound) shows how the sound can be decomposed into and rebuilt from sine wave partials.
D_F	Intrinsic dissonance of the spectrum F.
$D_F(c)$	Dissonance of the spectrum F at the interval c.
diatonic	A seven-note scale containing five whole steps and two half steps such as the common major and minor scales.
envelope	Evolution of the amplitude of a sound over time.
F	Name of a spectrum with partials at frequencies $f_1, f_2, ..., f_n$ and amplitudes $a_1, a_2, ..., a_n$.
f_i	Frequencies of partials.
fifth	A 700-cent interval in 12-tet, or a 3:2 ratio in JI.
FFT	Fast Fourier Transform, a clever implementation of the DFT.
FM	Frequency Modulation, when the frequency of a sine wave is changed, often sinusoidally.
formant	Resonances that may be thought of as fixed filters through which a variable excitation is passed.
fourth	A 500-cent interval in 12-tet, or a 4:3 ratio in JI.

GA	Genetic Algorithm, an optimization technique.
harmonic	Harmonic sounds have a fundamental frequency f and partials at integer multiples of f.
Hz	Hertz is a measure of frequency in cycles per second.
IAC	Interapplication ports that allow audio and MIDI data to be exchanged between applications.
inharmonic	The partials of an inharmonic sound are not at integer multiples of a single fundamental frequency.
JI	Just Intonation, the theory of musical intervals and scales based on small integer ratios.
JND	Just Noticeable Difference, smallest change that a listener can detect.
K	K means $2^{10} = 1024$. For example, a 16K FFT contains $16 \times 1024 = 16386$ samples.
ℓ_i	Loudness of the ith partial of a sound.
MIDI	Musical Interface for Digital Instruments, a communications protocol for electronic musical devices.
octave	Musical interval defined by the ratio 2:1.
partial	The partials (overtones) of a sound are the prominent sine wave components in the DFT representation.
periodic	A function, signal, or waveform $w(x)$ is periodic with period p if $w(x + p) = w(x)$ for all x.
RIW	Resampling with Identity Window, a technique for spectral mapping.
[S:]	Reference to the sound examples, see p. 399.
semitone	In 12-tet, an interval of 100 cents.
signal	When a sound is converted into digital form in a computer, it is called a signal.
sine wave	The "simplest" waveform is completely characterized by frequency, amplitude, and phase.
SMF	Standard MIDI File, a way of storing and exchanging MIDI data between computer platforms.
spectral mapping	Technique for manipulating the partials of a sound.
SPSA	Simultaneous Perturbation Stochastic Approximation, a technique of numerical optimization
steady state	The part of a sound that can be closely approximated by a periodic waveform.
-tet	Abbreviation for tone-equal-tempered. 12-tet is the standard Western keyboard tuning.
transient	That portion of a sound that cannot be closely approximated by a periodic signal.
[V:]	Reference to the video examples, see p. 411.
[W:]	Web references, see p. 413
waveform	Synonym for signal.
whole tone	In 12-tet, an interval of 200 cents.
xenharmonic	Strange musical "harmonies" not possible in 12-tet.
xentonal	Music with a surface appearance of tonality, but unlike anything possible in 12-tet.
$\oplus$	Pronounced *oh-plus*, this symbol indicates the "sum" of two intervals in the symbolic method of constructing spectra.

1

The Octave Is Dead . . . Long Live the Octave

1.1 A Challenge

The octave is the most consonant interval after the unison. A low C on the piano sounds "the same" as a high C. Scales "repeat" at octave intervals. These commonsense notions are found wherever music is discussed:

> The most basic musical interval is the octave, which occurs when the frequency of any tone is doubled or halved. Two tones an octave apart create a feeling of identity, or the duplication of a single pitch in a higher or lower register.[1]

Harry Olson[2] uses "pleasant" rather than "consonant":

> An interval between two sounds is their spacing in pitch or frequency... It has been found that the octave produces a pleasant sensation... It is an established fact that the most pleasing combination of two tones is one in which the frequency ratio is expressible by two integers neither of which is large.

W. A. Mathieu[3] discusses the octave far more poetically:

> The two sounds are the same and different. Same name, same "note" (whatever that is), but higher pitch. When a man sings nursery rhymes with a child, he is singing precisely the same song, but lower than the child. They are singing together, but singing apart. There is something easy in the harmony of two tones an octave apart - played either separately or together - but an octave transcends *easy*. There is a way in which the tones are identical.

[1] From [B: 66].
[2] [B: 123].
[3] [B: 104].

Arthur Benade[4] observes that the similarity between notes an octave apart has been enshrined in many of the world's languages:

> Musicians of all periods and all places have tended to agree that when they hear a tone having a repetition frequency double that of another one, the two are very nearly interchangeable. This similarity of a tone with its octave is so striking that in most languages both tones are given the same name.

Anthony Storr[5] is even more emphatic:

> The octave is an acoustic fact, expressible mathematically, which is not created by man. The composition of music requires that the octave be taken as the most basic relationship.

Given all this, the reader may be surprised (and perhaps a bit incredulous) to hear a tone that is distinctly dissonant when played in the interval of an octave, yet sounds nicely consonant when played at some other, nonoctave interval. This is exactly the demonstration provided in the first sound example[6] [S: 1] and repeated in the first video example[7] [V: 1]. The demonstration consists of only a handful of notes, as shown in Fig. 1.1.

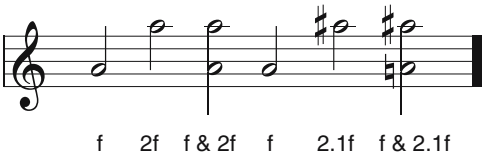


Fig. 1.1. In sound example [S: 1] and video example [V: 1], the timbre of the sound is constructed so that the octave between f and $2f$ is dissonant while the nonoctave f to $2.1f$ is consonant. Go listen to this example now.

A note is played (with a fundamental frequency $f = 450$ Hz[8]) followed by its octave (with fundamental at $2f = 900$ Hz). Individually, they sound normal enough, although perhaps somewhat "electronic" or bell-like in nature. But when played simultaneously, they clash in a startling dissonance. In the second phrase, the same note is played, followed by a note with fundamental at $2.1f = 945$ Hz (which falls just below the highly dissonant interval usually called the augmented octave or minor 9th). Amazingly, this second, nonoctave (and even microtonal) interval appears smooth and restful, even consonant; it has many

[4] [B: 12].
[5] [B: 184].
[6] Beginning on p. 399 is a listing of all sound examples (references to sound examples are prefaced with [S:]) along with instructions for accessing them with a computer.
[7] Beginning on p. 411 is a listing of all video examples (references to video examples are prefaced with [V:]) along with instructions for accessing them with a computer.
[8] *Hz* stands for *Hertz*, the unit of frequency. One Hertz equals one cycle per second.

of the characteristics usually associated with the octave. Such an interval is called a *pseudo-octave*.

Precise details of the construction of the sound used in this example are given later. For now, it is enough to recognize that the tonal makeup of the sound was carefully chosen *in conjunction with* the intervals used. Thus, the "trick" is to choose the spectrum or timbre of the sound (the tone quality) to match the tuning (the intervals desired).

As will become apparent, there is a relationship between the kinds of sounds made by Western instruments (i.e., harmonic[9] sounds) and the kinds of intervals (and hence scales) used in conventional Western tonal music. In particular, the 2:1 octave is important precisely because the first two partials of a harmonic sound have 2:1 ratios. Other kinds of sounds are most naturally played using other intervals, for example, the 2.1 pseudo-octave. Stranger still, there are inharmonic sounds that suggest no natural or obvious interval of repetition. Octave-based music is only one of a multitude of possible musics. As future chapters show, it is possible to make almost any interval reasonably consonant, or to make it wildly dissonant, by properly sculpting the spectrum of the sound.

Sound examples [S: 2] to [S: 5] are basically an extended version of this example, where you can better hear the clash of the dissonances and the odd timbral character associated with the inharmonic stretched sounds. The "same" simple piece is played four ways:

[S: 2] Harmonic sounds in 12-tet
[S: 3] Harmonic sounds in the 2.1 stretched scale
[S: 4] 2.1 stretched timbres in the 2.1 stretched scale
[S: 5] 2.1 stretched timbres in 12-tet

where *12-tet* is an abbreviation for the familiar 12-tone per octave equal tempered scale, and where the *stretched scale*, based on the 2.1 pseudo-octave, is designed specially for use with the stretched timbres. When the timbres and the scales are matched (as in [S: 2] and [S: 4]), there is contrast between consonance and dissonance as the chords change, and the piece has a sensible musical flow (although the timbral qualities in [S: 4] are decidedly unusual). When the timbres and scales do not match (as in [S: 3] and [S: 5]), the piece is uniformly dissonant. The difference between these two situations is not subtle, and it calls into question the meaning of basic terms like timbre, consonance, and dissonance. It calls into question the octave as the most consonant interval, and the kinds of harmony and musical theories based on that view. In order to make sense of these examples, *Tuning, Timbre, Spectrum, Scale* uses the notions of *sensory consonance* and *sensory dissonance*. These terms are carefully defined in Chap. 3 and are contrasted with other notions of consonance and dissonance in Chap. 5.

[9] Here *harmonic* is used in the technical sense of a sound with overtones composed exclusively of integer multiples of some audible fundamental.

1.2 A Dissonance Meter

Such shaping of spectra and scales requires that there be a convenient way to measure the dissonance of a given sound or interval. One of the key ideas underlying the sonic manipulations in *Tuning, Timbre, Spectrum, Scale* is the construction of a "dissonance meter." Don't worry—no soldering is required. The dissonance meter is a computer program that inputs a sound in digital form and outputs a number proportional to the (sensory) dissonance or consonance of the sound. For longer musical passages with many notes, the meter can be used to measure the dissonance within each specified time interval, for instance, within each measure or each beat. As the *challenging the octave* example shows, the dissonance meter must be sensitive to both the tuning (or pitch) of the sounds and to the spectrum (or timbre) of the tones.

Although such a device may seem frivolous at first glance, it has many real uses:

As an audio signal processing device: The dissonance meter is at the heart of a device that can automatically reduce the dissonance of a sound, while leaving its character more or less unchanged. This can also be reversed to create a sound that is more dissonant than the input. Combined, this provides a way to directly control the perceived dissonance of a sound.

Adaptive tuning of musical synthesizers: While monitoring the dissonance of the notes commanded by a performer, the meter can be used to adjust the tuning of the notes (microtonally) to minimize the dissonance of the passage. This is a concrete way of designing an adaptive or dynamic tuning.

Exploration of inharmonic sounds: The dissonance meter shows which intervals are most consonant (and which most dissonant) as a function of the spectrum of the instrument. As the *challenging the octave* example shows, unusual sounds can be profitably played in unusual intervals. The dissonance meter can concretely specify related intervals and spectra to find tunings most appropriate for a given timbre. This is a kind of map for the exploration of inharmonic musical spaces.

Exploration of "arbitrary" musical scales: Each timbre or spectrum has a set of intervals in which it sounds most consonant. Similarly, each set of intervals (each musical scale) has timbres with spectra that sound most consonant in that scale. The dissonance meter can help find timbres most appropriate for a given tuning.

Analysis of tonal music and performance: In tonal systems with harmonic instruments, the consonance and dissonance of a musical passage can often be read from the score because intervals within a given historical period have a known and relatively fixed degree of consonance and/or dissonance. But

performances may vary. A dissonance meter can be used to measure the actual dissonance of different performances of the same piece.

Analysis of nontonal and nonwestern music and performance: Sounds played in intervals radically different from those found in 12-tet have no standard or accepted dissonance value in standard music theory. As the dissonance meter can be applied to any sound at any interval, it can be used to help make musical sense of passages to which standard theories are inapplicable. For instance, it can be used to investigate nonwestern music such as the gamelan, and modern atonal music.

Historical musicology: Many historical composers wrote in musical scales (such as meantone, Pythagorean, Just, etc.) that are different from 12-tet, but they did not document their usage. By analyzing the choice of intervals, the dissonance meter can make an educated guess at likely scales using only the extant music. Chapter 11, on "Musicological Analysis," investigates possible scales used by Domenico Scarlatti.

As an intonation monitor: Two notes in unison are very consonant. When slightly out of tune, dissonances occur. The dissonance meter can be used to monitor the intonation of a singer or instrumentalist, and it may be useful as a training device.

The ability to measure dissonance is a crucial component in several kinds of audio devices and in certain methods of musical analysis. The idea that dissonance is a function of the timbre of the sound as well as the musical intervals also has important implications for the understanding of nonwestern musics, modern atonal and experimental compositions, and the design of electronic musical instruments.

1.3 New Perspectives

The dissonance curve plots how much sensory dissonance occurs at each interval, given the spectrum (or timbre) of a sound. Many common Western orchestral (and popular) instruments are primarily harmonic, that is, they have a spectrum that consists of a fundamental frequency along with partials (or overtones) at integer multiples of the fundamental. This spectrum can be used to draw a dissonance curve, and the minima of this curve occur at or near many of the steps of the Western scales. This suggests a relationship between the spectrum of the instruments and the scales in which they are played.

Nonwestern Musics

Many different scale systems have been and still are used throughout the world. In Indonesia, for instance, gamelans are tuned to five and seven-note

scales that are very different from 12-tet. The timbral quality of the (primarily metallophone) instruments is also very different from the harmonic instruments of the West. The dissonance curve for these metallophones have minima that occur at or near the scale steps used by the gamelans.[10] Similarly, in Thailand, there is a classical music tradition that uses wooden xylophone-like instruments called *renats* that play in (approximately) 7-tet. The dissonance curve for renat-like timbres have minima that occur near many of the steps of the traditional 7-tet Thai scale, as shown in Chap. 15. Thus, the musical scales of these nonwestern traditions are related to the inharmonic spectra of the instruments, and the idea of related spectra and scales is applicable cross culturally.

New Scales

Even in the West, the present 12-tet system is a fairly recent innovation, and many different scales have been used throughout history. Some systems, such as those used in the Indonesian gamelan, do not even repeat at octave intervals. Can *any* possible set of intervals or frequencies form a viable musical scale, assuming that the listener is willing to acclimate to the scale?

Some composers have viewed this as a musical challenge. Easley Blackwood's *Microtonal Etudes* might jokingly be called the "Ill-Tempered Synthesizer" because it explores all equal temperaments between 13 and 24. Thus, instead of 12 equal divisions of the octave, these pieces divide the octave into 13, 14, 15, and more equal parts. Ivor Darreg composed in many equal temperaments,[11] exclaiming

> the striking and characteristic moods of many tuning-systems will become the most powerful and compelling reason for exploring beyond 12-tone equal temperament. It is necessary to have more than one non-twelve-tone system before these moods can be heard and their significance appreciated.[12]

Others have explored nonequal divisions of the octave[13] and even various subdivisions of nonoctaves.[14] It is clearly possible to make music in a large variety of tunings. Such music is called *xenharmonic*,[15] strange "harmonies" unlike anything possible in 12-tet.

The intervals that are most consonant for harmonic sounds are made from small integer ratios such as the octave (2:1), the fifth (3:2), and the fourth (4:3). These simple integer ratio intervals are called *just* intervals, and they collectively form scales known as *just intonation* scales. Many of the just

[10] See Chap. 10 "The Gamelan" for details and caveats.
[11] [D: 10].
[12] From [B: 36], No. 5.
[13] For instance, Vallotti, Kirchenberg, and Partch.
[14] For instance, Carlos [B: 23], Mathews and Pierce [B: 102], and McLaren [B: 108].
[15] Coined by Darreg [B: 36], from the Greek *xenos* for strange or foreign.

intervals occur close to (but not exactly at[16]) steps of the 12-tet scale, which can be viewed as an acceptable approximation to these just intervals. Steps of the 19-tet scale also approximate many of the just intervals, but the 10-tet scale steps do not. This suggests why, for instance, it is easy to play in 19-tet and hard to play in 10-tet using harmonic tones—there are many consonant intervals in 19-tet but few in 10-tet.

New Sounds

The *challenging the octave* demonstration shows that certain unusual intervals can be consonant when played with certain kinds of unusual sounds. Is it possible to make *any* interval consonant by properly manipulating the sound quality? For instance, is it possible to choose the spectral character so that many of the 10-tet intervals became consonant? Would it then be "easy" to play in 10-tet? The answer is "yes," and part of this book is dedicated to exploring ways of manipulating the spectrum in an appropriate manner.

Although Western music relies heavily on harmonic sounds, these are only one of a multitude of kinds of sound. Modern synthesizers can easily generate inharmonic sounds and transport us into unexplored musical realms. The spectrum/scale connection provides a guideline for exploration by specifying the intervals in which the sounds can be played most consonantly or by specifying the sounds in which the intervals can be played most consonantly. Thus, the methods allow the composer to systematically specify the amount of consonance or dissonance. The composer has a new and powerful method of control over the music.

Consider a fixed scale in which all intervals are just. No such scale can be modulated through all the keys. No such scale can play all the consonant chords even in a single key. (These are arithmetic impossibilities, and a concrete example is provided on p. 159.) But using the ideas of sensory consonance, it is possible to adapt the pitches of the notes dynamically. For harmonic tones, this is equivalent to playing in simple integer (just) ratios, but allows modulation to any key, thus bypassing this ancient problem. Although previous theorists had proposed that such dynamic tunings might be possible,[17] this is the first concrete method that can be applied to any chord in any musical setting. *It is possible to have your just intonation and to modulate, too!* Moreover, the adaptive tuning method is not restricted to harmonic tones, and so it provides a way to "automatically" play in the related scale (the scale consisting of the most consonant intervals, given the spectral character of the sound).

[16] Table 6.1 on p. 101 shows how close these approximations are.
[17] See Polansky [B: 142] and Waage [B: 202].

New "Music Theories"

When working in an unfamiliar system, the composer cannot rely on musical intuition developed through years of practice. In 10-tet, for instance, there are no intervals near the familiar fifths or thirds, and it is not obvious what intervals and chords make musical sense. The ideas of sensory consonance can be used to find the most consonant chords, as well as the most consonant intervals (as always, sensory consonance is a function of the intervals and of the spectrum/timbre of the sound), and so it can provide a kind of sensory map for the exploration of new tunings and new timbres. Chapter 14 develops a new music theory for 10-tet. The "neutral third" chord is introduced along with the "circle of thirds" (which is somewhat analogous to the familiar circle of fifths in 12-tet). This can be viewed as a prototype of the kinds of theoretical constructs that are possible using the sensory consonance approach, and pieces are included on the CD to demonstrate that the predictions of the model are valid in realistic musical situations.

Unlike most theories of music, this one does not seek (primarily) to explain a body of existing musical practice. Rather, like a good scientific theory, it makes concrete predictions that can be readily verified or falsified. These predictions involve how (inharmonic) sounds combine, how spectra and scales interact, and how dissonance varies as a function of both interval and spectrum. The enclosed CD provides examples so that you can verify for yourself that the predictions correspond to perceptual reality.

Tuning and spectrum theories are independent of musical style; they are no more "for" classical music than they are "for" jazz or pop. It would be naive to suggest that complex musical properties such as style can be measured in terms of a simple sensory criterion. Even in the realm of harmony (and ignoring musically essential aspects such as melody and rhythm), sensory consonance is only part of the story. A harmonic progression that was uniformly consonant would be tedious; harmonic interest arises from a complex interplay of restlessness and restfulness,[18] of tension and resolution. It is easy to increase the sensory dissonance, and hence the restlessness, by playing more notes (try slamming your arm on the keyboard). But it is not always as easy to increase the sensory consonance and hence the restfulness. By playing sounds in their related scales, it is possible to obtain the greatest contrast between consonance and dissonance for a given sound palette.

1.4 Overview

While introducing the appropriate psychoacoustic jargon, Chap. 2 (the "Science of Sound") draws attention to the important distinction between what

[18] Alternative definitions of dissonance and consonance are discussed at length in Chap. 5.

we perceive and what is really (measurably) there. Any kind of "perceptually intelligent" musical device must exploit the measurable in order to extract information from the environment, and it must then shape the sound based on the perceptual requirements of the listener. Chapter 3 looks carefully at the case of two simultaneously sounding sine waves, which is the simplest situation in which sensory dissonances occur.

Chapter 4 reviews several of the common organizing principles behind the creation of musical scales, and it builds a library of historical and modern scales that will be used throughout the book as examples.

Chapter 5 gives an overview of the many diverse meanings that the words "consonance" and "dissonance" have enjoyed throughout the centuries. The relatively recent notion of sensory consonance is then adopted for use throughout the remainder of the book primarily because it can be readily measured and quantified.

Chapter 6 introduces the idea of a *dissonance curve* that displays (for a sound with a given spectrum) the sensory consonance and dissonance of all intervals. This leads to the definition of a *related* spectrum and scale, a sound for which the most consonant intervals occur at precisely the scale steps. Two complementary questions are posed. Given a spectrum, what is the related scale? Given a scale, what is a related spectrum? The second, more difficult question is addressed at length in Chap. 12, and Chap. 7 ("A Bell, A Rock, A Crystal") gives three detailed examples of how related spectra and scales can be exploited in musical contexts. This is primarily interesting from a compositional point of view.

Chapter 8 shows how the ideas of sensory consonance can be exploited to create a method of adaptive tuning, and it provides several examples of "what to expect" from such an algorithm. Chapter 9 highlights three compositions in adaptive tuning and discusses compositional techniques and tradeoffs. Musical compositions and examples are provided on the accompanying CD.

The remaining chapters can be read in any order. Chapter 10 shows how the pelog and slendro scales of the Indonesian gamelan are correlated with the spectra of the metallophones on which they are played. Similarly, Chap. 15 shows how the scales of Thai classical music are related to the spectra of the Thai instruments.

Chapter 11 explores applications in musicology. The *dissonance score* can be used to compare different performances of the same piece, or to examine the use of consonances and dissonances in unscored and nonwestern music. An application to historical musicology shows how the tuning preferences of Domenico Scarlatti can be investigated using only his extant scores.

Chapter 14 explores one possible alternative musical universe, that of 10-tet. This should only be considered a preliminary foray into what promises to be a huge undertaking—codifying and systematizing music theories for non-12-tet. Although it is probably impossible to find a "new" chord in 12-tet, it is impossible to play in n-tet without creating new harmonies, new chordal structures, and new kinds of musical passages.

Chapters 12 and 13 are the most technically involved. They show how to specify spectra for a given tuning, and how to create rich and complex sounds with the specified spectral content.

The final chapter sums up the ideas in *Tuning, Timbre, Spectrum, Scale* as exploiting a single perceptual measure (that of sensory consonance) and applying it to musical theory, practice, and sound design. As we expand the palette of timbres we play, we will naturally begin to play in new intervals and new tunings.

2

The Science of Sound

"Sound" as a physical phenomenon and "sound" as a perceptual phenomena are not the same thing. Definitions and results from acoustics are compared and contrasted to the appropriate definitions and results from perception research and psychology. Auditory perceptions such as loudness, pitch, and timbre can often be correlated with physically measurable properties of the sound wave.

2.1 What Is Sound?

If a tree falls in the forest and no one is near, does it make any sound? Understanding the different ways that people talk about sound can help get to the heart of this conundrum. One definition[1] describes the wave nature of sound:

> Vibrations transmitted through an elastic material or a solid, liquid, or gas, with frequencies in the approximate range of 20 to 20,000 hertz.

Thus, physicists and engineers use "sound" to mean a pressure wave propagating through the air, something that can be readily measured, digitized into a computer, and analyzed. A second definition focuses on perceptual aspects:

> The sensation stimulated in the organs of hearing by such vibrations in the air or other medium.

Psychologists (and others) use "sound" to refer to a perception that occurs inside the ear, something that is notoriously hard to quantify.

Does the tree falling alone in the wilderness make sound? Under the first definition, the answer is "yes" because it will inevitably cause vibrations in the air. Using the second definition, however, the answer is "no" because there are no organs of hearing present to be stimulated. Thus, the physicist says yes, the psychologist says no, and the pundits proclaim a paradox. The source of the confusion is that "sound" is used in two different senses. Drawing such distinctions is more than just a way to resolve ancient puzzles, it is also a way to avoid similar confusions that can arise when discussing auditory phenomena.

[1] from the American Heritage Dictionary.

Physical attributes of a signal such as frequency and amplitude must be kept distinct from perceptual correlates such as pitch and loudness.[2] The physical attributes are measurable properties of the signal whereas the perceptual correlates are inside the mind of the listener. To the physicist, sound is a pressure wave that propagates through an elastic medium (i.e., the air). Molecules of air are alternately bunched together and then spread apart in a rapid oscillation that ultimately bumps up against the eardrum. When the eardrum wiggles, signals are sent to the brain, causing "sound" in the psychologist's sense.

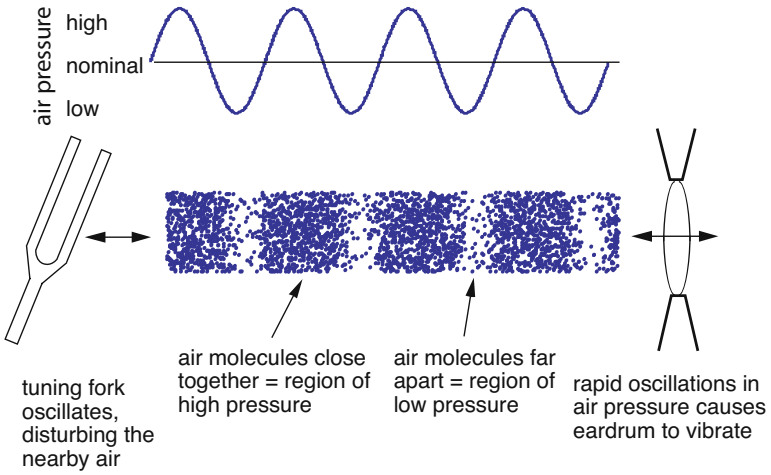


Fig. 2.1. Sound as a pressure wave. The peaks represent times when air molecules are clustered, causing higher pressure. The valleys represent times when the air density (and hence the pressure) is lower than nominal. The wave pushes against the eardrum in times of high pressure, and pulls (like a slight vacuum) during times of low pressure, causing the drum to vibrate. These vibrations are perceived as sound.

Sound waves can be pictured as graphs such as in Fig. 2.1, where high-pressure regions are shown above the horizontal line, and low-pressure regions are shown below. This particular waveshape, called a *sine wave*, can be characterized by three mathematical quantities: frequency, amplitude, and phase. The frequency of the wave is the number of complete oscillations that occur in one second. Thus, a sine wave with a frequency of 100 Hz (short for Hertz, after the German physicist Heinrich Rudolph Hertz) oscillates 100 times each second. In the corresponding sound wave, the air molecules bounce back and forth 100 times each second.

[2] The ear actually responds to sound pressure, which is usually measured in decibels.

The human auditory system (the ear, for short) perceives the frequency of a sine wave as its pitch, with higher frequencies corresponding to higher pitches. The amplitude of the wave is given by the difference between the highest and lowest pressures attained. As the ear reacts to variations in pressure, waves with higher amplitudes are generally perceived as louder, whereas waves with lower amplitudes are heard as softer. The phase of the sine wave essentially specifies when the wave starts, with respect to some arbitrarily given starting time. In most circumstances, the ear cannot determine the phase of a sine wave just by listening.

Thus, a sine wave is characterized by three measurable quantities, two of which are readily perceptible. This does not, however, answer the question of what a sine wave *sounds like*. Indeed, no amount of talk will do. Sine waves have been variously described as pure, tonal, clean, simple, clear, like a tuning fork, like a theremin, electronic, and flute-like. To refresh your memory, the first few seconds of sound example [S: 8] are purely sinusoidal.

2.2 What Is a Spectrum?

Individual sine waves have limited musical value. However, combinations of sine waves can be used to describe, analyze, and synthesize almost any possible sound. The physicist's notion of the spectrum of a waveform correlates well with the perceptual notion of the timbre of a sound.

2.2.1 Prisms, Fourier Transforms, and Ears

As sound (in the physical sense) is a wave, it has many properties that are analogous to the wave properties of light. Think of a prism, which bends each color through a different angle and so decomposes sunlight into a family of colored beams. Each beam contains a "pure color," a wave of a single frequency, amplitude, and phase.[3] Similarly, complex sound waves can be decomposed into a family of simple sine waves, each of which is characterized by its frequency, amplitude, and phase. These are called the *partials*, or the *overtones* of the sound, and the collection of all the partials is called the *spectrum*. Figure 2.2 depicts the *Fourier transform* in its role as a "sound prism."

This prism effect for sound waves is achieved by performing a spectral analysis, which is most commonly implemented in a computer by running a program called the Discrete Fourier Transform (DFT) or the more efficient Fast Fourier Transform (FFT). Standard versions of the DFT and/or the FFT are readily available in audio processing software and in numerical packages (such as `Matlab` and Mathematica) that can manipulate sound data files.

[3] For light, frequency corresponds to color, and amplitude to intensity. Like the ear, the eye is predominantly blind to the phase.

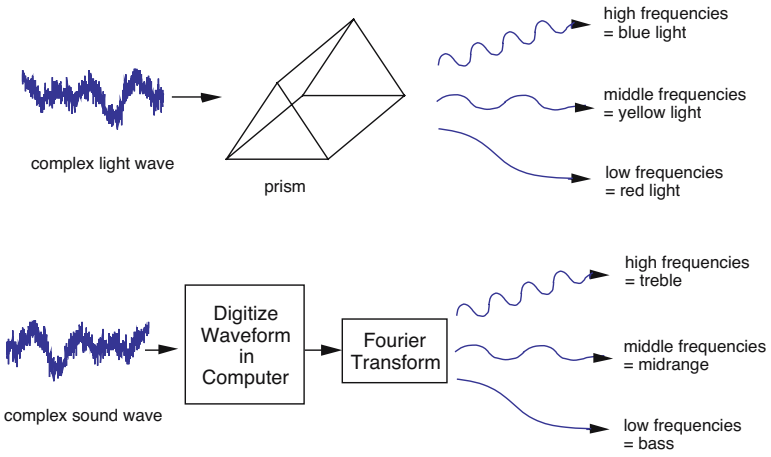


Fig. 2.2. Just as a prism separates light into its simple constituent elements (the colors of the rainbow), the Fourier Transform separates sound waves into simpler sine waves in the low (bass), middle (midrange), and high (treble) frequencies. Similarly, the auditory system transforms a pressure wave into a spatial array that corresponds to the various frequencies contained in the wave, as shown in Fig. 2.4.

The spectrum gives important information about the makeup of a sound. For example, Fig. 2.3 shows a small portion of each of three sine waves:

(a) With a frequency of 100 Hz and an amplitude of 1.2 (the solid line)
(b) With a frequency of 200 Hz and an amplitude of 1.0 (plotted with dashes)
(c) With a frequency of 200 Hz and an amplitude of 1.0, but shifted in phase from (b) (plotted in bold dashes)

such as might be generated by a pair of tuning forks or an electronic tuner playing the G below middle C and the G an octave below that.[4] When (a) and (b) are sounded together (mathematically, the amplitudes are added together point by point), the result is the (slightly more) complex wave shown in part (d). Similarly, (a) and (c) added together give (e). When (d) is Fourier transformed, the result is the graph (f) that shows frequency on the horizontal axis and the magnitude of the waves displayed on the vertical axis. Such magnitude/frequency graphs are called the *spectrum*[5] of the waveform, and they show what the sound is made of. In this case, we know that the sound is

[4] Actually, the G's should have frequencies of 98 and 196, but 100 and 200 make all of the numbers easier to follow.

[5] This is more properly called the *magnitude spectrum*. The *phase spectrum* is ignored in this discussion because it does not correspond well to the human perceptual apparatus.

composed of two sine waves at frequencies 100 and 200, and indeed there are two peaks in (f) corresponding to these frequencies. Moreover, we know that the amplitude of the 100-Hz sinusoid is 20% larger than the amplitude of the 200-Hz sine, and this is reflected in the graph by the size of the peaks. Thus, the spectrum (f) decomposes the waveform (d) into its constituent sine wave components.

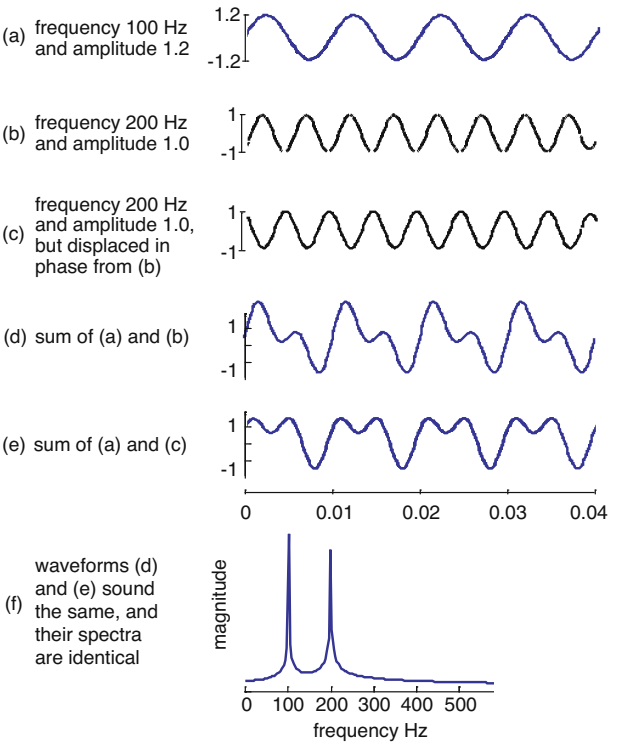


Fig. 2.3. Spectrum of a sound consisting of two sine waves.

This idea of breaking up a complex sound into its sinusoidal elements is important because the ear functions as a kind of "biological" spectrum analyzer. That is, when sound waves impinge on the ear, we hear a sound (in the second, perceptual sense of the word) that is a direct result of the spectrum, and it is only indirectly a result of the waveform. For example, the waveform in part (d) looks very different from the waveform in part (e), but they sound essentially the same. Analogously, the spectrum of waveform (d) and the spectrum of waveform (e) are identical (because they have been built from sine waves with the same frequencies and amplitudes). Thus, the spectral representation captures perceptual aspects of a sound that the waveform does

not. Said another way, the spectrum (f) is more meaningful to the ear than are the waveforms (d) and (e).

A nontrivial but interesting exercise in mathematics shows that any periodic signal can be broken apart into a sum of sine waves with frequencies that are integer multiples of some fundamental frequency. The spectrum is thus ideal for representing periodic waveforms. But no real sound is truly periodic, if only because it must have a beginning and an end; at best it may closely approximate a periodic signal for a long, but finite, time. Hence, the spectrum can closely, but not exactly, represent a musical sound. Much of this chapter is devoted to discovering how close such a representation can really be.

Figure 2.4 shows a drastically simplified view of the auditory system. Sound or pressure waves, when in close proximity to the eardrum, cause it to vibrate. These oscillations are translated to the *oval window* through a mechanical linkage consisting of three small bones. The oval window is mounted at one end of the cochlea, which is a conical tube that is curled up like a snail shell (although it is straightened out in the illustration). The cochlea is filled with fluid, and it is divided into two chambers lengthwise by a thin layer of pliable tissue called the basilar membrane. The motion of the fluid rocks the membrane. The region nearest the oval window responds primarily to high frequencies, and the far end responds mostly to low frequencies. Tiny hair-shaped neurons sit on the basilar membrane, sending messages toward the brain when they are jostled.

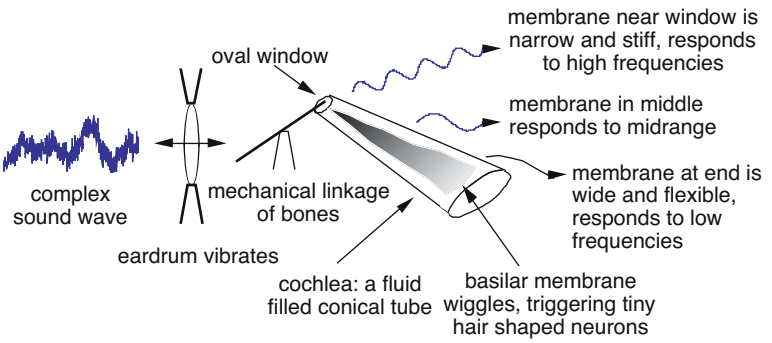


Fig. 2.4. The auditory system as a biological spectrum analyzer that transforms a pressure wave into a frequency selective spatial array.

Thus, the ear takes in a sound wave, like that in Fig. 2.3 (d) or (e), and sends a coded representation to the brain that is similar to a spectral analysis, as in (f). The conceptual similarities between the Fourier transform and the auditory system show why the idea of the spectrum of a sound is so powerful; the Fourier transform is a mathematical tool that is closely related to our perceptual mechanism. This analogy between the perception of timbre and

the Fourier spectrum was first posited by Georg Ohm in 1843 (see [B: 147]), and it has driven much of the acoustics research of the past century and a half.

2.2.2 Spectral Analysis: Examples

The example in the previous section was contrived because we constructed the signal from two sine waves, only to "discover" that the Fourier transform contained the frequencies of those same two sine waves. It is time to explore more realistic sounds: the pluck of a guitar and the strike of a metal bar. In both cases, it will be possible to give both a physical and an auditory meaning to the spectrum.

Guitar Pluck: Theory

Guitar strings are flexible and lightweight, and they are held firmly in place at both ends, under considerable tension. When plucked, the string vibrates in a far more complex and interesting way than the simple sine wave oscillations of a tuning fork or an electronic tuner. Figure 2.5 shows the first 3/4 second of the open G string of my Martin acoustic guitar. Observe that the waveform is initially very complex, bouncing up and down rapidly. As time passes, the oscillations die away and the gyrations simplify. Although it may appear that almost anything could be happening, the string can vibrate freely only at certain frequencies because of its physical constraints.

For sustained oscillations, a complete half cycle of the wave must fit exactly inside the length of the string; otherwise, the string would have to move up and down where it is rigidly attached to the bridge (or nut) of the guitar. This is a tug of war the string inevitably loses, because the bridge and nut are far more massive than the string. Thus, all oscillations except those at certain privileged frequencies are rapidly attenuated.

Figure 2.6 shows the fundamental and the first few modes of vibration for a theoretically ideal string. If half a period corresponds to the fundamental frequency f, then a whole period at frequency $2f$ also fits exactly into the length of the string. This more rapid mode of vibration is called the second partial. Similarly, a period and a half at frequency $3f$ fits exactly, and it is called the third partial. Such a spectrum, in which all frequencies of vibration are integer multiples of some fundamental f, is called *harmonic*, and the frequencies of oscillation are called the *natural modes of vibration* or *resonant frequencies* of the string. As every partial repeats exactly within the period of the fundamental, harmonic spectra correspond to periodic waveforms.

Compare the spectrum of the real string in Fig. 2.5 with the idealized spectrum in Fig. 2.6. Despite the complex appearance of the waveform, the guitar sound is primarily harmonic. Over 20 partials are clearly visible at roughly equal distances from each other, with frequencies at (approximately)

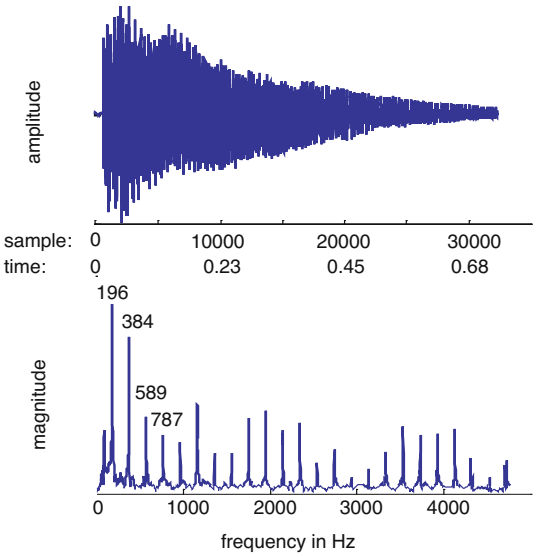


Fig. 2.5. Waveform of a guitar pluck and its spectrum. The top figure shows the first 3/4 second (32,000 samples) of the pluck of the G string of an acoustic guitar. The spectrum shows the fundamental at 196 Hz, and near integer harmonics at 384, 589, 787, ...

integer multiples of the fundamental, which in this case happens to be 196 Hz.

There are also some important differences between the real and the idealized spectra. Although the idealized spectrum is empty between the various partials, the real spectrum has some low level energy at almost every frequency. There are two major sources of this: noise and artifacts. The noise might be caused by pick noise, finger squeaks, or other aspects of the musical performance. It might be ambient audio noise from the studio, or electronic noise from the recording equipment. Indeed, the small peak below the first partial is suspiciously close to 60 Hz, the frequency of line current in the United States.

Artifacts are best described by referring back to Fig. 2.3. Even though these were pure sine waves generated by computer, and are essentially exact, the spectrum still has a significant nonzero magnitude at frequencies other than those of the two sine waves. This is because the sine waves are of finite duration, whereas an idealized spectrum (as in Fig. 2.6) assumes an infinite duration signal. This smearing of the frequencies in the signal is a direct result of the periodicity assumption inherent in the use of Fourier techniques. Artifacts and implementation details are discussed at length in Appendix C.

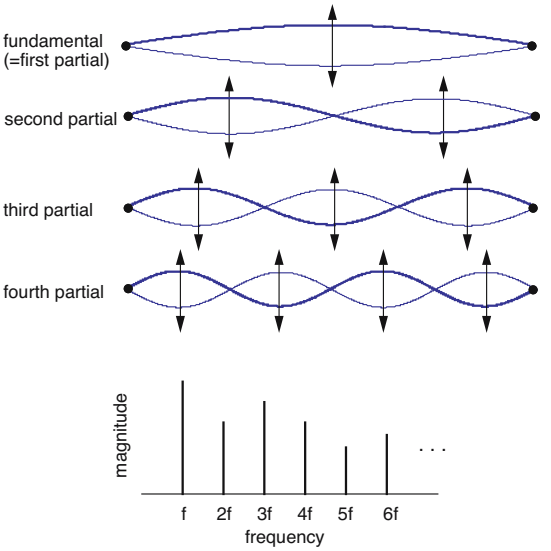


Fig. 2.6. Vibrations of an ideal string and its spectrum. Because the string is fixed at both ends, it can only sustain oscillations when a half period fits exactly into its length. Thus, if the fundamental occurs at frequency f, the second partial must be at $2f$, the third at $3f$, etc., as shown in the spectrum, which plots frequency verses magnitude.

Guitar Pluck: Experiment

Surely you didn't think you could read a whole chapter called the "Science of Sound" without having to experiment? You will need a guitar (preferably acoustic) and a reasonably quiet room.

Play one of the open strings that is in the low end of your vocal range (the A string works well for me) and let the sound die away. Hold your mouth right up to the sound hole, and sing "ah" loudly, at the same pitch as the string. Then listen. You will hear the string "singing" back at you quietly. This phenomenon is called *resonance* or *sympathetic vibration*. The pushing and pulling of the air molecules of the pressure wave set in motion by your voice excites the string, just as repetitive pushes of a child on a playground swing causes larger and larger oscillations. When you stop pushing, the child continues to bob up and down. Similarly, the string continues to vibrate after you have stopped singing.

Now sing the note an octave above (if you cannot do this by ear, play at the twelfth fret, and use this pitch to sing into the open string). Again you will hear the string answer, this time at the octave. Now try again, singing the fifth (which can be found at the seventh fret). This time the string responds, not at the fifth, but at the fifth plus an octave. The string seems to have suddenly developed a will of its own, refusing to sing the fifth, and instead jumping up

an octave. If you now sing at the octave plus fifth, the string resonates back at the octave plus fifth. But no amount of cajoling can convince it to sing that fifth in the lower octave. Try it. What about other notes? Making sure to damp all strings but the chosen one, sing a major second (two frets up). Now, no matter how strongly you sing, the string refuses to answer at all. Try other intervals. Can you get any thirds to sound?

To understand this cranky behavior, refer back to Fig. 2.6. The pitch of the string occurs at the fundamental frequency, and it is happy to vibrate at this frequency when you sing. Similarly, the octave is at exactly the second partial, and again the string is willing to sound. When you sing a major second, its frequency does not line up with any of the partials. Try pushing a playground swing at a rate at which it does not want to go—you will work very hard for very little result. Similarly, the string will not sustain oscillations far from its natural modes of vibration.

The explanation for the behavior of the guitar when singing the fifth is more subtle. Resonance occurs when the driving force (your singing) occurs at or near the frequencies of the natural modes of vibration of the string (the partials shown in Fig. 2.6). Your voice, however, is not a pure sine wave (at least, mine sure is not). Voices tend to be fairly rich in overtones, and the second partial of your voice coincides with the third partial of the string. It is this coincidence of frequencies that drives the string to resonate. By listening to the string, we have discovered something about your voice.

This is similar to the way Helmholtz [B: 71] determined the spectral content of sounds without access to computers and Fourier transforms. He placed tuning forks or bottle resonators (instead of strings) near the sound to be analyzed. Those that resonated corresponded to partials of the sound. In this way, he was able to build a fairly accurate picture of the nature of sound and of the hearing process.[6]

Sympathetic vibrations provide a way to hear the partials of a guitar string,[7] showing that they *can* vibrate in any of the modes suggested by Fig. 2.6. But *do* they actually vibrate in these modes when played normally? The next simple experiment demonstrates that strings tend to vibrate in many of the modes simultaneously.

[6] Although many of the details of Helmholtz's theories have been superseded, his book remains inspirational and an excellent introduction to the science of acoustics.

[7] For those without a guitar who are feeling left out, it is possible to hear sympathetic vibrations on a piano, too. For instance, press the middle C key slowly so that the hammer does not strike the string. While holding this key down (so that the damper remains raised), strike the C an octave below, and then lift up your finger so as to damp it out. Although the lower C string is now silent, middle C is now vibrating softly–the second partial of the lower note has excited the fundamental of the middle C. Observe that playing a low B will not excite such resonances in the middle C string.

Grab your guitar and pluck an open string, say the A string. Then, quickly while the note is still sounding, touch your finger lightly to the string directly above the twelfth fret.[8] You should hear the low A die away, leaving the A an octave above still sounding. With a little practice you can make this transition reliably. To understand this octave jump, refer again to Fig. 2.6. When vibrating at the fundamental frequency, the string makes its largest movement in the center. This point of maximum motion is called an *antinode* for the vibrational mode. Touching the midpoint of the string (at the twelfth fret) damps out this oscillation right away, because the finger is far more massive than the string. On the other hand, the second partial has a fixed point (called a *node*) right in the middle. It does not need to move up and down at the midpoint at all, but rather has antinodes at 1/4 and 3/4 of the length of the string. Consequently, its vibrations are (more or less) unaffected by the light touch of the finger, and it continues to sound even though the fundamental has been silenced.

The fact that the second partial persists after touching the string shows that the string must have been vibrating in (at least) the first and second modes. In fact, strings usually vibrate in many modes simultaneously, and this is easy to verify by selectively damping out various partials. For instance, by touching the string immediately above the seventh fret (1/3 of the length of the string), both the first and second partials are immediately silenced, leaving the third partial (at a frequency of three times the fundamental, the E an octave and a fifth above the fundamental A) as the most prominent sound. The fifth fret is 1/4 of the length of the string. Touching here removes the first three partials and leaves the fourth, two octaves above the fundamental, as the apparent pitch. To bring out the fifth harmonic, touch at either the 1/5 (just below the fourth fret) or at the 2/5 (near the ninth fret) points. This gives a note just a little flat of a major third, two octaves above the fundamental.

Table 2.1 shows the first 16 partials of the A string of the guitar. The frequency of each partial is listed, along with the nearest note of the standard 12-tone equal-tempered scale and its frequency. The first several coincide very closely, but the correspondence deteriorates for higher partials. The seventh partial is noticeably flat of the nearest scale tone, and above the ninth partial, there is little resemblance. With a bit of practice, it is possible to bring out the sound of many of the lower partials. Guitarists call this technique "playing the harmonics" of the string, although the preferred method begins with the finger resting lightly on the string and pulls it away as the string is plucked. As suggested by the previous discussion, it is most common to play harmonics at the twelfth, seventh, and fifth frets, which correspond to the second, third, and fourth partials, although others are feasible.

[8] Hints: Just touch the string delicately. Do not press it down onto the fretboard. Also, position the finger immediately over the fret bar, rather than over the space between the eleventh and twelfth frets where you would normally finger a note.

Table 2.1. The first 16 partials of the A string of a guitar with fundamental at 110 Hz. Many of the partials lie near notes of the standard equal-tempered scale, but the correspondence grows worse for higher partial numbers.

Partial Number	Frequency of Partial	Name of Nearest Note	Frequency of Nearest Note
1	110	A	110
2	220	A	220
3	330	E	330
4	440	A	440
5	550	$C\sharp$	554
6	660	E	659
7	770	G	784
8	880	A	880
9	990	B	988
10	1100	$C\sharp$	1109
11	1210	$D\sharp$	1245
12	1320	E	1318
13	1430	F	1397
14	1540	G	1568
15	1650	$G\sharp$	1661
16	1760	A	1760

As any guitarist knows, the tone of the instrument depends greatly on where the picking is done. Exciting the string in different places emphasizes different sets of characteristic frequencies. Plucking the string in the middle tends to bring out the fundamental and other odd-numbered harmonics (can you tell why?) while plucking near the ends tends to emphasize higher harmonics. Similarly, a pickup placed in the middle of the string tends to "hear" and amplify more of the fundamental (which has its antinode in the middle), and a pickup placed near the end of the string emphasizes the higher harmonics and has a sharper, more trebly tone.

Thus, guitars both can and do vibrate in many modes simultaneously, and these vibrations occur at frequencies dictated by the physical geometry of the string. We have seen two different methods of experimentally finding these frequencies: excitation via an external source (singing into the guitar) and selective damping (playing the harmonics). Of course, both of these methods are somewhat primitive, but they do show that the spectrum (a plot of the frequencies of the partials, and their magnitudes) is a real thing, which corresponds well with physical reality. With the ready availability of computers, the Fourier transform is easy to use. It is more precise, but fundamentally it tells nothing more than could be discovered using other nonmathematical (and more intuitive) ways.

A Metal Bar

It is not just strings that vibrate with characteristic frequencies. Every physical object tends to resonate at particular frequencies. For objects other than strings, however, these characteristic frequencies are often not harmonically related.

One of the simplest examples is a uniform metal bar as used in a glockenspiel or a wind chime.[9] When the bar is struck, it bends and vibrates, exciting the air and making sound. Figure 2.7 shows the first 3/4 second of the waveform of a bar and the corresponding spectrum. As usual, the waveform depicts the envelope of the sound, indicating how the amplitude evolves over time. The spectrum shows clearly what the sound is made of: four prominent partials and some high-frequency junk. The partials are at 526, 1413, 2689, and 4267 Hz. Considering the first partial as the fundamental at $f = 526$ Hz, this is f, $2.68f$, $5.11f$, and $8.11f$, which is certainly not a harmonic relationship; that is, the frequencies are not integer multiples of any audible fundamental. For bars of different lengths, the value of f changes, but the relationship between frequencies of the partials remains (roughly) the same.

The spectrum of the ideal string was explained physically as due to the requirement that it be fixed at both ends, which implied that the period of all sustained vibrations had to fit evenly into the length of the string. The metal bar is free at both ends, and hence, there is no such constraint. Instead the movement is characterized by bending modes that specify how the bar will vibrate once it is set into motion. The first three of these modes are depicted in Fig. 2.8, which differ significantly from the mode shapes of the string depicted in Fig. 2.6. Theorists have been able to write down and solve the equations that describe this kind of motion.[10] For an ideal metal bar, if the fundamental occurs at frequency f, the second partial will be at $2.76f$, the third at $5.4f$, and the fourth at $8.93f$. This is close to the measured spectrum of the bar of Fig. 2.7. The discrepancies are likely caused by small nonuniformities in the composition of the bar or to small deviations in the height or width of the bar. The high-frequency junk is most likely caused by impact noise, the sound of the stick hitting the bar, which is not included in the theoretical calculations.

As with the string, it is possible to discover these partials yourself. Find a cylindrical wind chime, a length of pipe, or a metal extension hose from a vacuum cleaner. Hold the bar (or pipe) at roughly 2/9 of its length, tap it, and listen closely. How many partials can you hear? If you hold it in the middle and tap, then the fundamental is attenuated and the pitch jumps up to the second partial—well over an octave away (to see why, refer again to Fig. 2.8). Now, keeping the sound of the second partial clearly in mind, hold

[9] Even though wind chimes are often built from cylindrical tubes, the primary modes of vibration are like those of a metal bar. Vibrations of the air column inside the tube are not generally loud enough to hear.

[10] See Fletcher and Rossing's *Physics of Musical Instruments* for an amazingly detailed presentation.

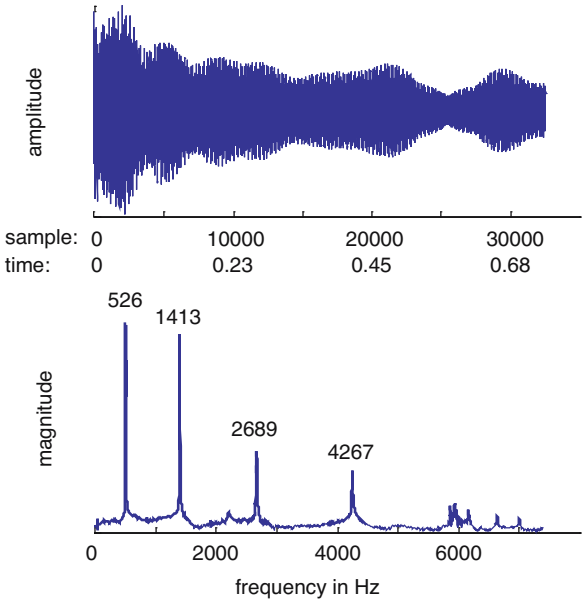


Fig. 2.7. Waveform of the strike of a metal bar and the corresponding spectrum. The top figure shows the first 3/4 second (32,000 samples) of the waveform in time. The spectrum shows four prominent partials.

and strike the pipe again at the 2/9 point. You will hear the fundamental, of course, but if you listen carefully, you can still hear the second partial. By selectively muting the various partials, you can bring the sound of many of the lower partials to the fore. By listening carefully, you can then continue to hear them even when they are mixed in with all the others.

As with the string, different characteristic frequencies can be emphasized by striking the bar at different locations. Typically, this will not change the locations of the partials, but it will change their relative amplitudes and, hence, the tone quality of the instrument. Observe the technique of a conga drummer. By tapping in different places, the drummer changes the tone dramatically. Also, by pressing a free hand against the drumhead, certain partials can be selectively damped, again manipulating the timbre.

The guitar string and the metal bar are only two of a nearly infinite number of possible sound-making devices. The (approximately) harmonic vibrations of the string are also characteristic of many other musical instruments. For instance, when air oscillates in a tube, its motion is constrained in much the same way that the string is constrained by its fixed ends. At the closed end of a tube, the flow of air must be zero, whereas at an open end, the pressure must

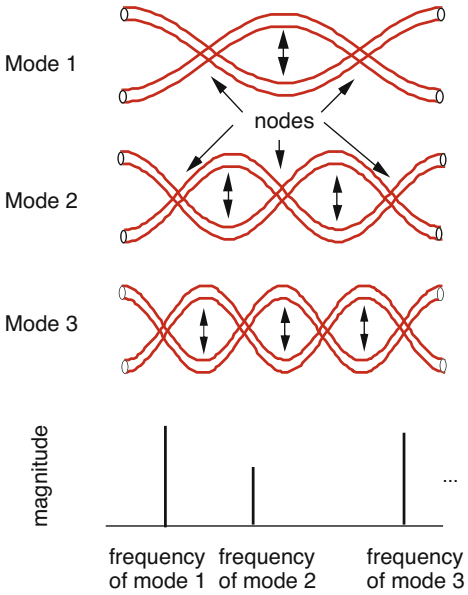


Fig. 2.8. The first three bending modes of an ideal metal bar and its spectrum. The size of the motion is proportional to the amplitude of the sound, and the rate of oscillation determines the frequency. As usual, the spectrum shows the frequencies of the partials on the horizontal axis and their magnitude on the vertical axis. Nodes are stationary points for particular modes of vibration. The figures are not to scale (the size of the motion is exaggerated with respect to the length and diameter of the bars).

drop to zero.[11] Thus instruments such as the flute, clarinet, trumpet, and so on, all have spectra that are primarily harmonic. In contrast, most percussion instruments such as drums, marimbas, kalimbas, cymbals, gongs, and so on, have spectra that are inharmonic. Musical practice generally incorporates both kinds of instruments.

Analytic vs. Holistic Listening: Tonal Fusion

Almost all musical sounds consist of a great many partials, whether they are harmonically related or not. Using techniques such as selective damping and the selective excitation of modes, it is possible (with a bit of practice) to learn to "hear out" these partials, to directly perceive the spectrum of the sound. This kind of listening is called *analytic* listening, in contrast to *holistic* listening in which the partials fuse together into one perceptual entity. When listening analytically, sounds fragment into their constituent elements. When listening holistically, each sound is perceived as a single unit characterized by a unique tone, color, or timbre.

Analytic listening is somewhat analogous to the ability of a trained musician to reliably discern any of several different parts in a complex score where the naive (and more holistic) listener perceives one grand sound mass.

[11] For more information on the modes of air columns, refer to Benade's *Fundamentals of Musical Acoustics*. See Brown ([B: 20] and [W: 3]) for a discussion of the inharmonicities that may originate in nonidealized strings and air columns.

When presented with a mass of sound, the ear must decide how many notes, tones, or instruments are present. Consider the closing chord of a string quartet. At one extreme is the fully analytic ear that "hears out" a large number of partials. Each partial can be attended to individually, and each has its own attributes such as pitch and loudness. At the other extreme is the fully holistic listener who hears the finale as one grand tone, with all four instruments fusing into a single rich and complex sonic texture. This is called the root or *fundamental bass* in the works of Rameau [B: 145]. Typical listening lies somewhere between. The partials of each instrument fuse, but the instruments remain individually perceptible, each with its own pitch, loudness, vibrato, and so on. What physical clues make this remarkable feat of perception possible?

One way to investigate this question experimentally is to generate clusters of partials and ask listeners "how many notes" they hear.[12] Various features of the presentation reliably encourage tonal fusion. For instance, if the partials:

(i) Begin at the same time (attack synchrony)
(ii) Have similar envelopes (amplitudes change similarly over time)
(iii) Are harmonically related
(iv) Have the same vibrato rate

then they are more likely to fuse into a single perceptual entity. Almost any common feature of a subgroup of partials helps them to be perceived together. Perhaps the viola attacks an instant early, the vibrato on the cello is a tad faster, or an aggressive bowing technique sharpens the tone of the first violin. Any such quirks are clues that can help the ear bind the partials of each instrument together while distinguishing viola from violin. Familiarity with the timbral quality of an instrument is also important when trying to segregate it from the surrounding sound mass, and there may be instrumental "templates" acquired with repeated listening.

The fusion and fissioning of sounds is easy to hear using a set of wind chimes with long sustain. I have a very beautiful set called the "Chimes of Partch,"[13] made of hollow metal tubes. When the clapper first strikes a tube, there is a "ding" that initiates the sound. After several strikes and a few seconds, the individuality of the tube's vibrations are lost. The whole set begins to "hum" as a single complex tone. The vibrations have fused. When a new ding occurs, it is initially heard as separate, but soon merges into the hum.

At the risk of belaboring the obvious, it is worth mentioning that many of the terms commonly used in musical discourse are essentially ambiguous. The strike of a metal bar may be perceived as a single "note" by a holistic listener, yet as a diverse collection of partials by an analytic listener. As the analytic

[12] This is an oversimplification of the testing procedures actually used by Bregman [B: 18] and his colleagues.
[13] See [B: 91].

listener assigns a separate pitch and loudness to each partial, the strike is heard as a "chord." Thus, the same sound stimulus can be legitimately described as a note or as a chord.

The ability to control the tonal fusion of a sound can become crucial in composition or performance with electronic sounds of unfamiliar timbral qualities. For example, it is important for the composer to be aware of "how many" notes are sounding. What may appear to be a single note (in an electronic music score or on the keyboard of a synthesizer) may well fission into multiple tones for a typical listener. By influencing the coincidence of attack, envelope, vibrato, harmonicity, and so on, the composer can help to ensure that what is heard is the same as what was intended. By carefully emphasizing parameters of the sound, the composer or musician can help to encourage the listener into one or the other perceptual modes.

The spectrum corresponds well to the physical behavior of the vibrations of strings, air columns, and bars that make up musical instruments. It also corresponds well to the analytic listening of humans as they perceive these sound events. However, people generally listen holistically, and a whole vocabulary has grown up to describe the tone color, sound quality, or timbre of a tone.

2.3 What Is Timbre?

If a tree falls in the forest, is there any timbre? According to the American National Standards Institute [B: 6], the answer must be "no," whether or not anyone is there to hear. They define:

> Timbre is that attribute of auditory sensation in terms of which a listener can judge two sounds similarly presented and having the same loudness and pitch as dissimilar.

This definition is confusing, in part because it tells what timbre is *not* (i.e., loudness and pitch) rather than what it is. Moreover, if a sound has no pitch (like the crack of a falling tree or the scrape of shoes against dry leaves), then it cannot be "similarly presented and have the same pitch," and hence it has no timbre at all. Pratt and Doak [B: 143] suggest:

> Timbre is that attribute of auditory sensation whereby a listener can judge that two sounds are dissimilar using any criterion other than pitch, loudness and duration.

And now the tree does have timbre as it falls, although the definition still does not specify what timbre is.

Unfortunately, many descriptions of timbral perception oversimplify. For instance, a well known music dictionary [B: 75] says in its definition of timbre that:

> On analysis, the difference between tone-colors of instruments are found to correspond with differences in the harmonics represented in the sound (see HARMONIC SERIES).

This is simplifying almost to the point of misrepresentation. Any sound (such as a metal bar) that does not have harmonics (partials lying at integer multiples of the fundamental) would have no timbre. Replacing "harmonic" with "partial" or "overtone" suggests a definition that equates timbre with spectrum, as in this statement by the Columbia Encyclopedia:

> [Sound] Quality is determined by the overtones, the distinctive timbre of any instrument being the result of the number and relative prominence of the overtones it produces.

Although much of the notion of the timbre of a sound can be attributed to the number, amplitudes, and spacing of the spectral lines in the spectrum of a sound, this cannot be the whole story because it suggests that the envelope and attack transients do not contribute to timbre. Perhaps the most dramatic demonstration of this is to play a sound backward. The spectrum of a sound is the same whether it is played forward or backward,[14] and yet the sound is very different. In the CD *Auditory Demonstrations* [D: 21], a Bach chorale is played forward on the piano, backward on the piano, and then the tape is reversed. In the backward and reversed case, the music moves forward, but each note of the piano is reversed. The piano takes on many of the timbral characteristics of a reed organ, demonstrating the importance of the time envelope in determining timbre.

2.3.1 Multidimensional Scaling

It is not possible to construct a single continuum in which all timbres can be simply ordered as is done for loudness or for pitch.[15] Timbre is thus a "multidimensional" attribute of sound, although exactly how many "dimensions" are required is a point of significant debate. Some proposed subjective rating scales for timbre include:

$$\begin{gathered}\text{dull} \longleftrightarrow \text{sharp}\\ \text{cold} \longleftrightarrow \text{warm}\\ \text{soft} \longleftrightarrow \text{hard}\\ \text{pure} \longleftrightarrow \text{rich}\\ \text{compact} \longleftrightarrow \text{scattered}\\ \text{full} \longleftrightarrow \text{empty}\\ \text{static} \longleftrightarrow \text{dynamic}\\ \text{colorful} \longleftrightarrow \text{colorless}\end{gathered}$$

[14] As usual, we ignore the phase spectrum.

[15] The existence of auditory illusions such as Shephard's ever rising scale shows that the timbre can interact with pitch to destroy this simple ordering. See [B: 41].

Of course, these attributes are perceptual descriptions. To what physically measurable properties do they correspond? Some relate to temporal effects (such as envelope and attack) and others relate to spectral effects (such as clustering and spacing of partials).

The attack is a transient effect that quickly fades. The sound of a violin bow scraping or of a guitar pick plucking helps to differentiate the two instruments. The initial breathy puff of a flautist, or the gliding blat of a trumpet, lends timbral character that makes them readily identifiable. An interesting experiment [B: 13] asked a panel of musically trained judges to identify isolated instrumental sounds from which the first half second had been removed. Some instruments, like the oboe, were reliably identified. But many others were confused. For instance, many of the jurists mistook the tenor saxophone for a clarinet, and a surprising number thought the alto saxophone was a french horn.

The envelope describes how the amplitude of the sound evolves over time. In a piano, for instance, the sound dies away at roughly an exponential rate, whereas the sustain of a wind instrument is under the direct control of the performer. Even experienced musicians may have difficulty identifying the source of a sound when its envelope is manipulated. To investigate this, Strong and Clark [B: 186] generated sounds with the spectrum of one instrument and the envelope of another. In many cases (oboe, tuba, bassoon, clarinet), they found that the spectrum was a more important clue to the identity of the instrument, whereas in other cases (flute), the envelope was of primary importance. The two factors were of comparable importance for still other instruments (trombone, french horn).

In a series of studies[16] investigating timbre, researchers generated sounds with various kinds of modifications, and they asked subjects to rate their perceived similarity. A "multidimensional scaling algorithm" was then used to transform the raw judgments into a picture in which each sound is represented by a point so that closer points correspond to more similar sounds.[17] The axes of the space can be interpreted as defining the salient features that distinguish the sounds. Attributes include:

(i) Degree of synchrony in the attack and decay of the partials
(ii) Amount of spectral fluctuation[18]
(iii) Presence (or absence) of high-frequency, inharmonic energy in the attack
(iv) Bandwidth of the signal[19]
(v) Balance of energy in low versus high partials

[16] See [B: 139], [B: 46], [B: 64], and [B: 63].
[17] Perhaps the earliest investigation of this kind was Stevens [B: 181], who studied the "tonal density" of sounds.
[18] Change in the spectrum over time.
[19] Roughly, the frequency range in which most of the partials lie.

(vi) Existence of formants[20]

For example, Grey and Gordon [B: 63] exchange the spectral envelopes[21] of pairs of instrumental sounds (e.g., a french horn and a bassoon) and ask subjects to rate the similarity and dissimilarity of the resulting hybrids. They find that listener's judgments are well represented by a three-dimensional space in which one dimension corresponds to the spectral energy distribution of the sounds. Another dimension corresponds to the spectral fluctuations of the sound, and they propose that this provides a physical correlate for the subjective quality of a "static" versus a "dynamic" timbre. The third dimension involves the existence of high-frequency inharmonicity during the attack, for instance, the noise-like scrape of a violin bow. They propose that this corresponds to a subjective scale of "soft" versus "hard" or perhaps a "calm" versus "explosive" dichotomy.

2.3.2 Analogies with Vowels

The perceptual effect of spectral modifications are often not subtle. Grey and Gordon [B: 63] state that "one hears the tones switch to each others vowel-like color but maintain their original ... attack and decay." As the spectral distribution in speech gives vowels their particular sound, this provides another fruitful avenue for the description of timbre. Slawson [B: 175] develops a whole language for talking about timbre based on the analogy with vowel tones. Beginning with the observation that many musical sounds can be described by formants, Slawson proposes that musical *sound colors* can be described as variable sources of excitation passed through a series of fixed filters. Structured changes in the filters can lead to perceptually sensible changes in the sound quality, and Slawson describes these modifications in terms of the frequencies of the first two formants. Terms such as laxness, acuteness, openness, and smallness describe various kinds of motion in the two-dimensional space defined by the center frequencies of the two formants, and correspond perceptually to transitions between vowel sounds. For instance, opening up the sustained vowel sound *ii* leads to *ee* and then to *ae*, and this corresponds physically to an increase in frequency of the first formant.

2.3.3 Spectrum and the Synthesizer

In principle, musical synthesizers have the potential to produce any possible sound and, hence, any possible timbre. But synthesizers must organize their

[20] Resonances, which may be thought of as fixed filters through which a variable excitation is passed.

[21] The envelope of a partial describes how the amplitude of the partial evolves over time. The spectral envelope is a collection of all envelopes of all partials. In Grey and Gordon's experiments, only the envelopes of the common partials are interchanged.

sound generation capabilities so as to allow easy control over parameters of the sound that are perceptually relevant to the musician. Although not a theory of timbral perception, the organization of a typical synthesizer is a market-tested, practical realization that embodies many of the perceptual dichotomies of the previous sections. Detailed discussions of synthesizer design can be found in [B: 38] or [B: 158].

Sound generation in a typical synthesizer begins with the creation of a waveform. This waveform may be stored in memory, or it may be generated by some algorithm such as FM [B: 32], nonlinear waveshaping [B: 152], or any number of other methods [B: 40]. It is then passed through a series of filters and modulators that shape the final sound. Perhaps the most common modulator is an envelope generator, which provides amplitude modulation of the signal. A typical implementation such as Fig. 2.9 has a four-segment envelope with attack, decay, sustain, and release. The attack portion dictates how quickly the amplitude of the sound rises. A rapid attack will tend to be heard as a percussive ("sharp" or "hard") sound, whereas a slow attack would be more fitting for sounds such as wind instruments which speak more hesitantly or "softly." The sustain portion is the steady state to which the sound decays after a time determined by the decay parameters. In a typical sample-based electronic musical instrument, the sustain portion consists of a (comparatively) small segment of the waveform, called a "loop," that is repeated over and over until the key is released, at which time the sound dies away at a specified rate.

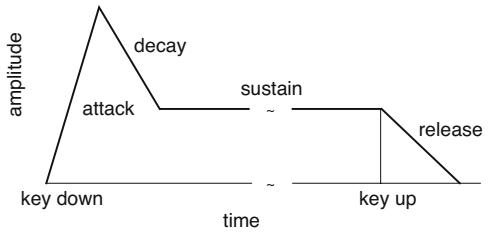


Fig. 2.9. The ADSR envelope defines a loudness contour for a synthesized sound. The attack is triggered by the key press. After a specified time, the sound decays to its sustain level, which is maintained until the key is raised. Then the loudness dies away at a rate determined by the release parameters.

Although the attack portion dictates some of the perceptual aspects, the steady-state sustained segment typically lasts far longer (except in percussive sounds), and it has a large perceptual impact. Depending on the underlying waveform, the sustain may be "compact" or "scattered," "bright" or "dull," "colorful" or "colorless," "dynamic" or "static," or "pure" or "rich." As most of these dichotomies are correlated with spectral properties of the wave, the design of a typical synthesizer can be viewed as supporting a spectral view of timbre, albeit tempered with envelopes, filters,[22] and modulators.

[22] One could similarly argue that the presence of resonant filters to shape the synthesized sound is a justification of the formant-based vowel analogy of timbre.

2.3.4 Timbral Roundup

There are several approaches to timbral perception, including multidimensional scaling, analogies with vowels, and a pragmatic synthesis approach. Of course, there are many other possible ways to talk about sounds. For instance, Schafer [B: 162] in Canada[23] distinguishes four broad categories by which sounds may be classified: physical properties, perceived attributes, function or meaning, and emotional or affective properties. Similarly, Erickson [B: 50] classifies and categorizes using terms such as "sound masses," "grains," "rustle noise," and so on, and exposes a wide range of musical techniques based on such sonic phenomena.

This book takes a restricted and comparatively simplistic approach to timbre. Although recognizing that temporal effects such as the attack and decay are important, we focus on the steady-state portion of the sound where timbre is more or less synonymous with stationary spectrum. Although admitting that the timbre of a sound can carry both meaning and emotion, we restrict ourselves to a set of measurable quantities that can be readily correlated with the perceptions of consonance and dissonance. These are largely pragmatic simplifications. By focusing on the spectral aspects of sound, it is possible to generate whole families of sounds with similar spectral properties. For instance, all harmonic instruments can be viewed as belonging to one "family" of sounds. Similarly, each inharmonic collection of partials has a family of different sounds created by varying the temporal features. As we will see and hear, each family of sounds has a unique tuning in which it can be played most consonantly.

Using the spectrum as a measure of timbre is like trying to make musical sounds stand still long enough to analyze them. But music does not remain still for long, and there is a danger of reading too much into static measurements. I have tried to avoid this problem by constantly referring back to sound examples and, where possible, to musical examples.

2.4 Frequency and Pitch

Conventional wisdom says that the perceived pitch is proportional to the logarithm of the frequency of a signal. For pure sine waves, this is approximately true.[24] For most instrumental sounds such as strings and wind instruments, it is easy to identify a fundamental, and again the pitch is easy to determine. But for more complex tones, such as bells, chimes, percussive and other inharmonic sounds, the situation is remarkably unclear.

[23] Not to be confused with Schaeffer [B: 161] in France who attempts a complete classification of sound.

[24] The mel scale, which defines the psychoacoustical relationship between pitch and frequency, deviates from an exact logarithmic function especially in the lower registers.

2.4.1 Pitch of Harmonic Sounds

Pythagoras of Samos[25] is credited with first observing that the pitch of a string is directly related to its length. When the length is halved (a ratio of 1:2), the pitch jumps up an octave. Similarly, musical intervals such as the fifth and fourth correspond to string lengths with simple ratios[26]: 2:3 for the musical fifth, and 3:4 for the fourth. Pythagoras and his followers proceeded to describe the whole universe in terms of simple harmonic relationships, from the harmony of individuals in society to the harmony of the spheres above. Although most of the details of Pythagoras' model of the world have been superseded, his vision of a world that can be described via concrete logical and mathematical relationships is alive and well.

The perceived pitch of Pythagoras' string is proportional to the frequency at which it vibrates. Moreover, musically useful pitch relationships such as octaves and fifths are not defined by differences in frequency, but rather by ratios of frequencies. Thus, an octave, defined as a frequency ratio of 2:1, is perceived (more or less) the same, whether it is high (say, 2000 to 1000 Hz) or low (250 to 125 Hz). Such ratios are called musical *intervals*.

The American National Standards Institute defines *pitch* as:

> that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high.

Because sine waves have unambiguous pitches (everyone orders them the same way from low to high[27]), such an ordering can be accomplished by comparing a sound of unknown pitch to sine waves of various frequencies. The pitch of the sinusoid that most closely matches the unknown sound is then said to be the pitch of that sound.

Pitch determinations are straightforward when working with strings and with most harmonic instruments. For example, refer back to the spectrum of an ideal string in Fig. 2.6 on p. 17 and the measured spectrum of a real string in Fig. 2.5 on p. 17. In both cases, the spectrum consists of a collection of harmonic partials with frequencies f, $2f$, $3f$, ..., plus (in the case of a real string) some other unrelated noises and artifacts. The perceived pitch will be f, that is, if asked to find a pure sine wave that most closely matches the pluck of the string, listeners invariably pick one with frequency f.

But it is easy to generate sounds electronically whose pitch is difficult to predict. For instance, Fig. 2.10 part (a) shows a simple waveform with a buzzy tone. This has the same period and pitch as (b), although the buzz is of a slightly different character. The sound is now slowly changed through (c) and

[25] The same guy who brought you the formula for the hypotenuse of a right triangle.

[26] Whether a musical interval is written as b:a or as a:b is immaterial because one describes the lower pitch relative to the upper, whereas the other describes the upper pitch relative to the lower.

[27] With the caveat that some languages may use different words, for instance, "big" and "small" instead of "low" and "high."

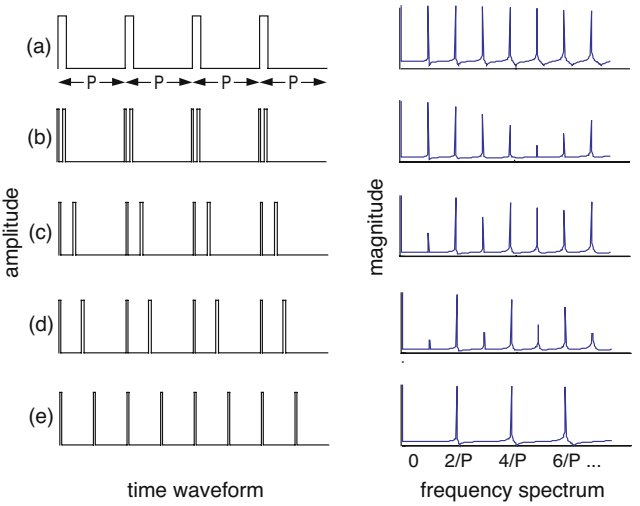


Fig. 2.10. (a) and (b) have the same period P and the same pitch. (c) and (d) change continuously into (e), which has period $\frac{P}{2}$. Thus, (e) is perceived an octave higher than (a). The spectra (shown on the right) also change smoothly from (a) to (e). Where exactly does the pitch change? See video example [V: 2].

(d) (still maintaining its period) into (e). But (e) is the same as (a) except twice as fast, and is heard an octave above (a)! Somewhere between (b) and (e), the sound jumps up an octave. This is demonstrated in video example [V: 2], which presents the five sounds in succession.

The spectra of the buzzy tones in Fig. 2.10 are shown on the right-hand side. Like the string example above, (a) and (e) consist primarily of harmonically related partials at multiples of a fundamental at $1/P$ for (a) and at $\frac{2}{P}$ for (e). Hence, they are perceived at these two frequencies an octave apart. But as the waveforms (b), (c), and (d) change smoothly from (a) to (e), the spectra must move smoothly as well. The changes in the magnitudes of the partials are not monotonic, and unfortunately, it is not obvious from the plots exactly where the pitch jumps.

2.4.2 Virtual Pitch

When there is no discernible fundamental, the ear will often create one. Such *virtual* pitch,[28] when the pitch of the sound is not the same as the pitch of any of its partials, is an aspect of holistic listening. Virtual pitch is expertly demonstrated on the Auditory Demonstrations CD [D: 21], where the "West-

[28] Terhardt and his colleagues are among the most prominent figures in this area; see [B: 195] and [B: 197].

minster Chimes" song is played using only upper harmonics. In one demonstration, the sounds have spectra like that shown in Fig. 2.11. This particular note has partials at 780, 1040, and 1300 Hz, which is clearly not a harmonic series. These partials are, however, closely related to a harmonic series with fundamental at 260 Hz, because the lowest partial is 260 times 3, the middle partial is 260 times 4, and the highest partial is 260 times 5. The ear appears to recreate the missing fundamental, and this perception is strong enough to support the playing of melodies, even when the particular harmonics used to generate the sound change from note to note.

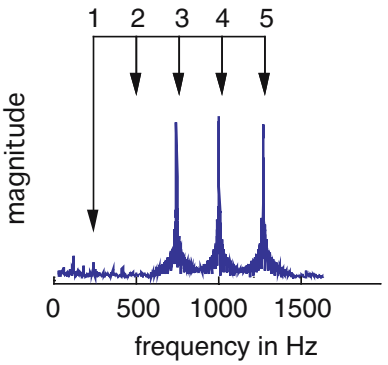


Fig. 2.11. Spectrum of a sound with prominent partials at 780, 1040, and 1300 Hz. These are marked by the arrows as the third, fourth, and fifth partials of a "missing" or "virtual" fundamental at 260 Hz. The ear perceives a note at 260 Hz, which is indicated by the extended arrow. See video example [V: 3].

The pitch of the complex tones playing the Westminster Chimes song is determined by the nearest "harmonic template," which is the average of the three frequencies, each divided by their respective partial numbers. Symbolically, this is $\frac{1}{3}(\frac{780}{3} + \frac{1040}{4} + \frac{1300}{5}) = 260$ Hz. This is demonstrated in video example [V: 3], which presents the three sine waves separately and then together. Individually, they sound like high-pitched sinusoids at frequencies 780, 1040, and 1300 Hz (as indeed they are). Together, they create the percept of a single sound at 260 Hz. When the partials are not related to any harmonic series, current theories suggest that the ear tries to find a harmonic series "nearby" and to somehow derive a pitch from this nearby series. For instance, if the partials above were each raised 20 Hz, to 800, 1060, and 1320 Hz, then a virtual pitch would be perceived at about $\frac{1}{3}(\frac{800}{3} + \frac{1060}{4} + \frac{1320}{5}) \approx 265$ Hz. This is illustrated in video example [V: 4], which plays the three sine waves individually and then together. The resulting sound is then alternated with a sine wave of frequency 265 Hz for comparison.

An interesting phenomenon arises when the partials are related to more than one harmonic series. Consider the two sounds:

(i) With partials at 600, 800, and 1000 Hz
(ii) With partials at 800, 1000, and 1200 Hz

Both have a clear virtual pitch at 200 Hz. The first contains the third, fourth, and fifth partials, whereas the second contains the fourth, fifth, and sixth partials. Sound example [S: 6] begins with the first note and ascends by adding 20 Hz to each partial. Each raised note alternates with a sine wave at the appropriate virtual pitch. Similarly, sound example [S: 7] begins with the second note and descends by subtracting 20 Hz from each partial. Again, the note and a sine wave at the virtual pitch alternate. The frequencies of all the notes are listed in Table 2.2. To understand what is happening, observe that each note in the table can be viewed two ways: as partials 3, 4, and 5 of the ascending notes or as partials 4, 5, and 6 of the descending notes. For example, the fourth note has virtual pitch at either

$$\frac{1}{3}\left(\frac{660}{3}+\frac{860}{4}+\frac{1060}{5}\right)\approx 215.6$$

or at

$$\frac{1}{3}\left(\frac{660}{4}+\frac{860}{5}+\frac{1060}{6}\right)\approx 171.2$$

depending on the context in which it is presented! Virtual pitch has been explored extensively in the literature, considering such factors as the importance of individual partials [B: 115] and their amplitudes [B: 116].

This ambiguity of virtual pitch is loosely analogous to Rubin's well-known face/vase "illusion" of Fig. 2.12 where two white faces can be seen against a black background, or a black vase can be seen against a white background. It is difficult to perceive both images simultaneously. Similarly, the virtual pitch of the fourth note can be heard as 215 when part of an ascending sequence, or it can be heard as 171 when surrounded by appropriate descending tones, but it is difficult to perceive both simultaneously.

Perhaps the clearest conclusion is that pitch determination for complex inharmonic tones is not simple. Virtual pitch is a fragile phenomenon that can be influenced by many factors, including the context in which the sounds are presented. When confronted with an ambiguous set of partials, the ear seems to "hear" whatever makes the most sense. If one potential virtual pitch is part of a logical sequence (such as the ascending or descending series in [S: 6] and [S: 7] or part of a melodic phrase as in the Westminster Chime song), then it may be preferred over another possible virtual pitch that is not obviously part of such a progression.

Fig. 2.12. Two faces or one vase? Ambiguous perceptions, where one stimulus can give rise to more than one perception are common in vision and in audition. The ascending/descending virtual pitches of sound examples [S: 6] and [S: 7] exhibit the same kind of perceptual ambiguity as the face/vase illusion.

Table 2.2. Each note consists of three partials. If the sequence is played ascending, then the first virtual pitch tends to be perceived, whereas if played descending, the second, lower virtual pitch tends to be heard. Only one virtual pitch is audible at a time. This can be heard in sound examples [S: 6] and [S: 7].

Note	First partial	Second partial	Third partial	Virtual Pitch ascending	Virtual Pitch descending
1	600	800	1000	200.0	158.9
2	620	820	1020	205.2	163.0
3	640	840	1040	210.4	167.1
4	660	860	1060	215.6	171.2
5	680	880	1080	220.9	175.3
6	700	900	1100	226.1	179.4
7	720	920	1120	231.3	183.6
8	740	940	1140	236.6	187.7
9	760	960	1160	241.8	191.8
10	780	980	1180	247.0	195.9
11	800	1000	1200	252.2	200.0

Pitch and virtual pitch are properties of a single sound. For instance, a chord played by the violin, viola, and cello of a string quartet is not usually thought of as having a pitch; rather, pitch is associated with each instrumental tone separately. Thus, determining the pitch or pitches of a complex sound source requires that it first be partitioned into separate perceptual entities. Only when a cluster of partials fuse into a single sound can it be assigned a pitch. When listening analytically, for instance, there may be more "notes" present than in the same sound when listening holistically. The complex sound might fission into two or more "notes" and be perceived as a chord. In the extreme case, each partial may be separately assigned a pitch, and the sound may be described as a chord.

Finally, the sensation of pitch requires time. Sounds that are too short are heard as a click, irrespective of their underlying frequency content. Tests with pure sine waves show that a kind of auditory "uncertainty principle" holds in which it takes longer to determine the pitch of a low-frequency tone than one of high frequency.[29]

2.5 Summary

When a tree falls in the forest and no one is near, it has no pitch, loudness, timbre, or dissonance, because these are perceptions that occur inside a mind. The tree does, however, emit sound waves with measurable amplitude, frequency, and spectral content. The perception of the tone quality, or timbre,

[29] This is discussed at length in [B: 99], [B: 61], and [B: 62].

is correlated with the spectrum of the physical signal as well as with temporal properties of the signal such as envelope and attack. Pitch is primarily determined by frequency, and loudness by amplitude. Sounds must fuse into a single perceptual entity for holistic listening to occur. Some elements of a sound encourage this fusion, and others tend to encourage a more analytical perception. The next chapter focuses on phenomena that first appear when dealing with pairs of sine waves, and successive chapters explore the implications of these perceptual ideas in the musical settings of performance and composition and in the design of audio signal-processing devices.

2.6 For Further Investigation

Perhaps the best overall introductions to the *Science of Sound* are the book by Rossing [B: 158] with the same name, *Music, Speech, Audio* by Strong [B: 187], and *The Science of Musical Sounds* by Sundberg [B: 189]. All three are comprehensive, readable, and filled with clear examples. The coffee-table quality of the printing of *Science of Musical Sound* by Pierce [B: 135] makes it a delight to handle as well as read, and it is well worth listening to the accompanying recording. Perceptual aspects are emphasized in the readable *Physics and Psychophysics of Music* by Roederer [B: 154], and the title should not dissuade those without mathematical expertise. Pickles [B: 133] gives *An Introduction to the Physiology of Hearing* that is hard to beat. The *Psychology of Music* by Deutsch [B: 41] is an anthology containing forward-looking chapters written by many of the researchers who created the field. The recording *Auditory Demonstrations* [D: 21] has a wealth of great sound examples. It is thorough and thought provoking.

For those interested in pursuing the acoustics of musical instruments, the *Fundamentals of Musical Acoustics* by Benade [B: 12] is fundamental. Those with better math skills might consider the *Fundamentals of Acoustics* by Kinsler and Fry [B: 85] for a formal discussion of bending modes of rods and strings (as well as a whole lot more). Those who want the whole story should check out the *Physics of Musical Instruments* by Fletcher and Rossing [B: 56]. Finally, the book that started it all is Helmholtz's *On the Sensations of Tones* [B: 71], which remains readable over 100 years after its initial publication.

3

Sound on Sound

All is clear when dealing with a single sine wave of reasonable amplitude and duration. The measured amplitude is correlated with the perceived loudness, the measured frequency is correlated with the perceived pitch, and the phase is essentially undetectable by the ear. Little is clear when dealing with large clusters of sine waves such as those that give rise to ambiguous virtual pitches. This chapter explores the in-between case where two sinusoids interact to produce interference, beating, and roughness. This is the simplest setting in which sensory dissonance occurs.

3.1 Pairs of Sine Waves

When listening to a single sine wave, amplitude is directly related to loudness and frequency is directly related to pitch. New perceptual phenomena arise when there are two (or more) simultaneously sounding sine waves. For instance, although the phase of a single sine wave is undetectable, the relative phases between two sine waves is important, leading to the phenomena of constructive and destructive interference. Beats develop when the frequencies of the two waves differ, and these beats may be perceived as sensory dissonance. Although the ear can resolve very small frequency changes in a single sine wave, there is a much larger "critical bandwidth" that characterizes the smallest difference between partials that the ear can "hear out" in a more complex sound. These ideas are explored in the next sections, and some simple models that capture the essence of the phenomena are described.

3.2 Interference

When two sine waves of exactly the same frequency are played together, they sound just like a single sine wave, but the combination may be louder or softer than the original waves. Figure 3.1 shows two cases. The sum of curves (a) and (b) is given in (c). As (a) and (b) have nearly the same phase (starting point), their peaks and valleys line up reasonably well, and the magnitude of the sum is greater than either one alone. This is called constructive interference. In contrast, when (d) and (e) are added together, the peaks of one are aligned with the troughs of the other and their sum is smaller than either alone, as

shown in curve (f). This is called destructive interference. Thus waves of the same frequency can either reinforce or cancel each other, depending on their phases.

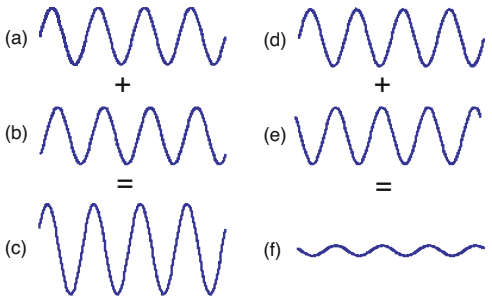


Fig. 3.1. Constructive and destructive interference between two sine waves of the same frequency. (a) and (b) add constructively to give (c), and (d) and (e) add destructively to give (f).

In Appendix A, trigonometriphiles will find an equation showing that the sum of two sine waves of the same frequency is always another sine of the same frequency, albeit with a different amplitude and phase. The equation even tells exactly what the amplitude and phase of the resulting wave are in terms of the phase difference of the original waves. These equations also describe (in part) the perceptual reality of combining sine waves in sound. Constructive interference reflects the common sense idea that two sine waves are louder than one. Destructive interference can be used to cancel (or muffle) noises by injecting sine waves of the same frequencies as the noises but with different phases, thus canceling out the unwanted sound. Sound canceling earphones from manufacturers such as Bose and Sennheiser use this principle, and some technical aspects of this technology, called active noise cancellation, are discussed in [B: 51].

3.3 Beats

What if the two sinusoids differ slightly in frequency? The easiest way to picture this is to imagine that the two waves are really at the same frequency, but that their relative phase slowly changes. When the phases are aligned, they add constructively. When the waves are out of phase, they interfere destructively. Thus, when the frequencies differ slightly, the amplitude of the resulting wave slowly oscillates from large (when in phase) to small (when out of phase).

Figure 3.2 demonstrates. At the start of the figure, the two sines are aligned almost perfectly, and the amplitude of the sum is near its maximum. By about 0.3 seconds, however, the two sine waves are out of sync and their sum is accordingly small. By 0.6 seconds, they are in phase again and the amplitude has grown, and by 0.9 seconds they are out of phase again and the

amplitude has shrunk. Thus, even though there are "really" two sine waves of two different frequencies present in the bottom plot of Fig. 3.2, it "looks like" there is only one sine wave that has a slow amplitude variation. This phenomenon is called *beating*.

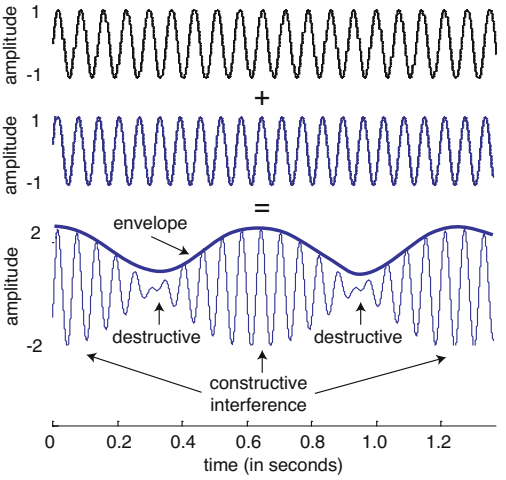


Fig. 3.2. The beating of two sine waves of close but different frequencies can be viewed as alternating regions of constructive and destructive interference. The bottom plot is the sum of the amplitudes of the two sinusoids above. The envelope outlines the undulations of the wave, and the beating occurs at a rate defined by the frequency of the envelope.

It may "look" like there is just one sine wave, but what does it "sound" like? Sound examples [S: 8] to [S: 10] investigate (and these are repeated in video examples [V: 5] to [V: 7]). The three examples contain nine short segments.

Examples [S: 8] and [V: 5]:

(i) A sine wave of 220 Hz (4 seconds)
(ii) A sine wave of 221 Hz (4 seconds)
(iii) Sine waves (i) and (ii) together (8 seconds)

Examples [S: 9] and [V: 6]:

(iv) A sine wave of 220 Hz (4 seconds)
(v) A sine wave of 225 Hz (4 seconds)
(vi) Sine waves (iv) and (v) together (8 seconds)

Examples [S: 10] and [V: 7]:

(vii) A sine wave of 220 Hz (4 seconds)
(viii) A sine wave of 270 Hz (4 seconds)
(ix) Sine waves (vii) and (viii) together (8 seconds)

The difference between the first two sine waves is fairly subtle because they are less than 8 cents[1] apart. Yet when played together, even this small difference

[1] There are 100 cents in a musical semitone. The *cent* notation is defined and discussed in Appendix B.

becomes readily perceivable as beats. The sound varies in loudness about once per second, which is the difference between the two frequencies. The fourth and fifth sine waves are noticeably distinct, lying about 39 cents apart. When played together, the perceived pitch is about 222.5 Hz. The beats are again prominent, beating at the much faster rate of five times each second. Again, the rate of the beating corresponds to the difference in frequency between sine waves.

In fact, it is not too difficult (if you like trigonometry) to show that the amplitude variation of the beats always occurs at a rate given by the difference in the frequencies of the sine waves. Appendix A gives the details. The result is[2]:

$$\left\{ \begin{array}{c} number\ of \\ beats\ per\ second \end{array} \right\} = \left\{ \begin{array}{c} frequency \\ of\ first\ wave \end{array} \right\} - \left\{ \begin{array}{c} frequency \\ of\ second\ wave \end{array} \right\}$$

Thus, the rate of beating decreases with the difference in frequency, and the beats disappear completely when the two sine waves are perfectly in tune. Because beats are often more evident than small pitch differences, they are used to tune stringed instruments such as the piano and guitar.

As the difference in frequency increases, the apparent rate in beating increases. A frequency difference of 1 Hz corresponds to a beat rate of 1 per second: 5 Hz corresponds to a beat rate of 5 times per second: 50 Hz corresponds to a beat rate of 50 times per second. But when the two sine waves of frequency 220 and 270 are played simultaneously, as in the ninth segment on the CD, there are *no beats at all*. Has the mathematics lied?

Don't lose the sound of the forest for the sound of falling trees.[3] Does the word "beats" refer to a physical phenomenon, or to a perception? If the former, then the mathematics shows that, indeed, the waveform in part (ix) of sound example [S: 10] exhibits beats at 50 Hz. But it is an empirical question whether this mathematical fact describes perceptual reality. There are two ways to "hear" part (ix). Listening holistically gives the impression of a single, slightly electronic timbre. Listening analytically reveals the presence of the two sine waves independently. As is audibly clear,[4] in neither case are there any beats (in the perceptual sense). Thus, the mathematical model that says that the beat rate is equal to the frequency difference is valid for perceptions of small differences such as 5 Hz, but fails for large differences such as 50 Hz.

Can the spectrum give any insight? Figure 3.3 shows time and frequency plots as the ratio of the frequencies of the two sine waves varies. When the ratio is large, such as 1:1.5, two separate peaks are readily visible in the spectral plot. As the ratio shrinks, the peaks grow closer. For 1:1.1, they

[2] If this turns out to be negative, then take its absolute value. There is no such thing as a negative beat.

[3] Recall the "paradox" on p. 11.

[4] Some people can also hear a faint, very low-pitched tone. This is the "difference frequency," which is due to nonlinear effects in the ear. See [B: 69] and [B: 140].

are barely discernible. For even smaller ratios, they have merged together and the spectrum appears to consist of only a single frequency.[5] A similar phenomenon occurs in the ear's "biological spectrum analyzer." When the waves are far apart, as in the sound example (ix), the two separate tones are clearly discernible. As they grow closer, it becomes impossible to resolve the separate frequencies. This is another property that the ear shares with digital signal-processing techniques such as the FFT.

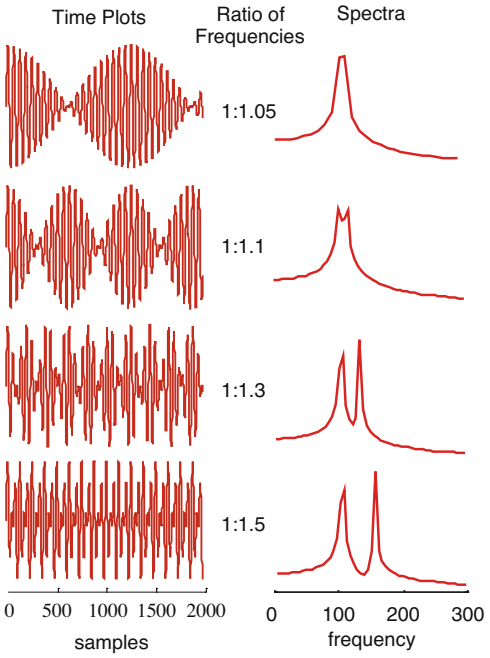


Fig. 3.3. Each plot shows a sum of two sine waves with frequencies in the specified ratios. Time plots show sample number versus amplitude, and spectral plots show frequency versus magnitude. Like the ear, the spectrum does not resolve partials when they are too close together.

3.4 Critical Band and JND

As shown in Fig. 2.4 on p. 16, sine waves of different frequencies excite different portions of the basilar membrane, high frequencies near the oval window and low frequencies near the apex of the conical cochlea. Early researchers such as Helmholtz [B: 71] believed that there is a direct relationship between the place of maximum excitation on the basilar membrane and the perceived pitch of the sound. This is called the "place" theory of pitch perception. When two tones are close enough in frequency so that their responses on the basilar membrane

[5] The resolving power of the FFT is a function of the sampling rate and the length of the data analyzed. Details may be found in Appendix C.

overlap, then the two tones are said to occupy the same *critical band*. The place theory suggests that the critical band should be closely related to the ability to discriminate different pitches. The critical band has been measured directly in cats and indirectly in humans in a variety of ways as described in [B: 140] and in [B: 212]. The "width" of the critical band is roughly constant at low frequencies and increases approximately proportionally with frequency at higher frequencies, as is shown in Fig. 3.4.

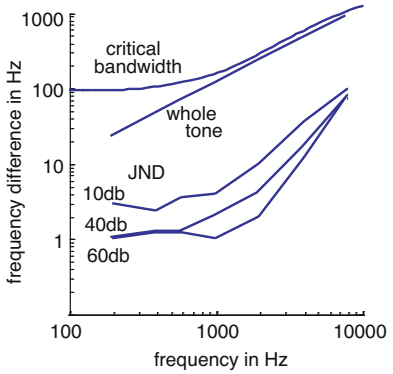


Fig. 3.4. Critical bandwidth is plotted as a function of its center frequency. Just Noticeable Differences at each frequency are roughly a constant percentage of the critical bandwidth, and they vary somewhat depending on the amplitude of the sounds. The frequency difference corresponding to a musical whole tone (the straight line) is shown for comparison. Data for critical bandwidth is from [B: 158] and for JND is from [B: 206].

The Just Noticeable Difference (JND) for frequency is the smallest change in frequency that a listener can detect. Careful testing such as [B: 211] has shown that the JND can be as small as two or three cents, although actual abilities vary with frequency, duration and intensity of the tones, training of the listener, and the way in which JND is measured. For instance, Fig. 3.4 shows the JND for tones with frequencies that are slowly modulated up and down. If the changes are made more suddenly, the JND decreases and even smaller differences are perceptible. As the JND is much smaller than the critical band at all frequencies, the critical band cannot be responsible for all pitch-detection abilities. On the other hand, the plot shows that JND is roughly a constant percentage of the critical band over a large range of frequencies.

An alternative hypothesis, called the "periodicity" theory of pitch perception, suggests that information is extracted directly from the time behavior of the sound. For instance, the time interval over which a signal repeats may be used to determine its frequency. In fact, there is now (and has been for the past 100 years or so) considerable controversy between advocates of the place and periodicity theories, and it is probably safe to say that there is not enough evidence to decide between them. Indeed, Pierce [B: 136] suggests that both mechanisms may operate in tandem, and a growing body of recent neurophysiological research (such as Cariani and his coworkers [B: 24] and [B: 25]) reinforces this.

Computational models of the auditory system such as those of [B: 111] and [B: 95] often begin with a bank of filters that simulate the action of the basilar membrane as it divides the incoming sound into a collection of signals in different frequency regions. Figure 3.5 schematizes a filter bank consisting of a collection of n bandpass filters with center frequencies $f_1, f_2, \ldots, f_n$. Typical models use between $n = 20$ and $n = 40$ filters, and the widths of the filters follow the critical bandwidth as in Fig. 3.4. Thus, the lower filters have a bandwidth of about 100 Hz and grow wider as the center frequencies increase.

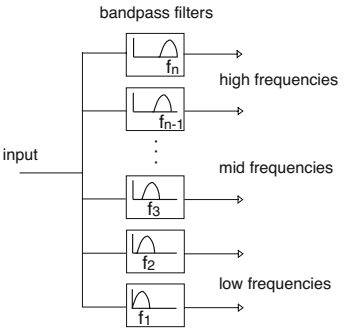


Fig. 3.5. The n filters separate the input sound into narrowband signals with bandwidths that approximate the critical bands of the basilar membrane.

The JND measures the ability to distinguish sequentially presented sine waves. Also important from the point of view of musical perception is the ability to distinguish simultaneously presented tones. Researchers have found that the ability to resolve concurrent tones is roughly equal to the critical band. That is, if several sine waves are presented simultaneously, then it is only possible to hear them individually if they are separated by at least a critical band. This places limits on how many partials of a complex tone can be "heard out" when listening analytically.

3.5 Sensory Dissonance

When listening to a pair of sine waves, both are readily perceptible if the frequencies are well separated. However, when the frequencies are close together, only one sine wave is heard (albeit with beats), due to the finite resolving power of the ear. What happens in between, where the ear is unsure whether it is hearing one or two things? Might the ear "get confused," and how would such confusion be perceived?

Sound example [S: 11] (and video example [V: 8]) investigate the boundary between these two regimes by playing a sine wave of frequency 220 Hz together with a wave of variable frequency beginning at 220 Hz and slowly increasing to 470 Hz. See Fig. 3.6 for a pictorial representation showing part of the waveform and typical listener reactions. Three perceptual regimes are evident. When the

sine waves are very close in frequency, they are heard as a single pleasant tone with slow variations in loudness (beats). Somewhat further apart in frequency, the beating becomes rapid and rough, dissonant. Then the tones separate and are perceived individually, gradually smoothing out as the tones draw further apart. Perhaps this perceived roughness is a symptom of the ear's confusion.

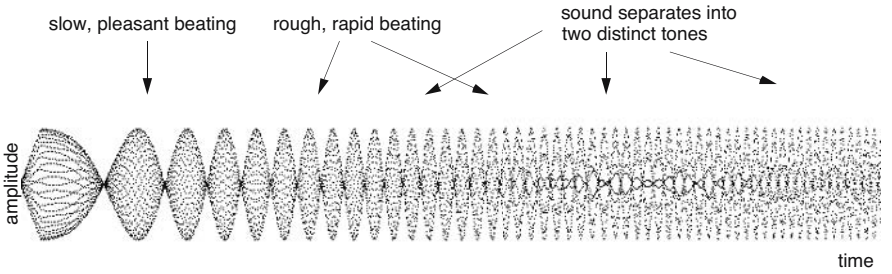


Fig. 3.6. Part of the waveform resulting from two simultaneous sine waves, one with fixed frequency of 220 Hz and the other with frequency that sweeps from 220 Hz to 470 Hz. Typical perceptions include pleasant beating (at small frequency ratios), roughness (at middle ratios), and separation into two tones (at first with roughness, and later without) for larger ratios. This can be heard in sound example [S: 11] and in video example [V: 8].

In an important experiment, Plomp and Levelt [B: 141] investigated this carefully by asking a large number of listeners to judge the consonance (euphoniousness, pleasantness) of a variety of intervals when sounded by pairs of pure sine waves.[6] The experiment is succinctly represented by the curves in Fig. 3.7, in which the horizontal axis represents the frequency interval between the two sine tones and the vertical axis represents a normalized measure of dissonance. The dissonance is minimum when both sine waves are of the same frequency, increases rapidly to its maximum somewhere near one-quarter of the critical bandwidth, and then decreases steadily back toward zero. In particular, this says that intervals such as the major seventh and minor ninth are almost indistinguishable from the octave in terms of sensory dissonance *for pure sine waves*. Such a violation of musical intuition becomes somewhat more palatable by recognizing that pure sine waves are almost never encountered in music.

Although this experiment was conducted with pairs of sine waves of fixed frequency, the results are similar to our observations from sound example [S: 11]. The same general trend of beats, followed by roughness and by a long smoothing out of the sound is apparent. The Plomp and Levelt curves have been duplicated and verified in different musical cultures (for instance,

[6] This experiment is discussed in more detail on p. 92.

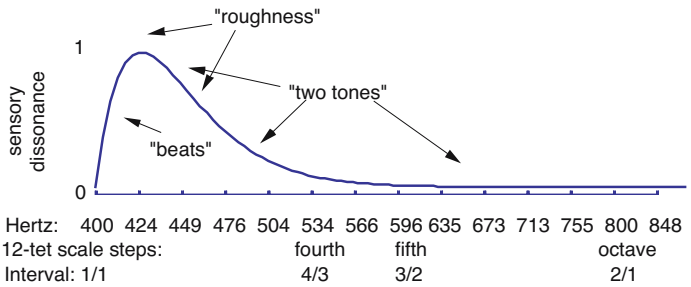


Fig. 3.7. Two sine waves are sounded simultaneously. Typical perceptions include pleasant beating (when the frequency difference is small), roughness (as the difference grows larger), and separation into two tones (at first with roughness, and later without) as the frequency difference increases further. The frequency of the lower sine wave is 400 Hz, and the horizontal axis specifies the frequency of the higher sine wave (in Hz, in semitones, and as an interval). The vertical axis shows a normalized measure of "sensory" dissonance.

Kameoka and Kuriyagawa [B: 79] and [B: 80] in Japan reproduced and extended the results in several directions), and such curves have become widely accepted as describing the response of the auditory system to pairs of sine waves. Figure 3.8 shows how the sensory dissonance changes depending on the absolute frequency of the lower tone.

The musical implications of these curves have not been uncontroversial. Indeed, some find it ridiculous that Plomp and Levelt used the words "consonance" and "dissonance" at all to describe these curves. "Everyone knows" that the octave and fifth are the most consonant musical intervals, and yet they are nowhere distinguishable from nearby intervals on the Plomp–Levelt

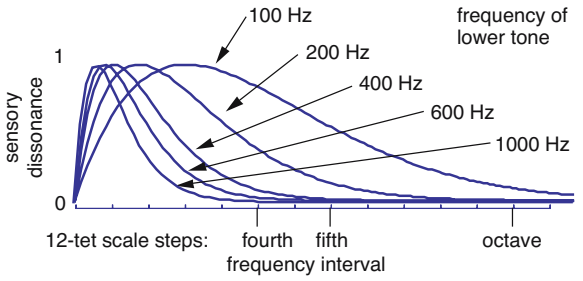


Fig. 3.8. Two sine waves are sounded simultaneously. As in Fig. 3.7, the horizontal axis represents the frequency interval between the two sine waves, and the vertical axis is a normalized measure of "sensory" dissonance. The plot shows how the sensory consonance and dissonance change depending on the frequency of the lower tone.

curves. We will have much more to say about this controversy in later chapters. Perhaps to defuse some of the resistance to their efforts, Plomp and Levelt were careful to call their axes *tonal* consonance and dissonance. Terhardt [B: 196] suggests the terms *sensory* consonance and dissonance, and we follow this usage.

One of the major contributions of the Plomp and Levelt paper was to relate the point of maximum sensory dissonance to the critical bandwidth of the ear. As the critical band varies somewhat with frequency, the dissonance curves are wider at low frequencies than at high, in accord with Fig. 3.8. Thus, intervals (like three semitones) that are somewhat consonant at high frequencies become highly dissonant at low frequencies. To hear this for yourself, play a major third in a high octave of the piano, and then play the same notes far down in the bass. The lower third sounds muddy and rough, and the higher third is clear and smooth. This is also consistent with musical practice in which small intervals appear far more frequently in the treble parts, and larger intervals such as the octave and fifth tend to dominate the lower parts.

3.6 Counting Beats

Perhaps the simplest way to interpret the sensory dissonance curves is in terms of the undulations of the amplitude envelope. Referring back to Fig. 3.7, the "slow pleasant beats" turn to roughness when the rate of the beating increases to around 20 or 30 beats per second.[7] As the frequencies spread further apart, they no longer lie within a single critical band[8]; the sine waves become individually perceptible and the sensory dissonance decreases. Thus, one way to create a model of sensory dissonance is to "count" the beats, to create a system that detects the amplitude envelope of the sound and then responds preferentially when the frequency of the envelope is near the critical number where the greatest dissonance is perceived.

One way to build such a model is to use a memoryless nonlinearity followed by a bandpass filter,[9] as shown in Fig. 3.9. The rectification nonlinearity

$$g(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{3.1}$$

leaves positive values unchanged and sets all negative values to zero. Combined with a low-pass filter, this creates an envelope detector[10] with an output that

[7] The peak of the dissonance curve in Fig. 3.7 occurs at about a semitone above 400 Hz, which is 424 Hz. Thus, the beat rate is 24 Hz when the dissonance is maximum.

[8] Figure 3.4 shows that a critical band centered at 400 Hz is a bit larger than 100 Hz wide.

[9] This is similar to an early model by Terhardt [B: 195].

[10] See Appendix C of [B: 76] for a discussion of envelope detectors.

rides along the outer edge of the signal. The bandpass filter is tuned to have maximum response in frequencies where the beating is most critical. Hence, its output is large when the beating is rough and small otherwise.

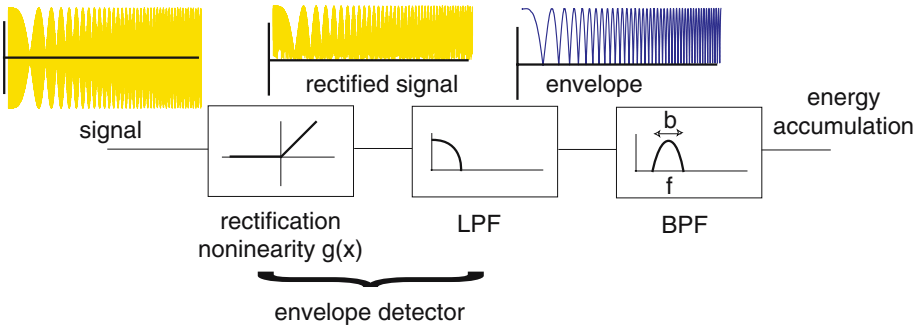


Fig. 3.9. The envelope detector outlines the beats in the signal and the bandpass filter is tuned to respond to energy in the 20 Hz to 30 Hz range where beating is perceived as roughest. Typical output of the model is shown in Fig. 3.10.

Typical output is shown in Fig. 3.10, which simulates the experiment of sound example [S: 11], where two sine waves of equal amplitude are summed to create the input; one is held fixed in frequency and the other slowly increases. The accumulated energy at the output of the model qualitatively mimics the sensory dissonance curve in Fig. 3.7. The detailed shape of the output depends on details of the filters chosen. For the simulation in Fig. 3.10, the LPF was a Remez filter with cutoff at 100 Hz and the BPF (which influences the detailed shape of the output signal) was a second-order Butterworth filter with passband between 15 and 35 Hz. This model is discussed further in Appendix G.

3.7 Ear vs. Brain

These first chapters have been using "the ear" as a synonym for "the human auditory system." Of course, there is a clear conceptual division between the physical ear (the eardrum, ossicles, cochlea, etc.) that acts as a transducer from pressure waves into neural impulses and the neural processes that subsequently occur inside the brain. It is not so clear, however, in which region various aspects of perception arise. For instance, the perception of pitch is at least partly accomplished on the basilar membrane, but it is also due in part to higher level processing.[11]

[11] Electrodes attached directly to the auditory nerves of deaf people induce the perception of a "fuzzy, scratchy" sound like "comb and paper"; see [B: 133].

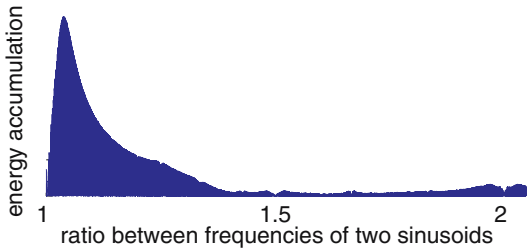


Fig. 3.10. Two sine waves of equal amplitude are input into the model of Fig. 3.9. The frequency of one sine is fixed at 400 Hz, and the other begins at 400 Hz and slowly increases to 850 Hz. The output of the bandpass filter (the energy accumulation) is largest when the beating is in the 20 to 30 Hz range.

To investigate whether the perception of roughness arises in the physical ear or in the brain, sound example [S: 12] repeats the previous track but with a binaural recording; the sine wave of fixed frequency is panned all the way to the right, and the variable sine wave is panned completely to the left. Listening normally through speakers, the two sides mix together in the air. But listening through headphones, each ear receives only one of the sine waves. If the perception of roughness originated exclusively in the physical ear, then no roughness should be heard. Yet it is audible, although the severity of the beating is somewhat reduced.[12] This suggests that perceptions of sensory dissonance are at least partly a mental phenomenon; that is, the signals from the two ears are combined in the neural architecture. As the effect is stronger when the waves physically mingle together (recall sound example [S: 11]), it is also likely that perceptions of roughness are due at least partly to the physical mechanism of the ear itself.

This chapter has considered the simple case of a pair of sine waves, where sensory dissonance is readily correlated with the interference phenomenon of beating. Later chapters return to this idea to build a more complete model that calculates the sensory dissonance of an arbitrary collection of sounds. Meanwhile, Chap. 4 turns to a consideration of musical scales and summarizes some of the many ways that people divide up the pitch continuum.

[12] Another way to listen to this sound example, suggested by D. Reiley, is to listen through the air and through headphones simultaneously. Plugging and unplugging the headphones as the example progresses emphasizes the dual nature of the perception: part "ear" and part "brain."

4

Musical Scales

People have been organizing, codifying, and systematizing musical scales with numerological zeal since antiquity. Scales have proliferated like tribbles in quadra-triticale: just intonations, equal temperaments, scales based on overtones, scales generated from a single interval or pair of intervals, scales without octaves, scales originating from arcane mathematical formulas, scales that reflect cosmological or religious structures, and scales that "come from the heart." Each musical culture has its own preferred scales, and many have used different scales at different times in their history. This chapter reviews a few of the more common organizing principles, and then discusses the question "what makes a good scale?"

4.1 Why Use Scales?

Scales partition the pitch continuum into chunks. As a piece of music progresses, it defines a scale by repeatedly exploiting a subset of all the possible pitch relationships. These repeated intervals are typically drawn from a small set of possibilities that are usually culturally determined. Fifteenth-century monks used very different scales than Michael Jackson, which are different from those used in Javanese gamelan or in Sufi Qawwali singing. Yet there are certain similarities. Foremost is that the set of all possible pitches is reduced to a very small number, five or six per octave for the monks, the major scale for Michael, either a five or seven-note nonoctave-based scale for the gamelan, and up to 22 or so notes per octave in some Arabic, Turkish, and Indian music traditions. But these are far from using "all" the possible perceptible pitches. Recall from the studies on JND that people can distinguish hundreds of different pitches within each octave.

Why does most music use only a few of these at a time? Most animals do not. Birdsong glides from pitch to pitch, barely pausing before it begins to slide away again. Whales click, groan, squeal, and wail their pitch in almost constant motion. Most natural sounds such as the howl of wind, the dripping of water, and the ping of ice melting are fundamentally unpitched, or they have pitches that change continuously.

One possible explanation of the human propensity to discretize pitch space involves the idea of categorical perception, which is a well-known phenomenon

to speech researchers. The brain tries to simplify the world around it. The Bostonian's "pahk," the Georgian's "paaark," and the Midwesterner's "park" are all interchangeable in the United States. Similarly, in listening to any real piece of music, there is a wide range of actual pitches that will be heard as the same pitch, say middle C. Perhaps the flute plays a bit flat, and the violin attacks a bit sharp. The mind hears both as the "same" C, and the limits of acceptability are far cruder than the ear's powers of resolution. Similarly, an instrumentalist does not play with unvarying pitch. Typically, there is some vibrato, a slow undulation in the underlying frequency. Yet the ear does not treat these variations as separate notes, but rather incorporates the perception of vibrato into the general quality of the tone.

Another view holds that musical scales are merely a method of classification that makes writing and performing music simpler. Scales help define a language that makes the communication of musical ideas more feasible than if everyone adopted their own pitch conventions. For whatever reasons, music does typically exploit scales. The next few sections look at some of the scales that have been historically important, and some of the ways that they have been generalized and extended.

4.2 Pythagoras and the Spiral of Fifths

Musical intervals are typically defined by ratios of frequencies, and not directly by the frequencies themselves. Pythagoras noted that a string fretted at its halfway point sounds an octave above the unfretted string, and so the octave is given by the ratio two to one, written $2/1$. Similarly, Pythagoras found that the musical fifth sounds when the length of two strings are in the ratio $3/2$, whereas the musical fourth sounds when the ratio of the strings is $4/3$.

Why do these simple integer ratios sound so special? Recall that the spectrum of a string (from Fig. 2.5 on p. 18 and Fig. 2.6 on p. 19) consists of a fundamental frequency f and a set of partials located at integer multiples of f. When the string is played at the octave (when the ratio of lengths is $2/1$), the spectrum consists of a fundamental at $2f$ along with integer partials at $2(2f) = 4f$, $3(2f) = 6f$, $4(2f) = 8f$, and so on, as shown in Fig. 4.1. Observe that all the partials of the octave align with partials of the original. This explains why the note and its octave tend to merge or fuse together, to be smooth and harmonious, and why they can easily be mistaken for each other. When the octave is even slightly out of tune, however, the partials do not line up. Chapter 3 showed how two sine waves that are close in frequency can cause beats that are perceived as a roughness or dissonance. In a mistuned octave, the n^{th} partial of the octave is very close to (but not identical with) the $2n^{th}$ partial of the fundamental. Several such pairs of partials may beat against each other, causing the characteristic (and often unwanted) out of tune sensation.

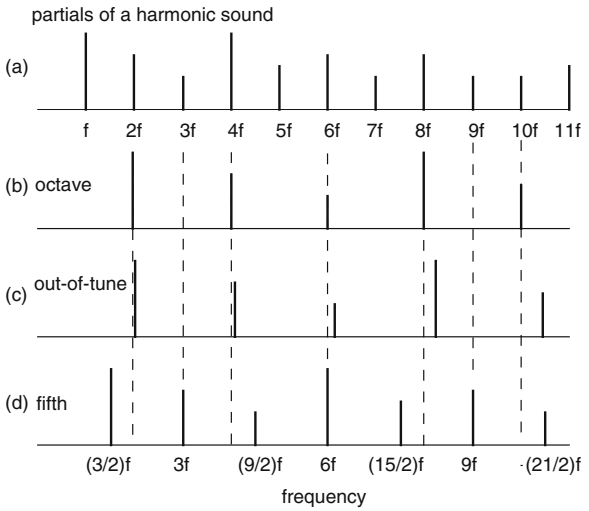


Fig. 4.1. A note with harmonic spectrum shown in (a) forms an octave, an out-of-tune octave, and a fifth, when played with (b), (c), and (d), respectively. Observe the coincidence of partials between (a) and (b) and between (a) and (d). In the out-of-tune octave (c), closely spaced partials cause beats, or roughness.

When a note is played along with its fifth, alternating partials line up. The partials that do not line up are far apart in frequency. As in the sensory dissonance curve of Fig. 3.7 on p. 47, such distinct partials tend not to interact in a significant way. Hence, the fifth also has a very smooth sound. As with the octave, when the fifth is mistuned slightly, its partials begin beating against the corresponding partials of the original note. Similarly, when other simple integer ratios are mistuned, nearby partials interact to cause dissonances. Thus, Pythagoras' observations about the importance of simple integer ratios can be viewed as a consequence of the harmonic structure of the string.

Using nothing more than the octave and the fifth, Pythagoras constructed a complete musical scale by moving successively up and down by fifths. Note that moving down by fifths is equivalent to moving up by fourths, because $(3/2)(4/3) = 2$. To follow Pythagoras' calculations, suppose that the (arbitrary) starting note is called C, at frequency 1. After including the fifth G at $3/2$, Pythagoras added D a fifth above G, which is $(3/2)(3/2) = (3/2)^2 = 9/4$. As $9/4$ is larger than an octave, it needs to be transposed down. This is easily accomplished by dividing by 2, and it gives the ratio $9/8$. Then add A with the ratio $(3/2)^3$, E at $(3/2)^4$, and so on (always remembering to divide by 2 when necessary to transpose back to the original octave). Alternatively, returning to the original C, it is possible to add notes spiraling up by fourths by adding F at $4/3$, $B\flat$ at $(4/3)^2$, and so on, again transposing back into the original octave. This process gives the *Pythagorean scale* shown in Fig. 4.2.

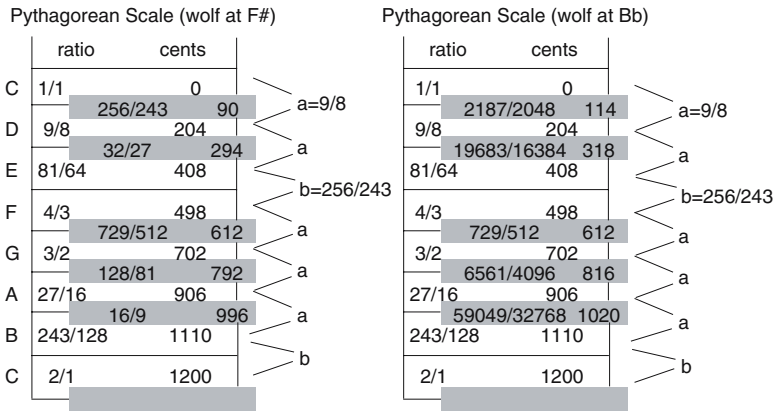


Fig. 4.2. In a Pythagorean scale, all intervals form perfect just fifths with the scale tone seven steps above except for one called the *wolf*. The Pythagorean diatonic (major) scale is shown on the white keys (labeled C, D, E, F, G, A, B, C) and the black keys show two possible extensions to a full 12-note system. The left-hand scale places the wolf on the $F\sharp$, and the right hand scale has the wolf at $B\flat$.

The seven-note Pythagorean scale in Fig. 4.2 is an early version of a *diatonic* scale. Diatonic scales, which contain five large steps and two small steps (whole tones and half tones), are at the heart of Western musical notation and practice [B: 53]. In this case, the scale contains the largest number of perfect fourths and fifths possible, because it was constructed using only the theoretically ideal ratios 3/2 and 4/3.

Much to Pythagoras' chagrin, however, there is a problem. When extending the scale to a complete tuning system (continuing to multiply successive terms by perfect 3/2 fifths), it is impossible to ever return to the unison.[1] After 12 steps, for instance, the ratio is $(3/2)^{12}$, which is $\frac{531441}{4096}$. When transposed down by octaves, this becomes $\frac{531441}{524288}$, which is about 1.0136, or one-quarter of a semitone (23 cents) sharp of the unison. This interval is called the *Pythagorean comma*, and Fig. 4.3 illustrates the Pythagorean "spiral of fifths."

The implication of this is that an instrument tuned to an exact Pythagorean scale, one that contained all perfect fifths and octaves, would require an infinite number of notes. As a practical matter, a Pythagorean tuner chooses one of the fifths and decreases it by the appropriate amount. This is called the *Pythgorean comma*, and the (imperfect) "fifth" that is a quarter semitone out of tune is called the *wolf* tone, presumably because it sounds bad enough

[1] To see that $(3/2)^n = 2^m$ has no integer solutions, multiply both sides by 2^n, giving $3^n = 2^{m+n}$. As any integer can be decomposed uniquely into primes, there can be no integer that factors into n powers of 3 and simultaneously into $m + n$ factors of 2.

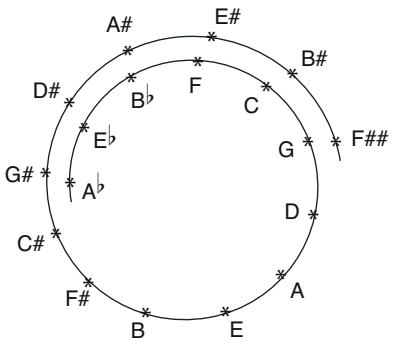


Fig. 4.3. In a Pythagorean scale built from all perfect fifths with ratios of $\frac{3}{2}$, the interval formed by 12 perfect fifths is slightly larger than an octave.

to make people howl. In the left-hand side of Fig. 4.2, the wolf fifth occurs between $F\sharp$ and the $C\sharp$ above.

To the numerologically inclined, the Pythagorean scale is a delight. First of all, there is nothing unique about the order in which the successive factors of a fourth and fifth are applied. For instance, the right-hand side of Fig. 4.2 shows a second Pythagorean scale with the wolf tone at $B\flat$. There are several ways to generate new scales based on the Pythagorean model. First, other intervals than the fifth and fourth could be used. For instance, let r stand for any interval ratio (any number between one and two will do), and let s be its complement (i.e., the interval for which $rs = 2$). Then r and s generate a family of scales analogous to the Pythagorean family. Of course, Pythagoras would be horrified by this suggestion, because he believed there was a fundamental beauty and naturalness to the first four integers,[2] and the simple ratios formed from them.

The Pythagorean scale can also be viewed as one example of a large class of scales based on tetrachords [B: 43], which were advocated by a number of ancient theorists such as Archytas, Aristoxenus, Didymus, Eratosthenes, and Ptolemy [B: 10]. A tetrachord is an interval of a pure fourth (a ratio of 4/3) that is divided into three subintervals. Combining two tetrachords around a central interval of 9/8 forms a seven-tone scale spanning the octave. For instance, Fig. 4.4 shows two tetrachords divided into intervals r, s, t and r', s', t'. When $r = r'$, $s = s'$, and $t = t'$, the scale is called an equal-tetrachordal scale. The Pythagorean scale is the special equal-tetrachordal scale where $r = r' = s = s' = 9/8$. A thorough modern treatment of tetrachords and tetrachordal scales is available in Chalmers [B: 31].

A third method of generating scales is based on the observation that the intervals between successive terms in the major Pythagorean scale are highly structured. As shown Fig. 4.2, there are only two distinct successive intervals, 9/8 and 256/243, between notes of the Pythagorean diatonic scale. Why not generate scales based on some other interval ratios r and s? For octave-based

[2] In the Pythagorean conception, the *tetraktys* was the generating pattern for all creation: politics, rhetoric, and literature, as well as music.

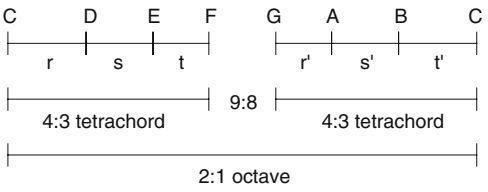


Fig. 4.4. Tetrachordal scales divide the octave into two 4:3 tetrachords separated by an interval of 9:8. The tetrachords are each divided into three intervals to form a seven-note scale, which is labeled in the key of C.

scales, this would require that there be integers n and m such that $r^m s^n = 2$. The simplest possible scale of this kind would have $s = r$, because then all adjacent notes would be equidistant.

4.3 Equal Temperaments

For successive notes of a scale to sound an equal distance apart, each interval must be the same. Letting s represent this interval, a scale with 12 equal steps can be written[3]

$$1, s, s^2, s^3, s^4, s^5, s^6, s^7, s^8, s^9, s^{10}, s^{11}, s^{12}.$$

If the scale is to repeat at the octave, the final note must equal 2. The equation $s^{12} = 2$ has only one real solution, called the twelfth root of two. It is notated $s = \sqrt[12]{2}$, and it is approximately 1.05946. A quick check with a calculator shows that multiplying 1.05946 times itself 12 times gives an answer (very close to) 2.

Although ratios and powers are convenient for many purposes, they can be cumbersome for others. An easy way to compare different intervals is to measure in *cents*, which divide each semitone into 100 equal parts, and the octave into 1200 parts. Figure 4.5 depicts one octave of a keyboard, and it shows the 12-tet tuning in ratios, in cents, and in the decimal equivalents. Given any ratio or interval, it is possible to convert to cents, and given any interval in cents, it is possible to convert back into a ratio. The conversion formulas are given in Appendix B.

The 12-tone equal-tempered scale (12-tet) is actually fairly recent.[4] With 12-tet, composers can modulate to distant keys without fear of hitting wolf tones. As the modern Western instrumental families grew, they were designed to play along with the 12-tet piano, and the tunings' dominance became a stranglehold. It is now so ubiquitous that many modern Western musicians and composers are even unaware that alternatives exist.

[3] The superscripts represent powers of s; hence, the interval between the n^{th} and $n + 1^{st}$ step is $s^{n+1}/s^n = s$.

[4] The preface to Jorgensen [B: 78] states that "the modern equal temperament taken for granted today as universally used on keyboard instruments did not exist in common practice on instruments until the early twentieth century... both temperament and music were tonal."

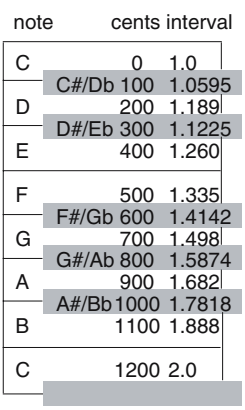


Fig. 4.5. The familiar 12-tone equal-tempered scale is the basis of most modern Western music. Shown here is one octave of the keyboard with note names, the intervals in cents defined by each key, and the decimal equivalents. The white keys (labeled C, D, E, F, G, A, B, C) form the diatonic C major scale, and the full 12 keys form the 12-tet chromatic scale.

This is not surprising, because most books about musical harmony and scales focus exclusively on 12-tet, and most music schools offer few courses on non-12-tet music, even though a significant portion of the historical repertoire was written before 12-tet was common. For instance, the standard music theory texts Piston [B: 137] and Reynolds and Warfield [B: 148] make no mention of any tuning other than 12-tet, and the word "temperament" does not appear in their indices. All major and minor scales of "classical music," the blues and pentatonic scales of "popular music," and all various "modes" of the jazz musician are taught as nothing more than subsets of 12-tet. When notes outside of 12-tet are introduced (e.g., "blues" or "bent" notes, glissandos, vibrato), they are typically considered aberrations or expressive ornaments, rather than notes and scales in themselves.

Yet 12 notes per octave is just one possible equal temperament. It is easy to design scales with an arbitrary number n of equal steps per octave. If r is the n^{th} root of 2 ($r = \sqrt[n]{2}$), then $r^n = 2$ and the scale

$$1, r, r^2, r^3, ...r^{n-1}, r^n$$

contains n identical steps. The calculation is even easier using cents. As there are 1200 cents in an octave, each step in n-tone equal temperament is $1200/n$ cents. Thus, each step in 10-tone equal temperament (10-tet) is 120 cents, and each step in 25-tet is 48 cents. Figure 4.6 shows all the equal temperaments between 9-tet and 25-tet. Because 12-tet is the most familiar, grid lines drawn at 100, 200, 300, ... cents provide a visual reference for the others.

The Structure of Recognizable Diatonic Tunings [B: 15] examines many equal-tempered tunings mathematically and demonstrates their ability to approximate intervals such as the perfect fifth. More important than the mathematics, however, are Blackwood's *12 Microtonal Etudes*[5] in each of the tunings between 13-tet and 24-tet, which demonstrate the basic feasibility of these tunings.

[5] See (and hear) [D: 4].

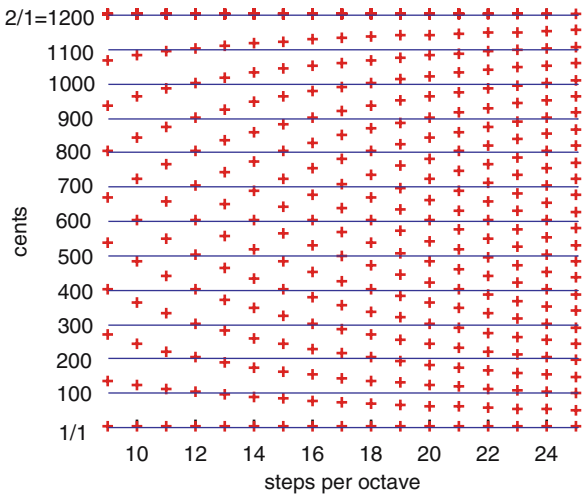


Fig. 4.6. Tuning of one octave of notes in the 9-tet, 10-tet, through 25-tet scales. The vertical axis proceeds from unison (1/1) to the octave (2/1). The horizontal lines emanate from the 12-tet scale steps for easy comparison.

It is fine to talk about musical scales and to draw interesting graphics describing the internal structure of tunings, but the crucial question must be: What do these tunings *sound* like? One of the major points of this book is that alternative tuning systems can be used to create enjoyable music. The accompanying CD contains several compositions in various equal temperaments, and these are summarized in Table 4.1. The pieces range from very strange sounding (*Isochronism* and *Swish*) to exotic (*Ten Fingers* and *The Turquoise Dabo Girl*) to reasonably familiar (*Sympathetic Metaphor* and *Truth on a Bus*). References marked with [S:] point to entries in the index of sound examples that starts on p. 399, where you can find instructions on how to listen to the files using a computer as well as more information about the pieces.

I believe that one of the main reasons alternative tunings have been underexplored is because there were few musical instruments capable of playing them. Ironically, the same keyboard instruments that saddled us with 12-tet for the past two and a half centuries can now, in their electronic versions, easily play in almost any tuning or scale desired.

Equal temperaments need not be based on the octave. A scale with n equal steps in every pseudo-octave[6] p is based on the ratio $r = \sqrt[n]{p}$. Again, this calculation is easier in cents. A pseudo-octave $p = 2.1$ defines an interval of 1284 cents. Dividing this into (say) 12 equal parts gives a scale step of 107 cents, a tuning that is explored in *October 21st* [S: 39]. Recall the "simple tune" of [S: 4]. This melody is developed further (and played in a variety

[6] $p = 2$ gives the standard octave.

Table 4.1. Musical compositions in various equal temperaments appearing on the CD-ROM.

Name of Piece	Equal Temperament	File	For More Detail
Swish	5-tet	`swish.mp3`	[S: 107]
Nothing Broken in Seven	7-tet	`broken.mp3`	[S: 117]
Pagan's Revenge	7-tet	`pagan.mp3`	[S: 116]
Phase Seven	7-tet	`phase7.mp3`	[S: 118]
March of the Wheel	7-tet	`marwheel.mp3`	[S: 115]
Anima	10-tet	`anima.mp3`	[S: 106]
Ten Fingers	10-tet	`tenfingers.mp3`	[S: 102]
Circle of Thirds	10-tet	`circlethirds.mp3`	[S: 104]
Isochronism	10-tet	`isochronism.mp3`	[S: 105]
The Turquoise Dabo Girl	11-tet	`dabogirl.mp3`	[S: 88]
Unlucky Flutes	13-tet	`13flutes.mp3`	[S: 99]
Hexavamp	16-tet	`hexavamp.mp3`	[S: 97]
Seventeen Strings	17-tet	`17strings.mp3`	[S: 98]
Truth on a Bus	19-tet	`truthbus.mp3`	[S: 100]
Sympathetic Metaphor	19-tet	`sympathetic.mp3`	[S: 101]
Dream to the Beat	19-tet	`dreambeat.mp3`	[S: 13]
Incidence and Coincidence	19-tet+12-tet	`incidence.mp3`	[S: 14]

of different pseudo-octaves) in *Plastic City* [S: 38]. One interesting pseudo-octave is $p = 2.0273$, which defines a pseudo-octave of 1224 cents, the amount needed to make 12 perfect 3/2 fifths.[7] Thus, the Pythagorean spiral of fifths can be closed by relaxing the requirement that the scale repeat each 2/1 octave. However, harmonic sounds clash dissonantly when played in 1224-cent intervals because of the almost coinciding partials. If the partials of the sounds are manipulated so as to realign them, then music in the 1224-cent pseudo-octave need not sound dissonant.

Moreno [B: 118] examines many nonoctave scales and finds that in some "n^{th} root of p" tunings the ratio p:1 behaves analogously to the 2:1 ratio in 12-tet. McLaren [B: 107] discusses the character of nonoctave-based scales and proposes methods of generating scales that range from number theory and continued fractions to the frequencies of vibrations of common objects. An interesting nonoctave scale was proposed independently by Bohlen [B: 16] on the basis of combination tones and by Mathews et al. [B: 101] on the basis of chords with ratios 3:5:7 (rather than the more familiar 3:4:5 of diatonic harmony). The resulting scale intervals are factors of the thirteenth root of 3 rather than the twelfth root of 2, and the *tritave*[8] plays some of the roles normally performed by the octave. Thus, $p = 3$ defines the pseudo-octave,

[7] Transposing $(\frac{3}{2})^{12}$ down (by octaves) to the nearest octave gives 1224 cents.
[8] An interval of 3/1 instead of the 2/1 octave.

and $r = \sqrt[13]{3}$ has 146.3 cents between each scale step. For more information, see the discussion surrounding Fig. 6.9 on p. 112.

It is also perfectly possible to define equal-tempered scales by simply specifying the defining interval. Wendy Carlos [B: 23], for instance, has defined the alpha scale in which each step contains 78 cents, and the beta scale with steps of 63.8 cents. Gary Morrison [B: 113] suggests a tuning in which each step contains 88 cents. This 88 cents per step tuning has 13.64 equal steps per octave, or 14 equal steps in a stretched pseudo-octave of 1232 cents. Many of these are truly xenharmonic in nature, with strange "harmonies" that sound unlike anything possible in 12-tet. As will be shown in subsequent chapters, a key idea in exploiting strange tunings such as these is to carefully match the tonal qualities of the sounds to the particular scale or tuning used. Two compositions on the CD use this 88 cent-per-tone scale: *Haroun in 88* [S: 15] and *88 Vibes* [S: 16].

4.4 Just Intonations

One critique of 12-tet is that none of the intervals are pure. For instance, the fifths are each 700 cents, whereas an exact Pythagorean 3/2 fifth is 702 cents. The imperfection of the wolf fifth has been spread evenly among all the fifths, and perhaps this small difference is acceptable. But other intervals are less fortunate. Just as the octave and fifth occur when a string is divided into simple ratios such as 2/1 and 3/2, thirds and sixths correspond to (slightly more complex) simple ratios. These are the *just* thirds and sixths specified in Table 4.2. For comparison, the 12-tet major thirds are 14 cents flat of the just values, and the minor thirds are 16 cents sharp.[9] Such discrepancies are clearly audible. Many music libraries will have a copy of Barbour [D: 2], which gives an extensive (and biased) comparison between just and equal-tempered intervals.

Table 4.2. The just thirds and sixths.

interval	ratio	cents
just minor third	6/5	316
just major third	5/4	386
just minor sixth	8/5	814
just major sixth	5/3	884

[9] The Pythagorean scale gives an even worse approximation. By emphasizing fourths and fifths, the thirds and sixths are compromised, and the Pythagorean major third 81/64 (408 cents) is even sharper than the equal-tempered third (400 cents). On the other hand, there are many ways to construct scales. For example, the Pythagorean interval $(\frac{3}{2})^8$, when translated to the appropriate octave, is almost exactly a just major third.

The *Just Intonation* (JI) scale appeases these ill-tempered thirds. Two examples are given in Fig. 4.7. The seven-note JI major scale in the top left is depicted in the key of C. The thirds starting on C, $C\sharp$, D, $D\sharp$, F, G, and $G\sharp$ are all just 5/4. As the fifths starting on C, $C\sharp$, F, G, and $G\sharp$ (among others) are perfect 3/2 fifths, all five form just major chords. Similarly, the JI scale on the bottom has five just minor chords starting on C, D, E, F, and A.

What do just intonations sound like? Sound examples [S: 17] through [S: 20] investigate. Scarlatti's Sonata K380 is first played in [S: 17] in 12-tet.[10] The sonata is then repeated in just intonation centered on C in [S: 18]. As it is performed in the appropriate key, there are no wolf tones. The overall impression is similar to the 12-tet version, although subtle differences are apparent upon careful listening. To clearly demonstrate the difference between these tunings, sound example [S: 19] plays in 12-tet and in just intonation simultaneously. Notes where the tunings are the same sound unchanged. Notes where the tunings differ sound chorused or phased and are readily identifiable.

The five pieces listed in Table 4.3 are performed in a variety of just intonation scales, which are documented in detail in [S: 23] through [S: 27]. These represent some of my earliest compositional efforts, and I prefer to recommend recordings by Partch [D: 31], Doty [D: 11], or Polansky [D: 34] to get a more complete idea of how just intonations can be used.

Table 4.3. Musical compositions in various just intonations appearing on the CD-ROM.

Name of Piece	File	For More Detail
Imaginary Horses	`imaghorses.mp3`	[S: 23]
Joyous Day	`joyous.mp3`	[S: 24]
What is a Dream?	`whatdream.mp3`	[S: 25]
Just Playing	`justplay.mp3`	[S: 26]
Signs	`signs.mp3`	[S: 27]

JI scales are sometimes criticized because they are inherently key specific. Although the above scales work well in C and in closely related keys (those nearby on the circle of fifths), they are notoriously bad in more distant keys. For instance, an $F\sharp$ major chord has a sharp third and an even sharper fifth (722 cents). Thus, it is unreasonable to play a piece that modulates from C to $F\sharp$ in JI. To investigate, sound example [S: 20] plays Scarlatti's K380 in just intonation centered on $C\sharp$ even though the piece is still played in the key of C. The out-of-tune percept is unmistakable in both the chords and the melody. When JI goes wrong, it goes very wrong. Barbour [D: 2] analogously plays a

[10] The musical score for K380 is shown in Fig. 11.3 on pp. 224 and 225. It is performed here (transposed down a third) in C major.

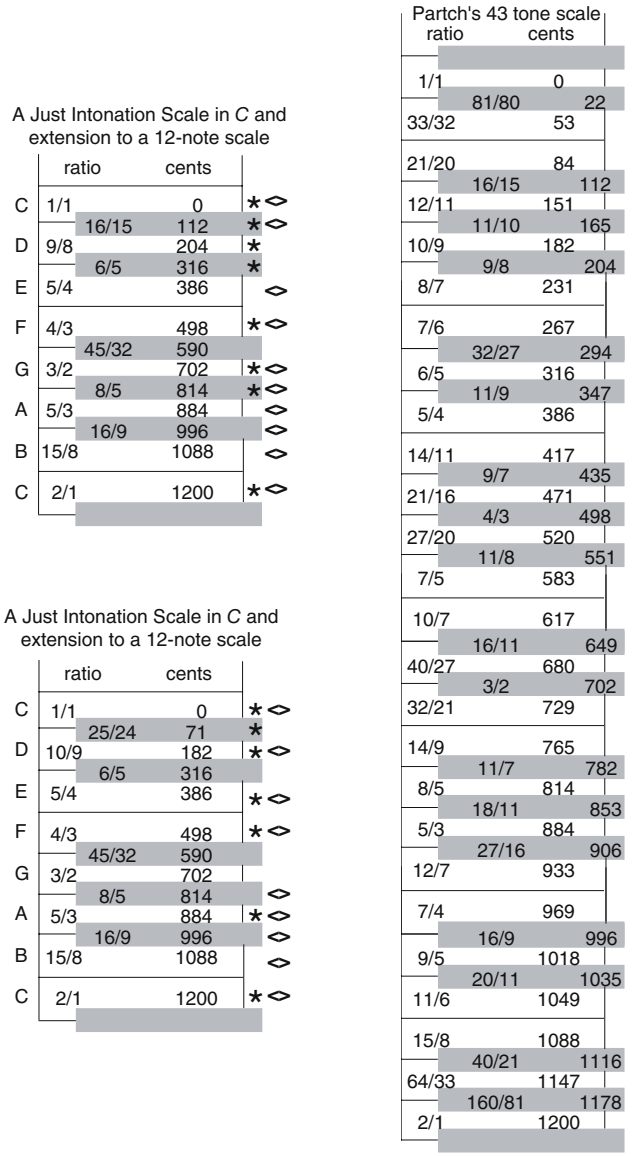


Fig. 4.7. The intervals in just intonation scales are chosen so that many of the thirds and fifths are ratios of small integers. Two JI diatonic scales are shown (labeled C, D, E, F, G, A, B, C) in the key of C; the black keys represent possible extensions to the chromatic 12-note setting. Each interval in the top JI major scale with a * forms a just major third with the note 4 scale steps above, and each note marked with <> forms a just fifth with the note 7 scale steps up. Similarly, in the bottom JI scale, each interval with a * forms a just minor third with the note 3 scale steps above, and each note marked with <> forms a just fifth with the note 7 scale steps up. Partch's 43-tone per octave scale contains many of the just intervals.

series of scales, intervals, and chords in a variety of tunings that demonstrate how bad JI can sound when played incorrectly. For instance, "Auld Lang Syne" is played in C in a just C scale, and it is then played in $F\sharp$ without changing the tuning. Barbour comments, "A horrible example—but instructive." It is a horrible example—of the misuse of JI. No practitioner would perform a standard repertoire piece in C just when it was written in the key of $F\sharp$.

There are several replies to the criticism of key specificity. First, most JI advocates do not insist that all music must necessarily be performed in JI. Simply put, if a piece does not fit well into the JI framework, then it should not be performed that way. Indeed, JI enthusiasts typically expect to retune their instruments from one JI scale to another for specific pieces. The second response is that JI scales may contain more than 12 notes, and so many of the impure intervals can be tamed. The third response involves a technological fix. With the advent of electronic musical instruments that incorporate tuning tables, it has become possible to retune "on the fly." Thus, a piece could be played in a JI scale centered around C, and then modulated (i.e. retuned) to a JI scale centered around $F\sharp$, without breaking the performance. This would maintain the justness of the intervals throughout. The fourth possibility is even newer. What if the tuning could be made *dynamic*, so as to automatically retune whenever needed? This is the subject of the "Adaptive Tunings" chapter.

The second criticism brought against JI is closely related to the first. Rossing [B: 158] explains that JI is impractical because an "orchestra composed of instruments with just intonation would approach musical chaos." Imagine if each instrumentalist required 12 instruments, one for each musical key! But it is only fixed pitch instruments like keyboards that are definitively locked into a single tuning. Winds, brass, and strings can and do change their intonation with musical circumstance. Where fixed pitch instruments set an equal-tempered standard, such microtonal inflections may be in the direction of equal temperament. But subtle pitch manipulations by the musician are heavily context dependent. Similarly, choirs sing very differently *a cappella* than when accompanied by a fixed pitch instrument.

The amusing and caustic book *Lies My Music Teacher Told Me* tells the first-hand story of a choir director who discovers justly intoned intervals, and trains his chorus to sing without tempering. Eskelin [B: 54] exhorts his choir to "sing *into* the chord, not through it," and teaches his singers to "lock into the chord," with the goal of tuning the sound "until the notes disappear." He describes a typical session with a new singer who is at first:

> reluctant and confused, and is convinced we are all a little crazy for asking him to sing the pitch out of tune. Eventually this defensiveness is replaced by curiosity, and finally the singer begins to explore the space outside his old comfort zone. When he experiences the peaceful calm that occurs when the note locks with [the] sustained root, the

> eyebrows raise, the eyes widen...another soul has been saved from the fuzziness of tempered tuning.

Whatever its practicality, JI concepts have been fertile ground for the creation of musical scales. For instance, scales can be based around intervals other than thirds, fifths, and octaves. Extending the JI vocabulary in this way leads to scales such as the 43-tone scale of Partch [B: 128] and to a host of 11 and 13-limit scales (those that use ratios with numerator and denominator less than the specified number). David Doty [B: 43] argues eloquently for the use of JI scales in his very readable *Just Intonation Primer*, and includes examples of many of the more important techniques for constructing JI scales. An organization called the Just Intonation Network has produced a number of interesting compilations, including *Rational Music for an Irrational World* and *Numbers Racket*, and numerous JI recordings are available from Frog Peak Music.[11]

4.5 Partch

Harry Partch was one of the twentieth century's most prolific, profound, opinionated, and colorful composers of music in just intonation. Partch developed a scale that uses 43 (unequal) tones in each octave. To perform in this 43-tone per octave JI scale, Partch designed and built a family of instruments, including a reed keyboard called the *chromelodeon*, the percussive *cloud chamber bowls*, the multistringed *kithara*, the *zymo-xyl* made from wine bottles, and the *mazda marimba* made from the glass of light bulbs. He wrote idiosyncratic choral and operatic music that mimicked some facets of ancient Greek performances and trained musicians to read and play his scores. Some of his recordings are available; both [D: 32] and [D: 31] have been recently reissued, and the Corporeal Meadows website [W: 6] contains photos of his instruments and up-to-date information on performances of his music.

Partch's scale, shown in Fig. 4.7, has the ability to maintain close approximations to many just intervals in many different keys. Also, the large palette of intervals within each octave provides the composer with far more choices than are possible in a smaller scale. For instance, depending on the musical circumstances and the desired effect, one might choose 7/4, 16/9, or 9/5 to play the role of dominant seventh, whereas the major seventh might be represented by 15/8 or 40/21. The melodic "leading tone" might be any of these, or perhaps 64/33 or 160/81 would be useful to guide the ear up into the octave. This scale, and Partch's theories, are discussed further in Sect. 5.3.

[11] See [B: 57] and [W: 13].

4.6 Meantone and Well Temperaments

Although many keyboards have been built over the centuries with far more than 12 keys per octave, none have become common or popular, presumably because of the added complexity and cost. Instead, certain tones on the 12-note keyboard were tempered to compromise between the perfect intervals of the JI scales and the possibilities of unlimited modulation in equal temperaments. Meantone scales aim to achieve perfect thirds and acceptable triads in a family of central keys at the expense of some very bad thirds and fifths in remote keys. They are typically built from a circle of fifths like the Pythagorean tuning, but with certain fifths larger or smaller than 3/2.

Figure 4.8 compares the Pythagorean, 12-tet, and two meantone tunings.[12] Each protruding spoke represents a fifth. A zero means that the fifth has a perfect 3/2 ratio, whereas a nonzero value means that the fifth is sharpened (if positive) or flattened (if negative) from 3/2. The Pythagorean tuning has zeroes everywhere except between the wolf, which is shown here between $G\sharp$ and $E\flat$. The -1 represents the size of the Pythagorean comma, and the sum of all the deviations of the fifths in any octave-based temperament must equal -1. In equal temperament, each fifth is squeezed by an identical $-1/12$. Quarter-comma meantone flattens each fifth by $-1/4$ and then compensates by creating a $+7/4$ wolf. This is done because a stack of four $-1/4$ tempered fifths gives a perfect 5/4 third.

Of course, there are many other possibilities. Figure 4.9 shows a number of historical well temperaments that aim to be playable (but not identical) in every key. Many of these scales are of interest because they are easily tuned by ear. Before this century, keyboardists typically tuned their instruments before each performance, and a tuning that is easy to hear was preferred over a theoretically more precise tuning that is harder to realize. In fact, as Jorgensen [B: 78] points out, equal temperament as we know it was not in common use on pianos as late as 1885.[13] This is at least partly because 12-tet is difficult to tune reliably.

But the interest in well temperings is more than just the practical matter of the ease of tuning. Each key in a well temperament has a unique tone color, key-color, or character that makes it distinct from all others. It was these characteristic colors that Bach demonstrated in his *Well Tempered Clavier*, and not (as is sometimes reported) the possibility of unlimited modulation in equal temperament. Many Baroque composers and theorists considered these distinctive modes an important element of musical expression, one that was sacrificed with the rise of 12-tet. Carlos [D: 7] performs pieces by Bach in various well temperaments. Katahn [D: 24] performs a stunning collection of piano sonatas in *Beethoven in the Temperaments*.

[12] The form of this diagram is taken from [B: 114].

[13] Ellis' measurements, reported in Helmholtz [B: 71], were accurate to about one cent.

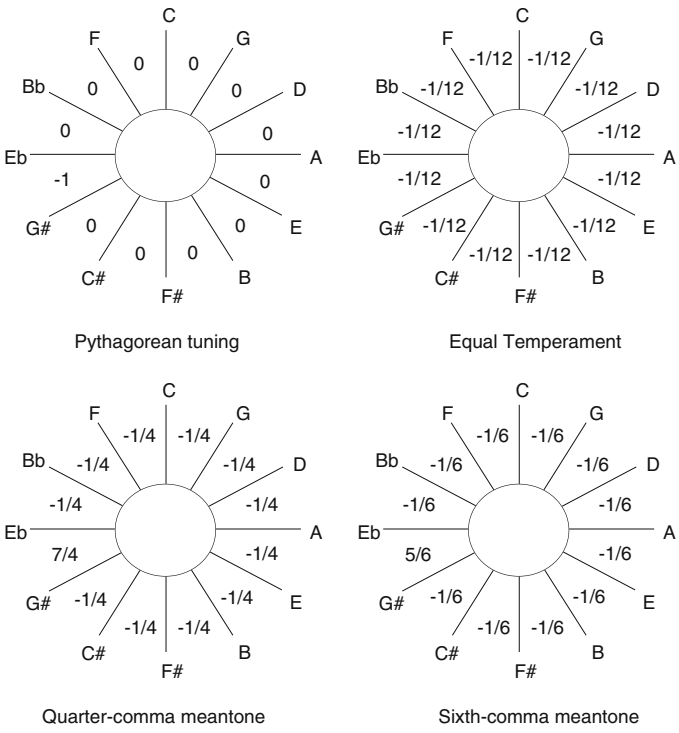


Fig. 4.8. Wheels of Tunings.

Two sound examples on the CD explore meantone tunings. Scarlatti's Sonata K380 is performed in the quarter comma meantone tuning centered in the key of C in [S: 21].[14] As in the JI performance, the effect is not overwhelmingly different from the familiar 12-tet rendition in [S: 17]. But when the meantone tuning is used improperly, the piece suffers (example [S: 22] uses the quarter comma meantone tuning centered on $C\sharp$).

4.7 Spectral Scales

Both the Pythagorean and the just scales incorporate intervals defined by simple integer ratios. Such ratios are aurally significant because the harmonic structure of many musical instruments causes their partials to overlap, whereas nearby out-of-tune intervals experience the roughness of beating partials. Another way to exploit the harmonic series in the creation of musical scales is to base the scale directly on the overtone series. Two possibilities are shown in Fig. 4.10. The first uses the eight pitches from the fourth octave of

[14] As in the previous examples [S: 17]–[S: 20], the piece is transposed to C major.

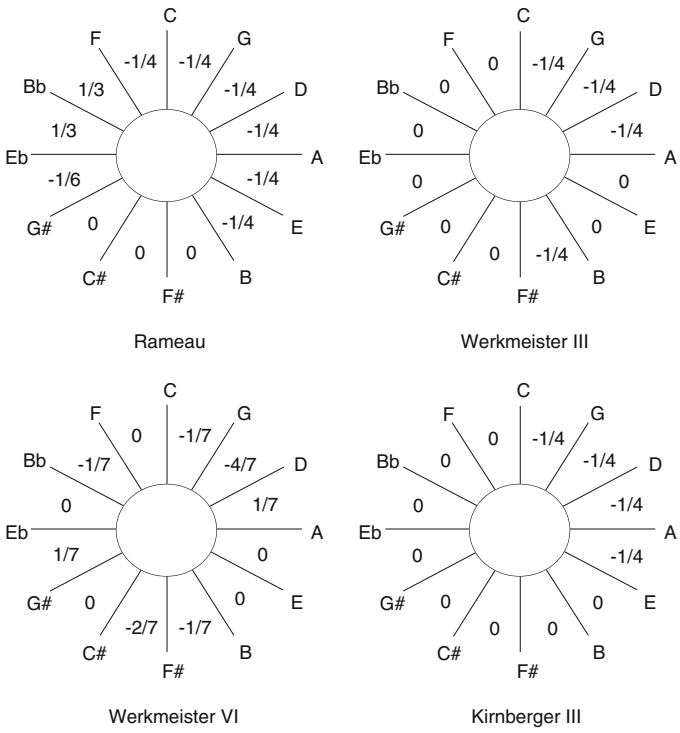


Fig. 4.9. Several well temperaments.

the overtone series, and the second exploits the 16 pitches of the fifth octave. Of course, many other overtone scales are possible because the sixth octave contains 32 different pitches (in general, the n^{th} octave contains 2^{n-1} pitches) and any subset of these can be used to define overtone scales.

Because the frequencies of the overtones are equally spaced arithmetically, they are not equally spaced perceptually. The pitches of the tones in a harmonic series grow closer together, and no two intervals between adjacent notes in the scale are the same. Moreover, each starting note has a different number of steps in its octave. This contrasts strongly with equal temperaments in which all successive intervals are identical and all octaves have the same number of steps. Nonetheless, overtone scales may be as old as prehistory. Tonometric measurements of pan pipes from Nasca, Peru suggest that the Nasca culture (200 BC to 600 AD) may have used an arithmetic overtone scale with about 43 Hz between succeeding tones, see [B: 67].

The "throat singing" technique ([B: 97], [D: 22], [D: 20]) allows a singer to manipulate the overtones of the voice. By emphasizing certain partials and de-emphasizing others, the sound may contain low droning hums and high

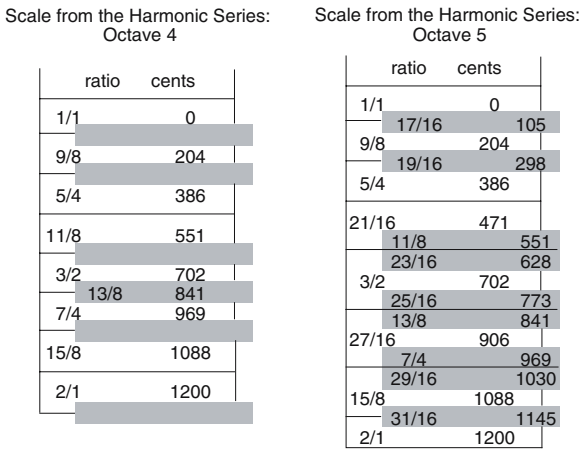


Fig. 4.10. All partials from the fourth octave of the harmonic series are reduced to the same octave, forming the scale on the left. Partials from the fifth octave of the harmonic series similarly form the scale on the right. The keyboard mappings are not unique.

whistling melodies simultaneously. Because the voice is primarily harmonic, the resulting melodies tend to lie on a single overtone scale.

Spectral composers such as Murail [B: 120] have attempted to build "a coherent harmonic system based on the acoustics of sound," which uses the "sound itself as a model for musical structure." One aspect of this is to decompose a sound into its constituent (sinusoidal) components and to use these components to define a musical scale. Thus, the scale used in the composition comes from the same source as the sound itself. When applied to standard harmonic sounds, this leads to overtone scales such as those in Fig. 4.10. More generally, this idea can be extended to inharmonic sounds. For example, the metal bar of Fig. 2.7 could be used to define a simple four-note scale. More complex vibrating systems such as drums, bells, and gongs can also be used to define corresponding "inharmonic" scales.

In Murail's *Gondwana* [D: 28], the sounds of bells (inharmonic) and trumpets (harmonic) are linked together by having the orchestral instruments play notes from scales derived from an analysis of the bells. In *Time and Again*, inharmonic sounds generated by a DX7 synthesizer are the catalyst for pitches performed by the orchestra. The orchestral instruments are thus used as elements to resynthesize (and augment) the sound of the DX7.

An interesting spectral technique is to tune a keyboard to one of the spectral scales, and to set each note to play a pure sine wave. Such a "scale" is indistinguishable from the "partials" of a note with complex spectrum, and it becomes possible to compose with the spectrum directly. As long as the sound

remains fused into a single perceptual entity, it can be heard as a flowing, constantly mutating complex timbre. When the sound is allowed to fission, then it breaks apart into two or more perceptual units. The composer can thus experiment with the number of notes heard as well as the tone quality. In Murail's *Désintégrations*, for example, two spectra fuse and fission in a series of spectral collisions. Such techniques are discussed at length in [B: 34].

As a composer, I find spectral scales to be pliant and easy to work with. They are capable of expressing a variety of moods, and some examples appearing on the CD are given in Table 4.4. These range from compositions using direct additive synthesis[15] (such as *Overtune* and *Pulsating Silences*) to those composed using spectral techniques and the overtone scales of Fig. 4.10 (such as *Free from Gravity* and *Immanent Sphere*). More information about the individual pieces is available in the references to the sound examples beginning on p. 399.

Table 4.4. Musical compositions in various spectral scales appearing on the CD-ROM.

Name of Piece	File	For More Detail
Immanent Sphere	`imsphere.mp3`	[S: 28]
Free from Gravity	`freegrav.mp3`	[S: 29]
Intersecting Spheres	`intersphere.mp3`	[S: 30]
Over Venus	`overvenus.mp3`	[S: 31]
Pulsating Silences	`pulsilence.mp3`	[S: 32]
Overtune	`overtune.mp3`	[S: 33]
Fourier's Song	`fouriersong.mp3`	[S: 34]

Spectral scales, even more than JI, tend to be restricted to particular keys or tonal centers. They contain many of the just intervals when played in the key of the fundamental on which they are based, but the approximations become progressively worse in more distant keys. Similarly, instruments tuned to overtone scales are bound to a limited number of related keys. For example, most "natural" (valveless) trumpets produce all their tones by overblowing, and they are limited to notes that are harmonics of the fundamental. These are inherently tuned to an overtone scale. Of course, many kinds of music do not need to modulate between keys; none of the pieces in Table 4.4 change key. Some do not even change chord. *Pulsating Silences* and *Overtune* do not even change notes!

[15] Where all sounds are created by summing a large collection of pure sine wave partials.

4.8 Real Tunings

Previous sections have described theoretically ideal tunings. When a real person tunes and plays a real instrument, how close is the tuning to the ideal? The discussion of just noticeable differences for frequency suggests that an accuracy of 2 or 3 cents should be attainable even when listening to the notes sequentially. When exploiting beats to tune simultaneously sounding pitches to simple intervals such as the octave and fifth, it is possible to attain even greater accuracy.[16] But this only describes the best possible. What is typical?

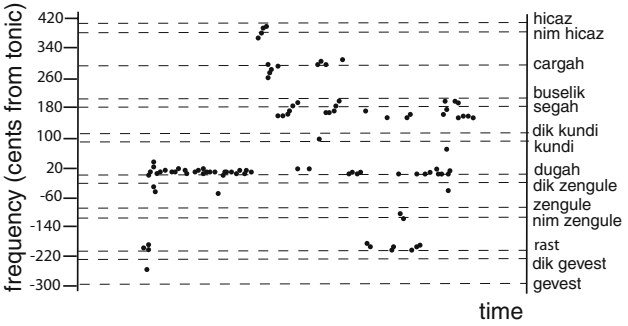


Fig. 4.11. Each note of the performance appears as a dot localized in time (the horizontal axis) and in frequency (the vertical axis). Theoretical note names of the Turkish tradition appear on the right. Figure used with permission [B: 4].

The actual tuning of instruments in performance is difficult to measure, especially in polyphonic music where there are many instruments playing simultaneously. Can Akkoç [B: 4] has recently transcribed the pitches of a collection of Turkish improvisations (*taksim*) played in a variety of traditional modes (*maqãmãt*) by acknowledged masters. Because these are played on a kind of flute (the *mansur ney* is an aerophone with openings at both ends), it is monophonic, and the process can be automated using a pitch-to-MIDI converter and then translated from MIDI into frequency. The results can be pictured as in Fig. 4.11, which plots frequency vs. time; each dot represents the onset of a note at the specified time and with the specified pitch. Observe the large cluster of dots near the tonic, the horizontal line labeled *dugah*. A large number of notes lie near this tonic, sometimes occurring above and sometimes below. Similarly, there are clusters of notes near other scale steps as indicated by the dashed lines. Interestingly, many pitches occur at locations that are far removed from scale steps, for instance, the cluster at the end halfway between *segah* and *dik kundi*. Thus, the actual performances are different from

[16] For instance, when matching two tones at 2000 Hz, it is possible to slow the beating rate below 1 beat per second, which corresponds to an accuracy of about half a cent.

the theoretical values. (Similar observations have also been made concerning Western performances.)

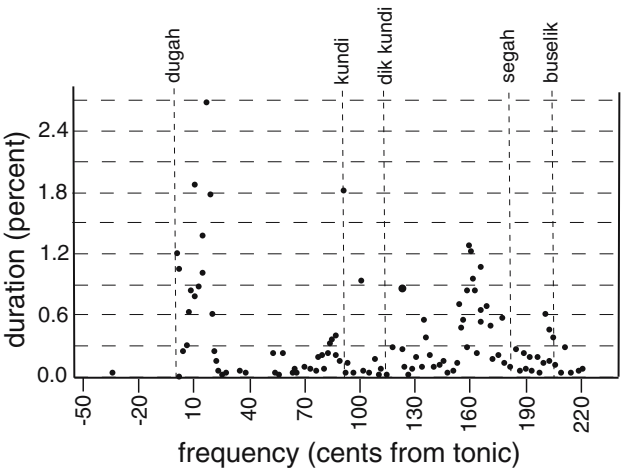


Fig. 4.12. Zooming into the region between dugah and busilek shows how the notes are distributed in pitch. Each dot represents the duration of all notes at the indicated frequency, as a percentage of the total duration of the piece. Figure used with permission [B: 4].

To try to understand this phenomenon, Akkoç replotted the data in the form of a histogram as in Fig. 4.12. In this performance, the longest time (about 2.7% of the total) was spent on a note about 10 cents above the tonic! The peaks of this plot can be interpreted as anchor tones around which nearby pitches also regularly occur. Akkoç interprets this stochastically, suggesting that master performers do not stick slavishly to predetermined sets of pitches, but rather deliberately play in distributions around the theoretical values. In one piece:

> two consecutive clusters are visited back to back at different points in time, and at each visit the musician has selected different sets of frequencies from the two clusters, thereby creating a variable micro scale...

Of course, the *mansur ney* is a variable intonation instrument, and it is perhaps (on reflection) not too surprising that the actual pitches played should deviate from the theoretical values. But surely an instrument like a modern, well-tuned piano would be tuned extremely close to 12-tet. This is, in fact, incorrect. Modern pianos do not even have real 2/1 octaves!

Piano tuning is a difficult craft, and a complex system of tests and checks is used to ensure the best sounding instrument. The standard methods begin

by tuning one note to a standard reference (say middle C) followed by all octaves of the C. Tuning then proceeds by fifths or by thirds (depending on the system), where each interval is mistuned (with respect to the just interval) by a certain amount. This mistuning is quantified by the number of beats per second that the tuner perceives. Jorgensen [B: 78], for instance, details several different methods for tuning equal temperament, and the instructions contain many statements such as "beating occurs at this high location between the nearly coinciding harmonics of the tempered interval below," "readjust middle C until both methods produce beats that are exactly equal," and "numbers denote beats per second of the test interval." At least part of the complexity of the tuning instructions occurs because beats are related linearly to frequency difference (and not frequency ratio, as is pitch). Thus, the expected number of beats changes depending on which octave is being tuned.

The deviation from 12-tet occurs because piano strings produce notes that are slightly inharmonic, which is heard as a moderate sharpening of the sound as it decays. Recall that an ideal string vibrates with a purely harmonic spectrum in which the partials are all integer multiples of a single fundamental frequency. Young [B: 208] showed that the stiffness of the string causes partials of piano wire to be stretched away from perfect harmonicity by a factor of about 1.0013, which is more than 2 cents. To tune an octave by minimizing beats requires matching the fundamental of the higher tone to the second partial of the lower tone. When the beats are removed and the match is achieved, the tuning is stretched by the same amount that the partials are stretched. Thus, the "octave" of a typical piano is a bit greater than 1202 cents, rather than the idealized 1200 cents of a perfect octave, and the amount of stretching tends to be greater in the very low and very high registers. This stretching of both the tuning and the spectrum of the string is clearly audible, and it gives the piano a piquancy that is part of its characteristic defining sound.

Interestingly, most people prefer their octaves somewhat stretched, even (or especially) when listening to pure tones. A typical experiment asks subjects to set an adjustable tone to an octave above a reference tone. Almost without exception, people set the interval between the sinusoids greater than a 2/1 octave. This craving for stretching (as Sundberg [B: 189] notes) has been observed for both melodic intervals and simultaneously presented tones. Although the preferred amount of stretching depends on the frequency (and other variables), the average for vibrato-free octaves is about 15 cents. Some have argued that this preference for stretched intervals may carry over into musical situations. Ward [B: 203] notes that on average, singers and string players perform the upper notes of the major third and the major sixth with sharp intonation.

Perhaps the preference for (slightly) stretched intervals is caused by constant exposure to the stretched sound of strings on pianos. On the other hand, Terhardt [B: 194] shows how the same neural processing that defines the sensa-

tion of virtual pitch[17] may also be responsible for the preference for stretched intervals. Although it may be surprising to those schooled in standard Western music that their piano is not tuned to real octaves, the stretching of octaves is a time-honored tradition among the Indonesian gamelan orchestras.

4.9 Gamelan Tunings

The gamelan, a percussive "orchestra," is the indigenous Indonesian musical traditions of Java and Bali. Gamelan music is varied and complex, and the characteristic shimmering and sparkling timbres of the metallophones are entrancing. The gamelan consists of a large family of inharmonic instruments that are tuned to either the five-note *slendro* or the seven-tone *pelog* scales. Neither scale lies close to the familiar 12-tet.

In contrast to the standardized tuning of Western music, each gamelan is tuned differently. Hence, the pelog of one gamelan may differ substantially from the pelog of another. Tunings tend not to have exact 2:1 octaves; rather, the octaves can be either stretched (slightly larger than 2:1) or compressed (slightly smaller). Each "octave" of a gamelan may differ from other "octaves" of the same gamelan.

An extensive set of measurements of actual gamelan tunings is given in [B: 190], which studies more than 30 complete gamelans. An average slendro tuning (obtained by numerically averaging the tunings of all the slendro gamelans) is

$$0,\ 231,\ 474,\ 717,\ 955,\ 1208$$

(values are in cents) which has a pseudo-octave stretched by 8 cents. The slendro tunings are often considered to be fairly close to 5-tet, although each gamelan deviates from this somewhat.

Similarly, an average pelog scale is

$$0,\ 120,\ 258,\ 539,\ 675,\ 785,\ 943,\ 1206,$$

which is a very unequal tuning that is stretched by 6 cents. The instruments and tunings of the gamelan are discussed at length in the chapter "The Gamelan," and detailed measurements of the tuning of two complete gamelans are given in Appendix L.

4.10 My Tuning Is Better Than Yours

It is a natural human tendency to compare, evaluate, and judge. Perhaps there is some objective criterion by which the various scales and tunings can be ranked. If so, then only the best scales need be considered, because it makes

[17] Recall the discussion on p. 35.

little sense to compose in inferior systems. Unfortunately, there are many different ways to evaluate the goodness, reasonableness, fitness, or quality of a scale, and each criterion leads to a different set of "best" tunings. Under some measures, 12-tet is the winner, under others 19-tet appears best, 53-tet often appears among the victors, 612-tet was crowned in one recent study, and under certain criteria nonoctave scales triumph. The next paragraph summarizes some of these investigations.

Stoney [B: 183] calculates how well the scale steps of various equal temperaments match members of the harmonic series. Yunik and Swift [B: 209] compare equal temperaments in terms of their ability to approximate a catalog of 50 different just intervals. Douthett et al. [B: 44] and van Prooijen [B: 144] use continued fractions to measure deviations from harmonicity for arbitrary equal temperaments. Hall [B: 68] observes that the importance of an interval depends on the musical context and suggests a least-mean-square-error criterion (between the intervals of n-tet and certain just intervals) to judge the fitness of various tunings for particular pieces of music. Krantz and Douthett [B: 88] propose a measure of "desirability" that is based on logarithmic frequency deviations, is symmetric, and can be applied to multiple intervals. As the criterion is based on "octave-closure," it is not dominated by very fine divisions of the octave. Erlich [B: 52] measures how close various just intervals are approximated by the equal temperaments up to 34-tet and finds that certain 10-tone scales in 22-tet approximate very closely at the 7-limit. Carlos [B: 23] searches for scales that approximate a standard set of just intervals but does not require that the temperaments have exact 2/1 octaves and discovers three new scales with equal steps of 78, 63.8, and 35.1 cents.

All of these comparisons consider how well one kind of scale approximates another. In an extreme case, Barbour [B: 10] essentially calculates how well various meantone and well-tempered scales approximate 12-tet and then concludes that 12-tet is the closest!

The search for sensible criteria by which to catalog and classify various kinds of scales is just beginning. Hopefully, as more people gain experience in composing in a variety of scales, patterns will emerge. One possibility is suggested in McLaren and Darreg [B: 109], who rate equal temperaments on a continuum that ranges from "biased towards melody" to "biased towards harmony." Perhaps someday it will be possible to reliably classify the possible "moods" that a given tuning offers. See [B: 36] for further comments.

4.11 A Better Scale?

Pythagoras felt that the coincidence of consonant intervals and small interval ratios were confirmation of deeply held philosophical beliefs. Such intervals are the most natural because they involve powerful mystical numbers like 1, 2, 3, and 4. Rameau [B: 145] considered the just intervals to be natural because they are outlined by the overtones of (many) musical sounds. Lou Harrison

says in his *Primer* [B: 70] that "The interval is just or not at all." "The best intonation is just intonation." For Harry Partch [B: 129], 12-tet keyboards are a musical straightjacket, "twelve black and white bars in front of musical freedom." From all of these points of view, the 12-tet tuning system is seen as a convenient but flawed approximation to just intervals, having made keyboard design more practical, and enabling composers to modulate freely.

Helmholtz further claimed that untrained and natural singers use just intervals, but that musicians, by constant contact with keyboards, have been trained (or brainwashed) to accept equal-tempered approximations. Only the greatest masters succeed in overcoming this cultural conditioning. Although logically sound, these arguments are not always supported by experimental evidence. Studies of the intonation of performers (such as [B: 4] and [B: 21]) show that they do not tend to play (or sing) in just intervals. Nor do they tend to play in Pythagorean tunings, nor in equal temperaments, exactly. Rather, they tend to play pitches that vary from any theoretically constructed scale.[18]

There are arguments based on numerology, physics, and psychoacoustics in favor of certain kinds of scales. There are arguments of expediency and ease of modulation in favor of others. While each kind of argument makes sense within its own framework, none is supported by irrefutable evidence. In fact, actual usage by musicians seems to indicate a considerable tolerance for mistunings in practical musical situations. Perhaps these deviations are part of the expressive or emotional content of music, perhaps they are part of some larger theoretical system, or perhaps they are simply unimportant to the appreciation of the music.

Almost every kind of music makes use of some kind of scale, some subset of all possible intervals from which composers and/or performers can build melodies and harmonies.[19] As the musical quality of an interval is highly dependent on the timbre or spectrum of the instruments (recall the "challenging the octave" example from the first chapter in which the octave was highly dissonant), *Tuning, Timbre, Spectrum, Scale* argues that the perceptual effect of an interval can only be reliably anticipated when the spectrum is specified. The musical uses of a scale depend crucially on the tone quality of the instruments.

Thus, a crucial aspect is missing from the previous discussions of scales. Justly intoned scales are appropriate *for harmonic timbres*. Overtone scales make sense *when used with sounds that have harmonic overtones*. Gamelan scales are designed for play *with metallophones*. Whether the scale is made from small integer ratios, whether it is formed from irrational number approximations such as the twelfth root of two, and whether it contains octaves

[18] Some recent work by Loosen [B: 98] suggests that musicians tend to judge familiar temperaments as more in-tune. Thus, violinists tend to prefer Pythagorean scales, and pianists tend to prefer 12-tet.

[19] The existence of sound collages and other textural techniques as in [D: 23], [D: 26], and [D: 43] demonstrates that scales are not absolutely necessary.

or pseudo-octaves (or neither) is only half of the story. The other half is the kinds of sounds that will be played in the scale.

5

Consonance and Dissonance of Harmonic Sounds

Just as a tree may crash silently (or noisily) to the ground depending on the definition of sound, the terms "consonance" and "dissonance" have both a perceptual and a physical aspect. There is also a dichotomy between attitude and practice, between the way theorists talk about consonance and dissonance and the ways that performers and composers use consonances and dissonances in musical situations. This chapter explores five different historical notions of consonance and dissonance in an attempt to avoid confusion and to place sensory consonance in its historical perspective. Several different explanations for consonance are reviewed, and curves drawn by Helmholtz, Partch, Erlich, and Plomp for harmonic timbres are explored.

5.1 A Brief History

Ideas of consonance and dissonance have changed significantly over time, and it makes little sense to use the definitions of one century to attack the conclusions of another. In his 1988 *History of 'Consonance' and 'Dissonance,'* James Tenney discusses five distinct ways that these words have been used. These are the melodic, polyphonic, contrapuntal, functional, and psychoacoustic notions of consonance and dissonance.

5.1.1 Melodic Consonance (CDC-1)

The earliest Consonance and Dissonance Concept (CDC-1 in Tenney's terminology) is strictly a melodic notion. Successive melodic intervals are consonant or dissonant depending on the surrounding melodic context. For instance, early church music was typically sung in unison, and CDC-1 refers exclusively to the relatedness of pitches sounded successively.

5.1.2 Polyphonic Consonance (CDC-2)

With the advent of early polyphony, consonance and dissonance began to refer to the vertical or polyphonic structure of music, rather than to its melodic

contour. Consonance became a function of the interval between (usually two) simultaneously sounding tones. Proponents of CDC-2 are among the clearest in relating "consonant" to "pleasant" and "dissonant" to "unpleasant." For instance, summing up the comments of a number of theorists from the thirteenth to the fifteenth century, Crocker [B: 35] concludes:

> These authors say, in sum, that the ear takes pleasure in consonance, and the greater the consonance the greater the pleasure; and for this reason one should use chiefly consonances...

Theorists were divided on the root cause of the consonance and dissonance. Some argued that the consonance of two tones is directly proportional to the degree to which the two tones sound like a single tone. Recall how the partials of simple ratio intervals such as the octave tend to line up, encouraging the two sounds to fuse together into a single perception. Other theorists focused on the numerical properties of consonant intervals, presuming, like the Pythagoreans, that the ear simply prefers simple ratios. As the simplest ratios are the unison, third, fourth, fifth, sixth, and octave, these were considered consonant and all others dissonant. These conflicting philosophies anticipate even further notions.

5.1.3 Contrapuntal Consonance (CDC-3)

Contrapuntal consonance defines consonance by its role in counterpoint. These are the "rules" that are familiar to music students today when learning voice-leading techniques. In a dramatic reversal of earlier usage, the fourth came to be considered a dissonance (except in certain circumstances) much as is taught today. Similarly, a minor third is considered consonant, whereas an augmented second is considered dissonant, even though the two intervals may be physically identical. Thus, it is the context in which the interval occurs that is crucial, and not the physical properties of the sound.

5.1.4 Functional Consonance (CDC-4)

Functional consonance begins with the relationship of the individual tones to a "tonic" or "root." Consonant tones are those that have a simple relationship to this fundamental root and dissonant tones are those that do not. This was crystallized by Rameau, whose idea of the *fundamental bass* roughly parallels the modern notion of the root of a chord. Rameau argues that all properties of:

> sounds in general, of intervals, and of chords rest finally on the single fundamental source, which is represented by the undivided string...

The "undivided" string in Fig. 5.1, which extends from 1 to A, sounds the fundamental bass. Half of the string, which vibrates at the octave, extends from 2 to A. One third of the string, which vibrates at the octave plus a fifth,

extends from 3 to A. Thus, Rameau identifies all of the familiar consonances by the distances on the string and their inversions. For example, suppose two notes form an interval of a major third (the region between 4 and 5 in the figure). These are both contained within the undivided string, which vibrates at the fundamental bass.

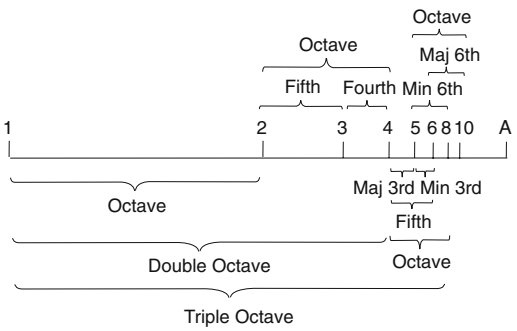


Fig. 5.1. Rameau illustrates the consonance of intervals on a vibrating string. If 1-A represents the complete string, 2-A is one half the string, 3-A is one third, and so on. The musical intervals that result from these different string lengths include all of the consonances. This figure is redrawn from [B: 145].

But Rameau's fundamental bass implies not only the static notion of the lowest note of a chord in root position, but also the dynamic notion of a succession of bass notes. Dissonances occur when the music has moved away from its root, and they set up an expectation of return to the root. Thus, functional dissonance is not a result of chordal motion, but rather its cause. This notion that dissonances cause motion is very much alive in modern music theory. For example, Walter Piston [B: 137], in *Harmony*, places himself firmly in this camp when he writes:

> A consonant interval is one which sounds stable and complete, whereas the characteristic of a dissonant interval is its restlessness and its need for resolution into a consonant interval... 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 sometimes erroneously assumed, its degree of unpleasantness to the ear.

5.1.5 Psychoacoustic Consonance (CDC-5)

The most recent concept of consonance and dissonance focuses on perceptual mechanisms of the auditory system. One CDC-5 view is called *sensory dissonance* and is usually credited to Helmholtz [B: 71] although it has been significantly refined by Plomp and Levelt [B: 141]. A major component of sensory dissonance is roughness such as that caused by beating partials; sensory consonance is then the smoothness associated with the absence of such

beats. Another component of psychoacoustic consonance, called *tonalness*, is descended from Rameau's fundamental bass through Terhardt's notions of harmony [B: 196] as extended by Parncutt [B: 126] and Erlich [W: 9]. A major component of tonalness is the closeness of the partials to a harmonic series; distonalness is thus increased as partials deviate from harmonicity.

CDC-5 notions of consonance and dissonance have three striking implications. First, individual complex tones have an intrinsic or inherent dissonance. From the roughness perspective, any tone with more than one partial inevitably has some dissonance, because dissonance is caused by interacting partials. Similarly from a tonalness point of view, as the partials of a sound deviate from a perfect harmonic template, the dissonance increases. These are in stark contrast to the earlier CDCs where consonance and dissonance were properties of relationships between tones.[1]

The second implication is that consonance and dissonance depend not just on the interval between tones, but also on the spectrum of the tones. Intervals are dissonant when the partials interact to cause roughness according to the sensory dissonance view. Similarly, intervals are increasingly dissonant as the partials deviate from harmonicity according to the tonalness view. In both cases, the exact placement of the partials is important.

The third implication is that consonance and dissonance are viewed as lying on a continuum rather than as an absolute property. In the earlier CDCs, a given interval is either consonant or dissonant. CDC-5 recognizes a continuum of possible gradations between consonance and dissonance.

The sensory notion of dissonance has no problem explaining the "challenging the octave" sound example [S: 1] of Chap. 1 (indeed, it was created from sensory considerations), and both sensory dissonance and tonalness have a firm basis in psychoacoustic experimentation (as discussed in Sect. 5.3.4). But these CDC-5 ideas are lacking in other respects. Perhaps the greatest strength of the contrapuntal and function consonance notions is that they provide comprehensive prescriptions (or at least descriptions) of the practice of harmony. They give guidance in the construction and analysis of polyphonic passages, and they explain how dissonances are crucial to the proper motion of musical compositions. In contrast, sensory dissonance and tonalness are static conceptions in which every collection of partials has some dissonance and there is not necessarily any relationship between successive clusters of sound in a musical sequence.

Mechanistic approaches to consonance are not without controversy and have been questioned from at least two perspectives. First, as Cazden [B: 28] points out, the ideas of psychoacoustic dissonance do not capture the functional idea of musical dissonance as restlessness or desire to resolve and the linked notion of consonance as the restful place to which resolution occurs. In essence, it becomes the responsibility of the composer to impose motion from

[1] Or of the relationship between a tone and the fundamental bass.

psychoacoustic dissonance to psychoacoustic consonance, if such a motion is desired.

Secondly, psychoacoustic experiments are tricky to conduct and interpret. Depending on the exact experimental setup, different effects may be emphasized. For example, many experiments address the relevance of beats and roughness to perceptions of intonation. Among these is Keisler [B: 81], who examines musicians' preferences to various "just" and "tempered" thirds and fifths by manipulating the partials of the sounds in a patterned way. Keisler concludes that beating is not a significant factor in intonation. Yet other studies such as Vos [B: 201], using different techniques, have found the opposite. Similarly, the fact of perception of virtual pitch is uncontested, and yet it sometimes appears as a strong and fundamental aspect (e.g., the Westminster chime song played by Houtsma [D: 21]), or it may appear fragile and ambiguous (as in sound examples [S: 6] and [S: 7]).

5.2 Explanations of Consonance and Dissonance

What causes these sensations of consonance and dissonance? Just as there are different paradigms for what consonance and dissonance mean, there are different ideas as to their cause: from numerological to physiological, from difference tones to differing cultures. Are there physical quantities that can be measured to make reasonable predictions of the perceived consonance of a sound, chord, or musical passage?

5.2.1 Small Is Beautiful

Perhaps the oldest explanation is the simplest: People find intervals based on small integer ratios more pleasant because the ear naturally prefers small ratios. Although somewhat unsatisfying due to its essentially circular nature, this argument can be stated in surprisingly many ways. Pythagoras, who was fascinated to find small numbers at the heart of the universe, was content with an essentially numerological assessment. Galileo [B: 58] wrote:

> agreeable consonances are pairs of tones which strike the ear with a certain regularity; this regularity consists in the fact that the pulses delivered by the two tones, in the same interval of time, shall be commensurable in number, so as not to keep the eardrum in perpetual torment, bending in two different directions in order to yield to the ever discordant impulses.

A more modern exposition of this same idea (minus the perpetual torment) is presented in Boomsliter and Creel [B: 17] and in Partch [B: 128]. Here, consonance is viewed in terms of the period of the wave that results when two tones of different frequency are sounded: The shorter the period, the more consonant the interval. Thus, 3/2 is highly consonant because the combined

wave repeats every 6 periods, whereas 301/200 is dissonant because the wave does not repeat until 60,200 periods.[2] In essence, this changes the argument from "the ear likes small ratios" to "the ear likes short waves." The latter forms a testable hypothesis, because the ear might contain some kind of detector that would respond more strongly to short repeating waveforms. In fact, periodicity theories of pitch perception [B: 24] and [B: 136] suppose such a time-based detector.

5.2.2 Fusion

The fusion of two simultaneously presented tones is directly proportional to the degree to which the tones are heard as a single perceptual unit. Recall from Fig. 4.1 on p. 52 that many of the partials of sounds in simple ratio intervals (such as the octave) coincide. The ear has no way to tell how much of each partial belongs to which note, and when enough partials coincide, the sounds may lose their individuality and fuse together. Stumpf [B: 188] determined that the degree of fusion of intervals depends on the simple frequency ratios in much the same way as consonance and hypothesized that fusion is the basis of consonance. The less willing a sound is to fuse, the more dissonant.

5.2.3 Virtual Pitch

Whereas Rameau's theories focus on physical properties of resonating bodies, Terhardt focuses on the familiarity of the auditory system with the sound of resonating bodies. This shifts the focus from the physics of resonating bodies to the perceptions of the listener. Terhardt's theory of virtual pitch [B: 197] is combined with a "learning matrix" [B: 195] (an early kind of neural network) to give the "principle of tonal meanings."

> By repeatedly processing speech, the auditory system acquires - among other Gestalt laws - knowledge of the specific pitch relations which... become familiar to the "central processor" of the auditory system ... This way, these intervals become the so-called musical intervals.

Terhardt emphasizes the key role that learning, and especially the processing of speech, plays in the perception of intervals. Different learning experiences lead to different intervals and scales and, hence, to different notions of consonance and dissonance.

One of the central features of virtual pitch is that the auditory system tries to locate the nearest harmonic template when confronted with a collection of

[2] On the other hand, the 12-tet equal fifth, whether considered as having infinite period or some very long finite period, is more consonant than other intervals such as 25/24, which have much shorter period. Thus, the theory cannot be so simple.

partials. This is unambiguous when the sound is harmonic but becomes more ambiguous as the sound deviates from a harmonicity. The idea of *harmonic entropy* (see [W: 9], Sect. 5.3.3, and Appendix J) quantifies this deviation, measuring the tonalness of an interval based on the uncertainty involved in interpreting the interval in terms of simple integer ratios.

5.2.4 Difference Tones

When two sine waves of different frequencies are sounded together, it is sometimes possible to hear a third tone at a frequency equal to the difference of the two. For instance, when waves of $f = 450$ Hz and $g = 570$ Hz are played simultaneously, a low tone at $g - f = 120$ Hz may also occur. These *difference tones* are usually attributed to nonlinear effects in the ear, and Roederer [B: 154] observes that "they tend to become significant only when the tones used to evoke them are performed at high intensity." Under certain conditions, difference tones may be audible at several multiples such as $2f - g$, $3g - 2f$, etc.[3] When f and g form a simple integer ratio, there are few distinct difference tones between the harmonics of f and the harmonics of g. For instance, if f and g form an octave, the difference tones occur at the same frequencies as the harmonics. But as the complexity of the ratio increases, the number of distinct difference tones increases. Thus, Krueger [B: 89] (among others) proposes that dissonance is proportional to the number of distinct difference tones; consonance occurs when there are only a few distinct difference tones.

Because both difference tones and beats occur at the same difference frequency $f - g$, it is easy to imagine that they are the same phenomenon, that difference tones are nothing more than rapid beats. This is not so. The essence of the beat phenomenon is fluctuations in the loudness of the wave, whereas difference tones are a result of nonlinearities, which may occur in the ear but may also occur in the electronic amplifier or loudspeaker system. Hall provides a series of tests that distinguish these phenomena in his paper [B: 69], "the difference between difference tones and rapid beats."

Difference tones are also similar to, but different from virtual pitch. Recall the example on p. 35 where three sine waves of frequencies 600, 800, and 1000 Hz generate both a virtual pitch at 200 Hz and a difference tone at 200 Hz. When the sine waves are raised to 620, 820, and 1020, the virtual pitch is somewhat higher than 200 Hz, whereas the difference tone remains at 200 Hz. For most listeners in most situations, the virtual pitch dominates emphasizing that difference tones can be subtle, except at high intensities. On the other hand, "false" difference tones can be generated easily in inexpensive electronic equipment by nonlinearities in the amplifier or speaker.

Difference tones can be readily heard in laboratory settings, and Hindemith [B: 72] presents several musical uses. In many musical settings, however, difference tones are not loud enough to be perceptually relevant and, hence, cannot

[3] In general, such higher order difference tones may occur at $(n + 1)g - nf$ for integers n.

form the basis of dissonance, as argued by Plomp [B: 138]. On the other hand, when difference tones are audible, they should be taken into account.

5.2.5 Roughness and Sensory Dissonance

Helmholtz's idea is that the beating of sine waves is perceived as roughness that in turn causes the sensation of dissonance. This sensory dissonance is familiar from Fig. 3.7 on p. 47, and this model can be used to explain why intervals made from simple integer ratios are perceptually special, as suggested by the mistuned octaves in Fig. 4.1 on p. 53.

The "challenging the octave" example (recall Fig. 1.1 on p. 2) demonstrates this dramatically. The partials of the inharmonic tone are placed so that they clash raucously when played in a simple 2/1 octave but sound smooth when played in a 2.1/1 pseudo-octave. Are these 2/1 and 2.1/1 intervals consonant or dissonant? It depends, of course, on the definition. Much of our intuition survives from CDC-2, where consonant and dissonant are equated with pleasant and unpleasant. Clearly, the 2.1/1 pseudo-octave is far more euphonious (when played with 2.1 stretched timbres) than the real octave. Modern musicians have been trained extensively (brainwashed?) with harmonic sounds. Because octaves are always consonant when played with harmonic sounds, the musician is likely to experience cognitive dissonance (at least) when hearing the 2.1/1 interval appear smoother than the 2/1 octave. This example is challenging to advocates of functional consonance (CDC-4) because it is unclear what the terms "key," "tonal center," and "fundamental root" mean for inharmonic sounds in non-12-tet scales. This is also a setting where the predictions of the tonalness model and the sensory dissonance model disagree, and this is discussed more fully in Sects. 6.2 and 16.3.

5.2.6 Cultural Conditioning

One inescapable conclusion is that notions of consonance and dissonance have changed significantly over the years. Presumably, they will continue to change. Cazden [B: 28] argues that the essence of musical materials cannot be determined by unchanging natural laws such as mathematical proportion, wave theories, perceptual phenomena, the physiology of hearing, and so on, because "it is not possible that laws which are themselves immutable can account for the profound transformations which have taken place in musical practice." Similarly, the wide variety of scales and tunings used throughout the world is evidence that cultural context plays a key role in notions of consonance and dissonance.

The importance of learning and cultural context in every aspect of musical perception is undeniable. But physical correlates of perceptions need not completely determine each and every historical style and musical idiosyncrasy as Cazden suggests; rather, they set limits beyond which musical explorations

cannot go. Surely the search for such limits is important, and this is discussed further in Sect. 16.3 "To Boldly Listen" in the final chapter.

Cazden also rightly observes that an individual's judgment of consonance can be modified by training, and so cannot be due entirely to natural causes. This is not an argument for or against any particular physical correlate, nor even for or against the existence of correlates in general. Rather, the extent to which training can modify a perception places limits on the depth and universality of the correlate.

The larger picture is that Cazden[4] is attacking excessive scientific reductionism in music theory, and in much of this he is quite correct. However, Cazden defines a consonant interval to be stable and a dissonant interval to be restless, an attack on the CDC-5 mindset using a CDC-4 definition. He states firmly that "consonance and dissonance do not originate on the level of properties of tones, but on the level of social communication," and hence, all such beat, fusion, and difference tone explanations are fundamentally misguided. Interpreting this to mean that questions of musical motion are not readily addressable within the CDC-5 framework, Cazden is correct. But this does not imply that such physiological explanations can offer nothing relevant to the perception of dissonances.

5.2.7 Which Consonance Explanation?

There are at least six distinctly different explanations for the phenomena of consonance and dissonance: small period detectors, fusion of sounds, tonalness and virtual pitch, difference tones, cultural conditioning, and beats or roughness. The difference tone hypothesis is the weakest of the theories because experimental evidence shows that it occurs primarily at high sound intensities, while dissonances can be clearly perceived even at low volumes.

The remaining possibilities each have strengths and limitations. Consonance and dissonance, as used in musical discourse, are complicated ideas that are not readily reducible to a single formula, acoustical phenomenon, or physiological feature. As we do not ultimately know which (if any) of the explanations is correct, a pragmatic approach is sensible: Which of the possible explanations for consonance and dissonance lead to musically sensible ideas for sound exploration and manipulation?

There is undoubtedly a large component of cultural influence involved in the perception of musical intervals, but it is hard to see how to exploit this view in the construction of musical devices or in the creation of new musics. On the other hand, as Terhardt [B: 195] points out, to whatever extent conventional musical systems are the result of a learning process, "it may not only be possible but even promising to invent new tonal systems." Chapters 7, 9, 14, and 15 do just this.

[4] In [B: 29] and [B: 30].

The importance of fusion in the general perception of sound is undeniable—if a tone does not fuse, then it is perceived as two (or more) tones. It is easy to see why a viable fusion mechanism might evolve: The difference between a pack of hyenas in the distance and a single hyena nearby might have immediate survival value. But its role in consonance is less clear. In the "Science of Sound" chapter, several factors were mentioned that influence fusion, including synchrony of attack, simultaneous modulation, and so on. Unfortunately, these have not yet been successfully integrated into a "fusion function" that allows calculation of a degree of fusion from some set of physically measurable quantities. Said another way, the fusion hypothesis does not (yet?) provide a physical correlate for consonance that can be readily measured. From the present utilitarian view, we therefore submerge the fusion hypothesis because it cannot give concrete predictions. Nonetheless, as will become clear when designing and exploring inharmonic sounds, ensuring that these sounds fuse in a predictable way is both important and nontrivial. Finding a workable measure of auditory fusion is an important arena for psychoacoustics work. See Parncutt [B: 126] for a step in this direction.

The small period hypothesis can only be sensibly applied to harmonic (i.e., periodic) sounds; it is not obvious how to apply it to music that uses inharmonic instruments. For example, the small period theory cannot explain why or how the pseudo-octaves of the "challenging the octave" experiment sound pleasant or restful (pick your favorite CDC descriptor) when played in the 2.1 stretched timbres. On the other hand, the roughness/sensory dissonance can be readily quantified in terms of the spectra of the sounds. Because a large class of interesting sounds are inharmonic, further chapters exploit the ideas of psychoacoustic consonance as a guide in the creation of inharmonic music. It is important to remember that this is just one possible explanation for the consonance and dissonance phenomenon. Moreover, the larger issue of creating "enjoyable music" is much wider than any notion of dissonance.

5.3 Harmonic Dissonance Curves

Early theorists focused on the consonance and dissonance of specific intervals within musical scales: Some are consonant and some are not. But there are an infinite number of possible pitches and, hence, of possible intervals. Are all of these other intervals perceived as dissonant? Helmholtz investigated this using two violins, one playing a fixed note and the other sliding up slowly. He found that intervals described by small number ratios are maximally consonant. Partch listened very carefully to his 43-tone-per-octave chromelodeon (a kind of reed organ) and learned to tune all the intervals by ear using the beating of upper partials. He found he could relate the relative consonances to small integer ratios. Erlich's tonalness quantifies the confusion of the ear as it tries to relate intervals to nearby small integer ratios. Plomp and Levelt use electronic equipment to carefully explore perceptions of consonance and disso-

nance. Again, they find that the intervals specified by small integer ratios are the most consonant. All four, despite wildly differing methods, mindsets, and theoretical inclinations, draw remarkably similar curves: Helmholtz's roughness curve, Partch's "one-footed bride," Erlich's harmonic entropy, and Plomp and Levelt's plot of consonance for harmonic tones.

5.3.1 Helmholtz and Beats

The idea of sensory consonance and dissonance was introduced[5] by Helmholtz in *On the Sensations of Tones* as a physical explanation for the musical notions of consonance and dissonance based on the phenomenon of beats. If two pure sine tones are sounded at almost the same frequency, then a distinct beating occurs that is due to interference between the two tones. The beating becomes slower as the two tones move closer together, and it completely disappears when the frequencies coincide. Typically, slow beats are perceived as a gentle, pleasant undulation, whereas fast beats tend to be rough and annoying, with maximum roughness occurring when beats occur about 32 times per second. Observing that any sound can be decomposed into sine wave partials, Helmholtz theorized that dissonance between two tones is caused by the rapid beating between the partials. Consonance, according to Helmholtz, is the absence of such dissonant beats.

To see Helmholtz's reasoning, suppose that a sound has a harmonic spectrum like the guitar string of Fig. 2.5 on p. 17, or its idealized version in Fig. 2.6 on p. 17. When such a sound is played at a fundamental frequency $f = 200$ (near the G below middle C), its spectrum is depicted in the top graph of Fig. 5.2. The same spectrum transposed to a fundamental frequency $g = 258$ is shown just below. Observe that many of the upper partials of f are close to (but not coincident with) upper partials of g. For instance, the fourth and fifth partials of f are very near the third and fourth partials of g. As partials are just sine waves, they beat against each other at a rate proportional to the frequency difference, in this case 26 Hz and at 32 Hz. Because both these beat rates are near 32 Hz, the partials interact roughly.

Assuming that the roughnesses of all interacting partials add up, the dissonance of any interval can be readily calculated. Figure 5.3 is redrawn from Helmholtz. The horizontal axis represents the interval between two harmonic (violin) tones. One is kept at a constant frequency labeled c', and the other is slid up an octave to c''. The height (vertical axis) of the curves is proportional to the roughness produced by the partials designated by the frequency ratios. For instance, the peaks straddling the valley at g' are formed by interactions between:

(i) The second partial of the note at g' and the third partial of c' (labeled 2:3 in the figure)

[5] Similar ideas can be found earlier in Sorge [B: 178].

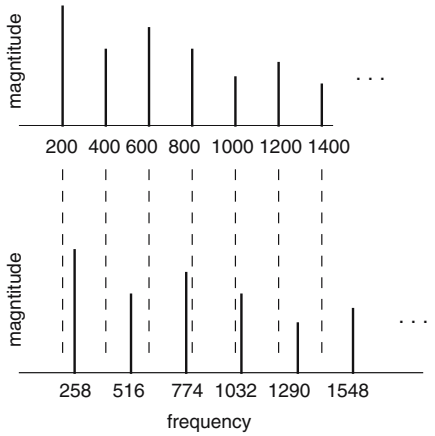


Fig. 5.2. A harmonic note at fundamental frequency $f = 200$ Hz is transposed to $g = 258$ Hz. When played simultaneously, some of the upper partials interact by beating roughly, causing sensory dissonance.

(ii) The fourth partial of the note at g' and the sixth partial of c' (labeled 4:6)
(iii) The sixth partial of the note at g' and the ninth partial of c' (labeled 6:9)

Other peaks are formed similarly by the beating of other pairs of interacting partials.

To draw these curves, Helmholtz makes three assumptions: that the spectra of the notes are harmonic, that roughnesses can be added, and that the 32 Hz beat rate gives maximal roughness. His graph has minima (intervals at which minimum beating occurs) near many of the just intervals, thus suggesting a connection between the beating and roughness of sine waves and the musical notions of consonance or dissonance. Helmholtz's work can be evaluated by comparing his conclusions with those of other notions of consonance and dissonance and by investigating his assumptions in more detail.

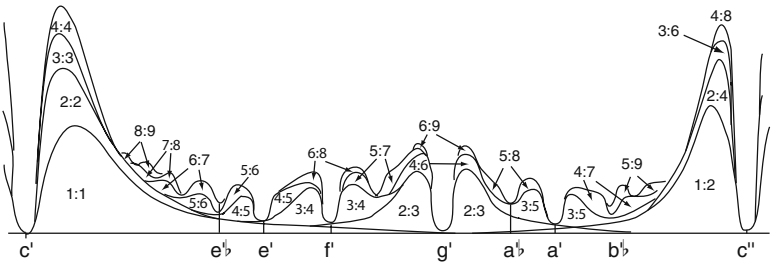


Fig. 5.3. Two pitches are sounded simultaneously. The regions of roughness due to pairs of interacting partials are plotted over one another, leaving only a few narrow valleys of relative consonance. The figure is redrawn from Helmholtz's *On the Sensation of Tone*.

For instance, does the 32 Hz beat rate for maximal roughness hold up under rigorous investigation? Do roughnesses really add up?

5.3.2 Partch's One-Footed Bride

Harry Partch was an eclectic composer and theorist who not only created a just 43-tone-per-octave musical scale, but also a family of instruments to play in this scale. In *Genesis of a Music*, Partch [B: 128] details how he tuned his chromelodeon reed organ by ear:

> To illustrate the actual mechanics of tuning, assume that the interval intended as 3/2 is slightly out of tune, so that beats are heard, perhaps two or three per second between the second partial of the "3" and the third partial of the "2" .... Hence we scratch the reed at the tip, testing continually, until the beats disappear entirely - that is, until the two pulsations are "commensurable in number" ... Experience in tuning the chromelodeon has proved conclusively that not only the ratios of 3 and 5, but also the intervals of 7, 9, and 11 are tunable by eliminating beats.

Although Partch is willing to use beats to tune his instruments, he maintains that consonance is purely a result of simple integer ratios. He states this in terms of the period of the resulting wave: The shorter the period, the more consonant the interval. This is reminiscent of Galileo, who viewed simple intervals like 3/2 as a pleasant bending exercise for the ear, but intervals like 301/200 as perpetual torment. Partch ridicules simple sine wave experiments (such as the kind used to explain sensory dissonance in the "Sound on Sound" chapter) in a section called "Obfuscation by the Moderns," although it is unclear from his writing whether he disbelieves the experimental results, or simply dislikes the conclusions reached.

However anachronistic his theoretical views, Partch was a careful listener. Using the chromelodeon, he classified and categorized all 43 intervals in terms of their comparative consonance, resulting in the "One-Footed Bride: A Graph Of Comparative Consonance," which is redrawn here as Fig. 5.4. Observe how close this is to Helmholtz's figure, although it is inverted, folded in half, and stood on end. Where Helmholtz draws a dissonant valley, Partch finds a consonant peak: All familiar JI intervals are present, and the octave, fourth, and fifth appear prominently.

In discussing the one-footed bride, Partch observes that "each consonance is a little sun in its universe, around which dissonant satellites cluster."[6] As a composer, Partch is interested in exploiting these suns and their planets. He finds four kinds of intervals: intervals of power, of suspense, of emotion, and of approach. Power intervals are the familiar perfect consonances recognized

[6] Helmholtz would claim that these dissonant clusters are caused by the beating of the same upper partials that allowed Partch to tune the instrument so accurately.

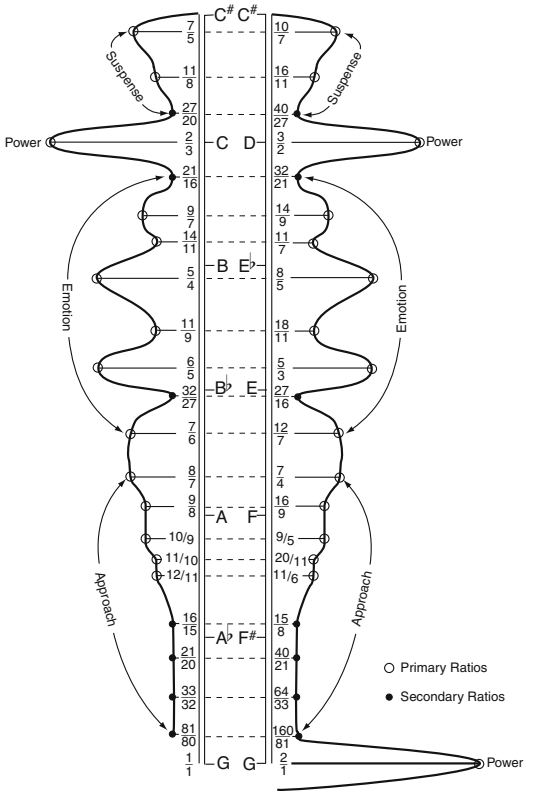


Fig. 5.4. Partch's graph of comparative consonance, the "One-Footed Bride," shows the relative consonance of each of the intervals in his 43-tone-per-octave just scale based on *G*. Four kinds of intervals are depicted: intervals of power, suspense, emotion, and approach. Figure is redrawn and used with permission [B: 128].

since antiquity. Suspenseful intervals are those between the fourth and the fifth that generalize the function of the tritone. A variety of thirds and sixths rationalize (in a literal sense) and expand on the kind of emotions normally associated with major and minor thirds and sixths. Finally, the intervals of approach are usually reserved for passing tones and melodic inflections.

Like Helmholtz, Partch observed little correlation between the notes of the 12-tet scale and the comparative consonance of the intervals. Of course, 12-tet scale steps can approximate many of the just ratios. But Partch was not a man to compromise or approximate, and he devoted his life to creating music and instruments on which to realize his vision of a just music that would not perpetually torment the ear. Fortunately, today things are much easier. Electronic keyboards can be retuned to Partch's (or any other scale) with the push of a button or the click of a mouse.

5.3.3 Harmonic Entropy

The discussion of virtual pitch (in Sect. 2.4.2) describes how the auditory system determines the pitch of a complex tone by finding a harmonic template

that lies close to the partials of the tone. If the fundamental (or root) of the template is low, then the pitch is perceived as low; if the root of the template is high, then the pitch is perceived as high. Often, however, the meaning of "closest harmonic template" is ambiguous, for instance, when there is more than one note sounding or when a single note has an inharmonic spectrum. Harmonic entropy, as introduced by Erlich [W: 9], provides a way to measure the uncertainty of the fit of a harmonic template to a complex sound spectrum. Erlich writes:

> There is a very strong propensity for the ear to try to fit what it hears into one or a small number of harmonic series, and the fundamentals of these series, even if not physically present, are either heard outright, or provide a more subtle sense of overall pitch known to musicians as the "root." As a component of consonance, the ease with which the ear/brain system can resolve the fundamental is known as "tonalness."

Entropy is a mathematical measure of disorder or uncertainty; harmonic entropy is a model of the degree of uncertainty in the perception of pitch. Tonalness is the inverse: A cluster of partials with high tonalness fits closely to a harmonic series and has low uncertainty of pitch and low entropy, and an ambiguous cluster with low tonalness has high uncertainty and hence high entropy. Recall that a single sound is more likely to fuse into one perceptual entity when the partials are harmonic. Similarly, holistic hearing of a dyad or chord as a unified single sound is strengthened when all of the partials lie close to some harmonic series.

In the simplest case, consider two harmonic tones. If the tones are to be understood as approximate harmonic overtones of some common root, they must form a simple-integer ratio with one another. One way to model this uses the Farey series $\mathcal{F}_n$, which contains all ratios of integers up to n. This series has the property that the distance between successive terms is larger when the ratios are simpler. Thus, $1/2$ and $2/3$ occupy a larger range than complex ratios such as $24/49$. For any given interval i, a probability distribution (a bell curve) can be used to associate a probability $p_j(i)$ with the ratios f_j in $\mathcal{F}_n$. The probability that the interval i is perceived as the jth member of the Farey series is high when i is close to f_j and low when i is far from f_j. The harmonic entropy (HE) of i is then defined in terms of these probabilities as

$$HE(i) = -\sum_j p_j(i) \log(p_j(i)).$$

When the interval i lies near a simple-integer ratio f_j, there will be one large probability and many small ones. Harmonic entropy is low. When the interval i is distant from any simple-integer ratio, many complex ratios contribute many nonzero probabilities. Harmonic entropy is high. A plot of harmonic entropy over a one-octave range is shown in Fig. 5.5 where the intervals are labeled in cents. Clearly, intervals that are close to simple ratios are distinguished by having low entropy values, whereas the more complex intervals have high

harmonic entropy. Details on the calculation of harmonic entropy can be found in Appendix J.

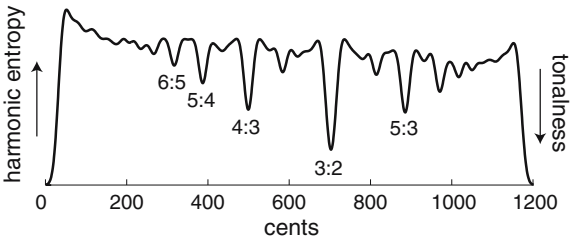


Fig. 5.5. Erlich's model of harmonic entropy predicts the tonalness (degree of certainty in the perception of the root) for various intervals. Some of the most tonal simple ratios are labeled.

5.3.4 Sensory Consonance and Critical Bandwidth

In the mid 1960s, Plomp and Levelt conducted a series of experiments on the sensations of consonance and dissonance. About 90 volunteers were asked to judge pairs of pure tones on a seven-point scale where 1 indicated the most dissonant and 7 the most consonant. The pairs were chosen so as to vary both the octave and the frequency ratios presented within the octave. The experiment was carefully devised: Each subject was tested individually, each subject only judged a few intervals so as to avoid interval recognition and fatigue, responses were tested for consistency (those who gave erratic results were discounted), and the subjects were allowed a preliminary series of intervals to familiarize them with the range of stimulus so they could make adequate use of the seven-point scale.

One of the most unique (and controversial) features of Plomp and Levelt's methodology was the use of musically untrained subjects. Previous studies had shown that musically trained listeners often recognize intervals and report their learned musical responses rather than their actual perceptions. An example is in Taylor's *Sounds of Music*, which presents Helmholtz's roughness curve along with a series of superimposed crosses that closely match the curve. These crosses are the result of a series of experiments in which sine waves were graded by subjects in terms of their harshness or roughness. As Taylor says, the close match "cannot be explained in terms of the beating of upper partials, because there are none!" However, the close match may be explainable by considering the musical background of his subjects.

To avoid such problems with learned responses, Plomp and Levelt chose to use musically naive listeners. Subjects who asked for the meaning of consonant were told *beautiful* and *euphonious*, and it can be argued that the experiment therefore tested the pleasantness of the intervals rather than the consonance. However, as most musically untrained people (and even many with training) continue to think in this CDC-2 manner, this was deemed an acceptable compromise.

Despite considerable variability among the subject's responses, there was a clear and simple trend. At unison, the consonance was maximum. As the interval increased, it was judged less and less consonant until at some point a minimum was reached. After this, the consonance increased up toward, but never quite reached, the consonance of the unison. This is exactly what we heard in sound example [S: 11] when listening to two simultaneous sine waves.

Their results can be succinctly represented in Fig. 3.7 on p. 46, which shows an averaged version of the dissonance curve (which is simply the consonance curve flipped upside-down) in which dissonance begins at zero (at the unison) increases rapidly to a maximum, and then falls back toward zero. The most surprising feature of this curve is that the musically consonant intervals are undistinguished—there is no dip in the curve at the fourth, fifth, or even the octave (in contrast to the learned response curves found by investigators like Taylor, which do show the presence of normally consonant intervals, even for intervals formed from pure sine waves).

Plomp and Levelt observed that in almost all frequency ranges, the point of maximal roughness occurred at about 1/4 of the critical bandwidth. Recall that when a sine wave excites the inner ear, it causes ripples on the basilar membrane. Two sine waves are in the same critical band if there is significant overlap of these ripples along the membrane. Plomp and Levelt's experiment suggests that this overlap is perceived as roughness or beats. Dependence of the roughness on the critical band requires a modification of Helmholtz's 32 Hz criterion for maximal roughness, because the critical bandwidth is not equally wide at all frequencies, as was shown in Fig. 3.4 on p. 44. For tones near 500 Hz, however, 1/4 of the critical band agrees well with the 32 Hz criterion.

Of course, these experiments gathered data only on perceptions of pure sine waves. To explain sensory consonance of more musical sounds, Plomp and Levelt recall that most traditional musical tones have a spectrum consisting of a root or fundamental frequency, along with a series of sine wave partials at integer multiples of the fundamental. If such a tone is sounded at various intervals, the dissonance can be calculated by adding up all of the dissonances between all pairs of partials. Carrying out these calculations for a note that contains six harmonically spaced partials leads to the curve shown in Fig. 5.6, which is taken from Plomp and Levelt [B: 141].

Observe that Fig. 5.6 contains peaks at many of the just intervals. The most consonant interval is the unison, followed closely by the octave. Next is the fifth (3:2), followed by the fourth (4:3), and then the thirds and sixths. As might be expected, the peaks do not occur at exactly the scale steps of the 12-tone equal-tempered scale. Rather, they occur at the nearby simple ratios. The rankings agree reasonably well with common practice, and they are almost indistinguishable from Helmholtz's and Partch's curves. Thus, an argument based on sensory consonance is consistent with the use of just intonation (scales based on intervals with simple integer ratios), at least for harmonic sounds.

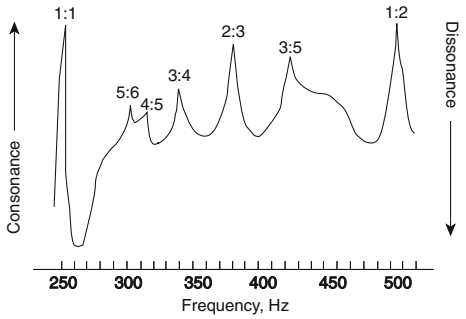


Fig. 5.6. Plomp and Levelt calculate the consonance for two tones, each with six harmonics. The first tone is fixed at a frequency of 250 Hz, and the second varies over an octave. Peaks of consonance occur at simple integer ratios of the fundamental frequency, where harmonics coincide. From Plomp and Levelt (1965).

5.4 A Simple Experiment

It is easy to experience sensory dissonance for yourself. Play a note on an organ (or some other sustained, harmonically rich sound) that is near the low end of your vocal range. While sounding the note loudly and solidly (turn off the vibrato, chorusing, and other effects), sing slightly above, slightly below, and then swoop right onto the pitch of the note. As you approach the correct pitch, you will hear your voice beating against the organ, until eventually your voice "locks into" the fundamental. It works best if you use little or no vibrato in your voice. Now repeat the experiment, but this time sing around (slightly above and slightly below) the fifth. Again, you will hear your voice beat (the second partial of your voice against the third partial of the organ) and finally lock onto the perfect fifth.

Now sing a major third above the sustained organ note, again singing slightly above and slightly below. Listen carefully to where your voice goes... does it lock onto a 12-tet third? Or does it go somewhere slightly flat? Listen carefully to the pitch of your locked-in voice. If you are truly minimizing the dissonance, then the fourth partial of your voice will lock onto the fifth partial of the organ. Assuming the organ has harmonic partials, you will be singing a just major third (a ratio of 5/4, or about 386 cents, instead of the 400 cent third in 12-tet). Can you feel how it might be tempting for a singer to synchronize in this way? By similarly exploring other intervals, you can build up your own personal dissonance curves. How do they compare with the curves of Helmholtz, Partch, and Plomp and Levelt?

In his amusing book *Lies My Music Teacher Told Me*, Eskelin [B: 54] describes this to his choir:

> If you do it slowly and steadily, you will hear the relationship between the two sounds changing as your voice slides up. It's a bit like tuning in stations on a radio dial (the old fashioned ones that had knobs to turn, not buttons to push). As you arrive at each "local station" it gradually comes into sharp focus and then fades out of focus as you go past it. What you are experiencing is called *consonance* and *dissonance*.

5.5 Summary

The words "consonance" and "dissonance" have been used in many ways throughout history, and many of these conflicting notions are still prevalent today. Psychoacoustic consonance provides a pragmatic working definition in the sense that it leads to physical correlates that can be readily measured. It is sensory dissonance that underlies the "dissonance meter" and the resulting applications of the first chapter. Although arguably the most important notion of dissonance today, sensory dissonance does not supplant previous notions. In particular, it says nothing about the important aspects of musical movement that functional consonance provides.

Helmholtz understood clearly that his roughness curve would be "very different for different qualities of tone." Partch realized that his one-footed bride would need to be modified to account for different octaves and different timbres, but he saw no hope other than "a lifetime of laboratory work." Plomp and Levelt explicitly based their consonance curve on tones with harmonic overtones. But many musical sounds do not have harmonic partials. The next chapter explores how sensory consonance can be used in inharmonic settings, gives techniques for the calculation of sensory dissonance, suggests musical uses in the relationship between spectrum and scale, and demonstrates some of the ideas and their limitations in a series of musical examples.

5.6 For Further Investigation

On the Sensations of Tone [B: 71] set an agenda for psychoacoustic research that is still in progress. Papers such as Plomp and Levelt's [B: 141] "Tonal Consonance and Critical Bandwidth" and the two-part "Consonance of Complex Tones and its Calculation Method" in Kameoka and Kuriyagawa [B: 79] and [B: 80] have expanded on and refined Helmholtz's ideas. *A History of 'Consonance' and 'Dissonance'* by Tenney [B: 192] provided much of the historical framework for the first section of this chapter, and it contains hundreds of quotes, arguments, definitions, and anecdotes. Although Partch's *Genesis of a Music* [B: 128] may not be worth reading for its contributions to psychoacoustics or to historical musicology, it is inspiring as a prophetic statement about the future of music by a musical visionary and composer.

6

Related Spectra and Scales

Sensory dissonance is a function of the interval and the spectrum of a sound. A scale and a spectrum are related *if the dissonance curve for the spectrum has minima (points of maximum sensory consonance) at the scale steps. This chapter shows how to calculate dissonance curves and gives examples that verify the perceptual validity of the calculations. Other examples demonstrate their limits. The idea of related spectra and scales unifies and gives insight into a number of previous musical and psychoacoustic investigations, and some general properties of dissonance curves are derived. Finally, the idea of the dissonance curve is extended to multiple sounds, each with its own spectrum.*

"Clearly the timbre of an instrument strongly affects what tuning and scale sound best on that instrument." W. Carlos [B: 23].

6.1 Dissonance Curves and Spectrum

Figures like Helmholtz's roughness curve and Plomp and Levelt's consonance curve (Figs. 5.3 and 5.6) on pp. 88 and 94 are called *dissonance curves* because they graphically portray the perceived consonance or dissonance versus musical intervals. Partch's one-footed bride (Fig. 5.4 on p. 90) is another, although its axis is folded about the tritone. Perhaps the most striking aspect of these harmonic dissonance curves is that many of the familiar 12-tet scale steps are close to points of minimum dissonance. The ear, history, and music practice have settled on musical scales with intervals that occur near minima of the dissonance curve.

> A spectrum and a scale are said to be *related* if the dissonance curve for that spectrum has minima at scale positions.

Looking closely, it is clear that the minima of the harmonic dissonance curves of the previous chapter do not occur at scale steps of the equal-tempered scale. Rather, they occur at the just intervals, and so harmonic spectra are related to just intonation scales.

The relatedness of scales and spectra suggests several interesting questions. Given a spectrum, what is the related scale? Given a scale, what are the related spectra? How can spectrum/scale combinations be realized in existing electronic musical instruments? What is it like to play inharmonic sounds in unfamiliar tunings?

6.1.1 From Spectrum to Tuning

Because dissonance curves are drawn for a particular spectrum (a particular set of partials), they change shape if the spectrum is changed: Minima appear and disappear, and peaks rise and fall. Thus, given an arbitrary spectrum, perhaps one whose partials do not form a standard harmonic series, this chapter explores how to draw its dissonance curve. The minima of this curve occur at intervals that are good candidates for notes of a scale, because they are intervals of minimum dissonance (or, equivalently, intervals of maximum consonance).

The crucial observation is that these techniques allow precise control over the perceived (sensory) dissonance. Although most statements are made in terms of maximizing consonance (or of minimizing dissonance), the real strength of the approach is that it allows freedom to sculpt sounds and tunings so as to achieve a desired effect. Sensory consonance and dissonance can be used to provide a perceptual pathway helpful in navigating unknown inharmonic musical spaces.

The idea of relating spectra and scales is useful to the electronic musician who wants precise control over the amount of perceived dissonance in a musical passage. For instance, inharmonic sounds are often extremely dissonant when played in the standard 12-tet tuning. By adjusting the intervals of the scale, it is often possible to reduce (more properly, to have control over) the amount of perceived dissonance. It can also be useful to the experimental musician or the instrument builder. Imagine being in the process of creating a new instrument with an unusual (i.e., inharmonic) tonal quality. How should the instrument be tuned? To what scale should the finger holes (or frets, or whatever) be tuned? The correlation between spectrum and scale answers these question in a concrete way.

6.1.2 From Tuning to Spectrum

Alternatively, given a desired scale (perhaps a favorite historical scale, one that divides the octave into n equal pieces, or one that is not even based on the octave), there are spectra that will generate a dissonance curve with minima at precisely the scale steps. Such spectra are useful to musicians and composers wishing to play in nonstandard scales such as 10-tet, or in specially fabricated scales. How to specify such spectra, given a desired scale, is the subject of the chapter “From Tuning to Spectrum.”

6.1.3 Realization and Performance

All of this would be no more than fanciful musings if there was no way to concretely realize inharmonic spectra in their related tunings. The next chapter "A Bell, A Rock, A Crystal" gives three examples of how to find the spectrum of an inharmonic sound, draw the dissonance curve, map the sound to a keyboard, and play. The process is described in excruciating detail to help interested readers pursue their own inharmonic musical universes.

6.2 Drawing Dissonance Curves

The first step is to encapsulate Plomp and Levelt's curve for pure sine waves into a mathematical formula. The curve is a function of two pure sine waves each with a specified loudness. Representing the height of the curve at each point by the letter d, the relationship can be expressed as:

$$d(f_1, f_2, \ell_1, \ell_2), \text{ where } \begin{array}{l} f_1 \text{ is the frequency of the lower sine} \\ f_2 \text{ is the frequency of the higher sine} \\ \ell_1 \text{ and } \ell_2 \text{ are the corresponding loudnesses} \end{array}$$

A functional equation using exponentials is detailed in Appendix E, and the mathematically literate reader may wish to digress to this appendix for a formal definition of the function d and of dissonance curves. But it is not really necessary. Simply keep in mind that the function $d(\cdot, \cdot, \cdot, \cdot)$ contains the same information as Fig. 3.8 on p. 47.

When there are more than two sine waves occurring simultaneously, it is possible to add all dissonances that occur. Suppose the note F has three partials at f_1, f_2, and f_3, with loudnesses ℓ_1, ℓ_2, and ℓ_3. Then the intrinsic or inherent dissonance D_F is the sum of all dissonances between all partials. Thus D_F is the sum of $d(f_i, f_j, \ell_i, \ell_j)$ as i and j take on all possible values from 1 to 3. Although it is not the major point of the demonstration, you can hear sounds with varying degrees of intrinsic consonance by listening holistically to sound example [S: 54]. The initial sound is dissonant, and it is smoothly changed into a more consonant sound.

The same idea can be used to find the dissonance when the spectrum F is played at some interval c. For instance, suppose F has two partials f_1 and f_2. The complete sound contains four sine waves: at f_1, f_2, cf_1, and cf_2. The dissonance of the interval is the sum of all possible dissonances among these four waves. First is the intrinsic dissonances of the notes $D_F = d(f_1, f_2, \ell_1, \ell_2)$ and $D_{cF} = d(cf_1, cf_2, \ell_1, \ell_2)$. Next are the dissonances between cf_1 and the two partials of F, $d(f_1, cf_1, \ell_1, \ell_1)$ and $d(f_2, cf_1, \ell_2, \ell_1)$, and finally the dissonances between cf_2 and the partials of F, $d(f_1, cf_2, \ell_1, \ell_2)$ and $d(f_2, cf_2, \ell_2, \ell_2)$. Adding all of these terms together gives the dissonance of F at the interval c, which we write $D_F(c)$. The dissonance curve of the spectrum F is then a plot of this function $D_F(c)$ over all intervals c of interest.

If you are thinking that there are a lot of calculations necessary to draw dissonance curves, you are right. It is an ideal job for a computer. In fact, the most useful part of this whole mathematical parameterization is that it is now possible to calculate the dissonance of a collection of partials automatically. Those familiar with the computer languages BASIC or `Matlab` will find programs for the calculation of dissonance on the CD and discussions of the programs in Appendix E.[1]

For example, running either of the programs from Appendix E without changing the frequency and loudness data generates the dissonance curve for a sound with fundamental at 500 Hz containing six harmonic partials. This is shown in Fig. 6.1 and can be readily compared with Helmholtz's, Plomp and Levelt's, and Partch's curves (Figs. 5.3, 5.4, and 5.6 on pp. 88, 90, and 94).

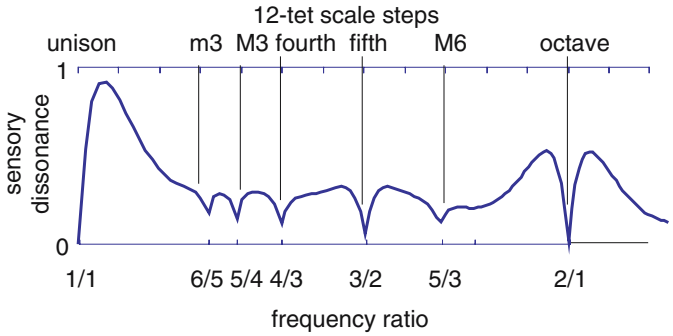


Fig. 6.1. Dissonance curve for a spectrum with fundamental at 500 Hz and six harmonic partials has minima that coincide with many steps of the Just Intonation scale and that coincide approximately with 12-tet scale steps, which are shown above for comparison.

Table 6.1 provides a detailed comparison among the 12-tet scale steps, the just intonation major scale, and the minima of the dissonance curve drawn for a harmonic timbre with nine partials. The JI intervals are similar to the locations of the minima of the dissonance curve. In particular, the minima agree with the septimal scales of Partch [B: 128] for seconds, tritones, and the minor seventh, but with the JI major scale for the major seventh. Minima occur at both the septimal and the just thirds.

One assumption underlying dissonance curves such as Fig. 6.1 is additivity, the assumption that the sensory dissonance of a collection of sine partials is the sum of the dissonances between all pairwise partials. Although this assumption generally holds as a first approximation, it is easy to construct examples where it fails. Following Erlich [W: 9], consider a sound with ratios

[1] A FORTRAN version, along with an alternative parameterization of the Plomp–Levelt curves can also be found in [B: 92].

Table 6.1. Notes of the equal-tempered musical scale compared with minima of the dissonance curve for a nine-partial harmonic timbre, and compared with the just intonation major scale from [B: 207]. Septimal (sept.) scale values from [B: 128].

Note Name	12-tet $r = \sqrt[12]{2}$	Minima of dissonance curve	Just Intonation	
C	$r^0 = 1$	1	1:1	unison
C♯	$r^1 = 1.059$		16:15	just semitone
D	$r^2 = 1.122$	1.14 (8:7 = sept. maj. 2)	9:8	just whole tone
E♭	$r^3 = 1.189$	1.17 (7:6 = sept. min 3)		
		1.2 (6:5)	6:5	just min. 3
E	$r^4 = 1.260$	1.25 (5:4)	5:4	just maj. 3
F	$r^5 = 1.335$	1.33 (4:3)	4:3	just perfect 4
F♯	$r^6 = 1.414$	1.4 (7:5 = sept. tritone)	45:32	just tritone
G	$r^7 = 1.498$	1.5 (3:2)	3:2	perfect 5
A♭	$r^8 = 1.587$	1.6 (8:5)	8:5	just min. 6
A	$r^9 = 1.682$	1.67 (5:3)	5:3	just maj. 6
B♭	$r^{10} = 1.782$	1.75 (7:4 = sept. min. 7)	16:9	just min. 7
B	$r^{11} = 1.888$	1.8 (9:5 = just min. 7)	15:8	just maj. 7
C	$r^{12} = 2$	2.0	2:1	octave

4:5:6:7 (this can be heard in sound example [S: 40]) and an inharmonic sound with ratios 1/7:1/6:1/5:1/4 (as in sound example [S: 41]). Both sounds have the same intervals,[2] and hence, the sensory dissonance is the same. Yet they do not sound equally consonant. Sound example [S: 42] alternates between the harmonic and inharmonic sounds, and most listeners find the harmonic sound more consonant. Thus, dissonance cannot be fully characterized as a function of the intervals alone without (at least) considering their arrangement. Accordingly, sensory dissonance alone is insufficient to fully characterize dissonance. In this case, the sound with greater tonalness (smaller harmonic entropy) is judged more consonant than the sound with lesser tonalness (greater harmonic entropy).

6.3 A Consonant Tritone

Imagine a spectrum consisting of two inharmonic partials at frequencies f and $\sqrt{2}f$. Because the $\sqrt{2}$ interval defines a tritone (also called a diminished fifth or augmented fourth in 12-tet), this is called the *tritone spectrum*. The dissonance curve for the tritone spectrum, shown in Fig. 6.2, begins with a minimum at unison, rapidly climbs to its maximum, then slowly decreases

[2] To be specific, the 4:5:6:7 sound example consists of sine waves at 400, 500, 600, and 700 Hz and contains the intervals 5/4, 3/2, 7/4, 6/5, 7/5, and 7/6. The inharmonic sound is made from sine waves at 400, 467, 560, and 700 Hz, and has the same intervals. Similar results appear to hold for harmonic sounds.

until, just before the tritone, it rises and then falls. There is a sharp minimum right at the tritone, followed by another steep rise. For larger intervals, dissonance gradually dies away. You can verify for yourself by listening to sound example [S: 35] that the perceived dissonance corresponds more or less with this calculated curve. Video example [V: 9] reinforces the same conclusion. Thus, the dissonance curve does portray perceptions of simple sweeping sounds fairly accurately. But it is not necessarily obvious what (if anything) such tests mean for more musical sounds, in more musical situations.

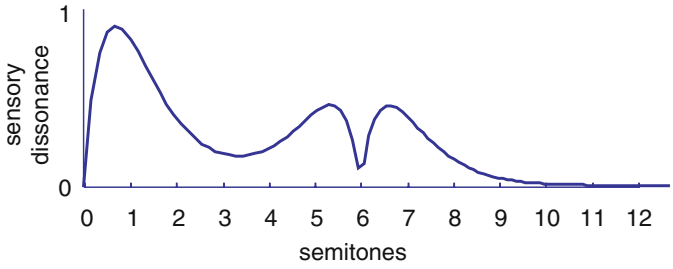


Fig. 6.2. Dissonance curve for an inharmonic spectrum with partials at f and $\sqrt{2}f$ has minima at 1.21 (between 3 and 4 semitones) and at 1.414, which is a tritone.

Sounds used in music are not just static sets of partials: they have attack, decay, vibrato, and a host of other subtle features. A more "musical" version of the tritone spectrum should mimic at least some of these characteristics. The "tritone chime" has the same tritone spectrum but with an envelope that mimics a softly struck bell or chime, and a bit of vibrato and reverberation. This chime will be used in the next two sound examples to verify the predictions of the dissonance curve.

Both the fifth (an interval of seven semitones) and the fourth (five semitones) lie near peaks of the tritone dissonance curve. Thus, the dissonance curve predicts that a chord containing both a fourth and a fifth should be more dissonant than a chord containing two tritones, at least when played with this timbre. To see if this prediction corresponds to reality, sound example [S: 36] begins with a single note of the tritone chime. It is "electronic" sounding, somewhat percussive and thin, but not devoid of all musical character. The example then plays the three chords of Fig. 6.3. The chords are then repeated using a more "organ-like" sound, also composed from the tritone spectrum. In both cases, the chords containing tritones are far more consonant than chords containing the dissonant fifths and fourths. The predictions of the dissonance curve are upheld. This demonstration is repeated somewhat more graphically in video example [V: 10].

But still, sound example [S: 36] deals with isolated chords, devoid of meaningful context. Observe that there is a broad, shallow minimum around 1.21,

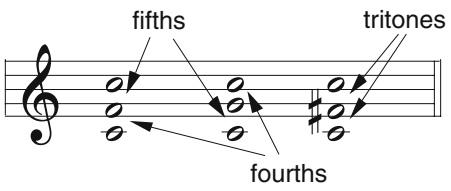


Fig. 6.3. Familiar intervals such as the fifth and fourth are dissonant when played using the "tritone chime." But chords containing tritones are consonant.

approximately a minor third. This suggests that the minor third is more consonant than the major third. Combined with the consonance of the tritone, this implies that a diminished chord (root, minor third, and tritone) should be more consonant than a major chord (root, major third, and fifth) when played with the tritone sound. Is this inversion of normal musical usage possible? Listen to sound example [S: 37], which places the tritone chime into a simple musical setting. The following two chord patterns are each repeated once:

(a) F major, C major, G major, C major
(b) C dim, D dim, $D\sharp$ dim, D dim

This is shown in musical notation in parts (a) and (b) of Fig. 6.4, where "dim" is an abbreviation for "diminished." Both patterns are played with the same simple chordal rhythm, but there is a dramatic difference in the sound. The major progression, which, when played with "normal" harmonic tones would sound completely familiar, is dissonant and bizarre. The diminished progression, which in harmonic sounds would be restless, is smooth and easy. The inharmonic tritone chime is capable of supporting chord progressions, although familiar musical usage is upended.

The final two tritone chime chord patterns, shown in (c) and (d) of Fig. 6.4, investigate feelings of resolution or finality. To my ears, (d) feels more settled, more conclusive than (c). Perhaps it is the dissonance of the major chord that

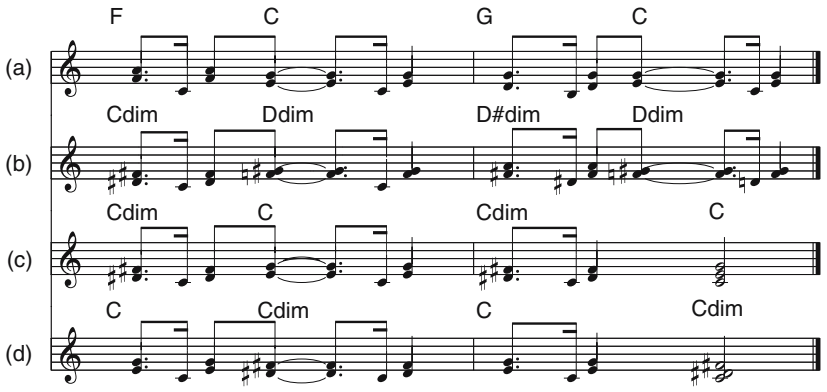


Fig. 6.4. Chord patterns using the tritone chime sound.

causes it to want to move, and the relative restfulness of the diminished chord that makes it feel more resolved. Essentially, the roles of the fifth and the tritone have been reversed. With harmonic sounds, the tritone leads into a restful fifth. With tritonic spectra, the fifth leads into a tranquil tritone.

Observe: We began by pursuing sensory notions of dissonance because it provided a readily measured perceptual correlate. Despite this, it is now clear (in some cases, at least) that sensory dissonance is linked with functional dissonance, the more musical notion, in which the restlessness, motion, and desire of a chord to resolve play a key role. Even in this simple two-partial inharmonic sound, chords with increased (sensory) dissonance demand resolution, whereas chords with lower (sensory) dissonance are more stable.

This two-partial tritone sound is not intended to be genuinely musical, because the tone quality is simplistic. The purpose of the examples is to demonstrate in the simplest possible inharmonic setting that ideas of musical motion, resolution, and chord progressions can make sense. Of course, the "rules" of musical grammar may be completely different in inharmonic musical universes (where major chords can be more dissonant than diminished, and where tritones can be more consonant than fifths), but there are analogies of chord patterns and strange inharmonic "harmonies." These are *xentonal*: Unusual tonalities that are not possible with harmonic sounds.

6.4 Past Explorations

As the opening quote of this chapter indicates, this is not the first time that the relationship between timbre and scale has been investigated, although it is the first time it has been explored in such a general setting. Pierce and his colleagues are major explorers of the connection between sound quality and tonality.

6.4.1 Pierce's Octotonic Spectrum

Shortly after the publication of Plomp and Levelt's article, Pierce [B: 134] used a computer to synthesize a sound designed specifically to be played in an eight-tone equal-tempered (8-tet) scale, to demonstrate that it was possible to attain consonance in "arbitrary" scales. Letting $r = \sqrt[8]{2}$, an *octotonic* spectrum can be defined[3] by partials at

$$1, r^{10}, r^{16}, r^{20}, r^{22}, r^{24}.$$

In the same way that 12-tet divides evenly into two interwoven whole-tone (6-tet) scales, the 8-tet scale can be thought of as two interwoven 4-tet scales, one containing the even-numbered scale steps and the other consisting of the

[3] Beware of a typo in Table 1 of [B: 134]: the frequency ratio of the second partial should be $r^{10} = 2.378$.

odd scale steps. As the partials of Pierce's octotonic spectrum fall on even multiples of the eighth root of two, the even notes of the scale form consonant pairs and the odd notes form consonant pairs, but they are dissonant when even and odd steps are sounded together.

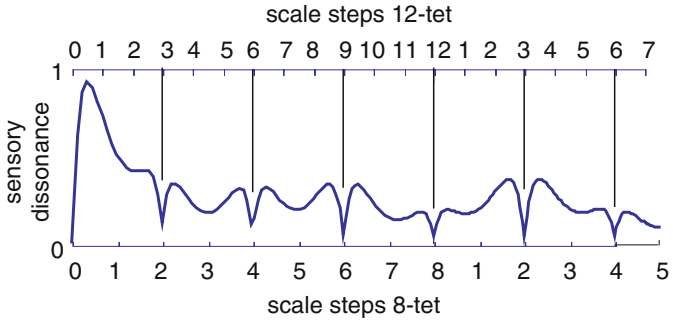


Fig. 6.5. Dissonance curve for Pierce's octotonic spectrum designed for play in the 8-tet scale. Minima occur at even steps of the 8-tet scale. The 12-tet scale steps are placed above for comparison. Every third step in 12-tet is the same as every second step in 8-tet.

This can be seen directly from the dissonance curve, which is shown in Fig. 6.5. The curve has minima at all even scale steps, implying that these intervals are consonant when sounded together. Although he does not give details, Pierce says "listeners report" that notes separated by an even number of scale steps are more concordant than notes separated by an odd number of scale steps.

The scale related to the octotonic spectrum consists of those scale steps at which minima occur. These are at ratios $1, r^2, r^4, r^6$, and r^8. Although this scale may appear completely foreign at first glance, observe how it lines up exactly with scale steps $1, 3, 6, 9$, and 12 of the 12-tet scale,[4] which is plotted above for handy reference. Thus, the primary consonant intervals in this octotonic scale are identical to the familiar minor third, tritone, and major sixth, and the octotonic spectrum is a close cousin of the tritone spectrum of the previous section. Again, conventional music theory has been upended, with consonant tritones and dissonant fifths, consonant diminished chords, and dissonant major chords.

To perform using Pierce's octatonic spectrum, I created a sound with the specified partials in which the loudnesses died away at an exponential rate of 0.9. A percussive envelope and a bit of vibrato help make it feel more like a natural instrument. First, I played in 12-tet. As expected, the tritones were far

[4] Using t to represent the 12-tet interval ratio $\sqrt[12]{2}$, this lining up occurs because $r^2 = t^3$, $r^4 = t^6$, $r^6 = t^9$, and of course $r^8 = t^{12} = 2$.

more consonant than the fifths, and the diminished chords were very smooth. Retuning the keyboard to 8-tet, the same diminished chords are present. In fact, that's all there is! In 8-tet with the octotonic spectrum, all even scale steps form one big diminished seventh chord (but a very consonant diminished seventh) and all odd scale steps form another diminished seventh. In a certain sense, music theory is very simple in this 8-tet setting: There are "even" chords and there are "odd" chords.[5] There are no major or minor chords, no leading tones, and no blues progressions–just back and forth between two big consonant diminished sevenths. Of course, related spectra and scales will not always lead to such readily comprehensible musical universes.

Pierce concludes on an upbeat note that, "by providing music with tones having accurately specified but inharmonic partials, the digital computer can release music from the tyranny of 12 tones without throwing consonance overboard."

6.4.2 Stretching Out

"Inharmonic" is as precise a description of a sound spectrum as "nonpink" is of light. As there are so many kinds of inharmonicity, it makes sense to start with sounds that are somehow "close to" familiar sounds. Recalling that the partials of a piano are typically stretched away from exact harmonicity (see Young [B: 208]), Slaymaker [B: 176] investigated spectra with varying amounts of stretch. The formula for the partials of harmonic sounds can be written $f_j = jf = f2^{\log_2(j)}$ for integers j. By replacing the 2 with some other number S, Slaymaker created families of sounds with partials at

$$f_j = fS^{\log_2(j)}.$$

When $S < 2$, the frequencies of the partials are squished closer together than in harmonic sounds, and the tone is said to be *compressed*. When $S > 2$, the partials are spread out like the bellows of an accordion, and the tone is *stretched* by the factor S. The most striking aspect of compressed and stretched spectra is that none of the partials occur at the octave. Rather, they line up at the stretched octave, as shown in Fig. 6.6. In the same way that the octave of a harmonic tone is smooth because the partials coincide, so the *pseudo-octave* of the stretched sound is smooth due to coinciding partials.

This is also readily apparent from the dissonance curves, which are plotted in Fig. 6.7 for stretch factors $S = 1.87$ (the pseudo-octave compressed to a seventh), $S = 2.0$ (normal harmonic tones and octaves), $S = 2.1$ (the pseudo-octave stretched by about a semitone), and $S = 2.2$ (the pseudo-octave stretched to a major 9th). In each case, the frequency ratio S is a pseudo-octave that plays a role analogous to the octave. Real 2:1 ratio octaves sound dissonant and unresolved when S is significantly different from 2, whereas

[5] Although even the even chords are decidedly odd.

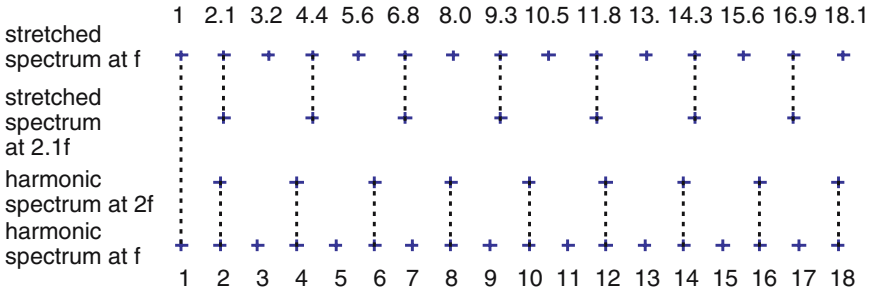


Fig. 6.6. Locations of partials are shown for four spectra. The partials of the 2.1 stretched spectrum at fundamental f have the same relationship to its 2.1 pseudo-octave (at fundamental $2.1f$) as the partials of the harmonic spectrum at fundamental f have to the octave at fundamental $2f$.

the pseudo-octaves are nicely consonant. This is where the "challenging the octave" sound example from the first chapter came from. A stretched sound with $S = 2.1$ was played in a 2.0 octave, which is dreadfully dissonant, as suggested by the lower left of Fig. 6.7. When played in its pseudo-octave, however, it is consonant.

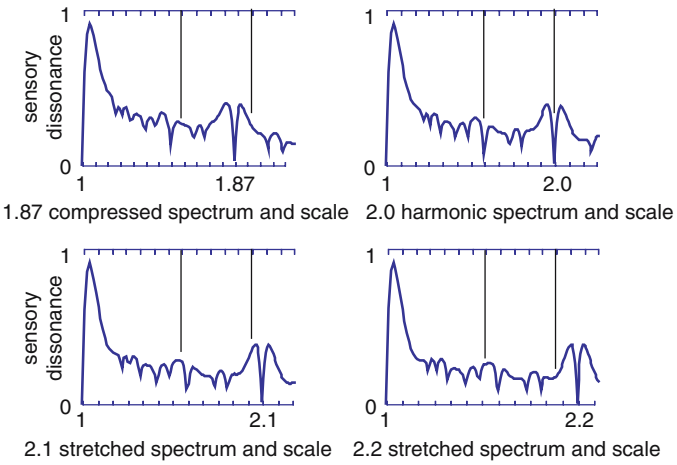


Fig. 6.7. Dissonance curves generated by stretched (and compressed) spectra have the same contour as the harmonic dissonance curve, but minima are stretched (or compressed) so that pseudo-octaves, pseudo-fifths, and so on, are clearly visible. The bottom axis shows 12 equal divisions of the pseudo-octave, and the top axis shows the standard 12-tet scale steps. Tick marks for the octave (frequency ratio of 2) and the fifth (frequency ratio 3/2) are extended for easy comparison. As usual, the dissonance axis is normalized.

Each of the curves in Fig. 6.7 has a similar contour, and minima of the dissonance curve occur at (or near) the 12 equal steps of the pseudo-octaves. A complete pseudo-just intonation of pseudo-fifths, pseudo-fourths, and pseudo-thirds is readily discernible, suggesting the possibility that music theory and practice can be transferred to compressed and stretched spectra, when played in compressed and stretched scales.

Is Stretched Music Viable?

There is a fascinating demonstration on the *Auditory Demonstrations CD* [D: 21] in which a four-part Bach chorale is played four ways:

(i) A harmonic spectrum in the unstretched 12-tet scale
(ii) A 2.1 stretched spectrum in the 2.1 stretched scale
(iii) A harmonic spectrum in the 2.1 stretched scale
(iv) A 2.1 stretched spectrum in the unstretched 12-tet scale

The first is normal sounding, if somewhat bland due to the simplicity of the nine partial "electric piano" timbre. The second version has no less sensory consonance, a result expected because all notes occur near minima of the dissonance curve. But the tone quality is decidedly strange. It is not easy to tell how many tones are sounding, especially in the inner voices. The notes have begun to lose tonal fusion. Although the sensory dissonance has not increased from (i), the tonalness aspect of dissonance has increased. The third and fourth versions are clangorous and dissonant in a spectacular way–like the extended versions of the "challenging the octave" demonstrations in sound examples [S: 2] to [S: 5].

Several experiments have investigated the uses and limitations of stretched tones in semimusical contexts. Mathews and Pierce [B: 100] tested subjects' ability to determine the musical key and the "finality" of cadences when played with stretched timbres. Three simple musical passages X, M, and T, were played in sequence $XMXT$, and subjects were asked to judge whether X was in the same key as M and T. Both musicians and nonmusicians were able to answer correctly most of the time. But when subjects were asked to rate the "finality" of a cadence and an anti-cadence, the stretched versions were heard as equally (not very) final. Mathews and Pierce observe that melody is more robust to stretching than harmony, and they suggest that the subjects in the key determination experiment may have used the melody to determine key rather than the chordal motion. The stretch factor used in these experiments was $S = 2.4$, which is well beyond where notes typically lose fusion. Thus, one aspect of musical perception (the finality of cadences) requires the fusion of tones, even though fusion may not be critical for others such as a sense of the "melody" of a passage. An alternative explanation is that notes of a melodic passage may fuse more readily when they are the focus of attention.

Perhaps the most careful examination of stretched intervals is the work of Cohen [B: 33], who asked subjects to tune octaves and fifths for a variety of

sounds with stretched spectra ranging from $S = 1.4$ to $S = 3.0$. Cohen observed two different tuning strategies: interval memory and partial matching. Some subjects consistently tuned the adjustable tone to an internal model or template of the interval, and they were able to tune to real octaves and fifths, despite the contradictory spectral clues. Others pursued a strategy of matching the partials of the adjustable tone to those of the fixed tone, leading to a consistent identification of the pseudo-octave rather than the true octave.

Plastic City: A Stretched Journey

In talking about Pierce's work on stretched tunings, Moore [B: 117] observes that Pierce uses traditional music, rather than music specifically composed around properties of the new sounds. Taking this as a challenge, I decided to hear for myself. First, I created about a dozen sets of sounds via additive synthesis[6] with partials stretched from $S = 1.5$ to $S = 3.0$. As expected, those with extreme stretching lost fusion easily, so I chose four sets of moderately stretched and compressed tones (with $S = 2.2,\ 2.1,\ 2.0$, and 1.87) that sounded more or less musical. When generating these sounds,[7] and when using the keyboard to add performance parameters such as attack and decay envelopes, vibrato, and so on, I was careful to keep the sounds strictly comparable: If I added vibrato or reverb to one sound, I added the same amount of vibrato or reverb to each of the other sounds. In this way, fair comparisons should be possible.

The resulting experiment, called *Plastic City*, can be heard in sound example [S: 38]. The structure of the piece is simple: The theme is played with harmonic tones (in standard 12-tet), then with the 2.2 stretched tones, then with the compressed 1.87 tones, and finally with the 2.1 stretched tones (each in their respective stretched scales, of course). The theme is based on a simple I V IV V pattern followed by I V I. It is unabashedly diatonic and has a clear sense of harmonic motion and resolution. The theme is repeated with each sound, and the second time a lead voice solos. At the end of the repeat, the theme disintegrates and scatters, making way for the next tuning.

> Now stop reading. Listen to *Plastic City* (sound example [S: 38] in the file `plasticity.mp3`), and make up your own mind about what parts work and what parts do not.

Most people find the entrance of the 2.2 tone extremely bizarre. Then, just as the ear is about to recover, the compressed tone begins a new kind of uneasiness. Finally, the entrance of the 2.1 tone is like a breath of fresh air after a torturous journey. The most common comment I have heard (besides

[6] Appendix D contains a discussion of additive synthesis.

[7] The sounds used in *Plastic City* contained between five and ten partials, with a variety of amplitudes with primarily percussive envelopes.

a sigh of relief) is that "now we're back to normal." But 2.1 stretched is really very far from normal—it contains no octaves, no fifths, no recognizable intervals at all. The octaves are out-of-tune by almost a semitone. This is the same amount of stretch used on the *Auditory Demonstrations* CD [D: 21] to show the loss of fusion with stretched tones. Yet in this context, 2.1 stretched can be heard as "back to normal."

Thus, 2.2 is stretched a bit too far, and 1.87 is squished a bit too much. The kinds of things you hear in *Plastic City* are typical of what happens when tones fission. It becomes unclear exactly how many parts are playing. It is hard to focus attention on the melody and to place the remaining sounds into the background. Chordal motion becomes harder to fathom. Of course, this piece is structured so as to "help out" the ear by foreshadowing using normal harmonic sounds. Thus, it is more obvious what to listen for, and by focusing attention, the "same" piece can be heard in the stretched and compressed versions, but it takes an act of will (and/or repeated listenings) before this occurs.

Perhaps the 2.1 version only sounds good in this context because the ear has been tortured by the overstretching and undercompressing. Sound example [S: 39], called *October 21st*, is a short piece exclusively in 2.1 stretched. The timbres are the same as used in *Plastic City* and in [S: 4], and here they sound bright, brilliant, and cheerful. The motion of the chord patterns is simple, and it is not difficult to perceive. Torture is not a necessary precondition to make stretched tones sound musical. Perhaps the most interesting aspect of this piece is its familiarity. I have played this for numerous people, and many hear nothing unusual at all.

What does it mean when a sound has been stretched or compressed "too far?" Perhaps the most obvious explanation is loss of fusion; that is, it is no longer heard as a single complex sound but as two or more simpler sounds. A closely related possibility is loss of tonal integrity; that is, the uncertainty in the (virtual) pitch mechanism has become too great. In the first case, the sound appears to bifurcate from one sound into two, whereas in the latter case, it appears to have a pitch that is noticeably higher (for stretched sounds) than the dominant lowest partial. Cohen's experiments [B: 33] are relevant, but it is not obvious how to design an experiment that clearly distinguishes these two hypotheses.

Moving beyond stretched versions of the 12-tet scale, it is not always possible to correlate inharmonic spectra and their related scales with standard music theory. The next example shows how a simple class of sounds (those with odd-numbered partials) can lead to a nonintuitive tuning based on 13 equal divisions of the "tritave" rather than 12 equal divisions of the octave.

6.4.3 The Bohlen–Pierce Scale

Pan flutes and clarinets (and other instruments that act like tubes open at a single end) have a spectrum in which odd harmonics predominate. For in-

stance, Fig. 6.8 shows the spectrum of a pan flute with fundamental frequency $f = 440$ Hz and prominent partials at about $3f$, $5f$, $7f$, and $9f$. Recall that the just intonation approach exploited ratios of the first few partials of harmonic tones to form the "pure" intervals such as the fifth, fourth, and thirds. A generalized just intonation approach to sounds with only odd partials would similarly exploit ratios of small odd numbers, such as 9/7, 7/5, 5/3, 9/5, 7/3, and 3/1.

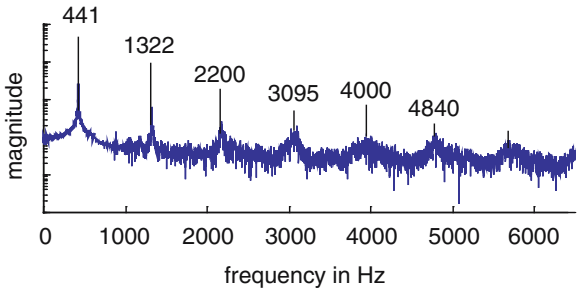


Fig. 6.8. Some instruments have spectra that consist primarily of odd-numbered partials. This pan flute has a fundamental at $f = 441$ Hz and prominent partials at (approximately) $3f$, $5f$, $7f$, $9f$, and $11f$.

Mathews and Pierce[8] observed that these ratios can be closely approximated by steps of a scale built from 13 equal divisions of the ratio 3/1 (the *tritave*). The most promising of these scales,[9] which they call the *Bohlen–Pierce* scale, contains nine notes within a tritave. Recall that when a harmonic sound is combined with its octave, no new frequency components are added, as was shown in Fig. 4.1. For spectra with only odd partials, however, the addition of an octave does add new components (the even partials), but the addition of a tritave does not. Thus, the tritave plays some of the same roles for spectra with odd partials that the octave plays for harmonic tones.

Mathews and Pierce analyze many of the possible chords in the tritave-based Bohlen–Pierce scale in the hope of determining if viable music is possible. Chords built from scale steps 0, 6, and 10 are somewhat analogous to major chords, and those built from 0, 4, and 10 have a somewhat minor flavor. When musicians and nonmusicians are asked to judge the consonance of the various chords, some interesting discrepancies originate. Naive listeners tend to judge the consonance of the chords more or less as indicated by the Plomp–Levelt models (i.e., to agree with the predictions of the dissonance curve). But musically sophisticated listeners judge some of the chords more dissonant than expected. On closer inspection, Mathews and Pierce found that these chords

[8] [B: 102], and see also Bohlen [B: 16].
[9] Built on steps 0, 1, 3, 4, 6, 7, 9, 10, and 12.

contained close (but not exact) approximations to standard 12-tet intervals. Thus, the musically trained subjects heard a familiar interval out of kilter, rather than an unfamiliar interval in tune. Recall that Plomp and Levelt had similar problems with highly trained musical subjects whose judgments of intonation were often based on their training rather than on what they heard.

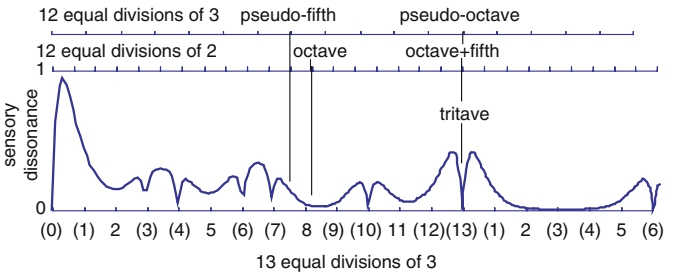


Fig. 6.9. Dissonance curve for the panflute spectrum with odd integer partials at f, $3f$, $5f$, $7f$, and $9f$. The bottom axis shows steps of the Bohlen–Pierce scale in parentheses, which are a subset of 13 equal divisions of 3. Observe how steps 3, 4, 6, 7, 10, and 13 occur at or near sharp minima of the dissonance curve. The top axes shows the familiar 12-tet scale steps as well as the $S = 3$ stretched scale.

Figure 6.9 shows the dissonance curve for spectra with odd partials such as the pan flute. Observe that the curve has many minima aligned with the Bohlen–Pierce scale: at steps 3, 4, 6, 7, 10, and 13. The tritave is very consonant, and all the intervals of the "major" and "minor" chords proposed by Mathews and Pierce (and their inversions) appear convincingly among the deepest of the minima. To facilitate comparison with previous scales, two additional axes appear at the top of the diagram. Note that the tritave is equal to a standard octave plus a fifth, but that virtually none of the other 12-tet scale steps occur near minima of the dissonance curve. Also, compare the Bohlen–Pierce tritave scale and the stretched scale with stretch factor $S = 3$. Although the pseudo-octave of the stretched scale is identical to the tritave, none of the other stretched scale steps coincide closely with minima.[10] Thus, the Bohlen–Pierce scale really is fundamentally different, and it requires a fundamentally new music theory. Unlike the tritone spectrum in 8-tet, this theory is not trivial or obvious. Three exploratory compositions in the Bohlen–Pierce scale can be heard on the CD accompanying *Current Directions in Computer Music Research* [B: 103].

[10] Stretched scales and spectra are fundamentally different from the Bohlen–Pierce scale and spectra with odd integer partials. A $S = 3$ stretched spectrum, for instance, has partials at f, $3f$, $5.7f$, $9f$, $12.8f$, etc.

6.5 Found Sounds

Each of the previous examples began with a mathematically constructed spectrum (the tritone spectrum, the octatonic spectrum, stretched spectra, spectra with odd partials) and explored a set of intervals that could be expected to sound consonant when played with that spectrum. The dissonance curve provides a useful simplifying tool by graphically displaying the most important intervals, which together form the scale steps. Each of the previous examples had a clear conceptual underpinning. But mathematical constructions are not necessary—the only concept needed is the sound itself.

McLaren [B: 107] is well aware of the need to match the spectrum with the scale, "Just scales are ideal for instruments that generate lots of harmonic partials" but when the instruments have inharmonic partials, the solution is to use "non-just non-equal-tempered scales whose members are irrational ratios of one another... [to] better fit with the irrational partials of most... instruments." Found sounds:

> remain one of the richest sources of musical scales in the real world. Anyone who has tapped resistor heat sinks or struck the edges of empty flower pots realizes the musical value of these scales... [11]

This section suggests approaches to tunings for "found" objects or other sounds with essentially arbitrary spectra. In this respect, dissonance curves can be viewed as a formalization of a graphical technique for combining sounds first presented by Carlos. Two concrete examples are worked out in complete detail.

6.5.1 Carlos' Graphical Method

The quote at the start of this chapter is taken from the article "Tuning: At the Crossroads" by Carlos [B: 23], which contains an example showing how the consonance of an interval is dependent on the spectrum of the instrument. Carlos contrasts a harmonic horn with an electronically produced inharmonic "instrument" called the *gam* with both played in octaves and in stretched octaves. The gam sounds more consonant in the pseudo-octave, and the horn sounds most consonant in the real octave. This is presented on the sound sheet (recording) that accompanies the article, and it is explained in graphical form.

Carlos' graphical method can be applied to almost any sound. Consider a struck metal bar, and recall that the bending modes (partials) are inharmonically related. This was demonstrated in Fig. 2.8 on p. 25, which shows the partials diagrammatically. When several metal bars are struck in concert, as might happen in a glockenspiel or a wind chime, longer bars resonate at lower frequencies than smaller bars, but the relationships (or ratios) between the various resonances remains the same. Figure 6.10 shows three bars with

[11] From McLaren [B: 107].

fundamentals at f_1, g_1, and h_1. The invariance of the ratios between partials implies that $\frac{f_2}{f_1} = \frac{g_2}{g_1} = \frac{h_2}{h_1}$ and that $\frac{f_3}{f_1} = \frac{g_3}{g_1} = \frac{h_3}{h_1}$.

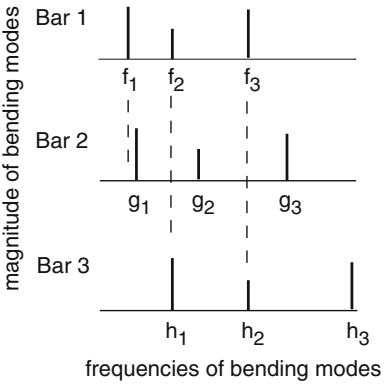


Fig. 6.10. Three metal bars of different lengths (that are otherwise identical) have the same pattern of bending modes (partials), but beginning at different base frequencies. When these partials coincide, as for bars 1 and 3, they achieve maximum sensory consonance. When they fail to coincide, like bars 1 and 2, dissonances originate.

When the partials of one bar fall close to (but not identical with) the partials of another, then the sound beats in a harsh and dissonant fashion. When the overtones coincide, however, the sound becomes smoother, more consonant. The trick to designing a consonant set of metal bars (wind chimes, for instance) is to choose the lengths so that the overtones overlap, as much as is possible. In the figure, bars 1 and 3 will sound smooth together, and bars 1 and 2 will be rougher and more dissonant.

Although this graphical technique of overlaying the spectra of inharmonic sounds and searching for intervals in which partials coincide is clear conceptually, it becomes cumbersome when the spectra are complex. Dissonance curves provide a systematic technique that can find consonant intervals for a given spectrum that is essentially independent of the complexity of the spectra involved.

6.5.2 A Tuning for Ideal Bars

There are many percussion instruments such as xylophones, glockenspiels, wind chimes, balophones, sarons, and a host of other instruments throughout the world that contain wood or metal beams with free (unattached) ends. Assuming that the thickness and density of the bar are constant throughout its length, the frequencies of the bending modes or partials can be calculated using a fourth-order differential equation given in *Fundamentals of Acoustics* by Kinsler and Fry [B: 85]. Assuming that the lowest mode of vibration is at a frequency f, and that the beam is free to vibrate at both ends, the first six partials are

$$f,\ 2.76f,\ 5.41f,\ 8.94f,\ 13.35f,\ \text{and}\ 18.65f$$

which are clearly not harmonically related.

Two octaves of the dissonance curve for this spectrum are shown in Fig. 6.11. Numerous minima, which define intervals of a scale in which the uniform bar instrument will sound most consonant, are spaced unevenly throughout the two octaves. Observe that there are only a few close approximations to familiar intervals: the fifth, the major third, and the second octave. The octave itself is fairly dissonant.

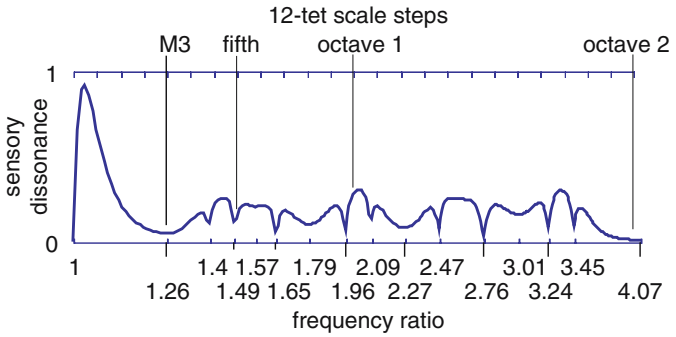


Fig. 6.11. Dissonance curve for a uniform bar has minima shown by tick marks on the lower axis. The upper axis shows 12-tet, with several intervals extended for easy comparison.

With so few intervals coincident with those of the 12-tet scale, how can such bar instruments be played in ensembles with strings, winds, and other harmonic instruments? First, most have a short, percussive envelope. This tends to hide the roughness, because beats take time to develop. Second, by mounting the bar in clever ways, many of the offensive partials can be attenuated. For instance, the bar is typically suspended from two points roughly two-ninths of the way from the ends. These points coincide with the nodes of the first partial. (In Fig. 2.8 on p. 25, these are the stationary points in the vibration pattern of the first partial.) As other partials require nonzero excursions at the 2/9 point, they rapidly die away. This is somewhat analogous to the way that guitarists play "harmonics" by selectively damping the fundamental, only here all partials but the fundamental are damped. To hear this for yourself, take a bar such as a long wind chime, and hold it in the middle (rather than at the 2/9 position). The fundamental will be damped, and the odd-numbered partials (at $2.76f$, $8.94f$ and $18.65f$) will be greatly exaggerated. Suspending at yet other points brings other partials into prominence.

Despite the short envelope and the selective damping of partials, the inharmonicity of bar instruments is considered a problem, and attempts to manipulate the contour and/or density of the bar to force it to vibrate more

harmonically[12] are common. The idea of related scales and spectra suggests an alternative. Rather than trying to manipulate the spectrum of the bar to fit a preexisting pattern, let the bar sound as it will. Play in the musical scale defined by the spectrum of the bar, the scale in which it will sound most consonant.

6.5.3 Tunings for Bells

Bell founders and carillon makers have long understood that there is an intimate relationship between the modes of vibration of a bell and how much in-tune certain intervals sound. Because bells are shaped irregularly, they vibrate in modes far more complex than strings or bars. The *Physics of Musical Instruments* by Fletcher and Rossing [B: 56] contains a fascinating series of pictures showing how bells flex and twist in each mode. The frequencies of these modes vary depending on numerous factors: the thickness of the material, its uniformity and density, the exact curvature and shape, and so on.

There is no theoretically ideal bell like there is an ideal rectangular bar, but bell makers typically strive to tune the lowest five modes of vibration (called the hum, prime, tierce, quint,[13] and nominal) so that the partials are in the ratios $0.5 : 1 : 1.2 : 1.5 : 2$. The tuning process involves carefully shaving particular portions of the inside of the bell so as to tame wanton modes without adversely effecting already tuned partials. Traditional church bells tuned this way are called "minor third" bells because of the interval 1.2, which is exactly the just minor third 6/5. Bell makers have recently figured out how to shape a bell in which the tierce becomes 1.25, which is the just major third 5/4. These are called "major third" bells.

Using dissonance curves, it is easy to investigate what intervals such bells sound most consonantly. The frequencies of the modes of vibrations of three bells are shown in Table 6.2. The partials of the ideal minor and major third bells are taken from [B: 94],[14] and the measured bell is from a D_5 church bell as investigated by [B: 132] and [B: 157]. The most noticeable difference between the minor and major bells is the tierce mode, which has moved from a minor to a major third. Inevitably, the higher modes also change. The measured bell gives an idea of how accurately partials can be tuned. The quint and undeciem are considerably different from their ideal values. There is debate about whether the stretched double octave is intentional (recall that stretching is preferred on pianos) or accidental.

The dissonance curves for these three bells are shown in Fig. 6.12, and the exact values of the minima are given in Table 6.3. Although bells cannot be made harmonic because of their physical structure, the close match between

[12] For instance, see [B: 124].

[13] Those who remember their Latin will recognize tierce and quint as roots for third and fifth.

[14] As reported in [B: 56].

Table 6.2. Partials of bells used in Fig. 6.12.

Name of Partial	Ideal Minor Third Bell	Measured Bell	Ideal Major Third Bell
hum	0.5	0.5	0.5
prime	1.0	1.0	1.0
tierce	1.2	1.19	1.25
quint	1.5	1.56	1.5
nominal	2.0	2.0	2.0
deciem	2.5	2.51	2.5
undeciem	2.61	2.66	2.95
duodeciem	3.0	3.01	3.25
upper octave	4.0	4.1	4.0

the just ratios and the minima of the dissonance curves suggests that bell makers tune their instruments so that they will be consonant with harmonic sounds. Such tuning is far more complex than simply tuning the fundamental frequency because it requires independent shaping of a large number of partials.

The dissonance curve for the measured bell is close to the ideal. Some extra minima have been introduced, and some of the deeper minima have been smeared by the slight misalignment of partials. The major third bell has accomplished its goal. In both octaves, the major third is very consonant, second only to the octave. Unfortunately, the consonance of the fifth has been reduced, and the minimum corresponding to the fifth has become noticeably flat. It is unclear whether or how much this effects the playability of the bell.

The literature on bells is vast, and either [B: 56] or [B: 157] can be consulted for an overview. The present discussion highlights the use of dissonance curves as a way of investigating what intervals sound consonant when played by a bell with a specified set of partials. An alternative is to try writing a piece of music emphasizing the inharmonic nature of the bell, an avenue pursued in the next chapter.

6.5.4 Tuning for FM Spectra

Frequency Modulation (FM) was originally invented for radio transmission. Chowning [B: 32] pioneered its use as a method of sound generation in digital synthesizers, and it gained popularity in the Yamaha DX and TX synthesizers. Sound is typically created in a FM machine using sine wave oscillators. By allowing the output of one sine wave (the modulator) determine the frequency of a second (the carrier), it is possible to generate complex waveforms with rich spectra using only a few oscillators. When the ratio of the carrier frequency to the modulator frequency is an integer, the resulting sound is harmonic, whereas noninteger ratios generate inharmonic sounds. In practice, these complex inharmonic sounds are often relegated to percussive or noise

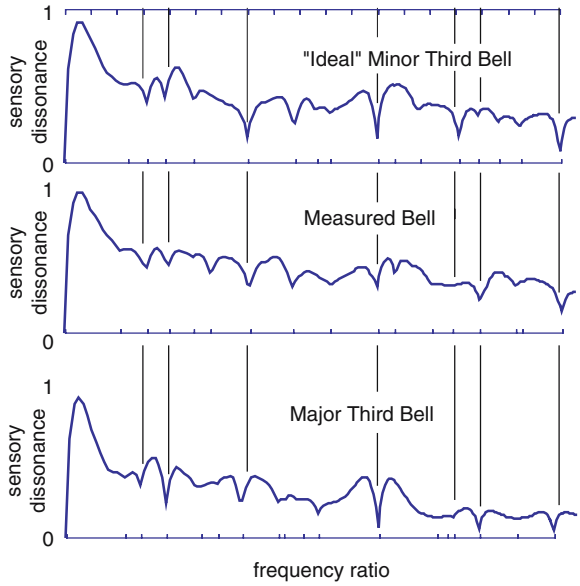


Fig. 6.12. Dissonance curve for an "ideal" minor third bell is compared with the dissonance curve of a real bell, and to the dissonance curve of the "major third" bell described by [B: 94]. The ideal has deep minima at many of the just ratios, and the minima for the real bell are skewed. The increase in consonance of the major third is apparent in both octaves of the lower plot, although the fifths have become slightly flat.

patches because they sound dissonant when played in 12-tet. Using the related scale allows such sounds to be played more consonantly.

For example, consider an FM tone with carrier-to-modulation ratio $c : m$ of $1 : 1.4$ and modulating index[15] $I = 2$. The frequencies and magnitudes of the resulting spectra are shown schematically in Fig. 6.13. The spectrum is clearly inharmonic, and the magnitude of the fundamental (at 500 Hz) is small compared with many of the partials. When programmed on a TX81Z (a Yamaha FM synthesizer), the sound is complex and somewhat noisy. Placing a slowly decaying "plucked string" envelope over the sound and a small amount of vibrato gives it a strange inharmonic flavor: more like a koto or shamisen than a guitar. There are few intervals in 12-tet at which this sound can be played without significant dissonance. The most consonant interval (when restricted to the 12-tet scale) is probably the minor seventh, although the fourth is also smooth. The fifth and octave are definitely dissonant.

[15] The way that the parameters c, m, and I relate to the frequencies and amplitudes of the partials of the resulting sound is complex, but formulas are available in [B: 32] and [B: 158].

Table 6.3. Minima of dissonance curves in Fig. 6.12.

Nearest Just Ratio	Ideal Minor Third Bell	Measured Bell	Ideal Major Third Bell
1/1	1.0	1.0	1.0
	1.15	1.13	1.14
6/5	1.2	1.2	1.18
5/4	1.25	1.26	1.25
4/3	1.33	1.33	1.35
		1.38	1.4
3/2	1.5	1.51	1.48
			1.6
			1.62
5/3	1.67	1.66	1.69
	1.75	1.8	1.75
2/1	2.0	2.0	2.0
	2.08	2.08	
	2.2	2.26	2.28
12/5	2.4	2.36	2.33
10/4	2.5	2.51	2.5
	2.62	2.72	2.72
	2.75	2.76	
3/1	3.0	3.01	2.95

Two octaves of the dissonance curve for this spectrum are plotted in Fig. 6.14, and it is readily apparent why there are so few consonant intervals in the 12-tet scale. Although there are numerous minima, almost none coincide with steps of the 12-tet scale, except for the fourth and minor seventh. But when retuned to the related "FM scale" with steps given by the minima of the figure, the sound can be played without excessive dissonance.

The reason for including this example is because it is likely that some readers will have access to an FM-based synthesizer. This is an easy source

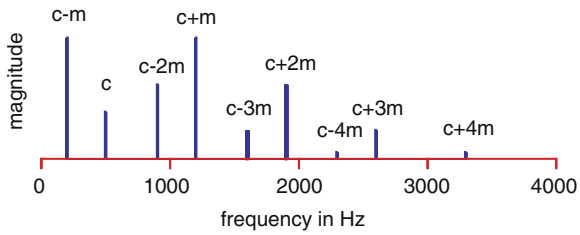


Fig. 6.13. Line spectrum showing the partials of the FM spectrum with $c : m$ ratio $1 : 1.4$ and modulating index $I = 2$. The "fundamental" was arbitrarily chosen at $c = 500$ Hz.

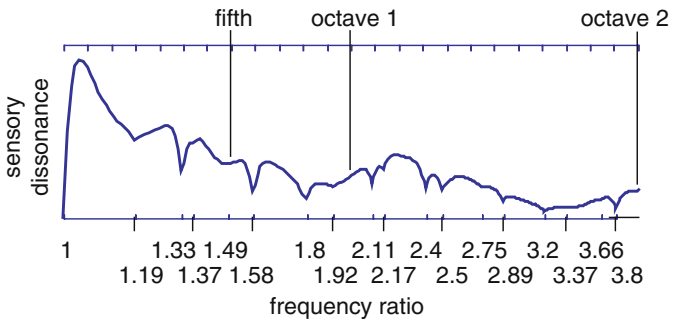


Fig. 6.14. Dissonance curve for the FM spectrum with $c : m$ ratio $1 : 1.4$ and modulating index $I = 2$ has minima shown by the tick marks on the bottom axis. The 12-tet scale steps are shown above for comparison.

of inharmonic sounds, and many units incorporate tuning tables so that the tuning of the keyboard can be readily specified. This particular timbre is, frankly, not all that interesting musically, but the procedure can be applied generally. Why not find the spectrum of your favorite (inharmonic) FM sound, and retune the synthesizer to play in the related scale? Working through an example like this is the best way to ensure you understand the procedure, and you may find yourself enthralled by a new musical experience.

6.6 Properties of Dissonance Curves

The shape of the dissonance curve is dependent on the frequencies (and magnitudes) of the components of the spectrum. Changing these frequencies (and magnitudes) changes the location and depth of the minima, which changes the scale in which the spectrum can be played most consonantly. The examples of the previous sections showed specific spectra and their related scales. In contrast, this section looks at general properties of dissonance curves by probing the mathematical model for internal structure and by exploring patterns in its behavior. Four generic properties are presented, although formal statements of these properties (and their proofs) are relegated to Appendix F. These properties place bounds on the number of minima of a dissonance curve, identify symmetries, and describe two generic classes of minima. These properties help give an intuitive feel for where minima will occur and how they change in response to changes in the frequencies and amplitudes of the partials.

Throughout this section, we suppose that the spectrum F has n partials located at frequencies $f_1, f_2, ..., f_n$.

Property 1: The unison is a minimum of the dissonance curve.

Recall that any nontrivial sound[16] has an inherent dissonance due to the interaction of its partials. The dissonance of the sound at unison consists of just this intrinsic dissonance, whereas other intervals also contain interactions between nonaligned partials. Details and caveats are given in Appendix F.

> *Property 2*: As the interval grows larger, the dissonance approaches a value that is no more than the intrinsic dissonance of the sound.

The second property looks at extremely large intervals where all partials of the lower tone fall below the partials of the upper tone. For large enough intervals, the interaction between the partials becomes negligible, and the dissonance decreases monotonically as the interval increases. In practical terms, a tuba and a piccolo may play together without fear of excess dissonance.

The next result gives a bound on the number of minima of a dissonance curve in terms of the complexity of the spectrum.

> *Property 3*: The dissonance curve generated by F has at most $2n^2$ minima that are located symmetrically (on a logarithmic scale) so that half occur for intervals between 0 and 1, and half occur for intervals between 1 and infinity.

There are really two parts to this property: a bound on the number of minima, and an assertion of symmetry. The easiest way to see (and hear) these is by example. Consider a simple spectrum with just two partials. As shown in Fig. 6.15, the dissonance curve can have three different contours depending on the spacing between the two partials:[17] The unison may be the only minimum, there may be an additional two steep minima, or there may be an additional two "broad" minima.

The middle graph of Fig. 6.15 shows the dissonance curve for a simple sound with two partials at f and $1.15f$. The dissonance begins at the unison, rises rapidly to its peak, and then plummets to a sharp minimum at 1.15. Dissonance then climbs again before sinking slowly toward zero as the two sounds drift apart. It is easy to understand this behavior in terms of the coincidence of the partials. Let r denote the ratio between the two notes. Near unity (for $r \approx 1$), the partials of f beat furiously against the corresponding partials of rf. When r reaches 1.15, the second partial of f aligns exactly with the first partial of rf, and the dissonance between this pair vanishes, causing the minimum in the curve. As r continues to increase, the previously aligned partials begin to beat, producing the second peak. For large r, both partials of f are separated from both partials of rf so that there is little interaction, and hence little dissonance.

[16] That is, any sound that contains more than a single partial. Only silence and a pure sine wave have zero dissonance.

[17] To make this figure clearer, the intrinsic dissonances have been subtracted out.

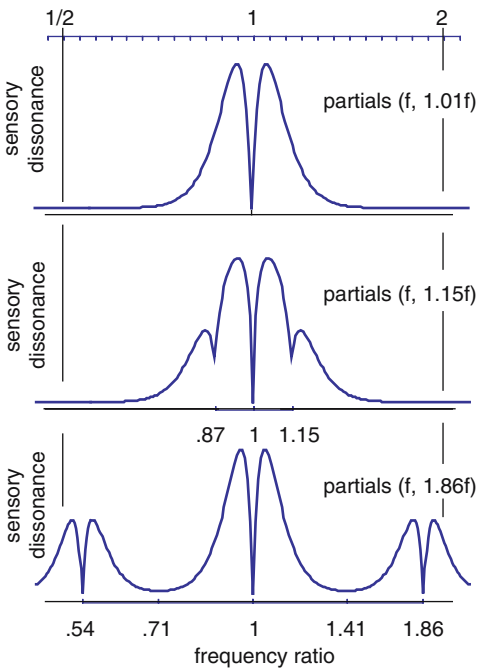


Fig. 6.15. Dissonance curves for spectra with two partials have three possible shapes: The partials may be too close together to allow any minima other than the unison (top), the minima may occur at the intervals defined by the ratios of the partials (middle), or there may also be "broad" minima due to the sparsity of partials (bottom). Observe the symmetry about the unison. Steps of the 12-tet scale are shown above for comparison.

Perhaps the most striking feature of this figure is its symmetry.[18] Suppose that instead of sliding the second tone up in frequency, it is shifted down; a similar scenario ensues. For $r \approx 1$, there is large dissonance. As r descends to 0.87 (which is the inverse of 1.15, that is, $\frac{1}{1.15} = 0.87$), the first partial of f aligns with the second partial of rf to cause a minimum. As r continues to descend, the rise and fall of dissonance occur just as before. In general, whenever there is a minimum at a particular value r^*, there is also a minimum at $1/r^*$. Thus, the range from 0 to 1 is a mirror image of the range from 1 to infinity, and they are typically folded together, as has been done for most of the dissonance curves throughout the book.

If the partials are too close together, there may be no minima other than the unison. The top graph in Fig. 6.15 shows the dissonance curve for a sound with partials at f and $1.01f$. At first thought, one might expect that $r = 1.01$ (and its inverse) should be minima. But the other partials are clustered nearby, and their combined dissonances are enough to overwhelm the expected minima. In essence, if the partials are clumped too tightly, minima can disappear.

Thus, minima may (or may not) occur when partials coincide. Minima can also occur when partials are widely separated. The bottom graph in Fig. 6.15 shows the dissonance curve for a sound with partials at f and $1.86f$. As

[18] The astute reader will note that the symmetry is not exact, because dissonance curves vary with absolute frequency. However, over much of the audio range, the curves are nearly symmetric.

expected, there are minima at 1.86 and its inverse 0.54, but there is also a new kind of "broad" minimum at 1.41 (and its inverse). This occurs because the partials are widely separated, so that for a large range of the ratio r, there is little significant interaction. Such minima are typically wide, and they tend to disappear for sounds with more than a few partials. The harmonic dissonance curve of Fig. 6.1 on p. 100, for instance, consists exclusively of minima caused by coinciding partials; the broad, in-between minima have been vanquished. This discussion foreshadows a property describing the two classes of minima: those caused by coinciding partials and those caused by widely separated partials.

> *Property 4: The principle of coinciding partials.* Up to n^2 of the minima occur at interval ratios r for which $r = f_i/f_j$ where f_i and f_j are partials of F. Up to n^2 of the minima are the broad type of the bottom curve in Fig. 6.15.

For example, spectra with three partials may have up to three minima at points where $r_1 f_1 = f_2$, $r_2 f_1 = f_3$, and $r_3 f_2 = f_3$, which are represented schematically in Fig. 6.16. Essentially, a minimum can occur whenever two of the partials coincide, and this property is called the principle of coinciding partials. Of course, other minima may exist as well. The top graph in Fig. 6.17 shows the dissonance curve for the spectrum $f, sf, s^4 f$, where $s = \sqrt[10]{2}$. Note that the three minima predicted by property 4 are at exactly the first and fourth scale degrees of the ten-tone equal-tempered scale, and at the difference frequency $s^3 f$. The bottom graph of Fig. 6.17 places the partials at $f, sf, s^6 f$, generating the expected scale steps at 1 and 6, and the difference frequency $s^5 f$ at 10-tet scale step 5. There is also a broad minimum between the third and fourth steps, which is a result of the distance between the partials sf and $s^6 f$.

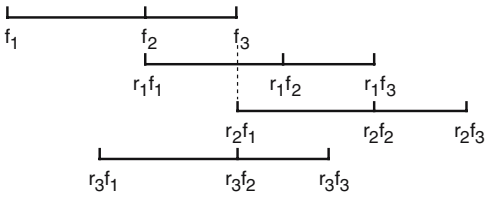


Fig. 6.16. Schematic representation of three possible local minima (at ratios r_1, r_2, and r_3) of a spectrum with partials at f_1, f_2, and f_3.

Properties 3 and 4 combine to give a fairly complete picture of the number and types of minima to expect. They are located symmetrically (on a logarithmic scale) so that half occur for intervals between 0 and 1, and half occur for intervals between 1 and infinity. No more than half of the minima are the broad type due to a paucity of partials. No more than half are the steep kind, which occur when partials coincide at intervals defined by ratios of the partials. Because the musically useful information is located in intervals

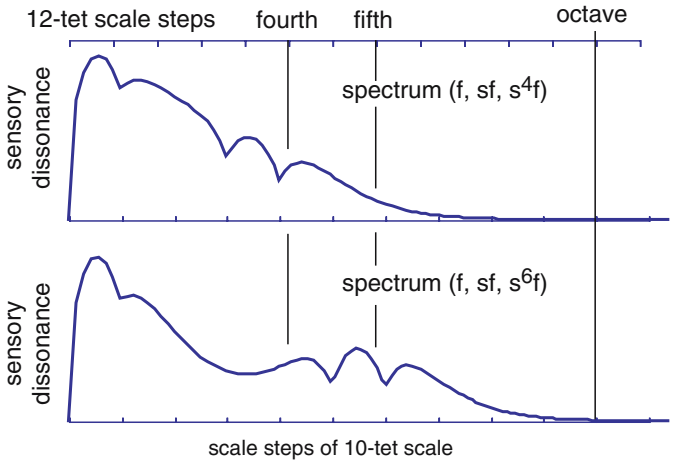


Fig. 6.17. Dissonance curves demonstrating local minima for spectra with three partials, with s defined as the tenth root of two. Observe that minima are coincident with scale steps of 10-tet and not with scale steps of 12-tet.

within a couple of octaves of unity, because the broad minima tend to vanish (except for sparse spectra), and because many minima are annihilated when partials are densely packed, typical dissonance curves exhibit far fewer than the maximum. In Fig. 6.1 on p. 100, for instance, there are only nine minima within the octave of interest, considerably fewer than the bound of 2×7^2.

Symmetry of the dissonance curves about one is not the same as repetition at the octave. For instance, the harmonic dissonance curve[19] has a minimum at 5/4, and the corresponding symmetric minimum occurs at 4/5. When translated back into the original octave between 1 and 2, this is 8/5, which is not a minimum. Thus, using the related scale under the assumption of octave equivalence is different, in general, from using the intervals of the dissonance curve plus their inverses. Depending on the musical context, either one or the other may be preferred.[20] Typically, the minima of a dissonance curve become sparser (further apart) for very high and for very low frequencies, implying that both low and high notes will be far apart when using the scale with inverses. This accords well with our perceptual mechanism because the majority of notes tend to cluster in the midrange where hearing is most sensitive.

Another consequence of the symmetry of dissonance curves is that the "inverse" of a spectrum will have the same dissonance curve as the spectrum. For example, subharmonic sounds are those defined by a frequency f, and the sub-

[19] Fig. 6.1 on p. 100.

[20] Octave equivalence is often assumed because it is generally easier to "map" to the keyboard, but this is a pragmatic and not a musical or perceptual preference.

harmonics $f/2$, $f/3, \ldots$. Such subharmonic sounds have the same dissonance curve and the same related scale as harmonic sounds.

6.7 Dissonance Curves for Multiple Spectra

The dissonance curves of the previous sections assumed that both notes in the interval had the "same" spectrum; that is, they differed only by a simple transposition.[21] As it is common to combine sounds of different tonal quality, it is important to be able to draw analogous dissonance curves for notes with different spectra.

Suppose the note F has partials at f_i with loudness a_i, and the note G has partials at g_j with loudness b_j. Then the dissonance between F and G is the sum of all dissonances $d(f_i, g_j, a_i, b_j)$, where the function[22] d represents the sensory dissonance between the pure sine wave partials at f_i and g_j as in Fig. 3.8 on p. 47, weighted by the loudnesses. Similarly, if G is raised (or lowered) by an interval s, then the dissonances $d(f_i, sg_j, a_i, b_j)$ are summed, whereas if F is raised (or lowered) by an interval r, then the dissonance is calculated[23] by summing all $d(rf_i, g_j, a_i, b_j)$.

For example, suppose that a sound F with four harmonic partials is played simultaneously with a sound G with three inharmonic partials at g, $1.515g$, and $3.46g$. The corresponding dissonance curve is shown in Fig. 6.18 over a region of slightly larger than an octave in both r and s. The curve is drawn with r and s on the same axis because they are essentially inverses; that is, the effect of playing F and transposing G by s is nearly the same[24] as playing G and transposing F by $r = 1/s$.

In this example, minima occur near many of the steps of 5-tet, which is shown on the top horizontal axis. There are minima when s is the first, second, and fifth steps of 5-tet, and when r is the first, third, and fourth steps. Together, this suggests that this pair of sounds may be sensibly played in 5-tet.

[21] The note with partials at f_i and loudness a_i, when transposed by an interval r, has partials at rf_i with the same loudness.

[22] Details of the function d can be found in Appendix E.

[23] An alternative approach is to combine the spectra of the two sounds, and then draw the (normal) dissonance curve. For instance, combining the F and G of Fig. 6.18 gives a "new" sound H with partials at h, $1.515h$, $2h$, $3h$, $3.46h$, and $4h$. The dissonance curve for this spectrum has many of the same features as Fig. 6.18, but it is not identical. For instance, when the sixth partial of the lower tone corresponds to the fourth partial of the higher tone (at the interval 4/3), the dissonance curve of H may have a minimum, depending on the loudness of the partials. There is no minimum at 4/3 in Fig. 6.18 however, because there are no pairs of partials in F and G with this 4/3 ratio.

[24] They differ only due to the absolute frequency dependence of dissonance, which is relatively small over moderate intervals.

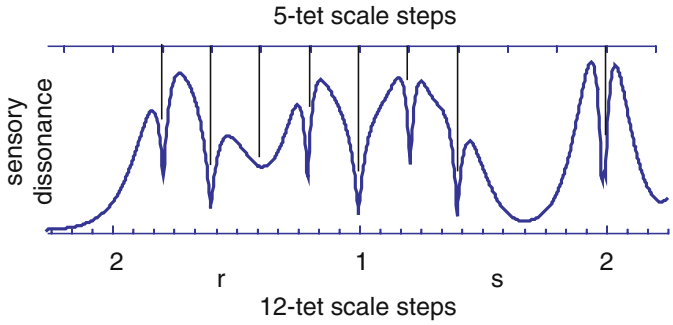


Fig. 6.18. Dissonance curve for sounds F (at interval r) and G (at interval s). F has four harmonic partials while G has three inharmonic partials at g, $1.515g$, and $3.46g$. The curve has many minima close to the steps of 5-tet, which is shown above for comparison.

Dissonance curves for multiple spectra have somewhat different properties than similar curves for sounds with a single spectrum. For instance, the unison is not always a minimum. Figure 6.19 shows the dissonance curve for two inharmonic sounds with partials at f, $1.7f$, and $2.84f$, and at g, $1.67g$, and $3.14g$. The deepest minimum occurs at the interval $s = 1.7$, where the first and second partials of F align with the second and third partials of G. The unison is not a minimum.

The second property, which says that dissonance must decrease as the intervals grow asymptotically large, is still valid. But the third property must be amended.

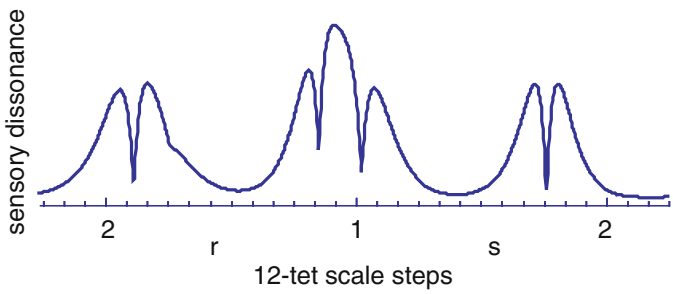


Fig. 6.19. Dissonance curve generated by two sounds F (with partials at f, $1.7f$, and $2.84f$) and G (with partials at g, $1.67g$, and $3.14g$). Loudness values for both sounds are 1, 5, and 5. Minima occur at $r = 1.1$, 1.37, and 1.85, and at $s = 1.02$, 1.33, and 1.7. The unison is not a minimum.

Property $3'$: The dissonance curve generated by F and G has at most $2nm$ minima, where n is the number of partials in F and m is the number of partials in G.

The symmetry of the curves about unity is lost, as shown in both Figs. 6.18 and 6.19. The principle of coinciding partials must also be modified.

Property $4'$: In the dissonance curve generated by F and G, up to nm of the minima occur at intervals r for which either $r = g_j/f_i$ or $r = f_i/g_j$, where f_i and g_j are the partials of F and G. Up to nm of the minima are the broad type of the bottom curve in Fig. 6.18.

Dissonance curves can give insight into how different kinds of sounds can be combined so as to control sensory consonance. This might find application, for instance, in a piece that combines several kinds of inharmonic sounds. Small manipulations of the pitches may lead to dramatic changes in the perceived dissonance of the combined sound, and dissonance curves can be used to reliably predict these changes.

6.8 Dissonance "Surfaces"

Dissonance curves can also be drawn for three note "chords." These can be readily pictured as dissonance surfaces where mountainous peaks are points of maximum dissonance, and valleys are locations of maximum consonance.

As usual, the total dissonance is calculated by adding the dissonances between all simultaneously sounding partials. The sensory dissonance of a sound F played in a chord containing the intervals 1, r, and s is[25]:

$$\left\{\begin{array}{c}Total\\Dissonance\\of\ Chord\end{array}\right\} = \left\{\begin{array}{c}Dissonance\\Between\\F\ and\ rF\end{array}\right\} + \left\{\begin{array}{c}Dissonance\\Between\\F\ and\ sF\end{array}\right\} + \left\{\begin{array}{c}Dissonance\\Between\\rF\ and\ sF\end{array}\right\}$$

Generalizations to m sounds, each with its own spectrum, follow the same philosophy, although in higher dimensions there is no simple way to draw pictures.

Figure 6.20 shows the dissonance "surface"[26] for a sound F consisting of six harmonic partials, as r and s are varied over a region slightly larger than an octave. The central rift, which is sandwiched by a range of high mountains near the diagonal, is the degenerate case where $r \approx s$. The two far edges of the surface (which are not clearly visible due to the angle of view) are where $r = 1$ (on the left) and $s = 1$ (around the back). As all three notes have the

[25] rF is the transposition of F by the interval r.

[26] Appendix E details how the surfaces are drawn.

same spectra, r and s are interchangeable and the surface is symmetric about the diagonal. Hence, the most interesting and musically useful information is contained in the foothills on the near side of the diagonal range.

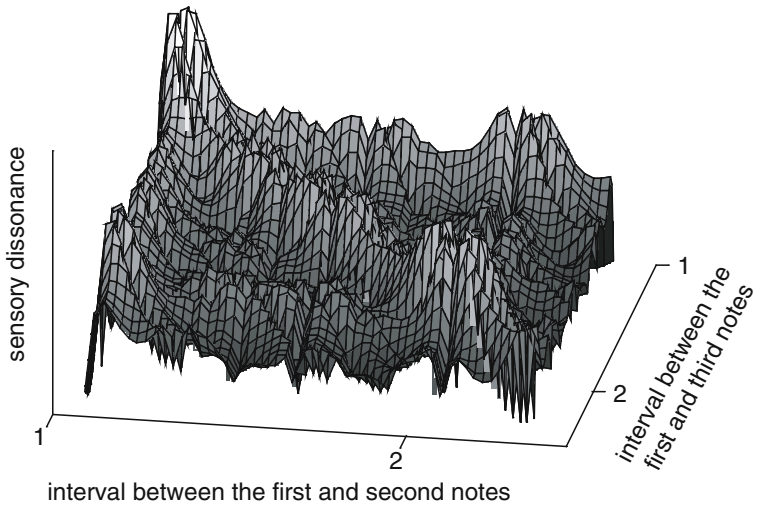


Fig. 6.20. Dissonance curve for a sound with six harmonic partials has minima at many intervals defined by small integer ratios. These form chords with maximum sensory consonance. Figure 6.21 shows the same data as a contour plot.

Although surface plots such as Fig. 6.20 give a broad overview of the landscape, it is not always easy to spot detailed features. The same information is displayed as a "contour plot," a topographic map of the dissonance landscape, in Fig. 6.21. The symmetry about the diagonal is readily apparent. The far and left-hand edges again represent the degenerate cases where $s \approx 1$ and $r \approx 1$, and the beaded strand on the diagonal is where $r \approx s$. In these regions, two of the three notes have merged.

Many of the just chords appear in the lower left half of the figure as prominent sinkholes in the dissonance wilderness. For instance, the arrows K and M in Fig. 6.21 indicate long narrow ravines at the perfect fifth in both the horizontal and vertical dimensions, that is, in both r and s. This ravine contains both the just major and just minor chords B and D. An angled string of minima for which the second and third notes are locked into a perfect fifth is indicated by the arrow L. This string intersects the ravine at the J chord, which is composed of two perfect fifths piled on top of each other.

The chord labeled A contains both a perfect fourth and a perfect fifth. Such "suspended" chords do not form a normal diatonic triad, and yet they are not unfamiliar. The chord G can be viewed as an inversion. Raising the fundamental of 1, r^5, r^{10} one octave gives r^5, r^{10}, r^{12}, which is a transposition

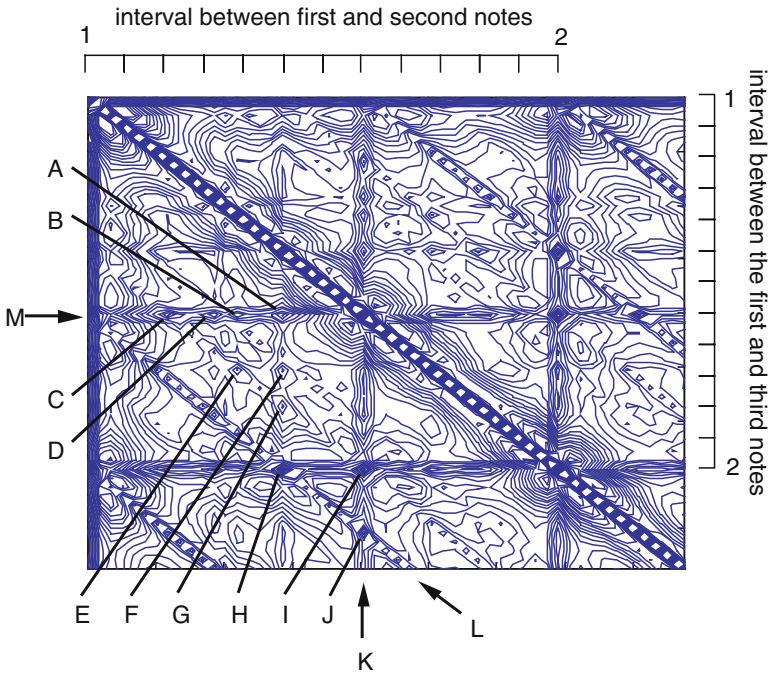


Fig. 6.21. Contour plot of the dissonance curve for three note chords with harmonic spectra. Several of the most important features are indicated. Tick marks on the axes indicate intervals of the 12-tet scale step. The chords labeled A-J are examined in more detail in Table 6.4.

Table 6.4. Minima of the dissonance surface for a sound with six harmonic partials occur at many of the just chords and at many of the simple integer ratios. Labels refer to regions on the contour plot for harmonic sounds in Fig. 6.21.

Label	Actual Minimum	Closest 12-tet scale steps $a = \sqrt[12]{2}$	Comment
A	1, 4/3, 3/2	1, a^5, a^7	suspended
B	1, 5/4, 3/2	1, a^4, a^7	just major
C	1, 9/8, 3/2	1, a^2, a^7	suspended
D	1, 6/5, 4/3	1, a^3, a^7	just minor
E	1, 5/4, 5/3	1, a^4, a^9	inversion of minor
F	1, 4/3, 5/3	1, a^5, a^9	inversion of major
G	1, 4/3, 16/9	1, a^5, a^{10}	string of fourths
H	1, 4/3, 2	1, a^5, 2	open fourth
I	1, 3/2, 2	1, a^7, 2	open fifth
J	1, 3/2, 9/4	1, a^7, a^{14}	string of fifths

of A. The chord C is also an inversion of A, as can be seen by lowering the highest note an octave. Similarly, E and F are inversions of the just major and minor chords.

It may at first appear strange that the intervals 9/8 and 16/9 appear in C and G, because the dissonance surface was generated by a harmonic sound containing only the first six partials. But the interval from 3/2 to 9/8 is exactly 4/3, and so the 9/8 interval is a byproduct of the consonance of the perfect fourth and the perfect fifth. Similarly, the 16/9 in G forms a perfect fourth with 4/3, and this suspended chord can be thought of as a "string of fourths." In fact, the string of fifths chord J is also an inversion of this same suspension, because lowering the highest note an octave gives the C chord.

The real purpose of this discussion is not to learn more about just intonation or about the traditional diatonic setting, because these have been explored extensively through the years. Rather, it is to demonstrate that in the familiar harmonic setting, features of dissonance curves and surfaces correspond closely with familiar musical objects. Hence, there is good reason to expect that in unfamiliar inharmonic contexts, analogous features can be used to predict and explore unfamiliar musical intervals, scales, and chords. An extended example is given in the chapter "Towards a 'Music Theory' for 10-tet."

6.9 Summary

Dissonance curves generalize the kinds of curves drawn by Helmholtz, Partch, and Plomp to sounds with inharmonic spectra. They give a graphic display of the intervals with the greatest sensory consonance (least sensory dissonance) for a given spectrum, and these intervals can be gathered into the *related scale*. Several previous investigations were highlighted, including the work of Mathews and Pierce and their colleagues, and the musical explorations of Carlos. Examples were drawn from ideal bars, bells, and FM synthesis. General properties of dissonance curves bound the number of minima, demonstrate the symmetry of the intervals about the unison, and classify them into those caused by coinciding partials and those that are a result of gaps in the partial structure. Extensions to multiple sounds with different spectra are straightforward. The next chapter explores three examples thoroughly.

7

A Bell, A Rock, A Crystal

To bring the relationship between tuning and spectrum into sharper focus, this chapter looks at three examples in detail: an ornamental hand bell, a resonant rock from Chaco Canyon, and an "abstract" sound created from a morphine crystal. All three are discussed at length, and each step is detailed so as to highlight the practical issues, techniques, and tradeoffs that originate when applying the ideas to real sounds making real music. The bell, rock, and crystal were used as the basis for three compositions: Tingshaw*,* The Chaco Canyon Rock*, and* Duet for Morphine and Cymbal*, which appear on the accompanying CD as sound examples [S: 43], [S: 44], and [S: 45].*

7.1 Tingshaw: A Simple Bell

By the tenth century BC, bells were used to accompany rituals, and they are among the oldest extant musical instruments. Bells can be made from metal, wood, clay, glass, and almost any other material that can be shaped to sustain oscillation. They range in size from tiny ornaments to monstrosities weighing several tons. Because of the great variety of materials, shapes, and sizes, bells are capable of a wide variety of tones and timbres. The typical bell sound is inharmonic, and its sound envelope (a rapid rise followed by a long slow decay) is probably its most distinctive feature.

This section uses one particular hand bell, and it derives the related scale using the dissonance curve. This scale is then "mapped" onto a standard keyboard, and some aspects of performance are considered. A musical composition called *Tingshaw* featuring this inharmonic bell played in its nonequal, nonoctave based scale, is presented in sound example [S: 43].

Despite the "scientific" flavor of much of the discussion in previous chapters, the translation from sound to scale is not a completely mechanical process. Decisions must be made that will ultimately shape the performance and playability of the sound and, hence, will help to mold the resulting music. To outline the complete procedure:

(i) Choose a sound
(ii) Find the spectrum of the sound
(iii) "Simplify" the spectrum

(iv) Draw the dissonance curve, and choose a set of intervals (a scale) from the minima
(v) "Create an instrument" that can play the sound at the appropriate scale steps
(vi) Play music

Each of these will now be discussed in detail, and the decisions and choices made for the tingshaw will be explained. Although someone versed in spectral analysis will find many aspects of this discussion familiar, there are a number of issues that are specific to the auditory setting.[1] I do not present this detail in the expectation that it would be useful to exactly duplicate my steps. Rather, over several years of working with this kind of material, I have run across certain problems and traps again and again. My hope is to post warnings near some of these traps.

7.1.1 Choose a Sound

Although obvious, this is the most crucial step of the procedure, because the character of everything in the music (from the character of the sound to the scale in which it will be played) are derived from the sound itself. Sounds may come from a musical synthesizer. They may be from "real" instruments such as bells, gongs, cymbals, and so on. They may originate from collisions between natural objects such as bricks, metal pans, scrap wood, rocks, or recyclables. They may be digitally generated by a computer program.

Although any sound can be used, not all sounds are equally useful. If the spectrum of the sound is too simple, then the related scale may be trivial. For instance, the tritone spectrum has a dissonance curve with only three minima, and hence, the related scale has only three notes; it will be hard to write a convincing melody with only three notes. On the other hand, if the spectrum of the sound is too complex, then the related scale may have hundreds or even thousands of notes. This extreme may also be impractical. Finally, an unexciting sound cannot be miraculously rejuvenated by playing it in the related scale. If the timbre is dull and uninteresting, then it will most likely lead to dull and uninteresting music.

For this example, I have chosen a small bell called tingshaw. It has a cheery little clang with a sharp attack and a long slow decay. The tingshaw was

[1] The musician may find all of these decisions and the incredible detail frightening. Recognize that I am trying to write it *all* down. Imagine if you were to try and document every step of the decision-making process when writing even a simple piece of music. You would need to explain why it is in 4/4 time, why one particular note is syncopated and another is not, why the viola line crosses the violin line (in violation of standard rules), and why you have allowed a parallel octave in one section but not another. There are many decisions for *each note*, and there are many, many notes! Rest assured that all of these decisions and detail would be enough to frighten even the hardiest of engineers.

sampled at the standard CD rate of 44100 Hz, and the sample was downloaded to a computer and stored in a file called ting.wav.

7.1.2 Find the Spectrum

There are many programs that can readily calculate the spectrum, but the accuracy and usefulness of the results are determined primarily by the sample rate, the number of samples analyzed, and the windowing procedure used. If you have never taken a spectrum before, you will want to read Appendix C, *Speaking of Spectra*, for an overview of the kinds of tradeoffs that are inherent in this process. The more competently these decisions are made, the more meaningful the results.

The tingshaw bell has a sharp attack followed by a long slow decay into inaudibility. The complete sound file contains about 120K samples, a little less than 3 seconds of sound.[2] Taking the FFT of the complete sound is a bad idea for two reasons. First, it is too long. Because the computation time for an FFT increases rapidly as the length of the signal increases, 120K points could take a long time. Second, the attack is very important to the sound, but it lasts only a few thousand samples. Even if the computation time was acceptable, the long decay would obscure the short attack because of the averaging effect of the FFT.

On the other hand, the FFT must not be too short. At least part of the decay portion of the sound must be present or the spectrum cannot represent the complete sound. Also, the accuracy will suffer. Recall (or read about it in Appendix C) that the width of the FFT frequency bins determines the precision with which the sinusoidal components can be pinpointed. As the width of the bins is proportional to the sampling rate divided by the length of the waveform, taking too small a portion of the wave leads to wide bins and poor estimates for the frequencies of the partials. Such inaccuracies can have serious consequences when defining the related scale.

As the just noticeable difference Fig. 3.4 on p. 44 showed, the ear is sensitive to changes in pitch as small as 2 or 3 Hz in the most sensitive regions below 1000 Hz. Thus, it is sensible to choose an FFT length that gives at least this accuracy. Using an FFT with length that is a power of two gives two choices: a 16K FFT with resolution of 2.69 Hz,[3] or a 32K FFT with a resolution of 1.35 Hz. To decide, I listened to the first 16K of the waveform and to the first 32K. The 16K segment seemed to capture enough of the sustained part of the sound.

To examine the effects caused by truncating the wave, I tried several different windowing strategies. The rectangular window and the hamming windows gave estimates for the most important frequencies that were several Hertz

[2] The duration is the length divided by the number of samples per second; thus, $\frac{120000}{44100} \approx 2.72$ seconds.

[3] $\frac{\text{sampling rate}}{\text{length of FFT}} = \frac{44100}{16384} = 2.69 =$ resolution in Hz.

apart. There are two sources of error: The hamming window attenuates the attack portion significantly, and the rectangular window simply truncates the signal after 16K samples. I reasoned that it was a good idea to leave the attack portion undisturbed, because this is where much of the important information resides. Because a signal has the same spectrum whether it is played forward or backward in time, I carefully selected a "middle point," and reversed the 16K waveform about this midpoint.[4] When plotted, the transition point was visually smooth (i.e., no large jump occurred in either the value of the signal or its slope), and so it seemed unlikely to greatly effect the results. Indeed, this gave a spectrum that differed by no more than 1.5 Hz from the original rectangular window, and so I decided to accept this as the "real" spectrum. Figure 7.1 shows an FFT of the first 16K samples of the sound file ting.wav, accomplished using a 32K FFT and a wave reversal "windowing" strategy.

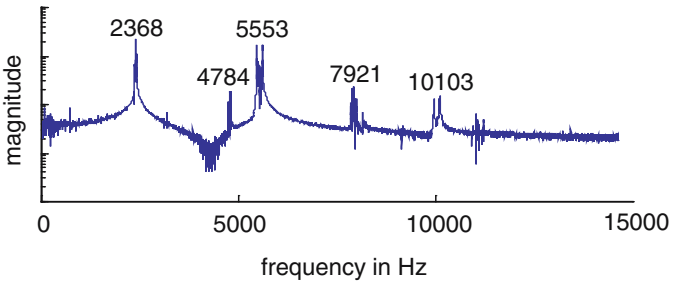


Fig. 7.1. Spectrum of the tingshaw bell with the most prominent spectral peaks labeled.

7.1.3 Simplify the Spectrum

The output of this FFT says that the first 3/8 second of the tingshaw sound consists of the first 16,386 harmonics of a fundamental at 1.35 Hz, each with a specified amplitude and phase. Despite the fact that this is literally true, it is useless.

A far better interpretation of Fig. 7.1 is that there are two dominant regions of spectral activity near 2370 and 5555, and three smaller peaks at 4784, 7921, and 10103. There is also a small cluster near 11300, and a couple of isolated peaks, at about 700 and 3200. It is important to try and select only the most significant peaks, without missing any, because spurious peaks may cause extra minima in the dissonance curve, whereas missing peaks may cause missing scale steps. Neither is good. Perhaps the best strategy is to analyze several different recordings and to choose only what is common among them.

[4] Various windowing strategies are discussed in Appendix C.

This approach is detailed in the next section in the discussion of the Chaco Rock. Unfortunately, the tingshaw bell went missing shortly after I recorded it, leaving only the one sample (and some great memories).

One way to get more information from limited data is to analyze it in different ways. I pursued two different strategies: multiple analysis and analysis by synthesis. One interesting and puzzling feature of the tingshaw spectrum Fig. 7.1 is that there are two separate peaks close to 5555. To investigate, I did a series of 4K spectral snapshots.[5] The snapshots suggested that there is really only one partial in any 4K segment, but that it is slowly changing in frequency from about 5570 down to about 5550 over the course of the sample. As 5550 is its steady-state value (as shown by FFTs taken with the attack portion of the sound stripped away), I settled on the single value 5553 to represent all of this activity. Using the same 4K snapshots shows that the peaks near 7921 are simpler: They merge into a single sinusoid as the sound progresses and remain centered at 7921 throughout.

The second way to try and understand more from a limited number of samples is a variation on a technique pioneered by Risset and Wessel [B: 151] in which the accuracy of an analysis is verified by resynthesizing the sound. If the analysis captures most of the important features of the sound, then the resynthesized sound will be much like the original. In the present context, I first resynthesized[6] the sound using the five major peaks, and then added in the smaller peaks near 700, 3200, and 11,300. Of course, the resynthesized sounds were not much like the tingshaw, but there was almost no perceptible difference between the two resynthesized sounds. This suggested that the extra smaller peaks were likely to have little effect on the overall sound.

Hence, I decided that the five inharmonically related peaks represent the primary constituents of the sound, and this simplified tingshaw spectrum is used to draw the dissonance curve. It is shown in Fig. 7.2.

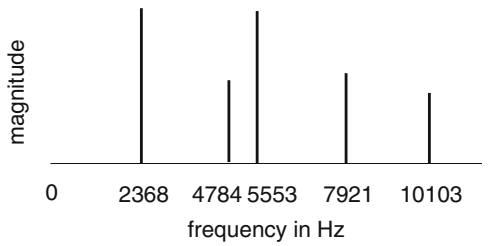


Fig. 7.2. Spectrum of the tingshaw bell simplified to show only the most prominent features.

A third method to help decide which are the most important spectral peaks might be called "analysis by subtractive synthesis." In this method, the FFT of the original sound is manipulated by removing a few suspicious

[5] To be specific, I used a 4K hamming window and evaluated the spectrum centered at samples 1K, 2K, 3K, ... , 15K.

[6] See the Appendix *Additive Synthesis* for details on the resynthesis procedure.

partials and then reconstructed using the inverse FFT. If there is little or no difference between the original and the reconstruction, then the removed partials must be of little importance to the overall sound. I did not actually need to use this technique on the tingshaw because I was already satisfied that I had located the most important spectral information, but it is a technique that has worked well in other situations.

7.1.4 Draw the Dissonance Curve

The simplified spectrum for the tingshaw shown in Fig. 7.1 can be entered into the dissonance calculating programs given in Appendix E, *How to Draw Dissonance Curves*, in a straightforward way. Setting the frequency vector and amplitude vectors

```
freq=[2368, 4784, 5553, 7921, 10103]
amp=[1.0, 0.5, 1.0, 0.6, 0.5]
```

gives the dissonance curve for the tingshaw shown in Fig. 7.3. This figure shows the dissonance curve from unison to just a bit more than two octaves. In the code, the algorithm increments by `inc=0.01` and the upper value is specified by the `range` variable, in this case 4.1. It is often a good idea, when first looking at the dissonance curve of a sound, to calculate the curve over a larger range to make sure nothing "interesting" happens at large values. For the tingshaw, there was one more bump and dip near 4.27, but it was small and seemed unimportant. As shown in the figure, the dissonance curve has minima unevenly spaced at

1, 1.16, 1.29, 1.43, 1.56, 1.66, 1.81, 2.02, 2.15, 2.35, 2.83, 3.34, and 4.08.

One way to choose the scale is to simply use these ratios (plus maybe the one at 4.27) to play the tingshaw. Another possibility is to also use the inverse ratios

1, 0.862, 0.775, 0.699, 0.641, 0.602, 0.552, 0.495, 0.465, 0.425, 0.353, 0.299, and 0.245,

which would result in a complete scale with almost twice as many notes. This is sensible because the dissonance curve is really symmetric about the unison (recall property number 3) and hence contains all of these inverse intervals as well.

But looking more carefully at the minima of the dissonance curve reveals an interesting pattern. If the minimum at 2.02 is thought of as a kind of pseudo-octave, then the intervals $2.15/2.02 = 1.16$, $3.34/2.02 = 1.65$, and $4.08/2.02 = 2.02$ are present in both pseudo-octaves. As these are the most prominent features in the second half of the curve, the tingshaw sound is closely related to the eight-note unequal stretched-octave scale

1, 1.16, 1.29, 1.43, 1.56, 1.66, 1.81, and 2.02.

This is the scale used in the piece *Tingshaw* on the accompanying recording.

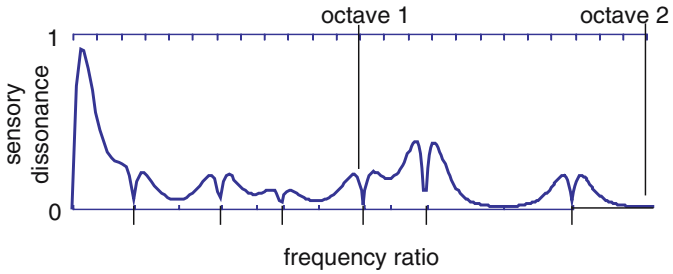


Fig. 7.3. Dissonance curve for the tingshaw bell. The minimum at 2.02 serves as a pseudo-octave, because some of the minima in the second pseudo-octave are aligned with those in the first. For example, $2.35/2.02 = 1.16$ and $3.34/2.02 = 1.65$ are found in both pseudo-octaves. Steps of the 12-tet scale are shown above for comparison.

7.1.5 Create an Instrument

Assuming adequate metal working skills and sufficient time, it would probably be possible to build a whole carillon of ting-bells: large ones to peal the deep notes and tiny ones to ring the highs. Exactly how to scale the proportions of the bell and how to choose appropriate materials so as to leave the timbral quality more or less unchanged are nontrivial issues, but with enough experimentation and dedication, these could likely be solved. This was exactly Harry Partch's situation when he found that his dream of playing in the 43-tone unequal scale could not be realized without instruments that could play in 43 tones per octave. Accordingly, he set out to build such instruments, and much of his career was devoted to instrument design, crafting, and construction. Until just a few years ago, embarking on a long and complex construction project would have been the only way to turn the ting-chime into reality.

Fortunately, today there is an easier way. Digital sampling technology is based on the idea of creating "virtual" instruments. Sound begins in a digital sampling keyboard[7] (a *sampler*) as a waveform stored in computer-like memory. This is processed, filtered, and modulated in a variety of ways, and then spread across the keyboard so that each key plays back the "same" sound, but at a different fundamental frequency. The (in)famous "dog-bark symphony" is a classic example where the vocalizations of man's best friend are tuned to a 12-tet scale and played as if it were a musical instrument. As general-purpose computers have become faster, software has become available for both synthesis and sample playback that can replace much of the external hardware.

[7] A detailed discussion of the design of samplers and other electronic musical instruments is well beyond the scope of this book. Sources such as De Furia [B: 38] provide an excellent introduction from a musicians perspective, and the engineer might wish to consult Rossing [B: 158] or DePoli [B: 40] for a more technological presentation.

The most exciting feature of many samplers (whether hardware or software) is that the user can specify both the waveform and the tuning; the sampler will then play back the chosen sound in the specified scale. In concrete terms, it is possible to transfer the sound file ting.wav from the computer into the sampler, and to then program the sampler so that it will play in the desired scale.[8] The musician can play the keyboard as a realistic simulation of a ting-carillon.

As the specifics of moving sound files from one machine to another are unique to the individual machines, they will not be discussed further: See your owners manual, software guide, or ask a friend. But one detail remains. Although we decided to use the eight-note unequal stretched-octave scale of the previous section, we did not decide how the scale steps were to be assigned to the keys of the keyboard. One possibility is to simply map successive scale tones to successive keys. Although this is often the most sensible strategy, in this particular case, there is a better way. As there are eight notes in the scale per pseudo-octave, and there are eight white notes per (normal, familiar) octave on the keyboard, the easiest mapping is the one shown in Fig. 7.4 in which each octave of the keyboard is used to play each pseudo-octave of the tingshaw scale.

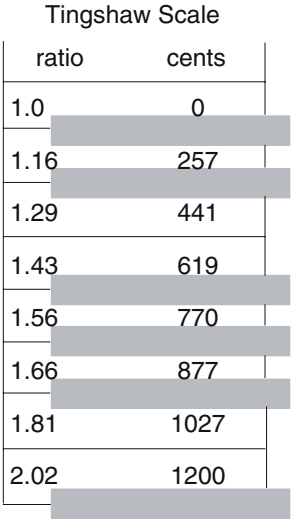


Fig. 7.4. Each pseudo-octave of the tingshaw scale can be readily mapped to the white keys on a standard keyboard.

[8] Transferring the wave file from the computer to the sampler can often be accomplished using software utilities available from the manufacturer or from third-party software companies. Each sampler has somewhat different internal specifications and limitations. For instance, some samplers only allow the pitch to be changed ± 1 semitone away from its 12-tet default value, whereas others allow arbitrary assignment of frequencies to keys of the keyboard. *Caveat emptor.*

7.1.6 Play Music

Most samplers have numerous options that let the musician manipulate certain features of the sound. Filters can be set to vary along with the note played, attack and decay parameters can be modulated by the key velocity (how rapidly the key is pressed), subtle pitch and timbral transformations can be programmed to respond to aftertouch (how hard the key is pressed), and reverberation and other effects can be added to simulate various auditory environments. All features of the sampler should be exploited, as seems appropriate to the sound.

For the tingshaw, I added a bit of reverberation to give the sound a more open feel, incorporated a subtle low-pass filter to subdue some harshness at the high end of the keyboard, and programmed the aftertouch to induce a delicate vibrato. Because the sound grew a bit mushy at the low end, I increased the speed of the attack for the lower notes. These are the kinds of modifications that any sound designer[9] would apply to make a more playable sound.

Now (finally!) comes the fun part. The tingshaw sound is spread across the keyboard in a virtual ting-carillon. Fingers are poised. This ting tolls for us.

7.2 Chaco Canyon Rock

The reddish rocks of Chaco Canyon (in New Mexico) produce colorful sounds as they scrape and clatter underfoot. They are musical, but inharmonic. They are resonant, but ambiguously pitched. While hiking the shale cliffs surrounding Chaco Canyon a few years ago, I was captivated by the music of these rocks. I hit them with sticks, struck them with mallets, and beat the rock against itself.

Figure 7.5 shows a typical sampled waveform. The large initial impact is rapidly damped, and the vibration is inaudible by 1/4 of a second. The shape of the waveform is irregular, although its envelope follows a smooth exponential decay. Using a digital sampler to pitch shift this sound across a keyboard creates a complete "lithophone" that sounds deep and resonant in the lower registers, natural in the middle range, and degenerates into a sharp plink when transposed into the far upper registers. The default operation of most samplers is to pitch shift the sound into the familiar 12-tet scale. Is this really the best way to tune a Chaco lithophone?

A little experimentation reveals that 12-tet works well for pieces that are primarily percussive, in which the sound envelope of one note dies away before

[9] I know of no single source containing a comprehensive discussion of sound design, although there are numerous articles spread throughout popular magazines such as *Electronic Musician* and *Keyboard* in which individual sound designers discuss their methods and philosophies.

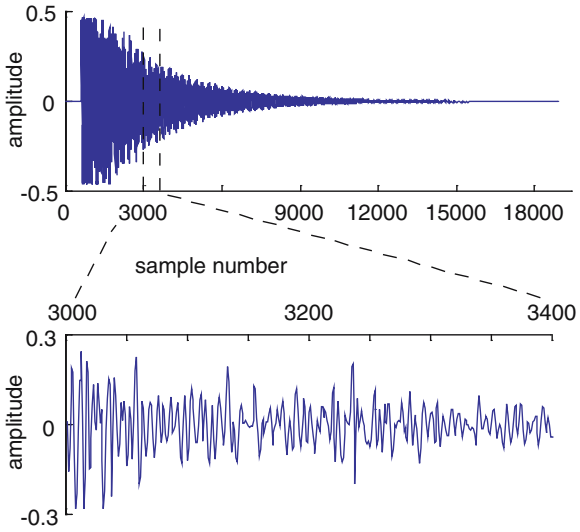


Fig. 7.5. Typical waveform of the Chaco rock when struck by a hard mallet. A small portion is expanded to make the irregularity of the waveform more apparent.

the next note begins. But denser pieces, and those with sustained tones[10] become increasingly dissonant, especially in the lower registers. This section details a systematic way to retune the pitches of the keyboard based on the spectrum of the rock sound so as to minimize the dissonance. The *Chaco Canyon Rock* (sound example [S: 44]) demonstrates many of the ideas.[11]

7.2.1 Find the Spectrum

Eventually, I settled on a favorite piece of rock. Roughly circular with a diameter of about 15 cm, it is less than a centimeter thick. It weighs 3 kilograms: lighter than it looks, but heavier than a cymbal of the same size. By striking it with different mallets in different places, it speaks in a remarkable variety of ways.

I sampled the rock 12 times[12] to try and capture the full range of its tonal qualities. Each sample was transferred to the computer, stored as a sound file, and analyzed by a 16K FFT. Most of the wavefiles (such as the one shown in Fig. 7.5 above) contained about 16K samples, and thus no windowing was needed. In a few cases, the wavefile was smaller than 16K samples. These were

[10] For instance, extreme time expansion can transform the sharp percussive envelope into a lengthy reverberation.

[11] This work on the Chaco rock was originally presented (in different form) at the *Synaesthetica* conference [B: 168].

[12] As before, at the standard rate of 44.1 KHz.

lengthened by zero padding, which augments the data with a string of zeroes. Three typical spectra are shown in the Fig. 7.6.

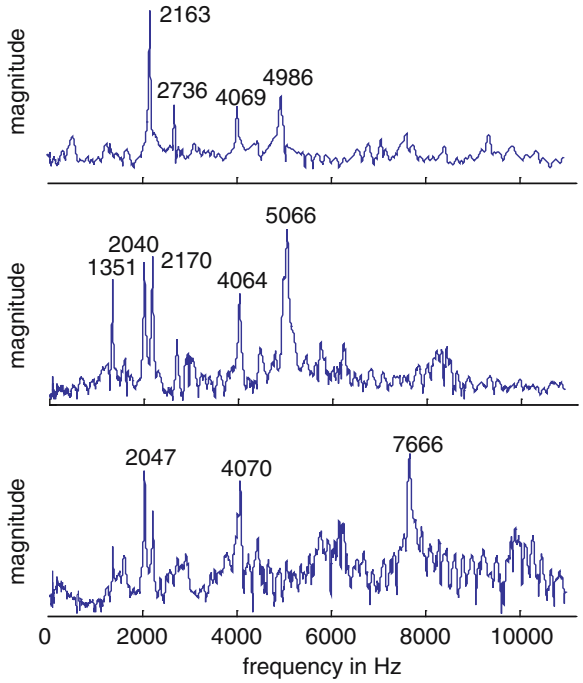


Fig. 7.6. Spectra of three different strikes of the Chaco Canyon rock.

7.2.2 Simplify the Spectrum

Each strike of the rock has a unique sound, and yet they are all clearly from the same source. The most constant mode (although rarely the loudest) is a high resonance near 4070 Hz. No matter how the rock is struck, no matter what mallet is used, this mode is audible. Other resonances occur in just one or two of the samples. For instance, the peak at 2736 in the top spectrum of Fig. 7.6 appears in only this one sample. Perhaps it was caused by the mallet, or perhaps this mode is very hard to excite, and I was lucky to find it. In either case, it is not a part of the generic sound of the rock.

Often, the loudest component of the sound is somewhere between 2040 and 2200. For instance, the most prominent partial in the top spectrum is at 2163. In the bottom spectrum, the dominant partial is at 2047, which may be reinforced by the (slightly flat) octave at 4070. At first, I thought these both represented a single dominant mode whose exact frequency varied somewhat with the situation. But by striking and listening carefully, it became clear that both really exist, as shown in the middle spectrum, where 2040 and 2170 are

present simultaneously. After playing around a bit, I realized that there are places on the rock face where it is possible to reliably predict which of these two modes will dominate. Moving the strike point back and forth causes the pitch of the rock to move up and down about a semitone. This makes sense because the ratio 2167/2040 is 105 cents. At least one of these two modes is present at all times, and this mode tends to determine the pitch. When both sound clearly, the pitch becomes more ambiguous.

As the partials near 5066 and 7666 are present in a number of samples other than the ones shown, they also form a part of the generic sound of the Chaco rock. The mode at 1351 is due to one particular edge of the rock. Whenever this edge is hit, the resonance at 1351 is excited. By striking elsewhere, the partial at 1351 is subdued.

Combining the above observations about the various modes of the rock, the "full" behavior can be approximated by forming the composite line spectrum in Fig. 7.7, which has spectral lines at 1351, 2040, 2167, 4068, 5066, and 7666. The exact values used for the amplitudes of the partials in the composite spectrum are somewhat arbitrary, but they are intended to reflect both the number of samples in which the mode appears and the amplitude of the partial within those samples.

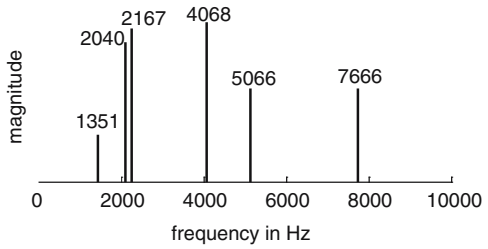


Fig. 7.7. The three spectra of the Chaco rock are combined to form a composite line spectrum that captures much of the acoustic behavior of the samples.

This is clearly not a harmonic sound, because the frequencies are not an integer multiple of any audible fundamental. The inharmonicity is evident to both the ear (the semitone between 2040 and 2167 is strikingly inharmonic) and to the eye (from the spectra).

7.2.3 Draw the Dissonance Curve

The composite spectrum for the Chaco rock shown in Fig. 7.7 can be entered into the dissonance calculating programs of the appendix in a straightforward way. Setting the frequency vector and amplitude vectors

```
freq=[1351, 2040, 2167, 4068, 5066, 7666]
amp=[0.2, 0.9, 0.9, 1.0, 0.5, 0.5]
```

gives the dissonance curve for the Chaco rock in Fig. 7.8, which shows the dissonance curve from unison to just a bit more than two octaves.

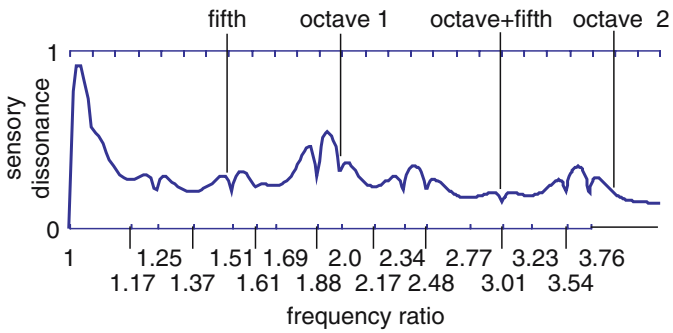


Fig. 7.8. Dissonance curve for the composite Chaco rock spectrum has 17 minima within a two-octave span. These are indicated by the tick marks on the horizontal axis. Upper axis shows 12-tet scale steps, with several extended for easy comparison.

Perhaps the most surprising features of this dissonance curve are the minima at the fifth, octave, and the octave plus fifth. A little thought (and some simple calculations) show that these are due to overlapping partials. When played at a ratio of 1.99, the 4068 partial of the lower tone coincides with the 2040 partial of the (almost) octave. When played at a ratio of 1.51, the 7666 partial of the lower tone coincides with the 5066 partial of the (almost) fifth. The minimum at 3.01 originates similarly from the coincidence of the 4068 and the 1351 partials.

Except for these familiar intervals, the inharmonic spectrum of the Chaco rock has a dissonance curve with minima that do not coincide with the notes of the 12-tet scale, and the most consonant intervals using the Chaco sound are different from the familiar consonant intervals defined by harmonic tones. Hence, the most consonant scale using the Chaco rock differs significantly from the familiar 12-tet scale.

7.2.4 Create an Instrument

Because it is illegal to remove material from a National Historical Site, quarrying rocks from Chaco Canyon and sculpting them into a giant lithophone is not feasible. Consequently, we will pursue a simulation strategy by building a virtual lithophone, which will be tuned by judicious use of the intervals from the dissonance curve.

Places where dips in the dissonance curve occur are intervals that sound most consonant. These points can be read directly from the figure and translated into their cent equivalents, which gives

0, 272, 386, 545, 713, 824, 908, 1093, 1200, 1472, 1572,
1764, 1908, 2030, 2188, and 2293.

Subtracting 1200 cents from each of the intervals in the second octave and rearranging shows that many of the intervals occur in both octaves, although some are markedly different:

0 272 386 545 713 824 908 1093
0 272 372 564 708 830 988 1093

Clearly, the final scale should contain the common intervals 0, 272, and 1093. Scale steps at 710 (a compromise between 708 and 713) and 827 (a compromise between 824 and 830) are sensible. As 908 and 988 are close to a semitone apart, it is reasonable to use both. Similarly, 545 and 564 differ significantly. As thirds are so important, we might also choose to use both 372 and 386 (which is exactly the just major third), giving three kinds of thirds: a flat minor third, a neutral third, and a just major third. This gives an 11-note scale. As it is much easier to play a tuning that repeats every 12 notes rather than 11, due to the physical layout of Western keyboards, perhaps we should add another note?

The largest step in the scale (by far) is the first interval of 272 cents. This seems like a reasonable place for an extra note because it might help to smooth a melody as it approaches or leaves the tonic. Recall from the previous discussion that it is possible to make the rock change pitch by about a semitone (105 cents) by striking it in different places. As this 105-cent interval naturally occurs within the stone, it is a reasonable "extra" interval. The full 12-note scale is defined in the Fig. 7.9, where the notes are shown mapped to a single octave of the keyboard from C to C.

Keyboard Layout for Chaco Tuning

interval	cents
1.0	0
1.063	105
1.17	272
1.24	372
1.25	386
1.37	545
1.385	564
1.507	710
1.612	827
1.69	908
1.77	988
1.88	1093
2/1	1200

Fig. 7.9. One possible keyboard layout for the Chaco lithophone repeats one full octave every 12 keys. Numbers give the tuning (in cents) of each key with respect to an arbitrarily chosen fundamental frequency f.

As the above discussion shows, there is nothing inevitable about this particular tuning. It is a compromise between faithfulness to the dissonance curve and finding a practical keyboard that is easy to play. Perhaps the most arbitrary decision in the whole process was to base the tuning on the octave. Although this is perfectly justified when focusing on the first octave, observe

that the second octave (marked "octave 2" in Fig. 7.8) does not occur at (or near) a local minimum.

7.2.5 Play Music

The performance molding capabilities of the sampler allow considerable freedom in sculpting the ultimate sound of the rock. Adding reverberation helps to counteract the rapid decay by creating a feeling of space. Imagine playing the lithophone in a hard-walled cavern where each stroke echoes subtly with its own reflection.

When playing the rock live, there are inevitable scraping and grating sounds as the mallet and rocks chafe and abrade. These "extraneous" sounds were mostly removed from the samples by careful sampling techniques, so that they would not influence the dissonance curve and the resulting scale. But now, to make the piece richer, I mixed them back in. Consequently, most of the rhythm track, and all of the rubbing and grating sounds were derived from the rock, albeit in a completely nontonal way.

To try and lighten the sound of the piece, I generated some noncorporeal (electronic) Chaco rocks. A number of interesting timbral variations are possible by using additive synthesis[13] in which the partial structure is specified from the composite spectrum of Fig. 7.7. These tend to be high and "electronic" sounding because they are much simpler than natural sounds, but they do help balance the heaviness of the raw rock samples. Because they are artificial, there is no constraint on their duration. In the first section of the piece, they are used as a soprano extension of the rock, whereas in the middle section they function more like an inharmonic rock organ.

Is music possible in such an idiosyncratic tuning, with such idiosyncratic timbres? Absolutely. Listen for yourself to the *Chaco Canyon Rock* in sound example [S: 44].

7.3 Sounds of Crystals

Sound is a kind of vibration, and there are many kinds of vibrations. For example, light and radio waves vibrate as they move through space. A stereo receiver works by translating electromagnetic vibrations into sound vibrations that you can hear. With such translations any type of vibration is a potential "sound." One kind of "noiseless" sound lurks in the molecular structure of everyday substances, and these sounds can be extracted using techniques of x-ray crystallography and additive synthesis.[14] Thus, the final example of this chapter begins with the "noiseless sound" of a crystal and realizes this in

[13] A program listing of a simple additive synthesis program is given in Appendix D.
[14] This idea was first reported in [B: 174].

a noisy, consonance-based way. The resulting piece, *Duet for Morphine and Cymbal*, appears in sound example [S: 45].

The simplest example of a noiseless sound is one that is pitched too low or too high for human ears to hear, like a dog whistle. Clearly, it is possible to record or sample a dog whistle, and to then play the sample back at a slower speed, thus lowering the pitch so that it can be heard. Another translation technique is employed by Fiorella Terenzi in *Music from the Galaxies* [D: 44]. Rather than beginning with a dog whistle, she uses digital recordings of the microwave radio emissions of various interstellar objects. These are slowed down until they are transposed into the audible range, and music (or at least sound) is created. Dr. Terenzi calls her work "acoustic astronomy." Amazingly enough, in Terenzi's work, outer space sounds just like you always thought it would.

7.3.1 Choose the Sound

There are other, less obvious noiseless sounds in nature. A technique called x-ray diffraction is a way of discovering and understanding the molecular structure of materials. The idea is to shine an x-ray beam (think of it as the beam of a flashlight) onto a crystalline structure. The x-rays, which vibrate as they move, pass through the crystal and are bent when they hit the atoms inside. Because of the pattern in which the atoms are arranged, the x-rays bend in a few characteristic directions.

This process, called diffraction, is at work in prisms and rainbows. When sunlight passes through a prism, it is broken apart into its constituent elements—the colors of the rainbow. Each color has a characteristic frequency, and each color is bent (or diffracted) through an angle that is proportional to that frequency. The same idea works with the diffraction of x-rays through crystals, but because the structure is more complicated, there is a correspondingly more complicated pattern, composed of beams of x-rays moving in different directions with different intensities.

These diffraction patterns are typically recorded and displayed graphically as a Fourier transform, a spectral chart that concisely displays the angle and intensity information. For example, the transform of the chemical bismuth molybdenum oxide ($Bi_2Mo_3O_{12}$) is shown in Fig. 7.10. The main scientific use of this technique is that each crystal has a unique transform, a unique signature. Unknown materials can be tested, and their transforms compared with known signatures. Often, the unknown material can be identified based on its transform, much as fingerprints are used to identify people.

In materials, any periodic physical structure (usually called a *crystal*) reflects electromagnetic energy (such as x-rays) in a characteristic way that can be decomposed into a collection of angles. The angle at which diffraction occurs quantifies the resonance point for vibrations in the crystal, although the vibrations here are of x-rays and not of air. Thus, the angle of the diffracted

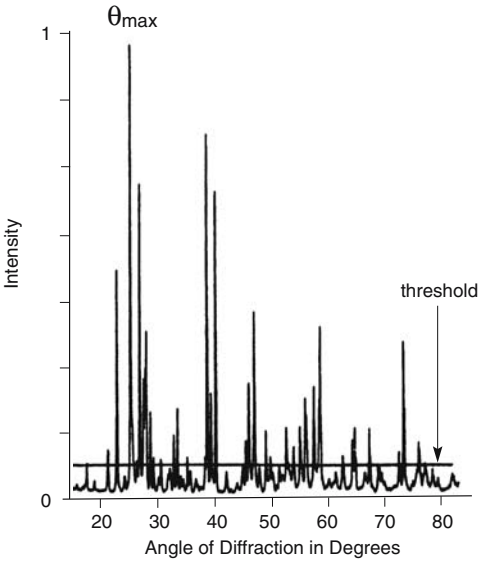


Fig. 7.10. This x-ray diffraction pattern is the (spatial) Fourier transform of the chemical bismuth molybdenum oxide. Using a simple mapping, it can be transformed into sound.

beam in crystallography plays a role similar to sine waves in sound, providing an analogy between the Fourier transform of the crystalline material and the Fourier transform of a sound. The intensity of the energy at each angle can be similarly translated into sound wave amplitudes. This then provides a basis for the mapping of x-ray diffraction data into sound data, and it defines a method of *auditory crystallography,* in which the spectrum of the crystal maps into the spectrum of a sound.

7.3.2 Find the Spectrum

A base frequency, or fundamental, must be chosen to realize the sound. This choice is probably best left to the performer by assigning various fundamentals to the various keys of a keyboard, allowing the "crystal tones" to be played in typical synthesizer fashion. In generating the sound data, the fundamental frequency is based on the angle, which has maximum intensity. Referring to Fig. 7.10, the largest spike occurs at an angle of about 25 degrees, which is labeled θ_{max}.

Each angle θ_i of the x-ray diffraction pattern can be mapped to a particular frequency f_i via the relation

$$f_i = \frac{\sin(\theta_{max})}{\sin(\theta_i)}$$

which transforms the x-ray diffraction angles into frequencies of sine waves. In general, angles that are less than θ_{max} are mapped to frequencies higher than the fundamental, whereas angles that are greater than θ_{max} are mapped

to lower frequencies. This feature of the mapping is responsible for much of the uniqueness of crystal sounds, because typical instrumental sounds have few significant partials below the fundamental. As both $\sin(\theta_i)$ and $\sin(\theta_{max})$ can take on any value between 0 and 1, f_i can be arbitrarily large (or small).

To see how the formula works, grab a calculator that has the sine function. For a θ_{max} of 25 degrees, calculate $\sin(\theta_{max}) = \sin(25) = 0.4226$ (if you get -0.1323, change from radians to degrees). To find the frequency corresponding to the spectral line at 41 degrees, calculate $\sin(41) = 0.6560$, and then divide $0.4226/0.6560 = 0.6442$. Thus, the frequency of this partial is 0.6442 times the frequency of the fundamental. For an A note at 440 Hz, this would be $440 \times 0.6442 = 283$ Hz.

The amplitude of each partial corresponds to the intensity of the θ_i, and it may be read directly from the graph. Referring to Fig. 7.10 again, the amplitude of the sine wave with frequency corresponding to an angle of 41 degrees is about 2/3 the amplitude of the fundamental. Designate the amplitude of the ith sine wave by a_i. Then the complete sound can be generated from the frequencies$f_1,\ f_2,\ f_3, ...$ with amplitudes $a_1,\ a_2,\ a_3, ...$ via the standard techniques of additive synthesis.

7.3.3 Simplify the Spectrum

As a practical matter, the number of different frequencies must be limited. The easiest method is to remove all angles with amplitudes below a given threshold. The threshold used for $Bi_2Mo_3O_{12}$, for example, is shown in Fig. 7.10. Using the formula of the previous section, the truncated x-ray diffraction pattern can be readily transformed into the set of partials shown in Fig. 7.11. The angle with the largest intensity in the diffraction pattern (about 25 degrees) corresponds to the partial with maximum amplitude, which appears at 950 Hz. Because the majority of larger angles in the diffraction pattern occur at angles larger than 25 degrees, the majority of partials in the resulting sound lie below 950 Hz. The clustering of partials near 500 Hz is perhaps the most distinctive feature of this sound.

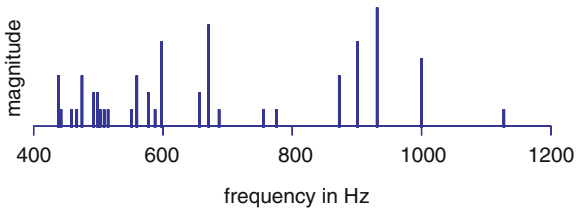


Fig. 7.11. The partials of the sound corresponding to the x-ray diffraction pattern for bismuth molybdenum oxide are tightly clustered.

It is feasible to create sounds from almost any material. Tom Staley and I [B: 174] experimented with a number of sound-materials, including glucose, tartaric acid, topaz, roscherite, reserpine, a family of Bismuth Oxides, cocaine, and THC.[15] One of my favorite sounding crystals was from morphine, and this sound is featured in the composition *Duet for Morphine and Cymbal*. There are numerous sources for x-ray diffraction data, which are available in technical libraries.

7.3.4 Dissonance Curve

Because crystal sounds like $Bi_2Mo_3O_{12}$[16] have a high intrinsic dissonance caused by tightly packed partials, the dissonance curves tend to be uniform, having neither deep minima nor large peaks. For instance, Fig. 7.12 shows that the dissonance curve for $Bi_2Mo_3O_{12}$ has eight minima within two octaves that are barely distinguishable from the general downward slope of the curve. Thus, no intervals are significantly more consonant than any others, and the rationale for defining the related scale via the dissonance curve vanishes.

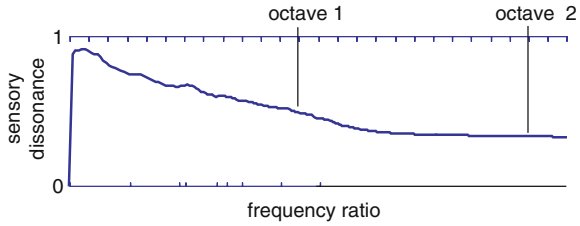


Fig. 7.12. Dissonance curve for bismuth molybdenum oxide has minima at the tick marks 1.2, 1.39, 1.42, 1.56, 1.61, 1.68, 1.89, and 2.13. The lack of any genuinely consonant intervals (no deep minima) suggests that these intervals might not produce a very convincing musical scale.

This problem with the dissonance curves of highly complex spectra is readily audible. Although the crystal spectra sound interesting, it is difficult to find any intervals at which the sounds can be reasonably played. Octaves, fifths, and the small dips in the dissonance curve all sound muddy in the lower registers, and clash disastrously in the higher registers. One solution is to return to the diffraction pattern and choose a higher threshold. This

[15] Listening to materials does not necessarily have the same effect as consuming them.

[16] I have used bismuth molybdenum oxide throughout this section to describe the process of transforming crystal data into sound (even though the musical composition is based on the spectrum of the morphine crystal) because I was unable to locate a clean x-ray diffraction graph for morphine.

will give a simpler spectrum and, hence, a more usable dissonance curve. The danger is that oversimplification may lose the essence of the original diffraction pattern.

Recall that points of minimum dissonance often develop because partials in two simultaneously sounding complex tones coincide, and that dissonance curves show the intervals at which a single sound can be played most consonantly. But if, as with the $Bi_2Mo_3O_{12}$ sound, there are no such intervals, another approach is needed. Perhaps consonance can be regained by *changing the spectrum along with the interval.* The simplest approach is to change the spectrum at each scale step, so that all partials coincide, no matter what scale steps are played. As will become clear, the total dissonance of any combination of scale steps need not exceed the intrinsic dissonance of the original sound.

7.3.5 Create an Instrument

Think of a "crystal instrument" in which each partial location defines a scale step. If the 25 partials of the bismuth molybdenum oxide sound of Fig. 7.11 are labeled $f_1,\ f_2, ..., f_{25}$, then the scale steps occur at precisely these frequencies. Construct a different spectrum at each scale step by choosing from among the remaining partials. For instance, the spectrum at f_1 might contain partials at $f_1,\ f_2,\ f_5,\ f_6,\ f_{10},\ f_{13},\ f_{16},\ f_{21}$, and f_{22}. Similarly, the spectrum at f_6 might contain $f_6,\ f_7,\ f_{13},\ f_{15},\ f_{17}$, and f_{20}. This is shown diagrammatically in Fig. 7.13, which displays a possible spectrum for each of the first 13 notes of the scale. Thus, each vertical stripe is a miniature line spectrum specifying the frequency and amplitude of the partials played when the key with "fundamental" f_i is pressed.

Observe that each spectrum contains a subset of the partials from the original crystal sound. When playing multiple notes, only partials that occur in the original sound are present, and hence, the dissonance cannot be significantly greater than the intrinsic dissonance of the original (it might increase somewhat because the partials in the combined sound can have different amplitudes than in the original). Each note contains only a small piece of the "complete" timbre, which is revealed only by playing various "chords" and tonal clusters.[17]

In terms of implementation, this is more complex than the previous two examples, because each key of the sampler must contain its own waveform (corresponding to the specified spectrum) and each spectrum must be created separately. Nevertheless, the process of generating 25 different spectra and assigning them to 25 different keys on the sampler is not particularly onerous, especially when much of the work can be automated by software.

[17] Essentially, the higher notes are pieces of a single grand über-chord. This is somewhat parallel to Rameau's fundamental bass, but for inharmonic sounds.

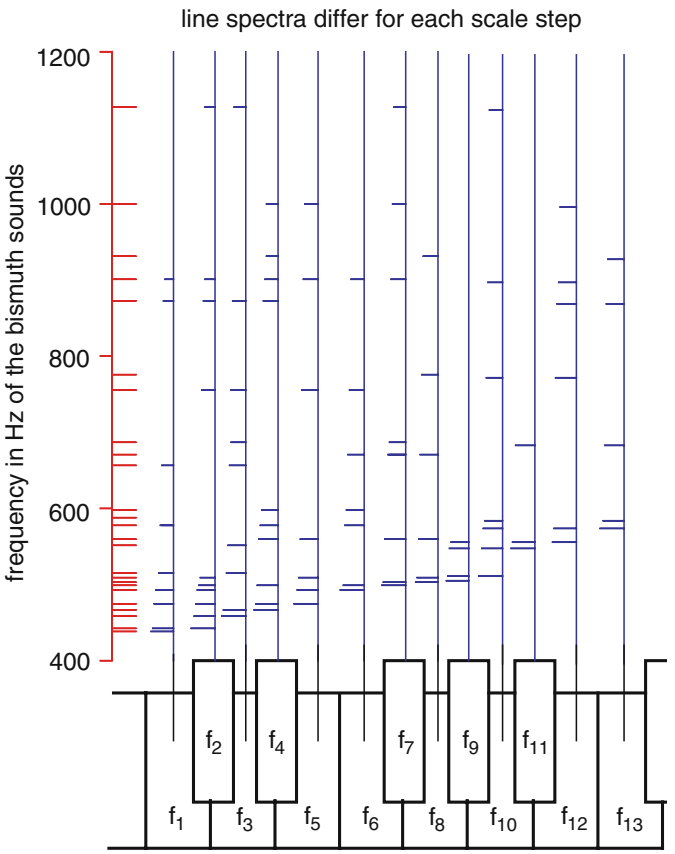


Fig. 7.13. The frequencies of the bismuth sound are used to construct a scale and a family of spectra consonant with that scale. Each scale step occurs with a fundamental f_i, and a possible line spectrum is shown for each.

7.3.6 Play Music

The most striking feature of crystal sounds is their inharmonicity. The spectra tend to be rich in frequencies within an octave of the fundamental because the major peaks of the diffraction pattern often lie in clusters. This is in stark contrast with conventional harmonic tones that consist of integer multiples of a single base frequency. Crystal spectra do not sound like standard musical instruments. A tempting analogy is with the inharmonic spectra of bells. When the crystal tones are struck, and the sound is allowed to die away slowly, they resonate much like a bell, although additive synthesis does not require the use of such a percussive envelope. Although some of the sounds (THC and roscherite, for instance) are very similar, most are distinct. Perhaps the closest comparison is with synthesizer voices with names like "soundtrack,"

"metal vapor," and "space pad," which give an idea of the subjective flavor of the sounds.

Because it has a distinct and complex quality, I chose to compose a piece using the sound of the morphine crystal, which was truncated so as to have 37 different partials. The 37-note partial-based scale was programmed into a sampler, and a "different" spectrum was assigned to each key, as in Fig. 7.13. The sounds were then looped, and performance parameters like modulation, aftertouch, and amplitude envelopes were added.

The keyboard is easy to play, although decidedly unfamiliar. As each note consists of partials aligned precisely with the partials of the crystal sound, it is almost impossible to hit "wrong" notes. Almost any combination of notes can be played simultaneously, creating unique tonal clusters. In essence, partial-based scales and spectra allow the performer to play with timbre directly, in a highly structured way. In the *Duet for Morphine and Cymbal*, complex clusters of tones are juxtaposed over a rhythmic bed supplied by the more percussive timbre of the cymbal. The bass line was created exactly as above, but with very simple spectra (only two or three partials per note) pitched well below the rest of the sound mass. Finally, a partial-based scale of pure sine waves was used for the melody lines.

7.3.7 The Sound of Data

Originally we had hoped that by listening to the sounds of crystalline structures, it would be possible to learn to identify the material from which the sound came, using the ear as an aid in data analysis. Although we have been unsuccessful in realizing this goal of auditory crystallography, "noiseless sounds" such as the spectral interpretation of x-ray diffraction data can provide a fruitful source of sounds and tunings. This gives a way to "listen" to crystal structures and to "play" the sounds of materials.

Imitative sound synthesis captures real sounds and places them inside musical machines. Audio crystallography begins with a conceptual sound (molecular resonances) that does not exist until it is mapped into the audio realm. There are many other sources of conceptual sound data. For instance, atomic resonances are often described via Fourier transforms, and they can be similarly converted to sound. At the other end of the time scale, planetary and stellar systems resonate and can be described using Fourier techniques.

Indeed, such explorations have already begun. Alexjander [B: 5] used transform data to generate musical scales in the article "DNA Tunings" and the CD *Sequencia* [D: 1], although the sounds used with these scales were standard synthesizer tones and acoustic instruments. Terenzi [D: 44] mapped data from radio telescopes into audio form. She comments, "The predominant microtonality of the galaxy is a fascinating aspect that could be explored... by creating new scales and timbres." Indeed, part of this book presents methods to carry out such exploration in a musical and perceptually sensible way.

7.4 Summary

In the pursuit of genuinely xenharmonic music that does not sacrifice consonance or depth of timbral material, this chapter presented three concrete examples of related tunings and spectra. The tingshaw bell and the Chaco rock showed how to take the spectrum of an existing sound, draw the dissonance curve, find the related scale, and build a playable "instrument." The crystal section showed how to take an arbitrary complex spectrum and to realize it in sound via a related partial-based scale.

Despite the odd timbres and scales, the resulting music gives an impression of tonality or key. It has the surface feeling of tonality, but it is unlike anything possible in 12-tet. McLaren comments[18]:

> *The Chaco Canyon Rock* bounces from one inharmonic "scale member" to another, producing an astonishing sense of consonance. The effect isn't identical to traditional tonality–yet it produces many of tonality's effects. One is instantly aware of "right" and "wrong" pitches, and there is a sense of spectral "progression."

We call such music *xentonal*.

With the intent of making this chapter a "how to" manual, no amount of detail was spared. Each of many agonizing compositional, technical, and creative decisions was discussed, the options weighed, and then one way was chosen. Other paths, other choices of analysis methods, windowing techniques, scale steps, performance parameters, keyboard mappings, and so on, would have led to different compositions. Thus, the complete process, as outlined in the above six steps, is not completely mechanical, and there are numerous technical and artistic pitfalls. Although the bell, the rock, and the crystal were used throughout as examples, the methods readily apply to any sound, although they are most useful with inharmonic sounds.

It is often desirable to augment the original sound with other complementary tones, and there are three approaches to creating new sounds that are fully consonant with the original. Additive synthesis has already been mentioned several times as one way to augment the timbral variation of a piece. The use of partial-based scales is not limited to sounds created from x-ray crystallography, and it can be readily applied in other situations. The third technique, called spectral mappings, is a way of transforming familiar instrumental sounds into inharmonic versions that are consonant with a desired "target" spectrum. This is discussed at length in the chapter "Spectral Mappings."

[18] In *Tuning Digest* 120.

8

Adaptive Tunings

Throughout the centuries, composers and theorists have wished for musical scales that are faithful to the consonant simple integer ratios (like the octave and fifth) but that can also be modulated to any key. Inevitably, with a fixed (finite) scale, some intervals in some keys must be compromised. But what if the notes of the "scale" are allowed to vary? This chapter presents a method of adjusting the pitches of notes dynamically, an adaptive tuning*, that maintains fidelity to a desired set of intervals and can be modulated to any key. The adaptive tuning algorithm changes the pitches of notes in a musical performance so as to maximize sensory consonance. The algorithm can operate in real time, is responsive to the notes played, and can be readily tailored to the spectrum of the sound. This can be viewed as a generalized dynamic just intonation, but it can operate without specifically musical knowledge such as key and tonal center, and it is applicable to timbres with inharmonic spectra as well as the more common harmonic timbres.*

8.1 Fixed vs. Variable Scales

A musical scale typically consists of an ordered set of intervals that (along with a reference frequency such as $A = 440$ Hz) define the pitches of the notes used in a given piece. As discussed at length in Chap. 4, different scales have been used in different times and places, and scales are usually thought of as being fixed throughout a given piece, and even throughout a complete repertoire or musical genre. However, even master performers may deviate significantly from the theoretically ideal pitches [B: 21]. These deviations are not just arbitrary inaccuracies in pitch, but they are an important expressive element. One way to model these pitch changes is statistically[1]; another is to seek criteria that govern the pitch changes. For example, the goal might be to play in a just scale that maximizes consonance even though the piece has complex harmonic motion. The key is to use a variable scale, an *adaptive tuning* that allows the tuning to change dynamically while the music is performed. The trick is to specify sensible criteria by which to retune.

[1] As suggested in [B: 4] and discussed in Sect. 4.8.

Imagine a trumpet player. When performing with other brasses, there is a temptation to play in the tuning that originates naturally from the overtones of the tubes. When performing with a fixed pitch ensemble, the temptation is to temper the pitches. Similarly, a violinist may lock pitch to the overtones of others in a string quartet but may temper toward 12-tet when playing with keyboard accompaniment. Some a capella singers (such as Barbershop quartets) are well known to deviate purposefully from 12-tet so as to lock their pitches together. Eskelin[2] advises his choral singers to "sing *into* the chord, not through it," to "lock into the chord." In all of these cases, performers purposely deviate from the theoretically correct 12-tet scale, adjusting their intonation dynamically based on the musical context. The goal of an adaptive tuning is to recapture some of these microtonal pitch variations, to allow traditionally fixed pitch instruments such as keyboards an added element of expressive power, to put a new musical tool into the hands of performers and composers, and to suggest a new theory of adaptive musical scales.

8.1.1 Approaches to (Re)tuning

The simplest kind of tuning that is responsive to the intervals in a piece uses a fixed scale within the piece but retunes between pieces. There is considerable historical precedent for this sensible approach. Indeed, harpsichordists regularly retune their instruments (usually just a few notes) between pieces. Carlos [B: 23] and Hall [B: 68] introduced quantitative measures of the ability of fixed scales to approximate a desired set of intervals. As different pieces of music contain different intervals, and because it is mathematically impossible to devise a single fixed scale in which all intervals are perfectly in tune, Hall [B: 68] suggests choosing tunings based on the piece of music to be performed. For instance, if a piece has many thirds based on C, then a tuning that emphasizes the purity of this interval would be preferred. An elegant early solution to the problem of comma drift in JI uses two chains of meantone a perfect fifth apart. This was proposed by Vicentino in 1555 [B: 199] and is explored in [W: 32]. The *Groven* System[3] allows a single performer to play three acoustic pianos that together are tuned to a 36-tone just scale.

8.1.2 Approaches to Automated (Re)tuning

With the advent of electronics, Polansky [B: 142] suggests that a "harmonic distance function" could be used to make automated tuning decisions, and points to the "intelligent keyboard" of Waage [B: 202] that uses a logic circuit to automatically choose between alternate versions of thirds and sevenths depending on the musical context. As early as 1970, Rosberger [B: 155] proposed a "ratio machine" that attempts to maintain the simplest possible integer ratio

[2] From [B: 54]. Discussed more fully on p. 63.
[3] Described at `http://vms.cc.wmich.edu/~code/groven`

intervals at all times. Expanding on this idea, Denckla [B: 39] uses sophisticated tables of intervals that define how to adjust the pitches of the currently sounding notes given the musical key of the piece. The problem is that the tables may grow very large, especially as more contextual information is included. A modern implementation of this idea can be found in the *justonic* tuning system [W: 14], which allows easy switching between a variety of scales as you play. Frazer has implemented a dynamic tuning in the Midicode Synthesizer [W: 11] that allows the performer to specify the root of the retuned scale on a dedicated MIDI channel. The *hermode* tuning [W: 15] "analyses chords and immediately adjusts the pitch of each note so that the prominent harmonics line up." Through its numerous sound examples, the website provides a strong argument for the use of tunings that can continuously adjust pitch. The method is discussed further in Sect. 8.2. Another modern implementation of a dynamic tuning is included in Robert Walker's *Fractal Tune Smithy* [W: 31], which microtonally adjusts the pitch of each new note so as to maximize the number of consonant dyads currently sounding.

Partch had challenged [B: 128] that "it is conceivable that an instrument could be built that would be capable of an automatic change of pitch throughout its entire range." The hermode tuning system is one response. Another approach is John deLaubenfels' [W: 7] spring-mass paradigm that models the tension between the currently sounding notes (as deviations from an underlying just intonation template) and adapts the pitches to relax the tension. This spring model, detailed in Sect. 8.3, provides a clear physical analog for the operation of adaptive tunings.

The bulk of this chapter realizes Partch's challenge using a measure of consonance as its "distance function" to change the pitches of notes dynamically (and in real time) as the music is performed.[4] As we will see, the strategy can maintain a desirable set of intervals (such as the small integer ratios) irrespective of starting tone, transpositions, and modulations. In addition, the adaptive tuning is responsive to the spectrum of the instruments as they are played. Recall that the dissonance function $D_F(\alpha)$ describes the sensory dissonance of a sound with spectrum F when played at intervals α. Values of α at which local minima of the dissonance function occur are intervals that are (locally) maximally consonant. The adaptive tuning algorithm calculates the (gradient of the) dissonance at each time step and adjusts the tuning of the notes toward the nearest minimum of the dissonance curve.

8.2 The Hermode Tuning

The hermode tuning, created in 1988 by Werner Mohrlok ([B: 48], [W: 15]), is a method of dynamically retuning electronic musical instruments in real time so as to remove tuning errors introduced by the equal-tempered scale.

[4] This first appeared in [B: 167], from which key elements of this chapter are drawn.

In order to help retain compatibility with standard instruments playing in standard tunings, the hermode tuning adjusts the absolute pitches so that the sum of the pitch deviations (in cents from the nominal 12-tet) is zero.

The process begins with an analysis of the currently sounding notes. For example, suppose that C, E, and G are commanded. The system detects the C major chord and consults a stored table of retunings, finding (in this case) that the E should be flattened by 14 cents and the G sharpened by 2 cents to achieve a justly intoned chord. All three notes are then raised in pitch so that the average deviation is zero, as illustrated in Fig. 8.1. In its normal operation, the analysis proceeds by reducing all notes to one octave, which greatly simplifies the tables needed to store the retuning information.

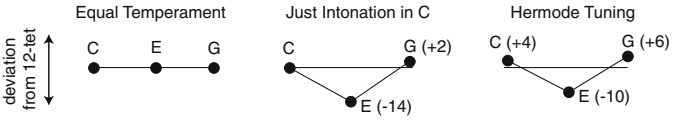


Fig. 8.1. The hermode tuning retunes chords to just intervals while centering the pitches so that the sum of all deviations is zero. This helps to maintain horizontal consistency and compatibility with standard instruments.

"Hermode" is a contraction and anglicization of *harmonischer modus*, which translates roughly as "modes of just intonation." Thus, the goal of the hermode system is to automatically retune the keyboard into a form of just intonation while retaining the ability to perform in concert with other instruments. For example, when the same note appears in successive chords, certain (vertical) intervals may be tempered to disguise the (horizontal) motion. In order to counteract possible drifts of the tuning, the hermode tuning does not allow the level of any chord pattern to be retuned more than ± 20 cents, which effectively limits the retuning of any given note to within ± 30 cents (except for some of the sevenths). Finally, when many notes are sounding simultaneously and the optimal tuning becomes ambiguous, the frequencies of the notes are controlled to the best horizontal line. A complete description of the hermode tuning can be found in Mohrlok's paper "The Hermode Tuning System," which is available electronically on the CD [W: 26].

The hermode tuning can operate in several modes. These provide different ways to ensure that the retuned pitches remain close to 12-tet and pragmatic features aimed at making the system flexible enough for real time use. Some of these are:

(i) A mode that only adjusts thirds and fifths
(ii) A mode that includes adaptation of sevenths
(iii) A mode that considers the harmonic center of a piece

(iv) A mode containing a depth parameter that allows the performer to use the hermode tuning at one extreme and equal temperament at the other extreme

The hermode tuning is currently implemented in the Waldorf "Q" synthesizer [W: 34], in the Access "Virus" [W: 33], in organs by Content [W: 5], and will soon be added to a number of software synthesizers. Theoretically, the hermode tuning generalizes just intonation in at least two senses. First, it is insensitive to the particular key of the piece; that is, the same tuning strategy "works" in all keys. Second, because the level at which the tunings are equalized (above and below equal temperament) is allowed to fluctuate with the music, there is no absolute tonal center.

8.3 Spring Tuning

To see why adaptive tunings are not completely straightforward to specify and implement, consider trying to play the simple four-note chord C, D, G, and A in a hypothetically perfect intonation in which all intervals are just. The fifths can be made just (each with 702 cents) by setting $C = 0$, $D = 204$, $G = 702$, and $A = 906$ cents.[5] But C to A is a sixth; if this is to be a just major sixth, it must be 884 cents.[6] Clearly, $884 \neq 906$, and there is a problem. Perfection is impossible, and compromise is necessary.

John deLaubenfels' approach [W: 7], developed in 2000, defines a collection of tuning "springs," one for each of the just intervals. As shown in Fig. 8.2, each spring connects two notes; the spring is at rest when the notes are at a specified just interval i. If the interval between the notes is wider than i, the springs pull inward to narrow it. If the notes are tuned too closely, the spring pushes the pitches apart. Once all pairs of notes are connected with appropriate springs, the algorithm simulates the tugging of the springs. Eventually, the system reaches equilibrium where the intervals between the notes have stabilized at a compromise tuning that balances all competing criteria.

For example, the right-hand side of Fig. 8.2 shows the four note-chord C, D, G, and A along with the appropriate assignments of desired intervals to springs. As the tuning of the fifths and sixths cannot all be pure simultaneously, the springs move the pitches slightly away from the just intervals. The exact values achieved depend on the strength of the springs; that is, the constants that specify the restoring force of the springs as a function of displacement. The spring tuning presumes that the "pain" caused by deviations in tuning (measured in cents) is proportional to the square of the pitch change. Thus, pain is analogous to energy (because the energy stored in a linear spring

[5] C to G is 702 cents and G to D is also 702 cents. Hence, C to D is 1404 cents, which is octave reduced to 204 cents. D to A is then $204 + 702 = 906$ cents.

[6] Recall Table 4.2 on p. 60.

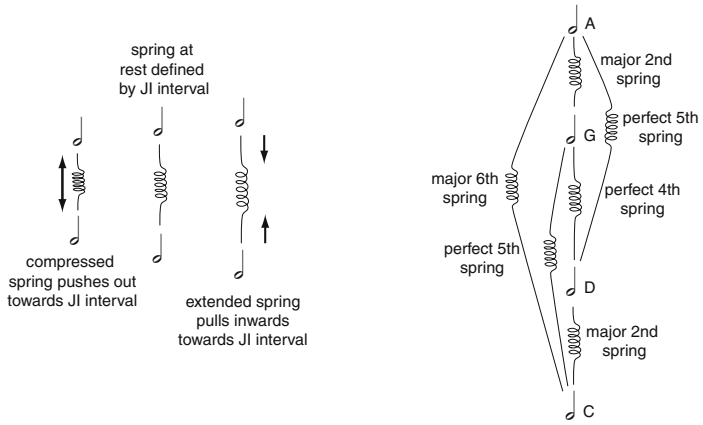


Fig. 8.2. Springs are at rest when the notes are at their assigned just intervals. Once all notes are connected by a network of springs (the right-hand network shows the four-note chord C, D, G, A and its springs), the algorithm simulates the pushing and pulling of springs. At convergence, a compromise tuning is achieved.

is proportional to the square of the displacement), and the goal of the spring tuning is to minimize the pain.

The mistuning of simultaneously sounding notes is only one kind of pain that can occur in a variable tuning. A second kind occurs when the same note is retuned differently at different times. This happens when the note appears in different musical contexts, i.e., in different chords, and it may be disconcerting in melody lines and in sustained notes when it causes the pitch to waver and wiggle. The third kind occurs when the whole tuning wanders up or down. All three of these issues are discussed in detail in the context of the adaptive tuning algorithm of Sect. 8.4.

For the spring tuning, there is an elegant solution: Assign new kinds of springs to deal with each new kind of pain. For example, Fig. 8.3 shows a collection of springs connected horizontally between successive occurrences of the same notes. Observe that these springs do not pull horizontally in time, but vertically in pitch. Strengthening the springs ensures less wavering of the pitches across time, but it pulls the vertical harmonies further from nominal. Weakening these springs allows more variation of the pitches over time and closer vertical harmonies. Similarly, "grounding" springs can be assigned to combat any tendency of the tuning to drift. This can be implemented by connecting springs from each note to the nearest 12-tet pitch (for instance).

Thus, there are three ways that the tuning can deviate from ideal and three kinds of springs: Across each vertical interval is a spring that pulls toward the nearest just ratio, horizontal springs control the instability of pitches over time, and grounding springs counteract any global wandering of the tuning.

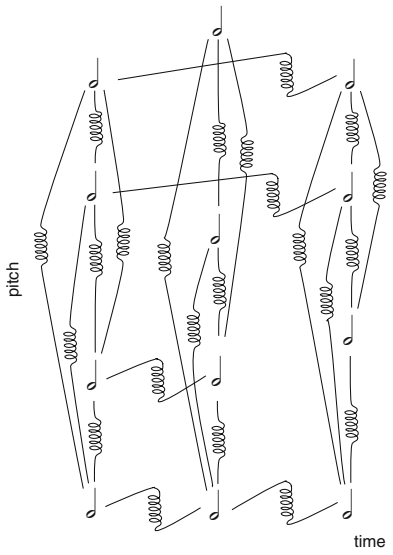


Fig. 8.3. When notes are allowed to vary in pitch, a C note in one chord may differ in pitch from the "same" C note in another. This wandering of pitches can be controlled by assigning a second set of springs between the same notes occurring at different (nearby) times. These springs are drawn vertically because they do not pull horizontally (in time), but only vertically (in pitch).

The model has several parameters that directly influence how the retuning proceeds:

(i) The strength of the vertical springs may differ for each interval type.
(ii) The strength of the horizontal strings may differ depending on the distance in time. Setting all horizontal springs completely rigid allows the same algorithm to find an "optimal" fixed tuning.[7]
(iii) The strength of the grounding springs may differ to specify the fidelity to the underlying fixed tuning.
(iv) The strength of the springs may be a function of the loudness of the notes.
(v) The time interval over which events are presumed to be simultaneous may be changed.
(vi) There may be a factor that weakens the horizontal springs when many notes are sounding.

The large number of parameters allows considerable flexibility in the implementation and may be changed based on individual taste. For example, a listener preferring pure intervals may de-emphasize the strength of the horizontal springs whereas a listener who dislikes wavering pitches may increase the strength of the horizontal springs. One thorny issue lies in the automatic

[7] In a preferred (non-real-time) application of the spring tuning, this "calculated optimum fixed tuning" (COFT) can be used as a starting point for further adaptation by tying the grounding springs to the COFT. This helps to lend horizontal consistency to the retuned piece. The COFT is analogous to the procedure applied to the Scarlatti sonatas in Sect. 11.2 using the consonance-based algorithm.

specification of which size or kind of spring should be assigned to each interval. For example, the just interval of a major second may be represented by the frequency ratio $\frac{10}{9}$, by $\frac{9}{8}$, or by $\frac{8}{7}$, depending on the musical context. In the spring tuning, this fundamental assignment must be made in a somewhat ad hoc manner, unless some kind of extra high-level logic is invoked. In one implementation, dissonances such as the major and minor seconds are not tied together with springs (equivalently, the spring constants are set to zero). A number of retunings of common practice pieces are available at deLaubenfels' personal web page, see [W: 7].

8.4 Consonance-Based Adaptation

Another way of creating an adaptive tuning is to calculate the sensory dissonance of all notes sounding at each time instant and to move the pitches so as to decrease the dissonance. Picture the mountainous contour of a dissonance curve such as Fig. 8.4. If the musical score (or the performer) commands two notes that form the interval α_1, then consonance can be increased by making the interval smaller. If the score commands α_2, the consonance can be increased by making the interval larger. In both cases, consonance is increased by sliding downhill, and dissonance is increased by climbing uphill. As the minima of the dissonance curve define the related scale, the simple strategy of always moving downhill provides a musically sensible way to automatically play in the related scale. This is the idea behind the adaptive tuning algorithm.

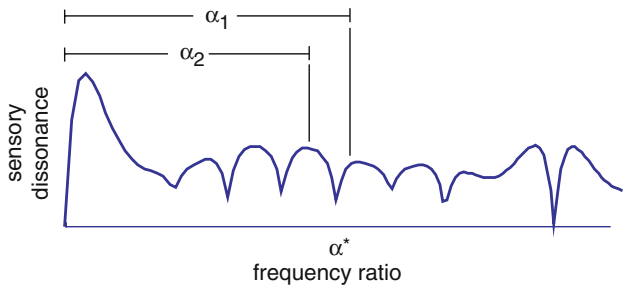


Fig. 8.4. Any interval between α_1 and α_2 is dynamically retuned by sliding downhill on the dissonance curve to the nearby local minimum at α^*. This adaptive tuning strategy provides a way to automatically play in the related scale.

The algorithm must have access to the spectra of the sounds it is to adjust because dissonance curves are dependent on the spectra. This information may be built-in (as in the case of a musical synthesizer or sampler that inherently "knows" the timbre of its notes), or it may be calculated (via a Fourier

transform, for instance). The algorithm adjusts the pitch of each note so as to decrease the dissonance until a nearby minimum is reached. This modified set of pitches (or frequencies) is then output to a sound generation unit. Thus, whenever a new musical event occurs, the algorithm calculates the optimum pitches so that the sound (locally) minimizes the dissonance.

There are several possible ways that the necessary adjustments can be carried out. Consider the simple case of two notes with pitches F_1 and F_2 (with $F_1 < F_2$). With no adaptive tuning, the interval F_2/F_1 will sound. The simplest adaptive strategy would be to calculate the dissonances of the intervals $F_2/F_1+\epsilon$ for various values of ϵ, (appropriate ϵ's could be determined by the bisection method, for instance). The point of minimum dissonance is given by that value of ϵ for which the dissonance is smallest. The pitches of F_1 and F_2 are then adjusted by an appropriate amount, and the more consonant interval sounded.

This simple search technique is inefficient, especially when it is necessary to calculate the dissonance of several simultaneous notes.[8] The *gradient descent* method [B: 205] is a better way to find the nearest local minimum of the dissonance curve. Suppose that m notes, each with spectrum F are desired. Let $f_1 < f_2 < ... < f_m$ represent the fundamental frequencies (pitches) of the notes. A *cost* function D is defined to be the sum of the dissonances of all intervals at a given time,

$$D = \sum_{i,j} D_F(\frac{f_i}{f_j}). \tag{8.1}$$

An iteration is then conducted that updates the f_i by moving downhill over the m dimensional surface D. This is

$$\left\{ \begin{array}{c} new \\ frequency \\ values \end{array} \right\} = \left\{ \begin{array}{c} old \\ frequency \\ values \end{array} \right\} - \{stepsize\}\{gradient\} \tag{8.2}$$

where the *gradient* is an approximation to the partial derivative of the cost with respect to the i^{th} frequency. The minus sign ensures that the algorithm descends to look for a local minimum (rather than ascending to a local maximum). More concretely, the algorithm is:

[8] The number of directions to search increases as 2^m, where m is the number of notes.

Adaptive Tuning Algorithm

do
  for $i = 1$ to m

$$f_i(k+1) = f_i(k) - \mu \frac{dD}{df_i(k)} \tag{8.3}$$

  endfor
until $|f_i(k+1) - f_i(k)| < \delta$ for all i

where k is an iteration counter. Thus, the frequencies of all notes are modified in proportion to the change in the cost and to the stepsize μ until convergence is reached, where convergence means that the change in all frequencies is less than some specified δ. Some remarks:

(i) δ should be chosen based on the tuning accuracy of the sound generation unit.
(ii) It may sometimes be advantageous to fix the frequency of one of the f_i and to allow the rest to adapt relative to this fixed pitch.
(iii) It is sensible to carry out the adaptation with a logarithmic stepsize, that is, one that updates the frequency in cents rather than directly in Hertz.
(iv) It is straightforward to generalize the algorithm to retune any number of notes, each with its own spectral structure.
(v) A detailed discussion of the calculation of $\frac{dD}{df_i(k)}$ is given in Appendix H.
(vi) There are many ways to carry out the minimization of D. An iterative algorithm is proposed because closed-form solutions for the minima are only possible in the simplest cases.
(vii) If desired, the adaptation can be slowed by decreasing the stepsize. Outputting intermediate values causes the sound to slide into the point of maximum consonance. This is one way to realize Darreg's vision of an "elastic" tuning [B: 36].

8.5 Behavior of the Algorithm

This section examines the adaptive tuning algorithm by looking at its behavior in a series of simple situations. Any iterative procedure raises issues of convergence, equilibria, and stability. As the adaptive tuning algorithm is defined as a gradient descent of the dissonance D, such analysis is conceptually straightforward. However, the function D is complicated, its error surface is multimodal, and exact theoretical results are only possible for simple combinations of simple spectra. Accordingly, the analysis focuses on a few simple

settings, and examples are used to demonstrate which aspects of these simple settings generalize to more complex (and hence more musically interesting) situations. The next few examples (which are formalized as theorems in Appendix H) show the close relationship between the behavior of the algorithm and the surface formed by the dissonance curve. In effect, the behavior of the algorithm is to adjust the frequencies of the notes so as to make a controlled descent of the dissonance curve.

8.5.1 Adaptation of Simple Sounds

The simplest possible case considers two notes F and G, each consisting of a single partial. Let f_0 and g_0 be the initial frequencies of the two sine wave partials, with $f_0 < g_0$, and apply the adaptive tuning algorithm. Then either

(i) f_k approaches g_k as k increases
(ii) f_k and g_k grow further apart as k increases

To see this graphically, picture the algorithm evolving on the single humped dissonance curve of Fig. 8.5. If the initial difference between f_0 and g_0 is small, then the algorithm descends the near slope of the hump, driving f_k and g_k closer together until they merge. If the difference between f_0 and g_0 is large, then the algorithm descends the far side of the hump and the dissonance is decreased as f_k and g_k move further apart. The two partials drift away from each other. (This is conceptually similar to the "parameter drift" of [B: 172], where descent of an error surface leads to slow divergence of the parameter estimates.) Together, (i) and (ii) show that the point of maximum dissonance (the top of the hump) is an unstable equilibrium.

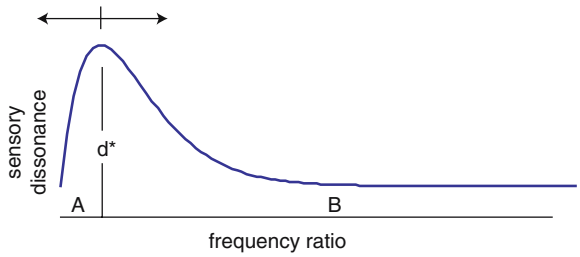


Fig. 8.5. Dissonance between two notes f and g, each a pure sine wave. There are two possible behaviors as the adaptive tuning algorithm is iterated, depending on the starting frequency. If g is in region A, then g ultimately merges with f. If g is in region B, then g and f ultimately drift apart.

For sounds with more complex spectra, more interesting (and useful) behaviors develop. Figure 8.6 shows how interlaced partials can avoid both drifting and merging. Suppose that the note F consists of two partials fixed at

frequencies f and αf with $\alpha > 1$, and that G consists of a single partial at frequency g_0 that is allowed to adapt via the adaptive tuning algorithm. Then:

(i) There are three stable equilibria: at $g = f$, at $g = \alpha f$, and at $g = (1 + \alpha)f/2$
(ii) If g_0 is much less than f, then g_k drifts toward zero
(iii) If g_0 is much greater than f then g_k drifts toward infinity

The regions of convergence for each of the possible equilibria are shown below the horizontal axis of Fig. 8.6. As in the first example, when g is initialized far below f or far above αf (in regions A or E), then g drifts away, and if g starts near enough to f or αf (in regions B or D), then g ultimately merges with f or αf.

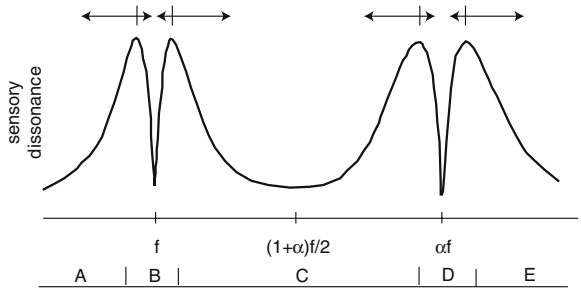


Fig. 8.6. Dissonance between a note with two fixed partials at f and αf, and a note with a single partial g, as a function of g. There are five possible behaviors as the adaptive tuning algorithm is iterated, depending on the starting frequency. If g begins in region A, then g drifts toward zero. If g begins in region B, then g merges with f. If g begins in region C, then g has a minimum at $\frac{(1+\alpha)f}{2}$. If g begins in region D, then g merges with αf. If g begins in region E, then g drifts toward infinity.

The interesting new behavior in Fig. 8.6 occurs in region C where g is repelled from both f and αf and becomes trapped at a new minimum at $\frac{(1+\alpha)f}{2}$. In fact, this behavior is generic—sandwiched partials typically reduce dissonance by assuming intermediate positions. This is fortunate, because it gives rise to many of the musically useful properties of adaptive tunings. In particular, sets of notes with interlaced partials do not tend to drift apart because it is difficult for partials to cross each other without a rise in dissonance.

To be concrete, consider two notes, F with partials at frequencies $(f_0, f_1, ...f_n)$ and G with partials at frequencies $(g_0, g_1, ...g_m)$. Suppose that g_i is sandwiched between f_j and f_{j+1},

$$f_j < g_i < f_{j+1},$$

and that all other partials are far away

$$f_{j-1} << f_j,\ f_{j+1} << f_{j+2}$$
$$g_{i-1} << f_j,\ f_{j+1} << g_{i+1}.$$

Then the dissonances (and their gradients) between g_i and the f_i are insignificant in comparison with the dissonances between g_i and the nearby frequencies f_j and f_{j+1}. Thus, g_i acts qualitatively like the g of Fig. 8.6 as it is adjusted by the adaptive tuning algorithm toward some intermediate equilibrium. Of course, the actual convergent value depends on a complex set of interactions among all partials, but g_i tends to become trapped, because approaching either f_j or f_{j+1} requires climbing a hump of the dissonance curve and a corresponding increase in dissonance.

8.5.2 Adapting Major and Minor Chords

As more notes are adapted, the error surface increases in dimension and becomes more complex. Notes evolve on an m-dimensional sheet that is pocketed with crevices of consonance into which the algorithm creeps. Even a quick glance at Appendix H shows that the number of equations grows rapidly as the number of interacting partials increases.

To examine the results of such interactions in a more realistic situation, Table 8.1 reports converged values (in Hertz, accurate to the nearest integer) for triads played with harmonic tones with varying numbers of partials. In each case, the algorithm is initialized with fundamental frequencies that correspond to the 12-tet notes C, $E\flat$, G (a minor chord) or to C, E, G (a major chord), and the algorithm is iterated until convergence. No drifting notes or divergence occurs because the partials of the notes are interlaced. In all cases, the fifth (the interval between C and G) remains fixed at a ratio of 1.5:1. For simple two and three partial notes, the major and minor chords merge, converging to a "middle third" that splits the fifth into two parts with ratios 1.21 and 1.24. With four partials, the middle third splits the fifth into two nearly equal ratios of 1.224.

Table 8.1. Converged major and minor chords differ depending on the number of harmonic partials they contain.

Initial notes in 12-tet	Initial frequencies	Converged frequencies (2–3 partials)	Converged frequencies (4 partials)	Converged frequencies (5–16 partials)
C	523	523	523	523
E♭	622	647	641	627
G	784	784	784	784
C	523	523	523	523
E	659	647	641	654
G	784	784	784	784

For notes with five or more partials (up to at least 16), the two initializations evolve into distinct musical entities. The major chord initialization converges to a triad with ratios 1.2 and 1.25, and the minor chord initialization converges to a triad with the inverted ratios 1.25 and 1.2. These are consistent with the minor and major thirds of the just intonation scale, suggesting that performances in the adaptive tuning are closely related to a just intonation when played with harmonic timbres of sufficient complexity. Thus, when the sounds have a harmonic spectra, the action of the adaptive tuning algorithm is consistent with just intonation.

8.5.3 Adapting to Stretched Spectra

When the spectra deviate from a harmonic structure, however, the justly tuned intervals are not necessarily consonant, and the adaptation operates so as to minimize the sensory consonance. In extreme cases, it is easy to hear that the ear prefers consonance over justness. A particularly striking example is the use of sounds with stretched (and/or compressed) spectra as in the *Challenging the Octave* sound example [S: 1] from Chap. 1.

Consider an inharmonic sound with partials at

$$f,\ 2.1f,\ 3.24f,\ 4.41f,\ \text{and}\ 5.6f$$

which are the first five partials of the stretched spectrum defined by

$$f_n = fA^{\log_2 n}$$

for $A = 2.1$. As shown in Table 8.2, an initial set of notes at C, E, G, C converges to notes with fundamental frequencies that are completely unrelated to "normal" 12-tet intervals based on the semitone $\sqrt[12]{2}$. The convergent values also bear no resemblance to the just intervals. Rather, they converge near notes of the stretched scale defined by the stretched semitone $\beta = \sqrt[12]{2.1}$. Thus, a major chord composed of notes with stretched timbres converges to a stretched major chord. Similarly, the minor chord converges to a stretched minor chord. Sound examples [S: 46] and [S: 47] demonstrate, first in the original 12-tet tuning and then after the adaptation is completed.

8.5.4 Adaptation vs. JI vs. 12-tet

As harmonic tones are related to a scale composed of simple integer ratios, using the adaptive tuning strategy is similar to playing in a Just Intonation (JI) major scale, at least in a diatonic setting. Significant differences occur, however, when the tonal center of the piece changes. Consider a musical fragment that cycles through major chords around the circle of fifths:

$$C\ G\ D\ A\ E\ B\ F\sharp\ C\sharp\ G\sharp\ D\sharp\ A\sharp\ F\ C$$

Table 8.2. Using five partial stretched timbres, the adaptive tuning algorithm converges to stretched major and minor chords. The chords in this table can be heard in sound examples [S: 46] and [S: 47].

Initial notes in 12-tet	Initial frequency of fundamental	Convergent values	Convergent ratios	Nearest stretched step $\beta = \sqrt[12]{2.1}$
C	523	508	1.0	$\beta^0 = 1$
E♭	622	616	1.21	$\beta^3 = 1.20$
G	784	784	1.54	$\beta^7 = 1.54$
C	1046	1067	2.1	$\beta^{12} = 2.1$
C	523	523	1.0	$\beta^0 = 1$
E	659	665	1.27	$\beta^4 = 1.28$
G	784	808	1.54	$\beta^7 = 1.54$
C	1046	1100	2.1	$\beta^{12} = 2.1$

For reference, this is performed in sound example [S: 48] in 12-tet. When played in JI in the key of C major,[9] as in sound example [S: 49], the progression appears very out-of-tune. This occurs because intervals in keys near C are just (or nearly so), whereas intervals in distant keys are not.[10] For instance, major thirds are harmoniously played at intervals of 5:4 in the keys near C, but they are sounded as 32:25 in A and E and as 512:405 in $F\sharp$. Some fifths are impure also; the fifth in the $D\sharp$ chord, for example, is played as 40:27 rather than the desired 3:2. Such inaccuracies are readily discernible to the ear and sound out-of-tune and dissonant. Problems such as this are inevitable for any non-equal fixed tuning [B: 68]. The adaptive tuning, on the other hand, is able to maintain the simple 5:4 and 3:2 ratios throughout the musical fragment because it does not maintain a fixed set of intervals. The circle of fifths is performed again in sound example [S: 50]; all chords are just and consonant.

One might consider switching from JI in C to JI in G to JI in D and so on, using the local musical key to determine which JI scale should be used at a given instant. This results in a performance identical to [S: 50].[11] This cures the immediate problem for this example. Unfortunately, it is not always easy to determine (in general) the proper local key of a piece, nor even to determine if and when a key change has occurred. The adaptation automatically adjusts the tuning to the desired intervals with no *a priori* knowledge of the musical key required. When used with harmonic timbres, it is reasonable to view the adaptive tuning as a way to continuously interpolate between an appropriate family of just intonations.

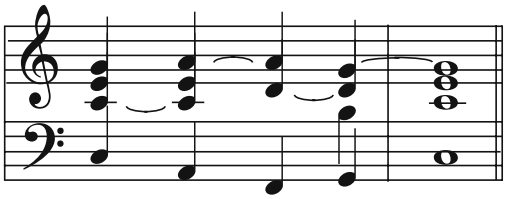
[9] Using the 12 note JI scale from Fig. 4.7 on p. 62.

[10] Such injustices shall not go unpunished!

[11] This is the approach taken by table-driven schemes such as the justonic [W: 14] tuning.

8.5.5 Wandering Tonics

A subtler problem[12] is that variable tunings may drift or wander. For example, Hall [B: 68] points out that if the chord pattern of Fig. 8.7 is played in JI with the tied notes held at constant pitch, then the instrument finishes lower than it begins. Equal temperament prevents this drift in tonal center by forcing the mistuning of many of the intervals away from their just small integer ratios. The adaptive tuning maintains the just ratios, and the tonal center remains fixed. This is possible because the pitches of the notes are allowed to vary microtonally. For instance, the C note in the second chord is played at 528 Hz, and the "same" note in the first chord is played at 523 Hz.



Frequencies when played in JI with held notes:	392.5 327 261.5 131	436 327 -261.5 109	-436 290.5 290.5 87	387.5 -290.5 242 96.5	-387.5 323 258.5 129
Frequencies when played in 12-tet:	392 329.5 261.5 131	440 329.5 261.5 110	440 293.5 293.5 87.5	392 293.5 247 98	392 329.5 261.5 131
Frequencies when played in adaptive tuning:	392.5 327 261.5 131	440 330 264 110	438.5 292 292 87.5	391 294 245 98	392.5 327 261.5 131
Ratios when played in adaptive tuning and in JI:	6/5 5/4 2/1	4/3 5/4 6/5	3/2 1/1 5/3	4/3 6/5 5/4	6/5 5/4 2/1

Fig. 8.7. An example of drift in Just Intonation: the fragment ends about 21 cents lower than it begins. 12-tet maintains the pitch by distorting the simple integer ratios. The adaptive tuning microtonally adjusts the pitches of the notes to maintain simple ratios and to avoid the wandering pitch. Frequency values are rounded to the nearest 0.5 Hz. The three cases are performed in sound examples [S: 51] to [S: 53].

[12] Gary Morrison, in the *Tuning Digest* (9/9/96), argues that wandering tonics can also be viewed as a feature of dynamic tunings that "have a fascinating musical effect."

Three renditions of Fig. 8.7 are played in sound examples [S: 51] to [S: 53]. In [S: 51], the phrase is played six times in just intonation. Because of the tied notes, the tuning drifts down about 21 cents each repeat. As the first and the final chords are identical, each repeat starts where the previous one ends. After five repetitions, it has drifted down about a semitone. The final rendition is played at the original pitch to emphasize the drift. For comparison, [S: 52] plays the same phrase in 12-tet; of course, there is no drift. Similarly, [S: 53] plays the phrase in adaptive tuning. Again there is no drift; yet all chords retain the consonance of simple integer ratios.

One of the major advantages of the 12-tet scale over JI is that it can be transposed to any key. The adaptive tuning strategy shares this advantage, as demonstrated by the circle of fifths example. Both 12-tet and the adaptive tuning can be played starting on any note (in any key). The 12-tet tuning has sacrificed consonance so that (say) all C notes can have the same pitch. As before, the adaptive tuning algorithm modifies the pitch of each note in each chord slightly to increase the consonance. Thus, the C note in the C chord has a (slightly) different frequency from the C note in the F chord, and from the (12-tet enharmonically equivalent) $B\sharp$ note in the $G\sharp$ chord.

When restricted to a single key (or to a family of closely related keys), JI has the advantage that it sounds more consonant than 12-tet (at least for harmonic timbres), because all intervals in 12-tet are mistuned somewhat from the simple integer ratios. The adaptive tuning shares this advantage with JI. Thus, the difference between an adapted piece and the same piece played in 12-tet is roughly the same as the difference between JI and 12-tet, for pieces in a single key when played with harmonic timbres. Whether this increase in consonance is worth the increase in complexity (and effort) is much debated, although the existence of groups such as the *Just Intonation Network* is evidence that some find the differences worthy of exploration.

When focusing on timbres with harmonic spectra, the adaptive spring tuning of Sect. 8.3 and the consonance-based adaptation have much the same effect, although the spring tuning requires more information because it must specify which just interval to assign to each spring. When the timbres are inharmonic, however, neither the spring tuning nor the table-driven models are appropriate.

8.5.6 Adaptation to Inharmonic Spectra

A major advantage of the adaptive tuning approach becomes apparent when the timbres of the instruments are inharmonic, that is, when the partials are not harmonically related. Consider a "bell-like" or "gong-like" instrument with the inharmonic spectrum of Fig. 8.8, which was designed for play in 9-tet using the techniques of Chap. 12. The dissonance curve is significantly different from the harmonic dissonance curve. The most consonant intervals occur at steps of the 9-tet scale (the bottom axis) and are distinct from the simple integer ratios.The 12-tet scale steps (shown in the top axis) do not closely

approximate most of these consonant intervals. Table 8.3 demonstrates the behavior of the adaptive tuning algorithm when used with this 9-tet tone. Pairs of notes are initialized at standard 12-tet; the algorithm compresses or expands them to the nearest minimum of the dissonance curve. In all cases, the converged values are intervals in 9-tet. Similarly, a standard major chord converges to the root, third, and fifth scale steps of the 9-tet scale.

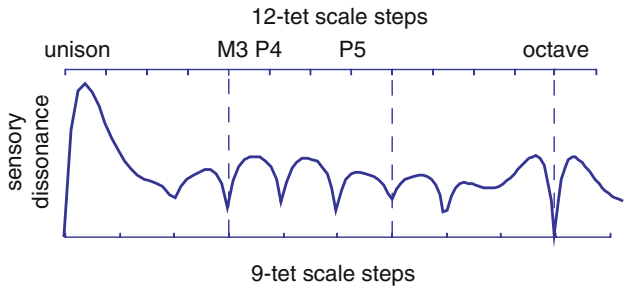


Fig. 8.8. Dissonance curve for an inharmonic timbre with partials at 1, β^9, β^{14}, β^{18}, β^{21}, β^{25}, β^{27}, and β^{30}, where $\beta = \sqrt[9]{2}$. This timbre is appropriate for 9-tet, because minima of the dissonance curve occur at many of the 9-tet scale steps (bottom axis) and not at the steps of the 12-tone scale steps (top axis). Observe that every third step in 9-tet is equal to every fourth step in 12-tet. This follows from the numerical coincidence that $(\sqrt[9]{2})^3 = (\sqrt[12]{2})^4$.

The adaptive tuning strategy can be viewed as a generalization of just intonation in two directions. First, it is independent of the key of the music being played; that is, it automatically adjusts the intonation as the notes of the piece move through various keys. This is done without any specifically "musical" knowledge such as the local key of the music. Second, the adaptive tuning strategy is applicable to inharmonic as well as harmonic sounds, thus broadening the notion of just intonation to include a larger palette of sounds. Recall that a scale and a timbre are said to be related if the timbre generates a dissonance curve with local minima at the scale steps. Using this notion of related scales and timbres, the action of the algorithm can be described succinctly:

> *The adaptive tuning algorithm automatically retunes notes so as to play in intervals drawn from the scale related to the timbre of the notes.*

8.6 The Sound of Adaptive Tunings

This section examines the adaptive tuning algorithm by listening to its behavior. Several simple sound examples demonstrate the kinds of effects possible.

Table 8.3. Using the 9-tet sound of Fig. 8.8, the adaptive tuning algorithm converges to minima of the related dissonance curve. The major chord converges to a chord with 9-tet scale steps 0, 3, and 5.

Initial notes in 12-tet	Initial frequency of fundamental	Convergent values	Convergent ratios	Nearest 9-tet step $\beta = \sqrt[9]{2}$
C	523	528	1.17	$\beta^2 = 1.17$
E♭	622	617		
C	523	528	1.26	$\beta^3 = 1.26$
E	659	659		
C	523	518	1.36	$\beta^4 = 1.36$
F	698	705		
C	523	513	1.47	$\beta^5 = 1.47$
F♯	739	755		
C	523	528	1.47	$\beta^5 = 1.47$
G	783	777		
C	523	523	1.59	$\beta^6 = 1.59$
G♯	830	830		
C	523	519	1.71	$\beta^7 = 1.71$
A	880	888		
C	523	527	1.26	$\beta^3 = 1.26$
E	659	664	1.47	$\beta^5 = 1.47$
G	783	774		

The compositions of Chap. 9 (see especially Table 9.1 on p. 189) demonstrate the artistic potential.

8.6.1 Listening to Adaptation

In sound example [S: 54], the adaptation is slowed so that it is possible to hear the controlled descent of the dissonance curve. Three notes are initialized at the ratios 1, 1.335, and 1.587, which are the 12-tet intervals of a fourth and a minor sixth (for instance, C, F, and $A\flat$). Each note has a spectrum containing four inharmonic partials at f, $1.414f$, $1.7f$, $2f$. Because of the dense clustering of the partials and the particular intervals chosen, the primary perception of this tonal cluster is its roughness and beating. As the adaptation proceeds, the roughness decreases steadily until all of the most prominent beats are removed. The final adapted ratios are 1, 1.414, and 1.703.

This is illustrated in Fig. 8.9, where the vertical grid on the left shows the familiar locations of the 12-tet scale steps. The three notes are represented by the three vertical lines, and the positions of the partials are marked by the small circles. During adaptation, the lowest note descends, and the higher two ascend, eventually settling on a "chord" defined by the intervals g, $1.41g$, and $1.7g$. The arrows pointing left show the locations of four pair of partials that are (nearly) coinciding.

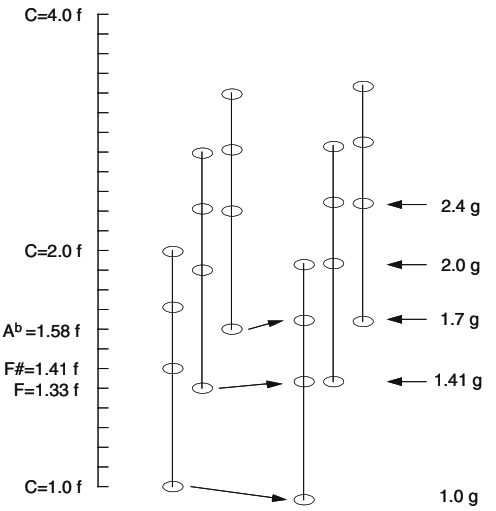


Fig. 8.9. Three notes have fundamentals at C, F, and $A\flat$, and partials at $1.0f$, $1.41f$, $1.7f$, and $2.0f$. After adaptation, the C at frequency f slides down to frequency g, and the other two notes slide up to $1.41g$ and $1.70g$. The arrows on the right emphasize the resulting four pairs of (almost) coinciding partials. Sound example [S: 54] demonstrates.

Sound example [S: 54] performs the adaptation three times at three different speeds. The gradual removal of beats is clearly audible in the slowest. When faster, the adaptation takes on the character of a sliding portamento. There is still some roughness remaining in the sound even when the adaptation is complete, which is due to the inherent sensory dissonance of the sound. The remaining slow beats (about one per second) are due to the resolution of the audio equipment.

There are two time scales involved in the adaptation of a musical passage. First is the rate at which time evolves in the music, the speed at which notes occur. Second is the time in which the adaptation occurs, which is determined by the stepsize parameter. The two times are essentially independent[13]; that is, the relative rates of the times can be chosen by the performer or composer. For instance, the adaptation can be iterated until convergence before the sound starts, as was done in Fig. 8.7 and sound examples [S: 50] and [S: 53]. Alternatively, intermediate values of the adaptive process can be incorporated into the performance, as was done in sound example [S: 54]. The resulting pitch glide can give an interesting elasticity to the tuning, analogous to a guitar bending strings into tune or a brass player lipping the sound to improve the intonation. Adaptation provides a kind of "intelligent" portamento that begins wherever commanded by the performer and slides smoothly to a nearby most- consonant chord. The speed of the slide is directly controllable and may be (virtually) instantaneous or as slow as desired.

[13] The inevitable time lag due to the computation of the algorithm can be made almost imperceptible by using a reasonably fast processor.

8.6.2 Wavering Pitches

When the two time rates are coupled incorrectly, there may be some unusual (and undesirable) effects. Several sound examples demonstrate using the first section of Domenico Scarlatti's harpsichord sonata K1. These are as follows:

(i) [S: 55]: Scarlatti's K1 sonata in 12-tet
(ii) [S: 56]: Scarlatti's K1 sonata with adaptation (incorrect stepsizes)
(iii) [S: 57]: Scarlatti's K1 sonata with adaptation.

The first two measures of the sonata are shown in Fig. 8.10. The first eight notes in all three are identical because only one note is sounding. When two voices occur simultaneously, both are adapted, and the adapted version differs from the 12-tet version. The most obvious change is during the trill at the end of the second measure, although subtler differences can be heard throughout.

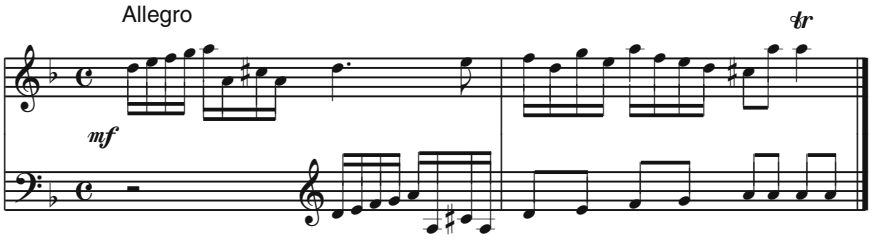


Fig. 8.10. Scarlatti's Sonata K1 is played in 12-tet, and with different speeds of adaptation. The first two measures are shown.

Sound example [S: 58] focuses attention on the second measure by playing all three versions one after the other. As written (and as heard in 12-tet), the trill alternates between A and $B\flat$, and it is accompanied by a slower repeated A an octave below. When adapted (assuming a harmonic spectrum for the harpsichord),[14] the behavior of the algorithm can best be described by reference to a dissonance curve for harmonic sounds (such as in Fig. 6.1 on p. 100). The octaves in the trill are unchanged, because the octave is a minimum of the dissonance curve. The interval between A and $B\flat$ does not fall on a minimum, and the adaptation moves downhill on the dissonance curve, pushing the notes apart to the nearby minimum that occurs at a ratio of 2.25 (which is just a bit more than an octave plus a whole tone). The algorithm essentially "splits the difference" by sharpening the $B\flat$ about 50 cents and simultaneously flattening the A about 50 cents. It is the rapid oscillation between the true A and the flat A that causes the wavering.

[14] The harpsichord is assumed to have nine harmonic partials where the i^{th} partial has amplitude 0.9^i. See Fig. 11.7 on p. 234.

Although the algorithm is moving each pair to the most consonant nearby interval, the overall effect is unlikely to be described as restful consonance. Rather, the rapid wiggling of the lower tone during the trill is probably confusing and disconcerting. This kind of wavering of the pitch can occur whenever rapidly varying tones occur over a bed of sustained sounds. Although this may be useful as a special effect, it is certainly not always desirable. The strangeness of the gliding of the adaptive tuning is especially noticeable when played using an instrumental sound like the harpsichord that cannot bend its pitch.

There are several different ways to fix the wavering pitch problem. The simplest is to adapt the notes with a slower time constant, like the elastic tuning of sound example [S: 54]. By adapting more slowly, the pitches of rapid trills such as in the second measure of the Scarlatti piece do not have time to wander far, thus reducing the waviness. Another solution is to adapt those notes that are already sounding more slowly than newer notes. This is implemented by making the stepsize corresponding to new notes larger than the stepsize corresponding to held notes. A third approach, using the idea of a musical "context" or "memory," is explored in Sect. 9.4.

To investigate this, the same two measures of the Scarlatti K1 sonata are played with new notes adapted ten times as fast as held notes. In sound example [S: 58](c), the wavering of the pitch beneath the trill is almost inaudible. A careful look at the adapted notes shows that the sustained A descends only about 10 cents, and the $B\flat$'s ascend almost 90 cents, again forming an interval of 2.25. Thus, the sustained A only wiggles imperceptibly and the $B\flat$ has risen to (almost) a B.

This example demonstrates that the use of the adaptive tuning can be at odds with a composers intent. Likely, Scarlatti meant for the dissonance of the trill to be part of the effect of the piece (else why write it?). By turning this dissonance into a slightly wavering series of consonances, this intent has been subverted, underscoring the danger of applying a musical transformation in a setting to which it is not appropriate. This example shows the behavior of the adaptive tuning algorithm in a particularly unfriendly setting. When many notes are sounding at once, new notes (such as the trill) become less likely to cause large wavering changes. Thus, the simple two note setting is the most likely place to encounter the wavering pitch phenomenon.

8.6.3 Sliding Pitches

In the adaptive tuning algorithm, whenever a new note occurs, all currently sounding notes are re-adapted. In some situations, like the Scarlatti example, this can cause an undesirable wavering pitch. In other situations, however, the pitches glide gracefully, smoothly connecting one chord to another. In yet other situations, the adaptation may cause new "chords" to form as the pitches change. Sound example [S: 59] contains six short segments:

(i) A single measure in 12-tet

(ii) The "same" measure after adaptation
(iii) The measure (i) followed immediately by (ii)
(iv) Another measure in 12-tet
(v) The "same" measure after adaptation
(vi) The measure (iv) followed immediately by (v)

Both (i) and (ii) start on a F major chord. The adapted version is slightly closer to a justly intoned chord, but this is probably imperceptible. The most obvious change occurs at the second beat. Although the 12-tet version simply continues to arpeggiate, one note of the adapted version slides up. Perhaps because this tone is moving against a relatively fixed background, it jumps out and becomes the "main event" of the passage. When the chord changes to G major at the third beat, an A note remains suspended. In the adapted version, this repels the sliding note, which moves back down to a G note on the third beat.

Thus, the adaptation has actually added something of musical interest. In fact, adaptation will sometimes change the "chord" being played. In parts (iv) and (v) of sound example [S: 59], one measure of a F chord is played in 12-tet, followed by its adapted version. Although the basic harmony remains fixed in the original 12-tet, the chord changes in the adapted version on the fourth beat. The change appears to be to a nearby, closely related chord, although in reality it is to a nearby microtonal variant of the original.

Sound example [S: 60], *Three Ears*, contains all the measures from sound example [S: 59]. Many other similar passages occur—the algorithm causes interesting glides and unusual microtonal adjustments of the notes, all within an "easy-listening" setting. The microtonal movement is done in a perceptually sensible fashion. In the Scarlatti examples [S: 58], the sliding pitches were a liability. In sound examples [S: 59] and in the *Three Ears*, they are exploited as a new kind of "intelligent" musical effect.

8.7 Summary

The adaptive tuning strategy provides a new solution to the long-standing problem of scale formation. Just intonations (and related scales) sacrifice the ability to modulate music through multiple keys, and 12-tet sacrifices the consonance of intervals. Adaptive tunings retain both consonance and the ability to modulate, at the expense of (real-time) microtonal adjustments in the pitch of the notes. The spring tuning provides a simple physical model of the stresses of mistunings, and the consonance-based adaptive tuning encodes a basic human perception, the sensory dissonance curves.

Adaptive tuning algorithms are implementable in software or hardware and can be readily incorporated into electronic music studios. Just as many MIDI synthesizers have built-in alternate tunings tables that allow the musician to play in various just intonations and temperaments, an adaptive tuning feature

could be readily added to sound modules. The musician can then effortlessly play in a scale that continuously adjusts to the timbre and the performance in such a way as to maximize sensory consonance. One concrete realization appears in Chap. 9.

The behavior of the adaptive tuning algorithm can be described in terms of notes continuously descending a complex multidimensional landscape studded with dissonant mountains and consonant valleys. These behaviors are described mathematically in Appendix H. For harmonic timbres, the adaptive tuning acts like a just intonation that automatically adjusts to the key of the piece, with no specifically musical knowledge required. For harmonic timbres, the action of the spring tuning and the consonance-based adaptations are similar. For inharmonic timbres, the adaptive tuning automatically adjusts the frequencies of the tones to a nearby minimum of the dissonance curve, providing an automated way to play in the scale related to the spectrum of the sound. Adaptive tunings are determined by the spectra of the sounds and by the piece of music performed; chords and melodies tend to become more "in tune with themselves."

9

A Wing, An Anomaly, A Recollection

The adaptive tuning of the last chapter adjusts the pitches of notes in a musical performance to minimize the sensory dissonance of the currently sounding notes. This chapter presents a real-time implementation called `Adaptun` *(written in the* `Max` *programming language and available on the CD in the* `software` *folder) that can be readily tailored to the timbre (or spectrum) of the sound. Several tricks for sculpting the sound of the adaptive process are discussed. Wandering pitches can be tamed with an appropriate* context*, a (inaudible) collection of partials that are used in the calculation of dissonance within the algorithm, but that are not themselves adapted or sounded. The overall feel of the tuning is effected by whether the adaptation converges fully before sounding (or whether intermediate pitch bends are allowed). Whether adaptation occurs when currently sounding notes cease (or only when new notes enter) can also have an impact on the overall solidity of the piece. Several compositional techniques are explored in detail, and a collection of sound examples and musical compositions highlight both the advantages and weaknesses of the method.*

9.1 Practical Adaptive Tunings

To bring the techniques of adaptive tunings into sharper focus, this chapter looks at several examples of the use of adaptation in tuning. In some (such as *Local Anomaly* [S: 79]), all notes adapt continuously and simultaneously. In others (such as *Wing Donevier* [S: 85]), all notes are adapted completely before they are sounded. *Recalled Opus* [S: 82] presents an adaptation of a (synthesized) string quartet in which a "context" is used to help tame excess horizontal (melodic) motion. Several compositions (which are listed in Table 9.1) are discussed at length, and steps are detailed to highlight the practical issues, techniques, and tradeoffs that develop when applying adaptive tunings.

The next section discusses the `Adaptun` software, and Sect. 9.3 details some of the simplifications to the basic algorithm of Chap. 8 that are used to make the program operate efficiently in real time. The use of a context is discussed in Sect. 9.4 as a way of imposing a kind of consistency on the adaptation

to reduce some of the melodic artifacts. The bulk of the chapter provides an extensive series of examples. Many of these are short snippets exploring some feature of the adaptive process, and many are complete compositions. The final section poses some of the aesthetic questions that arise in the use of adaptation in musical contexts.

9.2 A Real-Time Implementation in `Max`

Figure 9.1 shows the main screen of the adaptive tuning program `Adaptun`, which was first presented in [B: 171]. The user must first configure the program to access the MIDI hardware. This is done using the two menus labeled `Set Input Port` and `Set Output Port`, which list all valid MIDI sources and destinations. The figure shows the input `US-428 Port 1`, which is my hardware, and the output is set to ∞ `IAC Bus # 2`, which is an interapplication (virtual) port that allows MIDI data to be transferred between applications. The interapplication ports allow `Adaptun` to exchange data in real time with sequencers, software synthesizers, or other programs. In particular, the output of `Adaptun` can be recorded by setting the input of a MIDI sequencer to receive on the appropriate IAC bus.

In normal operation, the user plays a MIDI keyboard. The program rechannelizes and retunes the performance. Each currently sounding note is assigned a unique MIDI channel, and the adapted note and appropriate pitch bend commands are output on that channel. As the algorithm iterates, updated pitch bend commands continue to fine tune the pitches. The MIDI sound module must be set to receive on the appropriate MIDI channels with "pitch bend amount" set so that the extremes of ± 64 correspond to the setting chosen in the box labeled `PB value in synth`. The finest pitch resolution possible is about 1.56 cents when this is set to 1 semitone, 3.12 cents when set to 2 semitones, and so on.

There are several displays that demonstrate the activity of the program. First, the message box directly under the block labeled `Adapt` shows the normalized sensory dissonance of the currently sounding notes. The bar graph on the left displays the sensory dissonance as a percentage of the original sensory dissonance of the current notes. A large value means that the pitches did not change much, and a small value means that the pitches were moved far enough to cause a significant decrease in sensory dissonance. The large display in the center shows how many notes are currently adapting (how many pieces the line is broken into) and whether these notes have adapted up in pitch (the segment moves to the right) or down in pitch (the segment moves to the left). The screen snapshot in Fig. 9.1 shows the adaptation of three notes; two have moved down and one up. There is a wraparound in effect on this display; when a note is retuned more than a semitone, it returns to its nominal position. The number of actively adapting tones is also displayed numerically in the topmost message box.

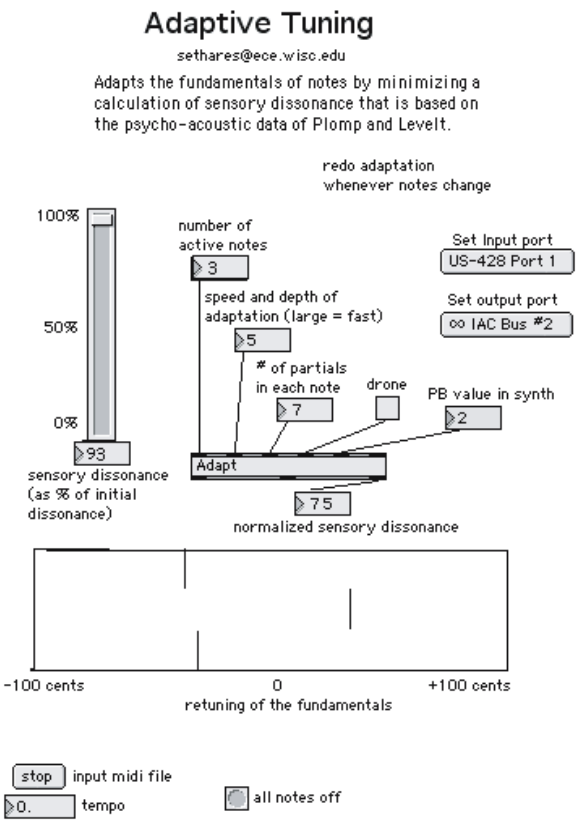


Fig. 9.1. Main screen of the adaptive tuning program `Adaptun`, implemented in the `Max` programming language.

The user has several options that can be changed by clicking on message boxes.[1] One is labeled `speed and depth of adaptation` in Fig. 9.1. This represents the stepsize parameter μ from (8.2) and (8.3) on p. 163. When small, the adaptation proceeds slowly and smoothly over the dissonance surface. Larger values allow more rapid adaptation, but the motion is less smooth. In extreme cases, the algorithm may jump over the nearest local minimum and descend into a minimum far from the initial values of the intervals. The relationship between the speed of adaptation and "real time" is complex, and it depends on the speed of the processor and the number of other tasks occurring simultaneously. The message box labeled `# of partials in each note`

[1] When a `Max` message box is selected, its value can be changed by dragging the cursor or by typing in a new value. Changes are output at the bottom of the box and incorporated into subsequent processing.

specifies the maximum number of partials that are used. (The actual values for the partials are discussed in detail in Sect. 9.3.)

There are two useful tools at the bottom of the main screen. The menu labeled `input MIDI file` lets the user replace (or augment) the keyboard input with data from a standard MIDI file. The menu has options to `stop`, `start`, and `read`. First, a file is `read`. When started, adaptation occurs just as if the input were arriving from the keyboard. The message box immediately below the menu specifies the tempo at which the sequence will be played. This is especially useful for older (slower) machines. A standard MIDI file (SMF) can be played (and adapted) at a slow tempo and then replayed at normal speed, increasing the apparent speed of the adaptation. Finally, the `all notes off` button sends "note-off" messages on all channels, in the unlikely event that a note gets stuck.

9.3 The Simplified Algorithm

In order to operate in real time (actual performance depends on processor speed), several simplifications are made. These involve the specification of the spectra of the input sounds, using only a special case of the dissonance calculation, and a simplification of the adaptive update.

The dissonance measure[2] in (8.1) on p. 163 is dependent on the spectra of the currently sounding notes, and so the algorithm (8.3) must have access to these spectra. Although it should eventually be possible to measure the spectra from an audio source in real time, the current MIDI implementation assumes that the spectra are known *a priori*. The spectra are defined in a table, one for each MIDI channel, and they are assumed fixed throughout the piece (or until the table is changed). They are stored in the collection[3] file `timbre.col`. The default spectra are harmonic with a number of partials set by the user in the message box on the main screen, although this can easily be changed by editing `timbre.col`. The format of the data reflects the format used throughout `Adaptun`; all pitches are defined by an integer

$$100 * (\text{MIDI Note Number}) + (\text{Number of Cents}). \tag{9.1}$$

For instance, a note with fundamental 15 cents above middle C would be represented as $6015 = 100 * 60 + 15$ because 60 is the MIDI note number for middle C. Similarly, all intervals are represented internally in cents: an octave is thus 1200 and a just major third is 386.

Second, the calculation of the dissonance is simplified from (8.1) by using a single "look-up" table to implement the underlying dissonance curves.[4] A

[2] This is further detailed in (H.2).

[3] In `Max`, a "collection" is a text file that stores numbers, symbols, and lists.

[4] This look-up table simplifies the implementation of (8.1) and (E.2) because no transcendental functions need be calculated.

nominal value of 500 Hz is used for all calculations between all partials, rather than directly evaluating the exponentials. In most cases, this will have little effect, although it does mean that the magnitude of the dissonances will be underestimated in the low registers and overestimated in the high. More importantly, the loudness parameters a_1 and a_2 are set to unity. Combined with the assumption of fixed spectra, this can be interpreted as implying that the algorithm operates on a highly idealized, averaged version of the spectrum of the sound.

The numerical complexity of the iteration (8.3) is dominated by the calculation of the gradient term, due to its complexity (which grows worse in high dimensions when there are many notes sounding simultaneously). One simplification uses an approximation to bypass the explicit calculation of the gradient. `Adaptun` adopts a variation of the simultaneous perturbation stochastic approximation (SPSA) method of [B: 180].[5] To be concrete, the function

$$g(f_i(k)) = \frac{D(f_i(k) + c\Delta(k)) - D(f_i(k) - c\Delta(k))}{2c\Delta(k)}$$

where $\Delta(k)$ is a randomly chosen Bernoulli ± 1 random vector, can be viewed as an approximation to the gradient $\frac{dD}{df_i(k)}$. This approximation grows closer as c approaches zero. The algorithm for adaptive tuning is then

$$f_i(k+1) = f_i(k) - \mu g(f_i(k)). \tag{9.2}$$

In the standard SPSA, convergence to the optimal value can be guaranteed if both the stepsize μ and the perturbation size c converge to zero at appropriate rates, and if the cost function D is sufficiently smooth [B: 179]. In the case of adaptive tunings, it is important that the stepsize and perturbation size *not* vanish, because this would imply that the algorithm becomes insensitive to new notes as they occur.

In the adaptive tuning application, there is a granularity to pitch space induced by the MIDI pitch bend resolution of about 1.56 cents. This is near to the resolving power of the ear (on the order of 1 cent), and so it is reasonable to choose μ and c so that the updates to the f_i are (on average) roughly this size. This is the strategy followed by `Adaptun`, although the user-chooseable parameter labeled `speed and depth of adaptation` gives some control over the size of the adaptive steps. Convergence to a fixed value is unlikely when the stepsizes do not decay to zero. Rather, some kind of convergence in distribution should be expected, although a thorough analysis of the theoretical implications of the fixed-stepsize version of SPSA remain unexplored. Nonetheless, the audible results of the algorithm are vividly portrayed in Sect. 9.5.

[5] This can also be viewed as a variant of the classic Kiefer–Wolfowitz algorithm [B: 84].

9.4 Context, Persistence, and Memory

Introspection suggests that people readily develop a notion of "context" when listening to music and that it is easy to tell when the context is violated, for instance, when a piece changes key or an out-of-tune note is performed. Although the exact nature of this context is a matter of speculation, it is clearly related to the memory of recent sounds. It is not unreasonable to suppose that the human auditory system might retain a memory of recent sound events, and that these memories might contribute to and color present perceptions. There are examples throughout the psychological literature of experiments in which subjects' perceptions are modified by their expectations, and we hypothesize that an analogous mechanism may be partly responsible for the context sensitivity of musical dissonance.

Three different ways of incorporating the idea of a musical context into the sensory dissonance calculation are suggested in [B: 173], in the hopes of being able to model some of the more obvious effects.

(i) The *exponential window* uses a one-sided window to emphasize recent partials and to gradually attenuate the influence of older sounds.
(ii) The *persistence model* directly preserves the most prominent recent partials and discounts their contribution to dissonance in proportion to the elapsed time.
(iii) The *context model* supposes that there is a set of privileged partials that persist over time to enter the dissonance calculations.

All three models augment the sensory dissonance calculation to include partials not currently sounding; these extra partials originate from the windowing, the persistence, or the context. A series of detailed examples in [B: 173] shows how each model explains some aspects but fails to explain others. The context model is the most successful, although the problem of how the auditory system might create the context in the first place remains unresolved.

To see how this might work, consider a simple context that consists of a set of partials at 220, 330, 440, and 660 Hz. When a harmonic note A or E is played at a fundamental of 220 or 330 Hz, many of their partials coincide with those of the context, and the dissonance calculation (which now includes the partials in the context as well as those in the currently sounding notes) is barely larger than the intrinsic dissonance of the A or E. When, however, a $G\sharp$ note is sounded (with fundamental at about 233 Hz), the partials of the note will interact with the partials of the context to produce a significant dissonance.

The context idea is implemented in `Adaptun` using a static "drone." The check box labeled `drone` enables a fixed context that is defined in the collection file `drone.col`. The format of the data is the same as in (9.1) above. For example, the drone file for the four-partial context of the previous paragraph is:

```
1, 4500;
2, 5202;
3, 5700;
4, 6402;
```

(The "02" occurs because the perfect fifth between 330 Hz and 220 Hz corresponds to 702 cents, not 700 cents as in the tempered scale.) When the `drone` switch is enabled, notes that are played on the keyboard (or notes that are played from the `input MIDI file` menu) are adapted with a cost function that includes both the currently sounding notes and the partials specified in the drone file. The drone is inaudible, but it provides a framework around which the adaptation occurs. Examples are provided in the next section.

9.5 Examples

This section provides several examples that demonstrate the adaptive tuning algorithm and explores the kinds of effects possible with the various options in `Adaptun`. Discussions of the compositional process and demonstrations of the artistic potential of the adaptive tunings are deferred until Sect. 9.6.

9.5.1 Randomized Adaptation

The motion of the adapting partials in sound example [S: 54] was shown pictorially in Fig. 8.9 on p. 174. When using `Adaptun` to carry out the adaptation (rather than (8.3), the true gradient algorithm), the final converged value of g may differ from run to run. This is because the iteration is no longer completely deterministic; the probe directions $\Delta(k)$ in (9.2) are random, and the algorithm will follow (slightly) different trajectories each time. The bottom of the dissonance landscape is always defined by the ratio of the fundamentals of the notes (in this case, g, $1.41g$, and $1.7g$) but the exact value of g may vary.

In most cases, the convergent values of the `Adaptun` algorithm will be the same as the converged values of the deterministic version. An exception occurs when the initial intervals happen to be maximally dissonant, that is, when they lie near a peak of the dissonance surface. The deterministic version will always descend into the same consonant valley, but the probe directions of `Adaptun`'s SPSA method may cause it to descend in either direction. This can be exploited as an interesting effect, as in the second adaptive study [S: 62] or the *Recalled Opus* [S: 82] where pairs of notes are repeatedly initialized near a dissonant peak and allowed to slide down: sometimes contracting to a unison and sometimes expanding to a minor third.

9.5.2 Adaptive Study No. 1

Sound example [S: 61] is orchestrated for four synthesized "wind" voices. When several notes are sounded simultaneously, their pitches are often changed

significantly by the adaptation. This is emphasized by the motif, which begins with a lone voice. When the second voice enters, both adapt, giving rise to pitch glides and sweeps. As the timbres have a harmonic structure, most of the resulting intervals are actually justly intoned because the notes adapt to align a partial of the lower note with some partial of the upper. By focusing attention on the pitch glides (which begin at 12-tet scale steps), this demonstrates clearly how distant many of the common 12-tet intervals are from their just counterparts.

Perhaps the most disconcerting aspect of the study is the way the pitches wander. As long as the adaptation is applied only to currently sounding notes, successive notes may differ: The C note in one chord may be retuned from the C note in the next. This can produce an unpleasant "wavy" or "slimy" sound. This effect is easy to hear in the long notes that are held while several others enter and leave. For instance, between 0:36 and 0:44 seconds (and again at 1:31 to 1:39), there is a three-note chord played. The three notes adapt to the most consonant nearby location. Then the top two notes change while the bottom is held; again all three adapt to their most consonant intervals. This happens repeatedly. Each time the top two notes change, the held note readapts, and its pitch slowly and noticeably wanders. Although the vertical sonority is maintained, the horizontal retunings are distracting.

The most straightforward way to forbid this kind of behavior is to leave currently sounding notes fixed as newly entering notes adapt their pitches. This can be implemented by setting the stepsize μ to zero for those fundamentals corresponding to held notes. Unfortunately, this does not address the fundamental problem; it only addresses the symptom that can be heard clearly in this sound example. A better way is by the introduction of the inaudible "drone," or context.

9.5.3 A Melody in Context

`Adaptun` implements a primitive notion of memory or context in its drone function. A collection of fixed frequencies are prespecified in the file `drone.col`, and these frequencies enter into the dissonance calculation although they are not sounded.

The simplest case is when the spectrum of the adapting sound consists of a single sine wave as in parts (a) and (b) of Fig. 9.2. The unheard context is represented by the dashed horizontal lines. Initially, the frequency of the note is different from any of the frequencies of the context. If the initial note is close to one of the frequencies of the context, then dissonance is decreased by moving them closer together. The note converges to the nearest frequency of the context, as shown by the arrow. In (b), the initial note is distant from any of the frequencies of the context. When both distances are larger than the point of maximum dissonance (the peaks of the curves in Fig. 3.8 on p. 47), then dissonance is decreased by moving further away. Thus, the pitch

is pushed away from both of the nearby frequencies of the context, and it converges to some intermediate position.

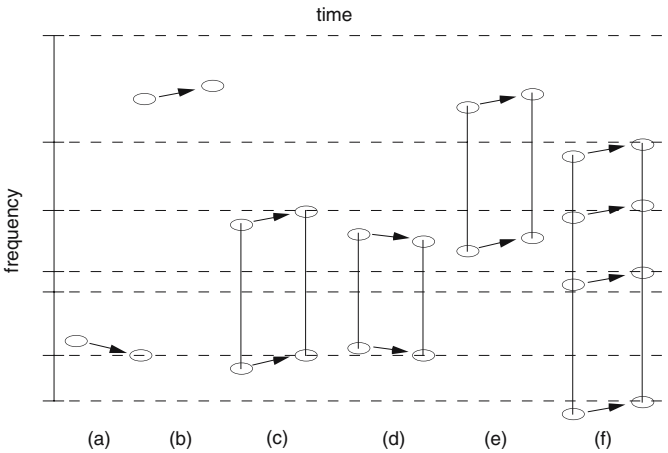


Fig. 9.2. The dashed horizontal grid defines a fixed "context" against which the notes adapt. When the note has a spectrum consisting of a single sine wave partial as in (a) and (b), then the note will typically adjust its pitch until it coincides with the nearest partial of the context as in (a), or else it will be repelled from the nearby partials of the context as in (b). When the spectrum has two partials, then the adaptation may align both partials as in (c), one as in (d), or none as in (e). In (f), the partials fight to align themselves with the context, eventually converging to minimize the beating.

Generally the timbre will be more complex than a single sine wave. Figure 9.2 shows several other cases. In parts (c), (d), and (e), the timbre consists of two sine wave partials. Depending on the initial pitch (and the details of the context), this may converge so that both partials overlap the context as in (c), so that one partial merges with a frequency of the context and the other does not as in (d), or to some intermediate position where neither partial coincides with the context, as in (e). Part (f) gives the flavor of the general case when the timbre is complex with many sine wave partials and the context is dense. Typically, some partials converge to nearby frequencies in the context and some will not.

To see how this might function in a more realistic setting, suppose that the current context consists of the note C and its first 16 harmonics. When a new harmonic note occurs, it is adapted not only in relationship to other currently sounding notes, but also with respect to the partials of the C. Because partials of the adapting notes often converge to coincide with partials in the context (as in part (f) of Fig. 9.2), there is a good chance that a partial of the note will align with a partial of the context. When this occurs, the adapted interval

will be just, formed from the small integer ratio defined by the harmonic of the note with that of the context.

Thus, the context provides a structure that influences the adaptation of all the sounding notes, like an unheard drone. In this way, it can give a horizontal consistency to the adaptation that is lacking when no memory is allowed.

9.5.4 Adaptive Study No. 2

The next example, presented in [S: 62], is orchestrated for four synthesized "violin" voices. Like the first study, the adaptive process is clearly audible in the sweeping and gliding of the pitches. For this performance, however, a context consisting of all octaves of C plus all octaves of G was used.[6]

The context encourages consistency in the pitches, maintaining (an unheard) template to which the currently sounding notes adapt. Although the study still contains significant pitch adaptation, the final resting places are constrained so that the adjusted pitches are related to the unheard C or G. Typically, some harmonic of each adapted note aligns with one of the octaves of the C or G template.

In several places throughout the piece, adjacent notes (of the 12-tet scale) are played simultaneously. For the specified timbres, this is near the peak of the dissonance curve. Depending on exactly which notes are played, the order in which they are played, and the vagaries of the random test directions $\Delta(k)$ of (9.2), sometimes the two pitches adapt to an interval at about 316 cents (a just minor third) by moving apart in pitch, and sometimes they merge into a unison at some intermediate pitch. In either case, the primary sensation is of the motion.

9.5.5 A Recollection

Many of the kinds of pitch slides and glides that are so obvious in the two adaptive studies are exploited in a more structured way in *Recalled Opus* [S: 82]. `Adaptun` was used to play four string voices (a synthesized "string quartet"). Each tone begins on a 12-tet pitch and adapts the pitches in real time. The action of the algorithm is unmistakable.

Because the string timbres are harmonic, the retuning converges primarily to various just intervals. When the pitches begin close to JI, such as a 12-tet fifth, the adjustment is only a few cents. But when the pitches begin far away from JI (such as a 12-tet minor second), the pitch sweeps are dramatic. All of the pitch bending is done by the algorithm in real time.[7] This piece provides a nonverbal and visceral demonstration of the differences between JI and 12-tet.

[6] The drone file contained all C's $2400, 3600, 4800, 6000, \ldots$ plus all G's $3100, 4300, 5500, 6700, \ldots$.

[7] The piece was not performed in one pass, several individual sections were recorded separately and then spliced together.

9.6 Compositional Techniques and Adaptation

Adaptive tunings are not constrained to any particular style of music, and the previous sound examples suggest that a number of interesting and unusual effects are possible. One avenue of exploration is perhaps obvious: Play with `Adaptun`, and allow happy accidents to occur. The adaptive studies and *Recalled Opus* [S: 82] were derived from spontaneous improvisations that crystallized into repeatable forms. *Persistence of Time* [S: 81] began with a three-against-two rhythmic bed, and repeated improvisation led to the final piece.

Table 9.1 lists the adaptively tuned pieces that appear on the CD along with three compositional parameters. The third column indicates whether a context was used during adaptation, as discussed in the previous section using the `drone` option in `Adaptun`. The fourth column specifies whether the algorithm was allowed to achieve full convergence before the notes are sounded (indicated by y) or whether all intermediate pitches were output (n). This can have a major impact on the sound and effect of the piece. For example, *Persistence of Time* does not have the kind of slimy undulating pitches that are so conspicuous in *Recalled Opus*. The column labeled "Adapt on Note-off" specifies whether the adaptation is redone when notes end (that is, each time the number of currently sounding notes changes) or whether adaptation occurs only when new notes begin. This is one of the reasons *Wing Donevier* sounds more steady than *Excitalking Very Much*.

Table 9.1. Several musical compositions appearing on the CD-ROM use adaptive tunings. Also indicated are whether a context was used, whether the algorithm was allowed to output intermediate pitches as it adapted (or only after convergence), and whether the adaptation was conducted at note-off events as well as note-on events.

Name of Piece	File	Context	Converge Fully	Adapt on Note-off	See
Adventiles in a Distorium	`adventiles.mp3`	y	n	y	[S: 74]
Aerophonious Intent	`aerophonious.mp3`	y	n	n	[S: 75]
Story of Earlight	`earlight.mp3`	n	n	n	[S: 76]
Excitalking Very Much	`excitalking.mp3`	y	y	n	[S: 77]
Inspective Liquency	`inspective.mp3`	n	n	y	[S: 78]
Local Anomaly	`localanomaly.mp3`	n	n	y	[S: 79]
Maximum Dissonance	`maxdiss.mp3`	n	y	n	[S: 80]
Persistence of Time	`persistence.mp3`	n	y	n	[S: 81]
Recalled Opus	`recalledopus.mp3`	y	n	y	[S: 82]
Saint Vitus Dance	`saintvitus.mp3`	n	n	y	[S: 83]
Simpossible Taker	`simpossible.mp3`	y	y	y	[S: 84]
Three Ears	`three_ears.mp3`	n	y	y	[S: 60]
Wing Donevier	`wing.mp3`	y	y	n	[S: 85]

With the exception of *Recalled Opus*, all of the pieces in Table 9.1 were created using a method of *adaptive randomization*, a compositional technique that is particularly appropriate for adaptive tunings. The adaptive randomization begins with a simple rhythmic pattern, adds complexity, orchestration, and timbral variety without regard for harmonic or melodic content, and then tames the dissonances by selective application of the adaptive tuning algorithm. The first step is to select an arbitrary pattern of notes triggering a set of synthesized sounds. As the pitches are essentially random, the sequence is wildly and uniformly dissonant. Application of the adaptive tuning algorithm perturbs the pitches of all currently sounding notes at each time instant to the nearest intervals that maximize consonance. Sometimes the dissonances are tamed and interesting phrases occur. By winnowing the results, separating desirable and undesirable elements, reorchestrating, and using the cut-and-paste operations available in modern audio editing software, strange and unusual pieces can be constructed.

There are many possible sources for musical patterns. They might be constructed mathematically (like the three-against-two pattern of *Persistence of Time*), they might be a complete piece (*Three Ears* was first composed in 12-tet and the adaptation imposed at a later stage), or they might be only a rhythm part (*Wing Donevier* began as a standard MIDI drum part[8] played in an aggressive seven beats per measure). The classical MIDI archive at [W: 4] contains thousands of MIDI files free for downloading, and there are many other sources on the web of both commercial and public domain libraries of MIDI files.

In order to demonstrate the technique concretely, Fig. 9.3 shows the first four measures of a standard MIDI drum track.[9] The information is displayed in a kind of "piano-roll" notation[10] in which the vertical axis represents MIDI note-number. Time proceeds along the horizontal axis. MIDI note events are shown in bold black. For drum tracks, there is a standard assignment of note numbers to instruments,[11] and the relevant ones (bass drum, snare, and three cymbals) are labeled on the left-hand side of the figure. This is performed as written in sound example [S: 63].

One of the interesting features of the MIDI standard is that note events are not necessarily tied to their default instrumentation. Sound example [S: 64], for instance, reorchestrates the four measures in Fig. 9.3 by assigning the lowest two notes to bass guitar (instead of bass drum and snare) and the upper notes to guitar (instead of cymbals) as indicated by the reassignment on the right-hand side. Even more useful than the reorchestration are the editing capabilities offered by modern software. Notes (and other MIDI events) can be

[8] From the Keyfax [W: 17] collection of drum tracks performed by Bill Bruford.
[9] Sequenced by Keyfax Software [W: 17] in the *Breakbeats* collection.
[10] Figures 9.3 through 9.6 show screen snapshots from *Digital Performer*, a commercial audio and MIDI sequencer [W: 20].
[11] Details of the MIDI file specification can be found at [W: 25].

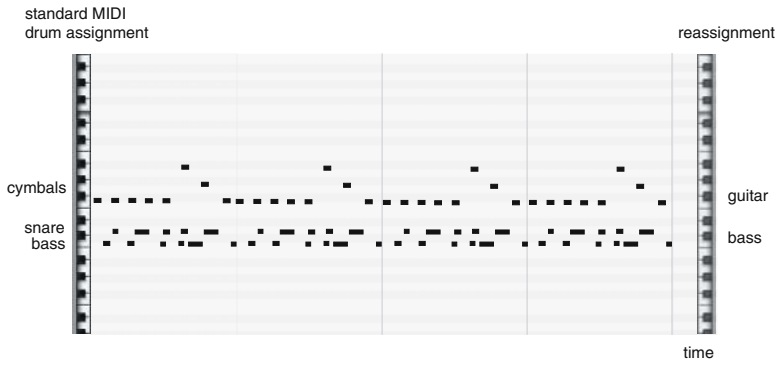


Fig. 9.3. A standard MIDI drum file can be played as a percussion part (sound example [S: 63] performs this sequence with the standard instruments indicated on the left), or it can be reorchestrated (sound example [S: 64] reassigns the notes to guitar and bass as indicated on the right).

rearranged in many ways using simple cut-and-paste techniques. Figure 9.4, for example, shows the same four measures as Fig. 9.3, with the upper notes (that were originally devoted to the cymbals) repeated, offset in pitch, and time-stretched by factors of two (one slower and one faster). As before, this can be performed on any desired set of instruments. Sound examples [S: 65] through [S: 67] demonstrate three simple variations.

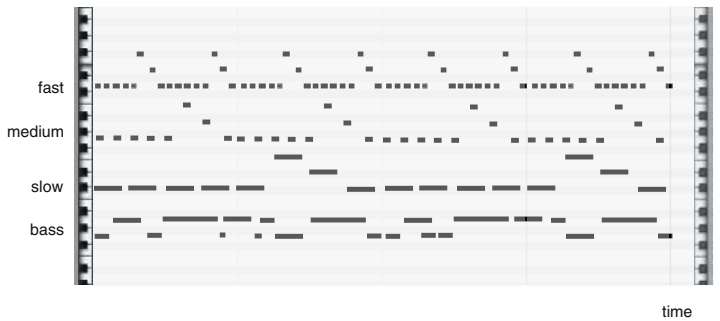


Fig. 9.4. The standard MIDI file in Fig. 9.3 is edited, creating more complex and interesting patterns. Sound examples [S: 65] through [S: 67] demonstrate.

When the instrumentation is finalized (in this case, using harmonic samples of guitar and bass), then the piece can be adapted. This is demonstrated in [S: 68] using the default settings in `Adaptun`. Compare this sparse example with the fully orchestrated *Simpossible Taker* [S: 84], which applied this same

method to a set of MIDI "hip-hop" drum patterns.[12] In order to tame some of the pitch sweeps, a context was used and all notes were allowed to converge fully. The remaining pitch glides are due to the adaptation of held notes. As all sounding notes readjust whenever a note enters or leaves, the held notes slide to their new "most consonant" pitch. This effect is already familiar from *Three Ears* [S: 60].

There are many other ways that MIDI data can be transformed to create interesting sequences. Figure 9.5 shows the data of Fig. 9.4 edited in several ways. The bass guitar part is randomized over an octave, creating a new bass line with considerable motion. Using the instrumentation of [S: 65], this can be heard in sound example [S: 69]. The "fast" line is also randomized and transposed, resulting in a rapid arpeggiation. This is performed in [S: 70] using the same guitar samples as in [S: 65]. Finally, the "slow" line of Fig. 9.4 is transposed up and randomized, creating a constrained random melody. Orchestrating the melody with a synthetic-sounding flute results in sound example [S: 71].

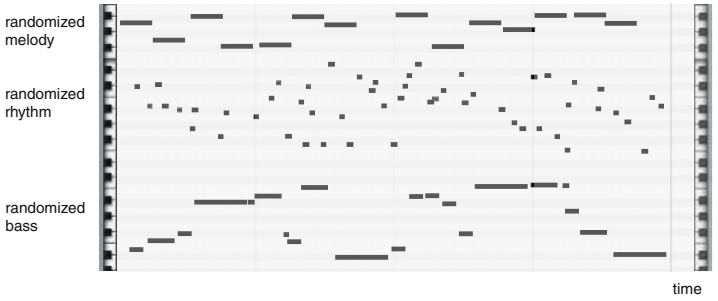


Fig. 9.5. The standard MIDI file in Fig. 9.4 is edited, creating more complex and interesting (randomized) patterns. Sound examples [S: 69] and [S: 70] demonstrate.

Although these are interesting in their own way, they can be combined with the adaptive process to create a large assortment of unusual effects. For example, sound example [S: 72] is an adapted version of [S: 71]. The sound is more aligned, almost lighter in the adapted version, although the pitch glides in the guitar may be disconcerting. Sound example [S: 73] repeats the same piece but using two methods to reduce the amount of pitch uncertainty: first by allowing the convergence to complete before outputting the notes, and then by disallowing adaptation when notes cease to sound. This technique is a template of many of the compositions in Table 9.1.

[12] Commercially available from [W: 17].

9.6.1 A Wing

Wing Donevier [S: 85] is named after a fictional captain who fell at the siege of Eriastur (itself a fictional medieval town). In 7/4 time, this piece began as a standard MIDI drum file from Keyfax Software [W: 17] in their Bill Bruford collection. The original is orchestrated solely for percussion and hence is nontonal, that is, in no particular key. It is recorded as a MIDI file, and so it is easy to assign different voices. A context consisting of all octaves of low C (65.4 Hz) and all octaves of low G (98 Hz) was used. The adaptive process moves the pitches of all notes so as to maximize the instantaneous sensory consonance between the currently sounding notes and the immutable context.

The result is still atonal, but not overly dissonant. Each vertical slice of time is fairly consonant, although melodically (horizontally) there are many small adjustments. After adaptation, the MIDI file was reorchestrated with bass, synth, and drums. The adaptation is allowed to converge completely before each note is sounded, and no adaptation is done when note-off events occur. Together, these choices remove most of the wavering pitches.

The screen snapshot in Figure 9.6 shows the sequence window of a combined audio/MIDI editor.[13] The numbers in the upper right represent measures. The small icons just below represent miniaturized versions of the MIDI tracks familiar from Figs. 9.3 though 9.5 that contain MIDI performance data. These are labeled by their instrumentation (bass, rhy1, rhy2, mel, etc.) and are sent to the IAC (interapplication MIDI) # 1 bus and hence to `Adaptun`. The return path uses IAC # 2, and this is record enabled so that the adapted data can be recorded for further editing and compositing. The adapted data are also output to "Unity," a software synthesizer.[14] Finally, the audio output of the synthesizer is sent to the digital to analog converters, which, in this case, is a Tascam US-428.

9.6.2 An Anomaly

Local Anomaly [S: 79] is another piece in which all notes were retuned adaptively beginning with a randomized MIDI drum file. The major timbres are again guitar-like (and hence primarily harmonic), but the use of the adaptation is quite different from both *Wing Donevier* and the string quartet *Recalled Opus*. Besides the obvious rhythmic intensity of the piece, the notes come rapidly. Rarely is a note held much longer than the time it takes it to converge to the nearby "most consonant" interval. As no context is used (and none of the 'cures' for wavering pitches are invoked), the pitch of each note is in constant motion.

Thus, one of the most prominent features of *Local Anomaly* is the pitch slides, which give an "elasticity" to the tuning analogous to a guitar bending

[13] The program is *Digital Performer* by Mark of the Unicorn [W: 20].
[14] Created by Bitheadz software [W: 2].

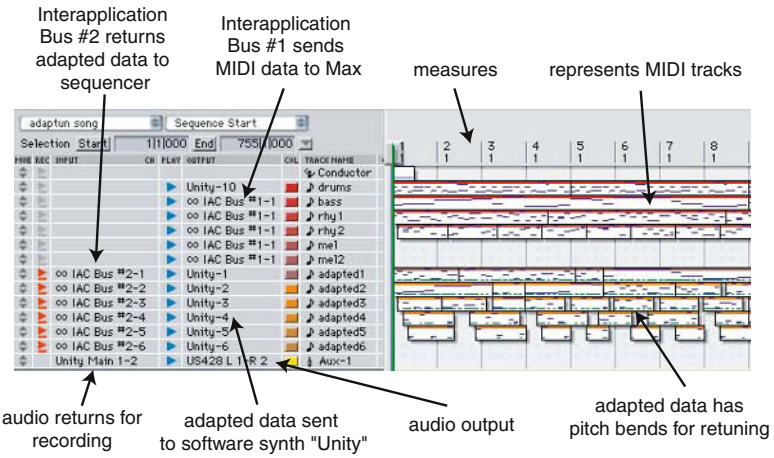


Fig. 9.6. This screen snapshot shows how MIDI information can be sent from the sequencer to Max (which is running `Adaptun`) and then returned to the sequencer for recording. The adapted MIDI data are then output to a software synthesizer so that the results can be heard.

strings into (or out of) tune. All glides in *Local Anomaly* are created by the adaptive process, which provides a kind of "intelligent" portamento that begins where commanded by the performer (or MIDI file) and slides smoothly to a nearby "most consonant" set of intervals. The tonal center in *Recalled Opus* was kept reasonably stable by careful composition. A context was used to ensure stability of *Wing Doneveir*. In contrast, the pitches fall where they may in *Local Anomaly* and there is no clear notion of musical "key." It is easy to hear the wriggling about of the tonal center (if indeed this piece can be said to have one). Perhaps it is better to think of it as having an "average" tonality that happens to have a large variance.

It is not easy to put these effects into words. The tonality is slinky and greasy, the drums funky and somewhat dark. The piece has an energetic minor cast. Even though there are both (just) major and (just) minor thirds throughout, the primary perception is of their wriggling around. There is a sense in which *Local Anomaly* "gets rid of scales and chords," bypassing any kind of fixed-pitch scales or tunings. At the same time, it is not without a considerable structure that is readily perceptible.

9.7 Toward an Aesthetic of Adaptation

The adaptive tuning strategy can be viewed as a generalization of just intonation in two respects. First, it is independent of the key of the music being played; that is, it automatically adjusts the intonation as the notes of the

piece move through various keys. This is done without any specifically "musical" knowledge such as the local "key" of the music, although such knowledge can be incorporated in a simple way via the context, the unheard drone. Second, although not stressed here, the adaptive tuning strategy is applicable to inharmonic as well as harmonic sounds. This broadens the notion of just intonation to include a larger palette of sounds. The adaptation provides a kind of "intelligent" portamento that begins where commanded by the performer and slides smoothly to a nearby "most consonant" set of intervals.

A shortcoming of the adaptive tuning approach is that sensory consonance is not a globally desirable property in music. Typically, a composer strives to move from consonance to dissonance and back again, and so indiscriminate application of the algorithm may, at least in principle, lead to pieces that lose appropriate dissonances. In practice, this may not be a large problem because it is always easy to increase dissonance by increasing the complexity of the sound, for example, by playing more notes. Alternatively, the algorithm could be applied selectively to places where consonance is most desired.

An extreme example occurs in *Maximum Dissonance* [S: 80], which, like its name, reverses the effect of the algorithm so as to maximize (rather than minimize) the sensory dissonance at each time instant. The piece is fairly difficult to listen to, especially at first, although it has a certain rhythmic vitality. Even with all of the dissonance, it cannot be said to be truly unlistenable (like the mismatched tuning/timbre combinations in sound examples [S: 3] and [S: 5]). This is probably because the dissonance is not uniform; it increases and decreases with the number of notes. With few notes, the algorithm can only increase the dissonance a small amount; with more notes, the algorithm is able to increase the dissonance significantly.

Considered as a group, perhaps the most obvious feature of the adaptively tuned pieces in Table 9.1 is the pitch glides—rarely do notes sustain without changing pitch. A sensible strategy when orchestrating such a piece is to use timbres that familiarly bend and slide: for example, violin and fretless bass rather than harpsichord and piano. Another technique that is used extensively in these pieces is hocketing; rather than playing a melodic passage with a single instrumental sound, each note is performed with a different sound. *Inspective Liquency* and *Aerophonious Intent* incorporate extensive hocketing. Pitch instabilities are not, however, an intrinsic property of the adaptive process, but rather a function of the particular program (i.e., `Adaptun`) used to carry out the adaptation. For example, pitch glides are absent from *Wing Donevier* and *Persistence of Time*.

The compositional technique of adaptive randomization begins with a pattern that is random melodically and harmonically (although not rhythmically). Complexity can be added to the sequence in many ways: duplicating notes and offsetting them in time or transposing in pitch, reversing patterns in time, randomizing or inverting pitches, quantizing, and so on. After orchestrating, some semblance of tonal order can be reimposed using the adaptive

tuning. Full pieces can be constructed by cut-and-paste methods. Of course, more traditional compositional methods may still be applied.

By functioning at the level of successions of partials (and not at the level of notes), the sensory dissonance model does not deal directly with pitch, and hence it does not address melody, or melodic consonance. Rasch [B: 146] describes an experiment in which:

> Short musical fragments consisting of a melody part and a synchronous bass part were mistuned in various ways and in various degrees. Mistuning was applied to the harmonic intervals between simultaneous tones in melody and bass... The fragments were presented to musically trained subjects for judgments of the perceived quality of intonation. Results showed that the melodic mistunings of the melody parts had the largest disturbing effects on the perceived quality of intonation...

Interpreting "quality of intonation" as roughly equivalent to melodic dissonance, this suggests that the misalignment of the tones with the internal template was more important than the misalignment due to the dissonance between simultaneous tones.

Such observations suggest why attempts to retune pieces of the common practice period into just intonation, adaptive tunings, or other theoretically ideal tunings may fail[15]; squeezing harmonies into just intonation requires that melodies be warped out of tune. If the melodic dissonance described by Rasch dominates the harmonic dissonance, then the process of changing tunings may introduce more dissonance, albeit of a different kind. This does not imply that it is impossible (or difficult or undesirable) to compose in these alternative tunings, nor does it suggest that they are somehow inferior; rather, it suggests that pieces may be more appropriately performed in the tunings in which they were conceived.

9.8 Implementations and Variations

There are several ways that adaptive tunings can be added to (or incorporated in) a computer-based music environment. These include:

(i) Software to manipulate Standard MIDI Files (or the equivalent). In such an implementation, the musician or composer generates a Standard MIDI File (SMF). The adaptive tuning algorithm is implemented as a software program that reads the SMF, adapts the tuning of the notes as described above, and writes a modified SMF file that can subsequently be

[15] The effort to improve Beethoven or Bach by retuning pieces to just intonation produced a sense that the music was "unpleasantly slimy" (to quote George Bernard Shaw when listening to Bach on Bosanquet's 53-tone per octave organ [B: 106]) or badly out of tune due to the melodic distortions.

played via standard sound modules or manipulated further by the musician/composer in a sequencer program.

(ii) A stand-alone piece of hardware (or software to emulate such hardware) that interrupts the flow of MIDI data from the controller (for instance, the keyboard), adapts the tuning, and outputs the modified notes.

(iii) The adaptive tuning strategy can be incorporated directly into the sound generation unit (the synthesizer or sampler).

(iv) Direct manipulation of digitized sound.

The software strategy (i) has the advantage that it may be simply and inexpensively added to any computer-based electronic music system. The disadvantage is that it is inherently not a real-time implementation. On the other hand, both the stand-alone approach (ii) and the built-in approach (iii) are capable of real time operation. `Adaptun` is in the second class. As the algorithm requires the spectra of the sounds, this must be input by the operator in both (i) and (ii). Of course, a frequency analysis module could be added to the software/hardware, but this would increase the complexity. The built-in solution (iii) does not suffer from any of these complications (indeed, the synthesizer inherently "knows" the spectrum of the sound it is producing) and is consequently preferred for MIDI implementation, although it would clearly require a commitment by musical equipment manufacturers.

The adaptive tuning can also be implemented in hardware (or software to emulate such hardware) that directly manipulates digitized sound. Such a device would perform an appropriate analysis of the sound (a Fast Fourier Transform, wavelet decomposition, or equivalent) to determine the current spectrum of the sound, run the adaptive algorithm to modify the spectrum, and then return the modified spectrum to the time domain with an inverse transform. The device could be operated off-line or in real time if sufficient computing resources were devoted to the task. Such an implementation is not, however, completely straightforward: it may be more of an adaptive "timbre" algorithm than an adaptive "tuning." This is an exciting area for future research.

Throughout Chaps. 8 and 9, the adaptive tuning algorithm has been stated in terms of an optimization problem based on dissonance curves solvable by gradient descent methods. Other algorithms are certainly possible. For instance, instead of laboriously descending the error surface, an algorithm might exploit the fact that the adaptation often converges to intervals that align the partials of simultaneously sounding notes. An algorithm that operated by simply lining up the partials would have much the effect of the consonance-based adaptation without much of the overhead. More generally, other optimization criteria based on other psychoacoustic measures of sound quality and solvable by other types of algorithms are also possible. For example, incorporating a virtual pitch model or a model of masking might allow the algorithm to function in a wider range of situations. Indeed, as the state of knowledge of psychoacoustic phenomena increases, new methods of adaptation seem likely.

9.9 Summary

Just as the theory of four taste bud receptors cannot explain the typical diet of an era or the intricacies of French cuisine, so the theories of sensory dissonance cannot explain the history of musical style or the intricacies of a masterpiece. Even restricting attention to the realm of sensory dissonance, the average amount of dissonance considered appropriate for a piece of music varies widely with style, historical era, instrumentation, and experience of the listener.

The intent of `Adaptun` is to give the adventurous composer a new option in terms of musical scale: one that is not constrained *a priori* to a small set of pitches, yet that retains some control over consonance and dissonance. The incorporation of the "context" feature helps to maintain a sense of melodic consistency while allowing the pitches to adapt to (nearly) optimal intervals.

This algorithm does not avoid the melodic artifacts associated with just intonation, but it automates intonation decisions. Perhaps more importantly, it can handle sounds with inharmonic spectra, such as bells, which fall outside conventional tuning theories.

10

The Gamelan

The gamelan "orchestras" of Central Java in Indonesia are one of the great musical traditions. The gamelan consists of a large family of inharmonic metallophones that are tuned to either the five-note slendro *or the seven-tone* pelog *scales. Neither scale lies close to 12-tet. The inharmonic spectra of certain instruments of the gamelan are related to the unusual intervals of the pelog and slendro scales in much the same way that the harmonic spectrum of instruments in the Western tradition is related to the Western diatonic scale.*

10.1 A Living Tradition

The gamelan plays many roles in traditional Javanese society: from religion and ceremony to education and entertainment. In recent years, recordings of gamelan music have become available in the West.[1] First impressions are often of an energetic, strangely shimmering sound mass punctuated with odd vocal gestures. The exotically tuned ensemble plays phrases that repeat over and over, with variations that slowly evolve through pieces of near symphonic length. A deep gong punctuates sections, and the music is driven forward by vigorous drumming and dynamic rhythmic articulations. Indeed, the word *gamelan* can be translated literally as "pounding of a hammer."[2]

The unique sounds are produced by an assortment of metallophones that include numerous gongs and xylophone or glockenspiel-like instruments of various sizes, timbres, and tones. At first glance, the *bonangs* and *kenongs* appear to be collections of upside-down pots and pans hit with sticks, and the *saron* players seem to pound a small collection of metal bars with hammers. As we will see, this is akin to viewing a Stradivarius as a wooden box with strings. Gamelan instruments are finely crafted, carefully tuned, and are the

[1] For instance, the excellent series from the World Music Library includes *Gamelan Gong Kebyar of Eka Cita* [D: 18], *Gender Wayang of Sukawati* [D: 19], the *Klênêngan Session of the Solonese Gamelan* [D: 25], *Gamelan Gong Gede of the Batur Temple* [D: 17], and the *Gamelan of Cirebon* [D: 16]. Other recordings are available from the Library of Congress (*Music for the Gods* [D: 29]), from CMP records (*Gamelan Batel Wayang Ramayana* [D: 15]), from Lyrichord (music of I. W. Sadra [D: 38]), and from Nonesuch (*Music from the Morning of the World* [D: 27].

[2] *gamel* means "hammer," and *-an* is a suffix meaning "action."

result of a long cultural tradition that values precision and refinement in its music, instruments, and musicians.

The first major study of the instruments, repertoire, and history of the gamelan ("the result of twenty-eight years' listening, collecting, and reflecting") was the landmark *Music in Java* [B: 90]. Kunst discusses the various instruments of the gamelan and the tuning systems and observes a difference in the listening aesthetic between the Javanese and the Western ear:

> of necessity a virtue was born: this partial discrepancy between vocally and instrumentally produced tones has developed unmistakably into an aesthetic element... a play of tensions alternately arising and disappearing... these discrepancies in intonation are to some extent satisfying to the Javanese ear.[3]

Kunst's love of the music and the people is obvious, and he catalogs a number of gamelan "themes" so that they would not be lost. Kunst offers a dire warning:

> Once again foreign influences are affecting it [gamelan music], but this time the interloper is... like a corrosive acid, like a transfusion from a different blood group, [which] attacks and destroys it in its profoundest essence... one can almost watch–or rather hear–native music degenerating day by day.

Fortunately, this apocalyptic vision has failed to materialize, and gamelan music has not only survived, but flourished.

There are many reasons why gamelan music challenges Western listeners. The timbre of the instruments is unusually bright and harsh. The scales and tunings are unfamiliar. Both the tunings and the timbres are discussed at length in later sections because they are easily quantifiable. But there are also profound differences in the basic structure of the music. In the *Guide to the Gamelan*, Sorrell [B: 177] describes the Javanese concept of an *inner melody* in the evocative passage:

> the concept of an inner melody... is the common basis of all the parts in the gamelan and yet which is not literally stated by any instrument. Rather, it is in the minds of the musicians. It is therefore felt, or, one may say, internally *sung*.

Thus, listening to and understanding the inner melody of a gamelan piece is different from listening to and understanding the outer melody of a symphony. In many traditional Western forms, the themes are stated, developed, and restated. In contrast, the gamelan performance presents many different ways of disguising the same underlying theme. An analogy may be fruitful. A syncopated rhythm has an underlying pulse. Although this pulse may never be

[3] A modern investigation of the perception of music among the Balinese can be found in [B: 82].

stated literally, it forms an essential part of the listening experience. To truly "understand" the syncopated rhythm, it is necessary to "hear" something that is not there!

10.2 An Unwitting Ethnomusicologist

There are as many different gamelan tunings as there are gamelans because instruments in the Indonesian musical tradition are not all tuned to a single standard reference scale. Rather, each instrument is tuned and timbrally adjusted to work in its own orchestral context; each instrument is created for and remains with a single ensemble. Each gamelan is tuned to its own variant of pelog or slendro. Every kettle of each bonang, every key of each saron, is hand shaped with hammer and file. The result is that a piece played on one gamelan inevitably differs in intonation, tone, and feel from the same piece played on another gamelan.

This presents an intriguing challenge. Recall that Western diatonic scales are intimately connected to[4] sounds with harmonic spectra. Perhaps a similar relationship exists between the pelog and slendro scales and the inharmonic sounds of the saron, bonang, gender, or gong. Further, perhaps the differences between the tunings of various gamelans can be explained in terms of the differences between the spectra of the various instruments.

An obvious starting point is to search the literature, and to correlate the spectra of the gamelan instruments with the tunings of the gamelans from which they come. Although several important studies over the years have documented the variation in the tunings of the gamelans, only one published article [B: 159] has detailed the spectra of any gamelan instruments, and this was not a complete study, even of the one gamelan. Of the metallophones, only the *jegongan* (a kind of Balinese gender) and the gong are studied. Clearly, more data are required.

Accordingly, I traveled to Indonesia between August and December 1995. A portable DAT machine and microphone[5] made it possible to carry everything needed for full fidelity recordings, which could be analyzed back in the lab. Gathering more data (i.e., recording each key of each instrument in the gamelan) was not straightforward. Although equipped for the technological task, I was underprepared for the social and cultural aspects. A few months of study of *Bahasa Indonesia* (the language) was adequate for basic survival, but it was not enough to conduct genuine interviews. Reading several books[6] on ethnomusicology (in general) and Indonesia (in particular) readied me for

[4] In the jargon of the previous chapters, "related to."

[5] Along with rechargeable batteries, a copy of *Everyday Indonesian*, and a backpack of essentials.

[6] Including the excellent general works by Merriam [B: 112] and Nettl [B: 121], and books specifically about the gamelan such as those by Kunst [B: 90], Sorrell [B: 177], and Tenzer [B: 193].

some of the issues I would confront, but it was not enough to provide ready answers.

In particular, it was difficult to approach gamelan masters with my request, in part because of the oddity of the task (usually people are more interested in gamelan performance and music than in the instruments), in part because of language difficulties, and in part because of property issues. Gamelans are often owned by the village, and it is considered improper for individuals to profit from public resources. This was further complicated by the diversity of Indonesian society; each region has its own customs and sense of propriety. Offering the gamelan master a small gift earmarked for the gamelan (to help with maintenance and upkeep) often seemed to be appropriate.

Eventually, I met Basuki Rachmanto at the University of Gadjah Mada in Yogyakarta, who became interested in the project, and helped find and record eight complete (pelog and slendro) gamelans. Basuki also introduced Gunawen Widiyanto, the son of a respected gamelan-smith in Surakarta. Gunawen arranged to record nine complete gamelans in the Surakarta area and helped me to interview several gamelan makers and tuners. Without the generous help of Basuki and Gunawen, it would have taken far longer to have accomplished far less. In addition, I am grateful to Ben Suharto of the ISI in Yogyakarta, and to Deni Hermawan of the STSI in Bandung for allowing me to record their "performance" gamelans.

10.3 The Instruments

Most of the idiophones of the gamelan are percussion instruments made from metal. They are struck with a variety of mallets that range from hard wood to woolen ball heads; harder mallets give a brighter tone with more high partials, and softer mallets return a more muted sound. Names of the instruments vary by region, and the names used here (gong, gender, saron, bonang, kenong, gambang) are common in the Central Javan cities of Yogyakarta and Surakarta.

Most of the instruments consist of a set of keys, kettles, or bells of definite shape, arranged on a wooden frame so that they may be readily struck by the performer. Each key is hand forged in a charcoal furnace. This is a slow, grueling process; a crew of three or four workers can beat a hot slab of metal into a rough bowl shape over the course of several hours.[7] Detailed shaping is done by hammer once it has cooled, and then the keys are polished. A complete set of keys is tuned by the master tuner using a hand file, although the final tuning is not done until all of the instruments are assembled.

Like most percussion instruments, the metallophones of the gamelan have inharmonic spectra. Each kind of instrument has its own idiosyncrasies, and the remainder of this section looks at each of the instruments in turn. All

[7] It takes 60 workers about 5 months to build a complete gamelan.

samples in this chapter are from either the Gamelan Swastigitha,[8] which is under the capable direction of Suprapto Atmosutijo, or from Gamelan Kyai Kaduk Manis, which was built for Pak Cokro (K.R.T. Wasitodiningrat), also of Yogyakarta. Gamelan Kyai Kaduk Manis was built in 1997, is in excellent condition, and hence is a good example of a modern gamelan, although it was modeled after one of the palace gamelans in Surakarta. Gamelan Swastigitha is considerably older, although it is certainly post-World War II.

10.3.1 Saron

Sarons are a kind of metal keyed xylophone. Each key is a solid rectangular chunk of bronze whose top has been rounded slightly, as in Fig. 10.1. Keys are suspended above a trough-shaped frame on two metal pins. Sarons appear in a large range of sizes (and hence pitches), and each usually has between six and nine keys.

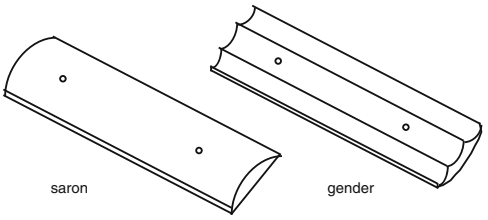


Fig. 10.1. Keys of the saron and gender act much like uniform metal bars, but details of their shape and contour cause important differences in the spectra of the sound.

Sarons are usually played with an interesting two-handed technique. First, the wooden hammer strikes a key at an angle so that the mass of the hammer does not interfere with the resonance. The player then mutes the key with the thumb and forefinger of the free hand by pinching it. Thus, at each moment, the player strikes a new note while damping the old. Fast passages are played by two (or more) players hocketing on matched instruments, that is, alternating notes in a predetermined way. The saron often plays the main theme, although it can also be heard playing a supporting role by syncopating or duplicating the main themes. Its keen, sparkling sound is one of the most characteristic timbres of the gamelan.

The sound, and hence the spectrum of the saron, varies somewhat from gamelan to gamelan, but the pitch is always determined by the fundamental. The spectra appear to come in two basic varieties. The simpler kind is shown in Fig. 10.2, which plots the spectra of two typical saron keys from gamelan Swastigitha.[9] The top spectrum has partials at f, $2.71f$, and $4.84f$, and the bottom spectrum has partials at f, $2.62f$, $4.53f$, $4.83f$, and $5.91f$. Over the

[8] Ngadinegawan MJ 3/122, Yogyakarta.

[9] Except where explicitly stated, all spectra in this chapter were computed using a 32K FFT. Each plot represents the behavior in the first 3/4 second of the sample.

whole set of instruments, four partials appear consistently. The median of these values is

$$f,\ 2.76f,\ 4.72f,\ \text{and}\ 5.92f$$

which may be taken as a kind of generic saron key for this gamelan. Observe that this is close to, but significantly different from, the spectrum of an ideal bar. In particular, the third and fourth partials of the ideal bar are $5.4f$ and $8.9f$, and the Swastigitha sarons are uniformly lower.

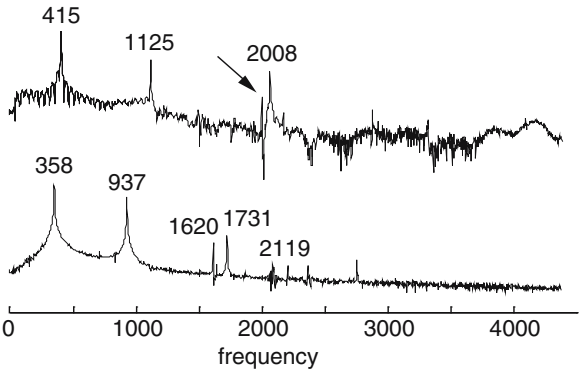


Fig. 10.2. Spectra of two typical keys of a saron from gamelan Swastigitha from Yogya.

The second kind of saron spectrum is exemplified by the sarons of Gamelan Kyai Kaduk Manis in Fig. 10.3, which have prominent partials at

$$f,\ 2.34f,\ 2.76f,\ 4.75f,\ 5.08f,\ 5.91f,\ \text{and at}$$
$$f,\ 2.31f,\ 2.63f,\ 4.65f,\ 5.02f,\ 6.22f.$$

Essentially, the partials near 2.7 and 4.8 have bifurcated so that a pair occurs where previously there was one. An idealized or generic version of the sarons of Gamelan Kyai Kaduk Manis is

$$f,\ 2.39f,\ 2.78f,\ 4.75f,\ 5.08f,\ 5.96f.$$

The origin of the bifurcated partials so prominent in the sarons of Kyai Kaduk Manis is not obvious. Perhaps they are caused by some impurity (or nonuniformity) in the brass, or perhaps from some accidental deviation in physical dimensions, but these seem unlikely because the intervals between the pairs are so consistent across the keys of all 11 sarons. Rather, it would appear that this timbre is intentional, that the tuner chose to encourage these closely spaced modes.[10] Indeed, referring back to the Swastigitha sarons, the higher

[10] Perhaps it is inherent in the rounded shape of the saron keys, or perhaps it is caused by some careful sculpting of the physical contour of the keys. If, for

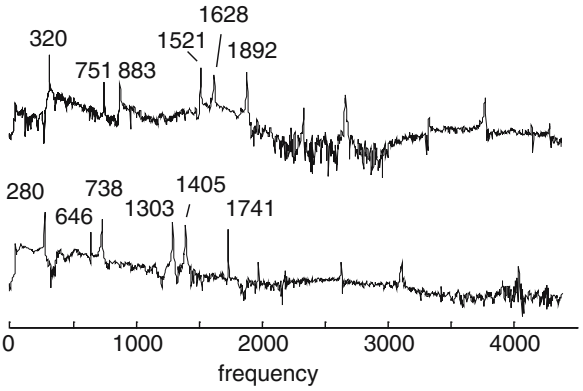


Fig. 10.3. Spectra of two typical keys of a saron from gamelan Kyai Kaduk Manis from Yogya.

of the two pairs are visible; they are prominent in the bottom spectrum, and the arrow in the upper spectrum points to a small, but observable bifurcated partial.

10.3.2 Gender

The gender is a metallophone with thin bronze keys (see Fig. 10.1) that are suspended above tubular resonators, much like a vibraphone. The air column vibrates in sympathy with certain partials, reinforcing the sound. When tuning a gamelan, the gender is usually tuned first, and all other instruments are tuned to the gender.

Genders are often played with soft disk-headed mallets, in such a way as to paraphrase and restate the melody. The padded mallet tends to give a soft, mellow sound. As the instrument resounds for a long time, the player usually mutes old notes with the heel of the hand while striking new notes. Larger (lower pitched) genders play slowly, and the smaller and higher pitched instruments move more rapidly.

The spectra of two typical gender hits are shown in Fig. 10.4. These have prominent partials at

$$f,\ 2.01f,\ 2.57f,\ 4.05f,\ 4.8f,\ 6.27f,\ \text{and}$$
$$f,\ 1.97f,\ 2.78f,\ 4.49f,\ 5.33f,\ 6.97f$$

which can be interpreted as a metal bar (the partials at or near $2.7f$ and $5.3f$) or as a modified saron bar (the partials at or near $2.7f$ and $4.8f$) in conjunction with harmonic partials at or near $2f$, $4f$, and $7f$. This makes

instance, one side of the key was slightly thinner than the other, then the two sides might vibrate at slightly different frequencies.

physical sense because the gender is a metal bar. The harmonic partials are likely due to the resonances of the air column.

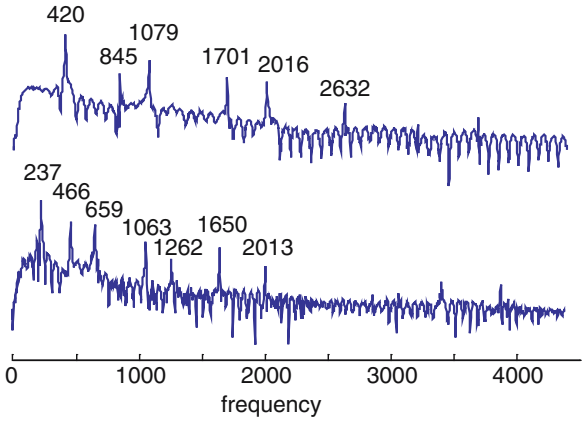


Fig. 10.4. Spectra of two typical gender hits.

In [B: 159], the resonances of four bars of a *jegongan* (a large Balinese gender) are found to be nearly identical to the resonances of an ideal bar. Presumably, these were measured without the air resonances, because there is no hint of the harmonic partials that are so prominent in Fig. 10.4.

10.3.3 Bonang

A bonang usually consists of two tiers of bronze kettles. Each kettle is shaped like a broad-rimmed gong as in Fig. 10.5, and it is suspended open side downward on two strings tied to a wooden frame. The player holds two hard, wrapped mallets, and strikes the protruding knobs on the top end. The kettles in a slendro bonang are often arranged antisymmetrically:

$$\begin{array}{ccccc} 6 & 5 & 3 & 2 & \dot{1} \\ 1 & \dot{2} & \dot{3} & \dot{5} & \dot{6} \end{array}$$

in the two ranks so that the performer can easily play (near-octave) pairs of notes. The dots indicate notes in the octave above or below.

A typical pelog bonang is similarly arranged:

$$\begin{array}{ccccccc} 4 & 6 & 5 & 3 & 2 & 7 & \dot{1} \\ 1 & \dot{7} & \dot{2} & \dot{3} & \dot{5} & \dot{6} & \dot{4} \end{array}$$

Kunst describes the musical function of the bonangs eloquently:

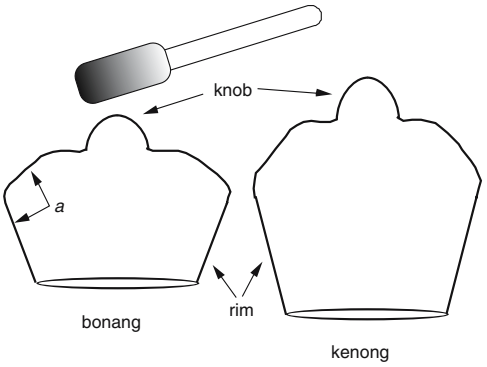


Fig. 10.5. The kettles of the bonang and kenong are shaped similarly, but the rim of the kenong is longer and the sound generally sustains longer.

> [the bonangs] devote themselves to the paraphrasing of the main theme. Now they anticipate it, now they analyze it into smaller values and imitate it in the octave. Then again, they syncopate it... then they fill up the melodic gaps with their penetrating tinkling sound.

As the bonang has a unique bell-like shape, there is no ideal to which it can be compared. The spectrum of three different bonang kettles have prominent partials at

$$f,\ 1.58f,\ 3.84f,\ 3.92f$$
$$f,\ 1.52f,\ 3.41f,\ 3.9f,$$
$$f,\ 1.46f,\ 1.58f,\ 3.47f,\ 3.71f,\ 4.12f,\ 4.49f$$

as shown in Fig. 10.6. The first two are typical and a good generic bonang spectrum is

$$f,\ 1.52f,\ 3.46f,\ 3.92f.$$

Many of the bonang kettles also demonstrate the behavior of bifurcating partials previously encountered in certain of the more complex saron keys. For instance, in the lower spectrum in Fig. 10.6, the partials at $1.46f$ and $1.58f$ might be interpreted as children of the generic bonang partial at $1.52f$, and those at $3.47f$ and $3.71f$ might be derived from the generic partial at $3.46f$.

The kenong is a kind of kettle with a larger rim that makes a clear and sustained sound. It is often used to subdivide the long gong phrases into smaller pieces, and hence it serves a primarily rhythmic function. Spectra of the kenong are similar to those of the bonang, despite the differences in shape.

10.3.4 Gong

Perhaps the most characteristic sound of the gamelan is the deep, dark strokes of the gong marking the end of each musical phrase. The largest gongs can have a diameter up to a meter, weigh 60 or more kilograms, and have a fundamental frequency of only 40 or 50 Hz. Gongs may come in a variety of shapes, and Fig. 10.7 shows a fairly common profile.

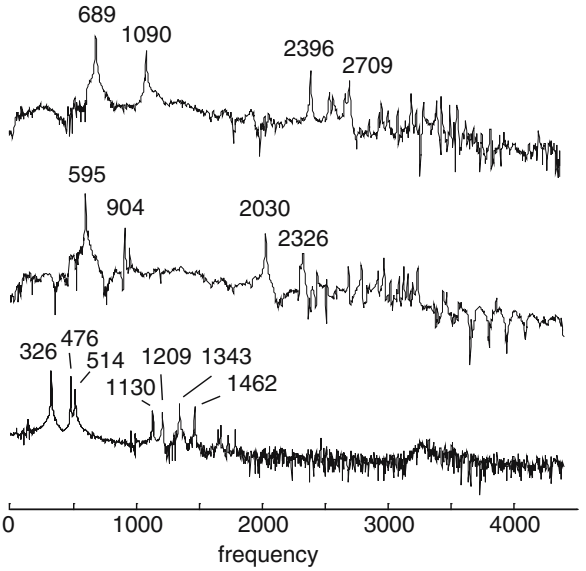


Fig. 10.6. Spectra of three typical bonang kettles.

According to tradition, gongs are of divine origin, and they were used as a signaling system among the Gods. Kunst [B: 90] reports that some gongs are protected by powerful beliefs; for instance, no European is allowed to touch the sacred gong at Lodaya. "One civil servant, who ventured nevertheless to touch it, died soon afterwards."

Without a doubt, the acoustic behavior of gongs is complicated. Figure 10.8 shows the first four seconds of a gong stroke, divided into 32K (3/4 second) segments. The first ten partials are at frequencies

$$90,\ 135,\ 151,\ 180,\ 241,\ 269,\ 314,\ 359,\ 538,\ 626$$

which is

$$f,\ 1.49f,\ 1.67f,\ 2f,\ 2.67f,\ 2.98f,\ 3.47f,\ 3.98f,\ 5.97f,\ 6.94f$$

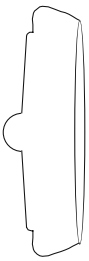


Fig. 10.7. The giant gongs of the gamelan have a rich deep sound that can last well over 30 seconds. "The sound of the gong, beaten heavily, rolls on its ponderous beats like the ocean tide." Quoted from Kunst [B: 90].

for $f = 90$ Hz, the perceived pitch. All of these partials are integer multiples of 15 Hz,[11] which is not directly perceptible. Equivalently, the "scale" formed by these ten partials (after reduction back into a single octave) is

$$1,\ 4/3,\ 3/2,\ 5/3,\ 7/4,\ 2$$

which is a simple just pentatonic scale.

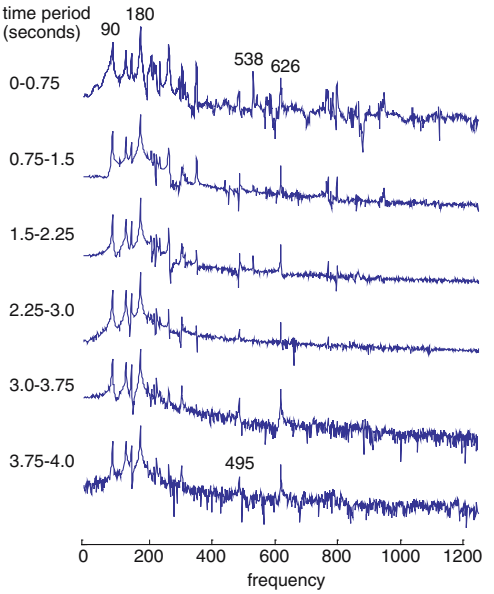


Fig. 10.8. Partials of the gong rise and fall as time evolves. Curves show the spectrum for successive time periods.

One interesting behavior is the rising and falling of partials as the sound evolves. For instance, consider the partial at 626 Hz, which slowly decays in amplitude until 3 seconds, when it suddenly begins to regain prominence. Similarly, the partial at 495 Hz falls and then grows, Such energy exchanges give the gong its characteristic evolving timbre—as if the partials of the gong are smoothly sweeping up and down the pentatonic scale.

Rossing and Shepherd [B: 159] suggest that the two prominent octave partials (at 90 and 180 Hz in this case) that determine the pitch arise from two axisymmetric modes of vibration and are tuned by careful control of the ratio of the mass of the dome to the total mass.

10.3.5 Gambang

The gambang is essentially a Javanese xylophone. Three or four octaves of wooden keys lie on soft cushions that are mounted on a wooden frame. The

[11] Rameau [B: 145] would have found this remarkable.

lower keys tend to be large and flat, and the higher keys are shorter and rounder. The sound is heavily damped, more of a plink than a dong. The spectra of typical gambang strikes are shown in Fig. 10.9. These are very close to the spectrum of an ideal bar, and hence the gambang is best thought of in this way.

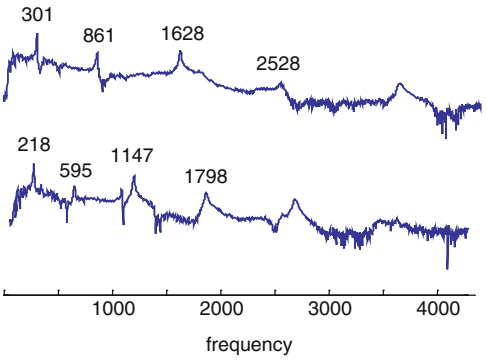


Fig. 10.9. Observe how these two hits of the gambang have spectra close to that of an ideal bar. The top has partials in the ratios 1, 2.86, 5.4, 8.4 and the bottom has partials in the ratios 1, 2.73, 5.26, 8.3.

10.3.6 Other Instruments

The *kendang* is a full-bellied wooden drum, not dissimilar to a conga drum. The head is traditionally made of buffalo skin that is stretched by means of rattan hoops. The kendang player is, more than anyone else, the conductor of the gamelan. Often, the kendang signals impending changes by stylized rhythmic messages, and subtle hand motions are used to indicate which parts are to be emphasized.

Besides the fixed pitch instruments of a typical gamelan, there are instruments that are often used in specific kinds of gamelan styles. In some styles, the theme is played[12] by the *rebab*, a two-stringed bowed lute with a heart-shaped body. By its nature, the rebab plays far more fluidly than the metallophones. The strings are often made from thin copper wire, and the bow is stretched taut by two fingers of the right hand, much like the Chinese *erhu*. There is no fingerboard as on a violin; rather, the strings are stopped by pressure from the fingers alone. Because the bow is applied near the bridge, the rebab has a more nasal quality than the violin. The spectrum of the sound is primarily harmonic, as expected from a stringed instrument.

The *suling* is an aerophone, an end blown bamboo tube with tone holes cut appropriately to sound in the pelog or the slendro scale. Air is forced to cross the wedge-shaped sound hole by means of an ingenious bamboo ring that encircles the mouthpiece. It is thus as easy to blow as a Western recorder. It

[12] Sometimes the rebab lags the "melody" (the *balungan*) slightly, and sometimes it anticipates.

is also easy to bend the pitches of notes by partially covering the holes, which allows the suling to imitate the call of a bird or the inflections of a voice in its richly ornamented parts. Like most instruments based on the resonance of air columns, the spectrum is primarily harmonic.

Finally, gamelan performances often include singing. This may be during an interpretation of the *wayang kulit* (shadow puppets), or it may represent a characters voice in a dramatic stage performance or a popular show. Thus, gamelan music includes several families of inharmonic instruments, each with their own character, and yet retains a basic compatibility with harmonic instruments such as the rebab, the suling, and the human voice.

10.4 Tuning the Gamelan

Gamelan tunings come in two flavors: the five-note slendro and the seven-note pelog. The earliest reported measurements of these tunings are from Kunst [B: 90], who observed that the interval between each note in a slendro scale is equal to 240 cents. This implies that slendro is similar to 5-tet:

note:	6	1	2	3	5	6
cents:		240	240	240	240	240

The naming of notes is only partially numerical. In slendro, there is no 4, and the scale is often considered to start (and end) on 6.

Pelog, according to Kunst, is more complex, consisting of seven unequal divisions of the octave:

note:	1	2	3	4	5	6	7	1
cents:		120	150	270	150	115	165	250

Unfortunately, Kunst's tone measurements were conducted using a monochord (a stretched string, to which the desired tones are compared by ear) and so are of limited accuracy. As more modern investigations show, the above scales are only part of the story.[13]

First, each gamelan is tuned differently. Hence, the pelog of one gamelan may differ substantially from the pelog of another. Second, tunings tend not to have exact 2:1 octaves. Rather, the octaves can be either stretched (slightly larger than 2:1) or compressed (slightly smaller). Third, each "octave" of a gamelan may differ from other "octaves" of the same gamelan. And fourth, there is usually some note that is common between the slendro and pelog scales of a given gamelan, although matching notes differ from gamelan to gamelan.

An extensive set of measurements is carried out in *Tone Measurements of Outstanding Javanese Gamelans in Yogyakarta and Surakarta* [B: 190], which

[13] Kunst also offers an explanation for the tunings of the gamelan in terms of von Hornbostel's theory of a cycle of "blown" (compressed) fifths.

gives the tunings of 70 gamelans.[14] The measurements were taken using an analog electronic system with an accuracy of about 1 Hz. The technique requires that all higher partials be filtered out, and so only the fundamentals are reported. This is completely adequate for measuring the tunings, because the pitches of the metallophones are determined by the fundamentals. Unfortunately, it means that information about the timbre (spectra) of the instruments has been lost.

Kunst measured the tuning of one saron in each gamelan, and extrapolated from that to the tuning of the whole gamelan. This was unfortunate for two reasons. First, tunings may differ somewhat depending on the register. Second, Kunst failed to observe that the tunings were not genuinely octave based. For instance, the notes 6 and D:6 (or $\dot{6}$ and 6) need not be in an exact 2:1 ratio. This latter fact is one of the most remarkable aspects of the gamelan tunings, at least from the octavo-centric Western viewpoint. The octave stretching (and compressing) is amply demonstrated in [B: 190], and pseudo-octaves ranging from 1191 to 1232 cents are reported.[15]

Another striking aspect of the data in [B: 190] is the accuracy to which gamelans are tuned. For instance, of the 11 instruments tuned to pitch 6 in the fifth register of Gamelan Kyahi Kanyutmesem (Table 3 of [B: 190]), all are within 3 Hz of 582. Eight are within 1 Hz of 580. It is therefore not a tenable position that gamelan octaves are stretched or compressed by accident, or by inability to tune the instruments accurately enough. Similarly, the differences in tuning between various gamelans are far greater than the variation within gamelans. The variety of gamelan tunings is intentional.

10.4.1 A Tale of Two Gamelans

This section examines the tunings of Gamelan Swastigitha and Gamelan Kyai Kaduk Manis in detail. The slendro tuning of Gamelan Swastigitha is shown in Table L.2 on p. 378, where the calculation of the fundamental of each key is accurate to about 1 Hz. With the exception of the gambang,[16] the tuning is extremely consistent. Different instruments in the same column have keys at the same pitch, and these rarely differ by more than 1 or 2 Hz. For example, the six metallophones at note 6 in register II are all between 471 and 472 Hz.

The last row of the table shows the median values within each column, and this represents an idealized tuning for this gamelan. Translating these values into cents and arranging by register shows the internal structure of this slendro scale:

[14] Originally published in Indonesian in 1972, this book has been recently translated into English.

[15] Carterette [B: 26] reanalyzes the data from [B: 190] and describes the stretching of the scales concretely by finding the best exponential fit.

[16] It may be that the gambang is harder to tune than the others because of its short envelope. It may also be that the wood becomes nicked, scratched, and detuned far more easily than the metallophones.

Gamelan Swastigitha: Slendro

register	intervals					"octave"
I	252	240	244	244	239	1219
II	233	249	243	235	246	1206
III	235	248	238	252	237	1210
average	240	246	242	241	241	1210

Each octave is stretched by an average of 10 cents. The scale is remarkably uniform; the mean difference of this scale from 5-tet is 2 cents, and the maximum error is 6. To place this in perspective, consider the just major scale of Table 6.1 (p. 101) and its approximation by 12-tet scale steps. The mean difference between these two is 8.8 cents, and the largest error is 16 cents.

Similarly, the slendro tuning of Kyai Kaduk Manis is given in Table L.3 on p. 378. Reformatting this into cents gives:

Gamelan Kyai Kaduk Manis: Slendro

register	intervals					"octave"
I	231	223	239	247	253	1193
II	237	237	238	234	250	1196
III	243	239	225	250	242	1199
average	237	233	234	244	248	1196

Again, the scale is very close to 5-tet (mean difference of 5.6 cents, maximum difference eight cents), but the octaves of this tuning are compressed slightly. All of these values fall well within the ranges observed in [B: 190].

Pelog tunings for the gamelans are given in Tables L.4 and L.5 on pp. 379 and 380. Rearranging the data gives:

Gamelan Swastigitha: Pelog

register	intervals							"octave"
I	100	145	301	121	99	162	261	1189
II	133	153	275	117	106	181	234	1199
III	123	166	269	119	119	173	238	1207
average	119	155	282	119	108	172	244	1199

and

Gamelan Kyai Kaduk Manis: Pelog

register	intervals							"octave"
I	166	161	267	119	119	171	237	1240
II	147	145	274	115	104	197	209	1191
III	158	154	258	96	154	180	206	1206
average	157	153	266	110	126	183	217	1212

Obviously, pelog is not an equal-tempered scale. Surjodiningrat et al. [B: 190] average the tunings from thirty pelog gamelans to obtain

120, 138, 281, 136, 110, 158, 263

but they are clear to state that this "does not mean the best but only the average."

In fact, a general pattern for pelog scales is

$$S_1,\ S_2,\ L_1,\ S_3,\ S_4,\ S_5,\ L_2,$$

where the S_i represent small intervals and the L_i represent large intervals.[17] The actual values of the S_i and L_i vary considerably among gamelans and even within the same gamelan, so this pattern cannot be taken too literally.

10.4.2 Conversations about Tuning

Why is your gamelan tuned this way? While traveling through Indonesia, I asked this question many times. People who tune gamelans, those who play, and those who build them were often willing to comment, and their answers ranged from practical tuning advice to mystical explanations, from detailed historical justifications to friendly ironic smiles that meant "what a silly question."

Before describing the responses, consider the question. If asked why the piano is tuned as it is, perhaps you would describe the historical progression from Pythagorean to equal temperaments, perhaps comment how 12-tet allows modulation through all of the keys, perhaps describe how the major scale originates from a juxtaposition of certain major triads, as an approximation to the harmonic series, or as a conjunction of tetrachords.[18] Similarly, it would be unreasonable to expect any kind of unanimity of answers about gamelan tuning.

The most common answer was to name a gamelan that had been used as a tuning reference, reflecting a common practice for the initial tuning of the gender. For instance, Pak Cokro, the master of Gamelan Kyai Kaduk Manis, said that it was referenced to a respected gamelan at the palace in Surakarta. "In ancient times it was necessary to tune the gender right in the palace," said Pak Cokro, "but in modern times most people use a tape recorder." A gamelan by Siswosumarto[19] was similarly referenced to the gamelan at the National Radio Station,[20] and a gamelan of Mulgo Samsiyo[21] was referenced to a gamelan at the University in Yogyakarta.[22] Mulgo Samsiyo uses an electronic tuning device to tune the genders. "All the others are the same as the genders," he said.

[17] This provides an interesting inversion of the diatonic scale defined by $L,\ L,\ S,\ L,\ L,\ L,\ S$.

[18] If you were reading this book, you might comment how 12-tet is an approximation to a scale related to sounds with a harmonic spectrum.

[19] Kaplingan Jatiteken Rt. 04/V. (Timor Bengawan Solo) Ds. Laban-Mojolaban Skh Surakarta.

[20] RRI, Surakarta.

[21] Dk. Gendengan Rt. 1/IV. Ds. Wirun Mojolaban, Sukoharjo-Jateng.

[22] ISI, Yogyakarta.

Suhirdjan,[23] a gamelan maker and tuner in Yogyakarta, described the tuning procedure. "You pick a scale and then tune the gender to that scale. Then all the other instruments are fit to the gender." I asked how the initial scale is chosen. "Just tune until it sounds right," he said. This sentiment was echoed far more poetically by Purwardjito,[24] an instructor at the Arts College in Surakarta, "Gamelans are tuned to nature. In the west you tune with your mind. In Indonesia, we tune with the heart."

Both Suhirdjan and Purwardjito are proficient with the techniques of tuning. Each described in detail the parts of the saron key that must be scraped to raise or lower the pitch, and these accord well with techniques used to tune xylophone keys.[25] The bonang family is trickier, but both agreed that filing from the outside of the rim tends to lower the pitch, and filing the inside has the opposite effect. Filing the knob on the outside also raises the pitch. The greatest factor, however, is the angle marked a in Fig. 10.5; smaller angles correspond to lower pitches, and larger angles correspond to high pitches. "This should only be changed in the gong factory, since it is dangerous to hammer a bronze kettle–it might crack." Purwardjito continues, "It's also important that the walls be uniform. When the thickness is uneven, the sound damps out much more quickly. We say the sound is drowning in water." Gongs are hard to tune. "You never know which way the pitch is going to go when you hit or file it," says Suhirdjan, "Each gong has its own personality."

Neither tuner uses beats when he tunes, although both are well aware of their existence. Towards the end of the interviews, I asked "a complicated question." Grabbing a bonang, I placed my hand so as to damp out all but the fundamental. After I hit it, I whistled the pitch of the fundamental. I then shifted the position of my hand so as to damp out all but the partial at about $1.5f$, and then highlighted the pitch of this partial[26] by whistling. There were two kinds of reactions. Some of the informers, like Suhirdjan, denied that there were two different pitches. "I hear both as the same pitch... or as different parts of the same pitch," he said. "It's like when you hit the same kettle softly, it is the same pitch as when you hit it hard. They are the same pitch, but different." Clearly, Suhirdjan is listening holistically. Very likely he tunes in a holistic way as well.

Purwardjito's reaction was different. First he laughed. Then he said, "Ah, I see. You mean the supporting[27] tone... There are many kinds of tuning. There is the tuning in the furnace, where you determine the shape. There is the fine tuning with file and hammer. When you tune the gender [to the reference scale], you only pay attention to the pitch. But when you tune the bonang,

[23] Condronegaran Mj. 1/951, Gedong Kiwo, Yogyakarta.
[24] STSI Surakarta. Jur-Karawitan, Kentingan Jebres.
[25] See, for instance, [B: 124].
[26] Which to my ear was now the dominant sound.
[27] Gunawen, who was translating the conversation, conferred with Purwardjito for several moments, searching for the right word, eventually settling on "supporting."

the kenong, or the gongs, you pay attention to the supporting tones." This kind of attention is analytical listening, and presumably Purwardjito tunes analytically as well.

10.5 Spectrum and Tuning

Just as Western theoreticians do not generally think in terms of correlating the spectrum of an instrument with its tuning, Indonesian gamelan tuners are unlikely to have developed their scales with a detailed awareness of the spectra of their instruments. Rather, they used their ears to create compatible scales and instruments.

A key tool in relating harmonic sounds to diatonic (just) scales is the dissonance curve. The partials of the sound are specified, and then the related scale is defined by the minima of the dissonance curve. Although gamelan tuners can tune with remarkable accuracy, the number of different partials they can reliably control is limited, usually only two to four.[28] Such sparse spectra lead to dissonance curves with only a few widely spaced minima, not enough to explain any of the extant scales. Thus, the situation for the gamelan is a bit more complex, because no single instrument has the appropriate spectrum.

One clue to the resolution of this dilemma is in the first quote in this chapter where Kunst spoke of the "discrepancy" between the vocal and instrumental tones of the gamelan. Another clue is that gamelan music includes several kinds of inharmonic instruments, and yet it retains compatibility with harmonic instruments such as the rebab, suling, and the human voice. Thus, gamelan scales can be viewed in terms of the spectra of two different instruments. That is, both pelog and slendro scales can be viewed as minima of the dissonance curve[29] generated by an inharmonic instrument in combination with a harmonic sound.

10.5.1 Slendro

Slendro is simpler than pelog both because it contains fewer notes and because it varies less from gamelan to gamelan. A generic bonang with partials at $f,\ 1.52f,\ 3.46f,\ 3.92f$ was experimentally derived in the previous sections. Drawing the dissonance curve for this spectrum F in combination with a harmonic spectrum G with partials at $g,\ 2g,\ 3g,\ 4g$ gives the dissonance curve[30] of Fig. 10.10.

Observe that many of the minima of this curve occur at or very near steps of the 5-tet scale, which are themselves very near the steps of typical slendro

[28] Usually only two to four partials are at consistent intervals throughout an instrument.

[29] The section "Dissonance Curves for Multiple Spectra" in the chapter "Related Spectra and Scales" details how such dissonance curves are drawn.

[30] All partials were assumed of equal magnitude.

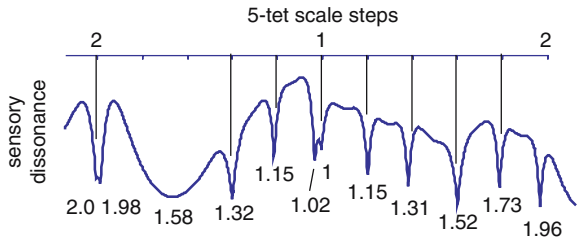


Fig. 10.10. Sounds F (a generic bonang) and G (a harmonic sound with four partials) generate a dissonance curve with many minima close to the steps of 5-tet, which is shown for comparison.

tunings. Thus, it is reasonable to interpret slendro tunings using the same principles as were used to derive the just scales as a basis of Western harmonic music. In fact, the deviation of slendro from 5-tet (and from the minima of the dissonance curve of Fig. 10.10) is smaller[31] than is the deviation of the just scale from 12-tet (and from the minima of the dissonance curve for harmonic sounds). In essence, the theory provides a better explanation for the slendro tunings than it does for Western tunings.

Besides the coincidence of the minima with scale steps, there are two notable features of this curve. First, there are three minima very close to the octave: at 1.96, 1.98, and 2.0. This variation in minima of the dissonance curve near the octave mirrors the variation in "octaves" of the slendro scales, and it may provide a hint as to why there is no single fixed octave in the slendro world. Second, observe the minimum at 1.02. With a fundamental of 100 Hz, this minimum would occur at 102 Hz, giving a beat rate of 2 per second. At a fundamental of 500 Hz, this minimum would occur at 510 Hz, with a beat rate of 10 Hz. This may be a hint as to the origin of the aesthetic of beats that the gamelan is famous for.

One objection to this analysis is that some arbitrary choices are made. For instance, why was G chosen to have four partials? Why not more? Why assume all partials are of equal importance (by assuming equal amplitudes)? Certainly, the particular values were chosen so that Fig. 10.10 was clear. Nonetheless, as in all dissonance curves, the fundamental features (in this case, the alignment of the minima with steps of the 5-tet scale) are relatively invariant to small changes in the assumptions. For instance, dropping a partial from G does not change any of the minima. Adding a partial to G causes another (extraneous) minimum to occur at 1.44. Deleting the partial at 3.92 from F causes the minima at 1.02 and 1.96 to disappear. Changing the amplitudes to more closely match the actual spectra only changes the height of the various minima, not their location. Indeed, the fundamental features are robust.

[31] Both in average and in maximum error.

10.5.2 Pelog

The pelog scale of one gamelan may differ substantially from the pelog of another. Thus, pelog is not as easily explained as slendro, which could be reasonably approximated by 5-tet.

One approach that appears fruitful is to combine the spectrum of the saron with a harmonic spectrum, in much the same way that slendro was approached as a combination of the bonang and a harmonic sound. To get a close match between the minima of the dissonance curve and the scale, however, it is not enough to use a saron averaged over all of the gamelans. Rather, the spectrum of the sarons actually used in the gamelan must be employed. For instance, a typical saron from gamelan Swastigitha was given in previous sections as $f,\ 2.76f,\ 4.72f,\ 5.92f$. Drawing the dissonance curve for this F along with a harmonic G containing five partials gives the dissonance curve of Fig. 10.11. Unlike the slendro scale, only half of this curve contains scale steps of the desired scale, so only this half is shown. Observe the close relation between the minima of the curve and the scale steps of the Swastigitha pelog scale.

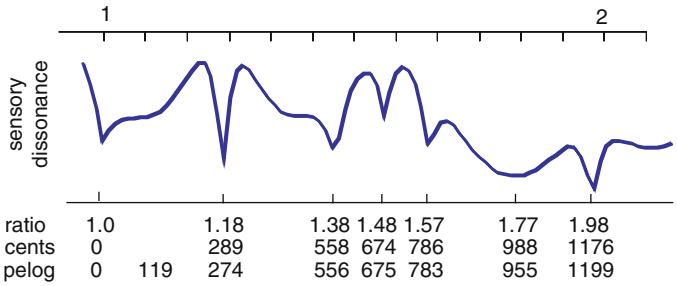


Fig. 10.11. Dissonance curve generated by the spectrum of the Swastigitha saron combined with a harmonic sound has minima near many of the scale steps of the Swastigitha pelog scale.

Although the first step of the scale is missing from the dissonance curve, the others are clearly present. Some of the scale steps are not aligned exactly, for instance, the second scale step is 289 cents on the curve but is averaged to 274 for the gamelan. Actual values over the three octaves of the gamelan are 245, 286, and 289, so the 289 is actually reasonable. The largest discrepancies occur in the last two steps. The next to last step is the only one that occurs on a broad minimum (the others all occur at the sharp, well-defined kind), and so it is not surprising that this value would have the largest variance. Indeed, the value of this step varies by more than 40 cents over the three octaves of the gamelan. The last step (near the octave) is understandable by the same mechanism as in the slendro scales. Looking over the whole curve (and not just this half), there are minima at 1.98, 2.0, and 2.14, and the three actual octaves

of the gamelan occur at 1.98, 1.99, and 2.01. Again, this may be interpreted in terms of the stretching and/or compressing of the octaves. Certainly, it is reasonable that the actual scales used should reflect the uncertainty of this placement of the "octave."

The sarons of Gamelan Kyai Kaduk Manis have somewhat more complex spectra, and the generic saron with partials at

$$f,\ 2.39f,\ 2.78f,\ 4.75f,\ 5.08f,\ 5.96f$$

can be combined with a sound G with five harmonics to give the dissonance curve of Fig. 10.12. This displays the same qualitative features as the previous figure: The first scale step is missing, and the seventh step (the octave) is not completely certain.[32] By a numerical coincidence, the next to last step is very close, but it again falls on a broad minimum and the exact value cannot be taken too seriously. Overall, however, the match between the minima of the dissonance curve and the measured values are good.

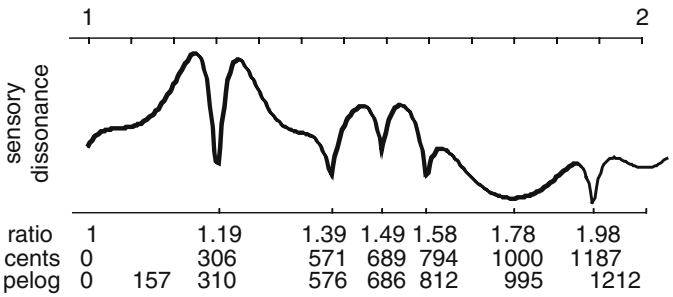


Fig. 10.12. Dissonance curve generated by the spectrum of the generic saron of Gamelan Kyai Kaduk Manis combined with a harmonic sound has minima near many of the scale steps of the Kyai Kaduk Manis pelog scale.

This does not imply that gamelan tuners actively listen to the partials of their instruments and sculpt them consciously so as to match the spectrum and the scale. Gamelan tuners view their task much differently; as a cycle of listening and filing that repeats until the gamelan "sounds right." Nonetheless, gamelan tuners like Suhirdjan, while listening holistically, do manipulate the partials as they tune. They do so in an intuitive way that is the result of a long period of apprenticeship, considerable skill in the techniques of tuning, and a deep insight into the way that gamelans "should" sound. Tuners like Purwardjito, by listening to the "supporting" tones as he tunes, may be listening and tuning more analytically. Purwardjito sees himself as tuning "from the heart." I believe him.

[32] There are again three "octaves" in the full curve. These occur at 1.98, 1.99, and 2.09.

10.6 Summary

A few general observations:

(i) In almost all cases, the lowest spectral peak is the largest. It is reasonable to call this lowest spectral peak the "fundamental," because it corresponds closely to its pitch.
(ii) The gamelan orchestras are "in-tune" with themselves in the sense that whenever two instruments occupy the same "note" of the scale, the fundamentals are rarely more than a few Hertz apart.
(iii) The relative amplitudes of the partials are heavily dependent on the angle, position, and force of the strike. The frequency of the partials is (comparatively) insensitive to the excitation.
(iv) The slendro tunings are very close to 5-tet, although the octave (or more properly, the pseudo-octave) of the scales are often slightly stretched or compressed from a perfect 2:1 octave.
(v) There are two classes of metallophones that are simple enough to understand: the bar-shaped instruments (saron and gender) and the kettle-shaped instruments (bonang and kenong). The acoustic behavior of the gongs, which is very complicated, is an area for further research.
(vi) The spectra of the bar-like instruments of the gamelan differ from the theoretically ideal bar. The differences are consistent enough to be considered purposeful.
(vii) The temporal evolution of the spectra of all bar-like instruments is simple... all partials decay. The higher partials decay faster.
(viii) There is no simple theoretical shape to which the spectrum of the kettle instruments can be compared. The partials of the keys are consistent across each gamelan.
(ix) The temporal evolution of the kettle spectra is more complex than that of the bar instruments. The cluster of high partials dies away quickly, whereas the partials near $1.5f$ grow (with respect to the fundamental) as time evolves, in many cases becoming the dominant (largest) partial and the most prominent part of the sound.

The method of dissonance curves can be used to correlate the spectra of instruments of the gamelan with the slendro and pelog scales in much the way that they can be used to correlate harmonic instruments with certain Western scales. The slendro scale can be viewed as a result of the spectrum of the bonang in combination with a harmonic sound, whereas the pelog scale can be (slightly less surely) viewed as resulting from a combination of the spectrum of the saron and a harmonic sound. Thus, gamelan scales exploit the unique features of the spectra of the inharmonic instruments of which they are composed, and yet retain a basic compatibility with harmonic sounds like the voice.

11

Consonance-Based Musical Analysis

The measurement of (sensory) consonance and dissonance is applied to the analysis of music using dissonance scores*. Comparisons with a traditional score-based analysis of a Scarlatti sonata show how the contour and variance of the dissonance score can be used to concretely describe the evolution of dissonance over time. Dissonance scores can also be applied in situations where no musical score exists, and two examples are given: a xenharmonic piece by Carlos and a Balinese gamelan performance. Another application, to historical musicology, attempts to reconstruct probable tunings used by Scarlatti from an analysis of his extant work.*

11.1 A Dissonance "Score"

There are many ways to analyze a piece of music. Approaches include the chord grammars and thematic processes of functional harmony as in Piston [B: 137], the harmonic and melodic tensions of Hindemith [B: 72], the harmonic and intervallic series of Schoenberg [B: 164], or in terms of the harmonic motion and structural hierarchy of Schenker [B: 163]. In most such musical analyses, the discussion of (functional) consonance and dissonance is based directly on the score, by an examination of the intervals, the harmonic context, and the tonal motion. This chapter introduces a way to explore the sensory consonance of a piece of music by calculating the performed dissonance at each time instant. The result is a graph called the *dissonance score* that shows how dissonance changes throughout the piece; the flow from consonance to dissonance (and back again) is directly displayed.

Consonance and dissonance are only one aspect of harmony, which is itself only one part of a complete analysis that must include melody and rhythm. Furthermore, sensory consonance and dissonance are not identical to the more traditional functional consonance and dissonance, and hence the dissonance score must be interpreted carefully. Nonetheless, the dissonance score is capable of conveying useful information that cannot be obtained in other ways. For instance, different performances of the same piece differ by virtue of the instruments used, idiosyncrasies of the musicians, and of the acoustic space in which the performance occurs. Dissonance scores reflect these differences and allow a comparison between various performances of the same piece. Dis-

sonance scores can also be drawn for music for which no musical score exists, and hence, they are applicable to a wider range of musics than those based on a formal score.

11.1.1 Drawing Dissonance Scores

Suppose that a musical piece has been recorded and digitized. The piece is partitioned into small segments, and the sensory dissonance of the sound in each segment is calculated by the techniques of the previous chapters. The dissonance score plots these values over time. Details are shown in Fig. 11.1.

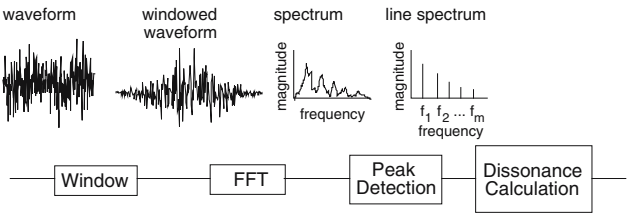


Fig. 11.1. Dissonance scores are calculated from a musical performance by windowing, applying an FFT, simplifying to a line spectrum, calculating the dissonance between all pairs of partials in the line spectrum, and then summing.

For example, one composer known for his innovative use of dissonance is the eighteenth century harpsichordist Domenico Scarlatti (1685–1757). Claude Roland-Manuel, in the liner notes to [D: 42], comments:

> Scarlatti's audaciously original harmonies, and his acciaccaturas–clusters and blocks of chords inherited from the Spanish guitar, taking dissonance almost to its ultimate limits...

Whether "ultimate" or not, there is no doubt that Scarlatti's sonatas were innovative in both their harmonic motion and their use of dissonance. They provide an interesting case study for the use of dissonance scores.

Figure 11.2 shows four versions of the dissonance score for the first half (40 measures) of Scarlatti's sonata[1] K380 in E major. In all cases, the horizontal axis represents time, which is indicated in measures by the numbers above the curves, whereas the vertical axis is the calculated[2] sensory dissonance. The top score was calculated from a standard MIDI file, assuming a single idealized harpsichord timbre for each note. Data for the other three performances were obtained by direct digital transfer from harpsichord performances on CD by [D: 30], [D: 37], and [D: 42] using the technique of Fig. 11.1.

[1] The prefix K stands for the harpsichordist Ralph Kirkpatrick, author of the standard catalog of Scarlatti's sonatas.

[2] In each curve, the point of maximum dissonance is normalized to unity.

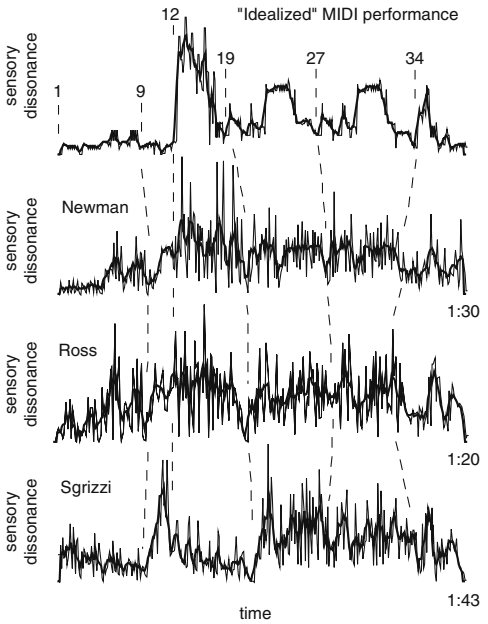


Fig. 11.2. Dissonance scores for several harpsichord performances of Scarlatti's sonata K380. Numbers indicate measures.

For the Scarlatti sonata, the data were partitioned into $L = 8K$ segments and the FFT of each segment was calculated. The most significant spectral frequencies (and their magnitudes) were then used to calculate the dissonance of each segment.[3] Each plotted point represents about 0.2 seconds, and the darker central lines are a moving average of the dissonance values over 10 points, or about 2 seconds. It is easy to plot the curves. But what do they mean?

11.1.2 Interpreting Dissonance Scores

To interpret the dissonance scores, it will help to correlate them with other, more traditional kinds of musical analysis. Figure 11.3 presents the musical score of the first 40 measures of Scarlatti's sonata K380. The piece begins with four repetitions (with slight variations in register and dynamics) of a I, V pattern, each ending in a trilled open fifth. These four repetitions appear in each of the dissonance scores as the first four little hills. In the idealized MIDI performance, the first pair of hillocks are identical and the second pair

[3] This simplification to the "most important" frequency components is not completely straightforward. An algorithm is discussed in Appendix C. Details of the calculations are given in Appendix E in equation E.6.

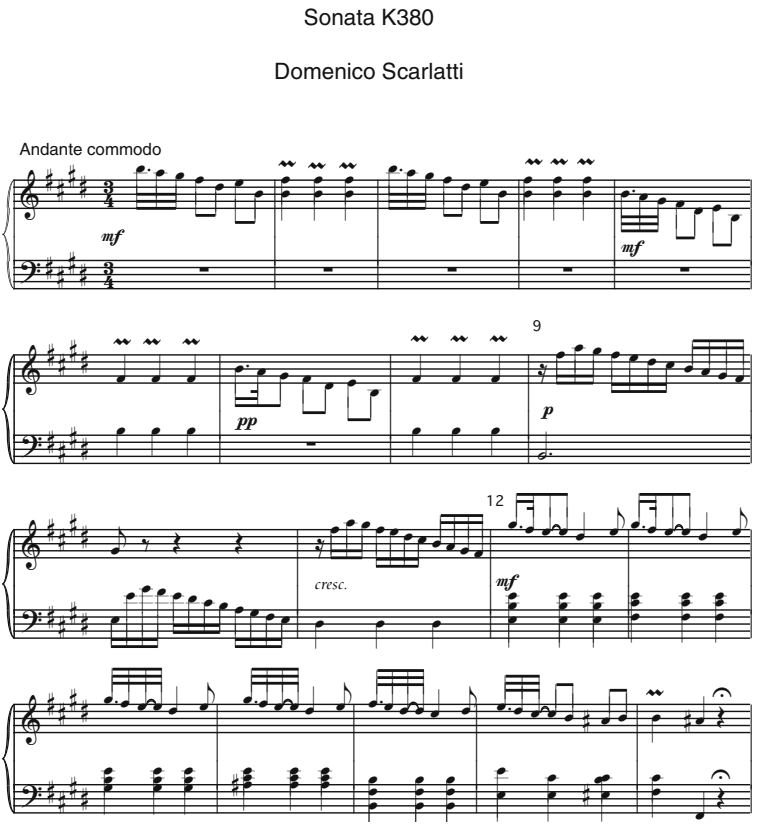


Fig. 11.3. Musical score to Scarlatti's sonata K380 (part one of two).

are identical, but larger. This reflects the fact that lower octaves have greater sensory dissonance than higher.[4] Measures 9 to 12 consist of descending runs that outline V, I, V. In the idealized performance, this is a short V-shaped segment, reflecting the fact that measures 9 and 11 contain bass notes, whereas the run is unaccompanied in the middle measure.

In measure 12, the melodic line begins the first of four repetitions. Underlying this repetitive figure is an E chord in measures 12 and 14, a $F\sharp$ dominant 7 in measure 13, and an $A\sharp$ diminished in measure 15. Although these may be mild compared with (say) passages from Stravinsky's *Rites of Spring*, they are considerably more dissonant than the previous sections. Besides the dissonance inherent in the bass clef chords, there is the $D\sharp$ neighboring tone in the melody, which forms a major seventh with the drone-like E. In addition, the $A\sharp$'s in the thirteenth and fifteenth measures form a repeated tritone.

[4] This is a direct result of the widening of the sine wave dissonance curves at lower frequencies.

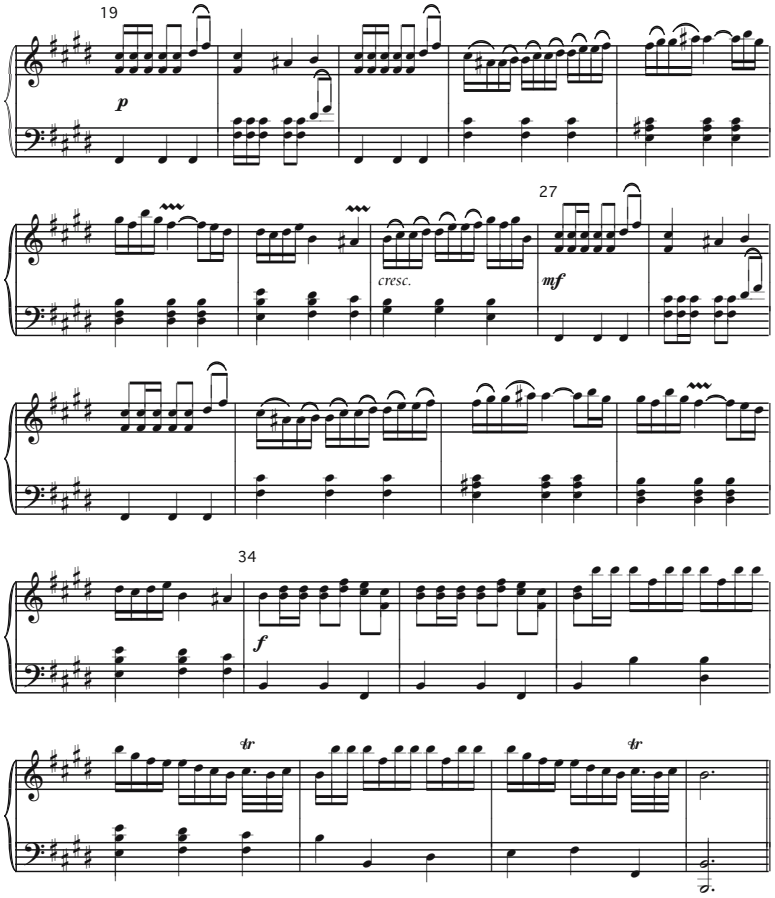


Fig. 11.3. Musical score to Scarlatti's sonata K380 (continued).

The dissonance of these four measures is clearly visible in the idealized MIDI performance as the large hump beginning at measure 12.

Scarlatti extricates himself from this dissonance by resolving from B major, through E major, and then to $F\sharp$, with a trilled suspension resolving down to the third. The melodic figure, which is transposed down twice, ties this to the previous four measures, and the journey into dissonance and back is completed by the end of measure 18. In the idealized performance, this return is apparent in the fluctuating low-level dissonances leading into measure 19.

Similarly, the remainder of the dissonance score can be interpreted in terms of the intervals, chords, and density of notes present in the original score. For instance, the two small bumps beginning at measure 19 are caused by the rhythmic "hunting horn" motif, whereas the large plateau starting at measure 23 is a result of the strong bass chords that again include an $A\sharp$ diminished.

When repeated at measures 27 and 31, the idealized dissonance score repeats almost exactly, just as in the musical score. When the first half of K380 ends in measure 40 by resolving to three octaves of B, the dissonance decreases toward zero. Thus, dissonance scores directly display some of the same qualitative information that can be interpreted indirectly from the musical score.

11.1.3 Comparing Dissonance Scores

Recall that sensory dissonance depends not only on the intervals, but also on the spectrum of the sound and its amplitude.[5] As dissonance scores can be drawn directly from a recorded performance, they can be used to compare different renditions of the same piece. For instance, where one performer might execute a phrase lightly, another might strike boldly. The brighter tone with more high harmonics will have greater dissonance, and it will appear differently on the dissonance score.

Figure 11.2 shows three different interpretations of the first half of K380 played by Newman, Ross, and Sgrizzi. Newman plays the "Magnum Opus Harpsichord" built by Hill and Tyre. At almost 11 feet, this lavishly illustrated three-manual instrument has five sets of strings and "may be the largest harpsichord ever constructed."[6] It has a full, lush sound. Ross plays the harpsichord of Anthony Sidey, which is a more traditional double-manual instrument. Sgrizzi plays the Neupert harpsichord at the Cathédrale San Lorenzo. Although the liner notes contain no information about the instrument, it clearly has at least two manuals, and the timbre of the two are different: One is bright, and the other is subdued and harp-like.

Performances of a piece can vary in many dimensions, including tempo, dynamics, tone color of the instrument, ornamentation, and properties of the recording environment such as reverberation, microphone placement, and equalization. These will all effect the dissonance score. For instance, a hall with large reverberation time (or equivalently, a long artificial reverberation added to the recording) will cause notes to sound longer. When sustained tones overlap, the dissonance increases because the spectra from all simultaneously sounding partials contributes to the dissonance calculation. Similarly, a faster rendition will tend to have more dissonance than a slower one, all else being equal, because successive notes overlap more. Although the dynamics of a harpsichord are relatively fixed (approximately the same force is applied each time a note is plucked), differences between instruments are significant, and differences between manuals and registers on the same instrument are inevitable. Thus, the performer has considerable control over nuances that effect the perceived dissonance of the rendition.

[5] Other factors being equal, a louder sound has greater sensory dissonance than a softer sound.

[6] According to the liner notes of [D: 30].

Dissonance scores display detailed information about the performance. For instance, the first eight measures appear as the first four bumps on the dissonance curves. Newman's version parallels the idealized MIDI performance; the first two bumps are both small, and the second two bumps are larger. Ross is similar, except that the fourth repetition is played with less dissonance than the third. The musical score marks dynamics for these phrases: *mf* for the first and third, *pp* for the second and fourth. Ross faithfully interprets these dynamic markings by reducing the dissonance.

In contrast, Sgrizzi decreases dissonance throughout the four phrases. The timbre of the instrument changes noticeably in the lower octave repetitions; presumably, Sgrizzi has changed manuals, and the effect is to decrease the dissonance despite the lowering of the octave. In measures 9 to 12, Sgrizzi returns to the brighter register. By playing these measures legato, the notes of the runs overlap, and these become the most dissonant passage in the piece.

One of the most obvious features of the dissonance scores is the rapid change in the instantaneous dissonances, which form a fuzzy halo about the averaged curve. These fluctuations can be quantified by calculating the sum squared deviation of the raw dissonance values from the averaged values. The standard deviations are:

Sgrizzi	0.124
Newman	0.133
Ross	0.155

In contrast to the human performances, the MIDI performance has very little fluctuation, with a standard deviation of only 0.063. This is because the MIDI dissonance score assumes an idealized harpsichord timbre containing exactly nine harmonic partials, an idealized instrument in which each note was identical except for transposition, and an idealized (quantized) performance.[7] Such a performance does not, of course, constitute an ideal performance, but it does provide a skeleton of the expected flow of consonance and dissonance throughout the piece.

Sgrizzi's low standard deviation is especially apparent in his careful handling of the dissonant chords in measures 12 through 19. Part of the low overall dissonance of this portion is likely due to the slow pace of the rendition, but the low variance also demonstrates a meticulous attention to the constancy of the musical flow. In contrast, Ross maintains both a high level of dissonance and a large variance throughout the phrase. This is due in part to the faster pace, but the high variance is caused by the rhythmic expression of the bass chords, which are played with deliberate attacks and an almost staccato articulation. The variance of Newman's performance is midway between Sgrizzi and Ross, but it is notable for its coherence. Observe how the third and fourth hills (measures 5–6 and 7–8) are almost exactly the same. Similarly, the "hunting horn" phrase in measures 19–27 is almost identical to

[7] The standard MIDI file is currently available on the Internet at the Classical Music Archives [W: 4].

the repeat in measures 27–34. Both Ross and Sgrizzi approach the two appearances of this motif differently. Ross builds tension by slowly incrementing the dissonance, whereas Sgrizzi slowly relaxes throughout the phrase.

Scarlatti's sonatas, although written for harpsichord, have often been adapted for piano, and many have been transcribed for classical guitar. Figure 11.4 shows the dissonance score for a performance of K380 on piano by [D: 33] and on guitar by [D: 14]. Pogorelich exploits the greater dynamic range of the piano to emphasize certain aspects of the piece. The first theme, for instance, follows Kirkpatrick's dynamic markings closely, and the dissonance follows the volume and the register. Pogorelich races through measures 12-19, but does so very softly. This controls the dissonance so that it peaks in the repeated hunting call of measures 19 and 27. This dissonance is due more to sheer volume than to the intervallic makeup of the chords. It is a sensible, although not inevitable, approach.

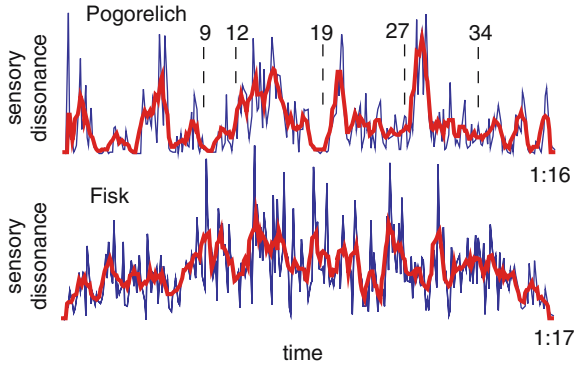


Fig. 11.4. Dissonance scores for two renditions of Scarlatti's sonata K380. Pogorelich performs on piano, and Fisk plays guitar.

Fisk's realization is almost as fast overall as Pogorelich's, but the tempo is more even. Where Pogorelich lingers in the first few measures and then charges through the next few, Fisk trods along with toe-tapping steadiness. Fisk's interpretation is unique among the performances because he treats the whole 40 measures as one long phrase. Observe how the dissonance score slowly rises and falls over the course of the piece, indicating this fluidity of motion. All other performances are segmented into (more or less) eight measure phrases, and the dissonance score rises and falls in synchrony. Although dissonance scores can give a quantitative assessment, they cannot pass judgment on the desirability of such interpretive decisions.

Dissonance scores must not be viewed carelessly. For instance, larger variance of the dissonance score might imply a more expressive performance, but it might also indicate a sloppiness of execution. Smaller variance points to

more careful control, perhaps more "technique," but it might also correspond to a more "mechanical" rendition. When comparing two dissonance scores of the same piece, the variation in dissonance due to the performance is more significant than the amount of dissonance, because both are normalized to unity. For instance, points of maximum or minimum dissonance might occur at different places, indicating those portions of the piece the performer wishes to emphasize or de-emphasize. Similarly, the contour of the dissonance curve carries much of the important information, but it requires an act of judgment to determine what contour is most desirable for a given piece.

Thus, dissonance scores can display unique information about a piece, and they may be used as an analytical tool to help concretely describe the motion from consonance to dissonance, and back again.

11.1.4 When There Is No Score

The dissonance score is not a notation, but a tool for analysis. Although it cannot supply as much information as a musical score, it is applicable in situations (to xenharmonic, aleatoric, serial, or ethnic musics, for example) where no scores exist and where traditional analytic techniques cannot be applied. To demonstrate the potential, this section briefly examines a short movement from Carlos' *Beauty in the Beast* and a segment from a Balinese gamelan performance. The dissonance scores are drawn, and they are related to various aspects of the music and the performances.

Beautiful Beasts

The title track of the symphonic *Beauty in the Beast* by Carlos [D: 5] is played in two xenharmonic scales. The *alpha* and *beta* scales are nonoctave-based tunings with equal steps of 78 and 63.8 cents, respectively. Although both scales can support recognizable triads, neither allows a standard diatonic scale, and neither repeats at the octave. Hence, it is not obvious how to apply standard analytical techniques, even if a score was available.

Figure 11.5 shows the dissonance score of the first 84 seconds of *Beauty in the Beast* along with the waveform, and an indication of how it might be divided into thematic sections. Section A is the "beast" motif, which is repeated with variations in A'. B is a soft transition section featuring wind chimes, which slowly builds into the "beauty" theme C. C' repeats the theme with melody, and in C'' the melody slowly fades into the background.

Both the beauty theme and the beast theme have an internal structure that is displayed by the dissonance score; each theme contains two dissonance bumps. In both A and A', the paired humps are roughly the same size. The bimodal structure of the beauty theme is less obvious because of the amplitude changes, which are apparent from the waveform. The long-term flow of the piece shows the characteristic motif of motion from consonance, through dissonance, and back again.

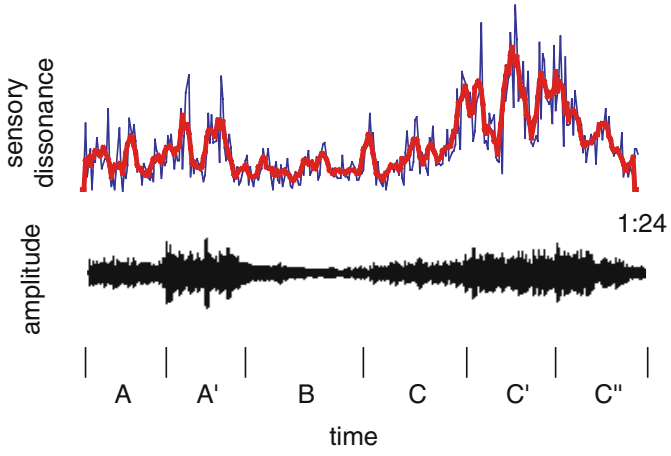


Fig. 11.5. First 1:24 of Carlos' *Beauty in the Beast* showing dissonance score, amplitude of waveform, and thematic structure.

The variance of the performance from its average is 0.955. Although this is smaller than any of Scarlatti performances (except for the idealized MIDI version), it would be rash to draw any conclusions from this. Perhaps the small variance is due to the synthesized nature of the work, which might lend precision to the performance. Perhaps it is due to the slower overall motion of the piece, or perhaps it is something inherent in the unusual tuning.

Gamelan Eka Cita

The gamelan, an "orchestra" of percussive instruments, is the primary indigenous musical tradition of Java and Bali. Music played by the gamelan is varied and complex, with styles that change over time and vary by place in much the way that styles wax and wane in the Western tradition. *Gong Kebyar*, which means "gong bursting forth," is a vibrant form of gamelan playing that began in Bali in the middle of the century, and it has flourished to become one of the dominant styles. Each year, the Bali State Arts Council sponsors the "All Bali Gong Kebyar Festival" in which gamelans from across the island compete. Eka Cita, an orchestra from the village of Abian Kapas Kaja near Denpassar, won the competition several years in a row, and a recording was made of their concert in [D: 18].

I. Wayan Beratha based *Bandrangan*, the second track on the CD, on the ritual spear dance *Baris Gede*. This energetic piece contains large contrasts in sound density, volume, and texture. The primary form of the piece consists of a short cycle, each beginning with a deep gong stroke, and each midpoint accented by a higher gong. The first 87 seconds (the complete piece is over 15 minutes) are displayed in Fig. 11.6, which shows the dissonance score and

the waveform. The cycles are marked by the grid at the bottom, and they are aligned with the primary gong hits. Many of the gong strokes are visible in the waveform, but they figure prominently throughout the segment even when they are not visible.

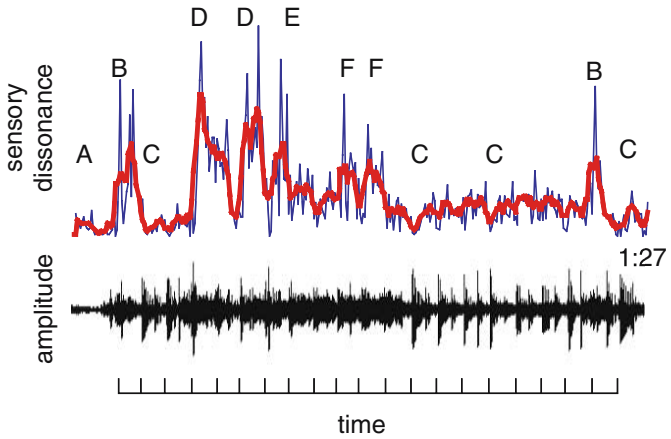


Fig. 11.6. First 1:27 of *Baris Gede Bandrangan* by I. Wayan Beratha, showing dissonance score, amplitude of waveform, and rhythmic structure.

Indonesia currently maintains a series of Institutes (called STSI[8] or ASTI) and Universities[9] that support and promote traditional culture, and they offer degrees in traditional music, dance, and painting, as well as courses in ethnomusicology and other "modernized" approaches to the study of the arts. Lacking immersion in the culture, it is difficult to analyze this (or any other) gamelan piece in more than a superficial manner. As with the analyses of Western music in the previous sections, the intention is to show how the technique of the dissonance score may be applied. Any conclusions drawn from this analysis must be considered tentative.

The first part of *Baris Gede Bandrangan*, shown in Fig. 11.6, can be thought of as containing several sections. *A* is a soft introduction that sets the pace. In *B*, the drummer (who is also the leader of the ensemble) crescendos, introducing the major "theme" in *C* along with the first gong strokes. These gong hits continue throughout the segment, delineating the cycles shown in the bottom grid. In *D*, a series of matching chords overlays the cycle, and this is repeated. In *E* and *F*, two different "melody" lines occur, again starting and stopping at cycle boundaries.

[8] Skola Tinggi Seni Indonesian.
[9] Such as Gadjah Mada University.

The dissonance score reflects some of these changes. Both dissonance peaks marked B are caused by the drum, which marks the beginning and/or end of a section. The peaks at D are a result of the raucous chording, and the smaller peaks E and F are produced by the rapid melodic motion of the higher pitched metallophones. Perhaps the most striking aspect of this score, at least in comparison with the Western pieces analyzed earlier, is that the dissonance peaks are episodic. That is, each cycle has a roughly constant dissonance, which changes abruptly at cycle boundaries.

In the pieces by Scarlatti and Carlos, the contour of the dissonance score delineates the major phrases as it slowly rises and falls. Apparently, in the gamelan tradition, (sensory) dissonance is used completely differently. Abrupt changes in dissonance are the norm, and these changes seem to reflect the entrance and exit of various instruments at cycle boundaries. If this pattern holds (for more than this single segment of a single composition), then this may be indicative of a fundamental difference in the musical aesthetic between the gamelan and Western traditions.

The standard deviation of the dissonance score of Fig. 11.6 about its mean (again, the average is drawn as the darker line) is 0.094. If this can be interpreted (as in the Western context) as a measure of the consistency of the performance, then this is a remarkable figure. It is considerably smaller than any of the Scarlatti performances, despite the fact that the gamelan is played by several musicians simultaneously.

11.2 Reconstruction of Historical Tunings

In 12-tet, there is no difference between various musical keys, there are no restrictions on modulation, and key tonality is not a significant structure in music. Three hundred years ago, the musical context was different. Until about 1780, keyboard instruments were tuned so that commonly used intervals were purer (closer to just) than less-used intervals. The resulting nonequal semitones gave a different harmonic color to each musical key, and these colors were part of the musical language of the time, both philosophically and practically. To understand the musical language of early keyboard composers, the tuning in which their music was conceived and heard is important.

However, few composers documented the exact tunings used in their music. Although there is sufficient historical evidence that the period and nationality of a composer can narrow the choice considerably, there are often significant variances between historically justifiable tunings for any specific piece of music. The tuning preferences of Domenico Scarlatti are particularly uncertain, because he was born and trained in Italy, but spent most of his career in Portugal and Spain, and did all of his significant composing while under strong Spanish influence. A method that might infer information concerning his tuning preferences solely from his surviving music would therefore be of value to musicians and musicologists.

This section discusses a quantitative method based on a measure of the sensory consonance and dissonance of the intervals in a tuning and their frequency of occurrence within the compositions. The presumption is that the composer would avoid passages using intervals that are markedly out-of-tune or dissonant (such as wolf fifths) except in passing, and would tend on average to emphasize those intervals and keys that are relatively pure. This investigation first appeared in an article co-authored with John Sankey called, "A consonance-based approach to the harpsichord tuning of Domenico Scarlatti" [B: 160], which finds tunings that minimize the dissonance over all intervals actually used by Scarlatti in his sonatas, and compares the results to several well-known historical tunings.

The method is equally applicable to other early keyboard composers. Barnes [B: 11] conducts a statistical analysis of the intervals that appeared in Bach's pieces to try and determine which tunings Bach was most likely to have used. This is similar in spirit to the present approach, but the optimization proceeds under a culture-dependent interval selection and classification scheme, rather than a psychoacoustic measure.

11.2.1 Total Dissonance

There are four basic steps to find the most consonant tuning for a piece (or collection of pieces) of music. These are:

(i) Specify the spectrum of each sound
(ii) Find (or count) the number of occurrences of each interval class, and weight by their duration
(iii) Choose an initial "guess" for the optimization algorithm
(iv) Implement a gradient descent (or other local optimization algorithm) to find the nearest "least dissonant" set of intervals

The bulk of this section describes these steps in detail.

As the Scarlatti sonatas were composed for harpsichord, a spectrum was chosen that approximates an idealized harpsichord string. The sound is assumed to contain 32 harmonic partials at frequencies

$$f,\ 2f,\ 3f, ...,\ 32f$$

where f is the fundamental. The amplitude of the partials is assumed to die away at a rate of $.75^n$, where n is the partial number. Surviving historical harpsichords vary considerably in these parameters. The low strings of some have more than 80 discernible partials, decreasing with an exponent as high as 0.9, whereas the high strings of others display as few as 8 partials with a more rapid decay. The amplitudes of the partials also vary due to the position at which the string is plucked (which may vary even on the same harpsichord), and from interactions among the strings. The chosen spectrum is a reasonable approximation to the average sound of a harpsichord in the portion of its range

in which a musician is most sensitive to questions of tuning. Three typical harpsichord timbres are shown in Fig. 11.7 for comparison.

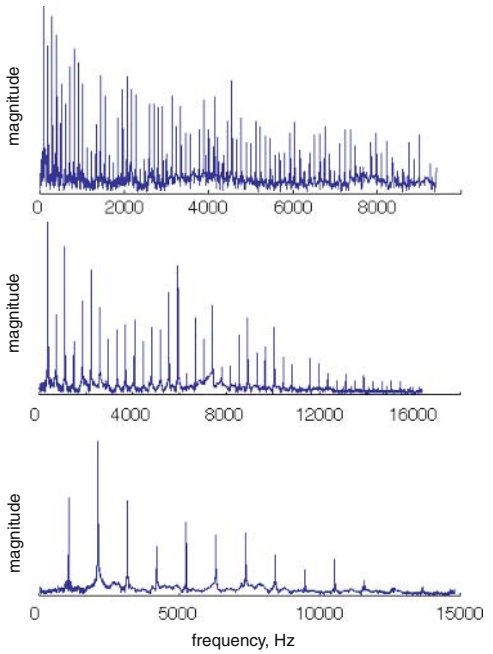


Fig. 11.7. Spectra of three notes of a harpsichord with fundamentals at 104 Hz, 370 Hz, and 1048 Hz (corresponding to notes $G\sharp$, $F\sharp$, and C). All partials lie close to a harmonic series, and the higher notes have fewer harmonics than the lower notes.

The sonatas of Scarlatti recordings have been encoded by John Sankey in Standard MIDI File (SMF) format,[10] which is a widely accepted standard for encoding the finger motions of a keyboard player as a function of time. These finger motions can be used to (re)synthesize the performance. A program was written to parse the SMF files and to collate the required information about frequency of occurrence of intervals and their duration in performance.

Recall that the sensory dissonance $D_F(f_j/f_i)$ between two notes with fundamentals f_i and f_j is the sum of all dissonances[11] between all pairs of sine wave partials. The Total Dissonance (TD) of a musical passage of m notes is defined to be the sum of the dissonances weighted by the duration over which the intervals overlap in time. Thus

$$TD = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} D_F(f_j/f_i)\ t(i,j)$$

where $t(i,j)$ is the total time during which notes i and j sound simultaneously. Although the amplitude of a single held note of a harpsichord decreases with

[10] The files are currently available on the Internet at [W: 4].
[11] See equation E.7 for details of the calculation of $D_F(f_j/f_i)$.

time, it increases significantly each time a succeeding note is played due to coupling via their shared soundboard. Given the high note rates in the sonatas, this rectangular sound intensity distribution is a reasonable approximation.

An n-note tuning based on the octave contains $n-1$ distinct intervals between 1:1 and 2:1. Observe that the TD for a musical composition depends on the tuning because the different intervals have different values of $D_F(f_j/f_i)$. By choosing the tuning properly, the total dissonance of the passage can be minimized, or equivalently, the consonance can be maximized. Thus, the problem of choosing the tuning that maximizes consonance can be stated as an optimization problem: Minimize the "cost" (the TD of the composition) by choice of the intervals that define the tuning. This optimization problem can be solved using a variety of techniques; perhaps the simplest is to use a gradient descent method. This is similar to the adaptive tuning method, but the TD maintains a history of the piece via the $t(i,j)$ terms. Adaptive tunings can be considered a special (instantaneous) case.

Let I_0 be the initial "tuning vector" containing a list of the intervals that define the tuning. A (locally) optimal I^* can be found by iterating

$$I_{k+1} = I_k - \mu \frac{dTD}{dI_k}$$

until convergence, where μ is a small positive stepsize and k is the iteration counter. The algorithm has "converged" when the change in each element of the update term has the same sign. Calculation is straightforward, although somewhat tedious. In most cases, the algorithm is initialized at the 12-tone equal-tempered scale; that is, I_0 is a vector in which all adjacent intervals are 100 cents.

A tuning for which a desired composition (or collection of compositions) has smaller TD is to be preferred as far as consonance is concerned. In the context of attempting to draw historical implications, the measure TD may provide reason for rejecting tunings (those that are overly dissonant) or reconsidering tunings (those with near-optimal values of TD). Such judgments cannot be made mechanically, they must be tempered with musical insight. The variation in values of the TD for different tunings is small, less than 1% between musically useful tunings, and are therefore expressed in parts per thousand ($^0/_{00}$) difference from 12-tet. A difference of $1^0/_{00}$ is clearly audible to a trained musical ear in typical musical contexts.[12]

Music of course does not consist solely of consonances. Baroque music is full of trills and similar features that involve overlapped seconds in real performance, and Scarlatti made heavy use of solidly overlapped seconds, deliberate dissonances, as a rhythmic device. Consequently, all intervals smaller than three semitones were omitted from the calculations of the TD. This had surprisingly little effect on the values of the convergent tunings; the precaution may be unnecessary with other composers.

[12] For this reason, a numerical precision of nine decimal places or greater is advisable for the calculations of TD.

11.2.2 Tunings for a Single Sonata

As harpsichords (in contrast to organs) were tuned frequently, usually by the performer, it is likely that composers might have changed their preferred tuning over the course of their lifetime, or used more than one tuning depending on the music to be played. Both of these are well documented in the case of Rameau. One way to investigate this is to initialize the tuning vector I_0 to the intervals of 12-tet, and find the optimum tunings I^* that minimize the TD for each sonata individually.

A histogram of all tunings obtained is shown in Fig. 11.8. The height of a bar shows the number of sonatas for which the optimum tuning contains a note of the given pitch. As can be seen, for most of the 11 pitches, there are two strong preferences. The location of the pure fifths[13] ascending and descending from C is shown below the frequency bars. The minimization process for samples as small as one sonata often "locks on" to the predominately nonunison minimum at pure fifths. This effect continues to dominate even when groups of up to ten sonatas are evaluated. Although baroque musicians often refined the tuning of their instruments before performing suites of pieces using a consistent tonality set, it is impractical to completely retune an instrument every 5 or 10 minutes, the length of a typical sonata pair with repeats and variations.

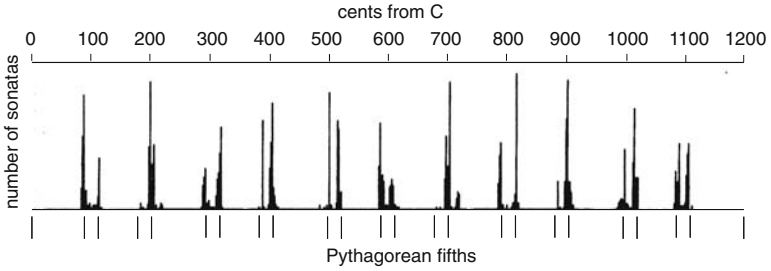


Fig. 11.8. The relative distribution of "optimal" tunings when considering each of the sonatas individually. Observe the clustering at the Pythagorean (pure) fifths.

The primary formal structure of most of the sonatas follows two symmetries: Tonalities are mirrored about a central double bar, and thematic material repeats after the double bar (although not always in exactly the same order). For example, K1 begins in D minor, progresses to A major at the double bar 14, and ends in D minor bar 31; thematically, bar 1 matches bar 14; 2–5, 22–25; 7, 17; 9, 18; 13, 31. One expects that Scarlatti's tuning(s) would have complemented and been consistent with these symmetries. Many of the single-sonata tunings found by this optimization method are not. For

[13] e.g., 702^n mod 1200 for $n = -11$ to 11.

example, bars 9 and 18 in K1 are symmetrically designed to strongly establish the tonalities D minor and F major, respectively, but the pure $D-A$ fifth on which bar 9 is based is inconsistent with the $F-C$ fifth of bar 18, a very audible 15 cents smaller than pure in this tuning. By comparison, these intervals differ by only 4 cents in the Vallotti A tuning. Using optimized tunings to retune sections of music of sonata length does not, therefore, seem to be a reliable guide to the practice of Scarlatti, nor to be useful in detecting changes in tuning preferences over his oeuvre.

11.2.3 Tunings for All Sonatas

When all of the sonatas are treated as a set, this kind of overspecialization to particular intervals does not occur, but there are a large number of minima of the TD within a musically useful range.

One tuning obtained while minimizing from 12-tet (labeled TDE in Table 11.1) has several interesting features. Many theorists, in the past and still today, consider the numerical structure of a scale to be important, often favoring just scales that consist of the simplest possible number of ratios. The 12-tet-refined tuning is one of this class: Take four notes $a = 1$, $b = 9/8$, $c = 4/3$, and $d = 3/2$. Then $d = 4b/3$ and $d = 9c/b$, so every note is just with respect to all others. Three such groups overlap to make a 12-note scale $C-D-F-G$, $E-F\sharp-A-B$, $A\flat-B\flat-D\flat-E\flat$. The tuning TDE found to be optimal for the sonatas contains two of these quartets. However, unlike many just tunings, this one is specially designed for use with an extended body of music, namely, the sonatas. There is no historical evidence that any influential performer or composer actually used such a tuning, but it is worth listening to by anyone wishing to hear the sonatas in a different but musical way. The technique of minimizing TD is a fertile source of new tunings for modern keyboard composers—there are many musically interesting tunings that have not been explored.

Table 11.1. Derived tunings. All values rounded to the nearest cent.

Label	cents										
TDE	98	200	302	402	506	605	698	800	900	1004	1104
TDA1	86	193	291	386	498	585	697	786	889	995	1087
TDA2	88	200	294	386	498	586	698	790	884	996	1084

The relative[14] TD of a number of tunings that are documented in the musical literature of Scarlatti's time are shown in Table 11.2. The tunings are defined in Table L.1 of Appendix L. Meantone tuning, in which all fifths are equal except one wolf fifth $G\sharp-E\flat$, was the most common tuning at the

[14] All TD values are normalized so that the TD of 12-tet is zero.

close of the Middle Ages. It was considered to be in the key of D, and it was modified steadily toward equal temperament by increasing the size of the equal fifths as time progressed. However, as only one note needs to be retuned to transpose any meantone tuning into the tuning for an adjacent fifth (e.g., to add or subtract one sharp or flat from the key signature), many performers did so to improve the sound of their favorite keys.

Table 11.2. Total Dissonance TD (in $^0/_{00}$ deviation from 12-tet) and strength s of various historical and derived tunings over all Scarlatti's sonatas.

Tuning	TD	s
12-tet	0	0
Bethisy	-0.4	4.1
Rameau b	-0.5	7.1
Werkmeister 5	-0.6	2.6
d'Alembert	-0.8	4.1
Barca	-1.0	2.4
Werkmeister 3	-1.9	3.1
Kirnberger 3	-1.9	3.4
Corrette	-2.2	6.8
Vallotti A	-2.5	2.9
Chaumont	-3.3	7.7
Rameau ♯	-4.0	7.1
1/4 Comma A	-5.8	10.3
Kirnberger 2	-6.0	4.5
TDE	-1.6	2.2
TDA1	-2.3	4.6
TDA2	-7.1	5.6

The TD for the set of all Scarlatti sonatas is shown in Fig. 11.9 versus the size of the equal fifths and the position of the wolf fifth. There is a sharp maximum with fifths 3.42 cents less than 12-tet when the wolf is between $E\flat$ and $B\flat$ or between $E\flat$ and $G\sharp$, precisely the medieval 1/4-comma tunings in the keys of A and D. There is another broader maximum with fifths 1.8 cents larger than 12-tet, which is close to the ancient Pythagorean tuning with pure fifths. The general shape of the meantone TD of the entire keyboard oeuvre of Scarlatti is, therefore, in accord with historical musical practice.

Many historical harpsichord tunings have been quantified by Asselin [B: 8]; the tunings used in this study are shown in Table L.1 of Appendix L. As the harpsichord scale has 11 degrees of freedom, it is desirable to characterize each tuning by a smaller number of musically useful parameters. The mean absolute difference between the various tunings and 12-tet gives a kind of "strength" parameter. Define

$$s(t) = \text{ mean } |c(k,e) - c(k,t)|$$

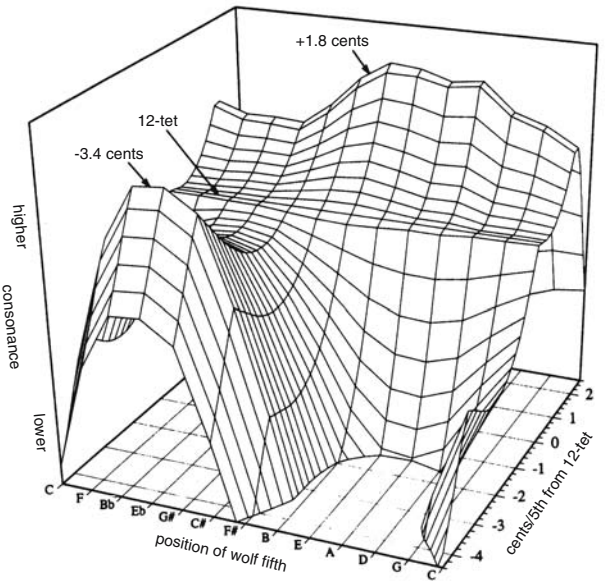


Fig. 11.9. The relation between consonance, size of equal fifths in a meantone-type tuning, and position of the unequal wolf fifth, for all sonatas as one unit.

where $c(k, e)$ is the pitch in cents of note k from the first note of the 12-tet scale, $c(k, t)$ the corresponding pitch of tuning t, and the function c has been normalized so that

$$\text{mean } c(k, e) = \text{ mean } c(k, t)$$

to remove the pitch scale dependence of the dissonance function. Historically, the value of $s(t)$ has decreased with time, from 10 cents for the medieval 1/4-comma meantone tuning to essentially zero for modern piano tunings. In general, a low value of s is associated with tunings that work in a wide variety of keys, a high value with tunings placing many restrictions on modulation.

Figure 11.10 plots the TD of each tuning (in $^0/_{00}$ of the TD of 12-tet) versus the strength of the tuning. If a series of meantone-type tunings in A is constructed, with the size of the equal fifths decreasing from 12-tet (100 cents) to 96 cents, the locus of TD and s is the solid line shown. (It is the same curve as that for the wolf between $E\flat$ and $B\flat$ in Fig. 11.9.) In Fig. 11.10, a decrease of both the TD and s represents an improvement in both consonance and in modulatability. A decrease in the TD associated with an increase in s requires a choice based on musical context, because any improvement in consonance will be offset by a reduction in the range of keys in which the consonance will occur.

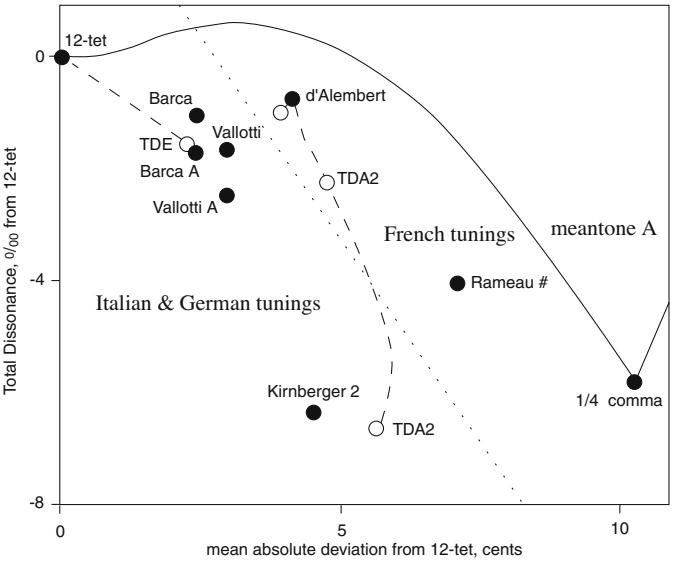


Fig. 11.10. The vertical axis plots the TD of all sonatas when played in the tunings of Table 11.2 as a percentage of the TD of all sonatas when played in 12-tet. The horizontal axis gives the mean absolute deviation of each tuning from the 12-tet scale.

In general, French tunings sought to purify the sound of major thirds, whereas Italian and German tunings were more closely derived from the fifth-based meantone. The two schools may be separated by the dotted line in Fig. 11.10; again, the TD is in accordance with historical knowledge. Both Italian tunings in A show superior consonance to those in D, and Rameau's "sharp" tuning has greater consonance than that in $B\flat$. (Modulated versions of any tuning have the same strength s.) The expectation from this figure is that Kirnberger 2 should be by far the best tuning for the sonatas, with meantone (1/4 comma) second except perhaps in some remote tonalities due to its strength. Next should be the sharp tuning of Rameau (again with possible difficulties in some tonalities), followed by Vallotti A, and then Barca A. Unfortunately, other factors intervene.

A primary phrase pattern widely used in Western music, and particularly by Scarlatti in the sonatas, is a gradual increase of musical tension culminating in a musical steady state (stasis) or a release of tension (resolution). Increasing pitch, volume, rapidity, harmonic density, and harmonic dissonance are techniques of increasing musical tension. A skilled composer will use these various techniques in a mutually supporting way, in consistent patterns. If, therefore, use of a particular tuning enhances the ebb and flow of musical tension, it may be the tuning that the composer used to hear music. As such a small proportion of potential intervals can be simultaneously in perfect tune

in one tuning, it is likely that an erroneous tuning at least occasionally results in a glaring mismatch of musical shape.

The TD predictions fail with the second tuning of Kirnberger when this tension structure is taken into account—the consonances in this tuning often fall in Scarlatti's relatively long tonal transition passages and all too frequently come to abrupt halts with unacceptably dissonant stases. For example, sonata K1 begins the second section with an A major triad ascent to an E in the treble, and then repeats the figure in the bass under the sustained E. With Kirnberger 2, $A - E$ is almost 11 cents smaller than just, one of the most dissonant fifths in the tuning. In both the Vallotti A and d'Alembert tunings, by comparison, $A - E$ is a bit less than 1/4 comma smaller than just, precisely right for an interim pause in the overall-upward passage of which the A to E phrase forms a part. Besides frequently failing the tension-topology criterion and the symmetry criteria discussed earlier, the 1/4-comma meantone tuning too often produces phrases that stay consistently out of tune for too long at a time (although obviously not long enough to affect the TD sufficiently), for example, the chromatic passages in bars 10–14 and 35–38 of K3. In fact, these bars together with their symmetric pair 58–63 and 84–87 cannot be played in consistent tune with any placement of a 1/4-comma-tuning wolf fifth.

However, although the tonal colors of Rameau are clearly in evidence, so are the consonances, which fall in the right places, and the tuning is particularly evocative in many of Scarlatti's slow plaintive melodic passages (K11, for example). The smooth matches of the Vallotti A tonal structure with those implicit in the music are very consistent, if unremarkable. The French tunings do indeed mostly have problems with dissonances in many places (the chromatic passages of K3, for example).

The historical instructions for some tunings are uncertain, even deliberately ambiguous, so modern numeric reconstructions may be slightly in error. This is almost certainly the case for the tuning of d'Alembert, which was described and redescribed in remarkably varied terms by several authors (e.g., Bethisy) of the time. The gradient algorithm was again applied to successively reduce the TD in small steps for the set of all sonatas, beginning with d'Alembert's tuning (instead of initializing with 12-tet), with the hope that this might correct minor errors in what is basically a good tuning. Two routes the algorithm took are shown by dashed lines in Fig. 11.10. The longer (right) curve shows the route when the only criterion for the change in I was lower TD. The shorter curve emanating from d'Alembert's tuning resulted when I was optimized for lower s and lower TD simultaneously. The first minimization proceeded well beyond the optimum musical point along the path, ending up at a tuning (TDA2 in Fig. 11.10) that made the most common intervals perfectly consonant but far too many lesser used musically important ones unacceptably dissonant (for example, the repeated high $D - A$ fifths of K1, 17 cents flat).

Furthermore, if this optimization from the d'Alembert tuning is applied individually to the few sonatas where the TDA1 tuning has residual difficul-

ties, a similar behavior is observed. At first, the sound improves, and then, with further iteration, the tuning becomes "overspecialized." For example, the fifths ending many phrases of K328, and the chords closing each half, are a bit more discordant with TDA1 than one would wish, although consistently so. Applying the refinement procedure for this sonata alone produces the tuning included in Table 11.1—the fifths and chords all improve in consonance compared with TDA1, without changing the sound of the rest of the sonata adversely or changing the basic color of the tuning. This is in accordance with historical practice, where a basic tuning would be "touched up" for a while to play a group of pieces that benefited from it (as opposed to the minimum-TD tunings that varied too much between sonatas to be practical).

11.3 What's Wrong with This Picture?

The music hall is austere—it is exactly the kind of place a Scarlatti or a Rameau might have played. The harpsichord is an immaculate reproduction made by the finest craftsmen from a historically authenticated model. The performer is well versed in the ornamentation and playing techniques of the period and is perhaps even costumed in clothes of the time. The music begins—*in 12-tet.*

What's wrong with this picture is the sound. 12-tet was not used regularly in Western music until well into the eighteenth century, and yet even performers who strive for authentic renditions often ignore this.[15] Perhaps this is excusable for Scarlatti, whose tuning preferences are uncertain, but no such excuse is possible for Rameau, whose treatise [B: 145] is one of the major theoretical works of his century. Imagine taking a serial piece by Schoenberg or Babbit, and "purifying" it for play in a major scale. Is the damage to Scarlatti's vision any less?

Although firm conclusions about tunings actually used by Scarlatti await his resurrection, the total dissonance of a large volume of music is a useful tool for studies of 12-tone keyboard tunings in a historical context, although it is insufficient by itself. Use of total dissonance to optimize a 12-tone tuning for a historical body of music can produce musically valuable results, but it must be tempered with musical judgment, in particular to prevent overspecialization of the intervals.

This chapter has shown how to apply the idea of sensory dissonance to musical analysis. For instance, there are many possible tunings in which a given piece of music might be performed. By drawing dissonance scores for different tunings (12-tet, just, meantone, adaptive, and so on), their impact can be investigated, at least in terms of the expected motion of dissonance.

[15] Few recordings of Scarlatti's sonatas are performed in nonequal tunings. There are dozens in 12-tet, many played on beautiful period harpsichords and boasting authentic-sounding blurbs on the cover.

Dissonance scores might also be useful as a measure of the "distance" between various performances. For instance, the area between the averaged curves of two renditions provides an objective criterion by which to say that two performances are or are not similar. One subtlety is that the dissonance scores must be aligned (probably by a kind of resampling) so that measures and even beats of one performance are coincident with corresponding measures and beats in the other. Most likely, this alignment must be done by hand because it is not obvious how to automatically align two performances when they differ in tempo.

12

From Tuning to Spectrum

The related scale for a given spectrum is found by drawing the dissonance curve and locating the minima. The complementary problem of finding a spectrum for a given scale is not as simple, because there is no single "best" spectrum for a given scale. But it is often possible to find "locally best" spectra, which can be specified as the solution to a certain constrained optimization problem. For some kinds of scales, such as ***n****-tet, properties of the dissonance curves can be exploited to directly solve the problem. A general "symbolic method" for constructing related spectra works well for scales built from a small number of successive intervals.*

12.1 Looking for Spectra

Given a tuning, what spectra are most consonant? Whether composing in n-tet, in some historical or ethnic scale, or in some arbitrarily specified scale, related spectra are important because they provide the composer and/or performer additional flexibility in terms of controlling the consonance and dissonance of a given piece.

For example, the Pythagorean tuning is sometimes criticized because its major third is sharp compared with the equal-tempered third, which is sharper than the just third. This excessive sharpness is heard as a roughness or beating, and it is especially noticeable in slow, sustained passages. Using a related spectrum that is specifically crafted for use in the Pythagorean tuning, however, can ameliorate much of this roughness. The composer or performer thus has the option of exploiting a smoother, more consonant third than is available when using unrelated spectra.

12.2 Spectrum Selection as an Optimization Problem

Any set of m scale tones specifies a set of $m-1$ intervals (ratios) $r_1, r_2, ..., r_{m-1}$. The naive approach to the problem of spectrum selection is to choose a set of n partials $f_1, f_2, ..., f_n$ and amplitudes $a_1, a_2, ..., a_n$ to minimize the sum of the dissonances over all $m - 1$ intervals. Unfortunately, this can lead to "trivial" timbres in two ways. Zero dissonance occurs when all amplitudes are

zero, and dissonance can always be minimized by choosing the r_i arbitrarily large. To avoid such trivial solutions, some constraints are needed.

Recall that the dissonance between two tones is defined as the sum of the dissonances between all pairs of partials, weighted by the product of their amplitudes. (Now would be an excellent time to review the section *Drawing Dissonance Curves* on p. 99 in the chapter *Relating Spectrum and Scale* if this seems hazy.) If any amplitude is zero, then that partial contributes nothing to the dissonance; if all amplitudes are zero, there is no dissonance. Thus, one answer to the naive minimization problem is that the dissonance can be minimized over all the desired scale steps by choosing to play silence—a waveform with zero amplitude! The simplest way to avoid this problem is to forbid the amplitudes a_i to change.[1]

Constraint 1: Fix the amplitudes of the partials.

A somewhat more subtle way that the naive minimization problem can fail to provide a sensible solution is a consequence of the second property of dissonance curves (see p. 121), which says that for sufficiently large intervals, dissonance decreases as the interval increases. Imagine a spectrum in which all partials separate more and more widely, sliding off toward infinity. Such infinitely sparse spectra minimize the dissonance at any desired set of scale steps and give a second "trivial" solution to the minimization problem. The simplest way to avoid this escape to infinity is to constrain the frequencies of all partials to lie in some finite range. The cost will then be reduced by spreading the partials throughout the set, while trying to keep it especially low at the scale steps r_i.

Constraint 2: Force all frequencies to lie in a predetermined region.

Fixing the amplitudes and constraining the frequencies of the partials are enough to avoid trivial solutions, but they are still not enough to provide good solutions. Although the resulting scale steps do tend to have reasonably small dissonance values, they often do not fall at minima of the dissonance curves. Consider an alternative "cost" that counts how many minima occur at scale steps. Minimizing this alternative cost alone would not be an appropriate criterion because it only reacts to the existence of minima and not to their actual value. But combining this with the original (constrained) cost encourages a large number of minima to occur at scale steps and forces these minima to have low dissonance.

The final revised and constrained optimization problem is as follows: With the amplitudes fixed, select a set of n partials $f_1, f_2, ..., f_n$ lying in the region

[1] Although not appealing, such a condition is virtually necessary. For instance, suppose the a_i for $i = 1, ..., n-1$ were fixed while a_n was allowed to vary. Then the cost could always be reduced by choosing $a_n = 0$. An alternative might be to fix the sum of the a_i, say, $\sum a_i = a^*$. Again, the cost could be reduced by setting $a_j = v^*$ and $a_i = 0$ for all $i \neq j$.

of interest so as to minimize the cost

$$C = w_1 \begin{pmatrix} \text{sum of dissonances} \\ \text{of the } m-1 \text{ intervals} \end{pmatrix} + w_2 \begin{pmatrix} \text{number of minima} \\ \text{at scale steps} \end{pmatrix}$$

where the w_1 and w_2 are weighting factors. Minimizing this cost tends to place the scale steps at local minima as well as to minimize the value of the dissonance curve. Numerical experiments suggest that weightings for which the ratio of w_1 to w_2 is about a 100 to 1 give reasonable answers.

12.3 Spectra for Equal Temperaments

For certain scales, such as the m-tone equal-tempered scales, properties of the dissonance curve can be exploited to quickly and easily sculpt spectra for a desired scale, thus bypassing the need to solve this complicated optimization problem.

Recall that the ratio between successive scale steps in 12-tet is the twelfth root of 2, $\sqrt[12]{2}$, or about 1.0595. Similarly, m-tet has a ratio of $s = \sqrt[m]{2}$ between successive scale steps. Consider spectra for which successive partials are ratios of powers of s. Each partial of such a sound, when transposed into the same octave as the fundamental, lies on a note of the scale. Such a spectrum is *induced* by the m-tone equal-tempered scale.

Induced spectra are good candidate solutions to the optimization problem. Recall from the principle of coinciding partials[2] that minima of the dissonance curve tend to be located at intervals r for which $f_i = rf_j$, where f_i and f_j are partials of the spectrum of F. As the ratio between any pair of partials in an induced spectrum is s^k for some integer k, the dissonance curve will tend to have minima at such ratios: these ratios occur precisely at steps of the scale. Thus, such spectra will have low dissonance at scale steps, and many of the scale steps will be minima: Both terms in the cost function are small, and so the cost is small.

This insight can be exploited in two ways. First, it can be used to reduce the search space of the optimization routine. Instead of searching over all frequencies in a bounded region, the search need only be done over induced spectra. More straightforwardly, the spectrum selection problem for equal-tempered scales can be solved by careful choice of induced spectra.

12.3.1 10-Tone Equal Temperament

As an example, consider the problem of designing sounds to be played in 10-tone equal temperament. 10-tet is often considered one of the worst temperaments for harmonic music, because the steps of the 10-tone scale are significantly different from the (small) integer ratios, implying that harmonic

[2] The fourth property of dissonance curves from p. 123.

tones are very dissonant. These intervals will become more consonant if played with specially designed spectra. Here are three spectra related to the 10-tet scale

$$f,\ s^{10}f,\ s^{17}f,\ s^{20}f,\ s^{25}f,\ s^{28}f,\ s^{30}f,$$
$$f,\ s^{7}f,\ s^{16}f,\ s^{21}f,\ s^{24}f,\ s^{28}f,\ s^{37}f, \text{and}$$
$$f,\ s^{10}f,\ s^{16}f,\ s^{20}f,\ s^{23}f,\ s^{26}f,\ s^{28}f,\ s^{30}f,\ s^{32}f,\ s^{35}f,\ s^{36}f,$$

where $s = \sqrt[10]{2}$. As expected, all three sound reasonably consonant when played in the 10-tet scale, and very dissonant when played in standard 12-tet. But each has its own idiosyncrasies.

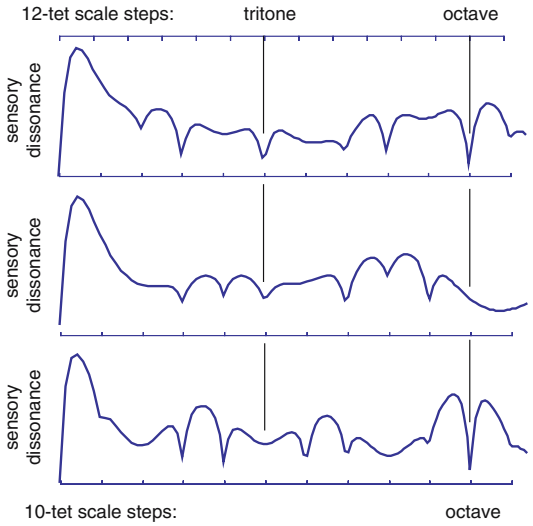


Fig. 12.1. Dissonance curves for spectra designed to be played in the 10-tone equal-tempered scale. Minima of the curves coincide with steps of the 10-tet scale and not with steps of 12-tet.

The dissonance curves of all three spectra are shown in Fig. 12.1, assuming the amplitude of the ith partial is 0.9^i. Observe that only the fifth scale step in 10-tet closely corresponds to any scale step in 12-tet; it is identical to the 12-tet tritone.[3] In all three spectra, the dissonance curve exhibits a minimum at the tritone, but only the top curve has a deep minimum there. This is caused by interaction of the partials at $s^{20}f$, $s^{25}f$, and $s^{30}f$, which differ by a tritone.

[3] This is because $(\sqrt[10]{2})^5 = (\sqrt[12]{2})^6$. In fact, the tritone is a feature of every octave based tuning with an even number of scale steps, because $(\sqrt[2r]{2})^r = \sqrt{2}$ for any r.

The dissonance curve for the middle spectrum has no minimum at the octave. This might be predicted by looking at the partials, because none of the pairs in this spectrum are separated by a factor of $s^{10} = 2$. On the other hand, both the top and bottom spectra have partials at $s^{10}f$, $s^{20}f$, and $s^{30}f$, which helps the octave retain its familiar status as the most consonant interval other than the unison. The middle spectrum would be less suitable for octave-based music than the others.

The top spectrum was chosen so that intervals 2, 3, 5, 7, 8, and 10 appear as ratios of the partials

$$\frac{s^{30}}{s^{28}} = s^2,\ \frac{s^{28}}{s^{25}} = s^3,\ \frac{s^{25}}{s^{20}} = s^5,\ \frac{s^{17}}{s^{10}} = s^7,\ \frac{s^{28}}{s^{20}} = s^8,$$

and several pairs differ by s^{10}. Consequently, these appear as minima of the dissonance curve and hence define the related scale. Similarly, when specifying the partials for the bottom spectrum, all 10 possible differences were included. Consequently, almost all scale steps occur at minima of the dissonance curve, except for the first scale step, which is formed by the ratio of the partials at s^{36} and s^{35}. This exception may occur because the interval s is close to one-half of the critical band,[4] or it may be because the amplitudes of the last two partials are significantly smaller than the others, and hence have less effect on the final dissonance.

Thus the three spectra have different sets of minima, and different related scales, although all are subsets of the 10-tet steps. Each spectrum has its own "music theory," its own scales and chords. Each sound plays somewhat differently, with the most consonant intervals unique to the sound: scale steps 3, 5, 8, and 10 for the top spectrum, but 3, 5, 7, and 9 for the middle. Moreover, keeping in mind that scale steps tend to have minima when the partials are specified so that their ratio is a scale step, it is fairly easy to specify induced spectra for equal temperaments, and to sculpt the spectra and scales toward a desired goal. Much of this discussion can be summarized by the observation that dissonance curves for induced spectra often have minima at scale steps. When the ratio of the partials is equal to a scale step, a partial from the lower tone coincides with a partial from the upper tone, causing the dip in the dissonance curve.

Of course, far more important than how the dissonance curves look is the musical question of how the resulting spectra and scales sound. The piece *Ten Fingers* on track [S: 102] of the accompanying CD uses the third 10-tet spectrum, and it exploits a number of possible chords. The particular tone quality used is much like a guitar, and the creation of such instrumental tones is discussed in the "Spectral Mappings" chapter. A possible "music theory" for such 10-tet sounds is presented in Chap. 14.

Observe that this sound has no problems with fusion as heard earlier with the 2.1 stretched (and certain other) spectra. Indeed, isolated notes of the

[4] Over a large range of fundamental f, s^{36} and s^{35} lie in the region where the critical band is a bit larger than a 12-tet whole step. See Fig. 3.4 on p. 44.

spectrum do not sound particularly unusual, despite their inharmonic nature. This is because the difference between the partials of this spectrum and the partials of a harmonic tone are not large. Looking closely at the locations of the partials shows that each one is as close as possible to an integer. In essence, it is as close to harmonic as a 10-tet-induced spectrum can be. Concretely:

$$s^{10} = 2,\ s^{16} \approx 3,\ s^{20} = 4,\ s^{23} \approx 5,\ s^{26} \approx 6,$$
$$s^{28} \approx 7,\ s^{30} = 8,\ s^{32} \approx 9,\ s^{35} \approx 10,\ \text{and}\ s^{36} \approx 11.$$

The overall effect is of music from another culture (or perhaps another planet). The chord patterns are clearly unusual, and yet they are smooth. The xentonal motion of the piece is unmistakable—there is chordal movement, resolution, and tensions, but it is not the familiar tonal language of Western (or any other) extant music.

How important is the sculpting of the spectrum? Perhaps just any old sound will be playable in 10-tet with such striking effect. To hear that it really does make a difference, track [S: 103] demonstrates the first few bars of *Ten Fingers* when played with a standard harmonic tone. When *Ten Fingers* is played with the related spectrum, many people are somewhat puzzled by the curious xentonalities. Most are decidedly uncomfortable listening to *Ten Fingers* played with a harmonic spectrum. The difference between tracks [S: 102] and [S: 103] is not subtle. The qualitative effect is similar to the familiar sensation of being out-of-tune. But the tuning is a digitally exact ten equal divisions of the octave, and so the effect might better be described as *out-of-spectrum*.

12.3.2 12-Tone Equal Temperament

Recall that most musical instruments based on strings and tubes are harmonic; their partials are closely approximated by the integer ratios of the harmonic series. Such spectra are related to the just intonation scale, and yet are typically played (in the West, anyway) in 12-tet. Although this is now considered normal, there was considerable controversy surrounding the introduction of 12-tet, especially because the thirds are so impure.[5] In terms of the present discussion, advocates of JI wish to play harmonic sounds in the appropriate related scale. An alternative is to design spectra especially for play in 12-tet.

As the above example moved the partials from their harmonic series to an induced 10-tet spectrum, the consonance of 12-tet can be increased by moving the partials away from the harmonic series to a series based on $s = \sqrt[12]{2}$. For instance, the set of partials

$$f,\ s^{12}f,\ s^{19}f,\ s^{24}f,\ s^{28}f,\ s^{31}f,\ s^{34}f,\ s^{36}f,\ s^{38}f$$

[5] For a discussion of this controversy, see [B: 198] or [B: 78]. This controversy has recently been revived now that the technical means for realizing JI pieces in multiple keys is available [B: 43].

is "almost" harmonic, but each of the integer partials has been quantized to its nearest 12-tet scale location. The effect on the dissonance curve is easy to see. Figure 12.2 compares the dissonance curve for a harmonic tone with nine partials to the 12-tet induced spectrum above (the amplitudes were the same in both cases). The dissonance curve for the induced spectrum has the same general contour as the harmonic dissonance curve but with two striking differences. First, the minima have all shifted from the just ratios to steps of the 12-tet scale: Minima occur at steps two through ten. Second, many of the minima are deeper and more clearly defined.

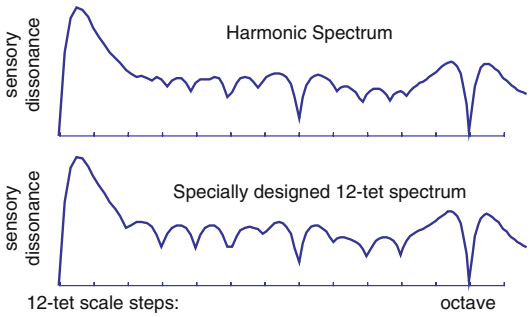


Fig. 12.2. Comparison of dissonance curve for harmonic spectrum with dissonance curve for spectrum with specially designed "12-tet" partials. Both spectra have nine partials, with amplitudes decreasing at the same exponential rate.

Thus, an alternative to playing in a just intonation scale using harmonic tones is to manipulate the spectra of the sounds so as to increase their consonance in 12-tet. To state this as an imprecise analogy: 12-tet with induced sounds is to 12-tet with harmonic sounds as just intonation with harmonic sounds is to 12-tet with harmonic sounds. Both approaches eliminate the disparity between 12-tet and harmonic tones, one by changing to the related scale, and the other by changing to related spectra.

Some electronic organs (the Hammond organ) produce induced 12-tet spectra using a kind of additive synthesis. Sound begins in 12 high-frequency oscillators. A circuit called a "frequency divider" transposes these 12 frequencies down by octaves, and these are combined as partials of the final sound. In effect, this quantizes the frequencies of the partials to steps of the 12-tet scale. Such organs are the first electronic example of matching spectrum and scale using induced timbres.

12.4 Solving the Optimization Problem

Minimizing the cost C of p. 247 is a n-dimensional optimization problem with a highly complex error surface. Fortunately, such problems can often be solved

adequately (although not necessarily optimally) using a variety of "random search" methods such as "simulated annealing" [B: 87] or the "genetic algorithm" [B: 65]. After briefly reviewing the general method, a technique for reducing the search space is suggested.

12.4.1 Random Search

In the simplest kind of "global optimization" algorithm, a spectrum is guessed, and its cost is evaluated. If the new cost is the best so far, then the spectrum is saved. New guesses are made until the optimum is found, or until some predetermined number of iterations has passed. Although this can work well for small n, it is inefficient when searching for complex spectra with many partials. For such high-dimensional problems, even the fastest computers may not be able to search through all possibilities. The algorithm can be improved by biasing new guesses toward those that have previously shown improvements.

12.4.2 Genetic Algorithm

The genetic algorithm (GA) is modeled after theories of biological evolution, and it often works reasonably well for the spectrum selection problem. Goldberg [B: 65] gives a general discussion of the algorithm and its many uses. The GA requires that the problem be coded in a finite string called the "gene" and that a "fitness" function be defined. Genes for the spectrum selection problem are formed by concatenating binary representations of the f_i. The fitness function of the gene $f_1, f_2, ..., f_n$ is measured as the value of the cost, and spectra are judged "more fit" if the cost is lower. The GA searches n-dimensional space measuring the fitness of spectra. The most fit are combined (via a "mating" procedure) into "child spectra" for the next generation. As generations pass, the algorithm tends to converge, and the most fit spectrum is a good candidate for the minimizer of the cost. Indeed, the GA tends to return spectra that are well matched to the desired scale in the sense that scale steps tend to occur at minima of the dissonance curve, and the total dissonance at scale steps is low. For example, when the 12-tet scale is specified, the GA often converges near induced spectra. This is a good indication that the algorithm is functioning and that the free parameters have been chosen sensibly.

12.4.3 An Arbitrary Scale

As an example of the application of the genetic algorithm to the spectrum selection problem, a desired scale was chosen with scale steps at 1, 1.1875, 1.3125, 1.5, 1.8125, and 2. A set of amplitudes was chosen as 10, 8.8, 7.7, 6.8, 5.9, 5.2, 4.6, 4.0, and the GA was allowed to search for the most fit spectrum. The frequencies were coded as 8-bit binary numbers with 4 bits for the integer

part and 4 bits for the fractional part. The best three spectra out of ten trial runs of the algorithm were

$$f,\ 1.8f,\ 4.9f,\ 14f,\ 9.87f,\ 14.81f,\ 6.4f,\ 12.9f,$$
$$f,\ 1.5f,\ 3.3f,\ 10.3f,\ 7.8f,\ 7.09f,\ 3.52f,\ 3.87f,\ \text{and}$$
$$f,\ 2.39f,\ 9.9275f,\ 7.56f,\ 11.4f,\ 4.99f,\ 6.37f,\ 10.6f.$$

The dissonance curve of the best spectrum is shown in Fig. 12.3. Clearly, these spectra are closely related to the specified scale, because minima occur at many of the scale steps. The cost function applies no penalty when there are extra minima, and each curve has a few minima more than were specified.

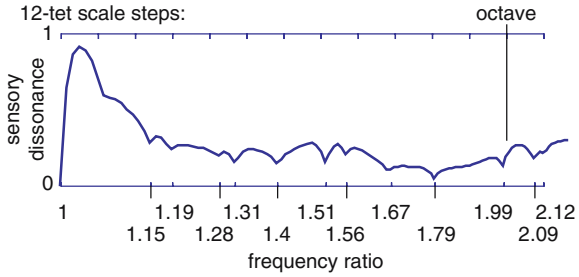


Fig. 12.3. Dissonance curve for the third spectrum has minima that align with many of the specified scale steps. The extra minima occur because no penalty (cost) is applied.

12.4.4 Reducing the Search Space

The algorithms suggested above conduct a structured random search for partials over all frequencies in the region of interest, and they calculate the dissonance of the intervals for each candidate spectrum. One way to simplify the search is to exploit the principle of coinciding partials (property four of dissonance curves[6] by restricting the search space to spectra containing intervals equal to the scale steps. For equal temperaments, this was as simple as choosing partial locations at scale steps, but in general, it is necessary to consider all possible intervals formed by all partials.

Let the candidate spectrum F have n partials at frequencies $f_1, f_2, ..., f_n$ with fixed amplitudes. Since scale steps can occur at any of the ratios of the f_i, let $r_{1,i} = \frac{f_{i+1}}{f_i}$ be all the ratios between successive partials, $r_{2,i} = \frac{f_{i+2}}{f_i}$ be the ratios between partials twice removed, and $r_{j,i} = \frac{f_{i+j}}{f_i}$ be the general terms. Any of the $r_{j,i}$ may become minima of the dissonance curve, and the

[6] Recall the discussion on p. 123.

problem reduces to choosing the f_i so that as many of the $r_{j,i}$ as possible lie on scale steps.

The inverse problem is more interesting. Given a scale S with desired steps $s_1, s_2, ..., s_m$, select an $r_{j,i}$ to be equal to each of the s_k. Solve backward to find the candidate partial f_i giving such $r_{j,i}$. The cost C of this spectrum can then be evaluated and used in the optimization algorithm. The advantage of this approach is that it greatly reduces the space over which the algorithm searches. Rather than searching over all real frequencies in a region, it searches only over the possible ways that the $r_{j,i}$ can equal the s_k.

To see how this might work in a simple case, suppose that a spectrum with $n = 5$ partials is desired for a scale with $m = 3$ steps. The set of all possible intervals formed by the partials $f_1, f_2, ..., f_5$ is:

$$\begin{array}{llll} r_{1,2} = \frac{f_2}{f_1} & r_{2,3} = \frac{f_3}{f_2} & r_{3,4} = \frac{f_4}{f_3} & r_{4,5} = \frac{f_5}{f_4} \\ r_{1,3} = \frac{f_3}{f_1} & r_{2,4} = \frac{f_4}{f_2} & r_{3,5} = \frac{f_5}{f_3} & \\ r_{1,4} = \frac{f_4}{f_1} & r_{2,5} = \frac{f_5}{f_2} & & \\ r_{1,5} = \frac{f_5}{f_1} & & & \end{array}$$

The desired scale steps are $(1, s_1, s_2, s_3)$. To choose a possible spectrum, pick one of the $r_{i,j}$ from each column, and set it equal to one of the s_k. For instance, one choice is

$$r_{1,4} = s_1,\ r_{2,4} = s_2,\ r_{3,5} = s_3,\ \text{and}\ r_{4,5} = s_2,$$

which leads to the following set of equations:

$$s_1 = \frac{f_4}{f_1},\ s_2 = \frac{f_4}{f_2},\ s_3 = \frac{f_5}{f_3},\ \text{and}\ s_2 = \frac{f_5}{f_4}$$

These can be readily solved for the unknowns f_i in terms of the known values of s_k. For this example, setting the first partial equal to some unspecified fundamental f gives

$$f_2 = \frac{s_1}{s_2} f,\ f_3 = \frac{s_1 s_2}{s_3} f,\ f_4 = s_1 f,\ \text{and}\ f_5 = s_1 s_2 f.$$

Assuming that the scale is to be octave based (i.e., that $s_3 = 2$), then the actual frequencies of the partials may be moved freely among the octaves. The cost of this spectrum is then evaluated, and the optimization proceeds as before.

12.5 Spectra for Tetrachords

The problem of finding spectra for a specified scale has been stated in terms of a constrained optimization problem that can sometimes be solved via iterative techniques. Although these approaches are very general, the problem is high

dimensional (on the order of the number of partials in the desired spectrum), the algorithms run slowly (overnight, or worse), and they are not guaranteed to find optimal solutions (except "asymptotically"). Moreover, even when a good spectrum is found for a given scale, the techniques give no insight into the solution of other closely related spectrum selection problems. There must be a better way.

This section exploits the principle of coinciding partials to transform the problem into algebraic form. A symbolic system is introduced along with a method of constructing related spectra. Several examples are given in detail, and related spectra are found for a Pythagorean scale and for a diatonic tetrachordal scale. A simple pair of examples then shows that it is not always possible to find such related spectra. The symbolic system is further investigated in Appendix I, where several mathematical properties are revealed.

Earlier in this chapter, the principle of coinciding partials was used to straightforwardly find spectra for 10-tet. Other equal temperaments are equally straightforward. To see why spectrum selection is more difficult for nonequal tunings, consider the Pythagorean diatonic scale, which was shown in Fig. 4.2 on p. 53 mapped to the "key" of C. Recall that this scale is created from a series of just 3/2 fifths (translated back into the original octave whenever necessary), and all seven of the fifths in the diatonic scale (the white keys) are just. An interesting structural feature is that there are only two successive intervals, a "whole step" of $a = 9/8$ and a "half step" of $b = 256/243$. This whole step is 4 cents larger than the equal-tempered version, whereas the half step is 10 cents smaller than in 12-tet.

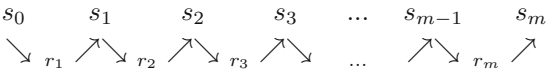
In attempting to mimic the "induced spectrum" idea of the previous sections, it is natural to attempt to place the partials at scale steps. Unfortunately, the intervals between scale steps are not necessarily scale steps. For instance, if one partial occurred at the seventh ($f_i = 243/128$) and the other at the third ($f_j = 4/3$), then a minimum of the dissonance curve might occur at $r = f_i/f_j = a^3 = 729/512$, which is not a scale step. Similarly, the ratio between a partial at 4/3 and another at 81/64 is $256/243 = b$, which again is not a scale step. Almost any nonequal scale has similar problems.

12.5.1 A Symbolic System

This section presents a symbolic system that uses the desired scale to define an operation that generates "strings" representing spectra, i.e., sets of partials. Admissible strings have all ratios between all partials equal to some interval in the scale, and thus they are likely to be related spectra, via the property of coinciding partials.

Basic Definitions

A desired scale S can be specified either in terms of a set of intervals $s_0, s_1, s_2, ..., s_m$ with respect to some fundamental frequency f or by the successive ratios $r_i = s_i/s_{i-1}$.



For instance, for the Pythagorean major scale of Fig. 4.2 of p. 53,

$$S = 1, 9/8, 81/64, 4/3, 3/2, 27/16, 243/128, 2/1,$$

and r_i is either $a = 9/8$ or $b = 256/243$ for all i. The intervals s_i in S are called the *scale intervals*.

A spectrum F is defined by a set of partials with frequencies at $f_1, f_2, ..., f_n$. The property of coinciding partials suggests that related spectra can be constructed by ensuring that the ratios of the partials are equal to scale steps. The following definitions distinguish the situation where all ratios of all partials are equal to some scale step, from the situation where all scale steps occur as a ratio of some pair of partials.

Complementarity: If for each i and j there is a k such that $\frac{f_i}{f_j} = s_k$, then the spectrum is called *complementary* to the scale.

Completeness: If for each k there is at least one pair of i and j such that $s_k = \frac{f_i}{f_j}$, then the spectrum is called *complete* with respect to the scale.

If a spectrum is both complete and complementary, then it is called *perfect* with respect to the given scale. Of course, scales and spectra need not be perfect to sound good or to be playable, and many scales have no perfect spectra at all. Nonetheless, when perfect spectra exist, they are ideal candidates.

An Example

The simplest nonequal scales are those with only a small number of different successive ratios. For example, one scale generated by two intervals a and b has scale intervals

$$s_0 = 1,\ s_1 = a,\ s_2 = ab,\ s_3 = a^2b,\ s_4 = a^2b^2,$$
$$s_5 = a^3b^2, \text{ and } s_6 = a^3b^3 = 2,$$

where a and b are any two numbers such that $a^3b^3 = 2$. Call this the ab-cubed scale. For the ab-cubed scale,

$$r_1 = a,\ r_2 = b,\ r_3 = a,\ r_4 = b,\ r_5 = a, \text{ and } r_6 = b.$$

To see how it might be possible to build a perfect spectrum for this scale, suppose that the first partial is selected arbitrarily at f_1. Then f_2 must be

$$af_1,\ abf_1,\ a^2bf_1,\ a^2b^2f_1,\ a^3b^2f_1, \text{ or } 2f_1$$

because any other interval will cause $\frac{f_2}{f_1}$ to be outside the scale intervals. Suppose, for instance, that $f_2 = a^2bf_1$ is selected. Then f_3 must be chosen so that $\frac{f_3}{f_1}$ and $\frac{f_3}{f_2}$ are both scale intervals. The former condition implies that f_3 must be one of the intervals above, whereas the latter restricts f_3 even further. For instance, $f_3 = a^3b^2f_1$ is possible because $\frac{a^3b^2f_1}{a^2bf_1} = ab$ is one of the scale intervals. But $f_3 = a^3b^3f_1$ is not possible because $\frac{a^3b^3f_1}{a^2bf_1} = ab^2$ is not one of the scale intervals. Clearly, building complementary spectra for nonequal scales requires more care than in the equal-tempered case where partials can always be chosen to be scale steps. For some scales, no complementary spectra may exist. For some, no complete spectra may exist.

Symbolic Computation of Spectra

This process of building spectra rapidly becomes complex. A symbolic table called the $\oplus$-table (pronounced "oh-plus") simplifies and organizes the choices of possible partials at each step. The easiest way to introduce this is to continue with the example of the previous section.

Let the scalar intervals in the ab-cubed scale be written $(1,0)$, $(1,1)$, $(2,1)$, $(2,2)$, $(3,2)$, and $(3,3)$, where the first number is the exponent of a and the second is the exponent of b. As the scale is generated by a repeating pattern, i.e., it is assumed to repeat at each octave, $(3,3)$ is equated with $(0,0)$. Basing the scale on the octave is not necessary, but it simplifies the discussion. The $\oplus$-table 12.1 represents the relationships between all scale intervals. The table shows, for instance, that the interval a^2b combined with the interval ab gives the scale interval a^3b^2, which is notated $(2,1) \oplus (1,1) = (3,2)$.

Table 12.1. $\oplus$-table for the ab-cubed scale.

$\oplus$	(0,0)	(1,0)	(1,1)	(2,1)	(2,2)	(3,2)
(0,0)	(0,0)	(1,0)	(1,1)	(2,1)	(2,2)	(3,2)
(1,0)	(1,0)	*	(2,1)	*	(3,2)	*
(1,1)	(1,1)	(2,1)	(2,2)	(3,2)	(0,0)	(1,0)
(2,1)	(2,1)	*	(3,2)	*	(1,0)	*
(2,2)	(2,2)	(3,2)	(0,0)	(1,0)	(1,1)	(2,1)
(3,2)	(3,2)	*	(1,0)	*	(2,1)	*

The $*$ indicates that the given product is not permissible because it would result in intervals that are not scalar intervals. Thus, $a^2b = (2,1)$ cannot be $\oplus$-added to $a = (1,0)$ because together they form the interval a^3b, which is not an interval of the scale. Observe that the "octave" has been exploited whenever the product is greater than 2. For instance, $(1,1) \oplus (3,2) = (4,3)$. When reduced back into the octave, $(4,3)$ becomes $(1,0)$ as indicated in the table, expressing the fact that $\frac{a^4b^3}{a^3b^3} = a^1b^0$. At first glance, this set of intervals

and the $\oplus$ operator may appear to be some kind of algebraic structure such as a group or a monad [B: 93]. However, common algebraic structures require that the operation be closed, that is, that any two elements (intervals) in the set can be combined using the operator to give another element (interval) in the set. The presence of the $*$'s indicates that $\oplus$ is not a closed operator.

Construction of Spectra

The $\oplus$-table 12.1 was constructed from the scale steps given by the *ab*-cubed scale; other scales S define analogous tables. This section shows how to use such $\oplus$-tables to construct spectra related to a given scale.

Let S be a set of scale intervals with unit of repetition or "octave" s^*. Let $T = [S,\ s^*+S,\ 2s^*+S,\ 3s^*+S, ...]$ be a concatenation of S and all its octaves. (The symbol "+" is used here in the normal sense of vector addition). Each element of s in S represents an equivalence class $s+ns^*$ of elements in T. Said another way, S does not distinguish steps that are one or more "octaves" s^* apart.

Example: For the *ab*-cubed scale,

$$S = [(0,0), (1,0), (1,1), (2,1), (2,2), (3,2)]$$

with octave $s^* = (3,3)$. Then

$$s^* + S = [(3,3), (4,3), (4,4), (5,4), (5,5), (6,5)],$$

$$2s^* + S = [(6,6), (7,6), (7,7), (8,7), (8,8), (9,8)],$$

and so on, and T is a concatenation of these.

The procedure for constructing spectra can now be stated.

Symbolic Spectrum Construction

(i) Choose t_1 in T, and let s_1 in S be the corresponding representative of its equivalence class.

(ii) For $i = 2, 3, ...,$ choose t_i in T with corresponding s_i in S so that there are $r_{i,i-j}$ with

$$s_i = s_j \oplus r_{i,i-j}$$

for $j = 1, 2, ..., i-1$.

The equation in the second step is called the $\oplus$-equation. The result of the procedure is a string of t_i, which defines a set of partials. By construction, the spectrum built from these partials is complementary to the given scale. If, in addition, all of the scale steps appear among either the s or the r, then the spectrum is complete and, hence, perfect.

The $\oplus$-equation expresses the desire to have all of the intervals between all of the partials $\frac{f_i}{f_j}$ be scale intervals. A set of s_j are given (which are defined by previous choices of the t_j). Solving this requires finding a single s_i such that the $\oplus$-equation holds for all j up to $i-1$. This can be done by searching all s_j columns of the $\oplus$-table for an element s_i in common. If found, then the corresponding value of $r_{i,i-j}$ is given in the leftmost column. Whether this step is solvable for a particular i, j pair depends on the structure of the table and on the particular choices already made for previous s_i. Solution techniques for the $\oplus$-equation are discussed at length in Appendix I.

It is probably easiest to understand the procedure by working through an example. One spectrum related to the the *ab*-cubed scale is given in Table 12.2. This shows the choice of t_i, the corresponding scale steps s_i (which are the t_i reduced back into the octave), and the $r_{i,k}$ that complete the $\oplus$-equation. As all of the s_i and $r_{i,k}$ are scale steps, this spectrum is complementary. As all scale steps can be found among the s_i or $r_{i,k}$, the spectrum is complete. Hence the spectrum of Table 12.2 is perfect for this scale. To translate the table into frequencies for the partials, recall that the elements t_i express the powers of a and b times an unspecified fundamental f. Thus, the first partial is $f_1 = a^3b^3f$, the second is $f_2 = a^5b^5f$, and so on.

Table 12.2. A spectrum perfect for the *ab*-cubed scale.

i	1	2	3	4	5	6	7	k
t_i	(3,3)	(5,5)	(6,6)	(9,8)	(10,9)	(11,10)	(13,12)	
s_i	(0,0)	(2,2)	(0,0)	(3,2)	(1,0)	(2,1)	(1,0)	
$r_{i,k}$		(2,2)	(1,1)	(3,2)	(1,1)	(1,1)	(2,2)	1
			(0,0)	(1,0)	(1,0)	(2,2)	(0,0)	2
				(3,2)	(2,1)	(2,1)	(1,1)	3
					(1,0)	(3,2)	(1,0)	4
						(2,1)	(2,1)	5
							(1,0)	6

12.5.2 Perfect Pythagorean Spectra

The Pythagorean major scale of Fig. 4.2 on p. 53 is constructed from two intervals a and b in the order a, a, b, a, a, a, b. Thus, the scale steps are given by:

1	a	a^2	a^2b	a^3b	a^4b	a^5b	$a^5b^2 = 2$
$(0,0)$	$(1,0)$	$(2,0)$	$(2,1)$	$(3,1)$	$(4,1)$	$(5,1)$	$(5,2) = (0,0)$

Typically, a^2b is a pure fourth. Along with the condition that $a^5b^2 = 2$, this uniquely specifies $a = 9/8$ and $b = 256/243$, and so the scale contains two

equal tetrachords separated by the standard interval 9/8. These exact values are not necessary for the construction of the perfect spectra that follow, but they are probably the most common. The $\oplus$-table for this Pythagorean scale is shown in Table 12.3. It is not even necessary that $(5, 2)$ be an exact octave; any pseudo-octave or interval of repetition will do.

Table 12.3. $\oplus$-table for the Pythagorean scale.

$\bigoplus$	(0,0)	(1,0)	(2,0)	(2,1)	(3,1)	(4,1)	(5,1)
(0,0)	(0,0)	(1,0)	(2,0)	(2,1)	(3,1)	(4,1)	(5,1)
(1,0)	(1,0)	(2,0)	*	(3,1)	(4,1)	(5,1)	*
(2,0)	(2,0)	*	*	(4,1)	(5,1)	*	*
(2,1)	(2,1)	(3,1)	(4,1)	*	(0,0)	(1,0)	(2,0)
(3,1)	(3,1)	(4,1)	(5,1)	(0,0)	(1,0)	(2,0)	*
(4,1)	(4,1)	(5,1)	*	(1,0)	(2,0)	*	*
(5,1)	(5,1)	*	*	(2,0)	*	*	*

Table 12.4. A spectrum perfect for the Pythagorean scale.

i	1	2	3	4	5	6	7	k
t_i	(5,2)	(8,3)	(10,4)	(12,4)	(14,5)	(15,5)	(17,6)	
s_i	(0,0)	(3,1)	(0,0)	(2,0)	(4,1)	(5,1)	(2,0)	
$r_{i,k}$		(3,1)	(2,1)	(2,0)	(2,1)	(1,0)	(2,1)	1
			(0,0)	(4,1)	(4,1)	(3,1)	(3,1)	2
				(2,0)	(1,0)	(5,1)	(0,0)	3
					(4,1)	(2,0)	(2,0)	4
						(5,1)	(4,1)	5
							(2,0)	6

Spectra can be assembled by following the procedure for symbolic spectrum construction, and one such spectrum is given in Table 12.4. Observe that all of the s_i and $r_{i,k}$ are scale steps, and that all seven scale steps are present among the s_i and the $r_{i,k}$. Hence, this spectrum is perfect for the Pythagorean scale. Assuming the standard values for a and b, this spectrum has its partials at

$$f,\ 2f,\ 3f,\ 4f,\ \frac{81}{16}f,\ \frac{27}{4}f,\ \frac{243}{32}f,\ \text{and}\ \frac{81}{8}f.$$

The first several partials are harmonic, and this is the "closest" perfect Pythagorean spectrum to harmonicity. For example, there are no suitable partials between $(12, 4) \approx 5$ and $(14, 5) = 6.75$ and thus no way to closely approximate the sixth harmonic partial $6f$. It is easy to check that $(13, 4)$ and $(14, 4)$ are not scale steps, and that $(13, 5) = (3, 1)$ forms the interval ab

with $(12, 4)$. As ab is not a scale step, $(13, 5)$ cannot occur in a complementary spectrum.[7]

The dissonance curve for this Pythagorean spectrum is shown in Fig. 12.4, under the assumption that the amplitude of the ith partial is 0.9^i. As expected from the principle of coinciding partials, this curve has minima that align with the scale steps. Thus, there are significant minima at the just fourth and fifths, and at the Pythagorean third 81/64 and the Pythagorean sixth 27/16, rather than at the just thirds and sixths as in the harmonic dissonance curve. This spectrum will not exhibit rough beating when its thirds or sixths are played in long sustained passages in the Pythagorean tuning. There are also two extra minimum that are shallow and broad. These are not due to coinciding partials. The exact location and depth of these minima changes significantly as the amplitude of the partials are changed. As is usual for such extra minima, they are only barely distinguishable from the surrounding regions of the curve. Thus, perfect spectra, as constructed by the symbolic procedure, do give dissonance curves with minima that correspond closely with scale steps of the desired scale.

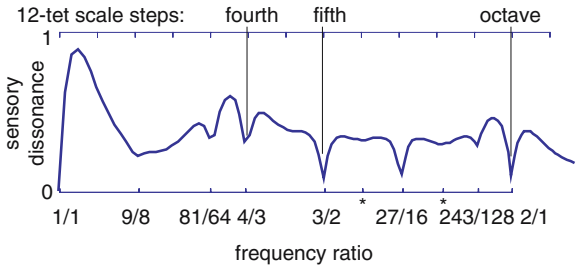


Fig. 12.4. Dissonance curve for the spectrum specially designed for play in the Pythagorean diatonic scale has minima at all of the specified scale steps. Two extra "broad" minima marked by stars are not caused by coinciding partials.

12.5.3 Spectrum for a Diatonic Tetrachord

A more general diatonic tetrachordal scale is constructed from three intervals a, b, and c in the order a, a, b, c, a, a, b. The scale steps are:

1	a	ab	a^2b	a^2bc	a^3bc	a^3b^2c	$a^4b^2c = 2$
$(0,0,0)$	$(1,0,0)$	$(1,1,0)$	$(2,1,0)$	$(2,1,1)$	$(3,1,1)$	$(3,2,1)$	$(4,2,1) = (0,0,0)$

[7] However, $(13, 5) = 6$ can be used if $(12, 4)$ is replaced by $(11, 4) = 9/2$. This would then sacrifice the accuracy of the fifth harmonic to increase the accuracy of the sixth. Tradeoffs such as this are common.

As before, a^2b is a pure fourth that defines the tetrachord. The new interval c is typically given by the interval remaining when two tetrachords are joined, and so $c = 9/8$. There are no standard values for a and b. Rather, many different combinations have been explored over the years. The $\oplus$-table for this diatonic tetrachordal scale is given in Table 12.5. As before, it is not necessary that $(4, 2, 1)$ be an exact octave, although it must define the intervals at which the scale repeats.

Table 12.5. $\oplus$-table for the specified tetrachordal scale.

$\bigoplus$	(0,0,0)	(1,0,0)	(1,1,0)	(2,1,0)	(2,1,1)	(3,1,1)	(3,2,1)
(0,0,0)	(0,0,0)	(1,0,0)	(1,1,0)	(2,1,0)	(2,1,1)	(3,1,1)	(3,2,1)
(1,0,0)	(1,0,0)	*	(2,1,0)	*	(3,1,1)	*	(0,0,0)
(1,1,0)	(1,1,0)	(2,1,0)	*	*	(3,2,1)	(0,0,0)	*
(2,1,0)	(2,1,0)	*	*	*	(0,0,0)	(1,0,0)	(1,1,0)
(2,1,1)	(2,1,1)	(3,1,1)	(3,2,1)	(0,0,0)	*	*	*
(3,1,1)	(3,1,1)	*	(0,0,0)	(1,0,0)	*	*	(2,1,1)
(3,2,1)	(3,2,1)	(0,0,0)	*	(1,1,0)	*	(2,1,1)	*

Table 12.6. A perfect spectrum for the specified tetrachordal scale.

i	1	2	3	4	5	6	7	k
t_i	(4,2,1)	(6,3,2)	(8,4,2)	(11,5,3)	(12,6,3)	(14,7,4)	(16,8,4)	
s_i	(0,0,0)	(2,1,1)	(0,0,0)	(3,1,1)	(0,0,0)	(2,1,1)	(0,0,0)	
$r_{i,k}$		(2,1,1)	(2,1,0)	(3,1,1)	(1,1,0)	(2,1,1)	(2,1,0)	1
			(0,0,0)	(1,0,0)	(0,0,0)	(3,2,1)	(0,0,0)	2
				(3,1,1)	(2,1,0)	(2,1,1)	(1,1,0)	3
					(0,0,0)	(0,0,0)	(0,0,0)	4
						(2,1,1)	(2,1,0)	5
							(0,0,0)	6

Spectra can be constructed by following the symbolic spectrum construction procedure, and one such spectrum is given in Table 12.6. Observe that all of the s_i and $r_{i,k}$ are scale steps and that all seven scale steps are present among the s_i or $r_{i,k}$. Hence, this spectrum is perfect for the specified tetrachordal scale.

In order to draw the dissonance curve, it is necessary to pick particular values for the parameters a, b, and c. As mentioned above, $c = 9/8$ is the usual difference between two tetrachords and the octave. Somewhat arbitrarily, let $b = 10/9$, which, combined with the condition that $a^2b = 4/3$ (i.e., forms a tetrachord), imply that $a = \sqrt{6/5}$. With these values, the spectrum defined in Table 12.6 is

$$f,\ 2f,\ 3f,\ 4f,\ 6.57f,\ 8f,\ 12f,\ \text{and}\ 16f,$$

and the resulting dissonance curve is given in Fig. 12.5 when the amplitude of the ith partial is 0.9^i. Minima occur at all scale steps except the first, the interval a. Although this may seem like a flaw, it is normal for small intervals (like the major second) to fail to be consonant; the Pythagorean spectrum of the previous section was atypical in this respect. Again, although a few broad minima occur, they are fairly undistinguished from the surrounding intervals. Thus, the symbolic method of spectrum construction has again found a spectrum that is well suited to the desired scale.

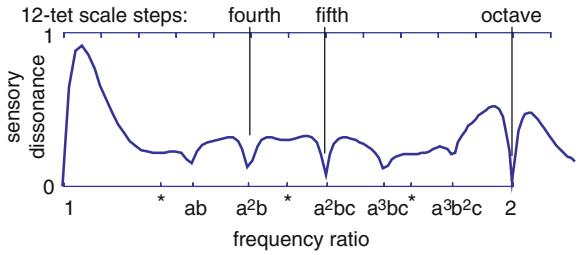


Fig. 12.5. The dissonance curve for the spectrum related to the diatonic tetrachord with $a^2 = \frac{6}{5}$, $b = \frac{10}{9}$, and $c = \frac{9}{8}$, has minima at all scale steps except for the first. The broad minima at the starred locations are not caused by coinciding partials.

12.5.4 When Perfection Is Impossible

The above examples may lull the unsuspecting into a belief that perfect spectra are possible for any scale. Unfortunately, this is not so. Consider first a simple scale built from three arbitrary intervals a, b, and c in the order a, b, c, a. The scale steps are:

$$\begin{array}{ccccc} 1 & a & ab & abc & a^2bc = 2 \\ (0,0,0) & (1,0,0) & (1,1,0) & (1,1,1) & (2,1,1) = (0,0,0) \end{array}$$

As suggested by the notation, $(2,1,1)$ serves as the basic unit of repetition that would likely be the octave. The $\oplus$-table for this scale is given in Table 12.7.

The difficulty with this scale is that the element $(1,1,0)$ cannot be combined with any other. The symbolic construction procedure requires at each step that the s_i be expressible as a $\oplus$-sum of s_j and some $r_{i,k}$. But it is clear that the operation does not allow $(1,1,0)$ as a product with any element (other than the identity) due to the column of $*$'s. In other words, if the interval $(1,1,0)$ ever appears as a partial in the spectrum or as one of the $r_{i,k}$, then the construction process must halt because no more complementary

Table 12.7. $\oplus$-table for the scale defined by three intervals in the order a, b, c, a.

$\bigoplus$	(0,0,0)	(1,0,0)	(1,1,0)	(1,1,1)
(0,0,0)	(0,0,0)	(1,0,0)	(1,1,0)	(1,1,1)
(1,0,0)	(1,0,0)	*	*	(0,0,0)
(1,1,0)	(1,1,0)	*	*	*
(1,1,1)	(1,1,1)	(0,0,0)	*	*

partials can be added. In this particular example, it is possible to create a perfect spectrum by having the element $(1, 1, 0)$ appear only as the very last partial. However, such a strategy would not work if there were two columns of $*$'s.

An extreme example for which no perfect spectrum is possible is a scale defined by four different intervals a, b, c, and d taken in alphabetical order. The scale steps are:

1	a	ab	abc	$abcd = 2$
$(0, 0, 0, 0)$	$(1, 0, 0, 0)$	$(1, 1, 0, 0)$	$(1, 1, 1, 0)$	$(1, 1, 1, 1) = (0, 0, 0, 0)$

As suggested by the notation, $(1, 1, 1, 1)$ serves as the basic unit of repetition that would typically be the octave. The $\oplus$-table for this scale is given in Table 12.8.

Table 12.8. $\oplus$-table for a simple scale defined by four different intervals.

$\bigoplus$	(0,0,0,0)	(1,0,0,0)	(1,1,0,0)	(1,1,1,0)
(0,0,0,0)	(0,0,0,0)	(1,0,0,0)	(1,1,0,0)	(1,1,1,0)
(1,0,0,0)	(1,0,0,0)	*	*	*
(1,1,0,0)	(1,1,0,0)	*	*	*
(1,1,1,0)	(1,1,1,0)	*	*	*

Partials of a complementary spectrum for this scale can only have intervals that are multiples of the octave $(1, 1, 1, 1)$ due to the preponderance of disallowed $*$ entries in the $\oplus$-table. The only possible complementary spectrum is $(0, 0, 0, 0)f$, $(1, 1, 1, 1)f$, $(2, 2, 2, 2)f$, and so on, which is clearly not complete, and hence not perfect. Thus, a given scale may or may not have perfect spectra, depending on the number and placement of the $*$ entries in the table.

12.5.5 Discussion

Do not confuse the idea of a spectrum related to a given scale with the notion of a perfect (complete and complementary) spectrum for the scale. The former is based directly on a psychoacoustic measure of the sensory dissonance of the

sound, and the latter is a construction based on the coincidence of partials within the spectrum. The latter is best viewed as an approximation and simplification of the former, in the sense that it leads to a tractable system for determining spectra via the principle of coinciding partials. But they are not identical.

Some scale intervals that appear in the spectrum (i.e., among the s_i or the $r_{i,k}$ of Tables 12.2, 12.4, or 12.6) may not be minima of the dissonance curve. For instance, the tetrachordal spectrum does not have a minimum at the first scale step even though it is complete. Alternatively, some minima may occur in the dissonance curve that are not explicitly ratios of partials. Three such minima occur in Fig. 12.5; they are the broad kind of minima that are due to wide spacing between certain pairs of partials.

The notion of a perfect spectrum shows starkly that the most important feature of related spectra and scales is the coincidence of partials of a tone—a result that would not have surprised Helmholtz. Perhaps the crucial difference is that related spectra take explicit account of the amplitudes of the partials, whereas perfect spectra do not. In fact, by manipulating the amplitudes of the partials, it is possible to make various minima appear or disappear. For instance, it is possible to "fix" the problem that the tetrachordal spectrum is missing its first scale step a by increasing the amplitudes of the partials that are separated by the ratio a. Alternatively, it is often possible to remove a minimum from the dissonance curve of a perfect spectrum by decreasing the amplitudes of the partials separated by that interval. Moreover, although a minimum due to coinciding partials may be extinguished by manipulating the amplitudes, its location (the interval it forms) remains essentially fixed. In contrast, the broad type minima that are not due to coinciding partials move continuously as the amplitudes vary; they are not a fixed feature of a perfect spectrum.

As the number of different intervals in a desired scale increases, it becomes more difficult to find perfect spectra; the $\oplus$-tables become less full (i.e., have more disallowed $*$ entries) and fewer solutions to the $\oplus$-equation exist. There are several simple modifications to the procedure that may result in spectra that are well matched to the given scale, even when perfection is impossible. One simple modification is to allow the spectrum to be incomplete. As very small intervals are unlikely to be consonant with any reasonable amplitudes of the partials, they may be safely removed from consideration. A second simplifying strategy is to relax the requirement of complementarity—although it is certainly important that prominent scale steps occur at minima, it is not obviously harmful if some extra minima exist. Indeed, if an extra minimum occurs in the dissonance curve but is never played in the piece, then its existence will be transparent to the listener.

A third method of relaxing the procedure can be applied whenever the scale is specified only over an octave (or over some pseudo-octave), in which case the completeness and complementarity need only hold over each octave. For instance, a partial t_i might be chosen even though it forms a disallowed

interval with a previous partial t_j, providing the two are more than an octave apart. Thus, judicious relaxation of various elements of the procedure may allow specification of useful spectra even when perfect spectra are not possible.

Perfect spectra raise a number of issues. For instance, a given nonequal scale sounds different in each key because the set of intervals is slightly different. How would the use of perfect spectra influence the ability to modulate through various keys? Certain chords will become more or less consonant when played with perfect spectra than when played with harmonic tones. What patterns of (non)harmonic motion are best suited to perfect spectra and their chords? Will perfect spectra be useful for some part of the standard repertoire, or will they be only useful for new compositions that directly exploit their strengths (and avoid their weaknesses)?

12.6 Summary

Given a spectrum, what is the related scale? was answered completely in previous chapters; draw the dissonance curve and gather the intervals at which its minima occur into a scale. This chapter wrestled with the more difficult inverse question: *Given a scale, what is the related spectrum?* One approach posed the question as a constrained optimization problem that can sometimes be solved using iterative search techniques. Reducing the size of the search space increases the likelihood that a good spectrum is found. The second approach exploits the principle of coinciding partials and reformulates the question in algebraic form.

Neither approach completely specifies a "best" spectrum for the given scale. Both stipulate the frequencies of the partials, but the optimization method assumes a set of amplitudes a priori, whereas the algebraic procedure leaves the amplitudes free. Thus, each answer gives a whole class of related spectra that may sound as different from each other as a trumpet from a violin or a flute from a guitar. Neither method gives any indication of how such sounds might be generated or created. One obvious way is via additive synthesis, but unless great care is taken, additive synthesis can result in static and lifeless sounds. An alternative is to begin with sampled sounds and to manipulate the partials so that they coincide with the desired perfect spectrum. This technique, called "spectral mapping," is discussed at length in the next chapter. A much more difficult question is how acoustic instruments might be given the kinds of deviations from harmonicity that are specified by perfect and related spectra.

13

Spectral Mappings

A spectral mapping is a transformation from a "source" spectrum to a "destination" spectrum. One application is to transform inharmonic sounds into harmonic equivalents. More interestingly, it can be used to create inharmonic instruments that retain much of the tonal quality of familiar (harmonic) instruments. Musical uses of such timbres are discussed, and forms of (inharmonic) modulation are presented. Several sound examples demonstrate both the breadth and limitations of the method.

13.1 The Goal: Life-like Inharmonic Sounds

A large number of different timbres can be created using only sounds with a harmonic spectrum. It should be possible to get at least as large a variety using inharmonic sounds. This chapter shows one way to make imitative inharmonic sounds, ones that seem to come from real instruments. This is how an inharmonic trumpet or guitar might sound.

Suppose a composer desires to play in some specified scale, say, in 11-tet. As familiar harmonic sounds are dissonant when played in 11-tet, it may be advantageous to create a new set of sounds, with spectra that cause minima of the dissonance curve to occur at the appropriate 11-tet scale steps. Figure 13.1, for example, shows the dissonance curve for a spectrum that has major dips at many of the locations of the 11-tet scale steps. This spectrum was designed using the techniques of the previous chapter, which specifies only a desired set of partials. But a complete spectrum consisting of magnitudes and phases must be chosen to draw the dissonance curve and to transform the sound into a time waveform for playback. In the figure, all partials are assumed equal, giving the sound a rich organish quality.

The most straightforward approach to the problem of sound synthesis from a specified set of partials is additive synthesis, such as described in Risset [B: 150], in which a family of sine waves of desired amplitude and phase are summed. Although computationally expensive, additive synthesis is conceptually straightforward. A major problem is that it is often a monumental task to specify all of the parameters (frequencies, magnitudes, and phases) required for the synthesis procedure, and there is no obvious or intuitive path to follow when generating new sounds. When attempting to create sounds for new scales, such as the 11-tet timbre above, it is equally challenging to choose

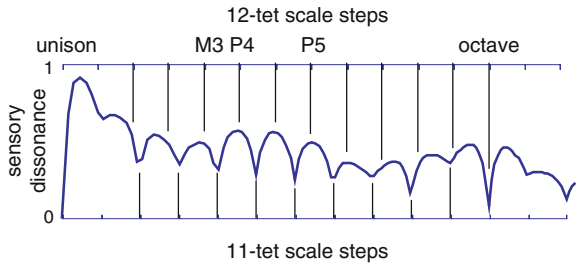


Fig. 13.1. Dissonance curve for the spectrum with equal amplitude partials at [1 a^{11} a^{17} a^{22} a^{26} a^{28} a^{31} a^{33} a^{35} a^{37} a^{38}], where $a = \sqrt[11]{2}$. The minima of this dissonance curve occur at many of the 11-tet scale steps (bottom axis) and not at the 12-tet scale steps (top axis).

these parameters in a musical way. Making arbitrary choices often leads to organ or bell-like sonorities, depending on the envelope and other aspects of the sound. Although these can be striking, they can also be limiting from a compositional perspective. Is there a way to create a full range of tonal qualities that are all related to the specified scale? For instance, how can "flute-like" or "guitar-like" timbres be built that are consonant when played in this 11-tet tuning?

A common way to deal with the vast amount of information required by additive synthesis is to analyze a desired sound via a Fourier (or other) transform, and then use the parameters of the transform in the additive synthesis. In such analysis/synthesis schemes, the original sound is transformed into a family of sine waves, each with specified amplitude and phase. The parameters are stored in memory and are used to reconstruct the sound on demand. In principle, the methods of analysis/synthesis allow exact replication of any waveform. Of course, the sound to be resynthesized must already exist for this procedure to be feasible. Unfortunately, 11-tet flutes and guitars do not exist.

Once a sound is parameterized, it is possible to manipulate the parameters. For example, the technique of Grey and Moorer [B: 64] interpolates the envelopes of harmonics to gradually transform one instrumental tone into another. Strong and Clark [B: 186] exchange the spectral and temporal envelopes among a number of instruments of the wind family and conduct tests to evaluate their relative significance. Probably the first parameter-based analysis/synthesis methods were the vocoder of Dudley [B: 45] and its modern descendant the phase vocoder of Flanagan and Golden[B: 55], which were designed for the efficient encoding of transmitted speech signals.

The consonance-based spectral mappings of this chapter are a kind of analysis/synthesis method in which the amplitudes and phases of the spectrum of the "source" sound are grafted onto the partials of a specified "destination" spectrum, which is chosen so as to maximize a measure of consonance (or more properly, to minimize a measure of dissonance). The goal is to relocate the

partials of the original sound for compatibility with the destination spectrum, while leaving the tonal quality of the sound intact. Musically, the goal is to modify the spectrum of a sound while preserving its richness and character. This provides a way to simulate the sound of nonexistent instruments such as the 11-tet flute and guitar. Figure 13.2 shows the spectral mapping scheme in block diagram form. The input signal is transformed into its spectral parameters, the mapping block manipulates these parameters, and the inverse transform returns the signal to a time-based waveform for output to a D/A converter and subsequent playback.

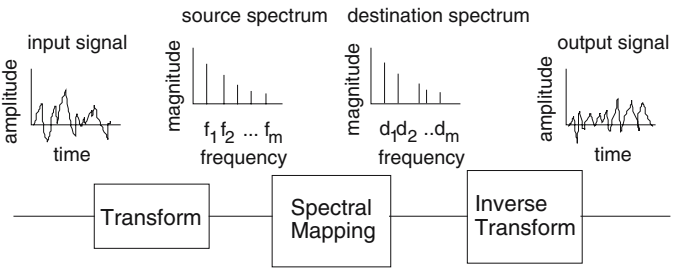


Fig. 13.2. Block Diagram of a transform-based analysis-synthesis spectral mapping. If the mapping is chosen to be the identity, then the input and output signals are identical.

13.2 Mappings between Spectra

A spectral mapping is defined to be a transformation from a set of n partials $s_1,\ s_2, ...,\ s_n$ (called the "source spectrum") to the partials $d_1,\ d_2, ...,\ d_n$ of the "destination spectrum" for which $T(s_i) = d_i$ for all i. Suppose that an N-point DFT (or FFT) is used to compute the spectrum of the original sound, resulting in a complex-valued vector X. The mapping T is applied to X (which presumably has partials at or near the s_i), and the result is a vector $T(X)$, which represents a spectrum with partials at or near the d_i. This is shown schematically in Fig. 13.3 for an "arbitrary" destination spectrum.

The simplest T is a "straight-line" transformation

$$T(s) = \left(\frac{d_{i+1} - d_i}{s_{i+1} - s_i}\right) s + \left(\frac{d_i s_{i+1} - d_{i+1} s_i}{s_{i+1} - s_i}\right) \qquad s_i \le s \le s_{i+1}.$$

Smoother curves such as parabolic or spline interpolations can be readily used, but problems occur with such direct implementations due to the quantization of the frequency axis inherent in any digital representation of the spectrum. For instance, if the slope of T is significantly greater than unity, then certain

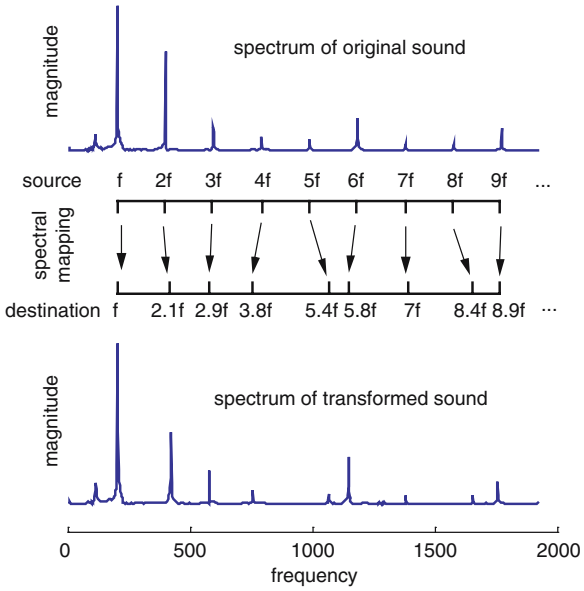


Fig. 13.3. Schematic representation of a spectral mapping. The first nine partials of a harmonic "source spectrum" are mapped into an inharmonic "destination spectrum" with partials at f, $2.1f$, $2.9f$, $3.8f$, $5.4f$, $5.8f$, $7f$, $8.4f$, and $8.9f$. The spectrum of the original sound (from the G string of a guitar with fundamental at 194 Hz) is transformed by the spectral mapping for compatibility with the destination spectrum. The mapping changes the frequencies of the partials while preserving both magnitudes (shown) and phases (not shown).

elements of $T(X)$ will be empty. More seriously, if the slope of T is significantly less than unity, then more than one element of X will be mapped into the same element of $T(X)$, causing an irretrievable loss of information. It is not obvious how to sensibly combine the relevant terms.

A better way to think of the spectral mapping procedure is as a kind of "resampling" in which the information contained between the frequencies s_i and s_{i+1} is resampled[1] to occupy the frequencies d_i to d_{i+1}. Resampling is a standard digital signal processing technique with a long history and a large literature. It generally consists of two parts, *decimation* and *interpolation*, which together attempt to represent the "same" information with a different number of samples.

[1] One implementation uses a polyphase algorithm with an anti-aliasing low-pass FIR filter incorporating a Kaiser window. The examples in this chapter filter ten terms on either side of x_i and use $\beta = 5$ as the window design parameter. These are the defaults of `Matlab`'s built in "resample" function. An alternative is to use $sinc(\cdot)$ interpolation as discussed in [W: 29].

One presumption underlying spectral mappings is that the most important information (the partials of the sound) is located at or near the s_i, and it is to be relocated as 'intact' as possible near the d_i. Figure 13.4 shows an exaggerated view of what occurs to a single partial when performing a straightforward resampling with a nonunity spectral map T. In essence, the "left half" of the spectrum becomes asymmetric from the "right half," and the transformed spectrum no longer represents a single sinusoid. This is a kind of nonlinear distortion that can produce audible artifacts.

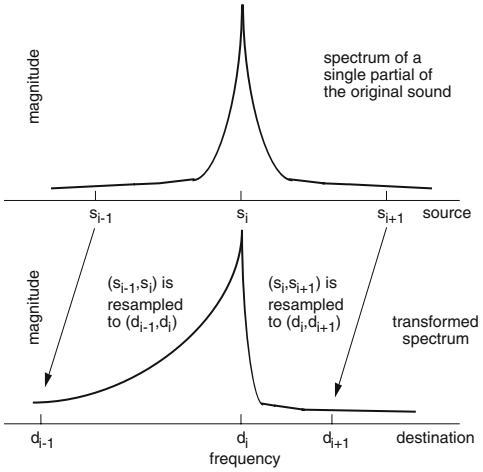


Fig. 13.4. Resampling causes asymmetries in the transformed spectrum that may cause audible anomalies.

One way to reduce this distortion is to chose a window of width $2w$ about the s_i that is mapped identically to a window of the same width about d_i. The remaining regions, between $s_i + w$ and $s_{i+1} - w$, can then be resampled to fit between $d_i + w$ and $d_{i+1} - w$. This is shown (again in exaggerated form) in Fig. 13.5. In this method of *Resampling with Identity Window* (RIW), the bulk of the most significant information is transferred to the destination intact. Changes occur only in the less important (and relatively empty) regions between the partials. We have found that window widths of about 1/3 to 1/5 of the minimum distance between partials to be most effective in reducing the audibility of the distortion.

Spectral mappings are most easily implemented in software (or in hardware to emulate such software) in a program:

```
input spectrum = FFT(input signal)
mapped spectrum = T(input spectrum)
output signal = IFFT(mapped spectrum)
```

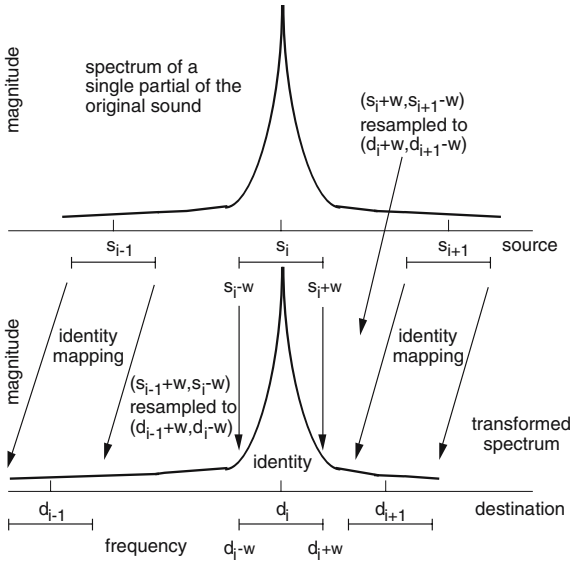


Fig. 13.5. Resampling with identity windows reduces the asymmetry of the transformed spectrum.

where the function $FFT()$ is the Discrete Fourier Transform or its fast equivalent, $IFFT()$ is the inverse, and the RIW spectral mapping is represented by T. Other transforms such as the wavelet or constant-Q transform [B: 19] might also be useful. Spectral mappings can be viewed as linear (but time-varying) transformations of the original signal. Let the signal be x, and let F be the matrix that transforms x into its DFT. Then the complete spectral mapping gives the output signal

$$\hat{x} = F^{-1}TF(x)$$

where T is a matrix representation of the resampling procedure. This is clearly linear, and it is time varying because the frequencies of signals are not preserved. Often T fails to be invertible, and the original signal x cannot be reconstructed from its spectrally mapped version $\hat{x}$.

There are many possible variations of T. For instance, many instrumental sounds can be characterized using formants, fixed linear filters through which variable excitation passes. If the original samples are of this kind, then it is sensible to modify the amplitudes of the resulting spectra accordingly. Similarly, an "energy" envelope can be abstracted from the original sample, and in some situations, it might be desirable to preserve this energy during the transformation. In addition, there are many kinds of resampling (interpolation and decimation), and there are free parameters (and filters) within each kind. Trying to choose these parameters optimally is a daunting task.

It may be more efficient computationally to implement spectral mappings as a filter bank rather than as a transform (a good modern approach to filter banks may be found in [B: 185]), especially when processing a continuous audio signal. This is diagrammed in Fig. 13.6, which shows a bank of filters carrying out the analysis portion of the procedure, a spectral mapping to manipulate the parameters of the spectrum, and a bank of oscillators to carry out the synthesis portion. This does not change the motivation or goals of the mappings, but it does suggest an alternative hardware (or software) approach.

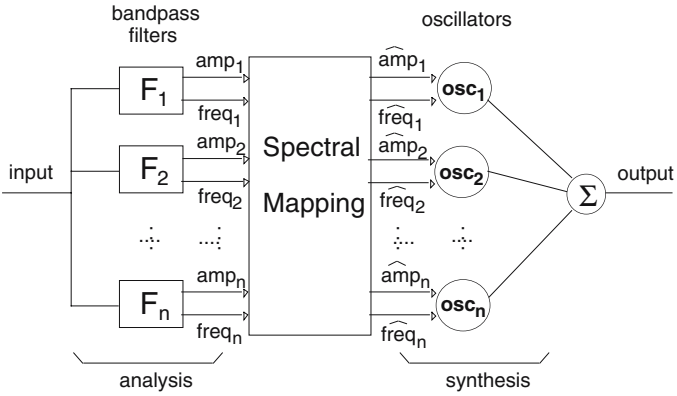


Fig. 13.6. A filter-bank implementation of spectral mapping. The input is bandpass filtered, and the signal is parameterized into n amplitude, phase, and frequency parameters. These are transformed by the spectral mapping, and the modified parameters drive n oscillators, which are summed to form the output.

13.2.1 Maintaining Amplitudes and Phases

The tonal quality of a harmonic sound is determined largely by the amplitudes of its sinusoidal frequency components. In contrast, the phases of these sinusoids tend to play a small role, except in the transient (or attack) portion of the sound, where they contribute to the envelope. The transformation T is specified so as to keep each frequency component (roughly) matched with its original amplitude and phase. This tends to maintain the shape of the waveform in the attack portion. For example, Fig. 13.7 shows a square wave and its transformation into the 11-tet timbre specified in Fig. 13.1. The first few pulses are clearly discernible in the mapped waveform. As the first few milliseconds of a sound are important in terms of the overall sound quality, maintaining the initial shape of the waveform contributes to the goal of retaining the integrity of the sound.

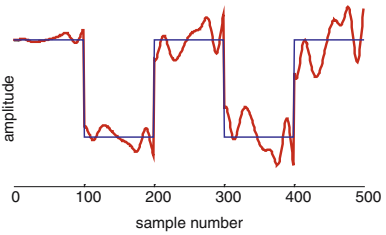


Fig. 13.7. A square wave and its transformation into a 11-tet version. Maintaining the phase relationships among the partials helps the attack portion retain its integrity.

13.2.2 Looping

A common practice in sample-based synthesizers is to "loop" sounds, to repeat certain portions of the waveform under user control. Periodic portions of the waveform are ideal candidates for looping. Strictly speaking, inharmonic sounds such as result from transformations like the 11-tet spectral mappings have aperiodic waveforms. Apparently, looping becomes impossible. On the other hand, the FFT induces a quantization of the frequency axis in which all frequency components are integer multiples of the frequency of the first FFT bin (for instance, about 1.3 Hz for a 32K FFT at a 44.1 KHz sampling rate). Thus, true aperiodicity is impossible in a transform-based system. In practice, it is often possible to loop the sounds effectively using the standard assortment of looping strategies and cross fades, although it is not uncommon for the loops to be somewhat longer in the modified waveform than in the original.

To be concrete, suppose that the original waveform contains a looped portion. A sensible strategy is to append the loop onto the end of the waveform several times, as shown in Fig. 13.8. This tends to make a longer portion of the modified waveform suitable for looping. It is also a sensible way of filling or padding the signal until the length of the wave is an integer power of two (so that the more efficient FFT can be computed in place of the DFT). The familiar strategy of padding with zeroes is inappropriate in this application. Figure 13.9, for instance, shows the results of three different mappings of the 4500 sample trumpet waveform of Fig. 13.8. Calculating the DFT and applying the 11-tet spectral mapping of Fig. 13.1 gives the waveform in Fig. 13.9(a). This version consists primarily of the attack portion of the waveform, and is it virtually impossible to loop without noticeable artifacts. An alternative is to extend the waveform to 8K samples by filling with zeroes. This allows use of the FFT for faster computation, but the resulting stretched waveform of Fig. 13.9(b) is no easier to loop than the signal in 13.9(a). A third alternative is to repeatedly concatenate the original looped portion until the waveform reaches the desired 8K length. The resulting stretched version contains a longer sustain portion, and it is correspondingly easier to loop.

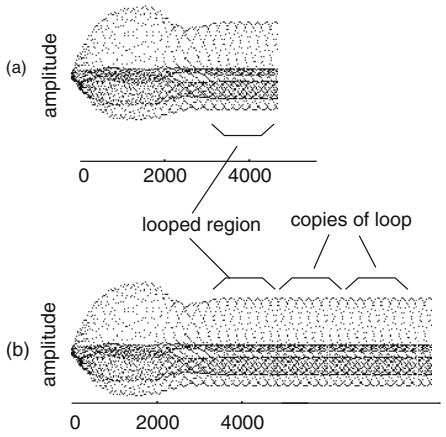


Fig. 13.8. (a) A 4500 sample trumpet waveform with looped region indicated. (b) The same waveform using a "fill with loop" rather than a "fill with zeroes" strategy to increase the length of the wave to 8K samples.

13.2.3 Separating Attack from Loop

The attack portion of a sound is often quite different from the looped portion. The puff of air as the flute chiffs, the blat of the trumpets attack, or the scrape of the violins bow are different from the steady-state sounds of the same instruments. Indeed, Strong and Clark [B: 186] have shown that it can often be difficult to recognize instrumental sounds when the attack has been removed.

Naive application of a spectral mapping would transform the complete sampled waveform simultaneously. Because the Fourier transform has poor time localization properties, this can cause a "smearing" of the attack portion over the whole sample, with noticeable side effects. First, the smearing can

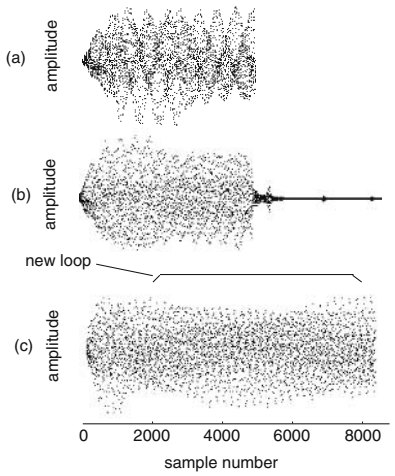


Fig. 13.9. Spectrally mapped versions of the trumpet waveform in Fig. 13.8. (a) Using a DFT of the original wave. (b) Using an FFT and the "fill with zeroes" strategy. (c) Using an FFT and the "fill with loop" strategy. Version (c) gives a longer, steadier waveform with more opportunity to achieve a successful loop.

sometimes be perceived directly as artifacts: a high tingly sound, or a noisy grating that repeats irregularly throughout the looped portion of the sound. Second, because the artifacts are nonuniform, they make creating a good loop of the mapped sound more difficult.[2]

Thus, a good idea when spectrally mapping sampled sounds (for instance, those with predefined attack and loop segments) is to map the attack and the loop portions separately, as shown in Fig. 13.10. The resulting pieces can then be pasted back together using a simple crossfade. This tends to maintain the integrity of the attack portion (it is shorter and less likely to suffer from phase and smearing problems), and to reduce artifacts occurring in the steady state.

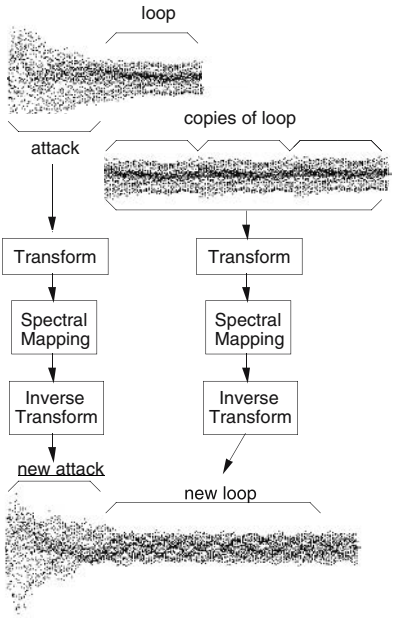


Fig. 13.10. Transforming the attack and steady-state (looped) portions separately helps to maintain the tonal integrity of the sound.

Often, a complete sampled "instrument" contains several different waveforms sampled in different pitch ranges and at different dynamic ranges. The creation of a spectrally mapped version should map each of these samples and then assign them to the appropriate pitch or dynamic performance level. In addition, it is reasonable to impose the same envelopes and other performance parameters such as reverb, vibrato, and so on, as were placed on the original samples, because these will often have a significant impact on the overall perception of the quality of the sound.

[2] Even the looping of familiar instrumental sounds can be tricky.

13.3 Examples

This section presents examples of spectral maps in which the integrity of the original sounds is maintained, and others in which the perceptual identity of sounds is lost. Examples include instruments mapped into a spectrum consonant with 11-tet and with 88-cet, a cymbal sound mapped so as to be consonant with harmonic sounds, and instruments mapped into (and out of) the spectrum of a drum. Spectrally mapped sounds can be useful in musical compositions, and Table 13.1 lists all of the pieces on the CD that feature sounds mapped into the specified scales.

Table 13.1. Musical compositions on the CD-ROM using sounds that are spectrally mapped into the specified scale.

Name of Piece	Scale	File	For More Detail
88 Vibes	88-cet	`vibes88.mp3`	[S: 16]
Anima	10-tet	`anima.mp3`	[S: 106]
Circle of Thirds	10-tet	`circlethirds.mp3`	[S: 104]
Glass Lake	tom-tom	`glasslake.mp3`	[S: 91]
Haroun in 88	88-cet	`haroun88.mp3`	[S: 15]
Hexavamp	16-tet	`hexavamp.mp3`	[S: 97]
Isochronism	10-tet	`isochronism.mp3`	[S: 105]
March of the Wheel	7-tet	`marwheel.mp3`	[S: 115]
Nothing Broken in Seven	7-tet	`broken.mp3`	[S: 117]
Pagan's Revenge	7-tet	`pagan.mp3`	[S: 116]
Phase Seven	7-tet	`phase7.mp3`	[S: 118]
Seventeen Strings	17-tet	`17strings.mp3`	[S: 98]
Sonork	harmonic	`sonork.mp3`	[S: 93]
Sympathetic Metaphor	19-tet	`sympathetic.mp3`	[S: 101]
Ten Fingers	10-tet	`tenfingers.mp3`	[S: 102]
The Turquoise Dabo Girl	11-tet	`dabogirl.mp3`	[S: 88]
Truth on a Bus	19-tet	`truthbus.mp3`	[S: 100]
Unlucky Flutes	13-tet	`13flutes.mp3`	[S: 99]

13.3.1 Timbres for 11-tone Equal Temperament

Familiar harmonic sounds may be dissonant when played in 11-tet because minima of the dissonance curve occur far from the desired scale steps. By using an appropriate spectral mapping, harmonic instrumental timbres can be transformed into 11-tet versions with minima at many of the 11-tet scale steps, as shown in Fig. 13.1. These can be used to play consonantly in a 11-tet setting. The mapping used to generate the tones in the sound example maps a set of harmonic partials at

$$f,\ 2f,\ 3f,\ 4f,\ 5f,\ 6f,\ 7f,\ 8f,\ 9f,\ 10f,\ 11f$$

to

$$f,\ r^{11}f,\ r^{17}f,\ r^{22}f,\ r^{26}f,\ r^{28}f,\ r^{31}f,\ r^{33}f,\ r^{35}f,\ r^{37}f,\ r^{38}f$$

where $r = \sqrt[11]{2}$ and f is the fundamental of the harmonic tone. All frequencies between these values are mapped using the RIW method.

Sound example [S: 86] (and video example [V: 11]) contain several different instrumental sounds that alternate with their 11-tet versions.[3]

(i) Harmonic trumpet compared with 11-tet trumpet
(ii) Harmonic bass compared with 11-tet bass
(iii) Harmonic guitar compared with 11-tet guitar
(iv) Harmonic pan flute compared with 11-tet pan flute
(v) Harmonic oboe compared with 11-tet oboe
(vi) Harmonic "moog" synth compared with 11-tet "moog" synth
(vii) Harmonic "phase" synth compared with 11-tet "phase" synth

The instruments are clearly recognizable after mapping into their 11-tet counterparts. There is almost no pitch change caused by this spectral mapping, probably because some partials are mapped higher, whereas others are mapped lower. Indeed, the third partial is mapped lower than its harmonic counterpart (2.92 vs. 3), but the fifth is higher (5.14 vs. 5). Similarly, the sixth is lower (5.84 vs. 6), but the seventh is higher (7.05 vs. 7).

Perhaps the clearest change is that some of the samples have acquired a soft high-pitched inharmonicity. It is hard to put words to this, but we try. In (i) it may almost be called a "whine." (ii) has a slight lowering of the pitch, as well as a feeling that "something else" is attached. (iii) has acquired a high "jangle" in the transition. It is hard to pinpoint any changes in (iv) and (vi). In (v), it becomes easier to "hear out" one of the partials in the mapped sound, giving it an almost minorish feel. The natural vibrato of (vii) appears to have changed slightly, but it is otherwise intact.

Despite the fact that all sounds were subjected to the same mapping, the perceived changes differ somewhat from sample to sample. This is likely an inherent aspect of spectral mappings. For instance, the bass has a strong third partial and a weak fifth partial compared with the other sounds. As the third partial is mapped down in frequency, it is reasonable to hypothesize that this causes the lowering in pitch. Because the fifth partial is relatively weak,

[3] The waveforms were taken from commercially available sample CD-ROMs and transferred to a computer running a `Matlab` program that performed the spectral mappings. After looping (which was done manually, with the help of *Infinity* looping software), the modified waveforms were sent to an Ensoniq ASR-10 sampler. The performances were sequenced and recorded to digital audiotape. In all cases, the same performance parameters (filters, envelopes, velocity sensitivity, reverberation, etc.) were applied to the spectrally mapped sounds as were used in the original samples.

it cannot compensate, as might occur in other sounds. Similarly, differing amplitudes of partials may cause the varying effects perceivable in (i)-(vii).

Such perceptual changes may be due to the way that inharmonicities are perceived. For instance, Moore [B: 115] examines the question of how much detuning is needed before an inharmonic partial causes a sound to break into two sounds rather than remain fused into a single percept. Alternatively, the changes may be due to artifacts created by the spectral mapping procedure. For instance, other choices of filters, windows widths, and so on, may generate different kinds of artifacts. Poorly implemented spectral mappings can introduce strange effects. For example, in some of the earliest experiments with spectral mappings, many sounds acquired a high-pitched jangling effect. The piece *Seventeen Strings* [S: 98] features these sounds, and the jangling provides an interesting high pitched background to the foreground harp. Although this may be acceptable in a single piece as a special effect, it is undesirable overall. This was the major impetus for separating the attack and looped portion of the sounds in the mapping procedure—separation reduces the artifacts significantly.

Isolated sounds do not paint a very good picture of their behavior in more complex settings. A short sequence of major chords are played in sound example [S: 87][4]:

(viii) Harmonic oboe in 12-tet
(ix) Spectrally mapped 11-tet oboe in 12-tet

As before, the individual sounds have only a small pitch shift. The striking difference between (viii) and (ix) shows that the "out-of-tune" percept may be caused by the structure of the partials of a sound, as well as by pitch or interval relationships. Sound example [S: 87](ix) is not literally "out-of-tune" because its fundamental is tuned to the accuracy of the equipment, which is about 1.5 cents. Rather, (ix) is "out-of-spectrum" or "out-of-timbre," in the sense that the partials of the sound interfere when played at certain intervals (in this case the 12-tet major third and fifth).

The next segments contain 11-tet dyads formed from scale steps 0-6 and 0-7, and culminate in a chord composed of scale steps 0-4-6.

(x) Harmonic oboe in 11-tet
(xi) Spectrally mapped 11-tet oboe in 11-tet

Examples (x) and (xi) reverse the situation from (viii) and (ix). Because of the extreme unfamiliarity of the intervals (observe that 11-tet scale steps 4 and 6 do not lie close to any 12-tet intervals), the situation is perhaps less clear, but there is a readily perceivable roughness of the 0-4-6 chord in (x) that is absent from (xi). Thus, after acclimation to the intervals, (xi) appears arguably less out-of-spectrum than (x).

[4] And presented in video format [V: 12].

Isolated chords do not show clearly what happens in genuine musical contexts. The piece, the *Turquoise Dabo Girl*, is played two ways:

Sound example [S: 88] in 11-tet with all sounds spectrally mapped.

Sound Example [S: 89] in 11-tet with the original harmonic sounds (first 16 bars only).

The "out-of-spectrum" effect of [S: 89] is far more dramatic than the equivalent isolated chord effect of (x), illustrating that the more musical the context, the more important (rather than the less important) a proper matching of the tuning with the spectrum of the sound becomes.

Hopefully, the *Turquoise Dabo Girl* also demonstrates that many of the kinds of effects normally associated with (harmonic) tonal music can occur, even in strange settings such as 11-tet, which is often considered among the hardest keys in which to play tonal music. Consider, for instance, the harmonization of the 11-tet pan flute melody that occurs in the "chorus." Does this have the feeling of some kind of (perhaps unfamiliar) "cadence" as the melody resolves back to its "tonic?" Does it not sound "in-tune" even though there is only one truly familiar interval (the octave) in the whole piece?

Observe that many of the subtle oddities in the mapped timbres (as noted in (i)-(vii) of sound example [S: 86]) seem to disappear when contextualized. Even with careful listening, it is difficult (impossible?) to hear the inharmonicities and artifacts that were so clear when presented in isolation. All the timbres used in the *Turquoise Dabo Girl* (except the percussion) appear in (i)-(vii). This may be due to a simple masking of the artifacts. It may also be due to a kind of "capture" effect, in which the artifact/inharmonicity of one note is captured by (or streamed with) other notes, and thus it becomes part of the musical flow. In either case, the lessening of tonalness (due to the inharmonicity) does not appear to play a large role in the *Turquoise Dabo Girl*, whereas the dissonance predictions of the sensory theory are readily upheld.

13.3.2 Spectrum of a Drum

The spectral mapping of the previous example changes the partials only moderately. In contrast, mapping from harmonic tones into the spectrum of a drum such as a tom tom changes the partials dramatically. The extreme inharmonicity of the sample is illustrated in Fig. 13.11, and the severe mapping is readily heard as drastic changes in the tonal quality and pitch of the transformed instruments. A harmonic spectrum at $g, 2g, 3g, 4g, 5g$ is mapped to $d, 1.67d, 2.46d, 3.2d, 3.8d$ (which is precisely $245, 410, 603, 786, 934$ for $d = 245$) using the RIW spectral mapping. Of the guitar, bass, trumpet, and flute, only the flute is recognizable, and even this is not without drastic audible changes. One listener remarked that the transformed sounds were "glassy—like a finger nail scratching across a glass surface." This description makes a certain

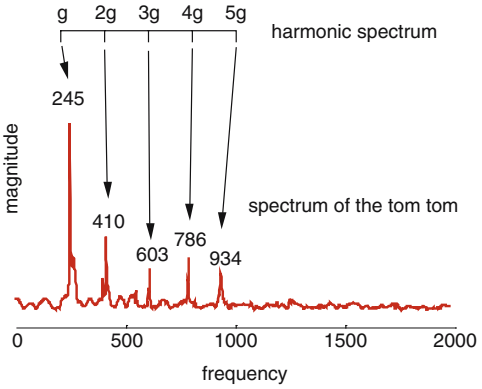


Fig. 13.11. A harmonic spectrum with fundamental g is mapped into the tom tom spectrum.

amount of physical sense, because glass surfaces and drums heads are both two-dimensional vibrating surfaces.

Sound example [S: 90] and video example [V: 13] contain several different instruments and their transformation into the spectrum of the tom tom shown in Fig. 13.11.

(i) Harmonic flute compared with tom tom flute
(ii) Harmonic trumpet compared with tom tom trumpet
(iii) Harmonic bass compared with tom tom bass
(iv) Harmonic guitar compared with tom tom guitar

Clearly, this spectral mapping causes a large change in the character of the sounds. As before, it is unclear what aspects of the resulting changes are due to the way inharmonic sounds are perceived, and what may be due to the details of the spectral mapping procedure. For instance, each of the sounds undergoes a pitch change, but the pitch change is different for each sound. Presumably this is because the partials of the mapped sounds inherit the amplitudes of the original sounds. This is consistent with virtual pitch theory where the ear picks out different "harmonic templates" (see Sect. 2.4.2 on p. 34) for each arrangement of amplitudes.

Again, it is hard to describe in words the kind of effects perceived. (i) has a noticeable pitch change, but it still sounds something like a flute. The trumpet undergoes a huge pitch change, and it gains a kind of glassy texture. The single note of the bass becomes a minorish chord, and the guitar pluck also gains a chord-like sound along with jangly artifacts.

Although the transformed timbres do not sound like the instruments from which they were derived, they are not necessarily useless. Sound example [S: 91], the *Glass Lake*, illustrates the transformed instruments (i)-(iv) played in the related scale, with steps defined by the dissonance curve of Fig. 13.12. This scale supports perceptible "chords," although they are not necessarily composed of familiar intervals. The piece is thoroughly xentonal.

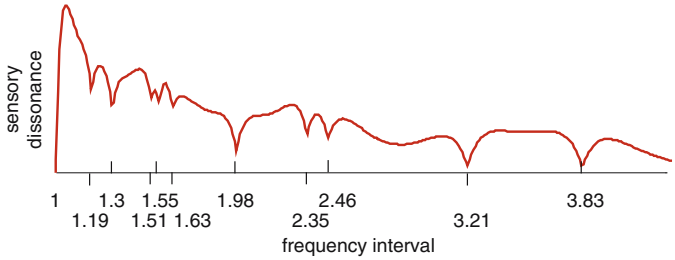


Fig. 13.12. The dissonance curve for the tom tom spectrum has an 11-note related scale that covers a little less than two octaves.

13.3.3 Timbres for 88-cet

Gary Morrison [B: 113] proposed a scale in which the interval between adjacent notes is 88 cents rather than 100 cents as in 12-tet. As 1200 is not divisible by 88, this scale has no real octaves. It can be interpreted as 14 equal divisions of a stretched pseudo-octave with 1232 cents, which corresponds to a ratio of $p = 2.0373$ to 1. One way to specify timbres for this scale is to map from a set of harmonic partials to a set of "88-cet" partials using the mapping

$$\begin{array}{cccccccccc} f & 2f & 3f & 4f & 5f & 6f & 7f & 8f & 9f & 10f \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ f & r^{14}f & r^{22}f & r^{28}f & r^{33}f & r^{36}f & r^{39}f & r^{42}f & r^{44}f & r^{47}f \end{array}$$

where $r = \sqrt[14]{2.0373}$ and f is the fundamental of the harmonic tone. The locations of the destination spectrum are taken from Table 13.2, although here the r is based on the pseudo-octave rather than the real octave. The dissonance curve for this timbre is shown in Fig. 13.13; observe that the curve has many minima at 88-cet scale steps (as expected) and no obvious relationship to the 12-tet scale steps shown above. The most consonant intervals occur at scale steps 1, 4, 6, 7, 9, 12, and 14. This is a good place to begin exploration of this unusual scale.

Two pieces demonstrate this timbre-scale combination in action. *Haroun in 88* [S: 15] is fully orchestrated with 88-cet flute, bass, trumpets, and synths. *88 Vibes* [S: 16] is performed on a spectrally mapped vibraphone.

13.3.4 A Harmonic Cymbal

The previous examples transformed familiar harmonic timbres into unfamiliar timbres and scales. This example uses spectral mappings to transform familiar inharmonic sounds into sounds maximally consonant with harmonic spectra. The spectrum of a cymbal contains many peaks spread irregularly through the whole audible range. For the chosen cymbal sample, the $N = 35$ largest

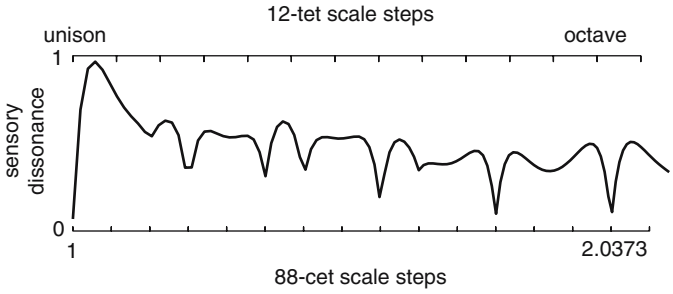


Fig. 13.13. The dissonance curve for the 88-cet spectrum has minima at many of the 88-cet scale steps, which are 14 equal divisions of the 2.0373 pseudo-octave.

peaks (labeled p_i, $i = 1, 2, ...N$) were fit to a "nearby" harmonic template $t_i = if$ by finding the fundamental f that minimizes

$$\sum_{i=1}^{N} (p_i - t_i)^2.$$

The solution is $f = \frac{\sum ip_i}{\sum i^2}$, and the p_i (source) and t_i (destination) define the spectral mapping. The transformed sound retains some of the noisy character of the original cymbal strike, but it has become noticeably more harmonic and has inherited the pitch associated with the fundamental f. The two brief segments in sound example [S: 92] are mirrored in video example [V: 14]:

(i) The original sample contrasted with the spectrally mapped version
(ii) A simple "chord" pattern played with the original sample, and then with the spectrally mapped version

The transformed instrument supports both chord progressions and melodies even though the original cymbal strike does not.

Sonork [S: 93] explores harmonic cymbals in a "prog-rock" setting. Except for the drums, all of the instruments in *Sonork* were created from spectrally mapped cymbals. The origin of the bass, synth, and lead lines is completely disguised. Some sounds in the quieter sections retain recognizable characteristics of the cymbals from which they derive, and some have gained a kind of fluttery underwater ambience from the spectral mapping.

Another example of the mapping of inharmonic instruments into tonal counterparts is presented in sound examples [S: 94] through [S: 96]. The first presents the original drum sound, which is clearly incapable of supporting melody or harmony. The second plays the spectrally mapped version of the drum into a harmonic sound; it has attained a character similar to a xylophone, and it readily supports both melody and harmony. The third example plays both simultaneously and is the most musical of the three.

13.4 Discussion

The discussion begins with a consideration of various aspects of timbral change, and then it suggests additional perceptual tests that might further validate (or falsify) the use of spectral mappings in inharmonic musical applications. Several types of inharmonic musical modulations are discussed.

13.4.1 Robustness of Sounds under Spectral Mappings

How far can partials be mapped before the sound loses cohesion or otherwise changes beyond recognition? It is clear from even a cursory listen that small perturbations in the locations of the partials (i.e., mappings that are not too distant from the identity) have little effect on the overall tonal quality of the sound. Flutes and guitars in 11-tet timbres retain their identity as flutes and guitars. The consistency of such sounds through various spectral mappings argues that perceptions of tonal quality are not primarily dependent on the precise frequency ratios of the partials. Rather, there is a band in which the partials may lie without affecting the "fluteness" or "guitarness" of the sound. Equivalently, the partials of such a sound can undergo a wide variety of mappings without significantly affecting its inherent tonal gestalt.

Besides the sounds demonstrated here, the author has spectrally mapped a large variety (over 100) of sounds into several different destination spectra, including stretched timbres with stretch factors from 1.5 to 3.0 (see [B: 176] and [B: 100] for a detailed discussion of stretched timbres), spectra designed to be consonant with n-tet for $n = 8, ..., 19$, and a variety of destination spectra derived from objects such as a tom tom, a bell, a metal wind chime, and a rock. Many of these are used in the compositions and studies described in Table 13.1. Overall, there is a wide variation in the robustness of individual sounds. For instance, the sound of a tom tom or cymbal survives translation through numerous mappings, some of them drastic. Only the flute still retains any part of its tonal identity when mapped into the tom tom spectrum of Fig. 13.11. Sounds like the guitar and clarinet can be changed somewhat without losing their tonal quality, surviving the transformation into the n-tet spectra but not into the more drastic tom tom spectrum. Other sounds, like the violin, are fragile, unable to survive even modest transformations. Thus, not all mappings preserve the perceptual wholeness of the original instruments, and not all instruments are equally robust to spectral mappings.

Using the RIW spectral mapping technique of the previous sections, the attack portion is mapped separately from the looped portion, which tends to maintain the character of the attack. As the envelope and other performance parameters are also maintained, changes in the timbral quality are likely due primarily to changes in the spectrum of the steady-state (looped) portion of the sound.

As a general rule, the change in timbral quality of instruments with complex spectra tends to be greater than instruments with relatively simple spectra. The flute and tom tom have fairly simple spectra (only four or five spectral peaks) and are the most robust of the sounds examined, retaining their integrity even under extreme spectral maps. Sounds with an intermediate number of significant spectral peaks, such as the guitar, bass, and trumpet, survive transformation through modest spectral mappings. In contrast, sounds like the violin and oboe, which have very complex spectra, are the most fragile sounds encountered, because they were changed significantly by a large variety of spectral mappings.

Perhaps the most familiar 'spectral mapping' is transposition, which modulates all partials up or down by a specified amount. As is well known, pitch transposition over a large interval leads to distortions in tonal quality. For instance, voices raised too far in pitch undergo "munchkinization." It should not be surprising that other spectral maps have other perceptual side effects.

13.4.2 Timbral Change

Is there a way to quantify the perceived change in a tone?

Even a pure sine wave can change timbre. Low-frequency sine waves are "soft" or "round," and high-frequency sine waves are "shrill" or "piercing." Thus, one aspect of timbral change is frequency dependent, which may be responsible for timbral changes caused by transposition. A second element of timbral change is the familiar notion that tonal quality changes as the amplitudes of the (harmonically related) partials change. This is likely responsible for the timbral differences between (say) a clarinet and a flute playing the same pitch. Spectral mappings suggest a third aspect of timbral change, that modification of the internal structure of a sound (i.e., a change in the intervals between the partials) causes perceptual changes in the sound. Depending on the spectral mapping (and the partials of the sound that is mapped), this may involve the introduction of (or removal of) inharmonicity.

Clearly, any measure of timbral change must account for all three mechanisms. It is reasonable to hypothesize that perceptions of change are:

(i) Proportional to the amount of transposition
(ii) Proportional to the change in amplitudes of the partials
(iii) Proportional to the change in the frequencies of the partials
(iv) Proportional to the decrease (or increase) in harmonicity (i.e., proportional to the change in tonalness)

Some general trends are suggested. Frequency shifts in a uniform direction (such as those of a stretched map, or in a transposition mapping) may not be as damaging to timbral integrity as those that shift some partials higher and others lower (like the 11-tet mapping). Sounds with greater spectral complexity (like the oboe) seem to undergo larger perceptual changes than simpler sounds like the flute.

To minimize the amount of perceptual change, the mapping T should be defined so that all slopes are as close to unity as possible, that is, so that the mapping is as near to the identity as possible, still consistent with the desire to minimize dissonance. For instance, when specifying timbres for n-tone octave-based equal temperaments, it is reasonable to place the partials at frequencies that are multiples of $r = \sqrt[n]{2}$ to ensure that local minima of the dissonance curve occur at the appropriate scale steps. A good rule of thumb is to define the mapping by transforming partials to the nearest power of r. Thus, an 11-tet timbre may be specified by mapping the first harmonic to r^{11} $(= 2)$, the second harmonic to r^{17} (≈ 3), the third harmonic to r^{22} $(= 4)$, and so on, as given in Fig. 13.1. Analogous definitions of timbres for scales between 5 and 23 are given in Table 13.2. The spectrum defined by

$$\begin{array}{cccccccccccc} f & 2f & 3f & 4f & 5f & 6f & 7f & 8f & 9f & 10f & 11f & 12f \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ fr^{p_1} & fr^{p_2} & fr^{p_3} & fr^{p_4} & fr^{p_5} & fr^{p_6} & fr^{p_7} & fr^{p_8} & fr^{p_9} & fr^{p_{10}} & fr^{p_{11}} & fr^{p_{12}} \end{array}$$

is an induced spectrum[5] for n-tet, where f is the fundamental, $r = \sqrt[n]{2}$, and the exponents p_i take on values from the nth row of Table 13.2.

Table 13.2. Definitions of the "nearest" induced spectra consonant with n-tone equal-tempered scales.

Steps per	Partials											
Octave	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	p_{11}	p_{12}
5	0	5	8	10	12	13	14	15	16	17	17	18
6	0	6	10	12	14	16	17	18	19	20	21	22
7	0	7	11	14	16	18	20	21	22	23	24	25
8	0	8	13	16	19	21	22	24	25	27	28	29
9	0	9	14	18	21	23	25	27	29	30	31	32
10	0	10	16	20	23	26	28	30	32	33	35	36
11	0	11	17	22	26	28	31	33	35	37	38	39
12	0	12	19	24	28	31	34	36	38	40	42	43
13	0	13	21	26	30	34	36	39	41	43	45	47
14	0	14	22	28	33	36	39	42	44	47	48	50
15	0	15	24	30	35	39	42	45	48	50	52	54
16	0	16	25	32	37	41	45	48	51	53	55	57
17	0	17	27	34	39	44	48	51	54	56	59	61
18	0	18	29	36	42	47	51	54	57	60	62	65
19	0	19	30	38	44	49	53	57	60	63	66	68
20	0	20	32	40	46	52	56	60	63	66	69	72
21	0	21	33	42	49	54	59	63	67	70	73	75
22	0	22	35	44	51	57	62	66	70	73	76	79
23	0	23	36	46	53	59	65	69	73	76	80	82

[5] The n-tet spectrum that lies closest to a harmonic spectrum.

13.4.3 Related Perceptual Tests

One way to investigate timbral change is to gather data from listener tests and apply a multidimensional scaling technique as in [B: 139]. For instance, Grey and Gordon [B: 63] swapped the temporal envelopes of the harmonics of instrumental tones and tested listeners to determine how different the modified sounds were from the originals. Such a study could be conducted for sounds formed from various spectral mappings, giving a quantitative way to speak about the degree to which sounds retain their integrity under spectral mappings. The clustering technique used by Grey and Gordon found three dimensions to the sounds, which were interpreted as a spectral dimension, a dimension that represents the amount of change in the spectrum over the duration of the tone, and a dimension determined primarily by the "explosiveness" or abruptness of the attack. Sounds that undergo modest spectral mappings are likely to change in the first dimension and to remain more or less fixed in the latter two. Instrumental sounds that are mapped so as to be consonant with 11-tet (say) sound far more like the original instrumental samples than they sound like each other. An interesting question is whether the spectrally mapped sounds might cluster into a "new" dimension.

The sound examples of this chapter suggest caution in the interpretation of results (such as the above), which rely on listening tests that lack musical context. Taken in isolation, 11-tet mapped trumpet sounds are very similar to harmonic trumpet sounds and thus should cluster nicely with harmonic trumpet timbres. But in a 12-tet musical context, the 11-tet trumpet will sound out of tune, for instance, when it is played in concert with harmonic instruments. Similarly, the harmonic trumpet will sound out of tune when played in 11-tet in an ensemble of 11-tet instruments. In this contextual sense, similarly mapped instruments should tend to cluster separately from harmonic instruments.

13.4.4 Increasing Consonance

Much of the current xenharmonic music is written in just intonations and other scales that are closely related to harmonic timbres. Many of the most popular equal temperaments (7, 17, 19, 21, and 31, for example) contain intervals that closely approximate the intervals of scales related to harmonic timbres. There is, of course, a body of work in tunings like 11-tet that are unrelated to harmonic timbres. Some of these pieces revel in their dissonance, emphasizing just how strange xenharmonic music can be.

Other composers have sought to minimize the dissonance. Bregman [B: 18] reports that the dissonance between a pair of sounds can be reduced by placing them in separate perceptual streams. This implies that musical parts that would normally be dissonant can sometimes be played without dissonance if the listener can be encouraged to hear the lines in separate perceptual streams. Skilled composers can coax sounds into streaming or fusing in several ways,

including large contrasts in pitch, tone color, envelope, and modulation. These techniques have not gone unexploited in xenharmonic music, and they can be viewed as a clever way of finessing the problem of dissonance. They are a solution at the compositional level.

Spectral mappings provide an alternative answer at the timbral level. It is possible to compose consonant music in virtually any tuning by redesigning the spectra of the instruments so that their timbre is related to the desired scale. Of course, it is not always desirable to maximize consonance. Rather, the techniques suggested here are a way to achieve increased contrast in the consonance and dissonance of inharmonic sounds when played in nonstandard tunings. Using spectra that have dissonance curves with minima at the scale steps allows these intervals to be as consonant as possible, thus giving the composer greater control over the perceived consonance.[6] That this is possible even for notorious scales such as 11-tet expands the range of possible moods or feelings in these scales.

13.4.5 Consonance-Based Modulations

Morphing from one set of related scales and timbres to another is a new kind of musical modulation. This might consist of a series of passages, each with a different tuning and timbre. For instance, a piece might begin with harmonic timbres in 12-tet, move successively through 2.01, 2.02, ... , 2.1 stretched octaves, and then return to harmonic sounds for the finale. Such consonance-based modulation can be extremely subtle, as in the modulation from 2.01 to 2.02 stretched. It can also be extremely dramatic, because it involves the complete timbre of the notes as well as the scale on which the notes are played. Alternatively, such modulations might move between various n-tet structures. By carefully choosing the timbres, the "same" instruments can play in different tunings and the dissonance can be tightly controlled.

It is also possible to morph from one spectrum to another in the evolution of a single sustained sound. This can be done by partitioning the waveform into a series of overlapping segments, calculating a Fourier transform for each segment, applying a different spectral mapping to each segment, and then rejoining the segments. Such consonance-based morphing of individual tones can be used to smooth transitions from one tuning/timbre pair to another, or it can be used directly as way to control timbral evolution.

At a point when the mapping becomes too severe, individual notes can lose cohesion and fission into a cluster of individually perceptible partials. Bregman [B: 18] suggests several methods of tonal manipulation that can be used to control the degree to which inharmonic tones fuse. Simultaneous onset times and common fluctuations in amplitude or frequency contribute

[6] It is easy to increase the dissonance by playing more notes or more tightly clustered chordal structures; the hard part is to decrease the dissonance without removing notes or simplifying the spectra.

to fusing, whereas independent fluctuations tend to promote fissioning. These can be readily used as compositional tools to achieve a desired amount of tonal coherence. For instance, a sound can be "modulated" from perceptual unity into a tonal cluster and then back again by judicious choice of such tools.[7] As spectral maps directly affect the amount of inharmonicity of a tone, a series of spectral maps can be used to approach or cross the boundaries of tonal fusion in a controlled manner.

Another form of modulation involves the boundary between melody and rhythm. For instance, when the cymbal of sound example [S: 92] is played using the original sample, it is primarily useful as a rhythm instrument. When the same sound is transformed into a harmonic spectrum, it can support melodies and harmonies. Consider a series of spectral mappings that smoothly interpolate between these two. At some point, the melodic character must disappear and the rhythmic character predominate. Careful choice of spectral mapping allows the composer to deliberately control whether the sound is perceived as primarily unpitched and rhythmic or as primarily pitched and harmonic, and to modulate smoothly between the two extremes.

13.5 Summary

Most of the sounds of the orchestra (minus certain members of the percussion family) and most of the common sounds of electronic synthesizers have harmonic spectra. As the tonal quality of sounds is not destroyed under many kinds of spectral mappings, whole orchestras of sounds can be created from inharmonic spectra. These sounds can retain much of the character of the sound from which they were derived, although they are not perceptually identical. For example, 11-tet sounds were created that clearly reflect their origin as guitar and flute samples. These are clearly perceived as instrumental in nature, and they can be played consonantly in 11-tet.

It is not necessary to abandon the familiar sound qualities of conventional musical instruments to play in unusual scales. The spectral mappings of this chapter provide a way to convert a large family of well-established, musically useful sounds into timbres that can be played consonantly in a large variety of scales. Musical tastes change slowly, and it can be difficult for audiences to appreciate music in which everything is new. The creation of "familiar" sounds that can be played in unusual scales may help to ease the transition to music not based on 12-tet.

Alternatively, extreme spectral mappings can be used to generate genuinely new sounds using familiar instrumental tones as raw material. When played in the related scales, these tend to retain familiar musical features such as consonance even though the timbres and intervals of the scale are unfamiliar.

[7] *Inharmonique* by Risset [D: 36] explores this type of modulation using an additive synthesis approach.

Spectral mappings can also be used to transform inharmonic sounds (such as certain cymbals and drums) into harmonic equivalents. Using these sounds, it is possible to play familiar chord patterns and melodies using this new class of harmonic percussion instruments. Consonance-based spectral mappings make it possible to explore a full range of tonal possibilities for many different spectra.

14

A "Music Theory" for 10-tet

Dissonance curves provide a starting point for the exploration of inharmonic sounds when played in unusual tunings by suggesting suitable intervals, chords, and scales. This chapter makes a first step toward a description of 10-tet, using dissonance curves to help define an appropriate "music theory." Most previous studies explore equal temperaments by comparing them with the just intervals or with the harmonic series. In contrast, this new music theory is based on properties of the 10-tet scale and related 10-tet spectra. Possibilities for modulations between 10-tet "keys" are evident, and simple progressions of chords are available. Together, these show that this xentonal 10-tet system is rich and varied. The theoretical ideas are demonstrated in several compositions, showing that the claimed consonances exist, and that the xentonal motions are perceptible to the ear.

14.1 What Is 10-tet?

In the familiar 12-tet, the octave is divided into 12 equal-sounding semitones, which are in turn divided into 100 barely perceptible cents. Instead, 10-tet divides the octave into *ten* equal sounding pieces. Each scale step contains 120 cents, which is noticeably larger than a normal semitone. Figure 14.1 shows how 12-tet and 10-tet relate.

Because the 10-tet intervals are unusual, it does not make sense to give them the familiar sharp and flat names: Instead we adopt an "alphabetical" notation in which each successive tone is labeled with a successive letter of the alphabet.[1] Thus, the scale begins with an A note, continues with B, and proceeds alphabetically through the J note.

The 10-tet tuning has no fifth, no third, no major seconds, and no dominant sevenths. The only interval common to both 10-tet and 12-tet (other than the octave) is the 600-cent interval normally called the tritone, augmented fourth, or diminished fifth. This is due to the numerical coincidence that:

[1] Although not an ideal solution to the notation problem, the alphabetical approach has the advantage that it can be readily applied to any tuning system that repeats at regular intervals.

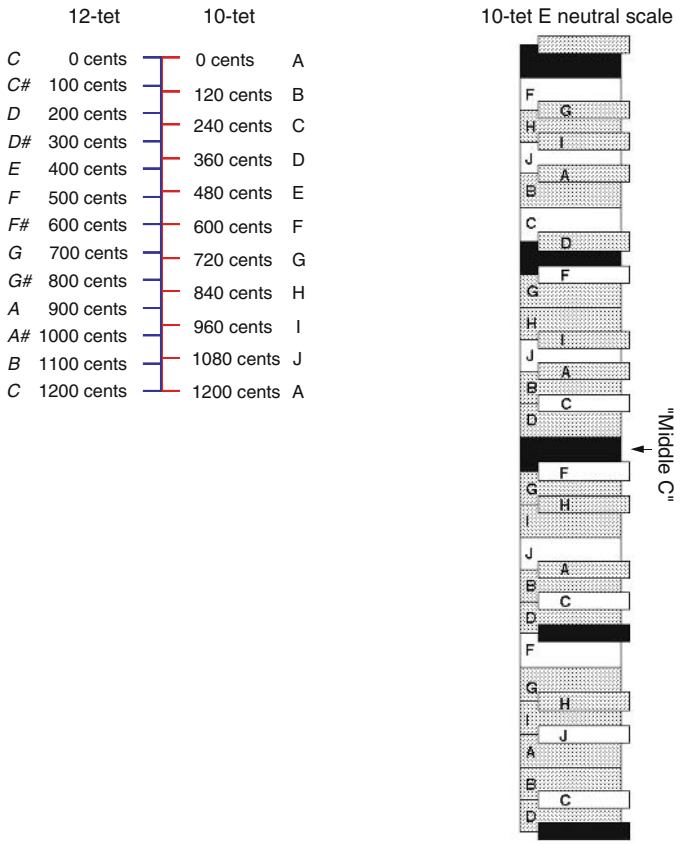


Fig. 14.1. The 10-tet and 12-tet scales have only the octave and tritone in common. When the scale steps of the 10-tet tuning are mapped consecutively along a standard keyboard, the octaves precess (as shown by the blackened keys). The black and gray keys combine to outline the *E*-neutral scale.

$$6 \text{ steps } * 100 \text{ cents } = 5 \text{ steps } * 120 \text{ cents } = 600 \text{ cents}$$

Although there are no major, minor, or seventh chords in 10-tet, there are new "chords" that do not have "real" thirds or fifths. All of the comforting scales and intervals have vanished, replaced by weird-sounding melodic intervals and even stranger xenharmonies. Nothing you learned in music class is true!

14.2 10-tet Keyboard

How shall the 10-tet scale be laid out across the keyboard? Unfortunately, the familiar 12-key-per-octave design is poorly suited to tunings like 10-tet. One

option is to choose a subset of the 12 keys, and to map the 10-tet pitches to only this subset, leaving two extra keys "empty." The primary advantage of this method is that each "octave" of keys still plays an octave. The disadvantage is that the normal flow of 10-tet steps is artificially interrupted by the silent keys.

The keyboard layout I prefer is one that assigns successive notes of the 10-tet scale to successive keys. With this 10-tet keyboard, a 10-tet chromatic scale encompasses only ten steps. If the scale starts at middle C, then it ends at the $B\flat$ key ten steps up or at the D key ten steps down. Thus, each interval normally fingered as a dominant seventh is actually an octave. Figure 14.1 shows how this nonoctave repetition plays out across the keyboard by blackening all E notes. Observe how the sounding octaves precess through the key-octaves at a rate of two keys per octave. This pattern can be exploited without great difficulty, given a bit of practice.

14.3 Spectra for 10-tet

If 10-tet is so cool, why don't more people already use it? The facile answer is that there are no 10-tet guitars, flutes, or pianos, hence no musicians versed to play in 10-tet, and no repertoire for them to perform. But there may be an underlying reason for this lack—that harmonic tones sound out-of-tune (or dissonant) when played in 10-tet. For instance, as shown in Fig. 14.1, the 10-tet interval from E to A is 720 cents. In contrast, a perfect 12-tet fifth is 700 cents. Hence, the 10-tet interval from E to A is likely to be heard as a sharp, out-of-tune 12-tet fifth. The full E neutral chord is even worse.

The problem is not simply that harmonic sounds are dissonant in 10-tet. As we know, the motion from consonance to dissonance (and back again) plays an important role in most music. The problem is that most of the intervals in 10-tet are dissonant, assuming harmonic sounds. It is thus very difficult to achieve the kinds of contrasts needed for tonal motion.

Using the ideas of the previous chapters, it is easy to design spectra for sounds that will appear consonant in the 10-tet intervals.[2] For instance, the dissonance curve for the mapping from a harmonic spectrum

$$\begin{array}{cccccccccccc} f & 2f & 3f & 4f & 5f & 6f & 7f & 8f & 9f & 10f & 11f & 12f \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ f & r^{10}f & r^{16}f & r^{20}f & r^{23}f & r^{26}f & r^{28}f & r^{30}f & r^{32}f & r^{33}f & r^{35}f & r^{36}f \end{array}$$

into a "10-tet spectrum" defined with $r = \sqrt[10]{2}$, is shown in Fig. 14.2. The minima of this curve are aligned with many of the 10-tet scale steps. Intervals such as the 720-cent "sharp fifth" and the 480-cent "flat fourth" need not sound dissonant and out-of-tune when played with sounds that have this spectrum, even though they appear very out-of-tune when played with normal harmonic sounds.

[2] Figure 12.1 on p. 248 contains three such spectra.

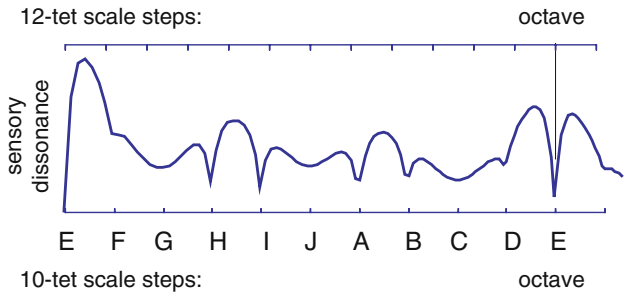


Fig. 14.2. The dissonance curve for a spectrum designed to be played in 10-tet. Minima coincide with many of the steps of the 10-tet scale and not with steps of 12-tet. The notes of the scale are named using the "alphabetical" notation, starting on E.

The above spectral mapping was applied to a sampled guitar, to create the "virtual 10-tet guitar" that is featured in the piece *Ten Fingers* in sound example [S: 102]. The overall impression of *Ten Fingers* is of a strange plucked instrument, like a sitar or a pipa, played in a musical style from an unknown musical tradition.

Close observation reveals that much of this piece centers around the 10-tet interval E to B (seven scale steps) and its inverse from B to E (three scale steps). These intervals are 360 and 840 cents, which are distinct from anything available in 12-tet, and dissonant when played with harmonic sounds. As often occurs, this dissonance is perceived primarily as an eerie out-of-tuneness, as demonstrated in sound example [S: 103], which plays the first few measures of *Ten Fingers* but with the original harmonic sampled guitar rather than with the spectrally mapped 10-tet version. More properly, this should be called "out-of-timbre" or "out-of-spectrum," because the actual tuning is precisely 10-tet. The contrast between examples [S: 102] and [S: 103] is not subtle.

14.4 10-tet Chords

Of course, 10-tet does not have major and minor chords. It does not have real I-IV-V progressions. It does not have a circle of fifths, because it does not really have "fifths." But there are chords, and these chords can be played in sensible musical progressions. These 10-tet sound patterns are just new kinds of progressions.

Dissonance curves suggest where to begin. Figure 14.2 shows that 10-tet scale steps 0, 3, 4, 6, 7, 9, and 10 occur at the narrow minima caused by coinciding partials. These are the most consonant intervals in this 10-tet setting. The most consonant chords are found by drawing the 3-D dissonance curve, which is shown in Fig. 14.3. As usual with such curves, the very highest

peaks (and the deepest valleys) occur near unisons. These create the two irregular far walls. The long bumpy strip along the diagonal is similarly caused by the (near) coincidence of the second and third notes. The most musically interesting areas of the terrain are the three smaller mountainous regions marked A, B, and C.

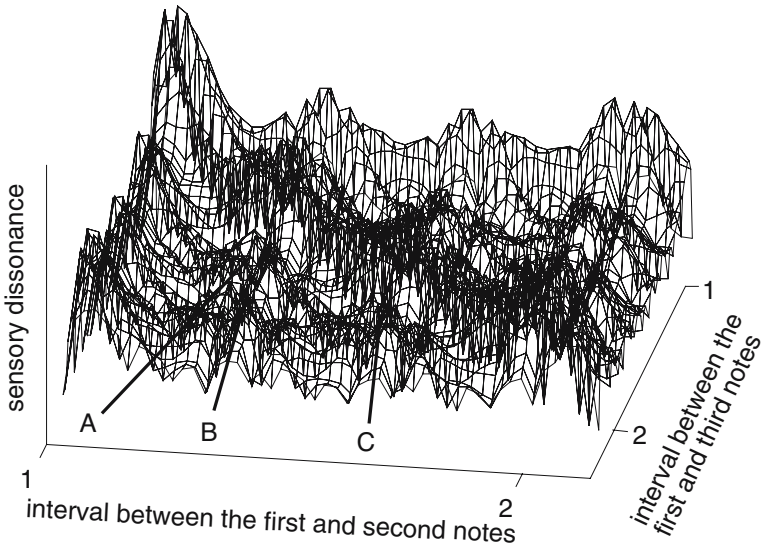


Fig. 14.3. Dissonance curve for three note chords using the spectrum designed for 10-tet has minima that define the most important 10-tet chords. Three regions of interest are indicated.

To get a closer look, the contour plot is drawn in Fig. 14.4, and the axes are labeled in increments of the steps of the 10-tet scale. The left edge and the bottom strip correspond to the two far walls of the 3-D version, whereas the jeweled stripe across the diagonal represents the second and third notes merging together. The three regions of interest are again labeled A, B, and C, and it is apparent that each of these regions actually contains three distinct minima. The intervals in these chords can be read directly from the figure. The chord featured in *Ten Fingers* appears in region C, containing the intervals 1, r^7, and 2. Its complement (the chord containing 1, r^3, and 2) is in region B.

The chords in region A are the most like standard triads. As r^6 is the closest 10-tet interval to a 12-tet fifth, the chord 1, r^3, r^6 is an obvious candidate.

14.4.1 Neutral Chords

Play middle C, the $E\flat$ above, and the $F\sharp$ above that. In the alphabetical notation for 10-tet, these are the E, H, and A notes.

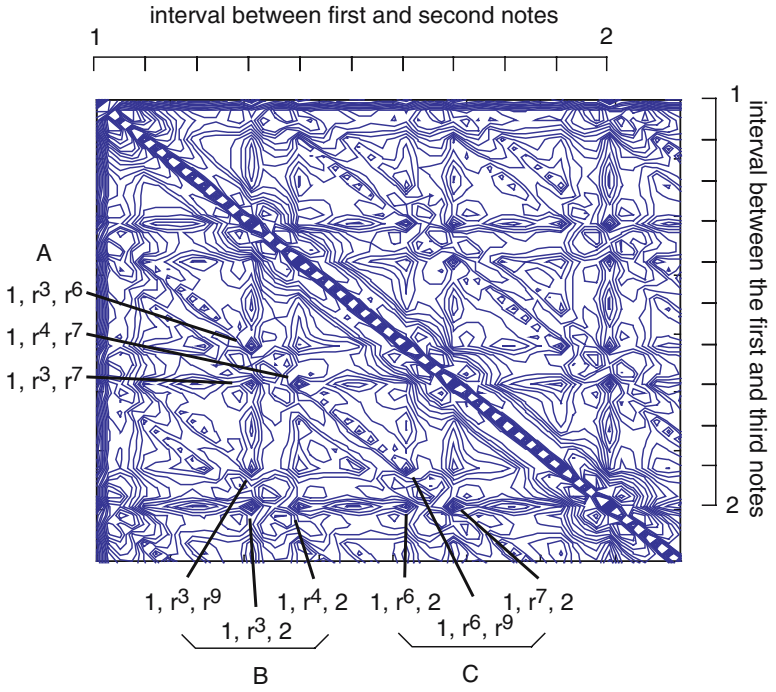
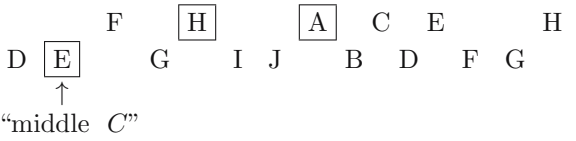
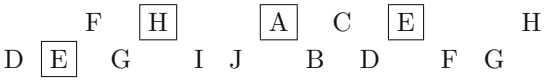


Fig. 14.4. Dissonance curve for three-note chords using the spectrum designed for 10-tet has minima that define the most important 10-tet chords. Three regions of interest are indicated.



Assuming that the timbre is built from the 10-tet spectrum given in the above spectral mapping, this will likely sound smooth, but a bit strange. The chord is completed by closing the octave with the $B\flat$ key above (but not below). This $B\flat$ key is the E an octave above the first E, because it is ten steps up. The complete chord



is called the E *neutral chord*.

Recall that a normal major chord begins on its root (say C), adds the third (the note E four semitones above the root) and then the fifth (the note G three semitones higher) to complete the C major chord C-E-G. In 10-tet,

the neutral chord begins on its root (say E), adds the note that is three 10-tet scale steps higher (the H note), and then the note that is three more 10-tet scale steps higher (completing the E neutral chord E-H-A). Of course, any note can be used as the root. As there are ten different notes, there are ten possible neutral chords.

In 12-tet, chords are called major or minor depending on whether the first interval in the chord is a major third (four semitones = 400 cents) or a minor third (three semitones = 300 cents). The interval used to build the neutral chord in 10-tet is three 10-tet scale steps, which is 360 cents. As 360 is about halfway between the major and minor thirds, it is neither major nor minor: hence the term "neutral."

Refer back to Fig. 14.4. There are three chords in region A that correspond to minima of the dissonance curve that are approximately equally deep. Perhaps there are other interesting chords or theoretical structures that can be built up around the 1, r^4, r^7 chord or the 1, r^3, r^7 chord. Unfortunately, this is not so, because all three are intimately related. For instance, suppose the root of the neutral chord was transposed an octave up, while leaving the other two tones fixed. Then the three tones would be in the relationship r^3, r^6, r^{10}, which is just a relabeling of 1, r^3, and r^7. Similarly, if the upper tone was transposed down an octave, the three tones would be in the relationship 1, r^4, r^7. Thus, all three chords in region A are different inversions of the "same" neutral chord.

14.4.2 Circle of Thirds

There is a very interesting and beautiful chord pattern in 10-tet that is analogous to (but very different from) the traditional circle of fifths.

Observe that by changing only one note, it is possible to modulate from the E neutral chord (containing E-H-A) to a B neutral chord (containing B-E-H). One way to finger this is to simply move the A to a B while holding the E and H constant. Thus, it is possible to move from the E chord

F [H] [A] C E H
D [E] G I J B D F G

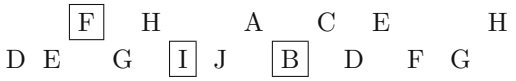
to the B chord

F [H] A C E H
D [E] G I J [B] D F G

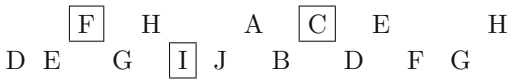
by moving only one finger. But now it is possible to modulate to an I chord (I-B-E) by raising the H note one step.

F H A C E H
D [E] G [I] J [B] D F G

Raising E to F



gives the F neutral chord, and raising B to C



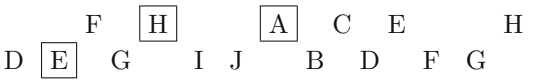
gives the C neutral chord... and so on. After 10 chord changes, the progression has moved

$$\mathrm{E} \to \mathrm{B} \to \mathrm{I} \to \mathrm{F} \to \mathrm{C} \to \mathrm{J} \to \mathrm{G} \to \mathrm{D} \to \mathrm{A} \to \mathrm{H} \to \mathrm{E}$$

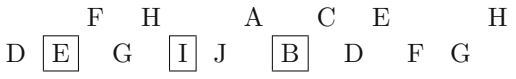
completely around the circle of thirds and back to its starting point. Because the root of each chord in this progression is a neutral third below the previous root, the complete cycle is called the circle of thirds. The song *Circle of Thirds* (sound example [S: 104]) plays around and around this circle of thirds: first fast, then slow, and then fast again.

14.4.3 “I-IV-V”

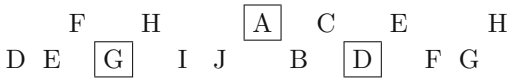
In 10-tet, the nearest interval to a fourth is 480 cents (instead of the familiar 500 cents) and the nearest interval to a fifth is 720 cents (instead of the normal 700 cents.) Thus, a I-IV-V progression is not really possible. But, using the flat fourth and sharp fifth in place of the familiar intervals does lead to musically sensible results. For instance, moving from E to I is as easy as playing



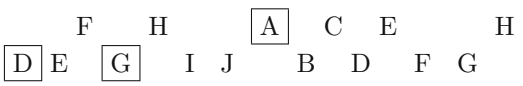
followed by



The A chord, which is only a few keys away, can be fingered either as



or as



These three chords form the basis of *Isochronism* [S: 105].

14.4.4 The Tritone Chord

The tritone, also called the augmented fourth and the diminished fifth, is an interval of 600 cents. It plays a very special role in conventional harmony when it appears in dominant seventh chords: It helps to define the finality of cadences, and it is often used as an "engine" that drives modulation from one key to another. For instance, the typical $V7 \to I$ progression

$$\text{tritone} \left\{ \begin{array}{l} F \to E \\ B \to C \end{array} \right\} \text{major third}$$
$$G \to G$$
$$D \to C$$

contains a tritone that resolves to a major third. Is there a 10-tet analog?

The tritone is the only interval (other than the octave) that is common to both the 10-tet and 12-tet systems. In fact, the tritone can function in much the same ways in the 10-tet system as it does in traditional harmony: It helps to define the finality of cadences and can be used to modulate between keys.

The chord that does this, called *the tritone chord*, is built from a root (say G), the note 5 steps above (B), and the note 3 steps above that (E).[3]

F H A C [E] H
D E [G] I J [B] D F G

This G tritone chord feels as if it wants to resolve. The most natural resolution is to move the lower note of the tritone up one step, the upper note of the tritone down one step, and to leave the third note fixed.

$$E \to E$$
$$\text{tritone} \left\{ \begin{array}{l} B \to A \\ G \to H \end{array} \right\} \text{neutral third}$$

Thus, the G tritone chord resolves to a E neutral chord.

F [H] [A] C [E] H
D E G I J B D F G

So far, the tritone chord has made a nice analogy with the dominant seventh chord of traditional harmony. But there is another kind of tritone chord that is built from a root (say D), the note 5 scale steps above (I) and the note 2 scale steps above (A).

F H [A] C E H
[D] E G [I] J B D F G

[3] Observe that this $5 + 3$ construction leaves only two steps until the octave. Thus, the note does have something of the character of a dominant seventh.

This tritone chord also wants to resolve. The bottom note of the tritone pulls upwards, the middle note of the tritone pushes down, and the third note remains fixed.

$$\text{tritone} \left\{ \begin{array}{c} A \to A \\ I \to H \\ D \to E \end{array} \right\} \text{neutral third}$$

so the (second kind of) D tritone also wants to resolve to the E neutral chord.

F [H] [A] C E H
D [E] G I J B D F G

Thus, in the 10-tet system, there are two different tritone chords, both of which function analogously to the dominant seventh chord of traditional harmony. There are two different ways to approach any given neutral chord, there are two different cadences resolving to any neutral chord, and there are consequently a far greater number of ways to modulate from one 10-tet key to another. So, although the 10-tet system lacks the dichotomy between minor and major chords,[4] it contains richer possibilities of modulation due to the greater number of tritone xentonalities.

14.5 10-tet Scales

The traditional major scale is intimately related to major chords. For instance, the C, F, and G chords contain exactly the notes of the C major scale. Similarly, one can think of building 10-tet scales from the notes of certain 10-tet chords.

One approach is to choose a neutral chord (say E with notes E-H-A) and the two tritone chords that lead to it (G with G-B-E, and D with D-I-A). Collecting all of these notes together gives the 7-note E neutral scale

F [H] [A] C [E] H
D [E] [G] [I] J [B] [D] F G

which is shown spread out across the keyboard in Fig. 14.1 on p. 291. Alternatively, one could begin with the analogs of I-IV-V (for instance, the E, I, and A neutral chords) and define the scale from these notes. This leads to the exact same 7-note scale. Finally, this scale is also the same as the minima of the dissonance curve (Fig. 14.2) with the addition of the G note.

14.6 A Progression

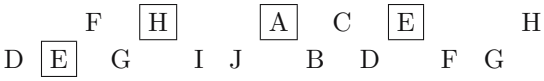
There are many ways to play in 10-tet. The use of 10-tet is not limited to any particular style of music—it is no more *for* jazz than it is *for* rock or

[4] Having only neutral chords.

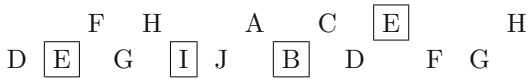
any other style. Think of it as an expansion of tonality. The 10-tet xentonal musical language is not intended to replace the familiar harmonic 12-tet, but to complement it. Lilies are not intended to replace roses, and the world would be a poorer place without either.

This section ends with a simple 10-tet chord pattern that I have grown fond of. It begins by moving back and forth between E and I. Then there is a short D tritone, followed by a G tritone, and finally a resolution back to E. Then repeat. It is simple, and maybe even a little catchy.

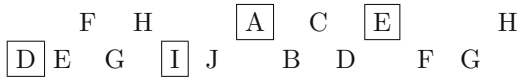
Begin by alternating the E chord



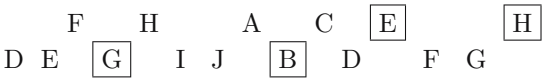
with the I chord.



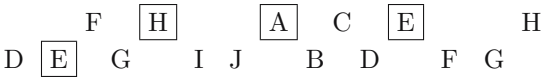
Then, the resolution begins with a D tritone chord (the second kind),



moves through the G tritone chord (the first kind)



and finally resolves back to E.



This chord pattern is used throughout *Anima* [S: 106], which also demonstrates that it is possible to sing in 10-tet.

14.7 Summary

Dissonance curves for a 10-tet spectrum were helpful in pinpointing useful intervals, chords, and scales. These can be combined in numerous ways into coherent patterns that, although unfamiliar, are perceivable as sensible xentonal progressions. “Neutral” chords occupy a place in 10-tet somewhat analogous to major chords in 12-tet, and two kinds of “tritone” chords can be used as engines of modulation and resolution, analogous to the familiar dominant

seventh chord. These are just a start; it would be impossible to exhaust an intricate system like 10-tet in a single chapter.

There is nothing magic about 10-tet, nor about this particular spectrum for 10-tet. Each of the n-tet tunings has its own kinds of related spectra, its own intervals and scales, its own chords and chord progressions, and its own character and moods.[5] There are new patterns of sound that can subtly (and not so subtly) entice and entrance, repel, and repulse. Unlike 12-tet, where it is virtually impossible to create a genuinely new chord pattern or scale, almost nothing is known about these n-tet worlds. Similarly, other divisions of the octave (and divisions of non-octaves as well) have their own timing, intervals, consonances, dissonances, and their own music theories. Each tuning has its own song to sing.

[5] Darreg [B: 36] was the first to point out the existence of these moods.

15

Classical Music of Thailand and 7-tet

Thai classical music is played on a variety of indigenous instruments (such as the xylophone-like renat *and* pong lang*) in a scale containing seven equally spaced tones per octave. This chapter shows how the timbres of these instruments (in combination with a harmonic sound) are related to the 7-tet scale, and then explores a variety of interesting sounds and techniques useful in 7-tet.*

15.1 Introduction to Thai Classical Music

Thai culture has been in contact with other civilizations for centuries. Thai music and instruments reflect influences from China, Indonesia, and India, as well as influences from the indigenous Khmer, who were conquered when the Thai invaded from southern China. The primary ensembles in Thai court music are a kind of percussion orchestra containing wooden xylophones (the *renat ek*, the lower pitched *renat thum*, the *pong lang*), gong-circles reminiscent of Javanese bonangs, melody instruments such as the *pi*, a multiple reed aerophone, the zither-like *jakeh*, and a variety of drums and cymbals.

Morton [B: 119] describes the music with evocative mixed metaphors:

> The sound of traditional Thai ensemble music might be likened to a stream... here and there little eddies and swirls come suddenly to the surface to be seen momentarily, then to disappear as suddenly... the various threads of seemingly independent melodies of the instruments bound together in a long never-ending wreath.

Morton is describing the technique of *polyphonic stratification* or heterophonic layering of parts in which variations of a single melody are played simultaneously on a number of different instruments. Some play faster, some slower, some syncopated, and some with elaborate ornamentation.

One striking aspect of traditional Thai music is that it is played in a scale that is very close to 7-tet. In the liner notes to [D: 12], Sorrell comments:

> Theoretically, the Thai scale has seven equidistant notes, which means that the intervals are "in the cracks" between our semitone and whole tone, and are equal, though in practice some are more equal than others!

A number of recordings of Thai music are currently available. *Instrumental Music of Northeast Thailand* [D: 45], *Classical Instrumental Traditions:*

Thailand [D: 9], and *Thailand-Ceremonial and Court Music* [D: 39] give an overview of the instrumental techniques, whereas *Sleeping Angel* [D: 12] and the *Nang Hong Suite* [D: 13] mix traditional music with modern music in both traditional and nontraditional styles.

This chapter explores the relationship between the 7-tet scale of Thai classical music and the timbres of the traditional instruments. As will be shown, two different timbres (that of an ideal bar like the renat and a harmonic sound) combine to create a dissonance curve that has minima at many of the 7-tet scale steps. Later sections show how to create "new" instrumental timbres with analogous spectra, and explore some compositional techniques for 7-tet.

15.2 Tuning of Thai Instruments

How close is the actual tuning of Thai instruments to the theoretical 7-tet scale? Many traditional Thai pieces begin with a musical figure played by the renats alone. This isolates the sound of the renat and makes it possible to measure the tuning with reasonable accuracy directly from musical recordings. The xylophone-like renat is ideal for this because it is a fixed pitch instrument unlike the aerophones and stringed instruments, whose pitches may vary each time a note is played.

The somewhat tedious is illustrated in sound example [S: 108], which begins with the first ten seconds of *Sudsaboun* from [D: 39], up to the point where the pi enters. Each of the seven notes present in this introduction are then separated (by a kind of audio cut-and-paste) and played individually. The pitch is determined by finding the sine wave that has the same pitch as the individual notes (recall that, for inharmonic instruments, this is how pitch is defined). The sound example alternates each struck note of the renat with the appropriate sine wave, and the frequencies for each are recorded in Table 15.1. These are then translated into cents (equating the lowest note with 0 cents) for comparison with the theoretical 7-tet scale.

Table 15.1. Tuning of the renats in *Sudsaboun* from [D: 39].

Note	Frequency (Hz)	Cents	7-tet
1	307	0	0
2	337	161	171
3	375	346	343
4	416	526	514
5	456	686	686
6	505	862	857
-	-	-	1028
7	614	1200	1200

By listening carefully to sound example [S: 108], it becomes clear that each of the renat strikes is not really a single note; rather, it is two notes being struck at an octave interval.

15.3 Timbre of Thai Instruments

The *pong lang* is a wooden xylophone-like instrument from Northeast Thailand. Like the boat-shaped *renat*, it is tuned to (approximately) 7-tet. The modes of vibration of keys of the pang lang and renat, like those of the Javanese gambang (recall Fig. 10.9), are very close to those of an ideal bar.[1] Figure 15.1 shows the spectrum of the pong lang taken from the introduction to *Lam Sithandon* on [D: 45].

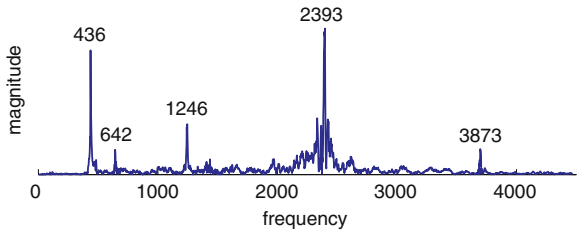


Fig. 15.1. The spectrum of a typical lower register strike of a pong lang has four partials close to those of an ideal bar.

The four largest partials compare closely to those of the ideal bar:

frequency Hz:	436	642	1246	2393	3873
ratio:	f	$1.47f$	$2.85f$	$5.48f$	$8.88f$
ideal bar:	f	–	$2.76f$	$5.4f$	$8.9f$

The spectra of higher pitched notes have less prominent higher partials: The partial near $8.9f$ disappears completely, and the partial near $5.4f$ is often greatly attenuated. The partial at 642 Hz (near $1.47f$) is somewhat anomalous. It occurs in several (but not most) of the spectral measurements of the pong lang but none of the renat spectra.

Section 6.7 shows how dissonance curves can be drawn when two sounds with nonidentical spectra are played. Combining the spectrum of an ideal bar (an idealized renat) with a harmonic sound G containing six partials (such as might result from the pi or jakeh) gives the dissonance curve shown in Fig. 15.2.

[1] The spectrum of the ideal bar is discussed in Chap. 2 (see p. 23 and Fig. 2.7), and scales for the ideal bar are shown in Fig. 6.11 on p. 115.

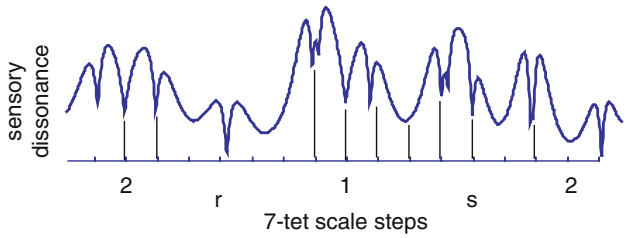


Fig. 15.2. An ideal bar and a harmonic sound with six partials generate a dissonance curve with many minima close to the steps of 7-tet, which is shown for comparison.

This dissonance curve has minima at or near all of the steps of the 7-tet scale, except for the fifth step (the nearest minimum to 1.64 is at 1.62, but it is one of the broad flat minima):

minima s	1.0	1.09	1.21	1.35	1.49		1.80	
minima r	1.0	1.11					1.81	2.0
7-tet ratio	1.0	1.10	1.22	1.35	1.49	1.64	1.81	2.0
7-tet cents	0	171	343	514	686	857	1028	1200

Hence this dissonance curve provides a concrete correlation between the spectrum of the traditional xylophone-like instruments and the 7-tet Thai scale.

As is obvious from even casual listening, Thai classical music is stylistically very different from Western music. It does not contain "harmonies" or "chords" in the Western sense. Rather, it is built linearly by juxtaposing a number of melody lines simultaneously. Often there is a single underlying melodic pattern that no single musician actually plays; the melody is stated (and restated with many kinds of variations) in a collective performance. Morton [B: 119] comments about the use of consonance and dissonance in Thai music:

> The motor power driving this type of music forward is the alternation of relative consonance at structural points of unison (or octaves) with relative dissonances between those points, through the idiomatic treatment of the lines.

How are these variations in consonance and dissonance achieved without harmony or chords? The various melodic lines overlap each other in very complex ways, and thus many different notes occur simultaneously. These clusters of notes clearly have different amounts of sensory dissonance, and this may be one source of the driving power Morton perceives in the music.

As the dissonance curve in Fig. 15.2 shows, the instruments can provide a range of consonances and dissonances as they combine the spectrum of an idealized xylophone with a harmonic spectrum. As more notes are added, the differences can be even more dramatic. To investigate this, Figs. 15.3 and

15.4 draw contour plots of the dissonance surfaces for three simultaneously sounding notes. These are analogous to the contour plots of Fig. 6.21 on p. 129.

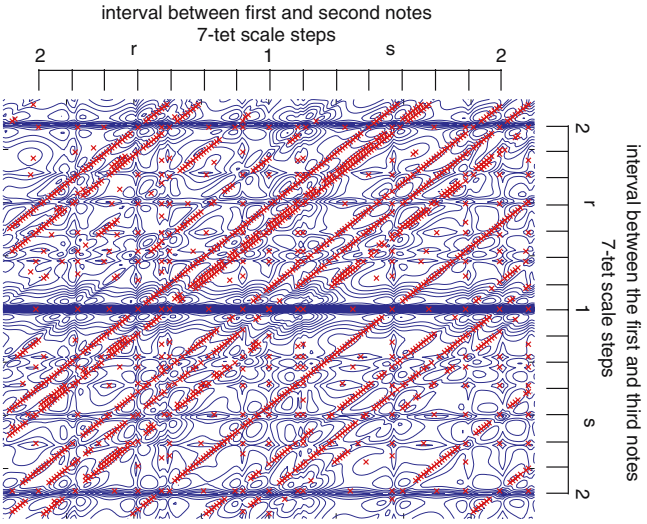


Fig. 15.3. This contour plot of a dissonance surface assumes three notes. The fixed note has a harmonic spectrum, the second has the spectrum of the ideal bar, and the third is harmonic. Minima of the dissonance curve occur at many of the scales steps of 7-tet, which is shown for reference on both axes. The x's represent locations where minima occur.

Dissonance surfaces are drawn assuming three notes, each with known spectrum. One note is held fixed, and the other two vary over a range of two octaves, from an octave below the fixed note to an octave above. As there are two different timbres to consider (that of the ideal bar and a harmonic spectrum), there are four possible surfaces depending on which spectra are assigned to which notes. In Fig. 15.3, for instance, the fixed note is harmonic, the second has the spectrum of the bar, and the third is harmonic. In Fig. 15.4, the fixed note is again harmonic, whereas the second and third both have the spectrum of the bar.[2]

The prominent horizontal stripe in Fig. 15.3 reflects the degenerate case where the first and third notes are tuned the same (in an interval of a uni-

[2] There are two other possibilities, and the corresponding figures are in `pdf` form on the CD in the folder `pdf/`. In the figure in `1bar2harm3bar.pdf`, the fixed note has the spectrum of the bar, the second note is harmonic, and the third has the spectrum of the bar. In the file `1bar2harm3harm.pdf`, the fixed note has the spectrum of the bar, whereas the other two are harmonic. These figures are qualitatively like Figs. 15.3 and 15.4, showing minima at many "chords" with intervals drawn from 7-tet.

son), and this gives (to close approximation) a copy of the one-dimensional dissonance curve in Fig. 15.2. Similarly, the horizontal stripes at $r = 2$ and $s = 2$ depict the situation where the two harmonic tones form octave intervals, again replicating the one-dimensional dissonance curve. In Fig. 15.4, the prominent diagonal stripe represents the degenerate case where the second and third notes (with identical spectra) are tuned the same and the stripe again repeats the one-dimensional dissonance curve.

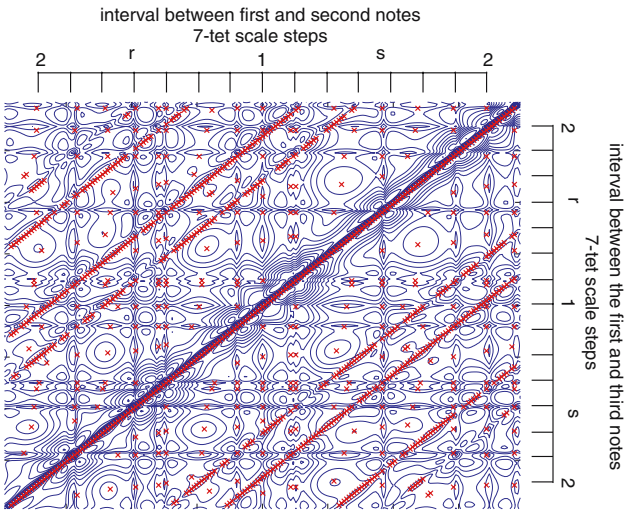


Fig. 15.4. This contour plot of a dissonance surface assumes three notes. The fixed note has a harmonic spectrum, and the two varying notes have the spectrum of the ideal bar. Minima of the dissonance curve occur at many of the scales steps of 7-tet, which is shown for reference on both axes. The x's represent locations where minima occur.

Far more interesting are the deep isolated minima that occur throughout the figures. For example, on Fig. 15.3, locate the fourth scale step between the first and second notes (the tick mark just below the letter r on the horizontal lattice). Looking down the graph reveals minima (marked by x's) at or near more than two-thirds of the scale steps. Similarly, many other columns (and rows) in both figures show a large number of highly consonant chords (more properly, three-note clusters) that use intervals in the 7-tet scale.

Let's oversimplify. Figures 15.3 and 15.4 show that, to a first approximation, almost any three-note cluster in 7-tet is reasonably consonant. So the contrast between consonance and dissonance that drives Thai music is unlikely to be caused by differences in the chordal structure. For example, numbering the notes of the 7-tet scale numerically, the dissonance of note clusters such

as $\begin{pmatrix}5\\3\\1\end{pmatrix}$, $\begin{pmatrix}6\\4\\1\end{pmatrix}$, and $\begin{pmatrix}5\\4\\1\end{pmatrix}$ does not differ greatly. Reinforcing this, there is no notion in Thai music theory that specific combinations of notes perform specific functions; thus, $\begin{pmatrix}5\\3\\1\end{pmatrix}$ does not necessarily play a different role than $\begin{pmatrix}6\\4\\1\end{pmatrix}$. This is very different from music of the common practice period where, for example, the tonic, dominant and subdominant serve highly prescribed and conventionalized roles.

This suggests that the contrast driving Thai music must arise in some other way. One possibility grows out of the layering of melodic lines (the polyphonic stratification). Consider a simplified example of a melody that repeats four notes 1 2 3 1 at three levels separated by a factor of two in tempo. The slowest layer performs the melody once during the time the middle layer plays it twice. Meanwhile, the fastest layer repeats the same melody four times. This can be represented schematically as

$$\begin{array}{ll} \text{fastest level:} & 1\ 2\ 3\ 1\ 1\ 2\ 3\ 1\ 1\ 2\ 3\ 1\ 1\ 2\ 3\ 1 \\ & 1-2-3-1-1-2-3-1- \\ \text{slowest level:} & 1---2---3---1--- \end{array} \tag{15.1}$$

where time proceeds horizontally. The initial three notes in unison are highly consonant. Similarly, the final stroke is consonant because it contains the last stroke of the fastest layer plus whatever sound remains from the 1's in the slower layers. In between is a rising and falling dissonance proportional (more or less) to the number of different notes sounding simultaneously. For this particular pattern, the greatest dissonance would occur at the second stroke (of the slowest layer) where all three different notes occur simultaneously. Thus, even in this highly idealized setting, there is a journey from consonance into dissonance and back again. This is dictated, not by chord placement or differences in dissonance between clusters, but by the temporal layering of the melodic lines.

To investigate this more concretely, the dissonance score[3] in Fig. 15.5 shows the first two minutes of *Lam Sithandon* [D: 45], which uses the "happy sounding *san* mode type" according to the liner notes. The introductory solo, played on the pong lang, is evident in the first large hump in the dissonance that culminates at about 14 seconds. The bulk of the analysis shows a large number of small peaks of varying heights that coincide with the phrase length. Each phrase is performed slightly differently: with different instruments, with different ornamentation, and with different density of orchestration. The drop in the dissonance at 80 seconds coincides with the end of the first major section

[3] Drawn using the method of Sect. 11.1.

and a return to the main theme. As Morton [B: 119] suggests, the relative consonance occurs at points of structural unison, and dissonance increases between.

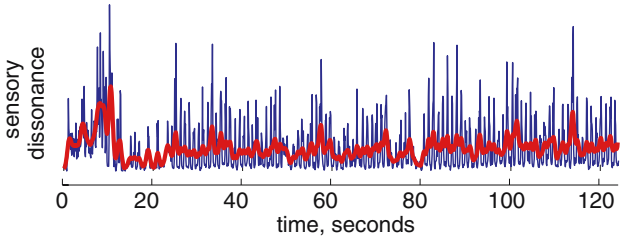


Fig. 15.5. Dissonance score for the first two minutes of *Lam Sithandon*. The dark line averages the raw dissonance calculations over 1 second.

15.4 Exploring 7-tet

Inspired by traditional Thai music, this section explores compositional techniques and sound design strategies for 7-tet. The first section discusses a variation on the spectral mapping techniques of Chap. 13 for the sculpting of a variety of instrumental sounds that have the same spectrum as an ideal bar. Succeeding sections discuss variations on the technique of polyphonic stratification that are applied to several musical compositions that can be heard on the accompanying CD.

15.4.1 Sounds for 7-tet

As the previous sections showed, two kinds of sounds combine to form dissonance curves with minima at steps of the 7-tet scale: harmonic sounds and bar sounds (those with the spectrum of an ideal bar). There is no shortage of interesting harmonic sounds, but there is no obvious source of timbres with the spectrum of a bar other than the bar instruments themselves (xylophone, glockenspiel, renat, gambang, and so on).

In principle, the spectral mapping approach of Sect. 13.2 (refer back to Fig. 13.3 on p. 270) can transform one spectrum into another by choosing a mapping from the source spectrum into the destination spectrum. This implicitly requires that there be the same number of partials in the destination as in the source. But the spectrum of a bar is sparse compared with (say) harmonic sounds; the first four partials of the bar (f, $2.76f$, $5.4f$, and $8.9f$) span the same range of frequencies as the first nine partials of a harmonic sound. A naive mapping like

$$\begin{array}{lccccc} \text{harmonic spectrum:} & f & 2f & 3f & 4f & \ldots \\ & \downarrow & \downarrow & \downarrow & \downarrow & \\ \text{spectrum of bar:} & f & 2.76f & 5.4f & 8.9f & \ldots \end{array}$$

can cause significant oddities in the resulting mapped sounds, more akin to the transformation from a harmonic sound into the spectrum of a tom-tom (sound example [S: 90]) than to the milder transformation into the nearby 11-tet spectrum (as in sound example [S: 86]).

One variation is to transform from the harmonic spectrum to the bar spectrum by mapping only the harmonic partials nearest the desired partial of the bar spectrum:

$$\begin{array}{lcccc} \text{harmonic spectrum:} & f & 3f & 5f & 9f \\ & \downarrow & \downarrow & \downarrow & \downarrow \\ \text{spectrum of bar:} & f & 2.76f & 5.4f & 8.9f \end{array}$$

But what happens to $2f$, $4f$, $6f$, $7f$, $8f$, and $10f$ and above? If they are left unchanged, then the sound is very likely to retain a large part of its harmonic character and it is no longer the kind of sound that is related to the 7-tet scale. Figure 15.6 suggests the simplest approach: to "simplify" the spectrum by removing the extra partials.

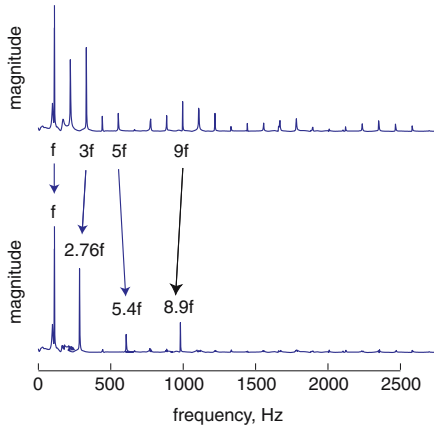


Fig. 15.6. Mapping rich harmonic sounds (such as this spectrum of a guitar pluck) into the spectrum of a bar can be done by simplifying the spectrum to contain only those partials nearest the destination. The resulting sound has (in this case) a bell-like ring.

For example, sound example [S: 109] plays several harmonic sounds and their mapped versions under the transformation of Fig. 15.6. Partials 1, 3, 5, and 9 are mapped using the resampling with identity window (RIW) method of Fig. 13.5, and the remaining partials are attenuated. Three instruments are demonstrated: three different notes of a bouzouki, three different notes of a harp, and a pan flute. Each harmonic tone is followed immediately by the 7-tet spectrally mapped tone, and it is easy to hear the differences. Overall there is some shift of the pitch and the sounds become simpler and cleaner,

more like the strike of a bell than the pluck of a guitar. The next sections place these sounds in their intended 7-tet musical context.

15.4.2 A Naive Approach to 7-tet

The seven equidistant tones of the 7-tet scale (which are compared with 12-tet in Fig. 15.7) lie outside the conventional tonal system. Indeed, with the exception of the octave, there are no familiar intervals. But as there are seven tones in the diatonic scale, perhaps 7-tet can be viewed as a regularization of the major (or minor) scale in which the alternating whole and half steps are equalized. Essentially this suggests a naive mapping

$$\begin{array}{lcccccccc} \text{diatonic scale:} & C & D & E & F & G & A & B & C \\ & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow \\ \text{7-tet scale:} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 1 \end{array} \tag{15.2}$$

which equates the seven equal steps of the 7-tet scale to the seven unequal steps of the diatonic scale.

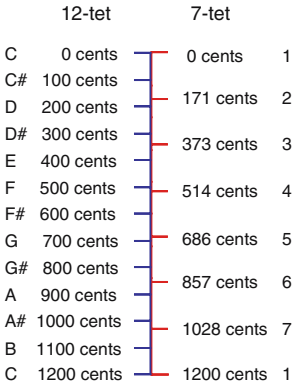


Fig. 15.7. The only interval that appears in both 7-tet and 12-tet is the octave. There is no easy way to exploit diatonic musical intuitions in the 7-tet tuning.

This idea is explored in several sound examples. The "simple theme" of [S: 2] is repeated in [S: 110]; first in 12-tet and then in 7-tet using the identification of notes in (15.2). It is played with harmonic timbres in [S: 110] and with bar timbres in [S: 111]. Scarlatti's K380 sonata (which has already been presented in a variety of historical tunings in sound examples [S: 17] through [S: 22]) is performed in 7-tet in [S: 112]. Both pieces sound flat (in literal and figurative senses) when transformed into 7-tet. Besides the uneasy out-of-tuneness is the problem of uniformity of dissonance: What begins in 12-tet as structural elements (for instance, the motion from I-IV-V-I in [S: 110]) is transformed into a series of tonal clusters with no distinguishable points of rest. Similarly, the melodic motions in [S: 112] appear aimless in 7-tet because they no longer end at a sensible place of repose. Whether the 7-tet version

of K380 is played with harmonic timbres (as in [S: 112]) or with spectrally mapped bar timbres (as in [S: 113]), it regains neither the normality nor the flow of the original. The idea of equating 7-tet to some subset of 12-tet is probably a mistake.

15.4.3 Composing in 7-tet

A wiser direction is to follow those with experience. Thai traditional music does not distinguish the functionality of different 7-tet chords, as [S: 110] through [S: 114] attempt. Rather, it exploits the possibilities of consonance and dissonance in 7-tet by rhythmic means, by superimposing various melodic lines. Denser lines give greater dissonance; sparser lines give greater consonance. Of course, this oversimplifies considerably, but it may be useful in the spirit of finding a reasonable rule of thumb.

Sound examples [S: 115] through [S: 118] explore this rule of thumb for 7-tet in a variety of ways. Inspired by the idea that there is not a large distinction in the dissonance of the various 7-tet chords,[4] *March of the Wheels* [S: 115] begins with a MIDI drum pattern, like the one shown in the piano role notation of Fig. 15.8. In this representation, time moves along the horizontal axis. Each row represents a different instrument (in the general MIDI drum definition, for instance, the row corresponding to $C1$ is the bass drum, $D1$ is the snare, and $F\sharp 1$, $C\sharp 2$, and $D\sharp 2$ are various kinds of cymbals). These are labeled. The relevant idea is to exploit the feature that such MIDI data can represent any kind of sound. In particular, the right-hand side of Fig. 15.8 shows one possible mapping from the MIDI data into a 7-tet scale. Thus, the (original) performance of a drum set is replaced event by event with a 7-tet instrument such as those of [S: 109].

If an interesting drum track is chosen, then there is a good chance that the resulting 7-tet performance will be rhythmically interesting. More variety can be added by changing the notes. Editing by hand is easy (although tedious), and many MIDI sequencers[5] have advanced editing capabilities that can manipulate the data in sophisticated ways. For example, Fig. 15.9 shows a selective randomization of the track in Fig. 15.8 in which the pitch of each note is randomized by a small amount. This preserves the register of the notes; the rhythmic pattern of the bass drum and snare becomes a bass line, and the cymbals are randomized within the more active upper registers. Such formal manipulations are ideal for generating segments or phrases that can be combined to create larger scale pieces. *March of the Wheels* [S: 115] is one such composition. By selective editing, it is easy to create denser and/or sparser sections that reliably increase or decrease the dissonance. Using cut-and-paste methods, whole sections can be constructed. By orchestrating with various timbres, repetitions can be disguised and differences can be unified. The wheel is repetitive, and yet has a clear sense of forward motion.

[4] In 7-tet, all chords are created equal!

[5] Such as Cakewalk for PC and Digital Performer for Mac.

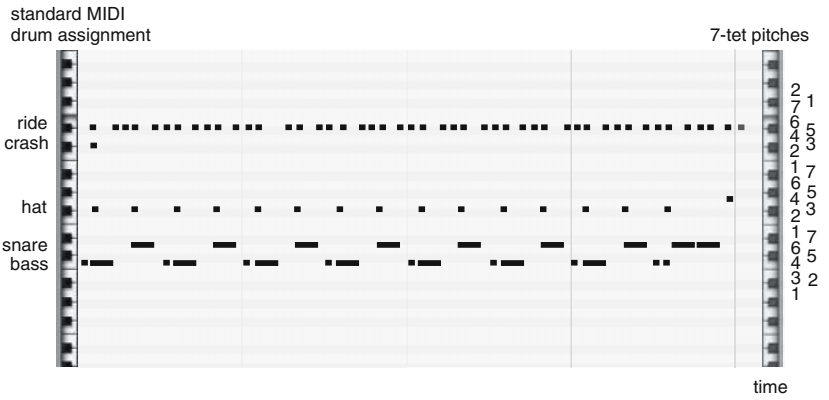


Fig. 15.8. A standard MIDI drum track is shown in piano roll notation. The track need not trigger drum sounds; the right margin suggests a possible mapping of the MIDI events into the seven tones of the 7-tet scale.

There is no need to begin the compositional process with a percussive track. *Pagan's Revenge* [S: 116] starts with a standard MIDI file of one of Niccolò Paganini's (1782–1840) *Caprices* (No. 24 as performed by D. Lovell) from the Classical MIDI archives [W: 4]. The translation from the original 12-tet file to 7-tet was the same as in Figs. 15.8 and 15.9: each 12-tet half step is mapped to a step of the 7-tet scale. Thus, the 7-tet version covers several more octaves than the original because each fifth (seven half steps) is converted into an octave. Even before editing and orchestration, the Caprice is utterly unrecognizable.

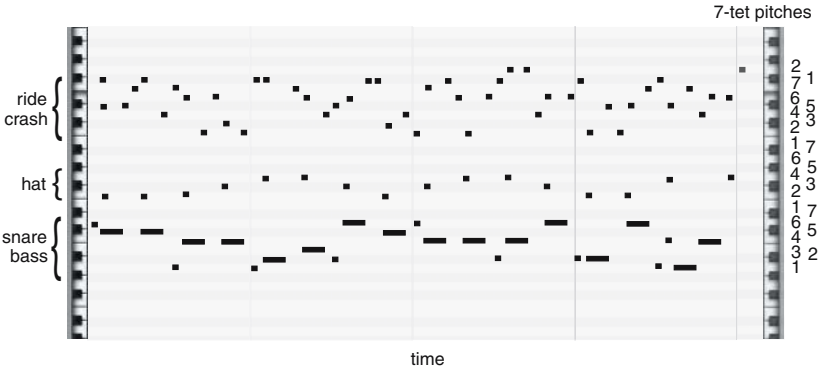


Fig. 15.9. The notes of the standard MIDI drum track in Fig. 15.8 are selectively randomized, creating more interesting "melodic" and "chordal" patterns.

The first half of the standard MIDI file worked well in 7-tet. After deleting the second half, I created "new" material by time-reversing the first half. This process is demonstrated in Fig. 15.10, which takes the first half of the drum sequence in Fig. 15.8, reverses it in time, and concatenates it to the end. This creates a point of rhythmic symmetry (the axis of time symmetry in Fig. 15.10). In *Pagan's Revenge*, the point of symmetry occurs midway through the piece at 1:58, forming a kind of musical palindrome in which the theme proceeds forward and then backward; eventually ending on the first note. The piece is lavishly orchestrated with a variety of sounds with spectra derived from both the bar and the harmonic series. Globally, there is a tension between the frenetic pace and the solemn, near ritual quality and depth of the timbres.

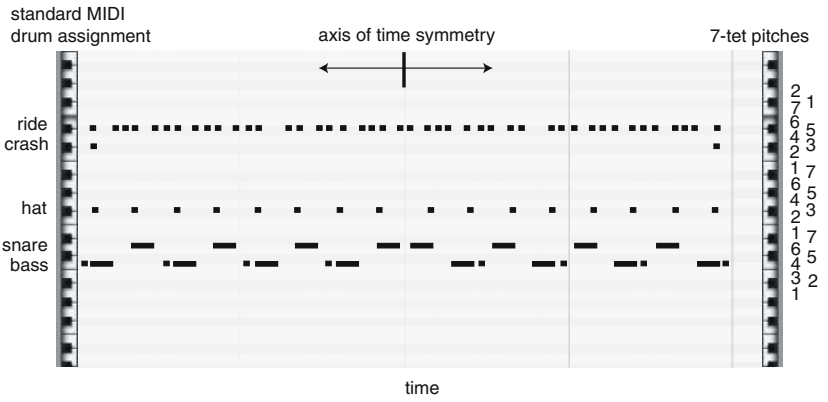


Fig. 15.10. The notes of the first half of the standard MIDI drum track in Fig. 15.8 are reversed in time, creating related but distinct rhythms.

The technique of polyphonic stratification interlocks melodic lines at different tempos, usually separated by a factor of two as schematized in (15.1). A modern technique pioneered by Steve Reich [D: 35] plays a single melodic line simultaneously at slightly different tempos. At first, the two lines are in-phase and the attacks are simultaneous. The faster version soon pulls ahead and anticipates the slower in a sequence of rapid double attacks. Later, the two break apart into a galloping rhythm. At the midpoint, the two are evenly spaced and are perceived as a hocketed melody moving twice as fast as the original tempo. As time proceeds, the same set of perceptions are repeated (although in reverse order) until they eventually resynchronize. This is shown schematically in Fig. 15.11, which indicates several regimes of rhythmic perception.

Nothing Broken in Seven [S: 117] applies this phasing idea in the 7-tet setting by playing the same isorhythmic six note melody throughout. *Phase*

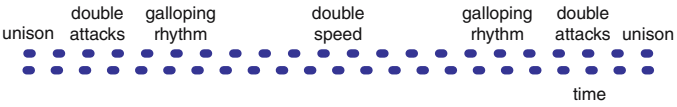


Fig. 15.11. Two rhythms performed at near identical tempos are perceived differently depending on their relative phase.

Seven [S: 118] uses an eight note melody. In both examples, the melody line is played against itself at five different tempos, two of which are speeded up (by 1% and 2%) and two of which are slowed down (also by 1% and 2%). This creates raw material that repeats fully only after several days. In order to create more manageable pieces, selected bits are culled, orchestrated using various 7-tet sounds, and then rejoined. In both cases, although the original pattern is monotonously simple, the result increases and decreases in complexity as the melodies phase against themselves. When there are five phasing lines, a very large number of "different" rhythms are perceptible.

15.5 Summary

The 7-tet tuning of Thai traditional music is related to the sounds of certain Thai instruments (those with the spectrum of an ideal bar and a harmonic spectrum) in much the same way that the tuning of the gamelan orchestras of Indonesia are related to the spectra of the traditional metallophones. The 7-tet musical universe is rich, although it is based on different principles than 12-tet. Chords do not have specified harmonic meanings or functions; rather, clusters of notes create dissonances that are proportional to the density of the sound. The technique of polyphonic stratification, in which different instruments perform various levels of rhythmic diminution over a structural melodic pattern, is the traditional way to create motion from consonance to dissonance (and back again) in the 7-tet system. But there are other ways, some of which are explored and illustrated in the compositions (especially [S: 115] through [S: 118]) of the previous section.

16

Speculation, Correlation, Interpretation, Conclusion

Tuning, Timbre, Spectrum, Scale *began with a review of basic psychoacoustic principles and the related notion of sensory dissonance, introduced the dissonance curve, and then applied it across a range of disciplines. Most of the book stays fairly close to "the facts," without undue speculation. This final chapter ventures further.*

16.1 The Zen of Xentonality

Max Mathews says in an interview in [B: 153]:

> It's clear that inharmonic timbres are one of the richest sources of new sounds. At the same time they are a veritable jungle of possibilities so that some order has to be brought out of this rich chaos before it is to be musically useful.

The organizing principle of this book, the relatedness of spectra and scales expressed in dissonance curves, brings order to this rich chaos by giving the composer control over the amount of sensory consonance or dissonance in a passage. By playing sounds in their related scales, it is possible to realize the entire range from unusual consonances to startling dissonances.

Risset [B: 149] comments:

> the interaction of the components of two (or more) such [inharmonic] tones can give rise to privileged "consonant" intervals that are not the octave and fifth... an intriguing relation exists between the inner structure of inharmonic sounds–which can be arbitrarily composed– and the melodic and harmonic relation between such sounds.

Dissonance curves give concrete form to this "intriguing relation." The spectrum/scale connection provides the same kind of xentonal framework for inharmonic sounds that tonality provides for harmonic sounds. These xentonal systems vary immensely. Some have few partials, few scale steps, and a simple music theory. Others have complex sounds and amazingly complex internal structures.

Although timbres with harmonic spectra are only one kind of sound, they thoroughly dominate the Western musical idiom. Modern electronic musical instruments are now capable of playing inharmonic sounds, and many include

some form of tuning table that allows the user to specify the pitch of the note played by each key. This makes it easy for the musician or composer to retune in any desired way.[1] It is now possible to play "any possible sound in any possible tuning."[2]

When working in an unfamiliar system, the composer cannot rely on musical intuition developed in the context of 12-tet. In 10-tet, for instance, there are no intervals near the familiar fifths or thirds, and it is not obvious what intervals and chords make musical sense. The deepest minima of the dissonance curve (or the dissonance surface) suggest intervals and chords, many of which can be used fruitfully in compositions.

Dissonance curves suggest that the formation of scales and the web of harmony is a collaboration between artistic invention and the timbre (or spectrum) of musical sounds. As the palette of accessible tones expands, the attractiveness of alternative musical scales and tunings increases. Most likely, they will slowly seep into public awareness along with the new timbral palettes afforded by computers, audio signal processing devices, and electronic musical instruments. Composers and musicians will slowly become more adept at moving between xentonal systems, just as they became more adept at modulation through keys when equal temperament first appeared.

Adaptive tunings constantly adjust the pitches of notes to minimize sensory dissonance, freeing music from any fixed scale: tonics wander, chords slither up and down, intervals compress and stretch in a patterned and fascinating way. No doubt there is an undiscovered art to composing with adaptive tunings just as there is an art to composing fugues or canons. As with many of the kinds of manipulations of spectrum and tunings in this book, this technology could be readily built into electronic keyboards, making the annoying calculations transparent to the musician.

16.2 Coevolution of Tunings and Instruments

The harmonic series is related to the just scales; the familiar 12-tet system can be viewed as a practical approximation to these just scales. Similarly, the spectrum of a Javanese bonang in combination with a harmonic tone generates a dissonance curve with minima near the steps of an idealized slendro scale. Pelog scales can be viewed as a result of the spectrum of a saron in combination with a harmonic sound. The 7-tet scale of Thai classical music can be derived by combining the spectrum of an ideal bar (an approximation to the spectrum of the renat) with a harmonic sound, as shown in Chap. 15. In each case, *the scales are related to the spectra of the instruments used by the culture.*

This leads to a musical chicken-and-egg paradox. Which came first, the tuning or the instruments?

[1] For a practical introduction to synthesizer retuning, see Aiken [B: 3].
[2] From the liner notes of Carlos' *Beauty in the Beast* [D: 5].

In biology, the process by which two interdependent species continuously adapt to changes in each other is called *coevolution*. For example, suppose that in order to more effectively catch flies, a species of frog evolves sticky tongues. Then, in order to avoid sticky tongues, a species of flies evolve slippery feet. The spectra of instruments and their tunings may have similarly coevolved. It is easy to imagine a scenario in which the spectrum of a sound influences the tuning of an instrument, which impacts the design of newer instruments, which in turn effects the tuning of the ensemble.

As any group of instruments that are played together must be tuned in some coherent way, once a tuning is established, only compatible new instruments are viable. The Western method of pitch standardization is one possible approach, and the Javanese method of tuning each gamelan ensemble as a distinct musical unit[3] is another. Perhaps this explains why the gamelan tradition has survived and thrived while other equally vibrant forms of music have been absorbed or co-opted. Because gamelan scales and timbres are so different from those of the West, they cannot be effectively combined in the same ensemble.

Perlman [B: 131] calls the belief that there is a natural, biological, or physical reason underlying the use of certain intervals and scales "intonational naturalism," and traces it though history:

> The seventeenth century scientist Christian Huygens conjectured that, since "the Laws of [Western] Musick are unchangeably fix'd by Nature," they should hold not only for the entire earth, but for the inhabitants of other planets as well.

Almost 300 years later, Bernstein [B: 14] echoes this, claiming that the laws of music apply not only pangalactically, but pantemporally as well:

> *All* music–whether folk, pop, symphonic, modal, tonal, atonal, polytonal, microtonal, well-tempered or ill-tempered, music from the distant past or imminent future–all of it has a common origin in the universal phenomenon of the harmonic series.

As we have seen, the harmonic series is by no means "universal." Harmonic sounds are only one kind of common sound; there are as many kinds of sounds as there are distinct kinds of vibrating objects. Musical systems have been built on many of these, and many others are undoubtedly possible.

The counter claim to intonational naturalism, that intervals and scales are purely a cultural construct, might be called "intonational relativism." After demonstrating the foolishness of discussing the gamelan in terms of just intonation and the harmonic series, Perlman [B: 131] examines the Javanese concept of *embat*, which refers to "any particular realization of a tuning system," although it can also refer to the intonational preferences and practices of individuals. Perlman summarizes:

[3] Gamelan instruments are not used separately, and the ensembles are not "mix-and-match."

> embat is a matter of feeling (*rasa*), not number; its source is the human voice, not necessary laws of nature; and it is individual,

echoing the beliefs of gamelan tuners who consider intonation to be a matter "of the heart."[4]

The naturalist vs. relativist debate in intonation resembles the "nature vs. nurture" controversy. The naturalist view claims that there is a physical, biological, acoustical, or psychoacoustical explanation for intervals and scales, whereas the relativist view denies that such an explanation exists. The analysis in *Tuning, Timbre, Spectrum, Scale* does not fit neatly into this classification, because it is neither fully naturalist nor fully relativist. To the extent that (sine wave) dissonance curves are universal across cultures, and to the extent that music exploits the contrast between sensory consonance and dissonance, the analysis is naturalistic. To the extent that particular instruments and tunings have coevolved along distinct paths in different cultures, it is relativistic.

Throughout history, many Eurocentric writers have described the music of other cultures as slowly evolving toward the "higher" Western forms, which are supposedly based on immutable laws of nature and the harmonic series. The fact that related spectra and scales apply cross culturally belies this, because the traditional musical instruments and scales of Indonesia and Thailand can be described in terms of the same "underlying laws" as Western instruments and scales. In fact, because the Asian forms use two spectra (rather than a single one as in the Western tradition), it is tempting to reverse the direction of the evolutionary arrow. *As Western music evolves to include more than one "kind" of sound, it may well take on more of the characteristics of the Asian traditions.*

16.3 To Boldly Listen

Are there limits to the kinds of sounds humans can appreciate as music?

There are obvious limits to perception. A "piece of music" that is never louder than −200 dB is inaudible.[5] The same piece played at 200 dB is not perceived as music, but as pain. A melody that always stays within a single JND of pitch is heard as a single tone. A symphony performed exclusively at megahertz frequencies is indistinguishable from silence. But assuming that such perceptual limits are not exceeded, are there limits to the human ability to appreciate sounds as music? Are there limits to possible musical styles?

The amazing diversity of musical cultures and styles to be found throughout the world shows that any such limits are very broad. The history of musical styles suggests constantly changing sensibilities of rhythmic, melodic, harmonic, tonal, and timbral materials, and it seems undeniable that there are musical styles, undreamed of today, that will develop in the future.

[4] Recall Purwardjito's comments on p. 215.

[5] Although John Cage did not perceive this as a limitation.

The only truly universal aspects of music are those based on biological or perceptual facts. By understanding the human auditory system, it should be possible to differentiate those aspects of music inherent in our nature from those that are learned. There are clear cultural biases toward certain kinds of sounds, certain kinds of rhythmic patterns, particular kinds of scales, but any true limits to appreciation must transcend cultural differences.

A simple analogy may help bring this into perspective. The "ear" (the ear canal, eardrum, oval window, basilar membrane, etc.) is like "hardware" that is relatively invariant from person to person and culture to culture. The "brain" (higher levels of auditory processing) is like programmable "software" that implements cultural conditioning. Those aspects dictated by the hardware are universal, whereas the software is rewritten with each new person in each new generation in each new culture. Thus, aspects of musical style that violate my software are unacceptable to me, but they may well be acceptable to someone from another time, place, or with a different background. On the other hand, aspects that violate the hardware are unacceptable to everyone.

In reviewing the sound examples presented here, there are two kinds of passages that may approach limits: those where the partials will not fuse together, and those where the spectrum is sufficiently mismatched from the tuning.

In the first, the notes have lost their perceptual integrity, each being perceived as two or more separate sounds. "Notes" have become "chords." Some compositions[6] in modern music have begun to exploit the boundary where notes fission and tonal clusters fuse, and it may be possible to learn to appreciate unfused sound masses, although they are not currently used in any common musical style.

In *Plastic City* (audio track [S: 38]), the same theme is played in 2.0, 2.2, 1.9, and 2.1 stretched and compressed tunings, each with related timbres. Although it is difficult for me to listen to the piece with naive ears, many people feel that 2.2 is stretched too far, and that 1.9 is compressed too much. After taking such torturous excursions, many first-time listeners hear the 2.1 stretched section and comment, "now we're back to normal, right?" although of course 2.1 stretched is far from "normal." After repeated exposure, however, the 2.2 and 1.9 sections become less strange, more capable of supporting perceptions analogous to chordal motion, yet each retains its own timbral character.

While recording these sections, a process that requires many listenings, I "heard" the passages as more tonally coherent than I typically do now. Moreover, I have learned to switch between perceptual modes (where I hear the piece as either a sound mass or as notes in a chord), although I have no way of knowing if either of these corresponds to a naive listener's perceptions. This argues against (lack of) fusion being a true limit to appreciation. In a musical culture that used various stretched timbres and tunings, members

[6] For instance, [D: 36] and [D: 8].

might develop such a switching strategy as part of normal listening. That I was able to overcome this aspect of my musical conditioning suggests that certain aspects of the fusion mechanism are part of the software of the brain.

The second candidate for a limit to appreciation is the mismatch between tuning and spectrum. In audio tracks [S: 2] to [S: 5], the same brief passage is played in standard and stretched 2.1 tunings, each with both standard and stretched timbres. When matched (i.e., 2.0 timbres with 2.0 tunings or 2.1 timbres with 2.1 tunings), the passage is inoffensive, if somewhat bland. The two mismatched segments, however, are more strident than inoffensive, more irritating than bland. Most likely this is because they are uniformly dissonant. The driving force behind many styles of music is the motion from consonance to dissonance and back again. In the mismatched versions, no such motion occurs, and so the piece appears static.

Similarly, the 10-tet piece *Ten Fingers* is a fine, if somewhat unusual sounding piece when played with related timbres. Most first-time listeners (in the United States) feel that it must be foreign, maybe "Indian." But when played with standard harmonic sounds, it takes on an out-of-tune character, which is more properly called out-of-spectrum. Even after numerous performances and listenings, it still sounds out-of-kilter, suggesting that the perceptual mechanism responsible for the essential wrongness of the mismatched tuning and spectra (i.e., sensory consonance and dissonance) is at least partially in the hardware of the brain.[7]

Whatever part of such perceptions that are in the hardware of the ear may provide limits to the human ability to appreciate sound passages, pointing toward aesthetic principles that may be directly correlated with a perceptual mechanism.

16.4 New Musical Instruments?

Tuning, Timbre, Spectrum, Scale has shown how several kinds of instruments in several different cultures follow a simple pattern; The instruments play pitches that correspond to minima of an appropriate dissonance curve. When designing and tuning new kinds of musical instruments, it may be advantageous to exploit this idea.

In the simplest case, the instrument will sound with a particular spectrum. The dissonance curve of this spectrum will have certain minima, and the instrument can be tuned to play these pitches. An orchestra of such instruments will then be able to play as consonantly as possible. If there are large intervals in the dissonance curve with no minima, then it may be advantageous to augment the scale with some intermediate pitches so that melodies can be more cogent.

[7] Indeed, recall that the binaural presentation of the original dissonance curve (audio track [S: 12]) can also be interpreted this way.

A slightly more complex scenario is when a new instrument (i.e., one with a "new" spectrum) is to be added to an existing orchestra. In this case, the dissonance curve can be drawn for the two spectra. The new instrument can be tuned to the appropriate minima, but the old instruments may also need to be adjusted for compatibility. This is the coevolutionary process in action.

The inverse problem is trickier. Given a desired spectrum, how can acoustic instruments be designed (or redesigned) so as to have that spectrum?

Strings: Uniform strings have harmonic partials as in a guitar or a piano. However, if the contour of the string is changed, or if the density of the string is not uniform, or if the string is weighted at strategic points, then the partials can deviate significantly from harmonicity. Devising a method for readily specifying the kinds of physical manipulations that correspond to useful spectral deviations is an important first step.

Air Columns: Instruments with a uniform air column make harmonic sounds and play in scales that are essentially overtones of a single fundamental (such as the unfingered scale of a cornet). When the column deviates from uniformity (for example, varying widths or flares or the addition of small air chambers), then the scale will change, but the spectrum remains primarily harmonic. On the other hand, many wind instruments like the saxophone can be played inharmonically using extended techniques such as multiphonics. How to (re)design such an instrument to encourage particular kinds of multiphonics is not obvious. Finding patterned ways to relate physical and spectral changes is an important area for the design of such inharmonic instruments.

Bars and Beams: Whether the bars are fixed at an end, or whether they are free to vibrate at both, bars and beams already have inharmonic partials. The exact placement of these partials is an interesting issue. Answers are available for only a handful of simple geometries.

Others: There are many kinds of oscillators and many kinds of resonators that can be used to create audible vibrations. Finding shapes and topologies that will generate a specific spectrum is no trivial task.

In some cases, modal frequencies can be determined from first principles. Perturbation methods can sometimes be applied. Finite element methods can almost always be applied, but they are not generalizable, because solving one problem does not usually give any insight into the solution of related problems. In short, the design of fine musical instruments is no easier now than it was in ancient times.

16.5 Silence Hath No Beats

Consonance and dissonance are only part of the musical landscape. Even in the realm of harmony (and ignoring musically essential aspects such as melody and rhythm), sensory consonance and dissonance do not tell the whole story. Indeed, progressions that are uniformly consonant tend to be uniformly dull. The distinction between sensory and functional consonance and dissonance is not insignificant. Although they often coincide (the minima of dissonance curves for harmonic timbres agrees with just scales, the dissonance score for the Scarlatti sonata correlates reasonably well with more standard analyses), they often do not. For instance, the functional consonance of a silent phrase is not meaningfully defined; yet silence has the greatest sensory consonance. Such extreme cases highlight limitations of the model.

Any model is based on abstractions that limit the scope of its conclusions. When relating an imprecise understanding of the human organism to a complex cultural activity, when relating an imperfect understanding of the auditory system to the complex behavior called music, limitations are manifest. Even at the simplest levels, much is unknown. For instance, when dealing with inharmonic sounds, the partials may fuse into one perceptual entity, or they may fission into many. Understanding this perceptual dichotomy is not trivial, and our ignorance is not for lack of effort. It underscores the gross nature of the additivity assumption in dissonance calculations; by clustering sounds differently, it is possible to change their apparent dissonance. Unfortunately, quantification of this phenomenon is well beyond the current state of psychoacoustic knowledge.

The model used throughout *Tuning, Timbre, Spectrum, Scale* uses linear combinations of the psychoacoustic data of Plomp and Levelt [B: 141]. Refinements such as the inclusion of masking effects or of amplitude effects[8] would enhance the model. In any case, the conclusions of the model (dissonance curves, surfaces, and scores) are qualitative rather than quantitative. It would be a mistake to place too much trust in small details and little dips in the curves: Only the major features that are readily audible need be taken seriously.

16.6 Coda

In retrospect, a connection between the way musical instruments sound and the way they are tuned seems obvious. Almost 100 years ago, Helmholtz recognized the connection between harmonic sounds and the just intervals of the diatonic scale. Because most Western instruments have primarily harmonic partials, theorists and composers tended to limit their theorizing and composing to musical structures based on this one "kind" of sound. But there are many "kinds" of sounds.

[8] For instance, the Fletcher–Munson curves.

It was not until the advent of electronic musical instruments that it became easy to create a variety of inharmonic sounds and to play them in a variety of scales and tunings. One conclusion is inescapable: Certain scales sound good with some timbres and not with others, and certain timbres sound good in some scales and not in others. *Tuning, Timbre, Spectrum, Scale* proposes a way to understand this relationship: to interpret "timbre" as "spectrum," and to interpret "sounds good" in terms of a measure of "sensory consonance." In this framework, dissonance curves codify those intervals that have the greatest (sensory) consonance *as a function of the spectrum of the sound*. It is now possible to systematically choose a tuning related to a given sound, or to choose a sound that is related to a given tuning. In both cases, the intervals are *in-tune* and *in-spectrum*. Compositions in nonstandard scales can easily enjoy contrasts in consonance and dissonance by proper sculpting of the spectra. Nonstandard sounds can be played consonantly or dissonantly by proper choice of interval.

Many nonwestern musical cultures use inharmonic instruments. In at least two cases (the Indonesian gamelan and the percussion orchestras of Thailand), the same kind of reasoning that relates harmonic sounds to just intonations can be used to relate the tone quality of the instruments to the nonwestern scales. Thus, the sensory dissonance approach enjoys a cultural independence that is rare in musical theories.

Appendices

The appendices contain information that does not fit well within the normal flow of the text.

A. Mathematics of Beats: trigonometric formulas describe how beats occur physically, in contrast to how they are perceived.
B. Ratios Make Cents: formulas (and computer programs) describe how to convert between two of the most common kinds of representations of musical intervals.
C. Speaking of Spectra: subtleties in the calculation of spectra and application of the FFT (Fast Fourier Transform program).
D. Additive Synthesis: a brief overview (and `Matlab` program.)
E. How to Draw Dissonance Curves: a theoretical presentation of how to parameterize dissonance curves and a description of `Matlab` programs that carry out the needed calculations.
F. Properties of Dissonance Curves: formal statements and demonstrations of the various results from Chap. 7 "Related Spectra and Scales."
G. Analysis of the time-domain sensory dissonance model of Sect. 3.6.
H. Behavior of Adaptation: details on the results presented in Chap. 8.
I. Symbolic Properties of $\oplus$-Tables: a method of solving the timbre selection problem, of finding a related timbre for a given tuning.
J. Harmonic Entropy: a measure of harmonicity.
K. Lyrics to Fourier's Song.
L. Tables of Scales: details several historical and gamelan tunings.

A

Mathematics of Beats

A basic trigonometric identity relates the sum of two sine waves to the product of a sine and cosine:

$$\sin(x) + \sin(y) = 2\cos(\frac{x-y}{2})\sin(\frac{x+y}{2}). \tag{A.1}$$

Suppose that two sine waves of the same frequency ω have a constant phase difference ϕ. Then the above identity implies that the sum of the two waves is expressible as

$$\sin(\omega t) + \sin(\omega t + \phi) = 2\ \cos(\frac{\phi}{2})\ \sin(\omega\, t + \frac{\phi}{2}), \tag{A.2}$$

which is a sine wave of frequency ω, amplitude $2\ \cos(\frac{\phi}{2})$, and phase $\frac{\phi}{2}$. When ϕ is near 0, the waves are in phase and the interference is *constructive*, because the amplitude of the sum is near its maximum at $\cos(0) = 1$. As ϕ increases, the amplitude decreases until at $\phi = \pi$, the amplitude has shrunk to zero. This is called *destructive* interference.

When the frequencies differ by an amount $\Delta\omega$, their sum is

$$\sin(\omega t) + \sin((\omega + \Delta\omega)t) = 2\cos(\frac{\Delta\omega}{2}t)\sin((\omega + \frac{\Delta\omega}{2})t). \tag{A.3}$$

When $\Delta\omega$ is small, the cosine term is slowly varying compared with the sine term, and the resulting signal can be viewed as a sine of frequency $\omega + \frac{\Delta\omega}{2}$ with a slowly varying envelope of frequency $\Delta\omega$.

B

Ratios Make Cents

This appendix presents formulas for conversion between ratios and cents. `Matlab` *functions are available on the CD to carry out the calculations.*

Cents were first introduced by Ellis (see his annotations to Helmholtz's *On the Sensations of Tone*) as a way of simplifying comparisons between various scales and temperaments. As perceptions of musical pitch are approximately proportional to the logarithm of the frequency (rather than the frequency itself), it is sensible to use a log-based measuring system. Ellis chose to define the octave as equal to 1200 cents,[1] and so it is necessary to scale by a factor of $\frac{1200}{\log(2)}$ when converting to cents.

ratio	$1:1$	$r:1$	$2:1$
log ratio	0	$\log(r)$	$\log(2)$
cents	0	$\left(\frac{1200}{\log(2)}\right)\log(r)$	1200

Said more simply, a cent is 1/100 of a semitone, and there are 100 cents in a semitone and 1200 cents in an octave.[2]

There are two reasons to prefer cents to ratios: Where cents are added, ratios are multiplied; and it is always obvious which of two intervals is larger when both are expressed in cents. For instance, an interval of a just fifth, followed by a just third is (3/2) (5/4) = 15/8, a just seventh. In cents, this is 702+386=1088. Is this larger or smaller than the Pythagorean seventh 243/128? Knowing that the latter is 1110 cents makes the comparison obvious.

Because ratios and cents ultimately contain the same information, it is possible to convert from one to the other. Given a ratio r, the number of cents is

$$c = \left(\frac{1200}{\log_{10}(2)}\right)\ \log_{10}(r) \approx 3986.314 \log_{10}(r),$$

[1] Others have chosen different conventions. For instance, 1000 steps per octave gives the "millioctave" system.

[2] In other words, one cent is equal to an interval of $\sqrt[1200]{2} \approx 1.00057779$ to 1.

where $\log_{10}$ is the logarithm[3] base 10.

To convert from cents back into ratios, let c be the number of cents. Then the ratio r is[4]

$$r = 10^{\left(\frac{c \ \log_{10}(2)}{1200}\right)} \approx 10^{0.00025086c}.$$

These formulas are the heart of the two `Matlab` functions `cent2rat.m`[5] and `rat2cent.m`,[6] which can be found on the CD in the `software` folder. As suggested by their names, these convert from ratios to cents and back again. Both are general enough to accept a vector of inputs. For instance, to find the cent equivalent of the JI major scale, enter the desired ratios as a vector

$$\mathtt{r} = [1, 9/8, 5/4, 4/3, 3/2, 5/3, 15/8, 2],$$

and then call the routine `rat2cent` by `c=rat2cent(r)`. The program should reply

$$\mathtt{c} = [0, 203.9, 386.3, 498, 702, 884.4, 1088.3, 1200].$$

As the two functions are inverses, entering `r=cent2rat(c)` gives back the JI major scale, although in decimal form.

---

[3] Any logarithm base can be used. For instance, with the natural log (often abbreviated "ln"), the formula becomes $c = \frac{1200}{\ln(2)} \ln(r) \approx 1731.234 \ln(r)$.

[4] Using natural logs, this is $r \approx \mathrm{e}^{0.000577623c}$.

[5] The `Matlab` function `cent2rat.m` converts from cents into (the decimal equivalent of) ratios:

```
function ratio=cent2rat(cents)
ratio=10.^((log10(2)/1200)*cents);
```

[6] The `Matlab` function `rat2cent.m` converts from ratios into cents:

```
function cents=rat2cent(ratio)
cents=1200/log10(2)*log10(ratio);
```

C

Speaking of Spectra

Beware thy methods of musical analysis. Their power to blind is proportional to their power to enlighten. – B. McLaren *in Tuning Digest 120.*

In the early part of the nineteenth Century, Jean Baptiste Joseph Fourier showed how any periodic signal (for instance, a sound with a steady tone) can be decomposed into (and rebuilt from)[1] a sum of sine wave partials. Such a decomposition is called the *spectrum* of the sound, and it is usually graphed with the frequency of each sine wave partial on one axis and the magnitude on the other. Although this is useful in many fields, it is particularly appropriate to analyze sounds in this way because the ear acts as a kind of "biological" spectrum analyzer.[2] When listening "analytically," so as to "hear out" the partials of a sound,[3] the ear carries out a similar decomposition, and the tonal quality of the sound can often be correlated with measurable features of the spectrum.

This is not the place for a technical discussion[4] of the mathematics of spectra, of Fourier transforms, nor of the details of how they are calculated using the FFT.[5] Rather, this appendix supposes the availability of a software routine or command to calculate the FFT and discusses the tradeoffs and compromises that are inherent when evaluating the spectrum of a sound. In other words, the focus is on how to use and interpret the FFT, rather than on worrying about how it works or the underlying mathematics.

[1] Appendix D details how to implement this rebuilding procedure.

[2] Different portions of the basilar membrane respond to different frequencies. Recall Fig. 2.4 on p. 16.

[3] Recall the discussion of analytic vs. holistic listening on p. 25.

[4] There are already many books in the engineering literature such as [B: 60] that do this quite well. The *Elements of Computer Music* by Moore [B: 117] has an extensive discussion of FFTs from a musical perspective and includes program listings in the C language. The *Digital Signal Processing Primer* of Steiglitz [B: 182] is less complete but equally compelling.

[5] The "Fast Fourier Transform" is the name of an efficient algorithm or computer program that carries out the necessary calculations to find the spectrum. Chapter 7 of [B: 76] has a comprehensive set of worked out examples and `Matlab` routines for spectral analysis.

A digitized sound is a string of real numbers (or *samples*) that represent the amplitude of the sound at each instant. Suppose that one period of a waveform contains N samples. The spectrum is found by applying the FFT, and the output of the FFT is a string of N complex numbers that are usually written as a magnitude and a phase.[6] The magnitude spectrum is important to the ear because it specifies the size of the sine wave partials of the sound. The phase spectrum is relatively unimportant in many applications because it is often impossible to hear the difference between two sounds that have the same magnitude spectrum, even if the phase spectra differ.

The FFT has two remarkable properties. First, it is invertible. This means that it is possible to calculate the spectrum from the waveform, or to calculate the waveform from the spectrum.[7] Said another way, the waveform and the spectrum contain the same information. Certain aspects of the sound are more clearly viewed in one form or the other. For instance, the envelope of the sound is clearer from the waveform, whereas the partials are clearer from the spectrum.

Second, the FFT is linear, implying that the FFT of the sum of two signals is the same as the sum of the FFT of the two signals separately. In symbols,

$$FFT(w + v) = FFT(w) + FFT(v),$$

where w and v are two signals. More generally, if a sound consists of a number of partials, then the FFT of the complete sound is equal to the sums of the FFTs of all partials. Thus, many of the subtleties of using and understanding the FFT occur even in the simplest setting when taking the FFT of a single sine wave.

C.1 Spectrum of a Sine Wave

When there is only a single partial in the sound, then the spectrum contains only this one partial. In an ideal setting, the spectrum of a pure sine wave is zero everywhere except at the frequency of the sine wave. But the actual FFT of a real sine wave is not exactly zero, and there are two different kinds of errors, roundoff (numerical) errors and artifacts ("edge effects"), that cause the representation of a sine wave to "leak" or "smear out" to other frequencies. Figure C.1 shows a portion of a sine wave in part (a) and its spectrum, as calculated by the FFT[8] in part (b). The frequency of the wave is given by the location of the peak in (b), and the balance of the spectrum, with magnitude about 10^{-15}, is due to numerical roundoff errors in the computations.

[6] The magnitude vector is symmetric about the midpoint, and the phase is antisymmetric about the midpoint. Thus, half of each vector is redundant and is typically discarded.

[7] This latter operation is often called the Inverse FFT, and it is abbreviated *IFFT*.

[8] The `Matlab` code used to generate (a) and (b) is:

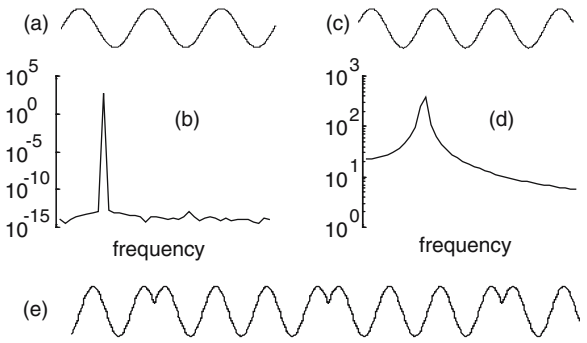


Fig. C.1. Figures (b) and (d) show the spectra of the sinusoidal segments in (a) and (c). Observe the wildly different scales of the two spectra; (b) is very close to zero except at the frequency of the sine wave, whereas (d) never sinks below 10. (e) shows several copies of (c) pasted together.

Contrast this with the sine wave shown in part (c) and its spectrum[9] in (d). The peak defining the frequency of the wave is again clearly visible, but the remainder of the spectrum only falls below 10 at high frequencies.

The sine waves (a) and (c) differ only slightly in frequency. What causes the dramatic difference in their spectra? As mentioned before, the FFT always assumes that the N samples represent exactly one period of a periodic waveform. Concatenating several copies of (a) does indeed give a longer sine wave. But concatenating several copies of (c) gives the waveform shown in (e), which is not at all sinusoidal. Thus, the spectrum (d) really shows how to decompose one period of the (nonsinusoidal) signal (e) into sine waves. It is unlikely that this is what was really intended when thinking of the frequency content of (c). Thus, there is a complex interplay between the periodicity of the waveform and the length of the FFT.

Given this, it might seem like a good idea to choose the length of the FFT to match the period of the partials. Unfortunately, this is almost never possible when analyzing real sounds, because choosing this length requires knowing the frequencies of the partials, and finding these frequencies is the reason for taking the FFT in the first place.

Think of it another way. The problem (the large magnitude at frequencies different from the "obvious" frequency of the sine wave) occurs because the "ends" do not line up; abrupt changes in the waveform cause the spectrum to smear. One way to force the ends to line up is to preprocess the data so

```
c=(2*pi)/128;            % c defines the frequency of the sine wave.
wave=sin(c*(0:1023));    % the sine wave is 1024 samples long.
plot(wave)               % generates the plot in part (a).
magspec=abs(fft(wave)); % ''FFT'' returns the FFT in complex form.
                         % ''abs'' takes the magnitude of the FFT.
semilogy(magspec(1:50)) % plots (b) with logarithmic vertical axis.
```

[9] Parts (c) and (d) were generated by identical code, except that the parameter c was changed slightly so that an integer number of periods do not fit into the sample length.

that it dies away to zero at both ends. Then, no matter what the underlying periodicity, there will be no abrupt changes in the waveshape.

One popular approach is to use a *Hamming* window,[10] which is shown in part (a) of Fig. C.2. Multiplying this window point by point times part (b) (which is the same waveform as in Fig. C.1(c)) gives the windowed version in part (c). The spectrum of (c) is shown in (d).

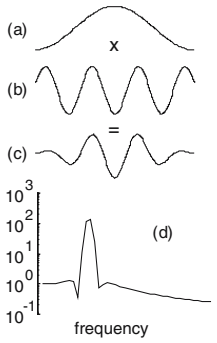


Fig. C.2. A hamming window (a) is multiplied point by point times a segment of a sinusoid (b), resulting in (c). The spectrum, shown in (d), has significantly lower sidelobes than in the unwindowed version, although the peak is somewhat wider.

Compare the spectrum of this windowed version with the spectrum of the unwindowed version in Fig. C.1(d). In both, the frequency of the sinusoid is given by the location of the peak. The windowed version has attenuated the smearing by a factor of almost 10, although the peak is about twice as wide. This is fairly typical of the windowing process.

When should a window be used? Windowing is unnecessary when dealing with a short isolated sound whose start and end are known. In a typical musical synthesizer or sampler, each sound has a well-defined start (attack) and a definite steady-state looped portion. As the loop is periodic, it is an ideal place to apply the FFT without windowing.[11] In many other circumstances, when a continuously changing signal is analyzed, windows are used to reduce end effects.[12] Figure C.3 shows this schematically. A series of offset windows in (a) are multiplied point by point times the waveform (b), giving the smaller segments (c). The segments can then be readily analyzed, giving spectral "snapshots" of the evolution of the partials of the sound.

End effects are a consequence of the fact that Fourier's theorem (and hence all techniques based on the Fourier transform) apply only to periodic

[10] Named after Richard Hamming, this is a single cycle of a scaled and shifted cosine wave. The formula is $h(t) = 0.54 - 0.46\cos(2\pi t/(N-1))$ for $0 \le t < N$. The Hamming window has been enshrined in a `Matlab` function called "hamming," but is only one of many possible windowing functions. Steiglitz [B: 182] and Moore [B: 117] discuss several alternatives, each with their own properties.

[11] The innards of a typical musical synthesizer are discussed on p. 31.

[12] Although it is true that windows help to reduce artifacts, it is worth remembering that this is, in effect, lying about the data.

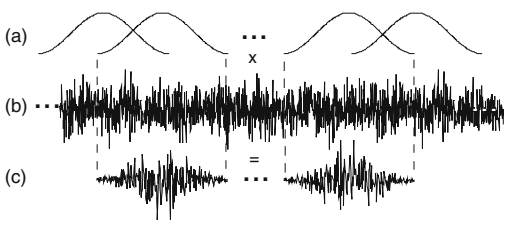


Fig. C.3. Overlapping windows applied to a continuos waveform give smaller segments that can be analyzed easily.

signals. To calculate the FFT of a "real" signal requires "pretending" that it is periodic with period equal to the length of the sample. Although this can often be done without gross distortion, careful choice of sample lengths and windowing techniques are needed to reduce the likelihood of misleading results.

C.2 Steady State Analysis

You somehow shake a waveform, and the partials come tumbling out.[13]

Consider a spectral analysis of the sound of a vibrating string that has a fundamental pitch of 100 Hz, approximately the G an octave below middle C. Assume the standard CD sampling rate of 44.1K samples per second, and that the sound of the string lasts about three seconds. This gives about 128K samples, and it is impractical to calculate an FFT of this length. The data should be broken up into chunks that can be analyzed separately. For example, 32K chunks representing 3/4 second of sound are reasonable.[14]

First, consider the simple case when the sample is very close to periodic, as occurs during the sustained steady-state portion of the sound. Because strings vibrate harmonically, there would ideally be a peak at 100 Hz, another at 200 Hz, another at 300 Hz, and so on, each with an appropriate amplitude. But the output of the FFT program does not look like this, not exactly. The FFT algorithm outputs a 32K magnitude vector and a 32K phase vector. As only half of each vector is meaningful, the remainder is discarded.

Each element in the (nonredundant) 16K magnitude vector represents the magnitude of a sine wave at some frequency. In this case, the first number represents the magnitude of the DC (zero frequency, or bias term). The second element represents the magnitude of the sine wave at

$$\frac{\text{sample rate}}{\text{sample length}} = \frac{44100}{32768} = 1.346 \text{ Hz}.$$

The next number is the magnitude of the sine wave at frequency 2.69 Hz. Thus, the output of the FFT cannot represent the sine wave at 100 Hz exactly,

[13] Paraphrased from Marion M. in *Tuning Digest 314*.
[14] For sounds that change more rapidly, smaller chunks should be used.

because there is no slot in this representation for 100 Hz. In fact, the 74th bin represents 99.59 Hz and the 75th slot represents 100.94 Hz, so the energy that should be at 100 Hz is spread out near the 74th and 75th slots. Similarly, none of the other "real" frequencies are exactly represented. This quantization of frequency is a direct result of the assumption that the signal is periodic, that it repeats every 32K. Of course, this is just a convenient fiction, because the signal from the string continues to die away for more than 128K samples.

Thus, there are two notions of "period," and this can be a source of confusion. First is the notion of the period of the fundamental and its harmonics. As the fundamental of the string is 100 Hz, there will also typically be string vibrations at 200 Hz, 300 Hz, 400 Hz, 500 Hz, and so on. The second notion of "period" that enters into the FFT analysis is that all frequencies of the analyzed signal appear to be multiples of 1.346 Hz, which is a direct result of the choice of a 32K FFT. Had the analysis used 8K FFTs, everything would have been a multiple of 5.38 Hz, and the representation of the 100 Hz fundamental would have been even worse. Thus, the resolution of the spectral analysis is directly proportional to the "width" of frequency bins, which determines how accurately the sine wave components can be represented. This is similar to the "smearing" observed when analyzing single sine waves in the previous section.

These two ideas of period suggest two interpretations of the spectral analysis. One is literally correct (but useless), and the other is an approximation (that is often useful). A literal interpretation of this FFT data suggests that the fundamental of the string is vibrating at 1.346 Hz, and that the 74th, 75th, 148th, 149th (and so on) harmonics are large. While literally true, this is not a particularly useful way to think of the vibrating string. Observe that using an 8K FFT, the same signal would be interpreted as a fundamental at 5.38 Hz along with some large harmonics: the 18th, 19th, 37th, 38th, and so on. Clearly, a true interpretation of the strings motion should not depend on the size of the FFT used in the analysis.

A better interpretation of the string data is as a fundamental between 99.59 Hz and 100.96 Hz, with a second partial near 200 Hz, and so on. But this does require that a judgment be made, because the location of the peaks must be determined. Although the peaks are obvious in some situations, in others there is ambiguity between peaks caused by the instrument (the string) and those due to noises, disturbances, and artifacts. A later section discusses an algorithm for automatic peak detection.

C.3 Analysis of the Attack

The previous section showed that Fourier analysis of a nearly periodic sound (such as the steady-state portion of the string vibrations) is feasible. Learning about the attack portion of a sound using Fourier analysis is trickier due to a kind of auditory uncertainty principle. The more accurately the frequency content of a sound is known, the harder it is to tell exactly when it occurs.

The more accurately specified an event is in time, the less can be said about the actual frequencies.

To see this in a simple setting, consider a sound that consists of a one-half second sinusoid with frequency 100 Hz followed by a one-half second sinusoid with frequency 200 Hz. Taking a single FFT over the complete wave shows two large peaks at 100 Hz and 200 Hz, along with smearing due to end effects and to the transition between the two halves. An FFT of the first half shows just the peak near 100 Hz (plus the inevitable artifacts), whereas an FFT of the second half shows just the peak at 200 Hz, again with artifacts. This is called the "averaging" property of the FFT and is inevitable when analyzing a sound that changes over time. Larger windows give more accurate locations for the partials,[15] but it becomes impossible to resolve when the various partials actually occur.

Because of this, a sensible strategy is to use several different FFTs on the same data. The larger FFTs help to resolve the actual frequencies, and the shorter FFTs help to locate when the partials occur. Such techniques are detailed in several places in Chap. 7 "A Bell, A Rock, A Crystal" in the context of analyzing the spectra of inharmonic musical sounds. The auditory uncertainty principle is also "discussed" in the last verse of Appendix K.

C.4 Pads and Windows

This section briefly describes a number of techniques for preprocessing the data before applying the FFT. None of these should be applied indiscriminately, but they may prove useful, especially when trying to analyze a single sound as accurately as possible.

Padding with Zeroes

The FFT requires that the number of samples be a power of two (or some highly composite number). One common technique is to "pad" the data with zeroes until the length reaches the next highest power of two. This can also increase the accuracy of the representation of the frequencies of the partials, because a longer FFT is used.

Reverse the Waveform

Another way to sensibly lengthen the waveform is to reverse and concatenate. Instead of taking the FFT of $s_1, s_2, ..., s_k$, the data can be augmented to

$$s_1, s_2, ..., s_{k-1}, s_k, s_{k-1}, s_{k-2}, ..., s_2, s_1.$$

[15] For instance, to the nearest 1.346 Hz for a 32K FFT instead of to the nearest 5.38 Hz for an 8K FFT.

The rationale for this is that the forward and reversed data have the same (magnitude) spectrum. If the "splice point" is chosen carefully so that the data varies smoothly near s_k, then the artifacts can be reduced.

One-Sided Window

When analyzing a sound (such as from a musical synthesizer or sampler) that has explicit attack and looped portions, no window should be applied to the loop. (Indeed, this is the one place where Fourier techniques shine—the loop genuinely is periodic.) The attack portion has a definite beginning, but its end mingles with the start of the loop. Applying a standard Hamming (or other symmetric) window to the attack portion will destroy much of the desired information at the start of the sound. Yet applying no window may encourage artifacts due to the abrupt change where the loop begins. A convenient compromise is to apply a one-sided window, that is, only the decaying (second) half of the window.[16] This leaves the initial portion unaltered, yet discourages artifacts caused by interface between the loop and attack portions.

C.5 Finding Spectral Peaks

Humans are very good at recognizing patterns. For instance, when looking at spectral plots such as Fig. 7.6 on p. 141, it is easy to visually "pick out" the most significant peaks, and in most cases, these peaks are indeed the most auditorily significant aspects of the sound. Machines are notoriously bad at this kind of task, for instance, reading text is a similar kind of pattern recognition problem that has not been completely solved, despite intense effort.

A naive approach to the "peak picking" problem is to find the largest term in the magnitude vector and call it the first peak, find the second largest element and call it the second peak, and so on. Unfortunately, few peaks are isolated outliers; they usually look like small mountains, with foothills and subpeaks. For example, the naive approach would find the highest peak in the middle spectrum of Fig. 7.6 on p. 141, at 5066 Hz, but it would then find the second highest element at 5063 Hz, and the third at 5069 Hz. A slightly more sophisticated approach would require that candidate peaks be larger than their immediate neighbors. But consider the complex of peaks near 5553 Hz on Fig. 7.1 of p. 134. Even a combination of the size and neighbor criteria would declare there to be many peaks here, even though only one (or maybe two) is sensible. Clearly, a more sophisticated approach is required.

The defining aspect of a peak is that it must be larger than the surrounding regions. The "competitive filtering" ideas of [B: 122] suggest dividing the search for peaks into three regions: to the left of the candidate peak, to the

[16] This can be analyzed as a zero (pre)padding, followed by application of a complete Hamming window, but it is simpler to implement directly as a half window.

right, and the value of the candidate peak itself. If the candidate is larger than (a constant times) the sum of the average to the left plus the average to the right, then a peak is successfully found. This simple algorithm can be effective, but there are two parameters that must be chosen. First is the constant, which is typically near one. This parameter is roughly proportional to the steepness of the peak, with larger values requiring steeper peaks. The second parameter is the length of the averages. This must be chosen based on the size of the FFT and using any a priori knowledge of how close together two peaks can be. For instance, if the frequencies of the FFT differ by 1.34 Hz (as in a 32K FFT) and the closest expected peaks are 50 Hz apart, then the averages should be taken over no more than 20 values to the left and right.

D

Additive Synthesis

A brief discussion of some `Matlab` *programs that implement additive synthesis and resynthesis.*

Additive synthesis is the process of summing a collection of sine wave partials so as to make a complex, and hopefully interesting, sound. For example, suppose we wish to generate sounds with the same partials (the same spectrum) as the Chaco rock of Fig. 7.6 on p. 141. The most important partials of the sound can be read directly from the figure or from the composite spectrum of Fig. 7.7 on p. 142. These are

$$1351,\ 2040,\ 2167,\ 4068,\ 5066,\ \text{and}\ 7666.$$

Letting these be the frequencies of the m partials and labeling them w_1 through w_m, a new sound can be built as

$$w(t) = \sum_{i=1}^{m} a_i \text{env}_i(t) \cos(w_i t + p_i),$$

where the a_i define the amplitudes associated with each partial and the p_i are some (usually arbitrarily specified) phases. The function $\text{env}_i(t)$ represents the envelope of partial i, and it can be chosen to help define the character of the sound. For instance, if all envelopes are constant, $\text{env}_i(t) = 1$, then the sound will be steady like an organ tone. Envelopes that die away exponentially, like $\text{env}_i(t) = e^{-t}$, tend to mimic the character of a struck, plucked, or percussive timbre.

By construction, the waveform $w(t)$ has partials at the w_i, and hence, it has a dissonance curve with minima at many of the same locations as the original sound. This is one way of generating "new" sounds that are compatible with an existing timbre. For instance, the high percussive tones in the *Chaco Canyon Rock* (audio track [S: 44]) were generated with exponentially decaying envelopes, and the sustained organish tones of the middle section were created using constant envelopes.

The `Matlab` program `addsynth.m`, which generates `.wav` files via additive synthesis, appears on the CD in the `software` folder. The frequencies

(in Hertz) are placed in the vector `freq` and the corresponding amplitudes and decay rates are specified in `amp` and `decay`.[1] The program generates a waveform `time` seconds long at a sampling rate `sr`. If there is a soundcard available on the computer, the sound can be previewed using the command

```
sound(wave, sr)
```

which plays the vector `wave` at the sampling rate `sr`. With its default parameters, `addsynth.m` generates a harmonic sound with five partials of equal amplitude. The sound is somewhat different each time `addsynth.m` is run because the decay rates change (due to the `randn` function in the definition of `decay`).

One common technique is to use data from the spectrum to resynthesize a sound. In the simplest case, the spectrum may be calculated and then transformed back into a waveform without loss of information. This is demonstrated in the `Matlab` program `resynth.m` (also available in the `software` folder of the CD), which calculates the spectrum of a sound and then carries out a direct resynthesis of the sound from the FFT decomposition. With no additional processing, the output `x` is identical to the input `y`, at least to numerical precision.

Alternatively, the sound can be sculpted or shaped as desired by manipulating the magnitude and/or phase values prior to the resynthesis. This would occur at the place in the code marked with the comment:

```
% Frequency domain processing goes here:
```

One possibility is to "move" the most prominent partials to make them compatible with some desired reference spectrum. This is the idea exploited in the "Spectral Mappings" chapter, although the more efficient inverse FFT is used instead of an additive resynthesis approach.

The programs given here are not computationally efficient; rather, they are intended to present the ideas as clearly as possible. For instance, a better way of carrying out additive synthesis is given in Steiglitz [B: 182], and a reasonable implementation of the related phase vocoder is presented in Moore [B: 117]. Finally, an important discussion of the impact of additive synthesis on electronic music is given in Risset [B: 150].

[1] The three vectors `freq` and `amp` and `decay` must all be the same length.

E

How to Draw Dissonance Curves

This appendix describes a parameterization of Plomp and Levelt's dissonance curves and computer programs that carry out the calculations. It is not necessary to follow the math in detail to make use of the computer programs. Contrariwise, it is not necessary to program the computer to understand the math.

The Plomp–Levelt curves of Fig. 3.7 on p. 46 can be conveniently parameterized by a model of the form

$$d(x) = \mathrm{e}^{-b_1 x} - \mathrm{e}^{-b_2 x} \tag{E.1}$$

where x represents the absolute value of the difference in frequency between two sinusoids, and the exponents b_1 and b_2 determine the rates at which the function rises and falls. Using a gradient minimization of the squared error between the (averaged) data and the curve $d(x)$ gives values of $b_1 = 3.5$ and $b_2 = 5.75$.[1]

The dissonance function $d(x)$ can be scaled so that the curves for different base frequencies and with different amplitudes are represented conveniently. If the point of maximum dissonance occurs at x^*, then the dissonance between sinusoids at frequency f_1 with loudness ℓ_1 and at frequency f_2 with loudness ℓ_2 (for $f_1 < f_2$) is

$$d(f_1, f_2, \ell_1, \ell_2) = \ell_{12}[\mathrm{e}^{-b_1 s(f_2 - f_1)} - \mathrm{e}^{-b_2 s(f_2 - f_1)}] \tag{E.2}$$

where

$$s = \frac{x^*}{s_1 f_1 + s_2} \tag{E.3}$$

and

$$\ell_{12} = \min(\ell_1, \ell_2). \tag{E.4}$$

The point of maximum dissonance $x^* = 0.24$ is derived directly from the model (E.1) above. The s parameters in (E.3) allow a single functional form

[1] An alternative parameterization of the Plomp–Levelt curves, proposed by Lafrenière [B: 92], replaces the difference between exponentials in (E.1) with $d(x) = \mathrm{e}^{-(\log(\beta x))^2}$, where β is chosen so that βx occurs at the point of maximum dissonance and where $x = \frac{f_2 - f_1}{f_1}$ is the normalized frequency. The resulting dissonance curves are qualitatively similar to the ones presented here, although the corners are more rounded. Another functional form that may also be useful in this context is $d(x) = x\mathrm{e}^{-\beta x}$.

to interpolate between the various curves of Fig. 3.8 on p. 47 by sliding the dissonance curve along the frequency axis so that it begins at f_1, and by stretching (or compressing) it so that the maximum dissonance occurs at the appropriate frequency. A least square fit was made to determine the values $s_1 = 0.021$ and $s_2 = 19$.

The form of equation (E.4) ensures that softer components contribute less to the total dissonance measure than louder components. For instance, if either ℓ_1 or ℓ_2 approaches zero, then ℓ_{12} decreases and the dissonance in (E.2) vanishes. Conversely, if the volume of the partials increases, the dissonance increases. This form is discussed in Appendix G, and is a refinement of the model in [B: 165], which assumed that the loudnesses were multiplicative.

Calculating loudness is not completely trivial as the discussions in [B: 85], [B: 154] and [B: 187] suggest. If $p(t)$ represents a simple harmonic planar wave with period T, then the effective pressure is the power

$$P_e = \sqrt{\frac{1}{T}\int_0^T p^2(t)dt}$$

of the wave. For a sine wave, $p(t) = A\sin(2\pi f_0 t + \phi)$ with frequency f_0 and amplitude A, $P_e = \frac{A}{\sqrt{2}}$. The sound pressure level in decibels (dB) is $\text{SPL} = 20\log_{10}(\frac{P_e}{P_{ref}})$, where P_{ref} is the standard reference of $20\mu\text{Pa}^2$ for SPL in air, which corresponds to the SPL of a barely audible sine wave of frequency 1000 Hz. Finally (and somewhat crudely), the loudness can be approximated as

$$\ell = \frac{1}{16}\, 2^{\frac{\text{SPL}}{10}}. \tag{E.5}$$

The loudness ℓ is measured in *sones*. The form of (E.5) originates from the observation that an increase of 10 dB corresponds (approximately) to a doubling of loudness. The fraction 1/16 normalizes the loudness so that 40 dB corresponds to one sone. More accurate models than (E.5) would include the effects of the Fletcher–Munson curves of equal loudness [B: 154], would sum the loudnesses differently depending on whether they occupy the same critical band, and would take into account masking effects.

To calculate the dissonance of more complex sounds, let F be a collection of n sine wave partials with frequencies $f_1 < f_2 < ... < f_n$ and loudnesses ℓ_j for $j = 1, 2, ..., n$. The partials will typically be displayed as the n-tuple $f_1, f_2, ..., f_n$. The dissonance of F can be calculated as the sum of the dissonances of all pairs of partials

$$D_F = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} d(f_i, f_j, \ell_i, \ell_j), \tag{E.6}$$

which is called the intrinsic or inherent dissonance of F. When two notes with spectrum F are played simultaneously at an interval α, the resulting sound

[2] One Pascal (Pa) is one N/m^2.

has a dissonance that is the same as that of a single timbre with frequencies f_i and αf_i by the additivity assumption. Thus, (E.6) can be used directly to calculate the dissonance between intervals (and chords) as well as the dissonance of isolated timbres. Defining the spectrum αF to contain the frequencies $\alpha f_1, \alpha f_2, ..., \alpha f_n$ (with loudnesses ℓ_j), the dissonance of F at an interval α is

$$D_F(\alpha) = D_F + D_{\alpha F} + \sum_{i=1}^{n} \sum_{j=1}^{n} d(f_i, \alpha f_j, \ell_i, \ell_j), \tag{E.7}$$

and the dissonance curve generated by the timbre F is defined as the function $D_F(\alpha)$ over all intervals of interest α.

The dissonance of a chord of three notes at the intervals $1, r$, and s can be similarly calculated by adding the dissonances between all partials

$$D_F(r, s) = D_F(r) + D_F(s) + D_{rF}(s/r),$$

where $D_F(r)$ is the dissonance of F at the interval r, $D_F(s)$ is the dissonance of F at the interval s, and $D_{rF}(s/r)$ is the dissonance between rF and sF. Generalizations to m sounds, each with their own spectrum, follow the same philosophy of calculating the sum of the dissonances between all simultaneously sounding partials.

Two computer programs that carry out these calculations are located in the `software` folder on the CD. The first, `Dissonance(Basic)`, is written in Microsoft's version of BASIC, and the other is in `Matlab`. Both programs encapsulate the equations of this section and can be used to draw dissonance curves for a timbre with `n` partials, at frequencies specified in the array `freq` with corresponding amplitudes in the array `amp`.

Some details of the implementation might help to follow the program logic. In the BASIC program, the `i` and `j` loops calculate the dissonance of the timbre at a particular interval `alpha`, and the `alpha` loop runs through all intervals of interest. The first few lines set up the frequencies and amplitudes of the timbre. The variable `n` must be equal to the number of frequencies in the timbre. Running the program with its default values generates the dissonance curve for a harmonic timbre with six partials. To change the start and end points of the intervals, use `startint` and `endint`. To make the intervals further apart, increase `inc`. All dissonance values are stored in the vector `diss`. Do not change `dstar` or any of the variables with numbers.

The `Matlab` programs are modular, one defining a `Matlab` function called `dissmeasure.m`, which calculates the dissonance of any set of partials `f` with loudness `amp` (the partials can be in any order). The main routine `dissmain.m` calls `dissmeasure.m` for each interval of interest to draw the dissonance curve. A FORTRAN version is also listed in [B: 92].

F

Properties of Dissonance Curves

For certain simple timbres, dissonance curves can be completely characterized. This appendix derives bounds on the number and location of minima of the dissonance curve and reveals some general properties, as discussed in Chap. 6. Two simplifications are made to streamline the discussion. A single dissonance function is assumed for all frequencies, and all partials are presumed to have unit amplitudes. Thus the simpler model (E.1) is used in place of the more complete model (E.2)-(E.4) whenever convenient.

When F is a spectrum with partials at frequencies $f_1, f_2, ..., f_n$, the intrinsic dissonance (in this simplified setting) is

$$D_F = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} d(f_i, f_j) \tag{F.1}$$

where $d(f_i, f_j)$ is really a function of a single variable; that is, $d(f_i, f_j) \equiv d(x)$ as defined in (E.1) with $x = \frac{|f_i - f_j|}{\min(f_i, f_j)}$, and where the last two (amplitude) terms of (E.2) are assumed unity. Because of the form of x, $d(\alpha f_i, \alpha f_j) = d(f_i, f_j)$, and so $D_F = D_{\alpha F}$ for any α. In other words, the simplification has removed the dependency on absolute frequency from the dissonance measure.

Using these notations, the dissonance curve (E.7) becomes

$$D_F(\alpha) = D_F + D_{\alpha F} + \sum_{i=1}^{n}\sum_{j=1}^{n} d(f_i, \alpha f_j). \tag{F.2}$$

The first result gives a precise statement of property two from p. 121, describing the behavior of the dissonance curve as the interval α grows large.

Theorem F.1. *For any timbre F with partials at $f_1, f_2, ..., f_n$,*
$\lim_{\alpha\to\infty} D_F(\alpha) = D_F + D_{\alpha F}$.

Proof: Clearly, $d(x) \to 0$ as $x \to \infty$. Thus, $d(f_i, \alpha f_j) \to 0$ for all i, j as $\alpha \to \infty$, which implies that the double sum in (F.2) approaches zero. Δ

Thus, the dissonance decreases as the interval α grows larger, approaching a value that is no more than the dissonances of the timbres D_F and $D_{\alpha F}$.

Various aspects of the dissonance curve (E.1) become important when investigating the locations of possible minima of the dissonance curve. Several of these are given here, most following from a direct application of calculus. Taking the derivative of (E.1), setting it to zero, and solving shows that the point of maximum dissonance occurs when

$$x^* = \frac{\ln(b_1/b_2)}{b_1 - b_2}. \tag{F.3}$$

Two partials f_i and f_j are said to be *separated by* x^* if

$$x = \frac{|f_i - f_j|}{\min(f_i, f_j)} > x^*.$$

The change in dissonance at $x = 0$ is

$$d'(0) = -b_1 \mathrm{e}^{-b_1 x} + b_2 \mathrm{e}^{-b_2 x}|_{x=0} = b_2 - b_1. \tag{F.4}$$

For $x > x^*$, the maximum change in the derivative occurs when $d'(x^{**})$ is minimum. As

$$d''(x) = b_1^2 \mathrm{e}^{-b_1 x} - b_2^2 \mathrm{e}^{-b_2 x}, \tag{F.5}$$

$x^{**} = \frac{2\ln(b_1/b_2)}{(b_1 - b_2)}$ is where the minimum occurs. After some simplification, the value of d' at x^{**} is

$$d'(x^{**}) = b_2 \left(\frac{b_1}{b_2}\right)^{\frac{2b_2}{b_1 - b_2}} - b_1 \left(\frac{b_1}{b_2}\right)^{\frac{2b_1}{b_1 - b_2}}. \tag{F.6}$$

When needed, the values $b_1 = 3.5$ and $b_2 = 5.75$ are used, so that $x^* \approx 0.22$, $d'(0) \approx 2.25$, $x^{**} \approx 0.44$, and $d'(x^{**}) \approx -0.292$, although generally $b_2 > b_1 > 0$ is enough.

The next result finds conditions under which the unison $\alpha = 1$ is a minimum of the dissonance curve $D_F(\alpha)$.

Theorem F.2. *Let F have partials $f_1 < f_2 < ... < f_n$ that are all separated by at least x^*. Then $\alpha = 1$ is a minimum of $D_F(\alpha)$.*

Proof: As D_F and $D_{\alpha F}$ are fixed and equal for all α, only the terms in the double sum (F.2) change the value of $D_F(\alpha)$. There are n terms of the form $d(f_i, \alpha f_i)$ in the sum, and for each of these there are $n - 1$ terms of the form $d(f_i, \alpha f_j)$ with $i \neq j$. We show that the change in $d(f_i, \alpha f_i)$ is greater than the sum of all changes in $d(f_i, \alpha f_j)$ for $i \neq j$ when α is suitably close to 1.

The change in $d(f_1, \alpha f_1)$ for $\alpha \approx 1$ is proportional to $d'(0)$, which is given in (F.4) as $b_2 - b_1$ (because $\alpha = 1$ corresponds to $x = 0$). The largest possible value for any of the $d(f_1, \alpha f_j)$ occurs when f_1 and αf_j define an x with $x = x^{**}$. Then $d'(x^{**})$ is given in (F.6). Because the f_j are assumed separated

by at least x^*, and because $x^{**} = 2x^*$, the next largest derivative is at most $d'(3x^*)$. We now claim that the sum of all derivatives $\sum_{i=1}^{n} |d'(ix^*)|$ is less than $d'(0)$. Observe that

$$d'(ix^*) = b_2 \left(\frac{b_1}{b_2}\right)^{\frac{ib_2}{b_1-b_2}} - b_1 \left(\frac{b_1}{b_2}\right)^{\frac{ib_1}{b_1-b_2}} \equiv b_2 t_2^i - b_1 t_1^i$$

and that

$$\sum_{i=2}^{n} |d'(ix^*)| \leq \sum_{i=1}^{\infty} |d'(ix^*)|.$$

As the $d'(ix^*)$ are all of the same sign, drop the $|\cdot|$. Combining the two previous expressions yields

$$\sum_{i=1}^{\infty} (b_2 t_2^i - b_1 t_1^i) = \frac{b_2 t_2}{1 - t_2} - \frac{b_1 t_1}{1 - t_1} \equiv t,$$

which is approximately $t = -0.758$. Since the f_j need not be spaced evenly, $\sum_{i=1}^{n} |d'(\cdot)|$ could be as large as $|t| + |d'(x^{**})| \approx 1.05$. In the general case, $d(f_i, \alpha f_j)$, the αf_j could occur both above and below the f_i; hence, the $\sum_{i=1}^{n} |d'(\cdot)|$ could be as large as $2(|t| + |d'(x^{**})|) \approx 2.1$. In all cases, the change in the diagonal terms $d(f_i, \alpha f_i)$ dominates the sum of the changes in all off-diagonal terms $d(f_i, \alpha f_j)$, giving the required inequality. Δ

The requirement in theorem F.2 that the partials be separated by x^* is sufficient but is certainly not necessary. If $n \leq 7$, then the same arguments show that no requirements are needed on the spacing of the f_i, because the change in each $d(f_i, \alpha f_i)$ is over seven times the largest possible value of the change in $d(f_i, \alpha f_j)$, for $i \neq j$ (i.e., $d'(0)/d'(x^{**}) \approx 7.7$).

Minima of dissonance curves tend to occur at ratios of the partials.

Theorem F.3. *Let timbre F have partials at f_1, f_2 that are separated by at least x^*. Then the dissonance curve $D_F(\alpha)$ has a minimum at $\alpha^* = f_2/f_1$.*

Proof: Let timbre G have partials $(g_1, g_2) = (\alpha f_1, \alpha f_2)$. Then $D_F = D_G = D_{\alpha F}$, and any change in $D_F(\alpha)$ must originate from the double sum in (F.2), which contains the terms $d(f_i, g_j)$ for $i = 1, 2$ and $j = 1, 2$. For $\alpha^* = f_2/f_1$, $(g_1, g_2) = (f_2, \alpha f_2)$. As α is perturbed from α^*, the contribution from the term $d(f_2, g_1) = d(f_2, \alpha f_1)$ increases, because at α^*, $\alpha^* f_1 = f_2$ and so $d(f_2, g_1) = d(f_2, f_2) = 0$. Thus, the result can be demonstrated by showing that the increase in $d(f_2, g_1)$ is greater than the decrease in the other three terms combined. The increase in $d(f_2, g_1)$ is proportional to $d'(0)$. As f_1 and f_2 are separated by x^*, the decrease in each of the other three terms is no greater than $d'(x^{**})$. As $|d'(0)| > 7|d'(x^{**})|$, this proves the desired result. Δ

Thus, the dissonance curve generated by a timbre with partials at f_1, f_2 has a minimum when $\alpha^* f_1 = f_2$. For example, for the timbre with partials at (500, 750), $\alpha^* = 1.5$. The result asserts that the timbre $\alpha^* F$, with frequencies (750,

1125) is locally a most consonant interval. In symbols, $D_F(\alpha^* - \epsilon) > D_F(\alpha^*)$ and $D_F(\alpha^* + \epsilon) > D_F(\alpha^*)$ for small ϵ. Thus, both (748, 1122) and (752, 1128) are less consonant than (750, 1125). This result is intuitively reasonable because when $\alpha f_1 \neq f_2$, the dissonance between the partials at αf_1 and f_2 is large, but when $\alpha f_1 = f_2$, this term disappears from the dissonance measure. Interestingly, the result can fail when f_1 and f_2 are too close.

Theorem F.4. *Let timbre F have partials f_1, f_2. Then there is a $\epsilon > 0$ such that for $|f_2 - f_1| < \epsilon$, the point $\alpha^* = f_2/f_1$ is not a minimum of $D_F(\alpha)$.*

Proof: Define G as in theorem F.3. Again, any change in $D_F(\alpha)$ is a result of the four terms in the sum of (F.2). For small $\epsilon > 0$, note that $d(f_1, g_1 + \epsilon) > d(f_1, g_1) > d(f_1, g_1 - \epsilon)$, $d(f_1, g_2 + \epsilon) > d(f_1, g_2) > d(f_1, g_2 - \epsilon)$, $d(f_2, g_2 + \epsilon) > d(f_2, g_2) > d(f_2, g_2 - \epsilon)$, and $d(f_2, g_1 + \epsilon) > d(f_2, g_1)$. On the other hand, $d(f_2, g_1 - \epsilon) > d(f_2, g_1) = d(f_2, f_2) = 0$. For small ϵ, the change in all four terms is approximately $\epsilon(b_2 - b_1)$ in magnitude. Thus, the dissonance value is decreased as G is moved ϵ closer to F, and $\alpha^* = f_2/f_1$ is not a minimum. Δ

In essence, if the partials f_1 and f_2 are too close, then the minimum at f_2/f_1 disappears. Theorem F.3 shows that a minimum occurs when partials coincide with each other. Minima can also occur when the partials are widely separated. For a two-partial timbre F, suppose that f_1 and f_2 are separated by at least $4x^*$. Then there is an interval of maximum dissonance near $\alpha f_1 = f_1 + x^*$, and another near $\alpha f_2 = f_2 - x^*$. Consequently, there must be a minimum for some α between $\alpha_L = (f_1 + x^*)/f_1$ and $\alpha_H = (f_2 - x^*)/f_2$. The full range of possible dissonance curves for two-partial timbres is shown in Fig. 6.15 on p. 121.

Theorem F.4 suggests that minima of the dissonance curve are unlikely for intervals smaller than about half the interval x^* at which maximum dissonance occurs. Plomp and Levelt estimate that x^* corresponds to a little less than 1/3 of the critical bandwidth. Thus, theorem F.4 predicts that scale steps closer together than about 1/6 of the critical bandwidth should be rare.

The next result describes minima of the dissonance curve for timbres with three partials.

Theorem F.5. *Let timbre F have partials f_1, f_2, f_3. Then there are $c_1 > 0$ and $c_2 > 0$ such that whenever f_1 and f_2 are separated by at least $x^* + c_1$, and f_2 and f_3 are separated by at least $x^* + c_2$, then minima of the dissonance curve occur at $\alpha_1 = f_2/f_1$, $\alpha_2 = f_3/f_1$, and $\alpha_3 = f_3/f_2$.*

Proof: Let G have partials $(g_1, g_2, g_3) = (\alpha f_1, \alpha f_2, \alpha f_3)$. Suppose first that $f_2 - f_1 > f_3 - f_2 + c_2$. Consider the candidate minimum α_1. For small ϵ, the most significant terms in $D_F(\alpha+\epsilon) - D_F(\alpha)$ are $d(f_2, g_1)$ and $d(f_3, g_2)$, because all others are separated by at least $x^* + c_2$. For $\epsilon > 0$, $d(f_2, g_1 + \epsilon) > d(f_2, g_1)$, $d(f_3, g_2 + \epsilon) > d(f_3, g_2)$, and $d(f_2, g_1 - \epsilon) > d(f_2, g_1)$. On the other hand, $d(f_3, g_2 - \epsilon) < d(f_3, g_2)$. But $d'(0) = b_2 - b_1$ and $d''(0) = b_1^2 - b_2^2 < 0$, so the slope is decreasing. Hence, $|d(f_2, g_1 - \epsilon)| > |d(f_3, g_2 - \epsilon)|$. Consequently, $D_F(\alpha_1 + \epsilon) > D_F(\alpha_1)$ and $D_F(\alpha_1 - \epsilon) > D_F(\alpha_1)$, showing that α_1 is a local

minimum. The case $f_3 - f_2 > f_2 - f_1 + c_1$ follows identically. The proofs for α_2 and α_3 are similar. Δ

Figures 6.16 and 6.17 on pp. 123 and 123 show theorem F.5 graphically. The final result specifies the maximum number of minima that a dissonance curve can have in terms of the complexity of the spectrum of the sound.

Theorem F.6. *Let timbre F have partials $f_1, f_2, ..., f_n$. Then the dissonance curve generated by F has at most $2n^2$ local minima.*

Proof: Consider the portion of $D_F(\alpha)$ due to the partial f interacting with a fixed partial f_j. For both very small α ($\alpha \approx 0$) and very large α ($\alpha \to \infty$), $d(\alpha f, f_j) \approx 0$. At $\alpha = f_j/f$, $d(\alpha f, f_j) = 0$. For the two intervals where αf and f_j are separated by x^* (one with $\alpha f < f_j$ and one with $\alpha f > f_j$), $d(\alpha f, f_j)$ attains its maximum value. Thus, f interacting with a fixed f_j has two maxima and one minima. Each f_i can interact with each f_j, and there are n^2 possible pairs. As $D_F(\alpha)$ consists of n^2 such curves added together, there are at most $2n^2$ maxima. Consequently, there can be no more than $2n^2$ minima. The two extreme minima at $\alpha = 0$ and $\alpha = \infty$ are not included. Δ

Despite the detail of this presentation, its main conclusion is not inaccessible: The most (musically) useful minima of the dissonance curve tend to be located at intervals α for which $f_i = \alpha f_j$, where f_i and f_j are arbitrary partials of the timbre F.

The theorems of this appendix assume that all partials are of equal amplitude. The effect of nonequal amplitudes is that some minima may disappear, some may appear, and others may shift slightly in frequency. Fortunately, these changes occur in a structured way. To be concrete, let the timbre F have partials $f_1, f_2, ..., f_n$ with amplitudes $a_1, a_2, ..., a_n$ and let $\hat{F}$ have the same set of partials but with amplitudes $1, 1, ..., 1$. As discussed above, the dissonance curve for $\hat{F}$ will have up to n^2 minima due to coinciding partials that occur at the intervals $\alpha_{ij} = f_i/f_j$. As the amplitudes a_j of F move away from unity, the depth of the dissonance curve at α_{ij} may change and the minima at some of the α_{ij} may disappear (an α_{ij} that is a minimum of $D_{\hat{F}}$ may not be a minimum of D_F), and other α_{ij} may appear (an α_{ij} that is not a minimum of $D_{\hat{F}}$ may be a minimum of D_F). Thus, amplitude variations of the partials tend to affect which of the α_{ij} happen to be minima. The dissonance curve also contains up to n^2 minima of the "broad" type. The location of these equilibria are less certain, because they move continuously with respect to variations in the a_j.

G

Analysis of the Time Domain Model

This appendix expands the model of Sect. 3.6 to account for more complex sounds and to reproduce the general dissonance curves (such as Figs. 6.1, 6.2, and 6.7) of Chap. 6. The model is then examined in some detail. This appendix is based on collaborative work with Marc Leman of IPEM [W: 16].

Recent time domain models of the pitch extraction mechanism (such as those of Patterson and Moore [B: 130] and Meddis [B: 111]) can successfully predict listeners' performance in a number of areas, including the pitch of the missing fundamental, pitch shift due to certain kinds of inharmonic components, repetition pitch, detection of the pitch of multiple tones sounding simultaneously, and musical applications such as harmony and tone center perception [B: 95]. These models typically consist of four steps:

(i) A critical band filtering that simulates the mechanical filtering in the inner and middle ear
(ii) A half wave rectification that simulates the nonlinear firing of hair cells
(iii) A periodicity extraction mechanism such as autocorrelation
(iv) A mechanism for aggregation of the within-band information

Similarly, the modeling of amplitude-modulation detector thresholds such as those of [B: 37] (and references therein) replace the third step (the pitch extraction schemes) with a "temporal modulation transfer function" and a "detector." The resulting systems can predict various masking effects and have been used to examine how the auditory system trades off spectral and temporal resolutions.

In contrast, models designed to predict the sensory dissonance of a collection of complex tones (such as in Chap. 6) typically begin with a spectral analysis that decomposes the sound into a collection of partials. When these partials are close to each other in frequency (but not identical), they beat in a characteristic way; when this roughness occurs at certain rates, it is called sensory dissonance. This appendix shows how sensory dissonance can be modeled directly in the time domain with a method that is closely related to the first two (common) steps of current pitch extraction and amplitude-modulation models.

The computational model of Sect. 3.6 contains an envelope detector followed by a bandpass filter. The simulations shown in Fig. 3.10 demonstrate

that the model can account for the dissonance curve generated from two pure sine waves. But this simple model breaks down when confronted with more complex wideband inputs. The source of the problem is that the envelope detector (the rectification nonlinearity followed by the LPF) only functions meaningfully on narrowband signals.[1] In keeping with (i)-(iv) above, Fig. G.1 suggests passing the input through a collection of bandpass filters (such as those in Fig. 3.5) that simulate the critical bands. This generates a series of narrowband signals to which the envelope detector can be applied, and it gives an approximation to the sensory dissonance within each critical band. The overall sensory dissonance can then be calculated by summing up all dissonances in all critical bands.

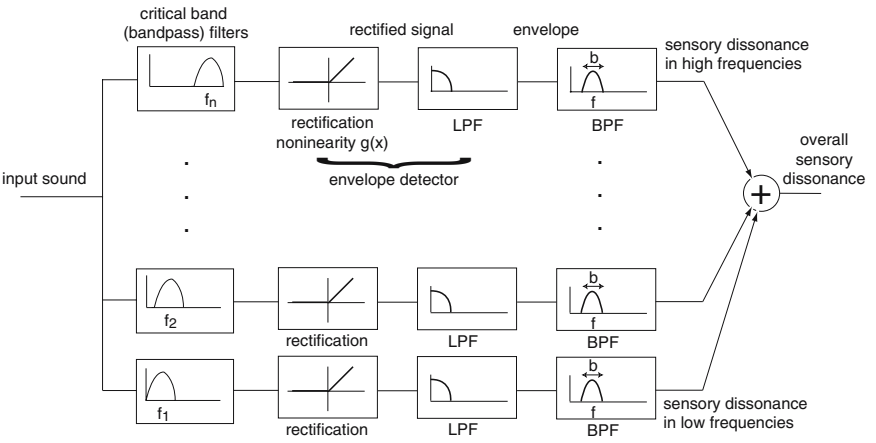


Fig. G.1. The n filters separate the input sound into narrowband signals with bandwidths that approximate the critical bands of the basilar membrane. The envelope detectors outline the beating within each critical band and the final bandpass filters accumulate the energy. Summing over all bands gives the overall sensory dissonance of the sound.

The core of the model lies in the rectification nonlinearity (where $g(x)$ is defined by equation (3.1) on p. 48). Physically, this originates from the hair cells of the basilar membrane, which are mechanically constrained to certain kinds of oscillation, and for which there is considerable neurophysiological evidence [B: 156]. The effect of the subsequent bandpass filtering is to remove both the lowest frequencies (which correspond perceptually to slow, pleasant beats and the sensation of loudness) and the higher frequencies (which correspond to the fundamentals, overtones, and summation tones). The energy of the signal in the passband is then proportional to the amount of roughness, or sensory dissonance due to the interactions of frequencies within the given

[1] This generic property of envelope detectors is discussed in [B: 76].

critical band. Summing these energies from all critical bands gives an overall measure of the sensory dissonance of the sound.

To see how this model works, consider the case where two sine waves at frequencies w_1 and w_2 pass through the same critical band filter at equal intensities. For w_1 near (but not equal) to w_2, this results in beats as shown in Fig. G.2(a). After passing through the rectification stage, this becomes the $r(t)$ as shown in G.2(b). To be concrete, suppose that the input $x(t)$ is the sum of the two sinusoids $\sin(w_1 t)$ and $\sin(w_2 t + \pi)$. The rectification nonlinearity $g(x)$ of (3.1) can be rewritten

$$g(x(t)) = \frac{1}{2}x(t) + \frac{1}{2}|x(t)|$$

and so

$$\begin{aligned} r(t) &= g(\sin(w_1 t) + \sin(w_2 t + \pi)) \\ &= \frac{1}{2}(\sin(w_1 t) + \sin(w_2 t + \pi)) + \frac{1}{2}|\sin(w_1 t) + \sin(w_2 t + \pi)| \\ &= \frac{1}{2}\sin(w_1 t) + \frac{1}{2}\sin(w_2 t + \pi) + |\sin(v_1 t)\sin(v_2 t + \frac{\pi}{2})| \end{aligned}$$

where $v_1 = \frac{w_1 - w_2}{2}$ and $v_2 = \frac{w_1 + w_2}{2}$ are assumed commensurate.

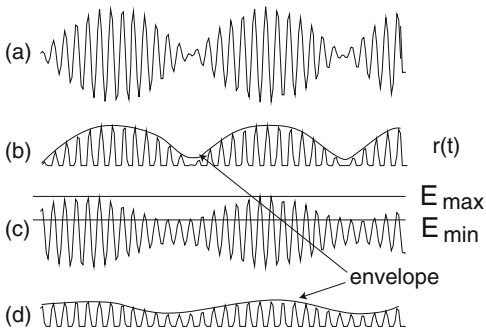


Fig. G.2. The beating of sine waves. (a) shows the sum of two sine waves of equal amplitude, which is rectified to give (b). (c) shows the sum of two sine waves of unequal amplitude, which is rectified to give (d).

Accordingly, the magnitude spectrum of $r(t)$ can be calculated as

$$\mathcal{F}\{r(t)\} = \frac{1}{2}\mathcal{F}\{\sin(w_1 t)\} + \frac{1}{2}\mathcal{F}\{\sin(w_2 t + \pi)\} + \mathcal{F}\{|\sin(v_1 t)|\} * \mathcal{F}\{|\sin(v_2 t + \frac{\pi}{2})|\},$$

where $*$ is the convolution operator. The Fourier series for $|\sin(v_1 t)|$ is

$$\frac{2}{\pi} - \frac{4}{\pi}\sum_{r=1}^{\infty}\frac{\cos(2rv_1 t)}{4r^2 - 1},$$

and so the magnitude spectrum consists of spikes at the even harmonics of v_1. Similarly, the Fourier series of $|\sin(v_2 t + \frac{\pi}{2})|$ has a magnitude spectrum

consisting of spikes at the even harmonics of v_2. As $w_1 \approx w_2$, $v_1 << v_2$ and the convolution of $\mathcal{F}\{|\sin(v_1 t)|\}$ with $\mathcal{F}\{|\sin(v_2 t + \frac{\pi}{2})|\}$ consists of a cluster of spikes near zero (these have magnitude $\frac{4}{\pi(4n^2-1)}$ at frequencies $2nv_1$) and similar clusters near nv_2 for all integers n.

From Fig. G.1, the rectification is followed by a bandpass filter with passband frequencies considerably less than w_1, w_2, and v_2. Hence, only the spikes near zero contribute significantly to the energy of $BPF\{r(t)\}$. Summing these terms over the frequency region of interest gives

$$d(v_1) = \sum_{\frac{f_1}{2v_1} \le n \le \frac{f_2}{2v_1}} \frac{4}{\pi(4n^2 - 1)}, \tag{G.1}$$

where f_1 and f_2 define the cutoff frequencies of the bandpass filter and $2v_1$ is the difference frequency. This function $d(v_1)$ represents the energy of the beating sinusoids within the critical band. Clearly, $d(v_1)$ is a function of the (difference between the) frequencies of the two input sine waves.

The following heuristic argument explains how (G.1), which provides a time domain analog of (E.2), qualitatively reproduces sensory dissonance curves. For $v_1 = 0$ (equivalently, $w_1 = w_2$), there are no terms in the sum and $d(v_1) = 0$. Consider fixing w_1 and varying w_2. As w_2 increases, v_1 increases and more terms (initially) enter into the sum (G.1), increasing $d(v_1)$. Eventually, however, v_1 increases past some critical value and the range $(\frac{f_1}{2v_1}, \frac{f_2}{2v_1})$ compresses so that fewer and fewer terms are summed in (G.1). Asymptotically, $d(v_1)$ returns to zero. Hence, $d(v_1)$ has a shape that is qualitatively like the measured dissonance curves such as shown in Fig. 3.7. The cutoff frequencies f_1 and f_2 of the bandpass filter must therefore be chosen so that the maximum of this sum occurs at the measured value d^* of maximum sensory dissonance.

Next, suppose that the two input waves are of unequal amplitudes,

$$s(t) = \alpha_1 e^{jw_1 t} + \alpha_2 e^{jw_2 t},$$

where again the frequencies of the (complex) sinusoids are w_1 and w_2, and $w_2 > w_1 >> w_2 - w_1$. If $B(w)$ represents the frequency response of the critical band (and other pre-rectification) filters then the signal entering the rectification is

$$\begin{aligned} &\alpha_1 B(w_1) e^{jw_1 t} + \alpha_2 B(w_2) e^{jw_2 t} \\ &= e^{jw_1 t}[\alpha_1 B(w_1) + \alpha_2 B(w_2) e^{j(w_2 - w_1)t}]. \end{aligned}$$

The $e^{jw_1 t}$ term is the "carrier" and the bracketed term is the envelope, which achieves its maximum and minimum at

$$E_{\max} = \frac{1}{2}(|\alpha_1 B(w_1)| + |\alpha_2 B(w_2)|)$$

$$E_{\min} = \frac{1}{2}(||\alpha_1 B(w_1)| - |\alpha_2 B(w_2)||)$$

as shown in Fig. G.2(c).

The previous analysis can now be repeated with $r(t)$ redefined as

$$r(t) = E_{\min}\ y(t) + (E_{\max} - E_{\min})\ x(t)y(t).$$

As the Fourier Series of a sum is the sum of the Fourier Series, the net effect is to increase the amplitudes of the spikes at nv_2 and to scale the sum in (G.1) by the constant $E_{\max} - E_{\min}$.

This weighting is incorporated into the dissonance model (E.2) by assuming that the roughness is proportional to the loudness of the beating. The amplitude of the beats is proportional to $E_{\max} - E_{\min}$, ignoring the effect of the filters $B(\cdot)$.[2] If $\alpha_1 > \alpha_2$, then $E_{\max} - E_{\min} = \frac{1}{2}(\alpha_1+\alpha_2) - \frac{1}{2}(\alpha_1-\alpha_2) = \alpha_2$. Similarly, if $\alpha_2 > \alpha_1$, $E_{\max} - E_{\min} = \frac{1}{2}(\alpha_1 + \alpha_2) - \frac{1}{2}(\alpha_2 - \alpha_1) = \alpha_1$. Hence $E_{\max} - E_{\min} = \min(\alpha_1, \alpha_2)$. Thus, the amplitude of the beating is given by the minimum of the two amplitudes.

As the disparity in the amplitudes of the partials increases, the dissonance $d(v_1)$ decreases and the maximum sensory dissonance occurs when the partials have equal amplitudes. Thus, the time-based model of sensory dissonance naturally accounts for the varying amplitudes of the partials of a sound.

To summarize this analysis: The time-based model of sensory dissonance can qualitatively reproduce the sensory dissonance curves such as are found in Plomp and Levelt [B: 141] and [B: 79] and makes concrete predictions regarding amplitude effects. Details of the shape of the dissonance curves will depend on the cutoff frequencies of the bandpass filters, their shape, and the integration time. As the model uses many of the building blocks of standard auditory models, it is not unreasonable to view sensory dissonance as a byproduct (or coproduct) of these neural elements.

[2] This is reasonable because the important beating (from the point of view of the dissonance calculation) is at the low frequencies near DC.

H

Behavior of Adaptive Tunings

This appendix derives concrete expressions for the update terms of the adaptive tuning algorithm and gives detailed statements and proofs of the results. The cost function

$$D = \sum_{i,j} D_F(\frac{f_i}{f_j}) \tag{H.1}$$

can be rewritten as

$$D = \frac{1}{2} \sum_{l=1}^{m} \sum_{k=1}^{m} \sum_{p=1}^{n} \sum_{q=1}^{n} d(a_p f_l, a_q f_k, v_p, v_q). \tag{H.2}$$

Only the terms in D that include f_i need to be considered when calculating the gradient $\frac{dD}{df_i}$. Thus, $\frac{dD}{df_i}$ is equal to

$$\frac{d}{df_i} \left[\frac{1}{2} \sum_{k=1}^{m} \sum_{p=1}^{n} \sum_{q=1}^{n} d(a_p f_i, a_q f_k, v_p, v_q) + \frac{1}{2} \sum_{k=1}^{m} \sum_{p=1}^{n} \sum_{q=1}^{n} d(a_p f_k, a_q f_i, v_p, v_q) \right]$$

$$= \sum_{k=1}^{m} \sum_{p=1}^{n} \sum_{q=1}^{n} \frac{d}{df_i} d(a_p f_i, a_q f_k, v_p, v_q) \tag{H.3}$$

because $d(f, g, v, w) = d(g, f, v, w)$ and the derivative commutes with the sums. Calculating the derivative of the individual terms $\frac{d}{df_i} d(f, g, v, w)$ in (H.3) is complicated by the presence of the absolute value and min functions in (E.2) and (E.3). The function is not differentiable at $f = g$ and changes depending on whether $f > g$ or $g > f$. Letting x^* be the point at which maximum dissonance occurs, define the function $\frac{d}{df} d(f, g, v, w)$ as

$$\begin{cases} \min(v, w) \left[\frac{-b_1 x^*}{(f s_1 + s_2)} e^{\left(\frac{b_1 x^* (f-g)}{f s_1 + s_2} \right)} + \frac{b_2 x^*}{(f s_1 + s_2)} e^{\left(\frac{b_2 x^* (f-g)}{f s_1 + s_2} \right)} \right] & \text{if } f > g \\ \min(v, w) \left[\frac{b_1 x^* (g s_1 + s_2)}{(f s_1 + s_2)^2} e^{\left(\frac{b_1 x^* (f-g)}{f s_1 + s_2} \right)} - \frac{b_2 x^* (g s_1 + s_2)}{(f s_1 + s_2)^2} e^{\left(\frac{b_2 x^* (f-g)}{f s_1 + s_2} \right)} \right] & \text{if } f < g \\ 0 & \text{if } f = g \end{cases}$$

which is a close approximation to the desired derivative. Then an approximate gradient is readily computable as the triple sum (H.3) of elements of the form $\frac{d}{df}d(f,g,v,w)$.

To streamline the results, the same simplifications and notations are made as in the previous appendices. The first theorem demonstrates the behavior of the algorithm when adapting two notes of equal loudness, each consisting of a single partial. Figure 8.5 on p. 165 shows this pictorially.

Theorem H.1. *Let f_0 and g_0 be the frequencies of two sine waves, with $f_0 < g_0$. Apply the adaptive tuning algorithm. Then*
(i) $g_0 > (1-s_1)f_0 - s_2$ implies that $|g_{k+1} - f_{k+1}| > |g_k - f_k|$ for all k,
(ii) $g_0 < (1-s_1)f_0 - s_2$ implies that $|g_{k+1} - f_{k+1}| < |g_k - f_k|$ for all k.

Proof: From the form of $\frac{d}{df}d(f,g,v,w)$, the updates for f and g are:

$$f_{k+1} = f_k - \frac{\mu x^*(g_k s_1 + s_2)}{(f_k s_1 + s_2)^2}\left[b_1 e^{\left(\frac{b_1 x^*(f_k - g_k)}{f_k s_1 + s_2}\right)} - b_2 e^{\left(\frac{b_2 x^*(f_k - g_k)}{f_k s_1 + s_2}\right)}\right]$$

$$g_{k+1} = g_k + \frac{\mu x^*}{(f_k s_1 + s_2)}\left[b_1 e^{\left(\frac{b_1 x^*(f_k - g_k)}{f_k s_1 + s_2}\right)} - b_2 e^{\left(\frac{b_2 x^*(f_k - g_k)}{f_k s_1 + s_2}\right)}\right]$$

The terms in brackets are positive whenever

$$\ln(b_1) + \frac{b_1 x^*(f_k - g_k)}{f_k s_1 + s_2} > \ln(b_2) + \frac{b_2 x^*(f_k - g_k)}{f_k s_1 + s_2}.$$

Rearranging gives

$$\frac{\ln(b_1) - \ln(b_1)}{b_1 - b_2} > \frac{x^*(f_k - g_k)}{f_k s_1 + s_2}.$$

As the left-hand side is equal to x^*, this can be rewritten

$$f_k s_1 + s_2 > f_k - g_k.$$

Thus, $g_k > (1-s_1)f_k - s_2$ implies that $g_{k+1} > g_k$. Similarly, $f_{k+1} < f_k$, which together show (a). On the other hand, if $g_k < (1-s_1)f_k - s_2$, an identical argument shows that $g_{k+1} < g_k$ and $f_{k+1} > f_k$ for all k. Δ

The next result is the theoretical counterpart of Fig. 8.6 on p. 166.

Theorem H.2. *Consider two notes F and G. Suppose that F consists of two partials fixed at frequencies f and αf with $\alpha > 1$, and that G consists of a single partial at frequency g_0 that is allowed to adapt via the adaptive tuning algorithm. Assuming that all partials are of equal loudness:*

(i) There are three stable equilibria: at $g = f$, at $g = \alpha f$ and at $g = (1+\alpha)f/2$.
(ii) If $g_0 << f$, then $|g_{k+1} - f| > |g_k - f|$ for all k.
(iii) If $g_0 >> \alpha f$, then $|g_{k+1} - \alpha f| > |g_k - \alpha f|$ for all k.

Proof: The total dissonance for this case includes three terms: $D_{total} = d(f,g) + d(f,\alpha f) + d(g,\alpha f)$. As α and f are fixed, $d(f,\alpha f)$ is constant, and minimizing D_{total} is the same as minimizing $d(f,g) + d(g,\alpha f)$. Using the simplified dissonance measure (E.1) in place of the more complete model (E.2)-(E.4), and assuming $f < g < \alpha f$, the update for g is

$$g_{k+1} = g_k - \mu \left[b_1 \mathrm{e}^{-b_1(\alpha f - g_k)} - b_2 \mathrm{e}^{-b_2(\alpha f - g_k)} - b_1 \mathrm{e}^{-b_1(g_k - f)} + b_2 \mathrm{e}^{-b_2(g_k - f)} \right].$$

This has an equilibrium when $\alpha f - g_k = g_k - f$, that is, when $g = \frac{(1+\alpha)}{2} f$. Calculation of the second derivative shows that it is positive at this point as long as $f/2(\alpha - 1) >> 1$, which holds for all reasonable f and α. Hence this is a stable equilibrium. (Note that if the complete model is used, then a much more complex update develops for g. This will have an equilibrium near, but not at, $(1+\alpha)f/2$.)

Due to the nondifferentiability of the dissonance function at $f = g$, it is not possible to simply take the derivative at this point. The strategy to show that $f = g$ is stable is to show that if $g = f + \epsilon$ for some small $\epsilon > 0$ then the update decreases g, whereas if $g = f - \epsilon$ for some small $\epsilon > 0$ then the update increases g. Supposing that $g > f$, and assuming that $f(\alpha - 1) >> 1$, the gradient is approximately

$$b_1 \mathrm{e}^{-b_1 f(\alpha-1)} - b_2 \mathrm{e}^{-b_2 f(\alpha-1)} - b_1 + b_2.$$

As b_2 is about twice the size of b_1, this is positive. Similarly, for $g = f - \epsilon$, the gradient is approximately

$$b_1 \mathrm{e}^{-b_1 f(\alpha-1)} - b_2 \mathrm{e}^{-b_2 f(\alpha-1)} + b_1 - b_2,$$

which is negative. Consequently, $f = g$ is a local stable point. The point where $\alpha f = g$ is analyzed similarly. Analogous arguments to those used in theorem H.1 show that for $g << f$, g decreases, and for $g >> \alpha f$, g increases. Δ

I

Symbolic Properties of $\oplus$-Tables

Although $\oplus$-tables do not form any recognizable algebraic structure, they do have several features that would be familiar to an algebraist. For instance, the tables have an identity element, the operation $\oplus$ is commutative, and it is associative when it is well defined. These are used to derive a set of properties that can help make intelligent choices in the symbolic timbre construction procedure.

Given any set of scale intervals S, the $\oplus$-table derived from S has the following characteristics.

Identity: The "octave" or unit of repetition s^* acts as an identity element, i.e.,

$$s^* \oplus s = s \oplus s^* = s \quad \forall s \in S.$$

Commutativity: The $\oplus$-table is symmetric, i.e.,

$$s_1 \oplus s_2 = s_2 \oplus s_1 \quad \forall s_1, s_2 \in S. \tag{I.1}$$

If one side of (I.1) is undefined (is "equal" to $*$), then so is the other. Commutativity of $\oplus$ follows directly from the commutativity of products of powers of real numbers.

Associativity: The $\oplus$ operator is associative whenever it is well defined. Thus

$$(s_1 \oplus s_2) \oplus s_3 = s_1 \oplus (s_2 \oplus s_3) \quad \forall s_1, s_2, s_3 \in S, \tag{I.2}$$

provided that both sides of (I.2) exist.

It is indeed possible for one side of (I.2) to exist but not the other.

Example: Consider the tetrachordal scale with $\oplus$-table 12.5 on p. 262. Observe that $((2,1,1) \oplus (1,0,0)) \oplus (2,1,0)$ is well defined and equals $(1,0,0)$, but that $(2,1,1) \oplus ((1,0,0) \oplus (2,1,0))$ does not exist because $(1,0,0) \oplus (2,1,0)$ is disallowed. To further emphasize how unusual this construction is, observe

that by commutativity, $(2,1,1) \oplus (1,0,0) = (1,0,0) \oplus (2,1,1)$. Substituting this in the above calculation gives $((1,0,0) \oplus (2,1,1)) \oplus (2,1,0)$, which is indeed equal to $(1,0,0) \oplus ((2,1,1) \oplus (2,1,0))$, because both sides are $(1,0,0)$.

The remaining properties of ⊕-tables concern "solutions" to the ⊕-equation defined in the symbolic timbre construction procedure

$$s_i = s_j \oplus r_{i,i-j}. \tag{I.3}$$

Recall that in the procedure, a set of s_j are given (which are defined by previous choices of the t_j). The goal is to find a single s_i such that the equation (I.3) is well defined for all j up to $i-1$. The properties of ⊕-tables can help pinpoint viable solutions to (I.3).

Theorem I.1. *Suppose that $s_j \in S$ have been chosen for all $j < k$. Let $\mathbf{S_j}$ be the set of all non-$*$ entries in the s_j column of the ⊕-table. Then for all $i \geq k$, s_i must be an element of $\bigcap_{j<k} \mathbf{S}_j$.*

Proof: First consider the case $i = k = 2$, with s_1 specified. Then (I.3) requires choice of s_2 such that $s_2 = s_1 \oplus r_{1,1}$ for some $r_{1,1}$. Such $r_{1,1}$ will exist exactly when $s_2 \in \mathbf{S_1}$. For $i > 2$, $s_i = s_1 \oplus r_{i,i-j}$ must be solvable, which again requires that $s_i \in \mathbf{S_1}$. The general case $s_i = s_j \oplus r_{i,i-j}$ is similarly solvable exactly when $s_i \in \mathbf{S_j}$. As this is true for every $j < k$, $s_i \in \bigcap_{j<k} \mathbf{S}_j$. Δ

Thus, when building timbres according to the procedure, the set $\mathcal{S}^k = \bigcap_{j<k} \mathbf{S}_j$ defines the allowable partials at the kth step. Clearly, $\mathcal{S}^k$ can never grow larger because $\mathcal{S}^k \supset \mathcal{S}^{k+1}\ \forall k$, and it may well become smaller as k increases. This demonstrates that the order in which the partials are chosen is crucial in determining whether a perfect timbre is realizable.

The easiest way to appreciate how the theorem I.1 simplifies (and limits) the selection problem is by example.

Example: In Table 12.1 on p. 257, once $s_i = (3,2)$ for some i, then for all $k > i$, s_k must be $(3,2)$, $(1,0)$, or $(2,1)$.

Example: In Table 12.3 on p. 260, once $s_i = (2,0)$ has been chosen, then for all $k > i$, s_k must be either $(2,0)$, $(4,1)$, or $(5,1)$. In particular, no s_k can be the identity $(0,0)$.

Corollary I.2. *Suppose that an element $\hat{s} \in S$ appears in every column of the ⊕-table. Then for any choice of s_j, $j < i$, (I.3) is always solvable with $s_i = \hat{s}$.*

Proof: As $\hat{s}$ is in every column of the table, $\hat{s} \in \mathbf{S}_j\ \forall j$ and hence $\hat{s} \in \bigcap_{j<k} \mathbf{S}_j$ for any k. Δ

In other words, for any $s \in S$, there is always a $r \in S$ such that $\hat{s} = s \oplus r$, and so $\hat{s}$ is always permissible.

Example: In Table 12.5 on p. 262, the identity $s^* = (0,0,0)$ appears in every column. Thus, it is always possible to choose a partial t_i with the equivalence class s^* at any step.

Suppose, on the other hand, that an element $\bar{s} \in S$ appears nowhere in the ⊕-table other than in the column and row of the identity. Then $\bar{s}$ cannot be used to define one of the s_i because $\bar{s} \notin \mathbf{S}_k$ for any k and so for any $s_i \neq s^*$, $s_i = \bar{s} + r$ has no solution. Although $\bar{s}$ cannot occur among the s_i, it is still possible that it might appear among the $r_{i,k}$. Indeed, it will need to in order to find a complete timbre.

Example: The element $\bar{s} = (2, 1)$ appears nowhere in ⊕-table 12.3 (from p. 260) defined by the Pythagorean scale. The timbre was made complete by ensuring that $\bar{s}$ appears among the $r_{i,k}$ of Table 12.4 of p. 260.

Another property of ⊕-tables is that elements are arranged in "stripes" from southwest to northeast. For instance, in Table 12.3 of p. 260, a stripe of $(4, 1)$ elements connects the $4, 1$ entry with the $1, 4$ entry. Similarly, a stripe of $(3, 1)$ elements connect the $3, 1$ with the $1, 3$ entries, although the stripe is broken up by a $*$. The fact that such (possibly interrupted) stripes must exist is the content of the next theorem.

Given an m note scale S, the entries of the corresponding ⊕-table can be labeled as a matrix $\{a_{j,k}\}$ for $j = 1, 2, ..., m$ and $k = 1, 2, ..., m$. Let P_i denote the ith stripe of the ⊕-table, that is, $P_i = \{a_{j,k}\}$ for all j and k with $j + k = i + 1$.

Example: For the Pythagorean ⊕-table:

$$P_1 = \{(0,0)\}, \quad P_2 = \{(1,0),\ (0,1)\}, \quad P_3 = \{(2,0),\ (2,0),\ (2,0)\},$$
$$P_4 = \{(2,1),\ *,\ *,\ (2,1)\}, \quad P_4 = \{(3,1),\ (3,1),\ *,\ (3,1),\ (3,1),\}, \quad \text{etc.}$$

Theorem I.3. *For each i, all non-$*$ elements of the stripe P_i are identical.*

Proof: By construction, the elements s_i and $s_{i+1} \in S$ are integer vectors, and they may be ordered so that

$$s_{i+1} = s_i + e_{j,i}\ \forall i, \tag{I.4}$$

where $e_{j,i}$ is a unit vector with zeroes everywhere except for a single 1 in the jth entry. Let $\Sigma(s_i)$ represent the sum of the entries in $s_i = (\sigma_1, \sigma_2, , ..., \sigma_p)$, i.e., $\Sigma(s_i) = \sum_{j=1}^{p} \sigma_j$, and let Σ^* represent the sum of the entries in the element that forms the unit of repetition. Because the ⊕ operation adds powers of the generating intervals,

$$\Sigma(s_j \oplus s_k) = \Sigma(s_j) + \Sigma(s_k) \pmod{\Sigma^*} \tag{I.5}$$

whenever $s_j \oplus s_k$ is well defined. Because of the ordering, the entries in the stripe P_i can be written

$$s_j \oplus s_k, \quad s_{j-1} \oplus s_{k+1}, \quad s_{j-2} \oplus s_{k+2}, \cdots$$

for all positive j and k with $j + k = i + 1$. Hence,

$$\Sigma(s_j \oplus s_k) = \Sigma(s_{j-1} \oplus s_{k+1}) = \cdots \tag{I.6}$$

whenever these are defined. From (I.4), $\Sigma(s_j) = \Sigma(s_k)$ implies that $s_j = s_j$. Hence (I.6) shows that $s_j \oplus s_k = s_{j-1} \oplus s_{k+1} = \cdots$ whenever the terms are defined, and hence all well-defined elements of the stripe are identical. Δ

This is useful because stripes define whether a given choice for the t_i (and hence s_i) is likely to lead to complete timbres. Suppose that $\tilde{s}$ is a candidate for s_i at the ith step. Whether $\tilde{s}$ will "work" for all previous s_j (i.e., whether $\tilde{s} = s_j \oplus r$ has solutions for all s_j) depends on whether $\tilde{s}$ appears in all corresponding $\mathbf{S}_j$. Theorem I.3 pinpoints exactly where $\tilde{s}$ must appear; at the intersection of the column $\mathbf{S}_j$ and the stripe containing $\tilde{s}$. Thus, the procedure can be implemented without conducting a search for $\tilde{s}$ among all possible columns.

A special case is when a column is "full," i.e., when it contains no $*$ entries.

Theorem I.4. *Let $\mathbf{S}_f$ be a full column corresponding to $s_f \in S$. Then $s_i = s_f \oplus r_i$ is solvable for all $s_i \in S$.*

Proof: As there are m entries in the column $\mathbf{S}_f$ and there are m different s_i, it is only necessary to show that no entries appear twice. Using the ordering (I.4) of the previous proof, $\mathbf{S}_f$ has elements

$$s_1 \oplus s_f, \quad s_2 \oplus s_f, \ \cdots, \ s_m \oplus s_f, \tag{I.7}$$

which are well defined by assumption. Now proceed by contradiction, and suppose that the ith and jth elements of (I.7) are the same, i.e., $s_i \oplus s_f = s_j \oplus s_f$. Then

$$\Sigma(s_i \oplus s_f) = \Sigma(s_j \oplus s_f) \pmod{\Sigma^*}$$

(where Σ and Σ^* were defined in the previous proof). This implies that

$$\Sigma(s_i) + \Sigma(s_f) = \Sigma(s_j) + \Sigma(s_f) \pmod{\Sigma^*}$$

which implies that $\Sigma(s_i) = \Sigma(s_j) \pmod{\Sigma^*}$. By the same argument as in the proof of theorem I.3, this implies that $s_i = s_j$. But each s_i appears exactly once in (I.7), which gives the desired contradiction. Δ

Thus, when a column is full, it must contain every element. In this case, equation (I.3) puts no restrictions on the choice of s_i. Let $\{s_j\}$ be all elements of S that have full columns. Then a $\oplus$-subtable can be formed by these $\{s_j\}$ that has no illegal $*$ entries. For example, Table 12.1 on p. 257 is generated by the the *ab*-cubed scale. The elements $(0,0)$, $(1,1)$, and $(2,2)$ have full columns and hence can be used to form a full $\oplus$-subtable. It is easy to generate perfect timbres for such full $\oplus$-subtables because equation (I.3) puts no restrictions on the choice of partials for a complementary timbre. Whether these extend to all elements of the scale, however, depends heavily on the structure of the non-full part of the table. Finding timbres for full subtables is exactly the same as finding timbres for equal temperaments, whose $\oplus$-tables have no disallowed $*$ entries. In fact, full $\oplus$-tables form a commutative group, which may explain why the equal-tempered case is relatively easy to solve.

All of the above properties were stated in terms of the columns of the ⊕-table. By commutativity, the properties could have been stated in terms of the corresponding rows.

From a mathematical point of view, the symbolic timbre selection procedure raises a number of interesting issues. The operation ⊕ defined here is not any kind of standard mathematical operator because of the disallowed $*$ entries. Yet ⊕-tables clearly have a significant amount of structure. For instance, any ⊕-table can be viewed as a subset of the commutative group of integer m vectors $(\sigma_1, \sigma_2, ..., \sigma_m)$ where the ith entry is taken mod n_i, from which certain elements have been removed. Can this structure be exploited? Another obvious question concerns the possibility of decomposing ⊕-tables in the same kind of ways that arbitrary groups are decomposed into normal subgroups. Might such a decomposition allow the building up of spectra for larger scales in terms of spectra defined for simpler scales?

J

Harmonic Entropy

Harmonic entropy is a measure of the uncertainty in pitch perception, and it provides a physical correlate of tonalness, one aspect of the psychoacoustic concept of dissonance. This Appendix shows in detail how to calculate harmonic entropy and continues the discussion in Sect. 5.3.3.

Harmonic entropy was introduced by Erlich [W: 9] as a refinement of a model by van Eck [B: 125]. It is based on Terhardt's [B: 196] theory of harmony, and it follows in the tradition of Rameau's fundamental bass [B: 145]. It provides a way to measure the uncertainty of the fit of a harmonic template to a complex sound spectrum. As a major component of tonalness is the closeness of the partials of a complex sound to a harmonic series, high tonalness corresponds to low entropy and low tonalness corresponds to high entropy.

In the simplest case, consider two harmonic tones. If the tones are to be understood as approximate harmonic overtones of some common root, they must form a simple-integer ratio with one another. One way to model this uses the Farey series $\mathcal{F}_n$ of order n, which lists all ratios of integers up to n. For example, $\mathcal{F}_6$ is

$$\frac{0}{1}, \quad \frac{1}{6}, \quad \frac{1}{5}, \quad \frac{1}{4}, \quad \frac{1}{3}, \quad \frac{2}{5}, \quad \frac{1}{2}, \quad \frac{3}{5}, \quad \frac{2}{3}, \quad \frac{3}{4}, \quad \frac{4}{5}, \quad \frac{5}{6}, \quad \frac{1}{1}.$$

A useful property of the Farey series is that the distance between successive terms is larger when the ratios are simpler. Let the jth element of the series be $f_j = \frac{a_j}{b_j}$. Then the region over which f_j dominates goes from the mediant[1] below to the mediant above, that is, from $\frac{a_{j-1}+a_j}{b_{j-1}+b_j}$ to $\frac{a_j+a_{j+1}}{b_j+b_{j+1}}$. Designate this region r_j. Figure J.1 plots the length of r_j vs. f_j for $\mathcal{F}_{50}$, the Farey series of order 50. Observe that complex ratios cluster together, and that the simple ratios tend to separate. Thus, simple ratios like 1/2, 2/3, and 3/4 have wide regions with large r_j, and complex ratios tend to have small regions with small r_j.

For any interval i, a Gaussian distribution (a bell curve) is used to associate a probability $p_j(i)$ with the ratio f_j in $\mathcal{F}_n$. The probability that interval i is perceived as a mistuning of the jth member of the Farey series is

$$p_j(i) = \frac{1}{\sigma\sqrt{2\pi}} \int_{t \in r_j} e^{-(t-i)^2/2\sigma^2} dt.$$

[1] Recall that the mediant of two ratios $\frac{a}{b}$ and $\frac{c}{d}$ is the fraction $\frac{a+c}{b+d}$.

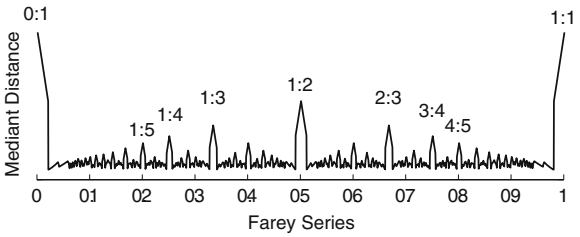


Fig. J.1. The mediant distances between entries (the length of the r_j) are plotted as a function of the small integer ratios f_j drawn from the Farey series of order 50. The simplest ratios dominate.

Thus, the probability is high when the i is close to f_j and low when i is far from f_j. This is depicted in Fig. J.2 where the probabilities that i is perceived as f_{j+1}, f_{j+2}, and f_{j+3} are shown as the three regions under the bell curve. Erlich refines this model to incorporate the log of the intervals and mediants, which is sensible because pitch perception is itself (roughly) logarithmic.

The harmonic entropy (HE) of i is then defined (parallel to the definition of entropy used in information theory) as

$$HE(i) = -\sum_j p_j(i) \log(p_j(i)).$$

When the interval i lies near a simple-integer ratio f_j, there will be one large probability and many small ones. Harmonic entropy is low. When the interval i is distant from any simple-integer ratio, many complex ratios contribute many nonzero probabilities. Harmonic entropy is high. A plot of harmonic entropy over an octave of intervals i (labeled in cents) appears in Fig. 5.5 on p. 92. This figure used $\mathcal{F}_{50}$ and $\sigma = 0.007$. Clearly, intervals that are close to simple ratios are distinguished by having low entropy, and more complex intervals have high harmonic entropy.

Generalizations of the harmonic entropy measure to consider more than two sounds at a time are currently under investigation; one possibility involves Voronoi cells. Harmonic series triads with simple ratios are associated with large Voronoi cells, whereas triads with complex ratios are associated with small cells. This nicely parallels the dyadic case. Recall the example (from p. 100 and sound examples [S: 40]–[S: 42]), which compares the clusters 4:5:6:7 with 1/7:1/6:1/5:1/4. In such cases, the harmonic entropy model tends to agree better with listener's perceptions of the dissonance of these chords than does the sensory dissonance approach. Paul Erlich comments that the study of harmonic entropy is a "public work in progress" at [W: 9].

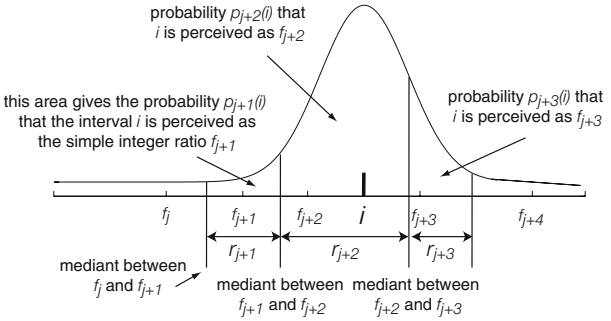


Fig. J.2. Each region r_{j+1} extends from the mediant between f_j and f_{j+1} to the mediant between f_{j+1} and f_{j+2}. The interval i specifies the mean of the Gaussian curve, and the probabilities $p_j(i)$ are defined as the disjoint areas between the axis and the curve.

K

Fourier's Song

Also known as Table 4.1: Properties of the Fourier Transform, Fourier's Song *was written by Bob Williamson and Bill Sethares "because we love Fourier Transforms, and we know you will too." Perhaps you have never taken a course where everything is laid out in a single song. Well, here it is...a song containing 17% of the theoretical results, 25% of the practical insights, and 100% of the humor of ECE330: Signals and Systems. The music is played in an additive (overtone) scale that consists of all harmonics of 100 Hz. It appears on the CD in* `sounds/Chapter04/fouriersong.mp3`*; see [S: 34]. There will be a test in the morning.*

Integrate your function times a complex exponential.
It's really not so hard you can do it with your pencil.
And when you're done with this calculation,
You've got a brand new function—the Fourier Transformation.

What a prism does to sunlight, what the ear does to sound,
Fourier does to signals, it's the coolest trick around.
Now filtering is easy, you don't need to convolve,
All you do is multiply in order to solve.

From time into frequency—from frequency to time

Every operation in the time domain
Has a Fourier analog – that's what I claim.
Think of a delay, a simple shift in time,
It becomes a phase rotation—now that's truly sublime!

And to differentiate, here's a simple trick.
Just multiply by $j\omega$, ain't that slick?
Integration is the inverse, what you gonna do?
Divide instead of multiply—you can do it too.

From time into frequency—from frequency to time

Let's do some examples... consider a sine.
It's mapped to a delta, in frequency—not time.
Now take that same delta as a function of time,
Mapped into frequency, of course, it's a sine!

Sine x on x is handy, let's call it a sinc.
Its Fourier Transform is simpler than you think.
You get a pulse that's shaped just like a top hat...
Squeeze the pulse thin, and the sinc grows fat.
Or make the pulse wide, and the sinc grows dense,
The uncertainty principle is just common sense.

Exercise K.1. Find as many Fourier transform pairs as you can in the lyrics to *Fourier's Song*.

Exercise K.2. Find as many properties of the Fourier transform in the lyrics to *Fourier's Song* as you can.

Exercise K.3. Mathematically define the function that looks like a "top hat" and explain why its transform is the sinc.

Exercise K.4. Explain what property of the Fourier transform is used in the last verse when the sinc "grows fat" and "grows dense." Why does this relate to the uncertainty principle?

L

Tables of Scales

This appendix provides tables of several historical and ethnic tunings. Others can be found throughout the text. A number of meantone tunings are defined on p. 65, and several well temperaments appear on p. 65. A large variety of tunings and scales are derived and defined throughout the chapter "Musical Scales."

Table L.1. Historical tunings, with all values rounded to the nearest cent.

Tuning	cents										
12-tet	100	200	300	400	500	600	700	800	900	1000	1100
1/4 Comma A	76	193	310	386	503	580	697	772	890	1007	1083
Barca	92	197	296	393	498	590	698	794	895	996	1092
Barca A	92	200	296	397	498	594	702	794	899	998	1095
Bethisy	87	193	289	386	496	587	697	787	890	993	1087
Chaumont	76	193	289	386	503	580	697	773	890	996	1083
Corrette	76	193	289	386	503	580	697	783	890	996	1083
d'Alembert	87	193	290	386	497	587	697	787	890	994	1087
Kirnberger 2	90	204	294	386	498	590	702	792	895	996	1088
Kirnberger 3	90	193	294	386	498	590	697	792	890	996	1088
Marpourg	84	193	294	386	503	580	697	789	890	999	1083
Rameau b	93	193	305	386	503	582	697	800	890	1007	1083
Rameau ♯	76	193	286	386	498	580	697	775	890	993	1083
Valloti	90	196	294	392	498	588	698	792	894	996	1090
Vallotti A	90	200	294	396	498	592	702	792	898	996	1094
Werkmeister 3	90	192	294	390	498	588	696	792	888	996	1092
Werkmeister 4	82	196	294	392	498	588	694	784	890	1004	1086
Werkmeister 5	96	204	300	396	504	600	702	792	900	1002	1098

Table L.2. Tuning of each slendro instrument of Gamelan Swastigitha. All values are rounded to the nearest Hertz.

Gamelan Swastigitha: Slendro

	I					II						III						
Instrument	6	1	2	3	5	6	1	2	3	5	6	1	2	3	5	6	1	2
gender	118	133	155	178	206	236	271											
gender	121	135	155	178	205	234	271	310	358	412	471	542	623	719				
gender						236	265	310	358	412	471	542	623	719	825	950	1093	1266
saron							272	310	358	412	472	544						
saron												544	626	719	828	951	1094	1268
bonang							271	308	355	413	472	544	622	717	825	954	1094	1250
bonang											472	545	622	717	825	954	1094	1268
kenong									357	412	472		623					
gambang						238	272	311	361	415	475	545	626	725	828	956	1106	1276
median	120	134	155	178	205	236	271	310	358	412	472	544	623	719	825	954	1094	1268

Table L.3. Tuning of each slendro instrument of Gamelan Kyai Kaduk Manis. All values are rounded to the nearest Hertz.

Gamelan Kyai Kaduk Manis: Slendro

	I					II						III						
Instrument	6	1	2	3	5	6	1	2	3	5	6	1	2	3	5	6	1	2
gender	120	140	160	183	210	241	279	320	367	420	480	557	639	733				
gender						241	279	320	366	420	482	556	638	733	838	968	1114	1279
gender	120	139	159	182	209	240	277											
saron						241	280	322	367	421	482	557						
saron						244	281	322	369	423	482	557						
saron											482	559	651	738	840	968	1113	
saron											484	560	643	738	841	978	1129	1283
saron											483	569	641	739	853	985	1139	
bonang							281	322	367	423	484	560	641	736	837	966	1114	1268
bonang												557	643	736	838	972	1113	1281
kenong						242		320	369	421	478	557						
gambang			155	180	206	237	275	319	366	415	474	556	637	725	844	961	1112	1266
median	120	140	159	182	209	241	279	320	367	421	482	557	641	738	840	968	1114	1278

Table L.4. Tuning of each pelog instrument of Gamelan Swastigitha. All values are rounded to the nearest Hertz.

Gamelan Swastigitha: Pelog																									
	I							II							III										
Instrument		1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4
gender	120	151	160	174		222	234		299	324	354		443	471		599	643	709								
gender							240		300	322	354		444	474		600	642	709		887	950		1203	1305	1414	
gender		151	160	174	207	222	236	258																		
saron									300	326	354	415	445	472	524											
saron																602	645	709	829	890	953	1052				
saron																							1205	1312	1427	1674
bonang									300	324	353	415	444	472	525	599	645	711	820	886	950	1042				
bonang																602	643	708	828	887	950	1052	1205	1311	1427	1676
gambang			157	178		215	234	258		328	354		444	471	522		645	712		892	961	1047				
median	120	151	160	174	207	222	235	258	300	324	354	415	444	472	524	600	644	709	828	887	950	1050	1205	1311	1427	1675

Table L.5. Tuning of each pelog instrument of Gamelan Kyai Kaduk Manis. All values are rounded to the nearest Hertz.

Gamelan Kyai Kaduk Manis: Pelog

	I									II							III									
Instrument	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3
gender	120		149	164	180		225	241		303	332	361		451	480		604	661	717							
gender			149	164	179	210	223	241	264																	
gender								241	266		334	359		452	479	537		661	717		891	972	1073		1311	1427
gender								240		304	332	361		451	480		606	662	717		892	972		1213	1307	1425
gender	120	135		166	180		226	241	269		332	361		452	480	538		661	717							
saron										306	334	362	423	452	482	540										
saron												362	421	452	483	538										
saron																	618	672	733	860	898	988	1082			
saron																	612	668	729	844	904	991	1082			
saron																						974	1116	1233		1453
saron																	608	665	727	838	892	977	1101			
bonang										310	336	362	424	445	482	538	606	668	728	844	892	973	1074			
bonang																	604	682	732	840	892	976	1077	1219	1323	1428
kenong								242			332	362		454	478	536	611									
median	120	135	149	164	180	210	225	241	266	305	332	361	423	452	480	538	607	665	727	844	892	975	1082	1219	1311	1428

B: Bibliography

References in the body of the text to the bibliography are coded with [B:] *to distinguish them from references to the discography, sound and video examples, and websites.*

[B: 1] *1/1 Journal of the Just Intonation Network*, San Francisco, CA. [A journal for devotees of Just Intonation.]

[B: 2] E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines*, Wiley-Interscience (1990). [A method of global optimization that mimics the relaxation of metals as they cool.]

[B: 3] J. Aiken, "Just intonation with synth tuning tables," Keyboard, Nov. (1994). [Use of synthesizer tuning tables.]

[B: 4] C. Akkoç, "Non-Deterministic Scales Used in Traditional Turkish Music," Journal of New Music Research 31, No. 4, Swets and Zeitlinger, Dec. (2002). Figures 4.11 and 4.12 used with permission. [Analyzes Turkish maqam performances under the assumption that the underlying scale structure is stochastic in nature.]

[B: 5] S. Alexjander, "DNA tunings," Experimental Musical Instruments VIII, No. 2, Sept. (1992). [Uses data from DNA sequences to generate interesting tunings.]

[B: 6] American National Standards Institute, "USA standard acoustic terminology," S1.1 (1960). [Even more dated than you might think.]

[B: 7] D. Arnold, Ed. *New Oxford Companion to Music*, Oxford University Press, Oxford, England (1984). [Standard music dictionary.]

[B: 8] P. Y. Asselin, *Musique et Temperament*, Editions Costallat (1985).

[B: 9] *Balugan*, the newsletter of the American Gamelon Society, Box A63, Hanover, NH. [Keeping up-to-date with the gamelan.]

[B: 10] J. M. Barbour, *Tuning and Temperament*, Michigan State College Press (1951). [Demonstrates the "superiority" of 12-tet over and over again.]

[B: 11] J. Barnes, "Bach's keyboard temperament," Early Music 7, No. 2 (1979). [Uses internal evidence from the *Well Tempered Clavier* to try and determine probable tunings used by Bach.]

[B: 12] A. H. Benade, *Fundamentals of Musical Acoustics*, Dover Pubs., New York (1990). [Contains a great presentation of modes of vibration and the way musical instruments work.]

[B: 13] K. W. Berger, "Some factors in the recognition of timbre," J. Acoust. Soc. Am. 36:1888 (1963).

[B: 14] L. Bernstein, *The Unanswered Question*, Harvard Univesity Press, Cambridge MA (1976). [Leonard Bernstein on the nature of music.]

[B: 15] E. Blackwood, *Structure of Recognizable Diatonic Tunings,* Princeton University Press, Princeton NJ (1985). [Mathematics of scales, with special emphasis on equal temperaments.]

[B: 16] H. Bohlen, "13 tonstufen in der duodezeme," Acoustica 39, 76-86 (1978). [Scales based on the 13*th* root of 3 rather than the 12*th* root of 2.]

[B: 17] P. Boomsliter and W. Creel, "The long pattern hypothesis in harmony and hearing," J. Music Theory 5, No. 2, 2-30 (1961).

[B: 18] A. Bregman, *Auditory Scene Analysis*, MIT Press, Cambridge, MA (1990). [Focuses on investigations of tonal fusion and auditory streaming. Clear and readable.]

[B: 19] J. C. Brown, "Calculation of a constant Q spectral transform," J. Acoust. Soc. Am. 89, 425-434 (1991). [Unfortunately, the transform is not invertible.]

[B: 20] J. C. Brown, "Frequency ratios of spectral components of musical sounds," J. Acoust. Soc. Am. 99, 1210-1218 (1996). [Discusses the inharmonicities that may arise in nonidealized strings and air columns.]

[B: 21] E. M. Burns and W. D. Ward, "Intervals, scales, and tuning," in *The Psychology of Music,* ed. Diana Deutsch, Academic Press, New York (1982). [Overview of experimental work in the psychology of tunings and scales.]

[B: 22] D. Butler, *The Musician's Guide to Perception and Cognition*, Macmillan, Inc., NY (1992). [Lucid discussion of psychoacoustics from a musician's perspective and an excellent set of demonstrations on CD.]

[B: 23] W. Carlos, "Tuning: at the crossroads," Computer Music Journal, Spring, 29-43 (1987). ["Clearly the timbre of an instrument strongly affects what tuning and scale sound best on that instrument."]

[B: 24] P. Cariani, "Temporal coding of periodicity pitch in the auditory system: an overview," Neural Plasticity 6, No. 4, 147-172 (1999). [Discusses various ways that neurons may encode information and applies this to the perception of pitch.]

[B: 25] P. Cariani, "A temporal model for pitch multiplicity and tonal consonance," ICMPC (2004). [Discusses a time-based (as opposed to frequency-based) model for virtual pitch and sensory consonance.]

[B: 26] E. C. Carterette, R. A. Kendall, and S. C. Devale, "Comparative acoustical and psychoacooustical analysis of gamelan instrument tones," J. Acoust. Soc. Japan (E) 14, No. 6, 383-396 (1993).

[B: 27] E. C. Carterette and R. A. Kendall, "On the tuning and stretched octave of Javanese gamelans," Leonardo Music Journal, Vol. 4, 59-68 (1994). [A readily available article based on careful measurements from [B: 190].]

[B: 28] N. Cazden, "Musical consonance and dissonance: a cultural criterion," J. Aesthetics 4, 3-11 (1945). [Argues that notions of consonance and dissonance are not determined by acoustical or physiological laws, but are culturally determined.]

[B: 29] N. Cazden, "Sensory theories of musical consonance," J. Aesthetics 20, 301-319 (1962). [Argues against the use of excessive reductionism in music theory, using theories of sensory consonance as a primary example.]

[B: 30] N. Cazden, "The definition of consonance and dissonance," International Review of the Aesthetics and Sociology of Music 11, No. 2, 123-168 (1980). [Begins with a CDC-4 definition, and proceeds to discredit CDC-5 notions of consonance and dissonance.]

[B: 31] J. Chalmers, Jr., *Divisions of the Tetrachord*, Frog Peak, Hanover NH (1993). [Everything about tetrachords and tetrachordal scales.]

[B: 32] J. M. Chowning, "The synthesis of complex audio spectra by means of frequency modulation," J. Audio Engineering Society 21, 526-534 (1973). [Creator of FM synthesis explains how it works.]

[B: 33] E. A. Cohen, "Some effects of inharmonic partials on interval perception," Music Perception 1, No. 3, 323-349, Spring (1984). [Musically trained subjects are asked to tune sounds with stretched spectra to octaves and fifths. Some subjects tune to real octaves and fifths, whereas others match partials.]

[B: 34] A. M. Cornicello, "Timbral Organization in Tristan Murail's 'Désintégrations' and 'Rituals,' " Brandeis University, May (2000). [Detailed study of pieces of one of the major spectral composers.]

[B: 35] R. Crocker, *Discant, Counterpoint, and Harmony* (1962).

[B: 36] I. Darreg, "Xenharmonic Bulletin," Vol. 1-15, 1975, available from [B: 57]. [Coins the word "xenharmonic" from the greek *xenos*, meaning strange or foreign. Proposes that each tuning has a characteristic "mood."]

[B: 37] T. Day and B. Kollmeier, "Modeling auditory processing of amplitude modulation II: spectral and temporal integration," J. Acoust. Soc. Am., 102(5), Nov. (1997).

[B: 38] S. De Furia, *Secrets of Analog and Digital Synthesis*, Third Earth Productions Inc., NJ (1986). [How electronic musical instruments work.]

[B: 39] B. Denckla, *Dynamic Intonation for Synthesizer Performance*, Masters Thesis, MIT, Sept. (1997). [Presents a way to automatically adjust the pitch of a synthesizer based on musical context.]

[B: 40] G. DePoli, "A tutorial on digital sound synthesis techniques," Computer Music Journal 7, 8-26 (1983). [Overview of many synthesis techniques.]

[B: 41] D. Deutsch, ed., *The Psychology of Music*, Academic Press Inc., San Diego, CA (1982). [Anthology containing forward-looking chapters written by many of the researchers who created the field.]

[B: 42] L. DeWitt and R. G. Crowder, "Tonal fusion of consonant musical intervals: the oomph in Stumpf," Perception and Psychophysics 41, No. 1, 73-84 (1987). [A modern approach that investigates confusions among a number of simultaneously presented tones by analyzing subject's response time. Data support associating consonance with detailed properties of the harmonic series.]

[B: 43] D. B. Doty, *Just Intonation Primer*, Just Intonation Network, San Fransisco, CA (1993). [Clear and readable introduction to just intonation.]

[B: 44] J. Douthett, R. Entringer, and A. Mullhaupt, "Musical scale construction: the continued fraction compromise," Utilitas Mathematica 42, 97-113 (1992). [Uses continued fractions to find "best approximations" for arbitrary equal temperaments.]

[B: 45] H. Dudley , "Remaking speech," J. Acoust. Soc. Am. 11, 169 (1939). [Earliest vocoding technology.]

[B: 46] D. Ehresman and D. Wessel, "Perception of timbral analogies," Rapports IRCAM 13/78 (1978). [Carries out a multidimensional scaling of musical timbres.]

[B: 47] *Electronic Musician*, Cardinal Business Media Inc., Emeryville, CA. [Popular magazine with concrete discussions of music technology and product information.]

[B: 48] B. Enders, "Stimmungswechsel—Variable Stimmungen per Computersteuerung," Das Musikinstrument 8, (1989). [An early description of the hermode tuning.]

[B: 49] R. Erickson, "Timbre and Tuning of the Balinese Gamelan," Soundings 14-15, Santa Fe, NM (1986). [Discussion of the stretched octaves and scales of the gamelan.]

[B: 50] R. Erickson, *Sound Structure in Music,* Univ. of California Press, Berkeley, CA (1975). [A taxonomy of sounds and musical examples.]

[B: 51] L. J. Eriksson, M. C. Allie, and R. A. Griener, "The selection and application of an IIR adaptive filter for use in active sound attenuation," IEEE Trans. on Acoustics, Speech, and Signal Processing ASSP-35, No. 4, Apr. (1987). [A signal 180^o out of phase with the disturbance signal is used to cancel out unwanted noises.]

[B: 52] P. Erlich, "Tuning, tonality, and twenty-two-tone temperament," Xenharmonikon 17, Spring (1998). [Evaluates the ability of a variety of equal temperaments to approximate a number of n-limit consonances.]

[B: 53] P. Erlich, "The forms of tonality," A .pdf version of this paper is available on the CD `TTSS/PDF/erlich-forms.pdf`. [Concepts of tone-lattices, scales, and notational systems for 5-limit and 7-limit music.]

[B: 54] G. Eskelin, *Lies My Music Teacher Told Me*, Stage 3 Pubs., Woodland Hills, CA (1994). [A choir director discovers untempered singing in this funny and vituperative book that bemoans the current state of musical instruction.]

[B: 55] J. L. Flanagan and R. M. Golden, "Phase vocoder," Bell System Technical Journal, 1493-1509 (1966). [A more modern vocoder implementation.]

[B: 56] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, Springer-Verlag, NY (1991). [The *whole* story of musical physics. Not for the mathematically illiterate.]

[B: 57] Frog Peak Music, Box 5036, Hanover, NH. See also [W: 13]. [Distributers of electronic and world music.]

[B: 58] G. Galilei, *Discorsi e dimonstrazioni matematiche interno a due nuove scienze attenenti alla mecanica ed i movimenti locali*, Elsevier, Lieden (1638). Translation from *Dialogues Concerning Two New Sciences* by H. Crew and A. de Salvio, McGraw-Hill, NY (1963).

[B: 59] J. M. Geary, "Consonance and dissonance of pairs of inharmonic sounds," J. Acoust. Soc. Am. 67, No. 5, 1785-1789 May (1980). [Inharmonic sounds generated from sideband modulation show consonance and dissonance based on positioning of their partials.]

[B: 60] R. M. Gray and J. W. Goodman, *Fourier Transforms*, Kluwer Academic Press, New York (1995). [Comprehensive modern approach to uses of Fourier transforms].

[B: 61] D. M. Green, "Discrimination of transient signals having identical energy spectra," J. Acoust. Soc. Am. 49, 894-905 (1970). [When does the ear act like a Fourier analyzer?]

[B: 62] D. M. Green, "Temporal acuity as a function of frequency," J. Acoust. Soc. Am. 54, No. 2, 373-379 (1973). [Below about 500 Hz, it is possible to hear the difference between two waveforms with identical magnitude spectra.]

[B: 63] J. M. Grey and J. W. Gordon, "Perceptual effects of spectral modifications on musical timbres," J. Acoust. Soc. Am. 65, No. 5, 1493-1500 (1978). [What do you get when you swap the spectral envelopes of a trumpet and a trombone? Uses multidimensional scaling to look for physical correlates of timbral perceptions.]

[B: 64] J. M. Grey and J. A. Moorer, "Perceptual evaluation of synthesized musical instrument tones," J. Acoust. Soc. Am. 62, 454-462 (1977). [Resynthesizes instrumental tones and looks at ways to simplify the analysis, for instance, straight line approximations to the envelopes of the partials.]

[B: 65] S. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, New York, NY (1989). [A technique for solving complex optimization problems via a kind of structured random search. Technique inspired by theories of evolution.]

[B: 66] *Grolier Multimedia Encyclopedia*, Grolier Electronic Publishing, Inc. (1995).

[B: 67] J. Haeberli, "Twelve Nasca panpipes: a study," Ethnomusicology, Jan. (1979). [Tonometric measurements of panpipes from Nasca suggest that the Nascans may have used an arithmetic (linear in frequency) scale.]

[B: 68] D. E. Hall, "Quantitative evaluation of musical scale tunings," Am. J. Physics, 543-552, July (1974). [Uses a least-mean-square-error criterion to judge the fitness of various tunings for particular pieces of music.]

[B: 69] D. E. Hall, "The difference between difference tones and rapid beats," Am. J. Phys. 49, No. 7, July (1981). [Presents a series of experiments that distinguish the phenomonon of beats (fluctuations in loudness) from the phenomenon of difference tones (caused by nonlinearities).]

[B: 70] L. Harrison, *Lou Harrison's Music Primer,* Peters, NY (1971). [Quotable snippets on compositional techniques.]

[B: 71] H. Helmholtz, *On the Sensations of Tones*, (1877). Trans. A. J. Ellis, Dover Pubs., New York (1954). [Still readable after all these years.]

[B: 72] P. Hindemith, *The Craft of Musical Composition*, (1937). Trans. A Mendel, Associated Music Publishers, New York, NY (1945). [Extensive survey of compositional techniques by one of the centuries great composers.]

[B: 73] C. Huygens, *The Celestial Worlds Discover'd*, reprint London: F. Cass (1968).

[B: 74] W. Hutchinson and L. Knopoff, "The acoustic component of western consonance," Interface 7, 1-29 (1978). [Builds a consonance/dissonance measure on the work of Plomp and Levelt and applies it to harmonic sounds.]

[B: 75] A. Jacobs, *Penguin Dictionary of Music*, Penguin Books LTD., Middlesex, England (1991).

[B: 76] C. R. Johnson, Jr. and W. A. Sethares, *Telecommunication Breakdown*, Prentice-Hall, Englewood Cliffs, NJ (2003). [Chapter 7 contains a useful introduction to frequency analysis using the FFT (along with extensive Matlab code), and Appendix C discusses envelope detectors.]

[B: 77] A. M. Jones, "Towards an assessment of the Javanese pelog scale," J. Society Ethnomusicology 7,No. 1 (1963). [Looks for patterns in the pelog tuning data of [B: 90].]

[B: 78] O. H. Jorgensen, *Tuning*, Michigan State University Press, East Lansing, MI (1991). [Big red Bible of tunings, with excellent instructions for aural and electronic tunings of the various historical temperaments.]

[B: 79] A. Kameoka and M. Kuriyagawa, "Consonance theory part I: consonance of dyads" J. Acoust. Soc. Am. 45, No. 6, 1452-1459 (1969). [Reproduction and extension of Plomp and Levelt's tonal consonance ideas.]

[B: 80] A. Kameoka and M. Kuriyagawa, "Consonance theory part II: consonance of complex tones and its calculation method" J. Acoust. Soc. Am. 45, No. 6, 1460-1469 (1969). [Reproduction and extension of Plomp and Levelt's tonal consonance ideas.]

[B: 81] D. F. Keisler, "The relevance of beating partials for musical intonation," Proceedings of the ICMC, Montreal (1993). [Beating does not contribute to the percept of "out-of-tuneness" according to these experiments on the intonation of thirds and fifths.]

[B: 82] E. J. Kessler, C. Hansen, and R. N. Shepard, "Tonal schemata in the perception of music in Bali and in the West," Music Perception, Winter 1984, Vol. 2, No. 2, 131-165 (1984). [Comparison of Balinese and Western musical perception.]

[B: 83] *Keyboard*, Miller Freeman Inc., San Fransisco, CA. [Popular magazine with concrete discussions of music technology and product information.]

[B: 84] J. Kiefer and J. Wolfowitz, "Stochastic Estimation of a Regression Function," Ann. Math. Stat. 23, 462466 (1952). [A standard way to avoid calculating derivatives in gradient optimization is to approximate using evaluations of the cost.]

[B: 85] L. E. Kinsler, J. R. Fry, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, John Wiley and Sons, New York (1982). [Excellent coverage of modes of vibration of strings, bars, drumheads, etc. Requires calculus for full appreciation.]

[B: 86] R. Kirkpatrick, *Domenico Scarlatti*, Princeton University Press (1953). [The standard catalogue of Scarlatti's sonatas.]

[B: 87] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," Science 220, No. 4598, May (1983). [A technique for solving complex optimization problems via a kind of structured random search. Technique inspired by cooling of metals.]

[B: 88] R. J. Krantz and J. Douthett, "A measure of reasonableness of equal-tempered musical scales," J. Acoust. Soc. Am. 95, No. 6, 3642-3650, June (1994). [A "desirability" function is used to investigate the ability of various equal temperaments to approximate just intervals.]

[B: 89] F. Krueger, "Differenztone und konsonanz," Arch. Ges. Psych. 2, 1-80 (1904). [Proposes difference tones as an explanation for perceptions of consonance and dissonance.]

[B: 90] J. Kunst, *Music in Java*, Martinus Nijhoff, The Hague, Netherlands (1949). [The first major study of Javanese music.]

[B: 91] G. Kvistad, *Woodstock Chimes*, Woodstock Percussion, Inc., Hudson Valley, NY.

[B: 92] V. Lafrenièr, *Notion d'acoustique musicale (Ètude de la courbe de dissonance)*, Ècole Polytechnique de Montreal (2001). [Explores an alternative parameterization of the Plomp–Levelt curves.]

[B: 93] S. Lang, *Algebra*, Addison-Wesley Pub. Co. (1965). [Big red book of abstract algebra.]

[B: 94] A. Lehr, "The designing of swinging bells and carillon bells in the past and present," Athanasius Foundation, Asten, The Netherlands. (1987). [How bells are tuned.]

[B: 95] M. Leman, *Music and Schema Theory: Cognitive Foundations of Systematic Musicology*, Springer-Verlag, Berlin (1995). [Systematic musicology melds perceptual psychology with musicology.]

[B: 96] M. Leman, "Visualization and calculation of the roughness of acoustical musical signals using the synchronization index model," *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, Verona, Italy, Dec. (2000).

[Explores a generalization of the sensory dissonance model of Sect. 3.6 and Appendix G.]

[B: 97] T. Levin and M. Edgerton, "The throat singers of Tuva," Scientific American, Sept. (1999). [An acoutsical analysis of throat singing.]

[B: 98] F. Loosen, "The effect of musical experience on the conception of accurate tuning," Music Perception 12, No. 3, 291-306, Spring (1995). [Musicians tend to judge the temperaments with which they are most familiar as the most "in tune."]

[B: 99] V. Majernick and J. Kaluzny, "On the auditory uncertainly relations," Acustica 43, 132-146 (1979). [The auditory uncertainty product, defined as the JND for frequency multiplied by the signal duration, is examined theoretically and experimentally.]

[B: 100] M. V. Mathews and J. R. Pierce, "Harmony and inharmonic partials," J. Acoust. Soc. Am. 68, 1252-1257 (1980). [A series of experiments using sounds with stretched spectra that attempts to distinguish among functional consonance, sensory consonance, and cultural conditioning.]

[B: 101] M. V. Mathews, L. A. Roberts, and J. R. Pierce, "Four new scales built on integer ratio chords," J. Acoust. Soc. Am. 75, S10(A) (1984).[Scales based on the $13th$ root of 3 rather than the $12th$ root of 2.]

[B: 102] M. V. Mathews, J. R. Pierce, A. Reeves, and L. A. Roberts, "Theoretical and experimental explorations of the Bohlen-Pierce scale," J. Acoust. Soc. Am. 84, 1214-1222 (1988). [A musical scale based on 13 equal divisions of the octave+fifth rather than 12 divisions of the octave.]

[B: 103] M. V. Mathews and J. R. Pierce, "The Bohlen-Pierce scale," in *Current Directions in Computer Music Research*, ed. M. V. Mathews and J. R. Pierce, MIT Press, Cambrige, MA (1991). [Much the same as previous article, but the accompanying CD has three short demonstration pieces in the Bohlen–Pierce scale.]

[B: 104] W. A. Mathieu, *The Musical Life*, Shambhala Publications, Inc., Boston MA (1994). [Personal narrative of a musical life. Charming and poetic.]

[B: 105] E. May, ed., *Music of Many Cultures*, University of California Press, Inc., Berkeley, CA (1980). [Survey of the varieties of music on planet Earth.]

[B: 106] McClure, "Studies in Keyboard Temperaments," GSJ, i, 2840 (1948).

[B: 107] B. McLaren, "The uses and characterisics of non-just non-equal-tempered Scales," Xenharmonikon 15 (1993). [Manifesto on the use of non-just non-equal-tempered scales.]

[B: 108] B. McLaren, "General methods for generating musical scales," Xenharmonikon 13, Spring (1991). [Numerical methods of generating a wide variety of scales.]

[B: 109] B. McLaren and I. Darreg, "Biases in xenharmonic scales," Xenharmonikon 13, Spring (1991). [Equal temperaments viewed on a continuum from "biased towards melody" to "biased towards harmony."]

[B: 110] C. McPhee, *Music in Bali*, Yale University Press (1966). [The standard reference for Balinese music.]

[B: 111] R. Meddis, "Simulation of mechanical to neural transduction in the auditory receptor," J. Acoust. Soc. Am. 79, No. 3, 702-711, March (1986). [A computational model of the workings of the auditory system.]

[B: 112] A. P. Merriam, *Anthropology of Music*, Northwestern University Press (1964). [Classic work on ethnomusicology, by one of the founders of the field.]

[B: 113] G. R. Morrison, "88 cent equal temperament," Xenharmonikon 15 (1993). [Explores possibilities of the 88-cet non-octave scale.]
[B: 114] G. R. Morrison, "EPS xenharmonic scales disk musicians guide" (1993). [Disk provides alternate tunings for Ensoniq EPS/ASR samplers, and manual provides overview of alternate tuning practice.]
[B: 115] B. C. J. Moore, R. W. Peters, and B. R. Glasberg, "Thresholds for the detection of inharmonicity in complex tones," J. Acoust. Soc. Am. 77, 1861-1867 (1985).
[B: 116] B. C. J. Moore, B. R. Glasberg, and R. W. Peters, "Relative dominance of individual partials in determining the pitch of complex tones," J. Acoust. Soc. Am. 77, 1853-1860 (1985).
[B: 117] F. R. Moore, *Elements of Computer Music*, Prentice Hall, Englewood Cliffs, NJ (1990). [Analyzing, processing, and synthesizing musical structures and sounds using computers.]
[B: 118] E. Moreno, "Embedding equal pitch spaces and the question of expanded chromas: an experimental approach," CCRMA Report No. STAN-M-93, Dec. (1995). [Transpositions by the ratio K:1 in an Nth root of K system can preserve the chroma of the pitch.]
[B: 119] D. Morton, "The Music of Thailand," in [B: 105].
[B: 120] T. Murail, "Spectres et lutins," *La Revue Musicale*, 421-424. Also "Spectra and pixies," in *Contemporary Music Review*, ed. T. Machover, Harwood Academic Publishers 1, No. 1, 157-170 (1984). [Seminal paper describing spectral composition, which attempts to use "sound itself as a model for musical structure."]
[B: 121] B. Nettl, *The Study of Ethnomusicology*, University of Illinois Press (1983). [Discussion of the central issues and problems in ethnomusicology.]
[B: 122] M. Niedzwiecki and W. A. Sethares, "Smoothing of discontinuous signals: the competitive approach," IEEE Trans. on Signal Processing 43, No. 1, 1-12, Jan. (1995). [An approach to the smoothing of discontinuous signals is justified by an extension of the Kalman filter to the nonlinear case.]
[B: 123] H. Olson, *Music, Physics, and Engineering*, Dover Pubs., New York (1967). [Technical details have not aged well.]
[B: 124] F. Orduna-Bustamante, "Nonuniform beams with harmonically related overtones for use in percussion instruments," J. Acoust. Soc. Am. 90, No. 6, 2935-2941 (1991). [How to manipulate the contour of a bar so as to increase the harmonicity of the partials.]
[B: 125] C. L. van Panthaleon van Eck, *J. S. Bach's Critique of Pure Music,* Buren, author, n.d. (1981).
[B: 126] R. Parncutt, *Harmony: A Psychoacoustic Approach*, Springer-Verlag, Berlin (1989). [Application of Terhadt's models of psychoacoustics to common practice music theory and composition with harmonic tones.]
[B: 127] R. Parncutt, "Applying psychoacoustics in composition: 'harmonic' progressions of 'inharmonic' sonorities," Perspectives of New Music, 1994. [Speculative application of Terhadt's models of psychoacoustics to music theory and composition for sounds with inharmonic spectra.]
[B: 128] H. Partch, *Genesis of a Music*, Da Capo Press, New York (1974). [Musical prophet cries in the wilderness. Read it.]
[B: 129] H. Partch, *Bitter Music*, Ed. Thomas McGeary, University of Illinois Press, Urbana (1991). [Partch's writings during the depression are a testament to human perseverance. They are musically fascinating and entertaining.]

[B: 130] R. D. Patterson and B. C. J. Moore, "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, ed. B. C. J. Moore, Academic Press, London (1986). [A computational model of the auditory system.]

[B: 131] M. Perlman, "American gamelan in the garden of eden: intonation in a cross cultural encounter," Musical Quarterly (1996). [Shows how just intonation cannot be applied to gamelan tunings. Intonational naturalism vs. intonational relativism.]

[B: 132] R. Perrin, T. Charnley, and J. de Pont, "Normal modes of the modern English church bell," J. Sound Vibration 102, 11-19 (1983). [Investigation of the modal structure of church bells.]

[B: 133] J. O. Pickles, *An Introduction to the Physiology of Hearing*, Academic Press, New York (1988). [Detailed yet readable introduction to the hardware of hearing.]

[B: 134] J. R. Pierce, "Attaining consonance in arbitrary scales," J. Acoust. Soc. Am. 40, 249 (1966). [Construction of a spectrum related to 8-tet.]

[B: 135] J. R. Pierce, *The Science of Musical Sound*, Scientific American Books, W. H. Freeman and Co., New York (1983). [One of the most clearly written, thoughtful, and beautifully illustrated books in musical science. The accompanying recording is well worth a listen.]

[B: 136] J. R. Pierce, "Periodicity and pitch perception," J. Acoust. Soc. Am. 90,No. 4, 1889-1893 (1991). [Tries to untangle the effects of the two pitch detection mechanisms, one operating on resolved partials and the other on periodicity.]

[B: 137] W. Piston, *Harmony*, 5th Edition, W. W. Norton & Co. Inc. (1987). [Standard reference on the common practice of composers in the 18th and 19th century.]

[B: 138] R. Plomp, "Detectability threshold for combination tones," J. Acoust. Soc. Am. 37, 1110-1123 (1965). [How loud are difference tones? Can they be heard in music?]

[B: 139] R. Plomp, "Timbre as a multidimensional attribute of complex tones," in *Frequency Analysis and Periodicity Detection in Hearing,* ed. R. Plomp and G. F. Smoorenburg, A. W. Sijthoff, Lieden (1970).

[B: 140] R. Plomp, *Aspects of Tone Sensation*, Academic Press, London (1976).

[B: 141] R. Plomp and W. J. M. Levelt, "Tonal consonance and critical bandwidth," J. Acoust. Soc. Am. 38, 548-560 (1965). [Relates tonal consonance/dissonance to the critical bandwidth. Extensive comments on Helmholtz and others consonance theories. If you want to see where many of the ideas in *Tuning, Timbre, Spectrum, Scale* come from, read this paper.]

[B: 142] L. Polansky, "Paratactical tuning: an agenda for the use of computers in experimental intonation," Computer Music Journal 11, No. 1, 61-68, Spring (1987). [Suggests need for a dynamic (adaptive) tuning.]

[B: 143] R. L. Pratt and P. E. Doak, "A subjective rating scale for timbre," Journal of Sound and Vibration 45, 317 (1976).

[B: 144] K. van Prooijen, "A Theory of Equal Tempered Scales," Interface 7, 45-56 (1978). [Uses continued fractions to measure deviations from harmonicity for arbitrary equal temperaments.]

[B: 145] J. P. Rameau, *Treatise on Harmony*, Dover Pubs., New York (1971), original edition (1722). [The first statement of "functional" consonance and dissonance.]

[B: 146] R. A. Rasch, "Perception of melodic and harmonic intonation of two-part musical fragments," Music Perception 2, No. 4, Summer (1985). [Compares effects of melodic mistuning and harmonic mistunings.]
[B: 147] J. W. S. Rayleigh, *The Theory of Sound*, Dover Pubs., New York (1945), original edition (1894). [Complete survey of the field of acoustics as of 1894 by one of the men who created it.]
[B: 148] W. H. Reynolds and G. Warfield, *Common-Practice Harmony*, Longman, New York (1985). [Standard music theory text.]
[B: 149] J. C. Risset, "The development of digital techniques: a turning point for electronic music?" Rapports IRCAM 9/78 (1978). [New directions in electronic music. Worth reading even though it's no longer new.]
[B: 150] J. C. Risset, "Additive synthesis," in *The Psychology of Music*, ed. Diana Deutsch, Academic Press, New York (1982). [How additive synthesis works.]
[B: 151] J. C. Risset and D. L. Wessel, "Exploration of timbre by analysis and synthesis," in *The Psychology of Music,* ed. Diana Deutsch, Academic Press, New York (1982). [How and why additive synthesis/resynthesis techniques have changed the way we look at musical sound.]
[B: 152] C. Roads, "A tutorial on nonlinear distortion or waveshaping synthesis," Computer Music Journal 3, No. 2, 29-34 (1979). [A technique for digital sound synthesis.]
[B: 153] C. Roads, "Interview with Max Mathews," Computer Music Journal 4, No. 4, Winter (1980). [Mathews says, "It's clear that inharmonic timbres are one of the richest sources of new sounds. At the same time they are a veritable jungle of possibilities so that some order has to be brought out of this rich chaos before it is to be musically useful."]
[B: 154] J. G. Roederer, *The Physics and Psychophysics of Music*, Springer-Verlag, New York (1994). [Emphasizes importance of perception. The title should not dissuade those without mathematical expertise.]
[B: 155] P. Rosberger, *The Theory of Total Consonance*, Associated University Presses Inc., Cranbury, NJ (1970). [Proposal for an adaptive tuning "ratio machine" that maintains simplest possible integer ratio intervals at all times.]
[B: 156] J. E. Rose, J. E. Hind, D. J. Anderson, and J. F. Brugge, "Some effects of stimulus intensity on response of auditory nerve fibers in the squirrel monkey," Journal of Neurophysiology 34, 685-699 (1971). [Among other things, includes solid physical evidence for the presence of nonlinearities in models that incorporate the hair cells of the basilar membrane.]
[B: 157] T. D. Rossing, *Acoustics of Bells*, Van Nostrand-reinhold, Stroudsberg, PA (1984). [Everything you always wanted to know about bells, and a bit more.]
[B: 158] T. D. Rossing, *The Science of Sound*, Addison Wesley Pub., Reading, MA (1990). [One of the best all around introductions to the Science of Sound. Comprehensive, readable, and filled with clear examples.]
[B: 159] T. D. Rossing and R. B. Shepherd, "Acoustics of gamelan instruments," Percussive Notes 19, No. 3, 73-83 (1982). [The only published investigation of the spectra of gamelan instruments.]
[B: 160] J. Sankey and W. A. Sethares, "A consonance-based approach to the harpsichord tuning of Domenico Scarlatti," J. Acoust. Soc. Am. 101,No. 4, 2332-2337, April (1997). [John's expertise in the Scarlatti-sphere was a boon, and the collaboration was great fun.]
[B: 161] P. Schaeffer, *Traite des Objets Musicaux*, Paris: Editions de Seuil (1968). [Attempts a complete classification of sound.]

[B: 162] R. M. Schafer, *The Tuning of the World*, Univ. of Pennsylvania Press, Philadelphia (1980). [Explores natural and artificial "soundscapes," both historical and modern.]
[B: 163] H. Schenker, *Five Graphic Music Analyses*, Dover Pubs., New York (1969). [A standard technique for the analysis of music.]
[B: 164] A. Schoenberg, *Structural Functions of Harmony*, Norton, New York (1954). [Standard technique for the analysis of music.]
[B: 165] W. A. Sethares, "Local consonance and the relationship between timbre and scale," J. Acoust. Soc. Am. 94, No. 3, 1218-1228, Sept. (1993). [Shows how to draw dissonance curves for arbitrary spectra, and how to construct spectra for arbitary scales.]
[B: 166] W. A. Sethares, "Relating tuning and timbre," Experimental Musical Instruments IX, No. 2 (1993). [A more popular version of the JASA article with similar name. No math.]
[B: 167] W. A. Sethares, "Adaptive tunings for musical scales," J. Acoust. Soc. Am. 96, No. 1, 10-19, July (1994). [How to use consonance/dissonance ideas to design adaptive or dynamic tunings that respond to the spectrum of the sound being played.]
[B: 168] W. A. Sethares, "Tunings for inharmonic sounds," Synaesthetica ACAT, Canberra, Australia, Aug. (1994). [Used the Chaco rock to demonstrate scale and spectrum issues.]
[B: 169] W. A. Sethares, "Specifying spectra for musical scales," J. Acoust. Soc. Am. 102, No. 4, Oct. (1997). [Presents a method of specifying the spectrum of a sound so as to maximize a measure of consonance with a given desired scale.]
[B: 170] W. A. Sethares, "Consonance-based spectral mappings," Computer Music Journal 22, No. 1, 56-72, Spring (1998).
[B: 171] W. A. Sethares, "Real-time adaptive tunings using MAX," Journal of New Music Research 31, No. 4, Dec. (2002). [Implementation of adaptive tunings in the MAX language.]
[B: 172] W. A. Sethares, D. A. Lawrence, C. R. Johnson, Jr., and R. R. Bitmead, "Parameter drift in LMS adaptive filters," IEEE Trans. on Acoustics, Speech, and Signal Processing ASSP-34, No. 4, August (1986). [Some problems that arise with gradient algorithms.]
[B: 173] W. A. Sethares and B. McLaren , "Memory and context in the calculation of sensory dissonance," Proc. of the Research Society for the Foundations of Music, Ghent, Belgium, Oct. (1998). [Investigates use of a "context" or "memory" function in the calculation of sensory dissonance.]
[B: 174] W. A. Sethares and T. Staley, "Sounds of crystals," Experimental Musical Instruments VIII, No. 2, Sept. (1992). [Uses data from x-ray crystallography to generate interesting sounds.]
[B: 175] W. Slawson, *Sound Color*, University of California Press, Los Angeles, CA. (1985). [What are the "laws of tone color?" What is timbre, and how can it be held constant when other aspects of a sound change?]
[B: 176] F. H. Slaymaker, "Chords from tones having stretched partials," J. Acoust. Soc. Am. 47, No. 2, 1469-1571 (1970). [Asks how musical stretched tones can be.]
[B: 177] N. Sorrell, *A Guide to the Gamelan*, Faber and Faber Ltd., London (1990). [A solid introduction to the Gamelan "orchestra:" its instruments, music, and traditions.]

[B: 178] G. A. Sorge, *Vorgemach der musicalischen composition* Verlag des Autoris, Lobenstein (1747). [First presentation of idea that beating of partials causes dissonance.]

[B: 179] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," IEEE Trans. Autom. Control 37, 332341 (1992). [Presents a way to avoid complex gradient calculations in optimization problems.]

[B: 180] J. C. Spall, "A one-measurement form of simultaneous perturbation stochastic approximation," Automatica 33, 109112 (1997). [Further development of [B: 179].]

[B: 181] S. S. Stevens, "Tonal density," J. Experimental Psychology 17 (1934). [Experiment requires observers to change a dial until two sounds have the same "tonal density."]

[B: 182] K. Steiglitz, *A Digital Signal Processing Primer*, Addison-Wesley Pub., Menlo Park, CA (1996). [Digital signal processing, with applications to computer music and digital audio. A more complete rendering of the "Speaking of Spectra" appendix.]

[B: 183] W. Stoney, "Theoretical possibilities for equal tempered systems" in *The Computer and Music* ed. H. B. Lincoln, Cornell University Press, Ithaca, NY (1970). [How well do the various equal temperaments approximate the harmonic series?]

[B: 184] A. Storr, *Music and the Mind*, Ballantine Books, New York (1992). [Why do people make music?]

[B: 185] G Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley College Press, Wellesley, MA (1996). [Solid introduction to theory and applications of wavelets and filter bank techniques.]

[B: 186] W. J. Strong and M. Clark, "Perturbations of synthetic orchestral wind instrument tones, J. Acoust. Soc. Am. 41, 277-285 (1967). [Take the spectrum of one instrument and graft it onto the envelope of another. What do you get?]

[B: 187] W. J. Strong and G. R. Plitnik, *Music, Speech, Audio*, Soundprint, Provo, UT (1992). [One of the best all around introductions to the Science of Sound. Comprehensive, readable, and filled with clear examples.]

[B: 188] K. Stumpf, "Konsonanz und dissonanz," Beitr. Akust. Musikwiss 1, 1-108 (1898). [Argues that consonance and dissonance are due directly to the degree of fusion of the sounds.]

[B: 189] J. Sundberg, *The Science of Musical Sounds,* Academic Press, New York (1991). [Excellent overview of the acoustics and psychoacoustics of musical sounds.]

[B: 190] W. Surjodiningrat, P. J. Sudarjana, and A. Susanto, *Tone Measurements Of Outstanding Javanese Gamelan In Yogyakarta And Surakarta,* Gadjah Mada University Press (1993). [Carefully documents the tuning of 70 different gamelans.]

[B: 191] C. Taylor, *Sounds of Music*, Charles Scribners Sons, New York (1976). [Read this book for its excellent series of simple experiments that demonstrate fundamental acoustical principles, not for its theoretical contributions.]

[B: 192] J. Tenney, *A History of 'Consonance' and 'Dissonance,'* Excelsior Music Pub., New York (1988). [Traces historical uses of the words "consonance" and "dissonance" from Medieval times to the present.]

[B: 193] M. Tenzer, *Balinese Music*, Periplus Editions, Inc. (1991). [Colorful and informative introduction to Balinese gamelan instruments, history, and repertoire.]

[B: 194] E. Terhardt, "Pitch shifts of harmonics, an explanation of the octave enlargement phenomenon," Proc. 7th Int. Congress Acoustics, Budapest, Hungary (1971). [Shows how the preference for stretching of octaves may be caused by the same mechanism that is responsible for virtual pitch.]

[B: 195] E. Terhardt, "Pitch, consonance, and harmony," J. Acoust. Soc. Am. 55, No. 5, 1061-1069, May (1974). [Notions of virtual pitch are combined with a "learning matrix" (an early kind of neural network) that is used to explain several auditory phenomena.]

[B: 196] E. Terhardt, "The concept of musical consonance: a link between music and psychoacoustics," Music Perception 1, No. 2, 276-295, Spring (1984). [Attempts to link virtual pitch notions with "harmony." When combined with sensory consonance, this provides a "complete" theory of musical consonance and dissonance.]

[B: 197] E. Terhardt, G. Stoll, and M. Seewann "Algorithm for extraction of pitch and pitch salience from complex tone signals," J. Acoust. Soc. Am. 71, No. 3, 679-688, March (1982). [Gives an algorithm for finding the virtual pitch of a complex tone.]

[B: 198] C. D. Veroli, *Unequal Temperaments*, Artes Graphicas Farro, Argentina (1978). [Emphasizes historical uses, musical characteristics, and the practice of unequal temperaments. Shows how to tune and fret instruments for nonequal playing.]

[B: 199] N. Vicentino "L'antica musica ridotta alla moderna prattica," Antonio Barre, Rome (1555), English translation: "Ancient music adapted to modern practice," eds. M. R. Maniates and C. V. Palisca, Yale University Press, New Haven, CT (1996). [A tuning consisting of two meantone chains that can function as an adaptive tuning.]

[B: 200] J. Vos, "The perception of pure and mistuned musical fifths and major thirds: thresholds for discrimination, beats, and identification," Perception and Psychophysics 32, 297-313 (1982).

[B: 201] J. Vos, "Subjective acceptability of various regular twelve-tone tuning systems in two-part musical fragments," J. Acoust. Soc. Am. 83, No. 6, 2383-2392 (1988).

[B: 202] H. Waage, "Liberation for the scale-bound consonance," 1/1 7, No. 3, March (1992). [This alternative to keyboards with a fixed tuning uses a "logic circuit" to determine how each note in a chord should be retuned.]

[B: 203] W. D. Ward, "Musical perception," in *Foundations of Modern Auditory Theory,* ed. J. V. Tobias, Academic Press, New York (1970). [Discusses (among other things) "subjective" pitch of musical intervals.]

[B: 204] E. G. Wever, "Beats and related phenomena resulting from the simultaneous sounding of two tones," Psychological Review 36, 402-418 (1929). [Gives an early history of beats.]

[B: 205] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," Proceedings of the IEEE 64, No. 8, 1151-1162, Aug. (1976). [A standard reference for gradient descent algorithms.]

[B: 206] C. G. Wier, W. Jesteadt, and D. M. Green, "Frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. 61, 178-184 (1977). [Finding the JND for frequency perception.]

[B: 207] S. R. Wilkinson, *Tuning In*, Hal Leonard Books (1988). [Popular book introducing alternatives to 12-tet. Focuses on making electronic synthesizers play in microtunings.]

[B: 208] R. Young, "Inharmonicity of plain piano wire," J. Acoust. Soc. Am. 24, 267-273 (1952). [The partials of piano wire are "stretched" by a factor of about 1.0013].

[B: 209] M. Yunik and G. W. Swift, "Tempered music scales for sound synthesis," Computer Music Journal, Winter (1980). [How well do the various equal temperaments approximate 50 different just intervals?]

[B: 210] D. Zicarelli, G. Taylor, et al., *Max 4.0 Reference Manual*, Cycling '74 (2001). [Manual for the Max programming language. See also [W: 22].]

[B: 211] E. Zwicker and H. Fastl, *Psychoacoustics*, Springer-Verlag, Berlin (1990). [Up to date account of many psychoacoustic measures and techniques, including an excellent overview of JND research.]

[B: 212] E. Zwicker, G. Flottorp, and S. S. Stevens, "Critical bandwidth in loudness summation," J. Acoust. Soc. Am. 29, 548 (1957). [Examines relationship between critical bandwith and JND.]

D: Discography

References in the body of the text to the discography are coded with [D:] *to distinguish them from references to the bibliography, sound and video examples, and websites.*

[D: 1] S. Alexjander, *Sequencia,* Science and the Arts, Berkeley, CA (1994). [Uses data from DNA sequences to generate interesting tunings.]

[D: 2] J. M. Barbour and F. A. Kuttner, *Theory and Practice of Just Intonation*, Musurgia Records, Jackson Heights, NY (1958). [This recording gives numerous examples of how bad Just Intonation can sound if played incorrectly. For instance, "Auld Lang Syne" is played in C in a Just C scale, and it is then played in $F\#$ without changing the tuning.]

[D: 3] J. M. Barbour and F. A. Kuttner, *Meantone Temperament in Theory and Practice*, Musurgia Records, Jackson Heights, NY (1958). [Meantone bad. Equal temperament good.]

[D: 4] E. Blackwood, *12 Microtonal Etudes for Electronic Music Media* (1976). [A sort of "ill-tempered synthesizer" with pieces in all equal temperaments from 13 to 24].

[D: 5] W. Carlos, *Beauty in the Beast*, SYNCD 200, Jem Records, Inc. South Plainfield, NJ (1986). ["Puts aside the traditional equally tempered scale, and also the standard acoustic and electronic timbres" to create one of the greatest xenharmonic pieces *so far*.]

[D: 6] W. Carlos, *Secrets of Synthesis*, CBS Records MK 42333 (1987). [Carlos introduces and explains synthesizer technology. In "Alternative Tunings–The Future," Carlos says, "... not only can we have any possible timbre but these can be played in any possible tuning... that might tickle our ears."]

[D: 7] W. Carlos, *Switched on Bach 2000*, Telarc Int. Co. CD-80323, Cleveland, OH (1992). [The classic album revisited. With modern synthesizer technology, Carlos performs in "authentic Bach tunings."]

[D: 8] J. Chowning, *Turenas, Stria, Phone, Sabelithe* WER 2012-50 Wergo, Mainz, Germany (1988). [Use of inharmonic materials in a "western" style.]

[D: 9] *Classical Instrumental Traditions: Thailand*, JVC World Sounds, VICG-5262, Tokyo, Japan (1993). [Focuses on solo pieces for a variety of indigenous Thai instruments.]

[D: 10] I. Darreg, *Detwelvulate*, Ivor Darreg Memorial Fund (1995). [Encourages use of non-12-tet tunings. Each tuning has its own "feel."]

[D: 11] D. Doty, *Uncommon Practice: Selected Compositions 1984-1995,* Frog Peak Music [B: 57]. [Compositions in just intonation.]

[D: 12] Fong Naam, *Sleeping Angel*, Nimbus Records, NI 5319 (1991). [Thai classical music is played in a close approximation to 7-tet.]
[D: 13] Fong Naam, *Nang Hong Suite*, Nimbus Records, NI 5332 (1992). [Thai funeral music, in 7-tet, is livelier than you might think.]
[D: 14] E. Fisk, *Baroque Guitar*, MusicMasters 0612-67130-2, Ocean, NJ (1993). [Scarlatti performed on classical guitar.]
[D: 15] *Gamelan Batel Wayang Ramayana*, CMP Records, NY CMP CD 3003 (1990). [Gamelan music accompanying the Ramayana saga.]
[D: 16] *Gamelan of Cirebon*, King Records, KICC 5130, Tokyo, Japan (1991). [An iron gamelan from Cirebon, played in the slendro tuning.]
[D: 17] *Gamelan Gong Gede of Batur Temple*, King Records, KICC 5153, Tokyo, Japan (1992). [A Balinese gamelan.]
[D: 18] *Gamelan Gong Kebyar of "Eka Cita," Abian Kapas Kaja*, King Records, KICC 5154, Tokyo, Japan (1992). [Award-winning gamelan from Denpassar, Bali.]
[D: 19] *Gender Wayang of Sukawati Village*, King Records, KICC 5156, Tokyo, Japan (1992). [The gamelan that accompanies the shadow puppet.]
[D: 20] The Gyuto Monks, *Freedom chants from the roof of the world,* Rykodisc (1989). [Overtone singing is common in the Tibetan tradition.]
[D: 21] A. J. M. Houtsma, T. D. Rossing, and W. M. Wagenaars, *Auditory Demonstrations* (Phillips compact disc No. 1126-061 and text) Acoustical Society of America, Woodbury NY (1987). [A wealth of great sound examples: thorough and thought provoking.]
[D: 22] Huun-Huur-Tu, "60 horses in my herd," Shanachie 64050 (1993). [Throat singing is integral to these traditional Tuvan songs.]
[D: 23] On the Edge, Selections of the 1996 International Computer Music Society, Hong Kong (1996).
[D: 24] E. Katahn, *Beethoven In The Temperaments*, Gasparo Records, No. 332 (1998). [Performances of several Beethoven piano sonatas in authentic temperaments.]
[D: 25] *Klênêngan Session of Solonese Gamelan*, King Records, KICC 5185, Tokyo, Japan (1994). [Gamelan from the palace (kraton) in Solo, played by musicians from the National Broadcasting Company (RRI).]
[D: 26] E. Lyon, *Red Velvet*, Smart Noise Records (1996) [Music that "hypernavigates a compressed informational world." Thanks, Eric.]
[D: 27] *Music from the Morning of the World,* Elekctra/Asylum/Nonesuch Records, 9 79196-2, Rockefeller Plaza, NY (1988). [Balinese gamelan and the Ramayana monkey chant.]
[D: 28] T. Murail, *Gondwana/Désintégrations/Time and Again*, performed by Y. Prin and P. Plissier, Salabert, Scd8902. [Spectral compositions.]
[D: 29] *Music for the Gods*, Ryko RCD 10315 (1992). [Recorded in 1941 and recently reissued. Compare the early sound of the gamelan with what it has become today.]
[D: 30] A. Newman, *Scarlatti Sonatas* NCD 60080, Newport Classic, RI (1989). [Scarlatti played on the "Magnum Opus" harpsichord, "maybe the largest harpsichord ever built."]
[D: 31] H. Partch, *The Bewitched*, Performed by members of the University of Illinois Musical Ensemble, CRI CD7001, 179 W. 74th St. NY (1990). [Partch's dance-satire is performed with a variety of his instruments tuned to his 43-tone just scale.]

[D: 32] H. Partch, *Music of Harry Partch*, CRI CD7000, New York (1989). [A "best of" Partch: new scales, new instruments, a new listening experience.]
[D: 33] I. Pogorelich, *Domenico Scarlatti Sonaten*, Deutsche Grammophon 435-855-2 (1992). [Scarlatti adapted for piano.]
[D: 34] L. Polansky, *Simple Harmonic Motion*, Artifact Recordings, Berkeley, CA (1994). [Works for instruments in just intonation.]
[D: 35] S. Reich, *Phase Patterns* Robi Droli/Newtone, No. 5018, (2000). [Exploits rhythmic phasing.]
[D: 36] J. C. Risset, *Sud, Dialogues, Inharmonique, Mutations*, INA C 1003, INA.GRM Paris, France (1987). [Use of inharmonic materials in a "western" context.]
[D: 37] S. Ross, *Scarlatti, Best Sonatas* Erato, 2292-45423-2, Erato-Disques, Radio France (1988). [Scarlatti recorded at the Chapelle du Chateau d'Assas.]
[D: 38] I. W. Sadra, *Karya*, Lyrichord LYRCD 7421. [New music from an influential Indonesian composer.]
[D: 39] *Thailand-Ceremonial and Court Music.*
[D: 40] W. A. Sethares, *Xentonality*, Odyssey Records XEN2001 (1997). [A variety of equal and unequal temperaments played with related timbres. Adaptively tuned and found-sound pieces. Thoroughly xentonal. Available from Frog Peak Music, Box 1052, Lebabnon NH 03766 and from amazon.com.]
[D: 41] W. A. Sethares, *Exomusicology*, Odyssey Records EXO2002 (2002). [A variety of equal and unequal temperaments played with related timbres. Adaptively tuned and found-sound pieces. Thoroughly xentonal. Available from amazon.com.]
[D: 42] L. Sgrizzi, *Vingt-quatre Sonates pour Clavecin,* Accord, 1491014, France (1984). [Scarlatti played on the harpsichord at the Cathedrale San Lorenzo.]
[D: 43] J. Teller, *My Inner Ear*, The Tyte Institute, Hesselogado 4,3 DC-2100, Copenhagen, Denmark. [Concert for three samplers in the spiral corridor of the Roundtower.]
[D: 44] F. Terenzi, *Music from the Galaxies,* Island Records, Inc., New York (1991). [Maps from interstellar radio telescope data into sound waves, creating interesting outer space sounds.]
[D: 45] *Instrumental Music of Northeast Thailand*, King Records, KICC 5124, Tokyo, Japan (1991). [*Pong lang* is a kind of wooden xylophone and a style of music.]

S: Sound Examples on the CD-ROM

The sound files on the CD-ROM are saved in the `.mp3` *format, which is readable using* `Windows Media Player` *or* `Quicktime`*. Navigate to* `TTSS/sounds/Chapter/` *and launch the* `*.mp3` *file by double clicking, or by opening the file from within the player. References in the body of the text to sound examples are coded with* [S:] *to distinguish them from references to the bibliography, discography, video examples, and web links. The sound examples may also be accessed using a web browser. Open the file* `TTSS/Contents.html` *in the top level of the CD-ROM and navigate using the html inteface.*

Sound Examples for Chapter 1

[S: 1] *Challenging the octave* (`challoct.mp3 0:24`). The spectrum of a sound is constructed so that the octave between f and $2f$ is dissonant while the nonoctave f to $2.1f$ is consonant. See p. 2 and video [V: 1].

[S: 2] *A simple tune* (`simptun1.mp3 0:47`). Harmonic timbres in the 12-tet scale set the stage for the next three examples. Chord pattern is taken from *Plastic City*, sound example [S: 38]. See pp. 3 and 322.

[S: 3] *The "same" tune* (`simptun2.mp3 0:47`). Harmonic timbres in the 2.1-stretched scale appear uniformly dissonant. See p. 3.

[S: 4] *The "same" tune* (`simptun3.mp3 0:47`). 2.1-stretched timbres are matched to the 2.1-stretched scale. See p. 3.

[S: 5] *The "same" tune* (`simptun4.mp3 0:47`). 2.1-stretched timbres in 12-tet appear uniformly dissonant. See p. 3.

Sound Examples for Chapter 2

[S: 6] *Virtual pitch ascending* (`virtpitchup.mp3 0:22`). Harmonic and inharmonic timbres alternate with sine waves at the appropriate virtual pitch. See Table 2.2 on p. 37 for a listing of all frequencies in this example.

[S: 7] *Virtual pitch descending* (`virtpitchdown.mp3 0:22`). Harmonic and inharmonic timbres alternate with sine waves at the appropriate virtual pitch. Comparing this example with [S: 6] shows how virtual pitch may be influenced by context. See Table 2.2 on p. 37 for a listing of all frequencies in this example.

Sound Examples for Chapter 3

[S: 8] *Beating of sine waves I* (`beats1.mp3 0:24`). See p. 41 and video [V: 5].
(i) A sine wave of 220 Hz (4 seconds)
(ii) A sine wave of 221 Hz (4 seconds)
(iii) Sine waves (i) and (ii) together (8 seconds)

[S: 9] *Beating of sine waves II* (`beats2.mp3 0:24`). See p. 41 and video [V: 6].
(iv) A sine wave of 220 Hz (4 seconds)
(v) A sine wave of 225 Hz (4 seconds)
(vi) Sine waves (iv) and (v) together (8 seconds)

[S: 10] *Beating of sine waves III* (`beats3.mp3 0:24`). See p. 41 and video [V: 7].
(vii) A sine wave of 220 Hz (4 seconds)
(viii) A sine wave of 270 Hz (4 seconds)
(ix) Sine waves (vii) and (viii) together (8 seconds)

[S: 11] *Dissonance between two sine waves* (`sinediss.mp3 1:06`). A sine wave of fixed frequency 220 Hz is played along with a "sine wave" with frequency that begins at 220 Hz and slowly increases to 470 Hz. See p. 45 and video [V: 8]. Figure 3.6 on p. 46 provides a visual representation.

[S: 12] *Dissonance between two sine waves: Binaural Presentation* (`sinedissbin.mp3 1:06`). The same as [S: 11], except the sine wave of fixed frequency is panned completely to the right and the variable sine wave is panned completely to the left. Using headphones will ensure that only one channel is audible to each ear. The dissonance percept is still present, although diminished. See p. 49.

Sound Examples for Chapter 4

[S: 13] *Dream to the Beat* (`dreambeat.mp3 5:28`). A 19-tet pop tune with a bass that beats like the heart. A microtonal love song. See p. 59.

[S: 14] *Incidence and Coincidence* (`incidence.mp3 5:23`). What happens when you play simultaneously in different tunings? Each note in this 19-tet melody is "harmonized" by a note from 12-tet, resulting in some unusual inharmonic sound textures. The distinction between "timbre" and "harmony" becomes confused, although the piece is by no means confusing. See p. 59.

[S: 15] *Haroun in 88* (`haroun88.mp3 3:36`). In all 12-tet instruments (like the piano), there are 100 cents between adjacent steps. *Haroun in 88* uses a tuning in which there are 88 cents between adjacent steps, a scale first explored by Gary Morrison [B: 113]. One feature of this scale is that it does not repeat at the octave; instead, it has 14 equal steps in a stretched "pseudo-octave" of 1232 cents. One way to exploit such "strange" tunings is to carefully match the tonal qualities of the sounds to the particular scale. See pp. 60, 277, and 283.

[S: 16] *88 Vibes* (`vibes88.mp3 3:47`). Also in the 88-cent-per-tone tuning, *88 Vibes* features a spectrally mapped "vibraphone." See pp. 60, 277, and 283.

[S: 17] *Sonata K380 by Scarlatti* (`k380tet12.mp3 1:29`). Performed in 12-tet in the key of C. See pp. 61 and 224.

[S: 18] *Sonata K380 by Scarlatti* (`K380JImajC.mp3 1:29`). Performed in just intonation centered in the key of C. See p. 61.

[S: 19] *Sonata K380 by Scarlatti* (`K380JIC+12.mp3 1:29`). Performed in just intonation centered in the key of C and 12-tet simultaneously. The notes where the differences are greatest stand out clearly. See p. 61.

[S: 20] *Sonata K380 by Scarlatti* (`K380JImajC+.mp3 1:29`). Performed in just intonation centered in the key of $C\sharp$. See p. 63.

[S: 21] *Sonata K380 by Scarlatti* (`K380JImeanC.mp3 1:29`). Performed in the quarter comma meantone tuning centered in the key of C. See p. 66.

[S: 22] *Sonata K380 by Scarlatti* (`K380JImeanC+.mp3 1:29`). Performed in the quarter comma meantone tuning centered in the key of $C\sharp$. See p. 66.

[S: 23] *Imaginary Horses* (`imaghorses.mp3 3:58`). This sequence contains the harmonic spectra of a piano and a "perc flute," which are matched to the simple integer ratios

1/1 6/5 4/3 3/2 8/5 9/5 2/1

to form a Just Intonation scale that was called "solemn procession" by Lou Harrison. The consequence is a piano and synth duet with galloping piano riff and bucking synth lines that does not sound solemn to me. See p. 61.

[S: 24] *Joyous Day* (`joyous.mp3 4:35`). This uses the just intonation

1/1 9/8 5/4 3/2 5/3 15/8 2/1

created by Lou Harrison. To my ears, it is a majestic, extra-major sounding tuning. See p. 61.

[S: 25] *What is a Dream?* (`whatdream.mp3 3:51`). Although the ancient Greeks did not record their music, they did write about it. They noticed the relationships between musical pitches and mathematical ratios. Some of the ancient scales fell into disuse, among them the "aeolic" scale, which uses the justly tempered pitches

1/1 9/8 32/27 4/3 3/2 128/81 16/9 2/1.

Lyrics expertly crafted by a non-ancient Greek, George Sethares. See p. 61.

[S: 26] *Just Playing* (`justplay.mp3 2:52`). In this piece, the 12 notes of the keyboard are mapped:

cents:	0	19	205	267	386	498	583	702	766	884	969	1088
mapped to:	C	$C\sharp$	D	$D\sharp$	E	F	$F\sharp$	G	$G\sharp$	A	$A\sharp$	B
interval:	1.0	1.011	1.125	1.167	1.25	1.33	1.4	1.5	1.56	1.67	1.75	1.87
ratio:	1/1	x/x	9/8	7/6	5/4	4/3	7/5	3/2	11/7	5/3	7/4	15/8

This includes all ratios of the JI major scale, along with a few extras. The small interval between C and $C\sharp$, for which there is no (small integer) just ratio, was used primarily for trills. See p. 61.

[S: 27] *Signs* (`signs.mp3 3:41`). One of the more prolific ancient Greeks (from the point of view of discovering and codifying musical scales) was Archytas, who lived about 400 B.C. Although his music has been lost, his tunings have survived. This song is played in one of Archytas' chromatic scales that is based on equal "tetrachords" (a set of four descending notes, see p. 55) with the intervals

28/27 243/224 32/27.

It is rather amazing that the sonorous beauty of scales such as this were surrendered by the European musical tradition for centuries in exchange for a keyboard that could be played equally in all keys. See p. 61.

[S: 28] *Immanent Sphere* (`imsphere.mp3 4:17`). Each note is an overtone of a single underlying fundamental. See p. 69.

[S: 29] *Free from Gravity* (`freegrav.mp3 3:28`). The melodic and harmonic motion conform to a simple additive scale, a regular lattice that organizes pitch space additively in frequency. See p. 69.

[S: 30] *Intersecting Spheres* (`intersphere.mp3 3:33`). The basic timbre is harmonic, and all partials of all tones are integer multiples of 50 Hz. The tuning is similarly a spectral scale consisting of all multiples of 50 Hz (although only a small subset are actually used.) The timbres were created using additive-style synthesis with the program Metasynth [W: 23], and the results were passed through various nonlinearities in Matlab [W: 21]. This causes many new overtones at ever higher frequencies that eventually hit the fold over frequency (22050 for normal CD recording) and begin descending. Because 22050 is divisible by 50, when the partials fold back, they still lie on the same 50 Hz lattice—they just augment (or decrease) the amplitude of the partials. So no matter how many nonlinearities are used, the sound remains within the same harmonic template. Much of the character (the "hair-raising on end") of the timbres is due to this unorthodox method of creating the sounds. See p. 69.

[S: 31] *Over Venus* (`overvenus.mp3 4:25`). This melody floats above a single low tone, playing on the multidimensional harmonics. See p. 69.

[S: 32] *Pulsating Silences* (`pulsilence.mp3 3:33`). A single living note that changes without moving, that grows while remaining still. Even if there was only one note, there would still be music. See p. 69.

[S: 33] *Overtune* (`overtune.mp3 3:54`). Additive synthesis can create very precise and clean sounds. All partials are from the same harmonic series. See p. 69.

[S: 34] *Fourier's Song* (`fouriersong.mp3 3:54`). Also known as *Table 4.1: Properties of the Fourier Transform*, this song was written by Bob Williamson and Bill Sethares "because we love Fourier Transforms, and we know you will too." Perhaps you have never taken a course where everything is laid out in a single song. Well, here it is...a song containing 17% of the theoretical results, 25% of the practical insights, and 100% of the humor of ECE330: Signals and Systems. The music is played in an additive (overtone) scale that consists of all harmonics of 100 Hz. See p. 69 or visit the web pages at [W: 8]. Lyrics appear in Appendix K.

Sound Examples for Chapter 6

[S: 35] *Tritone dissonance curve* (`tridiss.mp3 1:06`). This is the auditory version of Fig. 6.2. See p. 101 and video [V: 9].

[S: 36] *Tritone chime* (`trichime.mp3 0:37`). First, you hear a single note of the "tritone chime." Next, the chime plays the three chords from Fig. 6.3. The chords are then repeated using a more "organ-like" tritone timbre. See p. 102 and video [V: 10].

[S: 37] *Tritone chord patterns* (`trichord.mp3 0:52`). This sound example presents two chord patterns, each repeated once. Which passage appears more consonant, the major or the diminished?

(a) F major, C major, G major, C major
(b) C dim, D dim, $D\sharp$ dim, C dim

Which of the next two patterns feels more resolved?

(c) C dim, C major, C dim, C major
(d) C major, C dim, C major, C dim

Musical scores for these four segments are given in Fig. 6.4. See p. 103.

[S: 38] *Plastic City: A Stretched Journey* (`plasticity.mp3 6:00`). The "same" piece is played with harmonic sounds in 12-tet, with 2.2-stretched sounds, with 1.87-compressed sounds, and finally with 2.1-stretched sounds, all in their respective stretched or compressed tunings. See pp. 58, 109, and 321.

[S: 39] *October 21st* (`october21.mp3 1:42`). There are no real octaves (defined as a frequency ratio of 2 to 1) anywhere in this piece. The sounds in *October 21st* are constructed so that the octave between f and $2f$ is dissonant, whereas the nonoctave between f and $2.1f$ is consonant. Thus, the unit of repetition is a "stretched pseudo-octave" with a frequency ratio of 2.1 to 1. As the structure of the timbres are matched to the structure of the scale, these nonoctave intervals can be consonant, even as the (real) octave is dissonant. The same 2.1-stretched tones were demonstrated in [S: 4]. See pp. 58 and 110.

[S: 40] *A note with partials at 4:5:6:7* (`4567.mp3 0:08`). This note/chord is built from four sine wave partials with frequencies 400, 500, 600, and 700 Hz. See p. 100.

[S: 41] *A note with partials at 1/7:1/6:1/5:1/4* (`7654.mp3 0:08`). This note/chord is built from four sine wave partials with frequencies 400, 467, 560, and 700 Hz. See p. 100.

[S: 42] *4:5:6:7 vs. 1/7:1/6:1/5:1/4* (`4567_7654.mp3 0:16`). The two notes from sound examples [S: 40] and [S: 41] alternate. Which is more consonant? See p. 100.

Sound Examples for Chapter 7

[S: 43] *Tingshaw* (`tingshaw.mp3 4:03`). The tingshaw is a small handbell with a bright and cheerful ring, and it is played in a scale determined by the spectrum of the bell itself. *Tingshaw* is discussed extensively in Chap. 7. See p. 131.

[S: 44] *Chaco Canyon Rock* (`chacorock.mp3 3:38`). Piece based on the rock described at length in Chap. 7. See pp. 139 and 343.

[S: 45] *Duet for Morphine and Cymbal* (`morphine.mp3 3:21`). Each angle in an x-ray diffraction pattern can be mapped to an audible frequency, transforming a crystalline structure into sound. In this piece, complex clusters of tones derived from morphine crystal resonances are juxtaposed over a rhythmic bed supplied by the more percussive timbre of the cymbal. The mapping technique is described at length in Chap. 7. See p. 145.

Sound Examples for Chapter 8

[S: 46] *Adaptation of stretched timbres: minor chord* (`streminoradapt.mp3 0:06`). Stretched timbres play a 12-tet minor chord. After adaptation, this converges to the stretched minor chord detailed in Table 8.2. See p. 169.

[S: 47] *Adaptation of stretched timbres: major chord* (`stremajoradapt.mp3 0:06`). Stretched timbres play a 12-tet major chord. After adaptation, this converges to the stretched major chord detailed in Table 8.2. See p. 169.

[S: 48] *Circle of fifths in 12-tet* (`circle12tet.mp3 0:38`). The circle of fifths moves through all 12 keys, demonstrating one of the great strengths of 12-tet: reasonable consonance in all keys. See p. 168.

[S: 49] *Circle of fifths in C major just intonation* (`circleJICmaj.mp3 0:38`). The circle of fifths demonstrates one of the liabilities of JI: keys that are distant from the tonal center are unuseable. See p. 168.

[S: 50] *Circle of fifths in adaptive tuning* (`circleadapt.mp3 0:38`). Applying adaptation to the circle of fifths allows all chords to maintain the simple integer ratios, combining the best of 12-tet (modulation to all keys) with the consonance of JI. See p. 169.

[S: 51] *Syntonic comma example: JI* (`syntonJIdrift.mp3 0:43`). Each repeat of the phrase in Fig. 8.7 the tuning drifts lower. See p. 170.

[S: 52] *Syntonic comma example: 12-tet* (`synton12tet.mp3 0:21`). The phrase of Fig. 8.7 is performed in 12-tet. See p. 170.

[S: 53] *Syntonic comma example: adaptive tuning* (`syntonadapt.mp3 0:21`). The phrase of Fig. 8.7 does not drift yet maintains fidelity to the simple integer ratios when played in adaptive tuning with harmonic sounds. See p. 170.

[S: 54] *Listening to adaptation* (`listenadapt.mp3 0:32`). Each note has a spectrum containing four inharmonic partials at f, $1.414f$, $1.7f$, and $2f$. Three notes are initialized at the ratios 1, 1.335, and 1.587 (the 12-tet scale steps C, F, and $G\flat$) and allowed to adapt. The final adapted ratios are 1, 1.414, and 1.703. The adaptation is done three times:

(i) With extremely slow adaptation (very small stepsize)
(ii) Slow adaptation
(iii) Medium adaptation

See pp. 99 and 173.

[S: 55] *Scarlatti's K1 Sonata in 12-tet.* (`k001tet12.mp3 0:32`). The first phrase of the sonata. See Fig. 8.10 on p. 175.

[S: 56] *Scarlatti's K1 Sonata in adaptive tuning* (`k001adaptX.mp3 0:32`). Poor choice of stepsizes can lead to wavering pitches in the adaptive tuning. See Fig. 8.10 on p. 175.

[S: 57] *Scarlatti's K1 Sonata in adaptive tuning.* (`k001adapt.mp3 0:32`). Better choices of stepsizes can ameliorate the wavering pitches. See Fig. 8.10 on p. 175.

[S: 58] *Wavering pitches* (`waverpitch.mp3 0:21`). The second measure of Domenico Scarlatti's harpsichord sonata K1 is played three ways:

(i) Scarlatti's K1 sonata in 12-tet.
(ii) Scarlatti's K1 sonata with adaptation. Observe the wavering pitch underneath the trill at the end of the second measure.
(iii) Scarlatti's K1 sonata with adaptation, modified so that "new" notes are adapted ten times as fast as held notes. The wavering pitch is imperceptible.

See p. 175.

[S: 59] *Sliding pitches* (`slidepitch.mp3 0:45`). The kinds of pitch changes caused by the adaptive tuning algorithm are often musically intelligent responses to the context of the piece.

(a) A simple chord sequence from F major to G major is transformed by the adaptive tuning algorithm. The sliding pitch of one note stands out. Each measure is played separately, then together.
(b) The adaptive tuning algorithm "changes" the chord on the fourth beat.
See p. 176.

[S: 60] *Three Ears* (`three_ears.mp3 4:24`). As each new note sounds, its pitch (and that of all currently sounding notes) is adjusted microtonally (based on its spectrum) to maximize consonance. The adaptation causes interesting glides and microtonal pitch adjustments in a perceptually sensible fashion. Listen for the two previous segments from [S: 59]. Many similar effects occur throughout. See pp. 177, 189, and 191.

Sound Examples for Chapter 9

[S: 61] *Adaptive Study No. 1* (`adapt_study1.mp3 2:36`). Example of the pitch glides and wavering pitches using `Adaptun`. See p. 185.

[S: 62] *Adaptive Study No. 2* (`adapt_study2.mp3 2:28`). Using `Adaptun`'s context feature, the wandering of the pitch is reduced. See pp. 185 and 188.

[S: 63] *Compositional technique: example 1* (`breakdrums1.mp3 0:10`). A standard MIDI drum file from the Keyfax Software [W: 17] "Breakbeat" collection is performed using drum sounds. See Fig. 9.3 on p. 191.

[S: 64] *Compositional Technique: example 2* (`breakdrums2.mp3 0:10`). The same MIDI file as in [S: 63] is reochestrated with guitar and bass guitar. See p. 191.

[S: 65] *Compositional technique: example 3* (`breakmap1.mp3 0:20`). Editing the MIDI data in Fig. 9.3 leads to the sequence in Fig. 9.4 on p. 191. The original cymbal part is time stretched and offset in pitch.

[S: 66] *Compositional technique: example 4* (`breakmap2.mp3 0:20`). A variant of [S: 65]. See p. 191.

[S: 67] *Compositional technique: example 5* (`breakmap3.mp3 0:20`) Another variant of [S: 65]. See p. 191.

[S: 68] *Compositional technique: example 6* (`breakadapt1.mp3 0:23`). Adaptation the standard MIDI file of Fig. 9.4 using no context and default settings in `Adaptun`. See p. 191.

[S: 69] *Compositional technique: example 7* (`breakrand1.mp3 0:20`). The sequence in Fig. 9.4 and sound example [S: 65] is transformed by randomizing the bass line over an octave. See p. 192.

[S: 70] *Compositional technique: example 8* (`breakrand2.mp3 0:20`). Randomization of the "fast" line in Fig. 9.4 leads to this arpeggiated guitar. See p. 192.

[S: 71] *Compositional technique: example 9* (`breakrand3.mp3 0:20`). Randomization of the "slow" line in Fig. 9.4 leads to this synthesized melody. See p. 192.

[S: 72] *Compositional technique: example 10* (`breakadapt2.mp3 0:21`). After adaptation, example [S: 71] sounds very different. See p. 192.

[S: 73] *Compositional technique: example 11* (`breakadapt3.mp3 0:47`). Sound example [S: 71] is adapted with full convergence of the algorithm. The sound example is played twice: first without the melody, and then with. See p. 192.

[S: 74] *Adventiles in a Distorium* (`adventiles.mp3 4:46`). An adaptively tuned composition featuring frenetically distorted guitars. See p. 189.

[S: 75] *Aerophonious Intent* (`aerophonious.mp3 3:24`). An adaptively tuned composition orchestrated using an extreme form of hocketing. See p. 189.

[S: 76] *Story of Earlight* (`earlight.mp3 3:53`). An adaptively tuned recitation of whispers and flutes. See p. 189.

[S: 77] *Excitalking Very Much* (`excitalking.mp3 3:32`). An adaptively tuned conversation between a synthetic bass and a synthetic clarinet. See p. 189.

[S: 78] *Inspective Liquency* (`inspective.mp3 3:46`). An adaptively tuned piece where no note remains fixed. See p. 189.

[S: 79] *Local Anomaly* (`localanomaly.mp3 3:27`). This piece was created from a standard MIDI drum track, which was randomized and orchestrated using various percussive stringed sounds such as sampled guitars and basses. The extremely dissonant but highly rhythmic soundscape was input into `Adaptun`, and the notes adapted toward consonance. No context was used. See pp. 189 and 193.

[S: 80] *Maximum Dissonance* (`maxdiss.mp3 3:24`). Instead of minimizing the dissonance, this piece maximizes the dissonance at every time instant. See pp. 189 and 195.

[S: 81] *Persistence of Time* (`persistence.mp3 4:54`). Polyrhythms beat three against two, a paleo-futuristic audio conundrum where all intervals adapt to maximize instantaneous consonance. See pp. 189 and 189.

[S: 82] *Recalled Opus* (`recalledopus.mp3 3:45`). At each instant in time, these "violins" strive to minimize dissonance. See pp. 185, 189, and 193.

[S: 83] *Saint Vitus Dance* (`saintvitus.mp3 3:32`). Begin with a MIDI drum pattern. Use the pattern to trigger a sampled guitar sound; it is wildly dissonant because the pitches are essentially random. At each time instant, perturb the pitches of all currently sounding notes to the nearest intervals that maximize consonance. Thus is born an adaptively tuned dance.

[S: 84] *Simpossible Taker* (`simpossible.mp3 3:20`). An adaptively tuned composition that began as a hip hop drum pattern. See pp. 189 and 191.

[S: 85] *Wing Donevier* (`wing.mp3 3:17`). An adaptively tuned composition in seven beats per measure. See pp. 189 and 193.

Sound Examples for Chapter 13

[S: 86] *11-tet spectral mappings: before and after* (`tim11tet.mp3 1:20`). Several different instrumental sounds alternate with their 11-tet spectrally mapped versions:

(i) Harmonic trumpet compared with 11-tet trumpet
(ii) Harmonic bass compared with 11-tet bass
(iii) Harmonic guitar compared with 11-tet guitar
(iv) Harmonic pan flute compared with 11-tet pan flute
(v) Harmonic oboe compared with 11-tet oboe
(vi) Harmonic "moog" synth compared with 11-tet "moog" synth
(vii) Harmonic "phase" synth compared with 11-tet "phase" synth

See p. 277 and video [V: 11].

[S: 87] *12-tet vs. 11-tet* (`tim11vs12.mp3 0:37`). A short sequence of major chords are played:

(viii) Harmonic oboe in 12-tet

(ix) Spectrally mapped 11-tet oboe in 12-tet
(x) Harmonic oboe in 11-tet
(xi) Spectrally mapped 11-tet oboe in 11-tet

See p. 279 and video [V: 12].

[S: 88] *The Turquoise Dabo Girl* (`dabogirl.mp3 4:16`). Many of the kinds of effects normally associated with (harmonic) tonal music can occur, even in such strange settings as 11-tet (which is often considered among the hardest tunings in which to play tonal music). Consider, for instance, the harmonization of the 11-tet pan flute melody that occurs in the "chorus." Does this have the feeling of some kind of (perhaps unfamiliar) "cadence" as the melody resolves back to its "tonic?" Spectral mapping of the instrumental sounds allows such xentonal motion. See pp. 59 and 279.

[S: 89] *The Turquoise Dabo Girl (first 16 bars)* (`dabogirlX.mp3 0:29`). In 11-tet, but using unmapped harmonic sounds. The "out-of-timbre" percept is unmistakable. See p. 279.

[S: 90] *Tom Tom Spectral Mappings: Before and After* (`tomspec.mp3 0:37`). Several different instrumental sounds alternate with versions mapped into the spectrum of a tom tom:

(i) Harmonic flute compared with tom tom flute
(ii) Harmonic trumpet compared with tom tom trumpet
(iii) Harmonic bass compared with tom tom bass
(iv) Harmonic guitar compared with tom tom guitar

See p. 281 and video [V: 13].

[S: 91] *Glass Lake* (`glasslake.mp3 3:08`). Instruments that are spectrally mapped "too far" can lose their tonal integrity. When guitars, basses, and flutes are transformed into the partial structure of a drum (a tom tom), they are almost unrecognizable. But this does not mean that they are useless. All sounds in this piece (except for the percussion) were demonstrated in [S: 90]. The "tom tom" scale supports perceptible "chords," though the chords are not necessarily composed of familiar intervals. Tom Staley played a key role in writing and performing *Glass Lake*. See pp. 277 and 281.

[S: 92] *A harmonic cymbal* (`harmcym.mp3 0:23`). A cymbal is spectrally mapped into a harmonic spectrum. The resulting sound is pitched and capable of supporting melodies and chords.

(i) The original sample contrasted with the spectrally mapped version
(ii) A simple "chord" pattern played with the original sample, and then with the spectrally mapped version

See p. 282 and video [V: 14].

[S: 93] *Sonork* (`sonork.mp3 3:15`). The origin of each sound is a cymbal, spectrally mapped to nearby harmonic templates to create the bass, synth, and other instrumental sounds. See pp. 277 and 283.

[S: 94] *Inharmonic drum* (`inharmdrum.mp3 0:59`). This drum sound is incapable of supporting melody or harmony. See p. 283.

[S: 95] *Harmonic drum* (`harmdrum.mp3 1:29`). The drum sound from [S: 94] is spectrally mapped to the nearest harmonic template. It can now support both melody or harmony. See p. 283.

[S: 96] *Harmonic and inharmonic drum* (`harm+inharm.mp3 1:29`). The sounds from [S: 94] (the original inharmonic drum) and [S: 95] (the spectrally mapped version) are combined. See p. 283.

[S: 97] *Hexavamp* (`hexavamp.mp3 3:22`). A "classical" guitar is spectrally mapped into 16-tet and overdubbed with itself. See pp. 59 and 277.

[S: 98] *Seventeen Strings* (`17strings.mp3 3:22`). A sampled Celtic harp is transformed for compatibility with 17-tet. See pp. 59, 279, and 277.

[S: 99] *Unlucky Flutes* (`13flutes.mp3 3:51`). Flutes, guitars, bass, and keyboards are spectrally mapped into 13-tet. All instruments clearly retain their tonal identity, and yet sound harmonious even on sustained passages. Compare with the 13-tet demonstration on Carlos' *Secrets of Synthesis* [D: 6], which is introduced, "But the worst way to tune is probably this temperament of 13 equal steps." See pp. 59 and 277.

[S: 100] *Truth on a Bus* (`truthbus.mp3 3:22`). A 19-tet guitar piece that is unabashedly diatonic. If you were not listening carefully, you might imagine that this was a real guitar, tuned normally, and played skillfully. You would be very wrong. See pp. 277 and 59.

[S: 101] *Sympathetic Metaphor* (`sympathetic.mp3 3:59`). This guitar has 19 tones in each octave, and the melody dances pensively on a delicately balanced timbre. Peter Kidd plays the excellent fretless bass. See pp. 59 and 277.

Sound Examples for Chapter 14

[S: 102] *Ten Fingers* (`tenfingers.mp3 3:18`). Demonstrates the kind of consonance effects achievable in 10-tet. The guitar-like 10-tet timbre is created by spectrally mapping a sampled guitar into an induced spectrum. The full title of this piece is *If God Had Intended Us To Play In Ten Tones Per Octave, Then He Would Have Given Us Ten Fingers*. See pp. 59, 249, 277, 293, and 322.

[S: 103] *Ten Fingers: harmonic guitar* (`tenfingersX.mp3 0:28`). The first 16 bars of *Ten Fingers* [S: 102] are played with a harmonic (sampled) guitar. The *out-of-spectrum* effect is unmistakable. See p. 294.

[S: 104] *Circle of Thirds* (`circlethirds.mp3 3:41`). There is an interesting and beautiful chord pattern in 10-tet that is analogous to (but very different from) the traditional circle of fifths. This piece cycles around the *Circle of Thirds* over and over: first fast, then slow, and then fast again. See p. 297.

[S: 105] *Isochronism* (`isochronism.mp3 3:55`). When there are ten equal tones in each octave, special tone colors are needed to align the partials into consonant patterns. See p. 277 and p. 298 for a description of the 10-tet chord patterns.

[S: 106] *Anima* (`anima.mp3 4:03`). Uses modified timbres to effect a balance between coherence and chaos, between the obvious and the obscure. See p. 277. Exploits the 10-tet tritone chords described starting on p. 300.

[S: 107] *Swish* (`swish.mp3 3:20`). Timbres constructed in *Metasynth* swirl and mutate as the piece evolves in 5-tet, which is analogous to a wholetone scale inside 10-tet. See p. 59.

Sound Examples for Chapter 15

[S: 108] *Tuning of a classical Thai piece* (`thai7tet.mp3 0:28`). Demonstrates the procedure whereby the tuning of a piece can be found from the recording. Begins

with the first 10 seconds of *Sudsaboun* from [D: 39] and then separates the melody into individual notes, each of which is compared with a sine wave to determine its pitch. See Sect. 15.2 on p. 304.

[S: 109] *Comparison of harmonic sounds and their spectrally mapped 7-tet versions* (`7tetcompare.mp3 0:25`). Three instruments are demonstrated:

(i) Three different notes of a bouzouki
(ii) Three different notes of a harp
(iii) A pan flute

See pp. 311 and 313.

[S: 110] *Comparison between 7-tet and a 12-tet major scale* (`7vs12.mp3 1:19`). The theme of the simple tune from sound example [S: 2] is played first in 12-tet and then in 7-tet, using the "naive" mapping between 7-tet and the diatonic (major) scale defined in (15.2) and using harmonic timbres. See p. 312.

[S: 111] *Comparison between 7-tet and a 12-tet major scale* (`7vs12bar.mp3 1:19`). The theme of the simple tune from sound example [S: 2] is played first in 12-tet and then in 7-tet, using the "naive" mapping between 7-tet and the diatonic (major) scale defined in (15.2) with timbres have been mapped to the spectrum of an ideal bar. See p. 312.

[S: 112] *Scarlatti's K380 in 7-tet* (`K380tet7.mp3 1:29`). Using the "naive" mapping between 7-tet and the diatonic (major) scale of (15.2), Scarlatti's theme looses its harmonic meaning. More conventional tunings of K380 can be heard in sound examples [S: 17] through [S: 22]. The timbres are harmonic. See p. 312.

[S: 113] *Scarlatti's K380 in 7-tet* (`K380tet7bar.mp3 1:29`). Using the "naive" mapping between 7-tet and the diatonic (major) scale of (15.2), Scarlatti's theme loses its harmonic meaning. More conventional tunings of K380 can be heard in sound examples [S: 17] through [S: 22]. The timbres have been mapped to the spectrum of an ideal bar. See p. 312.

[S: 114] *Scarlatti's K380 in 12-tet* (`K380tet12bar.mp3 1:29`). This performance of K380 uses timbres that have been mapped to the spectrum of an ideal bar. See p. 312.

[S: 115] *March of the Wheels* (`marwheel.mp3 3:38`). The notes of a standard MIDI drum track are mapped into the 7-tet scale, creating the rhythmic foundation for this piece. The notes are randomized, creating a variety of serendipitous melodies. See pp. 59 and 313.

[S: 116] *Pagan's Revenge* (`pagan.mp3 3:55`). The notes of a standard MIDI file (Paganini's Caprice No.24 performed by D. Lovell) are mapped into 7-tet, creating the foundation for this piece. At the halfway point, the MIDI data in the file was time reversed so that the theme proceeds forward and then backward—finally ending on the first note. See pp. 59 and 315.

[S: 117] *Nothing Broken in Seven* (`broken.mp3 3:29`). A single six-note isorhythmic melody is repeated over and over, played simultaneously at five different speeds. See pp. 59 and 315.

[S: 118] *Phase Seven* (`phase7.mp3 3:41`). A single eight-note isorhythmic melody is repeated over and over, played simultaneously at five different speeds. See pp. 59 and 315.

V: Video Examples on the CD-ROM

The video files on the CD-ROM are saved in the `.avi` *format, which is readable using* `Windows Media Player` *or* `Quicktime`*. Navigate to* `TTSS/Videos/` *and launch the* `*.avi` *file by double clicking, or by opening the file from within the player. References in the body of the text to the video examples are coded with* [V:] *to distinguish them from references to the bibliography, discography, and sound examples. The video examples may also be accessed using a web browser. Open the file* `TTSS/Contents.html` *in the top level of the CD-ROM and navigate using the html inteface.*

[V: 1] *Challenging the Octave* (`challoct.avi 0:21`). See p. 2 and sound example [S: 1]. The spectrum of the sound is constructed so that the octave between f and $2f$ is dissonant while the nonoctave f to $2.1f$ is consonant.

[V: 2] *Pitch of Periodic Sounds* (`pitchclicks.avi 0:21`). See p. 33. The five buzzy sounds all have the same period; the pitch jumps up an octave somewhere between (a) and (e).

[V: 3] *Virtual Pitch of Harmonic Partials* (`virtpitch.avi 0:29`). See p. 35. Sine waves at frequencies 1040, 1300, and 1560 are presented individually and then together. With all three sounding, the primary percept is of a low buzzy sound at a pitch corresponding to 260 Hz.

[V: 4] *Virtual Pitch of Inharmonic Partials* (`virtpitchX.avi 0:30`). See p. 35. Sine waves at frequencies 1060, 1320, and 1580 are presented individually and then together. With all three sounding, the primary percept is of a low buzzy sound at a pitch corresponding to about 264 Hz, although this is less clear than when the partials are harmonically related, as in [V: 3].

[V: 5] *Beating of Sine Waves I* (`beats1.avi 0:23`). See p. 41 and sound example [S: 8].

[V: 6] *Beating of Sine Waves II* (`beats2.avi 0:23`). See p. 41 and sound example [S: 9].

[V: 7] *Beating of Sine Waves III* (`beats3.avi 0:23`). See p. 41 and sound example [S: 10].

[V: 8] *Dissonance Between Two Sine Waves* (`sinediss.avi 1:06`). See p. 45 and sound example [S: 11]. A sine wave of fixed frequency 220 Hz is played along with a "sine wave" with frequency that begins at 220 Hz and slowly increases to 470 Hz.

[V: 9] *Tritone Dissonance Curve* (`tridiss.avi 1:04`). See p. 101 and sound example [S: 35]. This is the auditory version of Fig. 6.2.

[V: 10] *Tritone Chime* (`trichime.avi 0:42`). See p. 102 and sound example [S: 36]. First, you hear a single note of the "tritone chime." Next, the chime plays the

three chords from Fig. 6.3. The chords are then repeated using a more "organ-like" tritone timbre.

[V: 11] *11-tet Spectral Mappings: Before and After* (`tim11tet.avi 1:15`). See p. 277 and sound example [S: 86]. Several different instrumental sounds alternate with their 11-tet spectrally mapped versions.

[V: 12] *12-tet vs. 11-tet* (`tim11vs12.avi 0:38`). See p. 279 and sound example [S: 87]. A short sequence of chords is played that compares spectrally mapped 11-tet sounds to harmonic sounds when playing chords drawn from the 11-tet scale.

[V: 13] *Tom Tom Spectral Mappings: Before and After* (`tomspec.avi 0:44`). See p. 281 and sound example [S: 90]. Several different instrumental sounds alternate with versions mapped into the spectrum of a tom tom:

[V: 14] *A Harmonic Cymbal* (`harmcym.avi 0:23`). See p. 282 and sound example [S: 92]. A cymbal is spectrally mapped into a harmonic spectrum—the resulting sound is pitched and capable of supporting melodies and chords.

W: World Wide Web and Internet References

This section contains all web links referred to throughout Tuning, Timbre, Spectrum, Scale. *References in the body of the text to websites are coded with* [W:] *to distinguish them from references to the bibliography, discography, and sound and video examples. The web examples may also be accessed using a web browser. Open the file* `TTSS/Links.html` *in the top level of the CD-ROM and navigate using the html interface.*

[W: 1] *Alternate tuning mailing list,* http://groups.yahoo.com/group/tuning/ [This group and [W: 18] continually discuss techniques of creating and analyzing music that is outside the Western tradition.]

[W: 2] *Bitheadz, Inc.*, http://www.bitheadz.com [Makers of audio tools such as the *Unity* software synthesizer.]

[W: 3] *How harmonic are harmonics?* http://www.phys.unsw.edu.au/~jw/ harmonics.html [Discussion of inharmonicities in strings and air column instruments.]

[W: 4] *Classical MIDI Archives*, http://www.classicalarchives.com/ [Thousands of standard MIDI files are available here free for listening, studying, and enjoying.]

[W: 5] *Content Organs*, http://www.content-organs.com [An organ maker that offers the hermode tuning in its organs.]

[W: 6] *Corporeal Meadows*, http://www.corporeal.com/ [Website devoted to Harry Partch. Partch's music, instruments, and personality are all profiled here.]

[W: 7] J. A. deLaubenfels, "Adaptive Tuning Web Site," http:// www.adaptune.com/ [Also, see John's personal web page at http://personalpages.bellsouth.net/j/d/jdelaub/jstudio.htm for sound examples and further details on the spring method of adaptive tuning.]

[W: 8] *ECE330: Signals and Systems* Prof. Sethares' class website for the course on Fourier transforms is:
http://eceserv0.ece.wisc.edu/~sethares/classes/ece330.html and the official university website is:
http://www.engr.wisc.edu/ece/courses/ece330.html

[W: 9] *P. Erlich on Harmonic Entropy*, http://tonalsoft.com/td/erlich/entropy.htm [Erlich discusses models of harmonic entropy in a series of posts to the Tuning Digest beginning in Sept. 1997.]

[W: 10] P. Erlich, "The forms of tonality," http://lumma.org/tuning/erlich/ Also available on the CD `TTSS/PDF/erlich-forms.pdf`. [Concepts of tone-lattices, scales, and notational systems for 5-limit and 7-limit music.]

[W: 11] P. Frazer, *Midicode Synthesizer*, http://www.midicode.com [Implements a method of dynamic retuning in a software syntheizer.]

[W: 12] *Freenote Music*, http://microtones.com/new.htm [Dedicated to microtonal guitars and recordings.]

[W: 13] *Frog Peak Music*, http://www.frogpeak.org/ [This composer's collective is a gold mine of alternatively tuned music.]
[W: 14] *The Justonic Tuning System*, http://www.justonic.com/ [Jutonic's pitch palette uses any 12-tone just, or harmonic scale to create a 3-dimensional array of tones that can be used to automatically retune a synthesizer as it plays.]
[W: 15] *The Hermode Tuning*, http://www.hermode.com/ [A form of automated tuning implemented in the Waldorf Virus C synthesizer. Website has good demonstrations of the uses of adaptive tunings.]
[W: 16] *Institute for Psychoacoustics and Music*, http://www.ipem.rug.ac.be/ [Part of the University of Ghent, IPEM is Belgium's premier center for electronic music.]
[W: 17] *Keyfax Software*, http://www.keyfax.com [Professionally recorded standard MIDI files.]
[W: 18] *Make Micro Music mailing list,* http:// groups.yahoo.com/ group/ MakeMicroMusic/ [This group and [W: 1] continually discuss techniques of creating and analyzing music that is outside the Western tradition.]
[W: 19] *Making Microtonal Music Website*, http://www.microtonal.org/ [A gathering point for people who are actively making microtonal music, and for those who would like to join them.]
[W: 20] *Mark of the Unicorn*, http://www.motu.com/ [Makers of music hardware and software,including *Digital Performer*, a MIDI and audio sequencer.]
[W: 21] *Matlab*, http://www.mathworks.com/ [General purpose programming language common in signal processing and engineering: "the language of technical computing."]
[W: 22] *Max 4.0 Reference Manual*, http://www.cycling74.com/products/dldoc.html [Website of Cycling '74, distributers of Max programming language. See also [B: 210].]
[W: 23] *Metasynth*, http://www.uisoftware.com/ [A powerful graphic tool for sound manipulation and visualization.]
[W: 24] *Microtonal Dictionary*, http://tonalsoft.com/ [Joseph Monzo's online dictionary of musical tuning terms is an excellent resource.]
[W: 25] *MIDI file formats described*, http://www.sonicspot.com/guide/midifiles.html
[W: 26] W. Mohrlok, *The Hermode Tuning System* [This provides a comprehensive description of the operation of the hermode tuning, and is available on the CD in `TTSS/pdf/hermode.pdf`.]
[W: 27] *Scala Homepage*, http://www.xs4all.nl/~huygensf/scala/ [Powerful software tool for experimentation with musical tunings.]
[W: 28] *Tuning, Timbre, Spectrum, Scale* http://eceserv0.ece.wisc.edu/~sethares/
[W: 29] Smith, J. O. "Bandlimited interpolation—interpretation and algorithm," http://ccrma-www.stanford.edu/ ~jos/resample/ [Excellent discussion of audio signal processing with focus on interpolation techniques.]
[W: 30] *John Starrett's Microtonal Music*, http:// www.nmt.edu/ ~jstarret/ microtone.html [Great resource for microtonal music, instruments, and tools.]
[W: 31] *Tune Smithy*, http://www.tunesmithy.connectfree.co.uk/ [A program for algorithmic music composition that includes extensive microtonal support and a dynamic tuning feature.]
[W: 32] *Vicentino's adaptive-JI of 1555*, http:// tonalsoft.com/ monzo/ vicentino/ vicentino.htm [Vicentino's "Second tuning of 1555" is composed of two chains of 1/4-comma meantone that can avoid comma drift.]

[W: 33] *Access "Virus" Synthesizer*, http://www.access-music.de/ [The hermode tuning is available in the Virus synthesizer.]
[W: 34] *Waldorf Synthesizers*, http://www.waldorf-music.de [First commercial implementation of an adaptive tuning.]

Index