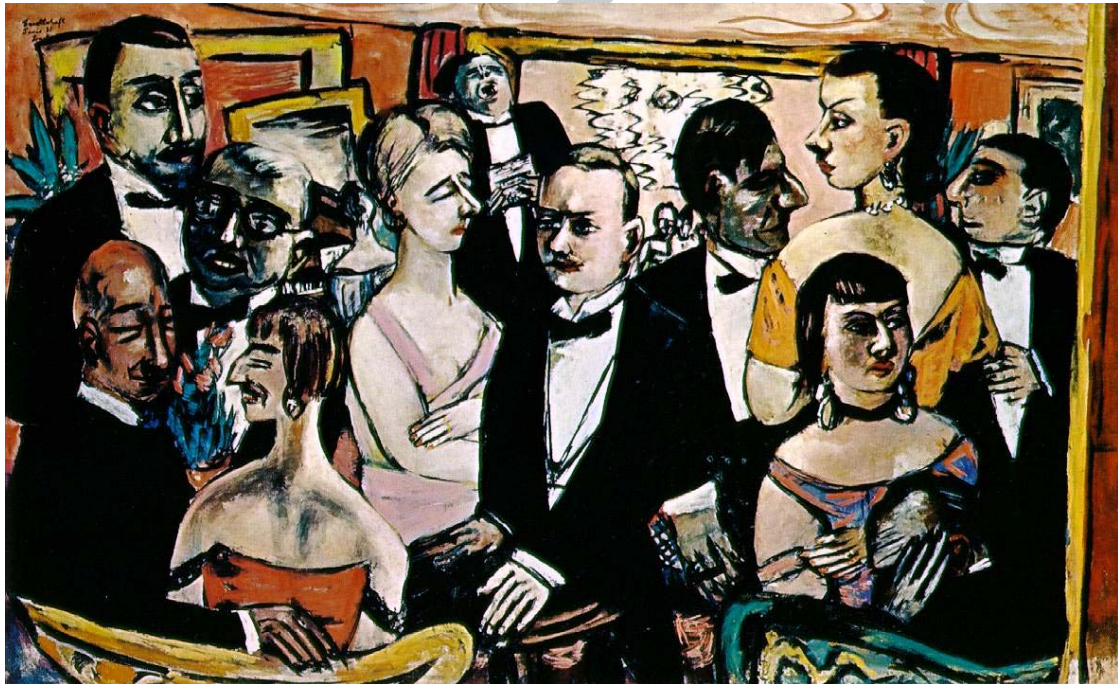


Sensation, Perception, Action – An Evolutionary Perspective

Chapter 8

Hearing 2: Complex Auditory Information



'Paris Society', Max Beckmann, 1931, Guggenheim Museum, New York

Hearing is so immensely important for humans because of its role in understanding spoken language. The eerily voiceless characters in Beckmann's painting allude to the fact that there are more crucial aspects to communication than auditory perception, but also signify the signal processing challenges in densely populated communication spaces.

Overview

To understand the perception of sound for situations that go beyond simple isolated tones, we need to consider the structure of sound patterns, and systematically describe the performance of hearing in a behavioural and ecological context. Complex sounds are described, and analysed, by means of ‘spectrograms’ which represent the sound intensity in different frequency bands as function of time. This technique provides us with the tools to study the advanced processing of complex acoustic information, such as human speech, laughter, or musical harmony. Spectrograms can be related to the equal-loudness contours that describe the range and perceived loudness of audible tones, in order to assess the human auditory experience – this is of particular importance when we compare the limits of hearing to the range of frequencies and loudness that used for acoustic communication. The spectral composition of human speech is thus discussed in the context of technical communication systems, such as telephones, and of impairments, such as partial hearing loss, and their consequences for organising behaviour. A second area of intense interest since many years is the perception of music, offering some challenges to the scientific mind – many of these questions have prompted new research efforts connected to the development of modern neuroscientific techniques in the last few years. Finally, we look into the question of how three-dimensional space can be explored with hearing, and how acoustic space is represented in the cortex. A particular challenge is the separation of sound sources in cluttered acoustic spaces with multiple complex sources, and the cocktail party effect is a stunning demonstration how efficiently the human sensory system deals with this problem. The chapter ends with a short comment on auditory illusions, using the example of Shepard’s eternally raising tone, which correspond to similar deceptions of the processing mechanisms in the human visual system.

Understanding speech

Hearing has not evolved to pick up simple sounds, or pure tones – these were only introduced in the previous chapter to illustrate the physical properties of sound and the fundamental principles of its encoding. In the real world, more often than not, acoustic events are rather complex combinations of such simple sounds, and the perception of complex sound is the topic of this chapter. Mathematically, according to the so-called ‘Fourier Theory’ (see chapter 3), each and every complex sound event can be composed by superposition of a set of pure tones (which are sinewave functions) with varying

frequency, phase, and amplitude (we saw a simple superposition in the last chapter in figure 7.3). So any acoustic event can be fully described by such a set of frequency components – in reverse, this means that it can be analysed in frequency channels like the neurons in the cochlear nerve tuned to particular frequencies (and the encoding filter banks in digital recorders). It also means that any acoustic stimulus can be generated by a sufficiently sophisticated set of frequency-specific resonators, as can be found in musical synthesizers. Unfortunately for the puzzled scientist, and fortunately for everyone of us as listener, the auditory system is not that simple, and individual tones are not just superimposed, but they interact in perception: this is called a non-linear system, which is much more interesting but also more substantially difficult to understand, and full of surprises. Nevertheless, the best way of describing sound is to show how it is made up from a combination of frequencies. The ‘spectrum’ of a tone, or sound, or noise, or any auditory event can be a simple or a complicated pattern of frequencies and intensities. This spectrum is usually modulated from one moment to the next, and therefore such a frequency composition is best described as function of time, which is called a ‘spectrogram’ (Moore 2003). The simple case of the three musical tones that are forming a chord can be seen in the schematic spectrogram sketched in figure 8.1a. You can see a succession of three different clusters with rising fundamental and harmonic frequency (greylevel corresponding to decreasing intensity of higher harmonics), which correspond to the three tones D, F, A (cf. figure 7.2).

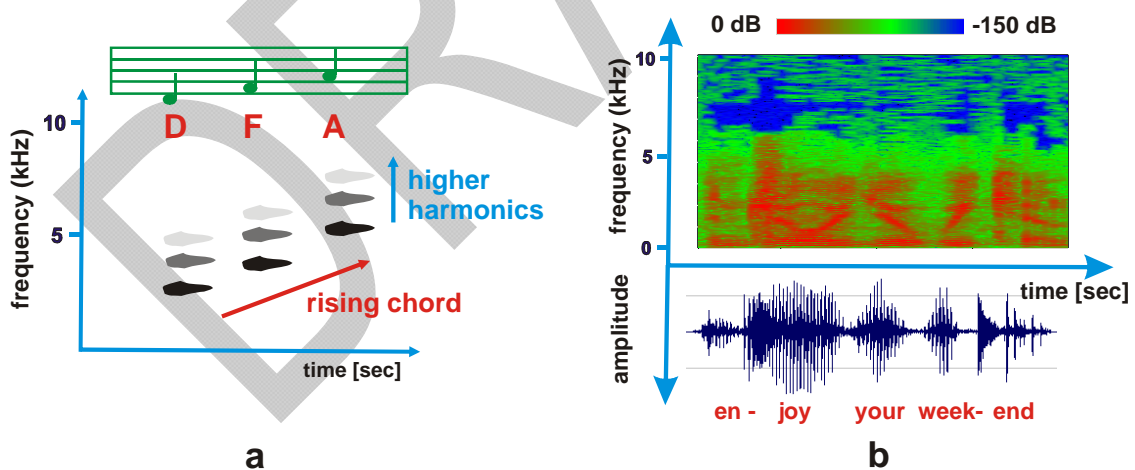


Figure 8.1: Spectrograms. (a) The representation of a simple rising chord; the composition of different frequencies and amplitudes is indicated by different greylevel patches as function of time, with a set of higher harmonics for each of the three tones, D, F, A. (b) Spectrogram of a male voice speaking the phrase 'enjoy your weekend' (intensities of individual frequency components are now shown in colour code); the spectrum contains a wide spread of

frequencies with characteristic rising and falling tones; at the bottom the corresponding waveform envelope is shown (measured as picked up by a microphone), mainly showing the dynamics in the volume of the voice.

Whereas the structure of such a simple chord is clear and well-ordered, we now need to become aware of the complexity of most common natural tones like spoken words, which is illustrated in figure 8.1b. Each syllable generates a complex pattern (spectrum) of frequency and intensity (colour-coded in this figure), which is modulated as function of time – the so called ‘dynamics’. At the bottom of figure 8.1b you can read the phrase that led to this particular spectrogram and the waveform (Plack 2005), which is the sound pressure track as it would be picked up with a microphone – perhaps you realize how much richer the information is that is given by the full spectrogram (frequency and amplitude composition, i.e. pitch and loudness), as compared to the waveform envelope (sound pressure maxima and minima only, without frequency information). This spectrogram represents the information from which your auditory system extracts the phrase ‘enjoy your weekend’, triggering a hopefully positive mood, and possibly giving you a signal to pack your bag and escape from the lecture hall! Now imagine how difficult it is to recognize the same word generated by different speakers, at different pitch and timbre, with different accents, or to detect the voice of a particular speaker from different phrases. If you ever worked with a voice recognition system, you may be able to appreciate the complexities inherent to this task, which your brain usually masters without any effort. The secret of speech perception lies in breaking down these complex patterns into simple spectrographic elements. The spectral envelopes of peak intensity at multiple frequencies are the ‘formants’ resulting from the characteristic resonance patterns of the human vocal system where the voice is generated. Formants with rising, or falling transitions can be mapped to ‘phonemes’, which are the smallest elements into which spoken language can be segmented. Phonemes then need to be recognized by the auditory system in spite of variability in average pitch, overall duration or intensity (Moore 2003). This closes the loop between speech production, related to the physics of the vocal tract generating elementary sounds with characteristic spectral and temporal envelopes, and speech perception based on neural mechanisms that decompose the resulting sound patterns into meaningful components. It will be interesting to investigate in more detail whether cortical pattern recognition mechanisms underlying such auditory processing resemble higher-level encoding strategies that are well studied in the visual system (King and Nelken 2009).

As can be seen in figure 8.1, the spectrum of human speech covers a certain range of the audible spectrum, and varies in volume to span several equal-loudness contours (cf. figure 7.11b), but is restricted to the centre of the full operating range of the human

auditory system (see figure 8.2). Normal speech only covers a region of the auditory response range approximately between 300 and 5000 Hz and between 40 and 70 dB, slightly different for males and females. It can be argued that the restriction to the centre of the audible range makes speech recognition less sensitive to noise and supports top-down mechanisms of perceiving speech in noise (Nahum et al 2008). How constrained this active range is in humans can be appreciated when we compare it with other biological systems, generating and picking up sounds which we cannot hear at all. Whales and elephants, for instance, are using infrasound – frequencies below 20 Hz – to communicate over large distances (e.g., Payne et al 1986). At the other end of the frequency range, bats, amongst other animals, are using ultrasound – frequencies above 20 KHz – for navigation and foraging (e.g., Neuweiler 1984). Within the limited range used by humans, vowel sounds are mainly found in the lower frequency region, whereas consonants cover almost the entire range (see figure 8.2). The specific location of sounds in the frequency-intensity space can vary for individual speakers, and its significance for speech recognition is a matter of debate, because there is evidence that the temporal envelope of a sound on its own can be very important for understanding spoken language (Shannon et al 1995).

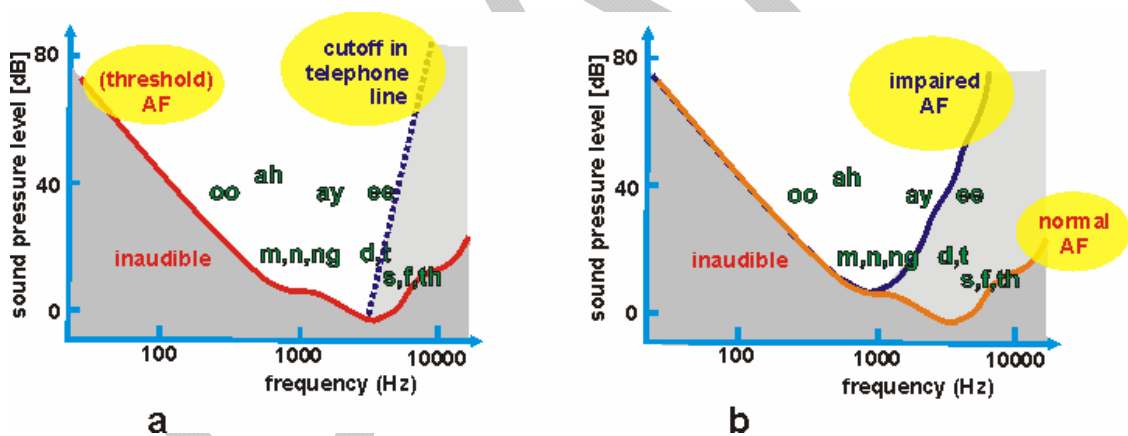


Figure 8.2: Frequency and intensity (sound pressure level) range of normal human speech. The red line shows the audibility function, which indicates detection threshold for pure tones (AF, cf. fig 7.11); the approximate location for vowels and consonants is shown by the green letters. (a) Cutting off higher frequencies (blue broken line) by bandwidth-limited technical devices, such as telephones, usually does not impair speech comprehension substantially. (b) Impairments of audibility, such as age-related loss of higher frequencies (blue solid line), can seriously affect communication.

Arguably of the most serious consequence of hearing impairments is the disruption of the ability to communicate effectively. Interestingly enough, some technical systems such as telephones restrict the range of frequencies to be communicated, and cut off the upper part of the spectrum that is crucial for detecting some consonants, such as 's', 'f', or 'th' (see figure 8.2a), and do so with minimal effects on speech recognition. Note that for a non-native English speaker, the notoriously difficult pronunciation of 'th' does not seem to matter when talking on the phone. The absence of disruptive effects in this context demonstrates the redundancy of speech communication. Obviously, this redundancy can be used by engineers to minimise the information transmitted by such technical systems whilst retaining the message of the sound that is produced to the receiver, which actually is the reason for limiting the frequency band transmitted through phone lines. Similar to the image compression techniques that were discussed in chapter 2, in modern systems the auditory information is compressed by digital technologies, such as mpeg encoding. This redundancy reduction helps to make better use of communication channels – for instance, moments of silence do not need to be transmitted to your mobile phone – and helps to save money, to the customer or the service provider. Equally, it is used to save memory space on audio devices by means of intelligent encoding of sound tracks, allowing you to store many more tracks on your iPod or any other media player.

A natural context in which auditory information is partially lost, without control of whether this information is redundant or essential, is hearing loss. There are many different types of auditory impairments, apart from complete deafness, some of which have profound impact on communication ability and therefore seriously affect the lifestyle and wellbeing of sufferers. Excessive noise exposure can lead to temporary threshold shifts (auditory fatigue) or permanent (and possibly partial) deafness, as discussed in chapter 7. The most common hearing impairment is presbycusis, the selective loss of high-frequency sensitivity with age (Patterson et al 1982). A typical pattern of presbycusis is shown in figure 8.2b, indicating how the sensitivity for many consonants and some vowels is lost, which reduces speech recognition substantially. Careful evaluation by the audiologist is required to identify the special needs of such a patient and to develop a treatment plan to minimize the impact of such a condition – sadly, presbycusis is not reversible because it is related to mechanical ageing of the inner ear. Because younger people are more sensitive to high frequencies, targeted sounds can be used that are inaudible for older folk, including special ring-tones for mobile phones that due to their high frequency open an exclusive communication channel for students but are kept from their teachers, or rather questionable attempts to utilize acoustic repellants against young ASBOs. A particularly annoying and mysterious hearing impairment is tinnitus: sufferers perceive continuous or intermittent humming or ringing in their ear, usually only in one, which cannot be stopped or suppressed. Only after long

exposure some patient learn to ignore the tone, and at some point starts losing any sensitivity in the affected frequency range, because they develop selective deafness for this tone. This sound is not just imagination; in some cases it can be picked up by a sensitive microphone placed in the ear canal, which suggests that in this condition the inner ear begins to oscillate autonomously and transmit sound into the outer ear (oto-emissions), but we are far from understanding the condition or developing a remedy or cure (Jastreboff and Hazell 1993). Given the consequences of any of these conditions for communication and social life, lot of research is targeted at the development of hearing aids. The diagram in figure 8.2 also makes clear that it is not sufficient just to increase the volume (as was done in early hearing aids), because for wide regions the threshold has been raised to unattainable values. More annoyingly, general amplification will intensify the noise as much as the tones which should be heard. What is needed is an intelligent device that amplifies and suppresses selectively, according to the needs of the individual patient – it seems to be difficult to match the sophistication of the human ear. For some forms of hearing impairments related to inner ear malfunction, cochlear implants – which basically replace the sensory transduction process by an electronic acoustic sensor that is connected to the auditor nerve – has been one of the more successful techniques to partially restore hearing function (Moore and Shannon 2009).

You may like it – or not: Music!

There are few sensory events that affect us so strongly and comprehensively as music. Why are these complex acoustic stimuli perceived so immediately, and why are they so closely connected to emotional responses (see figure 8.3)? More basically, what makes a sound pleasant? And what makes a sound unpleasant, or even aversive? Pleasant and aversive sounds are decoded effortless and fast, so do humans from all educational and cultural backgrounds have the same emotional responses to different categories of sound? At least within Western society, there seem to be some acoustical events which we all agree sound horrible: just think of finger nails scraping blackboards, or the squeaking of grinding disks (or the drill of your dentist)! The universality and the biological function of rapid, reliable categorising aversive sounds (Czigler et al 2007), often consistently between individuals, is still a matter of scientific debate (Cox 2008). On the other hand, the positive emotions are usually associated with experiencing music, which sometimes can elevate into near-mythical quality. The astronomer Johannes Kepler (1571 – 1630), when studying the movement of the planets at the nocturnal sky, was searching for the musical harmony of the spheres (Gingras 2003). Music is often believed to exert magical power, and a common thread of narrative is reaching out from antiquity (just think of their story of Orpheus who could bring rocks to life with the sound of his lyre) to modernity about the healing power of music. We may smile when we hear of the “letter

from the Reverend Dr Doddridge at Northampton, ... of one, who had no Ear to Music naturally, singing several Tunes when in a Delirium”, but being reported in the Proceedings of the Royal Society (Doddridge 1753) it certainly was regarded as solid empirical evidence about the very special nature of music. And we should never forget how popular the so-called Mozart effect was for many years, and similar ideas continue to be commercially exploited: There were various claims that listening to the music by Mozart would have surprising effects on various aspects of learning and development of humans (for the scientific end of this hype, see Rauscher et al 1993), which could raise IQ, and there are urban (or rural) myths that the milk yield can be boosted by playing Mozart to cows. Later research has shown that there are little if any consistent effects of Mozart on cognitive performance, although there is evidence that listening to music in general can have some beneficial effects on patients with epilepsy (Jenkins 2001).

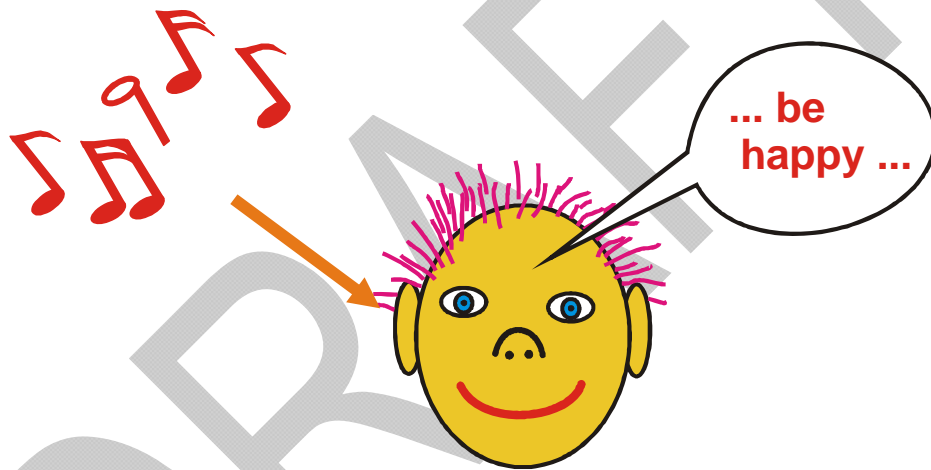


Figure 8.3: Music can have an immediate and substantial effect on our emotional well-being. Understanding the ‘magic power’ of such acoustic events in terms of information processing mechanisms is arguably one of the greatest challenges for researchers studying auditory perception.

Perhaps we should step back from the excitement of such glowing promises and ask more modest questions, which can be addressed with robust scientific methodology. Our knowledge of the basic function of the auditory system should provide us with some tools which could help us to study advanced levels of processing acoustic information, such as the perception of music. When it comes, however, to the scientific study of music, things get immediately complicated again – and we are confronted with a number of problems, such as the wide range of different methods to look at different aspects of music perception and production, the wide range of different music styles and demands, and the

different levels of involvement in music, which need to be investigated in their own right. First of all, music starts with music production, and music production requires a sophisticated coordination between sensory and motor activities, which involves the precise timing of actions in relation to percepts, and complex feedback mechanisms to control the musical output which require thorough knowledge of the instrument. It has been suggested that the perceptual features of music are integrated in the premotor cortex with appropriately timed and coordinated actions (Zatorre et al 2007). Therefore, when investigating the performance of the musician, it is very difficult to separate perceptual from motor control aspects. To complicate things even further, completely different approaches to music production become apparent when the reproduction of extensively rehearsed musical sequences in an orchestra is compared with the free improvisation in Jazz. During improvisation a dissociated pattern of activity is observed in the prefrontal cortex, and the deactivation of some areas and activation of others is interpreted as reflecting internally motivated processes that are not determined directly by the stimulus – patterns that you would expect for spontaneous improvisation. This incoherent activity pattern is accompanied by coherent activity in the sensorimotor area of the neocortex, related to planning and execution of finely controlled motor patterns, and a deactivation of limbic structures that are believed to drive motivation and emotion (Limb and Braun 2008).

One of the key aspects of music perception is related to the easiness and reliability with which we can recognize rhythms and melodies, musical motifs, individual singer's voices, instrumentations, style of individual composers or interpreters, musical genres and historical periods (Longuet-Higgins 1979). Although no performance or recording of a piece of music is exactly the same to any other, we immediately recognise Vivaldi's 'seasons' even if it is enslaved as a ring tone, or identify Paul McCartney's 'yesterday' even if it is weeping from the ageing elevator speaker in a noisy shopping mall. Although some amazing expertise can be developed to pick up the finest nuances in a tune, the basic task of recognising a tune can be accomplished even by Reverend Doddridge's acquaintance, as long as she is in a coma (see above). To classify and identify such a wide range of acoustic objects, the brain needs to combine representations of advanced musical features such as rhythm or tonality into 'conceptual structures' (Longuet-Higgins 1979). An attempt to break down the many processes involved in music perception into subunits that can be studied on their own, can be found in figure 8.4. This 'neurocognitive model' of music perception (Koelsch and Siebel 2005) joins together neuro-anatomical knowledge and electro-physiological responses to stimulation with music to develop a flow diagram of the processing steps that can account for low level feature and pattern recognition, being interpreted as syntax (structure), such as rhythmic grouping, melody, harmony, and link it to high-level mechanisms to extract semantics

(meaning) and emotional content that affects the autonomic nervous system. Although the arrangement of processing steps in this model is informed by the anatomical structure of the brain and by the timing of neural activity, the lines between the various boxes shown in figure 8.4, i.e. the connectivity of processing units, may appear a little strenuous and unspecific, as if each box is connected almost to each other, suggesting that the logic of sequential steps in a processing chain is not always clearly defined.

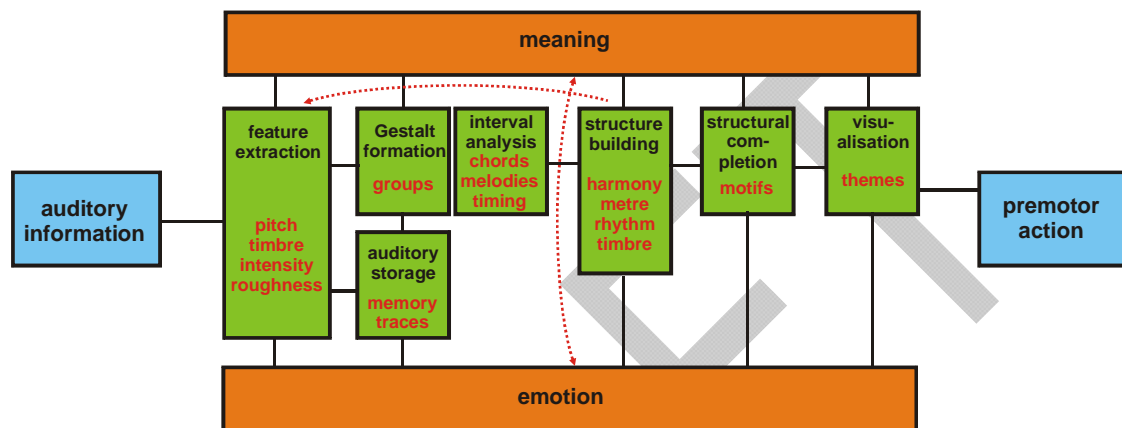


Figure 8.4: Neurocognitive model of music perception, adapted from (Koelsch and Siebel 2005), illustrating an attempt to break down the holistic experience of music into separate but interconnected processing units.

Looking into the functionality of any of the processing boxes shown in figure 8.4 we find some answers to the challenging questions about the detailed mechanisms, but many more fascinating, and unresolved, problems. Just as an example, what makes combinations of tones sound dissonant or consonant, what is the basis of harmony? There is a long history (and substantial inter cultural differences) to consider the frequency relationships between tone intervals that sound harmonic (see chapter 7), such as the idea that temporal patterns of activations show certain similarities for tones that are perceived as consonant, (e.g., Boomsalter and Creel 1961). Indeed, there can be temporal synchrony between auditory neurons that encode frequencies generated by tone pairs that are in harmonic or octave relationships, but such phase locking should also happen for simple frequency ratios which are not classified as harmonic intervals. An alternative idea is based on the frequency ratios of higher harmonics in naturally produced sounds such as speech, for instance 2:1, 3:2, 4:3, etc., are learnt from early childhood to support pitch detection, and that the familiarity with such frequency ratios determines what is perceived as similar or consonant (Terhardt 1974).

Looking at earlier steps of the music processing chain, it is known that there are comparatively simple grouping mechanisms related to the onset of brief acoustic stimuli, when signals that are repeatedly synchronised are grouped together (Darwin 1997). For the most early processing steps, which includes the perception of pitch, we should briefly discuss a phenomenon that has a particular importance for musicians, the perception of absolute pitch. Whereas everyone is very well able to distinguish the pitch of two tones, and reasonably adept in telling which is the higher one of the two, there are only a few individuals who can perceive the pitch of an isolated tone without any reference. That is the reason why musicians use a resonating instrument with constant pitch, such as a tuning fork, to anchor the tuning of their instrument, or set the initial pitch of their voice. Absolute pitch, however, is the ability to identify a tone by naming or singing it without referring to such an external standard. It is generally believed that this ability is inborn, and there are attempts to unravel the genetic basis of pitch recognition (Drayna et al 2001). On the other hand, there is evidence that there is a substantial involvement of long-term memory with particular importance of early infancy (Levitin and Rogers 2005) and that accuracy of pitch judgement can be improved by practise (Cuddy 1968).

Neuroscience has started to make a valuable contribution to our understanding of how music is perceived and produced. For instance, brain imaging techniques can be used to identify the neural substrate and some processing aspects of musical. Alternatively, by studying the impact of brain damage to the abilities of musicians, such as Maurice Ravel (Sergent 1993), the neuro-anatomy of musical function can be traced and related to perceptual and cognitive networks. The interaction of syntactical structures and semantic content in the processing of music (Koelsch and Siebel 2005) links music closely to language, and in tonal languages, such as Mandarin or Cantonese, pitch is not only used to alter the emotional vein of spoken words, but their meaning. Based on these observations, one could speculate about its evolutionary relationship – is music a form of communication that has evolved from speech, has language developed from basic musical utterances, or can we regard music and language independent developments that share some parts of the cortical hardware and some modules of information processing? The close link of music to the emotional content of vocal expressions, and the functional interactions between music and speech are indirectly demonstrated by the observation that the perception of emotional content in speech is facilitated in practicing musicians (Strait et al 2009). With growing level of expertise, musicians show stronger and stronger responses in subcortical structures which are involved in the communication of emotional states, in line with behavioural observations that musical training facilitates the processing of emotional contents of acoustic signals. Many aspects of music production and perception have been studied in their relation to cortical function with a range of imaging methods, such as the homogeneity of the spatial distribution of EEG activity

during listening to music (Bhattacharya and Petsche 2001), the overlap and separation of music and language in cortical processing (Steinbeis and Koelsch 2008), or the relationship between musical imagery and expertise (Herholz et al 2008). But it is perhaps the contribution of subcortical structures to music processing that is most surprising. Wong et al (2007) studied the encoding of pitch in the brainstem and found that this is more robust for musicians than for controls, which could be interpreted as top-down effects on the early neural representation shared by music and speech, from the cortex to the brainstem, which in turn could explain why musicians appear to be so good at learning languages!

Auditory space

To complicate things further for the brain that has to make sense of complex acoustic patterns, sounds are not coming out of the nowhere, but are generated and localised in the world around us - we need to consider acoustic space. Just as visual stimuli arise from a three-dimensional space, and objects can be placed at (almost) any location in this space, sound can originate from arbitrary locations in the very same three-dimensional space. In contrast to the visual system, which starts with two-dimensional images from which three-dimensional space is reconstructed (see chapter 5), the auditory system has to build a spatial representation from much poorer information: each ear samples sound at a single point in space and therefore essentially is a one-dimensional sensor handling with an information stream that does not contain any spatial information. The most important – but not the only – feature of hearing that allows us to explore auditory space is the use of two ears, which sample sounds from two separate positions in space (Plack 2005; Schnupp and Carr 2009). Comparing the sound sequences entering the two ears can be used to determine the location of a single sound source, and consequently to separate several simultaneous sound sources at different locations in the environment (see figure 8.5). In technical systems, such as recordings of concerts or sound tracks for movies, multiple microphones are positioned at various locations to recreate spatial sound effects. Additionally, there is a fundamental difference in the use of temporal information. Sound is a phenomenon that essentially requires time to be detected, because the most elementary information, the tone, is carried by waves, which are temporal modulations of air pressure, and the dynamics of these waveforms are crucial for the separation, grouping and recognition of acoustic objects such as phonemes in speech. In contrast, temporal change is not a prerequisite for visual perception, we have no difficulty to see static images that are flashed very briefly.



Figure 8.5: Auditory space. When you close your eyes and listen to your environment you can pick up many sounds as illustrated schematically in this sketch, the engine of a scooter, the rustling of leaves in the wind, the chopping of a helicopter, the tapping of a man bouncing a ball, the song of birds, and so forth. How does the brain recognise these sounds and localise them in space? As suggested by this image, and the instruction to close your eyes, interaction with other sensory systems such as vision supports auditory perception, and we will return to this issue in chapter 11.

This cursory overview of the information processing issues related to auditory space raises a number of questions, such as: can we hear in two, or three dimensions? how well can the auditory system localise objects? what are the limits of hearing several objects at the same time? Sound localization is based on spatial cues generated by the way in which sounds interact with the ears and the head, depending on the direction of the source (Moore 2003). Combining the information from both ears by ‘binaural processing’ is the most important mechanism to find horizontal position, or azimuth, of a sound source (Middlebrooks and Green 1991), which can make use of two different cues. (i) The head creates an acoustic shadow for sound waves, which leads to higher sound intensity in one ear than in the other (see figure 8.6a). At high frequencies, this difference in sound pressure can reach levels around 20 dB, and humans can use such interaural intensity differences (IID) to localize the azimuth of a sound source (Fedderson et al 1957). (ii) At low frequencies another mechanism is needed: the different distance between the sound source to the right and left ear, respectively, creates a difference in the arrival time of the sound wave (see figure 8.6b). From such interaural time differences (ITD), the human ear can detect delays between the two ears in the range of 0.01 – 0.6 msec, which in air

corresponds to a difference of about 3 mm between the distance between the sound source and the two ears (Stevens and Newman 1934). Both of these mechanisms critically depend on the separation of the two ears (about 15 cm), creating two distinct sampling points for sound waves. Note the similarity between auditory stereo (this name sounds familiar - your stereo system has two speakers!) and stereovision as discussed in chapter 5 (compare the geometry sketched in figure 8.6b and figure 5.6). The smallest angular difference between two positions of a sound source in a horizontal plane, or the ‘minimum audible angle (MAA), that can be achieved with these two mechanisms can be as small as a few degrees.

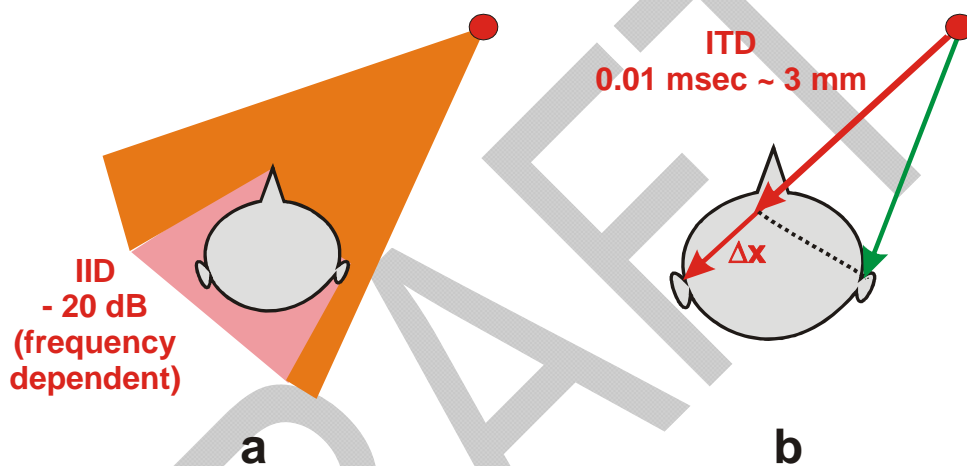


Figure 8.6: Binaural direction cues that can be used for localisation of sound in a horizontal plane. (a) Depending on the location of a sound source, the intensity reaching the two ears can differ considerably, generating an interaural intensity difference (IID). (b) The location of a sound source determines the distance to from the source to each ear (Δx), resulting in interaural time difference (ITD) between the two ears.

Neither IID nor ITD can be used to detect the vertical position, because all points in space that project with a similar angular difference on the two ears share a similar sound pressure and time difference. For estimating elevation, the pinnae are crucial (Middlebrooks and Green 1991), as can be easily demonstrated by the observation that performance is substantially reduced when using earphones. In more formal experiments, the sound reaching the two ears from loudspeakers at various elevations were recorded from the inside of the ear canals and then played back through headphones, and the specific spectra of these sounds allowed participants to identify where the loudspeakers had been positioned in the recordings, almost as accurately as in listening to the

loudspeakers directly (Butler and Belendiuk 1977). Furthermore, these experiments demonstrate that the individual shapes of the pinnae are important – performance was reduced when a participant heard the sounds recorded from someone else's ears. Humans are also able to estimate the distance of a sound source under some conditions, although little is known about the underlying mechanisms. There is evidence that the familiarity with the sound source, i.e. expectations about intensity and spectral composition, plays a crucial role in solving this task (Middlebrooks and Green 1991). In many of these more challenging situations strategies of active exploration are employed, by moving the head – this is a very simple trick used in many sensory systems to increase the number of spatial locations at which information is sampled (this point will be taken up more systematically in chapter 12).

This variety of different mechanisms to extract particular spatial components of sound sources is combined in the brain to generate a coherent auditory representation of auditory space in three dimensions, which is needed to navigate and locate objects in space when the visual system is unreliable or unavailable, for instance in the dark. Whereas most of the basic mechanisms underlying such location are well understood, there are still a number of open questions that need to be answered before we can appreciate comprehensively how spatial hearing works. One example of such problems is how the auditory system is adjusted to changes in the layout of the sensor – during development from infant to adult the size and shape of the head and ears changes considerable, which requires the brain to recalibrate its analysis algorithms. We now have a growing body of evidence that the capacity for recalibrating auditory localization continues well into adult life and involves a number of brain structures (King et al 2001), although we also need to be aware that adjusting to a hearing aid, for instance, needs a lot of time and patience. When studying such difficult questions, it is once again very helpful to look at other biological systems. In the case of spatial hearing, major progress was made from investigations of barn owls, which have evolved highly specialised hearing functions that enables them to hunt in the darkness using acoustic information (Konishi 1986).

So far, we have only considered the localisation of isolated sound sources in space. In real life we are surrounded by multiple sound sources, as sketched in figure 8.5, which need to be separated and grouped for perception. Perhaps one of the most astonishing phenomena is the ability to separate individual voices in a noisy crowded room, such as classroom during break, or a pub on Friday afternoon. The 'cocktail party effect' refers to the fact that, at least for younger people, it is easy to single out a particular conversation or one particular voice from the noisy background whilst ignoring other voices and conversations (Cherry 1953). After the previous sections, we should be puzzled how can this be achieved, because the wild mixture of wavefronts that is hitting the ears at the

same time has an overwhelming complexity. This is similar to trace individual ripples on the surface of a pond in heavy rain back to individual raindrops! Perhaps less surprising than the effect itself is the fact that it is not unique to humans but can be observed in other animals like penguins living in large colonies (Aubin 1998). Identifying individual sounds is a fundamental problem for sensory information processing that goes along with social lifestyle. The core of the cocktail party problem is masking (see chapter 7): the detection of a tone is impaired if another tone or noise is presented at the same time. Because masking depends on similarity in frequency and proximity in space, the process used to separate sound sources in space into individual streams can be described as binaural unmasking. If spatial distance or difference in frequency between competing sound sources increases, separation becomes easier. Because both of these cues are contained in the combined binaural information, sound sources can be separated by the interaction between the signals from the two ears. But this approach is only capturing the first processing steps, and it is clear that a range of high-level effects, like attention (Moray 1959), sequential grouping (Darwin 1997), the familiarity of a voice, its accent, the content of messages, or the language used (Arons 1992) contribute to the solution of this information processing task. Furthermore, in challenging situations like such cluttered acoustic scenes, sensory fusion (see chapter 12) is becoming increasingly important, with visual cues like facial movements or body language supporting the separation of conversations and voices.

Auditory illusions

Finally, let us briefly return to the topic of illusions, which was so helpful to demonstrate in the visual domain the mechanisms underlying sensory information processing. Are there auditory illusions? The simple answer is yes, but many of these are not as obvious as in the visual system. There are discussions, whether humans can experience auditory size constancy solutions, in which the relationship between perceived size and perceived distance determines what you hear, and it has been suggested that there is an auditory analogue to the Moon illusion (von Békésy 1949). Perhaps the most famous auditory illusion is Shepard's eternally rising tone (Shepard 1964). When a set of individual tones is continuously shifted to higher frequencies, whilst their amplitude is modulated by a constant envelope (see figure 8.7a), the overall average frequencies remains roughly constant, although the individual tones are rising. In such a conflict between the local change of frequencies and the global constancy of frequencies, human participants perceive a sound with a frequency that does not stop rising – this eternally rising tone is an acoustic object that resembles the impossible geometric objects, which we know from visual illusions like the eternally rising staircase (see figure 8.7b) that demonstrate the

conflict between a three-dimensional object and two-dimensional representation (cf. figure 5.2a).

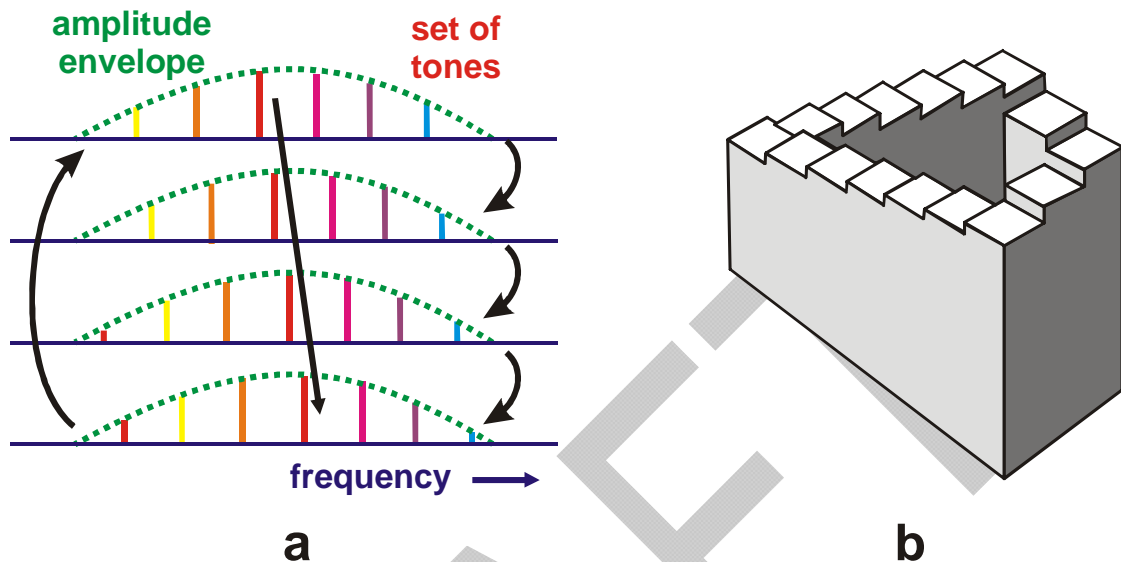


Figure 8.7: Shepard's eternally rising tone. (a) When a set of tones with different, constantly rising, frequencies (vertical lines with different colours) is modulated in amplitude by a constant envelope (green dotted line), a single sound is perceived, which appears to never stopping rising in frequency. (b) The visual analogue of this illusion is an impossible object, the eternally rising staircase.

Take home messages

- spoken words generate complex patterns of frequency and intensity, which can be described and analysed as spectrograms that display frequency-intensity distributions as function of time
- speech covers a considerable range of sound intensities and frequencies in the centre region of the audible spectrum; this finding helps us to appreciate the effects of natural and technical hearing impairments on communication
- sound can be localized by calculating intensity and phase differences between the two ears sound, and some subtle changes in the spectral properties of sound picked up by the outer ear

- in cluttered acoustic spaces auditory scenes, auditory localization and the separation and grouping of sound sources is particularly challenging, but still possible as demonstrated by the cocktail party effect
- can we generate auditory illusions? Yes, we can!

Discussion Questions

- Discuss the causes and consequences of hearing impairments.
- How are the mechanisms underlying speech perception related to the auditory processing of music ?
- Describe the two main auditory mechanisms of spatial localisation
- What is the auditory processing problem commonly referred to by the ‘cocktail party effect’?

References

- Arons B, 1992 "A review of the cocktail party effect" *Journal of the American Voice I/O Society* **12** 35-50
- Aubin T, 1998 "Cocktail-party effect in king penguin colonies" *Proceedings of the Royal Society B: Biological Sciences* **265** 1665-1673
- Bhattacharya J, Petsche H, 2001 "Universality in the brain while listening to music" *Proceedings of the Royal Society B: Biological Sciences* **268** 2423-2433
- Boomsalter P, Creel W, 1961 "The Long Pattern Hypothesis in Harmony and Hearing" *Journal of Music Theory* **5** 2-31
- Butler R, Belendiuk K, 1977 "Spectral cues utilized in the localization of sound in the median sagittal plane" *The Journal of the Acoustical Society of America* **61** 1264-1269
- Cherry E C, 1953 "Some experiments on the recognition of speech, with one and with two ears" *Journal of the Acoustical Society of America* **25** 975-979
- Cox T J, 2008 "Scraping sounds and disgusting noises" *Applied Acoustics* **69** 1195-2004
- Cuddy L L, 1968 "Practice Effects in the Absolute Judgment of Pitch" *The Journal of the Acoustical Society of America* **43** 1069-1076
- Czigler I, Cox T J, Gyimesi K, Horváth J, 2007 "Event-related potential study to aversive auditory stimuli" *Neuroscience Letters* **420** 251-256
- Darwin C J, 1997 "Auditory grouping" *Trends in Cognitive Sciences* **1** 327-333
- Doddrige D, 1753 "Postscript of a Letter from the Rev. Dr. Doddrige at Northampton, to Mr. Henry Baker F. R. S. of One, Who Had No Ear to Music Naturally, Singing Several Tunes When in a Delirium" *Philosophical Transactions (1683-1775)* **44** 596-596
- Drayna D, Manichaikul A, Lange M d, Snieder H, Spector T, 2001 "Genetic Correlates of Musical Pitch Recognition in Humans" *Science* **291** 1969-1972
- Feddersen W E, Sandel T T, Teas D C, Jeffress L A, 1957 "Localization of High-Frequency Tones" *The Journal of the Acoustical Society of America* **29** 988
- Gingras B, 2003 "Johannes Kepler's Harmonice mundi: A " Scientific" version of the Harmony of the Spheres" *JOURNAL-ROYAL ASTRONOMICAL SOCIETY OF CANADA* **97** 228-231

- Herholz S C, Lappe C, Knief A, Pantev C, 2008 "Neural basis of music imagery and the effect of musical expertise" *European Journal of Neuroscience* **28** 2352-2360
- Jastreboff P J, Hazell J W P, 1993 "A neurophysiological approach to tinnitus: Clinical implications" *British Journal of Audiology* **27** 7-17
- Jenkins J S, 2001 "The Mozart effect" *JRSM* **94** 170
- King A J, Nelken I, 2009 "Unraveling the principles of auditory cortical processing: can we learn from the visual system?" *Nat Neurosci* **12** 698-701
- King A J, Schnupp J W H, Doubell T P, 2001 "The shape of ears to come: dynamic coding of auditory space" *Trends in Cognitive Sciences* **5** 261-270
- Koelsch S, Siebel W A, 2005 "Towards a neural basis of music perception" *Trends in Cognitive Sciences* **9** 578-584
- Konishi M, 1986 "Centrally synthesized maps of sensory space" *Trends in Neuroscience* **4/86** 163-168
- Levitin D J, Rogers S E, 2005 "Absolute pitch: perception, coding, and controversies" *Trends in Cognitive Sciences* **9** 26-33
- Limb C J, Braun A R, 2008 "Neural Substrates of Spontaneous Musical Performance: An fMRI Study of Jazz Improvisation" *PLoS ONE* **3** e1679
- Longuet-Higgins H C, 1979 "The perception of music" *Proceedings of the Royal Society London* **B 205** 307-322
- Middlebrooks J C, Green D M, 1991 "Sound Localization by Human Listeners" *Annual Reviews in Psychology* **42** 135-159
- Moore B C J, 2003 *An Introduction to the Psychology of Hearing* (San Diego: Academic Press)
- Moore D R, Shannon R V, 2009 "Beyond cochlear implants: awakening the deafened brain" *Nat Neurosci* **12** 686-691
- Moray N, 1959 "Attention in dichotic listening: Affective cues and the influence of instructions" *The Quarterly Journal of Experimental Psychology* **11** 56-60
- Nahum M, Nelken I, Ahissar M, 2008 "Low-Level Information and High-Level Perception: The Case of Speech in Noise" *PLoS Biology* **6** e126
- Neuweiler G, 1984 "Foraging, Echolocation and Audition in Bats" *Naturwissenschaften* **71** 446-455

- Patterson R D, Nimmo-Smith I, Weber D L, Milroy R, 1982 "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold" *The Journal of the Acoustical Society of America* **72** 1788
- Payne K B, Langbauer W R, Thomas E M, 1986 "Infrasonic calls of the Asian elephant (*Elephas maximus*)" *Behav.Ecol.Sociobiol.* **18** 297-301
- Plack C J, 2005 *The Sense Of Hearing* (Mahwah, NJ: Lawrence Erlbaum Associates)
- Rauscher F H, Shaw G L, Ky K N, 1993 "Music and spatial task performance" *Nature* **365** 611-611
- Schnupp J W H, Carr C E, 2009 "On hearing with more than one ear: lessons from evolution" *Nat Neurosci* **12** 692-697
- Sergent J, 1993 "Music, the brain and Ravel" *Trends in Neurosciences* **16** 168-172
- Shannon R V, Zeng F G, Kamath V, Wygonski J, Ekelid M, 1995 "Speech Recognition with Primarily Temporal Cues" *Science* **270** 303
- Shepard R N, 1964 "Circularity in Judgments of Relative Pitch" *The Journal of the Acoustical Society of America* **36** 2346
- Steinbeis N, Koelsch S, 2008 "Comparing the Processing of Music and Language Meaning Using EEG and fMRI Provides Evidence for Similar and Distinct Neural Representations" *PLoS ONE* **3** e2226
- Stevens S S, Newman E B, 1934 "The Localization of Pure Tones" *Proceedings of the National Academy of Sciences of the United States of America* **20** 593-596
- Strait D L, Kraus N, Skoe E, Ashley R, 2009 "Musical experience and neural efficiency - effects of training on subcortical processing of vocal expressions of emotion" *European Journal of Neuroscience* **29** 661-668
- Terhardt E, 1974 "Pitch, consonance, and harmony" *The Journal of the Acoustical Society of America* **55** 1061-1069
- von Békésy G, 1949 "The Moon Illusion and Similar Auditory Phenomena" *The American Journal of Psychology* **62** 540-552
- Wong P C M, Skoe E, Russo N M, Dees T, Kraus N, 2007 "Musical experience shapes human brainstem encoding of linguistic pitch patterns" *Nat Neurosci* **10** 420-422
- Zatorre R J, Chen J L, Penhune V B, 2007 "When the brain plays music: auditory-motor interactions in music perception and production" *Nat Rev Neurosci* **8** 547-558