

## Chapter Seven

# Psychoacoustics and cognition for musicians

Richard Parncutt

Our experience of pitch, timing, loudness, and timbre in music depends in complex ways on physical measurements of frequency, time, amplitude, spectral envelope, and temporal envelope. Psychoacousticians, who investigate these dependencies, can shed light on basic musical questions. Psychoacoustics can explain the central role of frequency and time in music theory and notation, variations in the perceived consonance and frequency of occurrence of pitch–time structures, differences in consonance between simultaneous and successive intervals, why some structures are commoner than others, harmony and voice-leading conventions (including why parallels are avoided, why a leap is often followed by a step in the other direction, and why the interval between tenor and bass usually exceeds that between soprano and alto). In music performance, psychoacoustics can explain why musical intonation deviates from “pure” intervals, how we can recognise mistuned intervals, and why synthesising musical instrument sounds is so difficult.

Psychoacoustics is a subdiscipline of psychophysics dedicated to sound and hearing. Its experiential parameters include the pitch, loudness and timbre of musical tones. Experiential sound parameters depend on relevant physical parameters. Pitch differs from frequency, loudness differs from sound pressure level (SPL), and timbre differs from frequency spectra or temporal envelope. A pure tone’s pitch depends on its intensity (pitch shift)—it sounds slightly sharp or flat at extremes of intensity. The clarity of a short pure tone’s pitch depends on its duration; the more periods in the waveform, the clearer the pitch. In general, the exact pitches and saliencies (clarity) of the tone sensations within a complex sound depend on the frequencies and SPLs of all audible partials.

### **Is psychoacoustics relevant?**

Although psychoacoustics may seem unfamiliar, its musical relevance is obvious. If I play a CD and exclaim “Listen to this!”, I implicitly reference three different representations of music:

- The physical signal travelling from a loudspeaker to our ears
- Subjective experience—patterns of sensation and emotion
- Human culture (or knowledge)

Physics, experience and culture may be considered distinct but related representations of the phenomenon called “music”. The cultural meaning of

music depends on the relationship between the physical signal and our private experience of it. Psychoacoustics scrutinises such relationships.

Why is psychoacoustics not considered central to musical practice and musicianship study? While composition, performance, musicology, acoustics, and music theory help us to understand music from different perspectives, only the psychology and philosophy of music directly analyse musical experience, and only psychoacoustics systematically addresses the experience of musical structure.

### **Psychoacoustic perspectives on musical questions**

*Why do we usually experience a musical tone as a single entity, although the ear discriminates several partials?*

The ear is often presented with complex sounds comprising many partials. There is no simple relationship between the number of tones in the signal (physical) and the number heard (experiential). We must consider our interaction with our environment, where we experience important objects (an item of food, etc.) as unitary, corresponding to their functions for us (Gibson 1979), regardless of the complex structure of the visual, acoustic and other signals produced. Our experience is oriented toward sound sources—not the sounds themselves.

*Why do frequency and time (pitch and rhythm) dominate Western musical notation?*

While frequency, time and loudness are clearly defined and one-dimensional, timbre's definition is unclear. The perceptually important dimensions of timbre include brightness, roughness, and spectral flux (McAdams et al. 1995). This complexity effectively eliminates timbre as a potential axis in a graphic representation of music. Although loudness could be an axis label, it usually is not. Only two dimensions are available on paper, and we are more sensitive to small changes in pitch and rhythm: musicians adjust frequencies (intonation) and timing (rhythm) more precisely than dynamics and timbre. It is also easier to create cognitive hierarchies in pitch and time. Examples include major–minor tonality—with the tonic at the top of the hierarchy, then the perfect 5th outlined by tonic and dominant, the tones of the tonic triad, the diatonic tones and the chromatic tones (Lerdahl 2001); and meters such as 2/4—a superposition of isochronous rhythms or pulses at different hierarchical levels (Parncutt 1994).

*Why do some combinations of musical tones blend better than others?*

The consonance of simultaneous intervals in Western music depends primarily on three psychological factors: smoothness (lack of roughness), harmonic and familiarity. Roughness is the subjective impression of a

listener in a laboratory setting, and it has a physiological basis in the frequency analysis performed by the basilar membrane (Plomp and Levelt 1965). Harmonicity is the degree to which a sound's spectrum resembles a harmonic series, and is related to the phenomenon of perceptual fusion (Stumpf 1883, 1890; Parncutt and Hair 2011). Superposed on these psychoacoustic effects is the independent psychological effect of familiarity: we tend to prefer musical sounds and styles that we hear more often (Cazden 1945).

*Why don't cheap synthesisers sound like the original instruments, even when they correctly reproduce the relative amplitudes of the harmonics?*

The timbre of a musical tone does not depend primarily on the relative amplitudes of the partials (the spectral envelope). It also depends on the temporal envelope, which includes the amount of noise in the onset part of the tone (e.g. the scratchy noise at the start of a violin tone before the string starts to regularly vibrate); and the interaction between the temporal and spectral envelope, such as the quasi-independent amplitude trajectories of the harmonics of a trumpet tone. Moreover, these complex physical patterns vary from one tone in a musical scale to the next. The ear is remarkably sensitive to these details, so natural instrument sounds must be precisely imitated before listeners accept them as realistic. One approach to simulating musical instrument timbres is physical modelling of the instrument's underlying physics, including the interaction with the human performer.

*How is it possible to know that a pianist on a recording is playing loudly even when the recording is played quietly?*

The amplitudes of the partials of a musical tone tend to fall off at higher frequencies, and the slope of the spectral envelope depends on how loudly the instrument is played. Musical instruments played loudly generally produce more energy at high frequencies relative to low frequencies. Consider a piano tone. If the key is depressed quickly, the piano string will be bent more sharply by the hammer, and the resultant waveform (graph of sound pressure against time) will have sharper peaks, increasing the amplitude of high-frequency partials. This changes the timbre, which allows listeners to guess how loudly the piano was played, independent of the playback volume.

*How do we know what tones the members of a choir are intending to sing in an unfamiliar piece when they are singing seriously out of tune?*

Experiential scales such as color (in vision) and pitch (in hearing) can be perceived either continuously or categorically. In categorical perception, different shades of red are perceived as red, different intonations of C# are perceived as C#, and different tunings of a major third interval are

perceived as major thirds. Categorical perception allows listeners to accommodate mistunings in musical performances of as much as a quartertone or even a semitone, depending on context.

### **Frequency analysis in hearing**

How does the ear separate frequencies within musical tones and chords? Can this tell us something about musical structure? Can it explain why simultaneous intervals of 1 or 2 semitones (m2, M2) sound dissonant, and are therefore less common in Western music than 3 or 4 semitones (m3s and M3s)?

The cochlea is part of the inner ear—a tunnel inside the skull. It has two main functions: to transform sound from physical to neural signals for mental processing, and to perform a running frequency analysis of incoming sound. The signals sent from ear to brain are already analysed into different frequency ranges, and this auditory spectrum changes continuously as incoming sounds change. Each auditory nerve fibre connects to a hair cell on the basilar membrane within the cochlea. That hair cell is sensitive to a limited range of frequencies, whose centre is the cell's characteristic frequency and whose width (critical bandwidth) is like that of a bandpass filter. A hair cell whose critical frequency is 500 Hz responds most strongly to partials from approximately 460–540 Hz (Moore 2003), and more weakly to partials outside that range.

### **Evolution, environmental acoustics and uncertainty**

Auditory physiology is the result of prolonged biological evolution, during which organisms that responded better to their acoustical environment were more likely to survive. A complete explanation of musical dissonance (e.g. the avoidance of simultaneous second intervals in tonal music) begins with a consideration of the way prehumans (hominids and their mammal ancestors) interacted with their acoustical environments.

The probability that prehumans (or any other animals) survived long enough to reproduce partly depended on their ability to recognise objects on the basis of sound in an environment full of acoustic reflectors. Every sound reaching the ear is a mixture of direct and reflected sound, and reflections come from a variety of environmental objects.

Imagine talking to someone 5 metres away next to a cliff-face. You hear the direct sound of your own voice almost instantaneously, and the reflected sound about 30 milliseconds later ( $10\text{m}/330\text{ms}^{-1}$ ). You do not perceive someone 5 metres inside the wall—the ear ignores the reflection. The ear generally integrates sounds occurring within roughly 40–50 ms (a “temporal window”, Haas 1951). The listener is aware only of the earlier sound, a phenomenon known as the precedence effect. An “echo” can only be heard separately when it begins more than about 50 ms after the original sound, the exact delay depending on frequency (Terhardt 1998). The shorter the

temporal window, the more accurately we hear rhythms, but the less accurately we perceive frequencies. This trade-off, called the uncertainty principle, implies that there is no such thing as “the” frequency analysis of a sound.

When one sound makes another less audible, psychoacousticians speak of masking. Masking may be simultaneous (two simultaneous sounds mask each other) or successive. In simultaneous masking, one sound drowns out the other or the two seem to fuse—for example, when two pure tones lie within the same critical bandwidth. With successive masking, fast rhythmic events (approaching 50 ms apart) cannot be clearly distinguished; the first masks the second (forward masking) and the second masks the first (backward masking). If a musician wants to play a dotted rhythm very fast without the short note becoming inaudible, he or she must reduce the sharpness of the ratio (e.g. from 3:1 to 2:1). As investigations have demonstrated, musicians do just that (Friberg and Sundström 2002).

### **Auditory processing**

Sound is processed both physiologically (in the auditory periphery) and cognitively (in the brain). Both stages contribute to the auditory system’s recognition and assessment of sound sources. Apart from collecting sound and encoding it as neural firing patterns, the ear’s most important physiological function is running frequency analysis, without which recognising sound sources would be impossible. This analysis divides sound into frequency bands, from which the brain derives different sonic properties.

#### *Frequency*

Successive pure tones that are identical except for their frequency can be distinguished if the difference exceeds about 1/10 semitone. This just-noticeable difference is smaller at medium frequencies (where the ear is most sensitive, from about 300–2000 Hz), and larger for low or high frequencies. Since there are about 10 octaves or 120 semitones in hearing range, we can distinguish approximately 1000 frequencies.

#### *Amplitude*

Successive sounds that are identical except for their amplitude can be distinguished if the difference exceeds about 1 dB; again, the just-noticeable difference is smaller for medium amplitudes and larger for loud or soft sounds. We can hear intensity differences from  $10^{-12}$  Wm<sup>-2</sup> (0 dB SPL) to 1 Wm<sup>-2</sup> (120 dB SPL) under ideal conditions, so the ear can perceive approximately 100 different intensities. If the number of just-noticeable differences is a measure of sensitivity to a given parameter, the ear is about ten times more sensitive to frequency than to the amplitude of an isolated pure tone. This observation is consistent with musical notation: there are

about eight dynamic levels in common use (ppp to fff) and about eighty pitches (the 88 keys of the modern piano).

### *Timbre*

Timbre depends on a sound's temporal and spectral envelope. The importance of temporal envelope is often underestimated. Consider a recording of piano music. It still sounds piano-like when the spectral amplitude envelope is radically changed (e.g. using a graphic equaliser). But when the recording is played backwards, one hears a spooky organ-like sound which starts quietly and gradually gets louder, the volume accelerating to a peak at the end of each tone (where the hammer hit the string in the original recording).

### *Phase*

The addition of reflected sound to direct sound changes the waveform's shape so much that it would be difficult to distinguish timbres, and hence sound sources, from it alone. The proportion of incident energy absorbed by each environmental reflector depends on the frequency of each individual partial, so reflection and superposition radically and unpredictably change the phase relations among the partials. The auditory system has therefore evolved to be insensitive to monaural phase relationships (i.e. among partials picked up by one ear), focusing instead on other sound parameters (Terhardt 1988).

The ear is nevertheless sensitive to monaural phase relationships among partials in the attack or transient portions of a tone (Moore 2003), because in everyday acoustic environments only the attack portion of a sound can be perceived without interference from reflected sound, which arrives later. In this way, phase relationships can affect the pitch and timbre of piano bass tones (Galembi et al. 2001). The ear is also very sensitive to binaural phase relationships, due to their role in sound localisation.

### **Gestalt principles**

The auditory scene (Bregman 1993) represents the output of the first, physiological stage of sound processing, which becomes the input to the second, cognitive stage. It is visualised as a sonogram—a graph of frequency against time where the frequencies of individual partials (spectral frequencies, not fundamental frequencies) or noise bands are on the vertical axis, and the times at which they begin and end are on the horizontal axis. The brain reconstructs a sound's source by analysing the auditory scene, just as it reconstructs physical objects by analysing the visual scene.

We visually recognise objects by applying gestalt principles such as proximity and similarity. The gestalt principles are similar for hearing, because the boundaries of a visible object are analogous to the frequency–time trajectories of its audible partials (Terhardt 1998).

However, there are important differences between seeing and hearing. Consider the Gestalt principle of good continuation. Stepwise melodic motion in music tends to continue in the same direction, which listeners expect (Narmour 1990; Huron 2006). But listeners also expect a large leap to be followed by a step in the opposite direction—especially when the leap approaches the top or bottom of the melody’s tessitura (Huron 2001)—which contradicts good continuation. In the real world, sounds or partials that gradually rise in frequency soon reach a maximum that depends on physical parameters (size, density, tension...) of the sound-producing mechanism. Unable to go further, they must turn back. In another contradiction of good continuation, polyphonic musical parts do not normally cross over. If they did, it would be hard to hear one part rising or falling through the other, because the principle of proximity dominates the principle of good continuation at the intersection point.

### Psychoacoustics and music theory

Psychoacoustics can be used to explore links between everyday nonmusical sounds, the physiological systems humans have developed to perceive them, and musical sounds. In this way, we can explain the origins of familiar musical sound patterns. Much of what we perceive is based on our experience of the auditory world; the ear learns arbitrary sound patterns after repeated exposure to them.

| <b>Domain</b> | <b>Environmental source</b> | <b>Type of pattern</b> | <b>Typical values</b>  | <b>Musical manifestation</b> |
|---------------|-----------------------------|------------------------|------------------------|------------------------------|
| pitch         | voice                       | harmonic               | consonance = P8, P5... | Western harmony, tonality    |
| pitch         | voice                       | melodic                | step = M2              | melody                       |
| rhythm        | footsteps, heartbeat        | regular                | beat = 600 ms          | tempo, ritardando            |
| rhythm        | speech                      | irregular              | tone = 250 ms          | phrase, articulation         |

Table 1. Physiological bases of pitch–time structures in Western music

Table 1 presents some auditory universals and their correlates in Western music. Consider the first row, “pitch”. The human voice produces harmonic complex tones (voiced speech sounds). Western music theorists since Rameau have regarded the intervals among the lower partials (octaves, perfect 5ths, etc.) as the foundation of harmony and tonality. Psychoacoustic approaches to harmony since Helmholtz (1863) have considered interactions among the harmonics of tones in musical simultaneities, treating simultaneous and successive tones differently, and

separately considering percepts such as roughness and fusion. This aspect of Western harmony and tonality therefore has a psychophysical basis.

The second row of table 1 addresses pitch intervals between successive syllables in speech and between successive tones in melodies. The major 2nd is the most common successive interval across cultures—consistently more common than the minor 2nd (Vos and Troost 1989). Speech is different: larger intervals are more infrequent, so the minor 2nd category ( $\pm 1$  quarter tone) is more common than the major 2nd (Tierney et al. 2009). The difference involves categorical perception; a melodic minor 2nd in music is so small that the tones may be assigned to the same scale step, whereas in speech, the interval categories are less clearly defined.

Regarding the third row of table 1, a moderate musical tempo corresponds to about 100 beats per minute, or 600 ms (milliseconds) per beat. That corresponds to a typical heart rate (during moderate activity), or the interval between footfalls when walking moderately fast (Parncutt 1994).

The final row considers the average duration of a musical tone and the average inter-onset time of a speech syllable. The latter is 200–250 ms (we normally articulate 4–5 syllables per second), while musical tones are typically slightly longer, because music and speech have different functions: speech focuses on lexical communication, whereas music focuses on creating or altering emotional states.

### **Nature versus nurture**

The physiology of hearing is largely innate and almost identical for all individuals and cultures. The psychoacoustics of pitch, loudness and roughness is essentially culture-independent, being strongly coupled to auditory physiology. However, several interesting aspects of psychoacoustics are learned through environmental interaction.

Ecological approaches to music theory foreground environmental interaction: our ears learn by exposure to the pitch–time patterns of specific musical styles or the timbre of specific musical instruments. The auditory system may acquire information about the intervals in the harmonic series from environmental sounds before we can perceive the pitch at the fundamental (Terhardt 1988). Similarly, our perception of timbre depends on the timbres to which we have been exposed.

Is there a limit to the musical structures humans can understand and respond to? Opinions differ. Twentieth century modernist composers often assumed not—yet their failure to capture the public imagination suggests otherwise. Our world still generates an enormous diversity of musical styles across different societies. The degree to which humans can understand and respond to musical structures therefore appears partially limited by a mixture of physiological constraints (“nature”) and environmental



constraints (“nurture”). We may not respond to arbitrary sound patterns, but our physiology allows for remarkable diversity.

### Conclusion

Psychoacoustics explain a wide range of musical phenomena, and psychoacoustic theory can be learned without extensive training in the foundations of mathematics, physics and physiology. Compared with the time and energy required to learn a musical instrument, the fundamentals of psychoacoustics are easily mastered, suggesting that more psychoacoustic material should therefore be included in musicianship curricula.

### References

- Bregman, Albert S. 1993. “Auditory Scene Analysis: Hearing in Complex Environments.” In *Thinking in Sound: The Cognitive Psychology of Human Audition*, edited by Stephen McAdams and Emmanuel Bigand, 10–36. Oxford: Oxford University Press.
- Cazden, Norman. 1945. “Musical Consonance and Dissonance: A Cultural Criterion.” *Journal of Aesthetics and Art Criticism* 4(1):3–11.
- Friberg, Anders, and Andreas Sundström. 2002. “Swing Ratios and Ensemble Timing in Jazz Performance: Evidence for a Common Rhythmic Pattern.” *Music Perception* 19:333–49.
- Galembo, Alexander, Anders Askenfelt, Lola L. Cuddy, and Frank A. Russo. 2001. “Effects of Relative Phases on Pitch and Timbre in the Piano Bass Range.” *Journal of the Acoustical Society of America* 110:1649–66.
- Gibson, James J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Haas, Helmut. 1951. “On the Influence of a Single Echo on the Intelligibility of Speech.” *Acustica* 1:48–58.
- Helmholtz, Hermann von. 1863. *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. Brunswick: Vieweg.
- Huron, David. 2001. “Tone and Voice: A Derivation of the Rules of Voice-Leading from Perceptual Principles.” *Music Perception* 19:1–64.
- Huron, David. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: The MIT Press.
- Lerdahl, Fred. 2001. *Tonal Pitch Space*. New York: Oxford University Press.
- McAdams Stephen, Suzanne Winsberg, Sophie Donnadieu, Geert De Soete and Jochen Krimphoff. 1995. “Perceptual Scaling of Synthesized Musical Timbres: Common Dimensions, Specificities, and Latent Subject Classes.” *Psychological Research* 58:177–192.

- Moore, Brian C. J. 2003. *An Introduction to the Psychology of Hearing*, 5th edition. New York: Academic Press.
- Narmour, Eugene. 1990. *The Analysis and Cognition of Basic Melodic Structures*. Chicago: University of Chicago Press.
- Parncutt, Richard. 1994. "A Perceptual Model of Pulse Salience and Metrical Accent in Musical Rhythms." *Music Perception* 11:409–64.
- Parncutt, Richard and Graham Hair. 2011. "Consonance and Dissonance in Theory and Psychology: Disentangling Dissonant Dichotomies." *Journal of Interdisciplinary Music Studies*, 5:119–166.
- Plomp, Rene and Willem J. M. Levelt. 1965. "Tonal Consonance and Critical Bandwidth." *Journal of the Acoustical Society of America* 38:548–60.
- Popper, Karl R. and John C. Eccles. 1977. *The Self and its Brain*. Berlin: Springer.
- Stumpf, Carl. 1883 (Vol. 1) and 1890 (Vol. 2). *Tonpsychologie*. Leipzig: Hirzel.
- Terhardt, Ernst. 1988. "Psychophysikalische Grundlagen der Beurteilung musikalischer Klänge." In *Qualitätsaspekte bei Musikinstrumenten*, edited by Jürgen Meyer, 1–15. Celle: Moeck.
- Terhardt, Ernst. 1998. *Akustische Kommunikation*. Berlin: Springer.
- Tierney, Adam T., Frank A. Russo, and Aniruddh D. Patel. 2009. "Empirical Comparison of Pitch Patterns in Music, Speech and Birdsong." In *Proceedings of Acoustics '08*, 4723–28. Paris: Société Française d'Acoustique.
- Vos, Piet G. and Jim M. Troost. 1989. "Ascending and Descending Melodic Intervals: Statistical Findings and their Perceptual Relevance." *Music Perception* 6:383–96.