

Psychoacoustics and music perception

Richard Parncutt

Submitted to Bruhn, Kopiez, Lehmann, Oerter (2005?):
Musikpsychologie – das neue Handbuch

Stand: 03.06.2004

Psychoacoustics is commonly understood as the systematic, quantitative, empirical study of relationships between physical sound parameters and corresponding experiences, or between objective and subjective descriptions of sound.

If experiences are considered as reflections of underlying neural processes, psychoacoustics may be conceived of as involving relationships between sound and corresponding neural processes. This approach will be avoided in the present chapter for the following reasons. First, the focus here is on music, which can hardly be regarded as such (i.e. as art) if it is not *experienced*; so it is appropriate to regard experience as fundamental to musical psychoacoustics. Second, although neuropsychological research plays an increasingly important role in modern psychoacoustics, the connection between specific neurological mechanisms and (musical or other) experiences often remains unclear. In other words, the philosophical mind-body problem remains essentially unsolved. It may be avoided by restricting the investigation to acoustical measurements and subjective reports. Third, the neuropsychology of music perception is an important, rapidly growing field that is covered elsewhere in this volume. Here, the focus is on traditional psychoacoustical studies that are based on empirical data collected from listeners' descriptions of their experience.

What is the musical relevance of psychoacoustics? That is a central question, for if psychoacoustics were musically irrelevant, it would have no place in this volume. I will argue that psychoacoustics has both specific and general relevance, and that it has considerable, largely unrealized potential to contribute to musical and musicological discourse. An example of the specific relevance of psychoacoustics is the application of research on musical timbre in computer-assisted composition. For many years, academic musical forums such as the *Computer Music Journal* and the *International Computer Music Conference* have featured musical works by composers versed in the psychoacoustics of timbre perception. This body of knowledge played an important role in the emancipation of timbre in the compositional practice of the late twentieth century (e.g. at Ircam in Paris). A more specific example is work of the Canadian composer Sean Ferguson (2001), who incorporated the principles of pitch perception in music and speech originally developed by Terhardt (1998) and Parncutt (1989) into a computer-implemented compositional algorithm. Several works composed in this way have been performed by professional orchestras.

The general relevance of psychoacoustics for musicology may be seen from its position within musicology's underlying interdisciplinary structure, as a link not only between music acoustics and the psychology of music, but also between both these and music theory. These connections can be clarified by Karl Popper's concept of the "three worlds" (Popper & Eccles, 1977; Terhardt, 1998). Popper's World 1 is the physical world; in musical acoustics, it includes tones and noises (understood as

variations in air pressure) and their frequencies, sound pressure amplitudes or levels, phases, spectra, and so on. World 2 is the world of (raw, direct) experience and includes the experiential (perceived, subjective) parameters of pitch, loudness, timbre and apparent duration of musical tones. World 3 is the world of knowledge and information; for our purposes, it includes elements of music notation such as note names and note values, musical scores and works, computer programs, and any or all knowledge (e.g. about music, or about psychoacoustics). For example, this chapter belongs to World 3, but the paper on which it is printed and the computers in which it is stored belong to World 1, and the feelings experienced on reading it belong to World 2.

Frequency analysis

As a further example of the musical relevance of psychoacoustics, consider the effect of the ear's ability to separate simultaneous frequencies on the structure of typical sonorities in western music. Why, for example, are harmonic intervals of one or two semitones (minor and major seconds) less common than intervals of three or four semitones (minor and major thirds)?

The psychoacoustical term for the ear's ability to perceptually separate simultaneous pure tones is *critical bandwidth* or CBW. Among other things, CBW determines the number of audible partials¹ in harmonic complex tones; Plomp (1964) found this number to be about seven for harmonic complex tones with partials of equal amplitudes in the central range. Since the partials of real harmonic complex tones in music and speech usually do not have equal amplitudes, the number of audible partials varies considerably.

CBW has been defined and explained in several different ways. The most direct definition relates it to the peripheral physiology of pitch perception. The inner ear or cochlea has two main functions: to transform sound from a physical into a neural signal for processing by the brain, and to perform a *running frequency analysis* of the incoming sound. The signals sent from the ear to the brain are thus already analyzed into different frequency ranges, and this *auditory spectrum* changes continuously as incoming sounds change (hence *running*). Each fiber in the auditory nerve is connected to a hair cell on the basilar membrane within the cochlea. That hair cell is sensitive to a limited range of frequencies, whose center is the cell's *characteristic frequency* and whose width is the CBW at that frequency. Thus, the basilar membrane has a *tonotopic* (frequency-place) structure (Ohm's "place theory" of pitch). For example, a hair cell whose critical frequency is 500 Hz responds significantly to partials in the approximate range 460 Hz and 540 Hz, because CBW at 500 Hz is approximately 80 Hz (Moore, 2003). This hair cell also responds to frequencies outside that range, but the level of response gets weaker, the farther the incoming frequency lies from the critical frequency.

CBW values vary according to the operational definition of CBW, and different researchers make different assumptions regarding the shape of the filter and the cut-off point on each side of it (Moore, 2003). At lower frequencies, CBW is almost constant when measured in Hz or linear frequency and lies between 50 and 100 Hz.

¹ Other terms for "partial" in the psychoacoustic literature are "partial tones", "frequency components", "pure tone components" and simply "components".

At higher frequencies, CBW is almost constant when measured in semitones or log frequency and lies between 2 and 3 semitones.

The auditory physiology underlying CBW is the result of a long period of evolution, during which organisms that were better tuned to their acoustical environment were more likely to survive. A complete answer to the question posed above about the avoidance of harmonic second intervals in tonal music may therefore begin with a consideration of the way human ancestors (hominids and their mammal ancestors – hereafter called "prehumans") interacted acoustically with their environments.

The probability that prehumans would survive long enough to successfully reproduce depended in part on their ability to recognize objects on the basis of the sounds that they made. Every sound that reaches the ear is a mixture of direct and reflected sound. Reflections come from environmental objects of all kinds, including the ground. In typical (pre-) human acoustical environments, the effect of reflected sound is significant when a reflecting object is close by – say, within about 5 meters. Sound reflected from such an object reaches the ear about 30 ms (or $10 \text{ m} / 330 \text{ ms}^{-1}$) later. It is important to ignore reflected sound, for if reflected sound is mistaken for direct sound, a non-existent sound source will be perceived on the far side of the reflecting object. This suggests that the ability to ignore reflections is an important prerequisite for successful interaction with the acoustical world. The ear solves this problem by integrating the sounds occurring within roughly 40-50 ms of each other, i.e. by suppressing the echo (Haas, 1951). The listener is aware only of the earlier sound (*precedence effect*). An echo is only heard as such – i.e., distinct from the original sound – when it begins more about 50 ms after the original sound. In other words, the ear analyses incoming sounds within a *temporal window* of about 50 ms. The exact duration of the window depends on how it is operationally defined, and on its center frequency (Terhardt, 1998, Fig. 9.17, p. 255).

According to the *uncertainty principle* of frequency analysis² (Gabor, 1947), the shorter the temporal analysis window, the less accurately one can know the frequencies of tones within it. In a 50 ms window, for example, one cannot measure frequencies within less than about $\pm (1/50 \text{ ms})$ or 20 Hz.³ This is the main reason for the following three familiar phenomena. First, isolated pure tones below about 20 Hz do not evoke pitch. Second, simultaneous tones in the central musical range cannot be clearly distinguished unless they are separated by considerably more than a semitone (a semitone at 300 Hz corresponds to about 20 Hz). Third, in the central pitch range of tonal music, harmonic intervals of 1 or 2 semitones are avoided: the perception of *roughness* is associated with the ear's inability to clearly separate simultaneous tones that are close in frequency. A full explanation of each of these three phenomena is considerably more complex than this preliminary explanation; the duration of the ear's analysis window may nevertheless be regarded as the main or most parsimonious explanation.

Much of psychoacoustics has been concerned with the simple question of what is audible and what is not. In other words, what physical aspects of a sound have an effect on our experience of sound? The answer to this question depends to a large

² Here, "frequency analysis" means the same as "Fourier analysis" or "spectrum analysis" – the separation of a complex sound into its partials.

³ The mathematics is deliberately simplified. Depending on the exact operational definition of uncertainty of frequency, an additional (multiplicative) factor will be involved.

extent on the uncertainty principle, which is an unavoidable physical constraint on every kind of frequency analysis. In quantum mechanics, the uncertainty principle states that you cannot know the position and time coordinates of a small particle at the same time; the more precisely you know the position, the less precisely you know the time at which the particle is at that position, and vice-versa. In sound, and within the paradigm of classical mechanics, the uncertainty principle says that the more accurately you know the frequency of a partial, the less accurately you know when it starts and ends; a frequency can only be *exactly* determined if it has *infinite* duration. It follows that there is no such thing as *the* frequency analysis of a sound. The outcome of a frequency analysis routine always depends on arbitrary choices that determine the accuracy of temporal information by comparison to spectral information; the more accurate the temporal information, the less accurate the spectral information, and vice-versa. This trade-off is conceptualized in terms of the temporal window within which the analysis is assumed to take place. The longer the window, the smaller the critical bandwidth (CBW) – that is, more accurately the system can determine the partial frequencies of a steady-state sound.

The simplest and most direct way to measure the frequency-analysis characteristics of the auditory system is the classical masking experiment. Egan and Hake (1950) presented a constant pure tone and then switched on and off another pure tone whose level was adjusted by listeners until it was almost inaudible. This was exactly the same routine as the method of determining the threshold of hearing in quiet, except that the pure test tone was not presented against a silent background but against another pure tone. The resultant curve is called the *masked threshold*. It peaks at the frequency of the constant masking tone or "masker" and approaches the threshold in quiet at large distances from this frequency. Critical bandwidth (CBW) may be regarded as a measure of the width of this peak.

Categorical perception

Popper's World 1 is continuous – at least, if we assume that the physical world is Newtonian and can be accounted for by the principles of classical mechanics. This assumption is appropriate as long as we are talking about the direct interaction between humans and their everyday environment. World 2 is similarly assumed to be continuous. World 3 is made up of *discrete* units of information, and in this sense is different from Worlds 1 and 2.

The relationship between World 3 and the other worlds involves the psychological concept of *categorical perception* (Burns, 1999). As an example, consider the word "hat". This word can be uttered in many different ways. The /a/ can sound more or less similar to other vowels such as /i/ or (e/), more or less nasal, and longer or shorter. The /h/ can be stronger or weaker, or more or less guttural. The /t/ can be harder or softer, or more or less delayed. All these parameters depend on a range of factors including the speaker's regional accent or dialect, the word's syntactic and semantic context, and the physical and social context of the utterance. But the word is recognizable in spite the many possible colorations and contexts. Recognition immediately triggers a wide-ranging network of associations: thoughts about different kinds and functions of hats, the kinds of people who wear certain kinds of hats, related words and their lexical meanings, and so on. Comprehension may be said to have occurred as soon as (correct) associations begin to be triggered.

In Popper's approach, the process of understanding speech can be broken down into two different transitions, one from World 1 (the utterance as physical phenomenon – as changes in air pressure) to World 2 (the listener's raw experience of the sound, e.g. when hearing a foreign language and understanding nothing), and the other from World 2 to World 3 (meaning). It is also possible – both conceptually and empirically – to ignore the middle stage and regard the understanding of speech as a direct transition from World 1 to World 3; this corresponds to Gibson's (1979) concepts of *direct perception* and *ecological psychology*. Gibson regarded the experience of hearing a word as a byproduct of the direct process of extraction of information or meaning from the physical signal. All these considerations apply equally well to music as they do to speech. In a cognitive approach, the perception and meaning of music may be understood in two stages – from World 1 to World 2, then from 2 to 3. In an ecological approach, the perception and meaning of music may be understood to involve a direct transition from World 1 to World 3. In a cognitive approach, experiences mediate perception; in an ecological approach, they are mere byproducts of perception.

insert Figure WORLDS

Plan of figure:

as points of a triangle: World 1 physics, World 2 experience, World 3 information

2-way arrows between each pair of worlds: classical psychophysics, direct categorical perception, indirect categorical perception

Caption: A broad view of psychophysics based on Popper's three "worlds"

Research on categorical perception investigates the physical or experiential parameters corresponding to a given meaning. Where is the middle of the category, where are the category boundaries, and how do these depend on context? For example, what range of fundamental frequencies corresponds to the note F#4 in a performance of a classical string quartet movement in D major? Whereas classical psychoacoustics investigates quantitative relationships between Worlds 1 and 2, research on categorical perception investigates quantitative relationships between World 3 and the other two worlds (figure WORLDS). A broader definition of psychoacoustics might include all these kinds of research, i.e. all empirically examinable, quantitative relationships between all three worlds.

It is not possible to prove the existence or fundamental nature of Popper's three worlds. The advantage of the three-world approach is that it clarifies complex problems. The worlds can be regarded either as different kinds of reality, or as different ways of looking at a single reality. If we nevertheless assume that Popper's three worlds are fundamental, we may regard music either as a physical phenomenon (complex changes in air pressure), as a psychological phenomenon (a complex series of experiences), or as a form of information or knowledge (the score as a series of instructions from a composer to a listener, or collective knowledge about the structure of a piece of music as a kind of virtual score that is carried around in the heads of people who know the piece). In Popper's approach, all these conceptualizations of music are valid. Musical psychoacoustics may then be regarded either as a study of relationships between these different musical realities, or as a study of relationships between these different ways of looking at music. This definition of psychoacoustics is much broader than usual and suggests that a

psychoacoustical approach has considerable potential to bring together different aspects of musical scholarship that usually exist independently.

Classical psychoacoustics

The traditional domain of psychoacoustics is limited to *quantitative* (rather than qualitative) relationships between physical and experiential aspects of sound, and to *sensations* (rather than thoughts and emotions). Quantitative relationships involve numbers and are determined by measurement; such data can be analyzed statistically. The main perceptual parameters of a sound that are investigated quantitatively in psychoacoustics are pitch, loudness, timbre, and apparent duration (also called perceived or subjective duration). At first glance one might expect each perceptual parameter to depend only on its physical correlate. For example, one might expect pitch to depend only on frequency in Hertz, loudness only on sound pressure level in decibels (dB) and timbre only on spectral envelope, that is, on the relative amplitudes of the partials of a sound. In fact, each perceptual parameter depends in some way on *all* physical parameters, and it is one of the basic tasks of classical psychoacoustics to explore and measure these dependencies. For example, the pitch of a pure tone depends primarily on its frequency, but can also be affected by its sound pressure level and by the presence of other, simultaneous tones or noises (pitch shift); and timbre depends strongly on the temporal envelope as well as the spectral envelope of a sound. Concise accounts of the dependency of pitch, loudness, timbre and apparent duration on associated physical variables are presented together with instructive auditory illustrations by Houtsma et al. (1987).

Pitch

Pitch is defined as the perceived height of a sound and can be empirically quantified by presenting a listener with a sound (sometimes called the "experimental sound") followed by a pure tone (the "reference tone"), the two being separated by a short silence. The duration of both the experimental sound and the reference tone should be typical of pitched sounds in speech, music and everyday environments (say, 200-300 ms, since the average rate of speech phonemes is about 4 per second). To avoid memory loss, the silence between the two sounds in each experimental trial should also be quite short. The listener hears the sound-tone pair several times; between repetitions, s/he adjusts the frequency of the tone until its pitch is the same as the pitch of the sound. The frequency of the pure tone in Hertz is then taken as a measure of the pitch of the sound.

The surprising thing about this procedure is that the frequency of the reference tone often deviates systematically from all the frequencies that are physically present in the experimental sound. Consider the case of a steady-state complex sound comprising several partials with different frequencies and amplitudes. One might expect that the pitch of such a sound would correspond to the frequency of one of the partials – and sometimes it does. For example, the pitch of a harmonic complex tone⁴ usually corresponds to the frequency of the lowest harmonic partial or fundamental. If the pitch does not correspond to any partial, there are two possible reasons. One

⁴ A harmonic complex tone is a complex tone (that is, a tone comprising several partials), the frequencies of whose partials correspond to a harmonic series (e.g., 140 Hz, 280 Hz, 420 Hz...).

involves the phenomenon of *pitch shift*: the exact pitch of a pure tone depends on its amplitude (very loud pure tones sound slightly flat compared to very quiet pure tones with the same frequency) and on the presence of other tones (the pitch of a pure tone is pushed upward by a simultaneous tone or noise band at a slightly lower frequency, and pushed downward by a slightly higher frequency; Terhardt, 1998). The other involves *virtual pitch*: the pitch of a complex sound is typically determined by several partials whose frequencies correspond roughly to a harmonic pattern, i.e. to a subset of the harmonic series, even if there is no physical energy present at the fundamental of that pattern. That explains why the pitch of a harmonic complex tone is often (but not always) perceived to remain the same when the fundamental is physically removed – the well-known phenomenon of the "missing fundamental". Even when the fundamental is present, the pitch to which it corresponds is determined primarily by higher partials (except at relatively high fundamental frequencies).

How can the pitch at the missing fundamental be explained? Terhardt's solution was to divide pitch sensations into two kinds: *spectral* and *virtual*. A spectral pitch corresponds to a single partial, and a virtual pitch corresponds to a group of partials whose frequencies correspond roughly to part of a harmonic series.

Spectral pitches are determined physiologically (in the cochlea and auditory nerve) by a mixture of spectral and temporal information; the relative importance of each kind of information depends on the frequency of the tone. Spectral information is transmitted from the ear to the brain by the location on the basilar membrane of activated hair and nerve cells and by their relative degree of activation. Temporal information is transmitted by the neural firing patterns in those nerve fibers (the "volley" and "stimulus fine structure" theories of pitch; Moore, 2003).

Terhardt postulated that virtual pitches arise from the recognition of patterns of spectral pitches, enabled by neural networks (Laden, 1994). In that sense, virtual pitch perception is no different from other kinds of pattern recognition (such as vision). Empirical evidence consistent with this idea was provided by Houtsma and Goldstein (1972), who demonstrated that a virtual pitch can be created by harmonics presented simultaneously to different ears. Data collected by Zatorre (1988) suggested that neural networks underlying virtual pitch perception in the brain are located in and around the right Heschl's gyri.

Contrary to Terhardt's theory, virtual pitches (as defined above) can also arise from interactions between nearby partials of a complex tone that are inaudible due to mutual masking (Moore, 2003). Such partials do not, by definition, produce spectral pitches. However, the virtual pitches produced by such inaudible partials are typically weak (i.e., their *perceptual salience* is low). Thus, the most important pitches that we hear in our everyday lives and in music may indeed be virtual pitches that result from the recognition of harmonic patterns among spectral pitches.

The recognition of harmonic patterns among spectral pitches presumably played a role in the (pre-) historical development of (western) musical syntax. The *roots of chords*, which began to influence western musical syntax in the late middle ages (well before the concept of root had been developed by music theorists), can be predicted on the basis of virtual pitch theory. The Terhardt-inspired model of Parncutt (1988, 1993) derives five "root-support intervals" from the first 10 harmonics by applying the

principle of octave equivalence to the intervals between the fundamental and first 10 harmonics of a harmonic complex tone. These intervals are octave/unison, fifth, major third, minor seventh and major second/ninth. By appropriately weighting these relative to each other and incorporating them into a simple pattern-recognition algorithm, it is possible to predict the various possible roots of all familiar chords including the minor triad, solving a problem that has occupied music theorists for centuries.

insert figure CURVES

Loudness

The (perceived) loudness (German: *Lautheit*) of a pure tone depends strongly on both its frequency and its sound pressure level. This dependency is illustrated by the well-known *curves of equal loudness* of Fletcher and Munson, as shown in Figure CURVES. Consider for example a pure tone of variable frequency in a very quiet environment, synthesized on high-quality audio equipment capable of holding SPL exactly constant over a wide frequency range. If the SPL is held constant at 40 dB, the tone will be clearly audible at 1000 Hz but almost inaudible at 100 Hz. The *threshold of hearing in quiet* is thus not simply a given SPL in dB, but a curve on a graph of SPL against frequency. It is a curve of equal loudness for the special case where loudness is zero. It is plotted by presenting to an experimental subject repeated short bursts of a pure tone at a range of different frequencies. At each frequency, the listener adjusts the tone's intensity until the tone is barely audible.

The loudness of a complex sound additionally depends on its bandwidth (the range of frequencies that it covers). Consider a sound whose partials have constant SPL and different frequencies. As their frequencies are gradually shifted away from each other, the sound gradually becomes louder – even though the total SPL of the sounds remains constant (sound demonstration: Houtsma et al., 1987).

The widely accepted explanation for this effect is that loudness depends on the number of excited auditory filters or critical bands (Moore, 2003).

Roughness

Like pitch and loudness, roughness is an experiential (subjective) parameter. To measure it, listeners must listen to sounds in an experiment and rate how rough they sound. One of the most musically interesting findings of psychoacoustics is the relationship between perceived roughness and CBW.

insert Figure BEATING

Consider a sound consisting of two pure tones of equal amplitude and variable frequency. If the two frequencies are close to each other, we hear *beating* – familiar from pianos in which the two or three strings corresponding to each note are out of tune with each other. This effect is also called *amplitude modulation* or AM, because the amplitude of the temporal envelope of the sound, which we obtain by joining together the peaks in the waveform (from one cycle to the next, see Figure BEATING) gradually rises and falls with a frequency of its own, called the *beat* or

modulation frequency. The modulation frequency is equal to the difference between the frequencies of the original two tones.

Roughness is a sensation (sensory experience) that occurs when beating is so fast that the individual beats cannot be distinguished – faster than about 20 beats per second. The strength or salience of the roughness sensation produced by a beating pair of pure tones depends on both the carrier frequency and the modulation frequency or the signal. Roughness is maximum near the threshold for hearing two separate pitches, corresponding to some 1/4 to 1/3 of a CBW (depending on which measure of CBW you take, and the frequency register of the tones). As the pure tones are moved further apart, the sensation of roughness gradually becomes weaker until for intervals greater than about one CBW it disappears altogether (Plomp & Levelt, 1965).

Roughness explains, at least to some extent, why certain sonorities are more common than others in tonal western music (cf. data of Eberlein, 1994) and in this way establishes a link between certain music-theoretical conventions and the physiology of the auditory periphery. Western music developed according to an arbitrary, culture-specific principle that equates roughness with dissonance. Dissonance is of course not only about roughness – it also has a strong learned component. But roughness evidently contributed to the historical development of western musical syntax, including the functional relationships between consonances and dissonances within that vocabulary (remembering that consonance/dissonance had different meanings in different periods, Tenney 1988). The harmonic intervals of a minor or major second and, to a lesser extent, their inversions (major and minor sevenths) and compounds (minor and major ninths) are particularly rough in the central musical pitch range. In tonal music, sonorities containing these intervals are treated as dissonances requiring resolution, and are less common than other sonorities. This accounts for the higher dissonance and lower prevalence of triads with suspended fourths, and of seventh and ninth chords.

Another perceptual parameter that evidently influenced the evolution of western tonal syntax is perceptual *fusion* (DeWitt & Crowder, 1987).⁵ Fusion happens when a sound is perceived as a single entity that is produced by a single source. In western tonal music, sonorities that fuse perceptually tend to sound more consonant – independent of roughness. For reasons explained above under "pitch", such sonorities typically contain intervals from the lower reaches of the harmonic series, especially octaves, fifths and fourths.

The predominance of the major and minor triads in western tonal music can most easily be explained by assuming that roughness and fusion are the most important perceptual components of consonance and dissonance. The consonance of these chords is evidently due to a combination of two factors: the lack of major or minor seconds (the roughest intervals) and the presence of a perfect fifth or fourth (promoting fusion). No other combination of three tones within an octave of the chromatic scale satisfies both these criteria (Parncutt, 1988).

In the "common practice" tonal western music of the 18th and 19th centuries, major-minor (dominant) seventh chords were more prevalent than major seventh and minor seventh chords, which in turn were more prevalent than diminished and half-

⁵ Fusion is an example of auditory grouping and is subject to gestalt principles (table GESTALT).

diminished seventh chords (data of Eberlein, 1994). This suggests that the principle of fusion may have dominated the principle of roughness in determining the prevalence of sonorities in that period. The principle of fusion favors the major-minor seventh chord because all its tones correspond to lower elements of the harmonic series⁶ – at least within the limits of categorical perception as they are known to operate in the perception of (approximately) harmonic patterns of partials within complex sounds. The major-minor seventh chord is more common than the major and minor seventh chords, even though it is rougher than them (due to the tritone interval between the major third and minor seventh). The diminished seventh chord is not as rare as one might expect from its failure to fuse; it avoids the roughness of the major second interval, and in this sense is more consonant than all other seventh chords.⁷

The roughness of a chord also depends to a large extent on its *voicing* (inversion, spacing, doubling) and on other physical parameters of the tones (octave register, relative amplitude, spectral and temporal envelope). In particular, roughness explains why intervals between the upper voices of musical sonorities are typically smaller than those between lower voices (Plomp & Levelt, 1965; Huron, 2001): CBW in semitones is larger at lower frequencies. The conclusions of the previous paragraph are valid only when one takes an average over many typical voicings and realizations of each chord type.

Timbre

Roughness is an important element of the more complex perceptual parameter known as timbre or tone color. Timbre differs from the other basic perceptual properties of a tone in that it is *multidimensional*: it is always possible to say which of

⁶ The root, third, fifth and seventh of the chord correspond to harmonics 4, 5, 6 and 7 respectively. According to Terhardt's theory of harmonic pattern recognition and virtual pitch perception, the equally tempered minor seventh above the 4th harmonic is perceived to correspond to the 7th harmonic and therefore to belong to the harmonic complex tone, even though it is about 1/3 semitone sharp, because the harmonic pitch pattern recognition procedure normally accommodates such mistunings (Terhardt, 1998).

⁷ Traditional music theory offers different explanations for the relative prevalence of chords. Consider first the role of *voice leading* between successive chords. One might predict that chords are (or become) prevalent if they are easily approached or quitted, or if they can be generated by well-known (medieval or renaissance) principles of voice leading. A problem with this approach is that *all* chords can be approached and quitted in a variety of ways that conform to historical rules of voice-leading. A systematic analysis of these possibilities may fail to yield clear preferences for specific chords. A second possible explanatory theory involves *symmetry* within the chromatic scale. The diminished seventh chord, when presented out of context, does not clearly suggest a tonal center (perhaps the most likely center in "common practice" tonal music is a semitone above the bass, but there are many other possibilities). That is because the chord is symmetrical and does not match any of the diatonic scales, all of which are asymmetrical (which is necessary to enable the recognition of tonal centers). According to this theory, symmetrical chords should be less common and therefore more dissonant, which seems to explain the rarity of augmented triads, diminished sevenths and harmonies containing all tones of the whole-tone scale. This theory is problematic in two respects. First, the dissonance of these chords can more directly and simply be explained in terms of roughness. Second, the combination of two theories – voice-leading and symmetry – yields a counter-theory whose predictions tend in the opposite direction. Symmetrical chords allow for a wide range of resolutions: for example, any diminished seventh chord can act as a pivot between any key and any other key. On this basis, we would expect diminished seventh chords to be quite prevalent. Since symmetrical chords may be predicted to be both rare and prevalent, the inclusion of symmetry in a predictive model of chordal prevalence is problematic.

two pitches is higher or which of two sounds is louder, but not which of two sounds is more "timbral". Instead, one must break timbre down into its components or "dimensions" and ask which of two timbres is richer, brighter, noisier, more nasal, and so on.

A considerable amount of psychoacoustical research (many citing the classical study of Grey, 1977) has attempted to uncover the most important or fundamental dimensions of timbre using a technique called *multidimensional scaling of (dis-)similarity judgments*. In this approach, a limited set of sounds are compared with each other in all possible combinations and both orders of each pair. For example, comparing 10 sounds with each other in all possible combinations yields 100 pairs, of which 10 are identical. The data for the 90 non-identical pairs may be averaged over two possible orders of presentation to give 45 numbers, which are then input to a mathematical procedure called multidimensional scaling (MDS). The output is a map or *Euclidean space* in which the 10 timbres are arranged so that the distance between them is inversely proportional to their similarity. The map may have any number of dimensions; the correspondence between map distances and input data improves as more dimensions are added. Usually there is no significant gain when moving from three dimensions to four, so results are usually displayed in 3-D space. The MDS algorithm does not specify the positions of the axes, which must be determined subjectively by the experimenter (analogous to adding a compass to a geographical map to show the North-South and East-West dimensions). These axes can then be labeled with some physical or perceptual attribute such as the amount of energy at high frequencies relative to low (the *spectral energy distribution*) or its perceptual correlate *brightness*; the amount and kind of noise in the sound onset or its perceptual correlate *attack*; or the way in which partial amplitudes change relative to each other during the first few hundred milliseconds, which is critical for the sound of many musical instruments.

Research on the dimensions of timbre is based on the widely accepted negative definition of timbre, according to which timbre covers all aspects of the experience of a steady-state sound that are not already accounted for by pitch, loudness and subjective duration. In other words, if two steady-state sounds have the same pitch, loudness and subjective duration, and they sound different, then that difference is entirely due to timbre. This definition is not entirely satisfactory, because it ignores the *function* of timbre. Timbre tells us about the identity and state of sound sources. For example, it allows us to distinguish cars from buses, and trumpets from trombones. The relationship between a sound source and its timbre is an example of categorical perception; everything in the above section on categorical perception applies also this relationship.

Auditory processing

Sound is processed both physiologically (in the auditory periphery) and neurocognitively (in the brain). Both stages contribute to achieving the main aim and function of the auditory system, which is to recognize sound sources and provide information about their state. The survival and well-being of most animals including humans depends considerably on their ability to recognize and describe objects based the sounds that they make, either alone (when heard in the darkness, without any other cues such as smell or touch) or in conjunction with other senses.

Physiology

Apart from collecting sound and encoding it as neural firing patterns, the most important physiological function of the ear is running frequency analysis. Without this, it would be impossible to recognize sound sources.

To understand why this is so, consider again the consequences of the fact that all sounds reaching human ears are superpositions of direct and reflected sound (Terhardt, 1998). The addition of reflected sound to direct sound typically changes the shape of the waveform so much that it would be very difficult to distinguish different timbres, and hence different sound sources, on the basis of the shapes of the sound waves picked up at the two ears. For example, odd-numbered harmonics with specific amplitudes add to form a square wave – but only if the phase relationship between them is correct. If not, the resultant waveform is nothing but square. Adding a direct sound to its reflection can completely and unpredictably change the phase relationships between the partials and hence completely and unpredictably change the shape of the resultant wave.

In this sense, phase information does more harm than good when it comes to the identification of sound sources in typical human environments. The auditory system has therefore evolved to be relatively insensitive to phase. This aim was difficult to achieve, since the auditory system (especially the cochlea, auditory nerve and brain) is a physical device whose various moving parts and related electrical impulses are physically phase-locked to each other (Langner, 1992). The success with which the auditory system ignores phase in everyday signals may be demonstrated by resynthesizing everyday sounds such as speech and music from running frequency analyses in which all phase information has been discarded, and showing that such signals cannot be distinguished from the originals from which they were derived (Heinbach, 1988). Phase is important for sound localisation: the phase difference between the two ears is an important cue for determining the direction from which a sound arrives on a horizontal plane. A sudden change in phase relationship between the partials of a complex tone can temporarily affect their relative salience (Moore, 2003).

The evolutionary-ecological argument presented here is consistent with the experimental findings that the ear is sensitive to phase relationships between components in attack or transient portions of a tone but not in steady state tones (summarized by Moore, 2003). In everyday acoustic environments, only the attack portion of a sound can be perceived without interference from reflected sound. Therefore, the phase relationships in the attack portion can contain reliable information about the sound source. For the same reason, the ear should only be sensitive to phase relationships in sounds heard against a quiet background or with a high signal to noise ratio.

Regarding amplitude, different surfaces reflect different amounts of sound at different frequencies; for example, carpet absorbs more high than low frequencies. The amplitudes of tone components are not, therefore, entirely reliable sources of information about sound sources. The ear therefore needs to be able to recognize sound sources in spite of quite large variations in the (relative) amplitudes of their partials. For example, we can easily communicate by telephone, and enjoy music played on poor sound equipment, even though in both cases the lower and higher

frequencies of the original signal are weak or entirely absent. But if frequencies within these signals are changed, their character changes drastically.⁸

The ear is highly sensitive to frequency, because frequencies are not changed at all by reflection – provided that (1) the source, reflector and receiver are all stationary or at least moving much more slowly than the speed of sound relative to one another, and (2) changes in frequency are gradual by comparison to the duration of the ear's frequency analysis window – a condition that is usually fulfilled in everyday environments (Terhardt, 1998). Under these conditions, temporal patterns (rhythms) are not changed by reflections either, so the ear is also very sensitive to these, and also uses them to identify sound sources.

Thus, the ear solves the problem of source identification in an environment that includes reflecting surfaces by subjecting all incoming sounds to a running frequency analysis. This allows it to separate phase, amplitude and frequency information, to discard the phase information, and to focus on frequencies and rhythms. The importance of frequency and time for auditory perception is reflected by conventional music notation, in which the vertical axis corresponds approximately to the fundamental frequency of musical tones and the horizontal axis approximately to time. It is also reflected by graphical representations of the *auditory scene* (Bregman, 1993).

Insert figure RUNNING ((z.B. Bregman Fig. 2.1b))

Figure RUNNING shows a running frequency analysis of a brief vocal utterance, with frequency plotted on the vertical axis and time on the horizontal axis. If all parts of this figure are assumed perceptible (in the sense that changes in any part of the figure can be perceived), and if the auditory system reconstructs the source of the sound on the basis of the figure's details, the figure may be regarded as an *auditory scene* – analogous to a visual scene. The auditory scene may be regarded as the output of the first, physiological stage of sound processing and the input to the second, cognitive stage. The separation of auditory processing into two stages is controversial (Gibson, 1979), but it does make the system easier to understand.

Cognition. In a visual scene, we recognize objects by applying the Gestalt principles listed in table GESTALT.⁹ The *Gestalten* or patterns in question are generally defined by the perceptible boundaries or edges of an object, here called its *contours*. Obviously, recognition of an object is only possible if its contours are grouped together. The importance of contours for object recognition is clear from the ease with which objects can be recognized in black-and-white cartoons.

⁸ The ear is not very sensitive to small changes of intensity; the threshold for successive sounds is about between 0.5 and 1 dB, or roughly 20% of intensity. Instead, the ear is sensitive to a very wide range of intensities, from as little as 10^{-12} Wm⁻² (0 dB SPL) right up to 1 Wm⁻² (120 dB SPL).

⁹ The basic principles of auditory scene analysis expounded by Bregman (1993) are closely related to the gestalt principles. His *Regularity 1* ("unrelated sounds seldom start or stop at exactly the same time") is an example of the proximity principle: simultaneous sounds group because they are proximate in time. His *Regularity 2* (gradualness of change) is none other than the principle of good continuation. *Regularity 3* is the principle of closure applied to the frequency relationships between the harmonic partials of a harmonic complex tone. *Regularity 4* (partials tend to change in the same way at the same time) corresponds to the principle of common fate.

Name	visual	auditory or musical
proximity	An object's contours tend to be physically close to each other.	The tones of a melody are close to each other in pitch and time; if not, the melody breaks perceptually into fragments (Noorden, 1975). The tones of a chord fuse when their onsets are synchronous (temporal proximity) and not too widely spaced (pitch proximity).
similarity	An object's contours tend to look similar to each other.	The tones of a melody are similar in timbre; if not, the melody breaks up perceptually into fragments (Wessel, 1979).
closure	Some of an object's contours may be imperceptible due to occlusion or masking by other objects.	Harmonic complex tones fuse perceptually even if one or more partials (including the fundamental) are physically missing or inaudible.
common fate	An object's contours tend to move in synchrony with and at the same speed as each other, when the object moves.	When the frequencies and/or amplitudes of the partials of a complex tone move in parallel (e.g. in a musical <i>vibrato</i> , in which frequency and/or amplitude ratios are held constant), the tone tends to fuse perceptually, even if the spectrum is not harmonic.
good continuation	An object's contours tend to be smooth (straight, or with a large radius of curvature) and not to change direction suddenly. ¹⁰	This principle applies to continuations following melodic steps but not following large leaps, which are typically followed by a change in direction (Huron, 2001).

Table GESTALT. Names and explanations of some well-known gestalt principles in visual and auditory perception.

The gestalt principles are similar for seeing and hearing, with some interesting exceptions. The boundaries of a visible object are analogous to the frequency-time trajectories or contours of audible partials (Terhardt, 1998). The principles of proximity, similarity, closure and common fate explain important aspects of both visual and auditory (including music) perception. The principle of good continuation is not as strong in hearing as in vision. Consider the following two cases. First, melodies do not normally cross over each other, because if they do, it is hard to hear one rising or falling through the other. If the tones of the melodies differ only in pitch, i.e. if other parameters such as timbre, loudness and articulation are held constant, the principle of proximity prevails over the principle of good continuation: at the point where the melodies cross, tones are grouped because they are close to each other in pitch and time and not because they proceed steadily in a given direction. Thus, we hear a lower melody rise to a peak and fall again, and a higher melody fall to a valley and rise again (Deutsch, 1999). Second, stepwise melodic motion in music tends to

¹⁰ The physical principle of conservation of momentum leads observers to expect that heavy objects will not suddenly stop or start moving. This effect may be related to the gestalt principle of good continuation or regarded as a separate principle.

continue in the same direction, and listeners also have this expectation. But the same listeners tend to expect a (larger) 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). Here, knowledge or expectations about a melody's tessitura override the principle of good continuation.

This difference between hearing and vision suggests that the gestalt principles primarily reflect regularities in the physical world that are relevant for human survival and well-being (Gibson's 1979 *affordances*), and are not necessarily based on underlying cross-modal cognitive processes. Their acquisition may involve either *ontogeny* due to exposure to similar stimuli within the life-span of an individual (physiologically enabled by self-organizing neural networks, Laden, 1994) or *phylogeny* across many lifespans guided by the Darwinian principle of survival of the fittest (physiologically enabled by the genes involved in brain development).¹¹ The visual principle of good continuation is acquired as the visual system catches onto the following regularity: boundaries of distant objects tend to continue in the same way and direction in spite of interruptions due to occlusion by closer objects. This principle does not apply in hearing, because 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. In other words, in the perceptible physical world, the partial frequencies of complex tones are subject to physical constraints to which the visible boundaries of physical objects are not. Other *Gestalten* in hearing and in music have no visual correlates whatsoever; for example, the tonality of a piece of music may be regarded as a *Gestaltqualität* that is learned from exposure to tonal music (Krumhansl, 1990).

Conclusion

Psychoacoustic theory can be applied in many theoretical and practical areas of music and musicology, of which some have been addressed in this chapter. These areas include:

- music theory: rules of harmony and voice leading, history of tonal syntax
- composition: conceptual basis, computer-based tools
- historical and cultural musicology: understanding musical experience
- teaching of musicology and music performance
- music recording technology
- design and manufacture of musical instruments
- room acoustics: design of concert halls

¹¹ Bregman (1993) distinguished between *primitive* (or innate) principles and learned *schemas*. In specific cases it is difficult to know the extent to which a principle is innate or learned. As in so many other domains, the nature-nurture question remains unanswered.

Literature

- Bregman, A. S. (1993). Auditory scene analysis: Hearing in complex environments. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 10-36).
- Burns, E. M. (1999). Intervals, scales, and tuning. In D. Deutsch (Ed.), *Psychology of music* (2nd. ed., pp. 215-264). San Diego, CA: Academic.
- Deutsch, D. (1999). Grouping mechanisms in music. In Deutsch, D. (Ed.), *Psychology of music* (2nd. ed., pp. 299-348). San Diego, CA: Academic.
- DeWitt, L. A., & Crowder, R. G. (1987). Tonal fusion of consonant musical intervals: The Oomph in Stumpf. *Perception & Psychophysics*, 41, 73-84.
- Eberlein, R. (1994). *Die Entstehung der tonalen Klangsyntax*. Frankfurt: Lang.
- Egan, J. P., & Hake, H. W. (1950). On the masking pattern of a simple auditory stimulus. *Journal of the Acoustical Society of America*, 22, 622-630.
- Ferguson, S. (2001). *Concerto for piano and orchestra*. Doctoral thesis, Faculty of Music, McGill University, Montreal, Quebec.
- Gabor, D. (1947). Acoustical quanta and the theory of hearing. *Nature*, 159, 591-594.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Grey, J. M. (1977). Multidimensional scaling of musical timbres. *Journal of the Acoustical Society of America*, 61, 1270-1277.
- Haas, H. (1951). On the influence of a single echo on the intelligibility of speech. *Acustica*, 1, 48-58.
- Heinbach, W. (1988). Aurally adequate signal representation: The part-tone-time pattern. *Acustica*, 67, 113-121.
- Houtsma, A. J. M., Goldstein, J.L. (1972). The central origin of the pitch of complex tones: Evidence from musical interval recognition. *Journal of the Acoustical Society of America*, 51, 520-529.
- Houtsma, A. J. M., Rossing, T. D., & Wagenaars, W. M. (1987). *Auditory demonstrations* (CD and booklet). Acoustical Society of America.
- Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception*, 19, 1-64.
- Krumhansl, C. L. (1990). *Cognitive foundations of tonal music*. New York: Oxford University Press.
- Laden, B. (1994). A parallel learning model of musical pitch perception. *Journal of New Music Research*, 23, 133-144.
- Langner, G. (1992). Periodicity coding in the auditory system. *Hearing Research*, 60, 115-142.
- Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). New York: Academic Press.
- Noorden, L. van (1975). *Temporal coherence in the perception of tone sequences*. Doctoral dissertation, Institute for Perception Research, Eindhoven, NL.
- Parncutt, R. (1988). Revision of Terhardt's psychoacoustical model of the root(s) of a musical chord. *Music Perception*, 6, 65-94.
- Parncutt, R. (1989). *Harmony. A psychoacoustical approach*. Heidelberg: Springer.
- Parncutt, R. (1993). Pitch properties of chords of octave-spaced tones. *Contemporary Music Review*, 9, 35-50.
- Plomp, R. (1964). The ear as frequency analyzer. *Journal of the Acoustical Society of America*, 36, 1628-1636.
- Plomp, R., & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, 38, 548-560.

- Popper, K. R., & Eccles, J. C. (1977). *The self and its brain*. Berlin: Springer.
- Tenney, J. (1988). *A history of 'consonance' and 'dissonance'*. Excelsior, New York.
- Terhardt, E. (1998). *Akustische Kommunikation*. Berlin: Springer.
- Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal*, 3, 45-52.
- Zatorre, R. J. (1988) Pitch perception of complex tones and human temporal-lobe function. *Journal of the Acoustical Society of America*, 84, 566-572.