12 The psychology of electronic music

PETRI TOIVIAINEN

Psychology provides an important base from which to understand music, and is very relevant for electronic music in particular, where psychological theories have even inspired new compositional explorations. Furthermore, in analysing and composing electronic music, traditional music theory is often not applicable. There is no conventional score available on which the analysis of the music could be based, for the music does not rest solely on certain standard notated pitch structures and rhythmic frameworks, but encompasses timbre, spatialisation and other general auditory parameters. An appreciation of the role of aural cognition is vital for a true engagement with this field, where any sounding object is fair game.

The purpose of this chapter is to provide an introduction to perceptual and cognitive processes of music that are fundamental for understanding electronic music. The chapter begins with a discussion of the neuroscientific basis of the auditory system. This is followed by a discussion of low-level phenomena of audition, including the localisation of sound sources, masking, auditory stream segregation and the perception of timbre. Next, the perception of pitch is tackled, with a discussion on its relation to alternative tunings. Finally, basic notions of rhythm perception are introduced. For each of these parts, electronic music examples illustrating the perceptual principles will be given. Any and all principles expounded in this chapter might be taken up and profitably investigated by electronic musicians.

The neuroscientific basis of audition

The auditory system can be partitioned into three processing stages (Pickles 1988; Moore 1997). These are the auditory periphery, the auditory pathway and the auditory cortex. The auditory periphery consists of the outer, middle and inner ear, the auditory pathway of the tracts connecting the ear and the auditory cortex, and the cortical level primarily of the temporal lobes at the left and right sides of the brain.

The ear transforms the mechanical energy of sound vibrations into nerve impulses. The *outer ear* consists of the pinna and the ear canal, whose function is to collect and amplify the energy of the sound and lead it to the tympanic membrane. The ear canal also acts as a closed tube resonator

https://doi.org/10.1017/CCOL9780521868617.014 Published online by Cambridge University Press

and amplifies frequencies in the range of 2–5 kHz. The oscillations of the tympanic membrane are led to the inner ear via three bones (ossicles) of the middle ear, the malleus (hammer), incus (anvil) and stapes (stirrup). The vibrations of the ossicles enter the cochlea, a spiral structure of the inner ear, through the oval window.

The cochlea performs a transformation of the mechanical vibration into electrical impulses. This is carried out through the movement of the *basilar membrane*, which bends rows of hair cells beneath it. The bending of the hair cells gives rise to electric impulses that encode information about the periodicity and intensity of the sound.

Due to its mechanical properties, the basilar membrane acts as a frequency analyser. More specifically, the stiffness of the membrane varies along its length, causing the front end (base) of the membrane to resonate with higher frequencies, and its rear end (apex) with lower frequencies.

The *auditory pathway* leads neural impulses from the cochlea to the cortex. It also contains nuclei that carry out preliminary analysis of the sound signal with regard to, for instance, its intensity and spatial origin. The *cochlear nucleus* sharpens the frequency information contained in the neural signals. The *inferior colliculus* plays a role in sound source locating. The *thalamus* is considered as the 'gateway to cortex' and a 'gatekeeper of conscious experience' (Llinas *et al.* 1998); all neural information to the cortex passes through the thalamus.

The *auditory cortex* is one of the most folded parts of the brain. It is this part of the auditory system where identification and segregation of auditory objects occurs; memory-based sound processing also takes place in the cortex. Although the left and right cortices are mostly similar, some functional differences have been observed. In particular, the left hemisphere has been found to be dominant in rhythmic processing, whereas the right one is more relevant for pitch processing (Zatorre 2003).

There is still much unknown about the functioning of the cortex, but research in this area is active, and there will certainly be many implications of this research for musicians in the future. For instance, revealing the neural determinants of musical emotions is useful for understanding the elements of music that affect listeners' mood. Despite the present incomplete knowledge on the functioning of the brain, there have been several applications to use the activity on the cortex to produce music. In particular, brainwaves measured with electroencephalogram (EEG), have been used to control electronic synthesisers. Pioneers in this field include Richard Teitelbaum and David Rosenboom. For instance, in his work *In Tune* (1967), Teitelbaum combined amplified EEG signals with sounds of heartbeat and breath. In his composition *Ecology of the Skin* (1970), Rosenboom used ten live EEG performers to interactively generate immersive sonic environments.

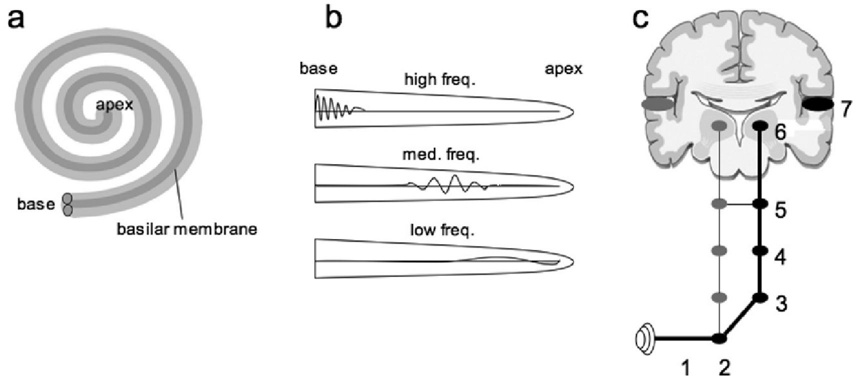

Figure 12.1 (a) Schematic presentation of the cochlea; (b) Excitation pattern of basilar membrane for high, medium, and low frequencies; (c) Auditory pathway; 1. Auditory nerve; 2. Cochlear nucleus; 3. Superior olive; 4. Lateral lemniscus; 5. Inferior colliculus; 6. Thalamus; 7. Auditory cortex

Several computational models of the auditory processing have been proposed and implemented as computer algorithms. For instance, the IPEM Toolbox (Leman, Lesaffre and Tanghe 2001) contains an implementation of models for pitch, sensory dissonance, onset detection, beat and metre, and timbre characteristics. Such models could have various applications as, for instance, artificial ears that could be used by composers to make perceptual analysis of their music. It must be noted, however, that a well-grounded theory of auditory processing only exists for the peripheral level, whereas models of the subsequent levels are much more speculative.

Localisation

Spatialisation has played an important role in electronic music throughout its history. Humans have a remarkable ability to localise sound sources accurately and rapidly. Sound localisation can be divided into three components: localisation of azimuth, elevation and distance. In what follows, each of these components is discussed in turn.

Localisation of azimuth refers to identifying the direction of the sound source on the horizontal plane. The two main cues used in this process are the *Interaural Time Difference* (ITD), and the *Interaural Level Difference* (ILD). The ITD is caused by the difference in the time it takes for the sound wave to reach the two ears. This is illustrated in Fig. 12.2a. The ILD, on the other hand, is caused by the fact that the ear that is more distant from the sound source receives less sound energy due to the head's shadow. This is demonstrated in Fig. 12.2b. To determine the azimuth, the auditory system uses both ITD and ILD information.

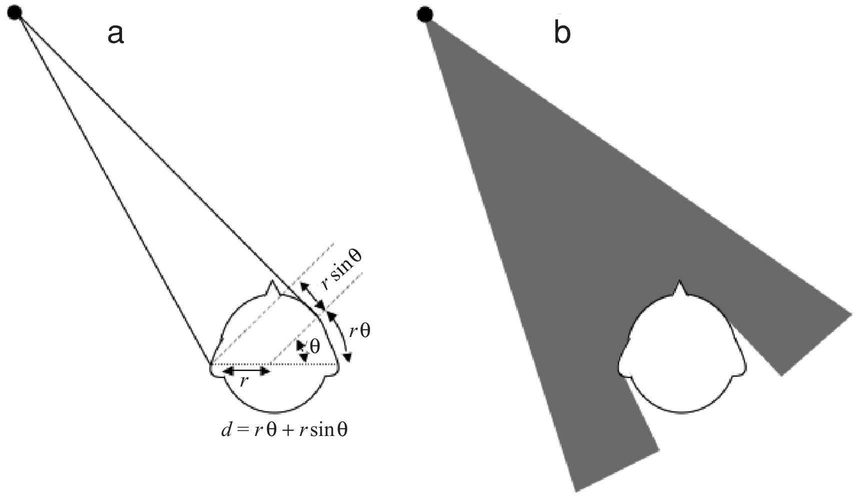

Figure 12.2 (a) Interaural Time Difference. The distance from the sound source (black circle) to the right ear is $d = r\,\theta + r\,\sin\theta$ longer than to the left ear, where r denotes the radius of the head and θ the azimuth of the sound source. (b) Interaural Level Difference. Head shadowing reduces the intensity of the sound arriving to the right ear

The *localisation of elevation* of a sound source is less accurate than that of azimuth. Still, we can easily tell whether the sound source is ahead, above or behind us. In this process, there are no binaural cues (ITD, ILD) to rely on; rather, the most important cues are related to the spectral shape of the perceived sound. In particular, the received sound spectrum is modified by reflections from the pinna that depend on elevation.

For *localisation of distance*, the loudness of a sound source is an evident cue, but cognitive knowledge about the quality of the particular sound has to be applied to utilise this cue. For instance, shouting from a long distance can have a higher perceived loudness than whispering from a short distance. A further cue is motion parallax, which refers to the fact that translational movement of the listener causes larger azimuth change for nearby objects than for distant ones. Further cues, sometimes used for musical purposes, are the loudness ratio between the direct and the reverberant sound (Zahorik 2002), and the brightness of timbre. More specifically, high intensity of reverberation gives an impression of a distant sound source. Furthermore, a dark (low-pass filtered) timbre may give an impression of a distant source, because high frequencies attenuate faster than low ones in the air.

One of the first electroacoustic compositions, Edgard Varèse's *Poème électronique*, was presented at the Brussels World's Fair in 1958. The audio part of this multimedia composition consisted of a three-track tape recording, each track of which was distributed dynamically to 425 speakers through an eleven-channel sound system with twenty amplifier channels. Karlheinz

Stockhausen has used spatial sound as an integral part of his work. He produced the first true quadraphonic composition for electronic sounds, *Kontakte* (1960). In this work, Stockhausen used a turntable system with a rotating loudspeaker mechanism, recording to four-channel tape via four microphones spaced around the table, creating an effect of sounds orbiting around the audience.

With methods of digital signal processing it is possible to spatialise sounds accurately by simulating the ITDs and ILDs as well as filtering and adding reverberation. John Chowning (1971) described techniques for the simulation of moving sound sources that are based on the Doppler effect as well as reverberation effects. The Doppler effect refers to the change of frequency caused by the movement of the sound source relative to the listener, and can be used compositionally to create an illusion of movement. Chowning's composition *Turenas* (1972) is a quadraphonic work based on these techniques. In the domain of digital signal processing, various techniques of sound spatialisation have been developed since and applied by composers of electroacoustic music. These include ambisonics, holophonics, wave-field synthesis, Dolby 5.1 Surround, and Digital Theater Systems (DTS) as well as binaural systems based on head-related transfer function (HRTF). Today's efficient computers and sound-processing software allow relatively easy experimentation with sound spatialisation effects. This issue is further discussed in chapter 13 of this book by Natasha Barrett.

Masking

Masking refers to a phenomenon whereby a signal with a low intensity (the maskee) is made inaudible by a stronger signal (the masker). There are two types of masking: simultaneous and temporal. In *simultaneous masking*, the strength of masking depends on the frequency content of the two signals. For instance, with pure tones, the masking effect is stronger the closer the frequency of the maskee is to that of the masker. Moreover, masking is stronger for frequencies above that of the masker than below. Simultaneous masking is caused by the overlap of excitation patterns on the basilar membrane. The range of frequencies within which masking occurs for a given frequency is referred to as the *critical bandwidth*. This bandwidth is about 90 Hz wide for sounds below 200 Hz, and increases to about 900 Hz for frequencies around 5000 Hz. The degree to which complex sounds mask other sounds, or are masked by other sounds, depends, in addition to their intensity, on their spectral content. For instance, a higher intensity difference is needed for a sinusoidal masker to mask a noise-like maskee than is needed for a noise-like masker to mask a sinusoidal maskee.

https://doi.org/10.1017/CCOL9780521868617.014 Published online by Cambridge University Press

Temporal masking refers to the phenomenon whereby a soft tone is masked by a louder tone that occurs shortly before (post-masking) or after (pre-masking) the soft tone. For a constant difference in the intensities of the masker and the maskee, post-masking has been found to have a longer range than pre-masking. Typically, post-masking occurs within about 50–200 msec (milliseconds) after the removal of the masker, whereas the range of pre-masking is about one tenth of this (Zwislocki 1978). All three forms of masking are utilised in Perceptual Audio Coding, such as MP3, MiniDisk, and Ogg Vorbis.

The phenomenon of masking has a number of implications to compositional practice. For instance, a loud sound object in a composition may make weaker sound objects in nearby frequencies inaudible; assigning the sound objects different frequency ranges will diminish the effect of masking. For instance, a melodic line is often played in a higher register than the accompaniment to make it better audible. The implication of temporal masking is that the audibility of less loud parts of the musical material can be improved by placing them at different temporal locations than louder parts. Such techniques are applied by many musicians and mix engineers in their practice.

Auditory streaming

Our auditory system has a remarkable capacity to make sense of the sounds we receive from our environment. In particular, it can extract from the sound signals we receive meaningful chunks of information that correspond to real-world activities. For instance, in a room full of people talking to each other we can usually without any trouble concentrate on the speech of a single person. In other words, our perceptual system is capable of extracting meaningful *auditory streams* from the perceived sound signals. When listening to music, we also tend to hear the auditory information as a collection of streams, such as parts in counterpoint, melodic lines, inner voices, bass lines, and accompaniment.

The research on the formation of auditory streams has a long history. The founders of Gestalt psychology, such as Paul Ehrenfels (1890) and Max Wertheimer (1923), initially presented musical examples to support their notions. The main ideas of Gestalt psychology can be summarised into a few principles. The ones that are most relevant with regard to auditory streaming are the following:

Principle of proximity: objects that are close to each other tend to be grouped together

https://doi.org/10.1017/CCOL9780521868617.014 Published online by Cambridge University Press

Principle of similarity: objects that share similar characteristics tend to be grouped together
Principle of good continuation: there is a preference for continuous forms
Principle of closure: objects that seem to form closed entities tend to be grouped together
Principle of common fate: objects that move together tend to be grouped together

In addition to the Gestalt psychologists, the problem of auditory streaming has been studied by researchers such as Hermann von Helmholtz, Carl Stumpf, Jay Dowling, Diana Deutsch, Leon van Noorden, David Wessel and Stephen McAdams. The most influential work in this field is the book *Auditory Scene Analysis* by Al Bregman (1990). In this book, Bregman presents detailed accounts of various processes involved in auditory streaming, many of which are based on the notions originally presented by the Gestalt psychologists. Bregman distinguishes between two main types of processes that are involved in auditory streaming. These are sequential integration and spectral integration.

Sequential integration refers to the putting together of events that follow one another in time. Musical events can coalesce into a single stream if they are sufficiently proximal in time and/or pitch. Moreover, the closer in time two events are, the more proximal in pitch they have to be in order to be integrated into a single stream. An example of this dependence is graphically depicted in Fig. 12.3a. The dependence between pitch and temporal proximity in stream formation was quantified by van Noorden (1977), who defined the temporal coherence and fission boundaries for auditory streaming. An example of auditory stream formation can be found, for instance, in Andean pipe music where musicians play alternate notes that, due to their pitch proximity, are perceived as a single stream. Sequential integration can also be based on similarity in timbre or loudness, so for instance, if a musical passage contains tones played by two instruments, these tend to be heard as two separate streams.

Spectral integration refers to integrating components that occur at the same time in different parts of the spectrum. There are a number of principles that govern this phenomenon. First, we tend to group frequency components by harmonicity. More specifically, components sharing the same fundamental are likely to come from the same source, and are grouped together. Second, we group frequency components by onset. This means that frequency components that have proximal onset times are likely to come from the same source, and are grouped together. Finally, we group frequency components based on similarity of their temporal evolution. For instance, spectral components sharing the same frequency modulation

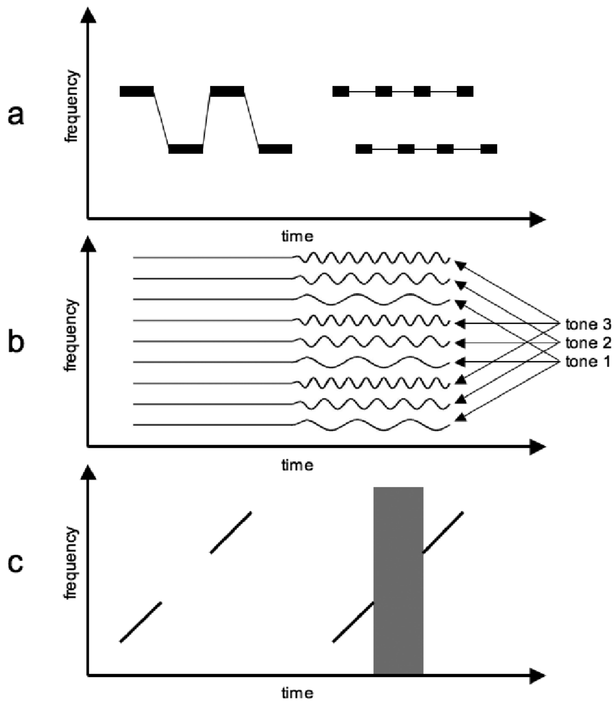

Figure 12.3 (a) Grouping by proximity. A slow sequence of tones with alternating frequencies (left) is perceived as a single stream; the same sequence played twice as fast is perceived as two separate streams; (b) Grouping by common fate. A collection of partials with no frequency modulation is perceived as a single tone (left); when frequency modulation is introduced, tones with similar FM pattern are grouped together, resulting in a percept of three separate tones (right); (c) Principle of 'old-plus-new heuristic'. Two frequency slides are perceived as separate tones (left); when a tone burst is played between the slides, they are perceived as a single tone

pattern (such as vibrato) are likely to come from the same source, and are grouped together (Cook 1999). An example of this phenomenon is shown in Fig. 12.3b. The last two principles are instances of the Gestalt principle of common fate.

A further principle involved in spectral integration is what Bregman refers to as the *'old-plus-new heuristic'*. This refers to perceptual continuation of an old sound at the presentation of a more complex sound. In other words, if part of a present sound can be interpreted as being a continuation of a previous sound, the auditory system tends to make this interpretation. An example of this principle is illustrated in Fig. 12.3c.

Perception of auditory streams plays a crucial role in the parsing of musical compositions. Based on the principles described above we hear, for instance, separate voices in a musical work. In much electronic music we cannot talk about melodic lines or even pitch, but the music still has temporal and spectral dimensions, to which the streaming principles apply.

As a result, in the total mass of sound we perceive layers that are independent from each other and that at some points fuse into a single percept.

Similarities based on timbre can be used to compose streams. On the other hand, introducing timbral differences within a sequential stream may add interest. This has been used, for instance in *Klangfarbenmelodien* by Schoenberg and Webern. In a melodic line, subsequent tones are played with different instruments. The rapid changes in timbre interfere with the smooth sequential integration, thus bringing interest to the music.

Very often a clear segregation of two streams in music is desired. To obtain this, one can, for instance, introduce differences between the spectral content of the two streams. A classical example of this is the singer's formant, a spectral bulge around the frequency of 3 kHz that helps the singer to be distinguished from the ensemble. Furthermore, the principle of common fate can be applied by introducing individual vibrato patterns to the streams. A further means is to add minor temporal deviations to otherwise simultaneous events. Even a slight asynchrony renders the two streams perceptually separate. Sound spatialisation provides a further means to separate streams: sound sources are segregated more easily when they are placed at different directions in the sound field than when they appear to come from the same direction.

Timbre perception

The timbre of sound is a complex phenomenon. There have been a number of studies aiming at extracting the most salient acoustic attributes affecting the perception of timbre (most aimed at studying monophonic instrument tone colour, rather than general sound objects). A widely used method is *similarity rating* (SR): subjects are asked to rate, on a given scale, the dissimilarity of all possible pairs in the set of stimuli. *Multidimensional scaling* (MDS) is then used to map the tones onto a low-dimensional space – frequently referred to as the timbre space. Most of these studies have found that a three-dimensional timbre space represents the dissimilarity ratings to a sufficient degree of precision (e.g. Grey 1977; McAdams *et al.* 1995). The first dimension in the MDS solution has been found to correlate with the *spectral centroid*, which corresponds to the perceived brightness of a tone. The second dimension in the MDS solution relates to the *attack time*, that is, the time it takes for the amplitude of the tone to reach the maximal value. As regards the third dimension, there is more discrepancy between the studies. It has been associated with spectral variation over time or spectral irregularity.

Most of the work on timbre has concentrated on single musical tones, that is, *monophonic timbre*, whereas the overall timbre of a musical piece, also referred to as the *polyphonic timbre*, has received less attention. Recently, however, there has been increased interest in the study of this phenomenon (Aucouturier 2006). Much of this work has been carried out in the field of Music Information Retrieval, where it has been found that various descriptors of polyphonic timbre can be leveraged, for instance, for the automatic classification of audio. A practical application of polyphonic timbre analysis is the Shazam query-by-example system for recognition of music via mobile phone. The system is based on a method of audio fingerprinting; relative positions of peaks in the spectrum of the audio query are located and converted into a fingerprint, which is then matched with the fingerprints of the pieces in the database.

A wider notion of timbre has played a vital role in electroacoustic music. For instance, the French composer Pierre Schaeffer (1910–95), the inventor of *musique concrète*, started from concrete sounds such as voices, noises, as well as sounds of prepared and conventional instruments, experimented with them, and abstracted them into musical compositions. Examples of such works by him are *Suite pour quatorze instruments* (1949), which is based on timbral transformations of orchestral instruments, and *Symphonie pour un homme seul* (1951), co-composed by Pierre Henry, which employed, among other things, the sounds of the human body. The composition *Atmosphères* (1961) by the Hungarian composer György Ligeti is a classic example of timbral composition. Written for a large orchestra, it abandons the concepts of melody, harmony and rhythm, while concentrating solely on the timbre of the sound produced. *Atmosphères* opens with a large cluster chord comprising every tone in the chromatic scale over a range of five octaves. Because of the pitch proximity of adjacent components in the cluster, the auditory systems cannot resolve every tone. Therefore, what is perceived is the timbre of the sound mass rather than a chord.

Advanced methods of digital signal processing allow operations on the microscopic level of sound structure, dubbed *microsound* by Curtis Roads (2001). Granular synthesis uses small pieces of sound, typically with a length between 1 and 50 msec, to build sound textures, or clouds. The first composer to use this technique was Iannis Xenakis in his compositions *Analogique A et B* (1958–9). Extended notions of timbre and sound materials are discussed in the next chapter in particular.

Pitch perception and alternative tunings

Pitch is a perceptual attribute of a tone that depends mainly on its frequency content. For complex harmonic tones, the perceived pitch is equal to the fundamental frequency of the complex tone, that is, the frequency of the first partial. For inharmonic tones, such as a bell tone, the perceived pitch may, however, be unclear. In fact, such tones often elicit a percept of several simultaneous pitches. Terhardt *et al.* (1982) proposed a model of *pitch salience* according to which this attribute depends on the degree of harmonicity of the tone, that is, the degree of coincidence of the subharmonics of the partials.

The *place theory* of pitch perception, originally suggested by George von Békésy (1963), states that the pitch percept can be explained by the locus of maximal vibration on the basilar membrane. The *rate theory* (Seebeck, 1843), on the other hand, states that the neural firing patterns encode the periodic structure of auditory stimuli. According to a commonly accepted view, both place and rate information are used by the auditory system to determine pitch, with rate information dominating for low frequencies and place information for high frequencies.

Traditionally, most music is comprised of a collection of discrete pitches, or *scales*, instead of a continuum of them. Furthermore, it is common that the scales repeat themselves after an octave, or a frequency ratio of two. Many studies have indicated that tones an octave apart are perceived as highly similar (Burns 1999) – a phenomenon referred to as *octave equivalence*. Conventionally, the number of pitches per octave has been between five and seven (Carterette and Kendall 1999). The repetition of pitch intervals after each octave is, however, dropped in some scales, such as the Bohlen–Pierce scale (see below). Moreover Iannis Xenakis uses non-octave scales in his works, such as *Tetora for string quartet* (1990), and discusses them in his writings on sieve structures (Xenakis 1992).

Scales are derived from *tuning systems*. In particular, a scale is a subset of a tuning system, often uneven in the sense that it consists of pitch intervals of varying sizes. There is a long tradition in the development of various tuning systems in Western culture, starting from Pythagoras (582–507 BC), who observed that consonant musical intervals produced by a vibrating string were associated with simple integer ratios of string length. In the Western world, the 12 Tone Equal Tempered tuning (or 12TET tuning) is the most prevalent nowadays. Although it is the most studied tuning system from both music-theoretical and perceptual points of view, it must be noted that it is not universally accepted.

The 12 Tone Equal Tempered tuning consists of intervals of equal size with a frequency ratio of $2^{1/12}$ between subsequent tones. It provides an approximation of tunings based on simple frequency ratios, while allowing for modulation between keys. It is possible to construct equally tempered scales beyond the 12TET tuning by using a frequency ratio of $2^{1/k}$, where k can be any integer. Many of these alternative tunings have been proposed to provide an approximation for a tuning based on simple frequency ratios. For instance, the 19TET tuning was used by the composer Guillaume Costeley as early as the sixteenth century. This tuning serves as an approximation for mean-tone temperament. The 22TET tuning, proposed by the nineteenth-century English scholar R. H. M. Bosanquet, provides an approximation of the five-limit just intonation, that is, a tuning consisting of frequency ratios that can be expressed with the primes 2, 3 and 5. Many composers, such as Charles Ives and Krzysztof Penderecki, have composed music using the 24TET tuning, consisting of half-semitone intervals. Higher-order equal temperaments that have been proposed include the 31TET, 53TET and 72TET tuning.

One great advantage of computers is the ease with which they allow exploration of alternative tuning systems. The 12TET tuning was originally adopted in order to allow flexible modulations between keys, for which the just intonations were not suitable. Today's computer technology allows the use of adaptive tunings, in which the just intonation is adapted to fit with the current key (see, e.g., Sethares 1994, 1999).

It is not necessary that an equally tempered tuning be based on the octave. Generally, one can design equally tempered tunings using a frequency ratio of $p^{1/k}$, where p can be any number. For instance, the Bohlen–Pierce tuning is a 13TET tuning based on a *tritave*, or a frequency ratio of $p=3$. This tuning can be seen as an approximation of a just intonation system based only on ratios of odd whole numbers, therefore being appropriate for timbres containing only odd harmonics. Composers who have utilised the Bohlen–Pierce scale in their works include Charles Carpenter (e.g., *Frog à la Pêche*, 1994), Juan Reyes (e.g., *ppP*, 1999–2000) and Richard Boulanger (e.g., *Solemn Song for Evening*, 1990).

Rhythm perception

The auditory system can accurately detect the temporal structure of the sounds it receives. In the shortest timescale, there are certain thresholds that are useful for electronic musicians. In binaural hearing, temporal differences as short as twenty microseconds can be detected. For monaural hearing, the threshold of simultaneity for clicks is about two milliseconds and for musical tones about twenty milliseconds. Echoes are discriminated if their temporal distance from the direct sound is at least fifty to sixty milliseconds.

Much music is organised so as to contain temporal periodicities that evoke a percept of regularly occurring *pulses*, or *beats*. The ability to infer beat and metre from music is one of the basic activities of musical cognition. It is a rapid process: after having heard only a short fraction of music, we are able to develop a sense of beat and tap our foot along with it. Even if the music is rhythmically complex, containing a range of different time values and probably syncopation, we are capable of inferring the different periodicities and synchronising to them.

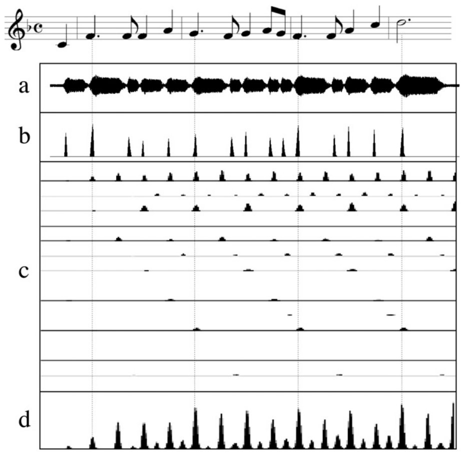

Figure 12.4 Response of a resonating oscillator bank to an excerpt from the Scottish folk melody 'Auld Lang Syne'. a) waveform; b) output of an onset detector; c) outputs of resonating oscillators; d) summed output of all oscillators. Notice the form of the summed output indicating a hierarchical structure of beat strengths.

A rhythmical sequence usually evokes a number of different pulse sensations, each of which has a different perceptual salience. The salience of a given pulse sensation depends on a number of factors related to the surface and structural properties of music. These factors include the frequency of tone onsets that coincide with the pulse (Palmer and Krumhansl 1990), and the *phenomenal accents* of these notes (Lerdahl and Jackendoff 1983). Phenomenal accents arise from changes in surface properties of music such as pitch, duration, loudness and timbre. For instance, a long note is usually perceived as more accented than a short one (Parncutt 1994).

A further factor that affects the salience of pulse sensation is the *pulse period*. According to a number of studies (Fraisse 1982; Parncutt 1994; van Noorden and Moelants 1999), the most salient pulse sensations have a period of approximately 600 msec, the region of greatest salience being between 400 and 900 msec. One should note that this range of periods corresponds roughly to the speed of some basic human activities such as heartbeat, locomotion, and infant sucking.

In much Western music, the perceived pulses are often hierarchically organised, and consist of at least two simultaneous levels, whose periods have an integer ratio. This gives rise to a percept of regularly alternating strong and weak beats, a phenomenon referred to as *metre* (Cooper and Meyer 1960; Lerdahl and Jackendoff 1983). In Western music, the ratio of the pulse lengths is usually limited to 1:2 (duple metre) and 1:3 (triple metre). It must be noted, however, that this kind of hierarchical organisation of pulses does not exist in all music. Examples of types of music that do not possess such metrical structure are found in Norwegian Hardanger fiddle music, Lappish *yoiks*, West African polyrhythmic percussion music and Eastern European *aksak* dance music. Electronic music allows the possibility to explore non-standard and alternative metrical structures which diverge from Western common practice to a wider sphere of human music-making, whilst still grounded in psychological percepts.

The perception of pulse and metre has been modelled computationally using a variety of different approaches (Gouyon and Dixon 2005; Collins 2006; also see chapter 10). Fig. 12.4 displays the response of a model of resonating oscillators to an excerpt of music. Whenever a note onset occurs, oscillators that are in synchrony with it are excited (Large and Kolen 1994; Toiviainen 1998).

Conclusion

This chapter has reviewed some of the important aspects of music perception and cognition that can be regarded as useful for gaining better understanding of the perception of electronic music. Although the different musical elements have been discussed separately, their interaction may also play a role in music perception, though this aspect has been less studied.

Most of the processes discussed in this chapter are relatively low-level ones. Higher-level aspects of cognition that are relevant include the perception of form and the effect of memory. In general, these processes are less well understood than the low-level ones. This is due to the increased effect of individual and cultural background on these processes and the consequent larger inter-individual variation.

Obviously, due to space limitations, many of these aspects have been described on a rather superficial level. To gain a better understanding of these processes, the interested reader is directed to works such as Deutsch (1999), Snyder (2000), and McAdams and Bigand (1993). Perceptual principles discussed in this chapter are intimately related to many issues that arise in other chapters of this companion.

Electronic music provides a range of devices for the production of sound and musical material that is far more extensive than that available with more traditional instruments. While these tools facilitate versatile expression of musical ideas, they also make it possible to produce musical material that exceeds the capacity of the human perceptual system. Therefore, for an electronic musician, being aware of the capabilities and limitations of human auditory processing is crucial for efficient communication of musical ideas and exploration of new grounds.