



Tuning, Timbre, Spectrum, Scale

Second Edition

William A. Sethares



Springer

CD-ROM



Included

William A. Sethares

Tuning, Timbre, Spectrum, Scale

September 16, 2004

Springer

Berlin Heidelberg New York

Hong Kong London

Milan Paris Tokyo

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 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?

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.

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
August 2004

William A. Sethares

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

X Acknowledgments

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

Prelude	V
Acknowledgments	IX
Variables, Abbreviations, Definitions	XVII
1 The Octave Is Dead . . . Long Live the Octave	1
<i>Introducing a dissonant octave—almost any interval can be made consonant or dissonant by proper choice of timbre.</i>	
1.1 A Challenge	1
1.2 A Dissonance Meter	3
1.3 New Perspectives	5
1.4 Overview	8
2 The Science of Sound	11
<i>What is sound made of? How does the frequency of a sound relate to its pitch? How does the spectrum of a sound relate to its timbre? How do we know these things?</i>	
2.1 What Is Sound?	11
2.2 What Is a Spectrum?	13
2.3 What Is Timbre?	26
2.4 Frequency and Pitch	31
2.5 Summary	36
2.6 For Further Investigation	36
3 Sound on Sound	39
<i>Pairs of sine waves interact to produce interference, beating, roughness, and the simplest setting in which (sensory) dissonance occurs.</i>	

3.1	Pairs of Sine Waves	39
3.2	Interference	39
3.3	Beats	40
3.4	Critical Band and JND	42
3.5	Sensory Dissonance	45
3.6	Counting Beats	47
3.7	Ear vs. Brain	49
4	Musical Scales	51
	<i>Many scales have been used throughout the centuries.</i>	
4.1	Why Use Scales?	51
4.2	Pythagoras and the Spiral of Fifths	52
4.3	Equal Temperaments	56
4.4	Just Intonations	60
4.5	Partch	63
4.6	Meantone and Well Temperaments	64
4.7	Spectral Scales	66
4.8	Real Tunings	69
4.9	Gamelan Tunings	72
4.10	My Tuning Is Better Than Yours	72
4.11	A Better Scale?	73
5	Consonance and Dissonance of Harmonic Sounds	75
	<i>The words “consonance” and “dissonance” have had many meanings. This book focuses primarily on sensory consonance and dissonance.</i>	
5.1	A Brief History	75
5.2	Explanations of Consonance and Dissonance	79
5.3	Harmonic Dissonance Curves	84
5.4	A Simple Experiment	91
5.5	Summary	91
5.6	For Further Investigation	92
6	Related Spectra and Scales	93
	<i>The relationship between spectra and tunings is made precise using dissonance curves.</i>	
6.1	Dissonance Curves and Spectrum	93
6.2	Drawing Dissonance Curves	95
6.3	A Consonant Tritone	97
6.4	Past Explorations	100
6.5	Found Sounds	108
6.6	Properties of Dissonance Curves	115

6.7	Dissonance Curves for Multiple Spectra	119
6.8	Dissonance “Surfaces”	121
6.9	Summary	124
7	A Bell, A Rock, A Crystal	127
	<i>Three concrete examples demonstrate the usefulness of related scales and spectra in musical composition.</i>	
7.1	Tingshaw: A Simple Bell	127
7.2	Chaco Canyon Rock	135
7.3	Sounds of Crystals	141
7.4	Summary	147
8	Adaptive Tunings	149
	<i>Adaptive tunings modify the pitches of notes as the music evolves in response to the intervals played and the spectra of the sounds employed.</i>	
8.1	Fixed vs. Variable Scales	149
8.2	The Hermode Tuning	151
8.3	Spring Tuning	153
8.4	Consonance-Based Adaptation	155
8.5	Behavior of the Algorithm	158
8.6	The Sound of Adaptive Tunings	166
8.7	Summary	170
9	A Wing, An Anomaly, A Recollection	171
	<i>Adaptation: tools for retuning, techniques for composition, strategies for listening.</i>	
9.1	Practical Adaptive Tunings	171
9.2	A Real-Time Implementation in Max	172
9.3	The Simplified Algorithm	174
9.4	Context, Persistence, and Memory	175
9.5	Examples	176
9.6	Compositional Techniques and Adaptation	180
9.7	Toward an Aesthetic of Adaptation	185
9.8	Implementations and Variations	187
9.9	Summary	188
10	The Gamelan	191
	<i>In the same way that Western harmonic instruments are related to Western scales, so the inharmonic spectrum of gamelan instruments are related to the gamelan scales.</i>	

10.1 A Living Tradition	191
10.2 An Unwitting Ethnomusicologist	192
10.3 The Instruments	194
10.4 Tuning the Gamelan	202
10.5 Spectrum and Tuning	207
10.6 Summary	211
11 Consonance-Based Musical Analysis	213
<i>The dissonance score demonstrates how sensory consonance and dissonance change over the course of a musical performance. What can be said about tunings used by Domenico Scarlatti using only the extant sonatas?</i>	
11.1 A Dissonance “Score”	213
11.2 Reconstruction of Historical Tunings	223
11.3 What’s Wrong with This Picture?	233
12 From Tuning to Spectrum	235
<i>How to find related spectra given a desired scale.</i>	
12.1 Looking for Spectra	235
12.2 Spectrum Selection as an Optimization Problem	235
12.3 Spectra for Equal Temperaments	237
12.4 Solving the Optimization Problem	241
12.5 Spectra for Tetrachords	244
12.6 Summary	255
13 Spectral Mappings	257
<i>How to relocate the partials of a sound for compatibility with a given spectrum, while preserving the richness and character of the sound.</i>	
13.1 The Goal: Life-like Inharmonic Sounds	257
13.2 Mappings between Spectra	259
13.3 Examples	266
13.4 Discussion	273
13.5 Summary	278
14 A “Music Theory” for 10-tet	279
<i>Each related spectrum and scale has its own “music theory.”</i>	
14.1 What Is 10-tet?	279
14.2 10-tet Keyboard	280
14.3 Spectra for 10-tet	281
14.4 10-tet Chords	282
14.5 10-tet Scales	288
14.6 A Progression	288
14.7 Summary	289

15	Classical Music of Thailand and 7-tet	291
	<i>Seven-tone equal temperament and the relationship between spectrum and scale in Thai classical music.</i>	
15.1	Introduction to Thai Classical Music	291
15.2	Tuning of Thai Instruments	292
15.3	Timbre of Thai Instruments	293
15.4	Exploring 7-tet	298
15.5	Summary	303
16	Speculation, Correlation, Interpretation, Conclusion	305
16.1	The Zen of Xentonality	305
16.2	Coevolution of Tunings and Instruments	306
16.3	To Boldly Listen	308
16.4	New Musical Instruments?	310
16.5	Silence Hath No Beats	311
16.6	Coda	312
	Appendices	313
A.	Mathematics of Beats: <i>Where beats come from</i>	315
B.	Ratios Make Cents: <i>Convert from ratios to cents and back again</i>	317
C.	Speaking of Spectra: <i>How to use and interpret the FFT</i>	319
D.	Additive Synthesis: <i>Generating sound directly from the sine wave representation: a simple computer program</i>	327
E.	How to Draw Dissonance Curves: <i>Detailed derivation of the dissonance model, and computer programs to carry out the calculations</i>	329
F.	Properties of Dissonance Curves: <i>General properties help give an intuitive feel for dissonance curves</i>	333
G.	Analysis of the Time Domain Model: <i>Why the simple time domain model faithfully replicates the frequency domain model</i>	339
H.	Behavior of Adaptive Tunings: <i>Mathematical analysis of the adaptive tunings algorithm</i>	345
I.	Symbolic Properties of \oplus -Tables: <i>Details of the spectrum selection algorithm</i>	349
J.	Harmonic Entropy: <i>A measure of harmonicity</i>	355
K.	Fourier's Song: <i>Properties of the Fourier transform</i>	359
L.	Tables of Scales: <i>Miscellaneous tunings and tables</i>	361
	B: Bibliography	365
	D: Discography	377

S: Sound Examples on the CD-ROM	381
V: Video Examples on the CD-ROM	393
W: World Wide Web and Internet References	395
Index	397

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. 365.
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. 377.
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, \dots, f_n and amplitudes a_1, a_2, \dots, 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.

XVIII Variables, Abbreviations, Definitions

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 i th 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. 381.
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. 393.
[W:]	Web references, see p. 395
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 <i>oh-plus</i> , this symbol indicates the “sum” of two intervals in the symbolic method of constructing spectra.

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 common-sense 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 Olson² 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. Mathieu³ 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.

Arthur Benade⁴ observes that the similarity between notes an octave apart has been enshrined in many of the world’s languages:

¹ From [B: 66].

² [B: 123].

³ [B: 104].

⁴ [B: 12].

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 Storr⁵ 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 example⁶ [S: 1] and repeated in the first video example⁷ [V: 1]. The demonstration consists of only a handful of notes, as shown in Fig. 1.1.

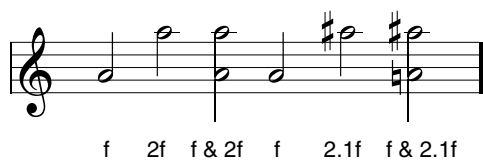


Fig. 1.1. In sound example [S: 1] and video example [V: 1], the timbre of the sound is constructed so that the octave between f and $2f$ is dissonant while the nonoctave f to $2.1f$ is consonant. Go listen to this example now.

A note is played (with a fundamental frequency $f = 450$ Hz⁸) 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 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).

⁵ [B: 184].

⁶ Beginning on p. 381 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. 393 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.

As will become apparent, there is a relationship between the kinds of sounds made by Western instruments (i.e., harmonic⁹ 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.

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

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

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.

¹⁰ See Chap. 10 “The Gamelan” for details and caveats.

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 octave¹³ 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 intervals occur close to (but not exactly at¹⁶) 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,"

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.

¹⁶ Table 6.1 on p. 97 shows how close these approximations are.

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. 153.) 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).

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

¹⁷ See Polansky [B: 142] and Waage [B: 202].

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 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”

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

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.

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 definition¹ 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.

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

¹ from the American Heritage Dictionary.

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

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.

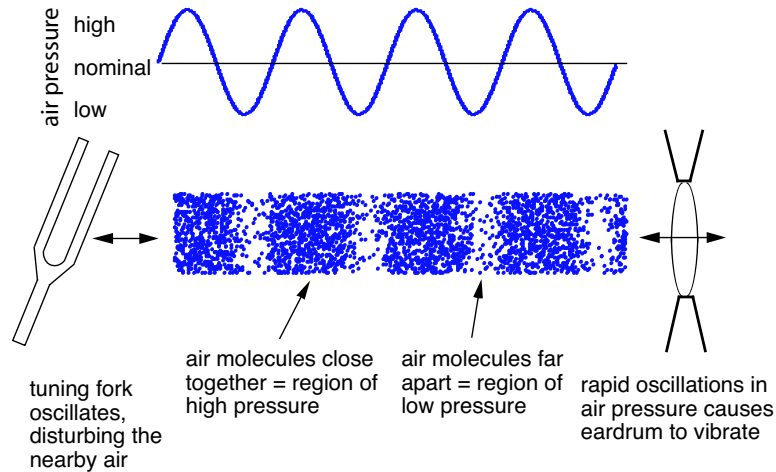


Fig. 2.1. Sound as a pressure wave. The peaks represent times when air molecules are clustered, causing higher pressure. The valleys represent times when the air density (and hence the pressure) is lower than nominal. The wave pushes against the eardrum in times of high pressure, and pulls (like a slight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

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.

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)

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

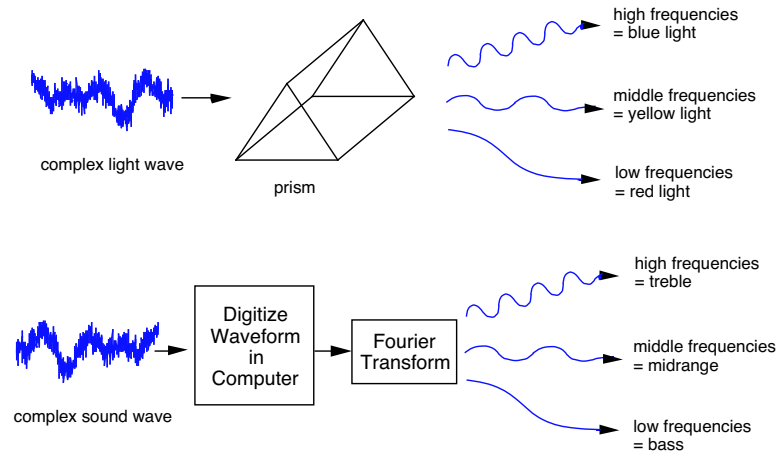


Fig. 2.2. Just as a prism separates light into its simple constituent elements (the colors of the rainbow), the Fourier Transform separates sound waves into simpler sine waves in the low (bass), middle (midrange), and high (treble) frequencies. Similarly, the auditory system transforms a pressure wave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

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 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.

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 spec-

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.

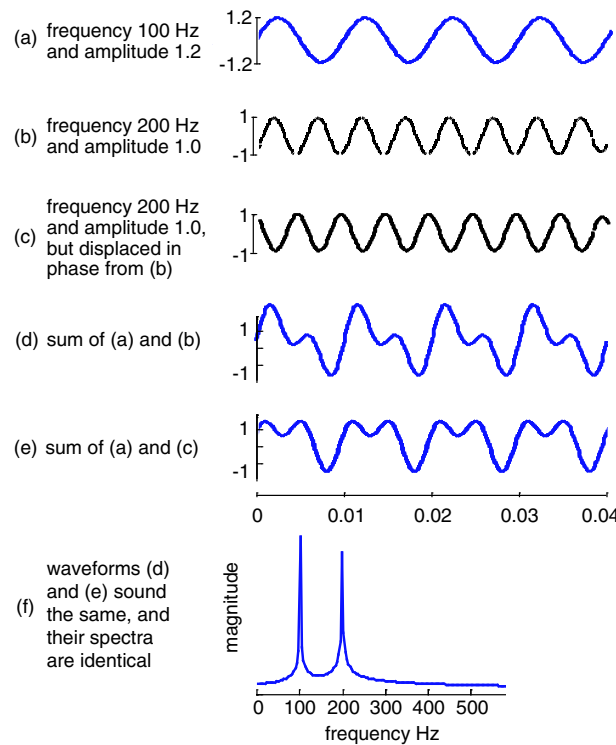


Fig. 2.3. Spectrum of a sound consisting of two sine waves.

trum 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.

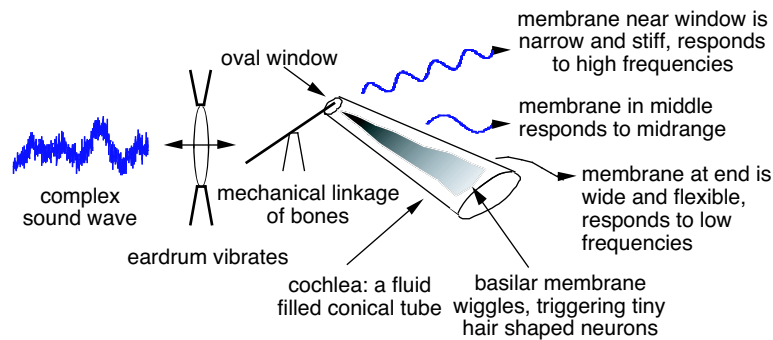


Fig. 2.4. The auditory system as a biological spectrum analyzer that transforms a pressure wave into a frequency selective spatial array.

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.

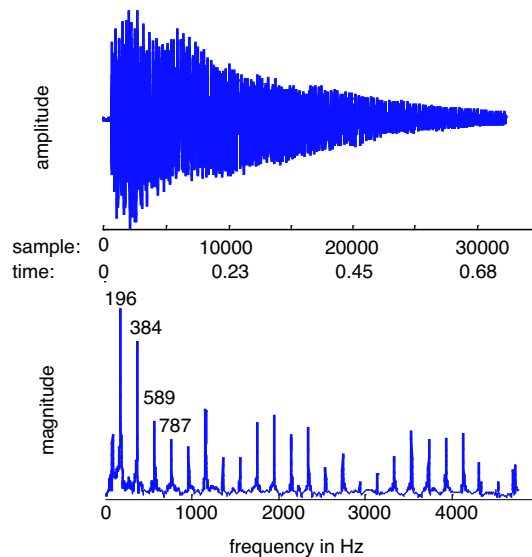


Fig. 2.5. Waveform of a guitar pluck and its spectrum. The top figure shows the first $3/4$ second (32,000 samples) of the pluck of the G string of an acoustic guitar. The spectrum shows the fundamental at 196 Hz, and near integer harmonics at 384, 589, 787, . . .

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.

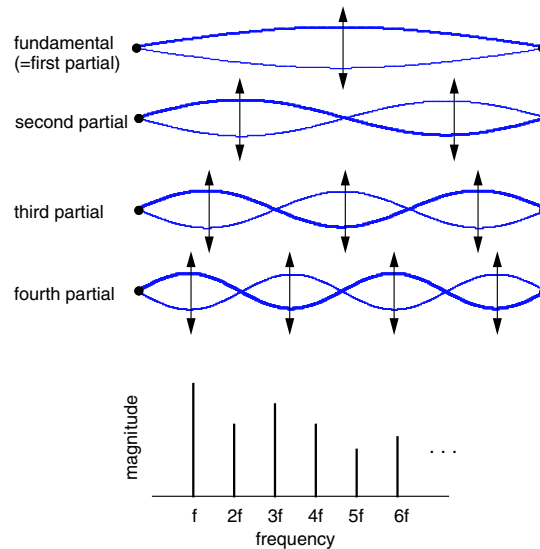


Fig. 2.6. Vibrations of an ideal string and its spectrum. Because the string is fixed at both ends, it can only sustain oscillations when a half period fits exactly into its length. Thus, if the fundamental occurs at frequency f , the second partial must be at $2f$, the third at $3f$, etc., as shown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

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) 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.

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.

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

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.

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.

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.

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	<i>A</i>	110
2	220	<i>A</i>	220
3	330	<i>E</i>	330
4	440	<i>A</i>	440
5	550	<i>C</i> ♯	554
6	660	<i>E</i>	659
7	770	<i>G</i>	784
8	880	<i>A</i>	880
9	990	<i>B</i>	988
10	1100	<i>C</i> ♯	1109
11	1210	<i>D</i> ♯	1245
12	1320	<i>E</i>	1318
13	1430	<i>F</i>	1397
14	1540	<i>G</i>	1568
15	1650	<i>G</i> ♯	1661
16	1760	<i>A</i>	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

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.

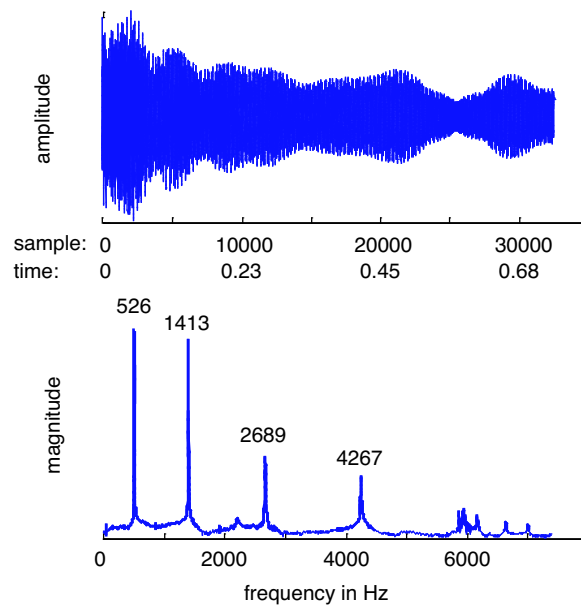


Fig. 2.7. Waveform of the strike of a metal bar and the corresponding spectrum. The top figure shows the first 3/4 second (32,000 samples) of the waveform in time. The spectrum shows four prominent partials.

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 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.

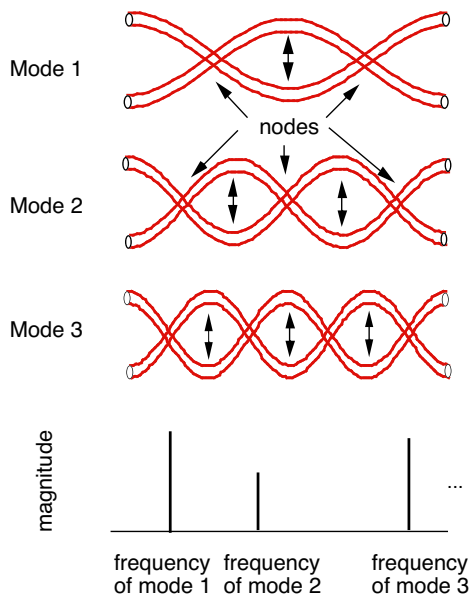


Fig. 2.8. The first three bending modes of an ideal metal bar and its spectrum. The size of the motion is proportional to the amplitude of the sound, and the rate of oscillation determines the frequency. As usual, the spectrum shows the frequencies of the partials on the horizontal axis and their magnitude 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).

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 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.

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.

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.

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 listener assigns a separate

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

¹³ See [B: 91].

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:

dull \longleftrightarrow sharp
 cold \longleftrightarrow warm
 soft \longleftrightarrow hard
 pure \longleftrightarrow rich
 compact \longleftrightarrow scattered
 full \longleftrightarrow empty
 static \longleftrightarrow dynamic
 colorful \longleftrightarrow colorless

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

¹⁴ As usual, we ignore the phase spectrum.

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

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 jurors 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 studies¹⁶ 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 fluctuation¹⁸
- (iii) Presence (or absence) of high-frequency, inharmonic energy in the attack
- (iv) Bandwidth of the signal¹⁹
- (v) Balance of energy in low versus high partials
- (vi) Existence of formants²⁰

For example, Grey and Gordon [B: 63] exchange the spectral envelopes²¹ 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

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.

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.

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 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.

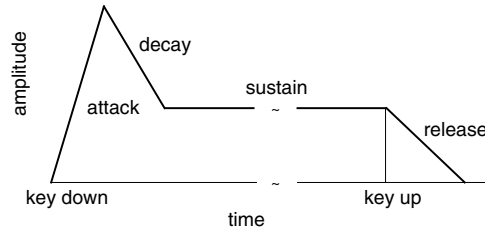


Fig. 2.9. The ADSR envelope defines a loudness contour for a synthesized sound. The attack is triggered by the key press. After a specified time, the sound decays to its sustain level, which is maintained until the key is raised. Then the loudness dies away at a rate determined by the release parameters.

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.

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 Canada²³ 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

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.

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

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.

2.4.1 Pitch of Harmonic Sounds

Pythagoras of Samos²⁵ 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 ratios²⁶: 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.

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

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.

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 high²⁷), 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. 18 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 $f, 2f, 3f, \dots$, 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 .

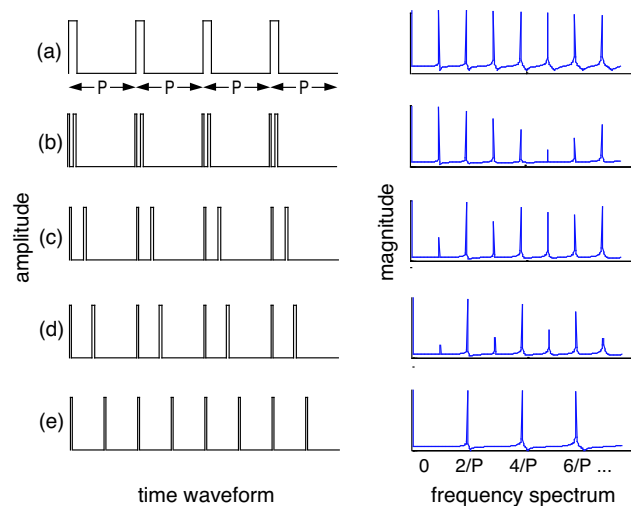


Fig. 2.10. (a) and (b) have the same period P and the same pitch. (c) and (d) change continuously into (e), which has period $\frac{P}{2}$. Thus, (e) is perceived an octave higher than (a). The spectra (shown on the right) also change smoothly from (a) to (e). Where exactly does the pitch change? See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

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

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 (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 “Westminster 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.

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.

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

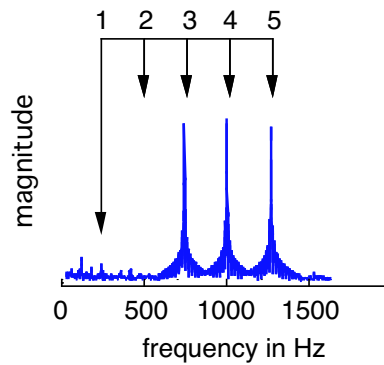


Fig. 2.11. Spectrum of a sound with prominent partials at 780, 1040, and 1300 Hz. These are marked by the arrows as the third, fourth, and fifth partials of a “missing” or “virtual” fundamental at 260 Hz. The ear perceives a note at 260 Hz, which is indicated by the extended arrow. See video example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

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.

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

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.

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 per-



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.

ceived 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, 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

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

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.