Auditory Perception

Andrzej Miśkiewicz

Contents

1. The auditory system – sensory fundamentals of hearing

1.1. Basic functions of the auditory system

Hearing is the process of receiving and perceiving sound through the auditory system – the organ of hearing. This process involves two kinds of phenomena: (1) *sensory effects* related with the functioning of the organ of hearing as a receiver of acoustic waves, transducer of sound into neural electrical impulses, and a conduction pathway for signals transmitted to the brain; (2) *cognitive effects* related with a variety of mental processes involved in sound perception. The present chapter discusses the functioning of the auditory system as a sensory receiver of acoustic waves and the cognitive processes of sound perception are presented in chapter 2.

The auditory system of most mammals is made up of two basic parts known as the *peripheral* and the *central auditory system*. The peripheral system comprises a symmetric pair of organs including the ear and the auditory nerve – a bundle of nerve fibers that carry auditory information between the ear and the brain. The central auditory system consists of the auditory areas in the auditory cortex and a network of neural pathways and nuclei that connect the peripheral system with the auditory cortex. The basic function of the peripheral auditory system is to conduct the sound waves from outside the ear to its inner parts, detect the physical characteristics of the waves and encode this information into neural electric potentials transmitted to auditory areas in the auditory cortex . Electrical stimulation of the auditory cortex gives rise to auditory sensations evoked in response to the physical stimulation of ear. The central auditory system also serves a regulatory function in auditory transduction of sound in the peripheral system.

1.2. Structure of the ear

Outer ear

The ear is a complex sensory organ divided into three structures: the *outer ear*, the *middle ear*, and the *inner ear* (Fig. 1.1). The outer ear is made up of three basic parts: the *pinna* (also called the *auricle*), the *auditory canal* (the *auditory meatus*), and the *tympanic membrane* (the *eardrum*).

The main function of the outer ear is collection of sound waves from the environment and transmission to the eardrum which separates the outer ear from the middle ear. As a result of its non-symmetrical shape and various ridges and folds, the pinna modifies the spectrum of incoming sound. The spectral shaping caused by the pinna depends on the direction of the incoming sound and enhances our ability to localize sounds. The shape of each individual's pinna puts a distinctive imprint on the sound spectrum.

The auditory canal is a resonant cavity and causes an enhancement of sound within a frequency range of 1500-7000 Hz, with a maximum of about 10 dB at about 3 kHz. This spectral shaping boosts our hearing sensitivity in that frequency range and improves the intelligibility of speech sounds.

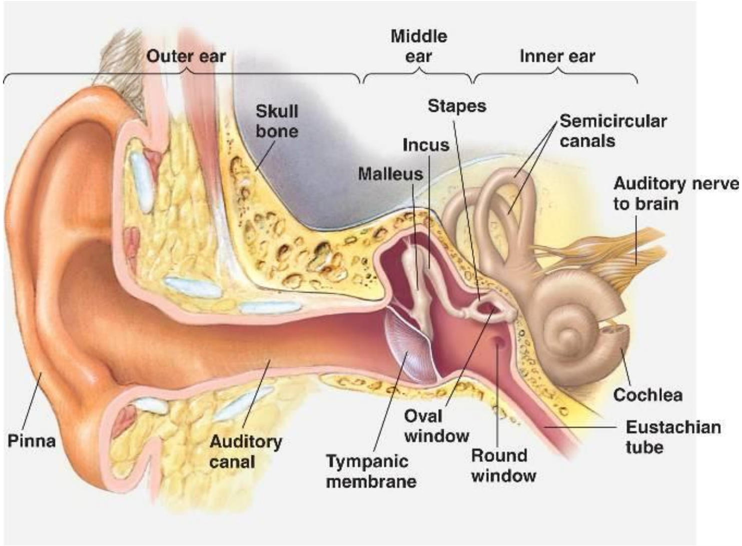

Fig. 1.1. The main structures of the human ear. From *Principles of Biological Science,* http://bio1152.nicerweb.com/Locked/media

Middle ear

The middle ear is an air-filled cavity housing three tiny ossicles : the *malleus* (also called the *hammer*), the *incus* (*anvil*), and the *stapes* (*stirrup*) (Fig. 1.2). The middle ear conducts the vibrations of the tympanic membrane to the *oval window*, which forms the boundary between the middle and inner ear.

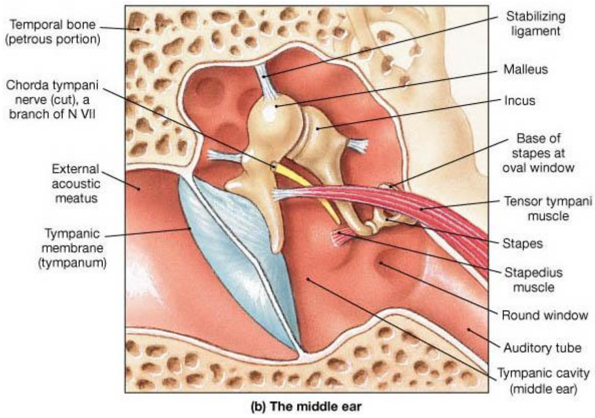

Fig. 1.2. The ossicular chain in the middle ear. From droualb.faculty.mjc.edu

The middle ear acts as a mechanical impedance-matching device or transformer that increases the energetic efficiency of sound transmission from the eardrum to the oval window. Its energetic efficiency is a combined result of the difference in effective areas of the eardrum and the stapes footplate and—to a lesser extend—of the lever action of the ossicular chain.

The *Eustachian tube* seen in Fig. 1.2, also known as the *auditory tube*, connects the middle ear with the nasopharynx. Its main function is to equalize pressure between the middle ear cavity and the atmosphere.

Another function of the middle ear is protection of the inner ear from damage that might be caused by high-intensity sounds. The middle ear contains two muscles: the *tensor tympani* attached to the malleus and the *stapedial muscle* attached to the stapes. When the eardrum is stimulated with high-intensity sounds the muscles contract and reduce the amplitude of the vibrations transmitted to the inner ear. This contraction is called the *acoustic reflex*. The activation of the acoustic reflex is too slow to attenuate impulsive sounds, such as gunshots and percussion instrument sounds therefore impulsive sounds are particularly hazardous to the auditory system and may cause damage to the inner ear. Acoustic reflex is also activated before vocalization and reduces the audibility of speech and other self-generated sounds.

Transmission of sound through the eardrum and the ossicular chain, in the way described above, is called *air conduction*. Sound travels to the inner ear also through the bones of the skull. This effect, known as *bone conduction*, does not play any important role in typical conditions of hearing, as the threshold of hearing in bone conduction is by about 50-60 dB higher than in air conduction. The spectral characteristics of bone conduction and air conduction pathways are different therefore we hear our voice differently to how others hear it. This difference is apparent when we listen to recordings of our own voice. When our voice is captured with a microphone the bone conduction pathway is eliminated and does not influence the sound timbre.

Bone conduction is used for sound transmission in certain kinds of hearing aids and in audio communication systems designed for use in special conditions, for example, under water. In such systems the sound conduction path bypasses the outer and the middle ear and the vibrations are conducted to the bones of the skull through two transducers placed on two sides of the skull, usually on the mastoid part of the temporal bone. A communication system based on bone conduction may also operate in the opposite way. A vibration sensor may be used as a contact microphone to detect the bone vibrations of a talker and convert them into an electrical audio signal transmitted through a communication system.

Inner ear

The *inner ear* is placed in the temporal bone and contains the *semicircular canals* and the *cochlea* (Fig. 1.3). The semicircular canals are the sensory organ of balance and do not play any role in hearing. The auditory portion of inner ear is the cochlea. The cochlea is a bony structure forming a coiled tunnel, about 35 mm long, shaped like the spiral shell of a snail. The part of the cochlea near the oval window is called the *base* and the other end is known as the *apex*.

The cochlea is divided along its structure into three adjacent canals filled with fluids, separated from each other by membranes. Two of the canals, called the *scala vestibuli* and the *scala timpani* are filled with perilymph. The third canal, placed between them, known as the *cochlear duct,* contains endolymph. The three canals are seen on the cross-section of the cochlea shown in Fig. 1.3.

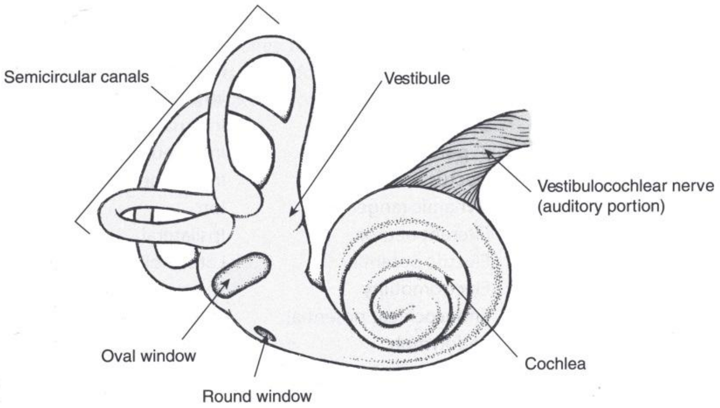

Fig. 1.3. An illustration showing the two main structures of the inner ear: the semicircular canals and the cochlea. From Emanuel and Letowski (2009).

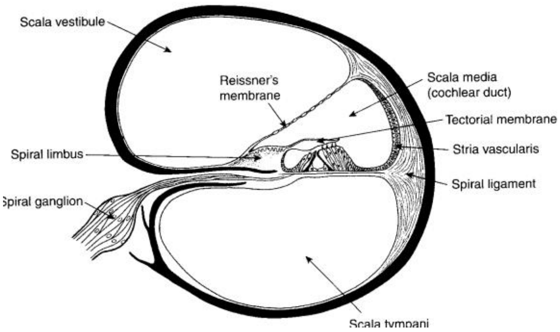

Fig. 1.4. Cross-section of the cochlea. From Emanuel and Letowski (2009).

The two cochlear fluids have different ionic composition: the perilymph is rich in sodium ions and has a low concentration of potassium ions and the endolymph has a high concentration of potassium ions and contains virtually no sodium. This ionic imbalance provides an energy supply for the mechanism of triggering electric neural potentials in the inner ear, in response to sound waves. The two kinds of fluids are kept separate by two membranes along the cochlea: *Reissner's membrane* and the *basilar membrane*. Scala vestibuli and scala timpani are connected at the apical end of the cochlea by an opening, known as the *helicotrema*.

The movement of the stapes footplate, back and forth against the oval window, elicits pressure waves that travel through the fluid along the cochlea. The pressure difference across the cochlear chambers causes the basilar membrane to vibrate and produces a pressure wave travelling along the cochlea, from the base to the apex (see Fig. 1.5).

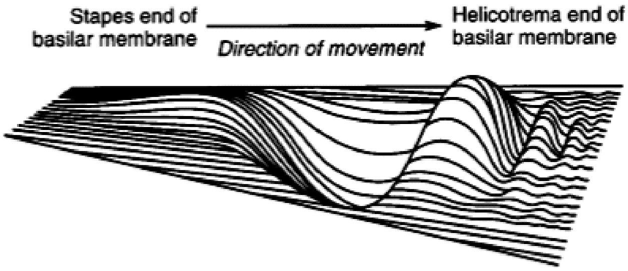

Fig. 1.5. A schematic illustration of an uncoiled basilar membrane showing a travelling wave produced by the pressure exerted by the oval window on the fluid in the scala vestibuli.
From www.ssc.education.ed.ac.uk

The basilar membrane supports a very delicate and complex structure of the *organ of Corti*. The organ of Corti is the part of the inner ear in which vibrations of the tympanic membrane, conducted to the inner ear through the ossicle chain and the oval window, are transduced into electrical neural impulses, or action potentials. The organ of Corti contains support cells and thousands of tiny hearing receptors called the *hair cells*, arranged in bundles, in several rows along the basilar membrane. Above the stereocilia lies a gelatinous layer called the *tectorial membrane*.

A view of the arrangement of hair cells on the basilar membrane is shown in Fig. 1.6. The inner hair cells are arranged in one row and the outer hair cells in up to five rows. In total, the human organ of Corti contains about 3500 inner hair cells and 12000 outer hair cells. From each hair cell projects a bundle of hairs, or *stereocilia*. The number of stereocilia protruding from a hair cell is approximately 40 in inner hair cells and 140 in outer hair cells. The stereocilia of the outer hair cells form V-shaped clusters and those of the inner hair cells are arranged in formations resembling a flattened letter U (Fig. 1.6).

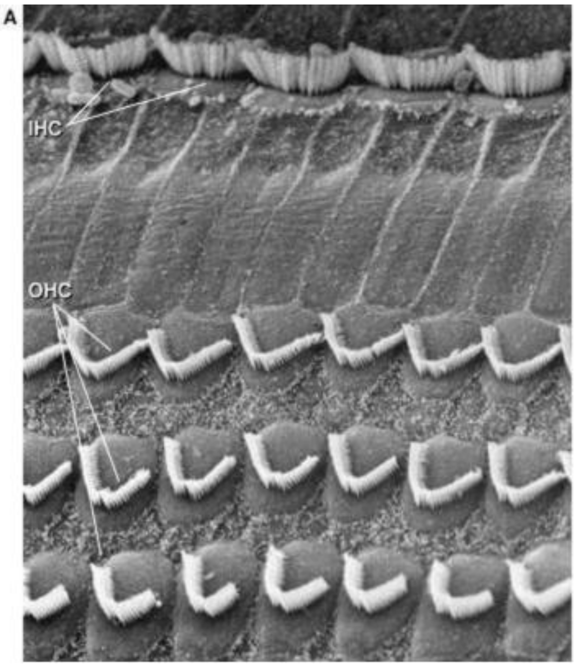

Fig. 1.6. Scanning electron micrograph showing the arrangement of the inner and the outer hair cells on the basilar membrane. From Harrison et al. (1988).

Hairs cells are connected with the brain through a network of neurons. Depending on their function, the neurons are divided into three groups: *afferent neurons*, *efferent neurons*, and *interneurons*. Afferent neurons transmit information from the organ of Corti to the brain. Efferent neurons carry information in the opposite direction, from the brain to the organ of Corti. The function of interneurons is to provide communication between afferent and efferent pathways. The bundle of nerve fibers connecting the cochlea with the central auditory system is called the *cochlear nerve* or the *auditory nerve*. The cochlear nerve is part of the *vestibulocochlear nerve.* The other part of the vestibulocochlear nerve, the *vestibular nerve*, transmits balance information from the semicircular canals to the brain.

Tops of the stereocilia contact the tectorial membrane. Vibrations of the basilar membrane create shearing force which causes the stereocilia to bend. The bending of the hair cells caused by the upward and downward displacement of the basilar membrane is shown in Fig. 1.7. When the basilar membrane moves towards the tectorial membrane the deflection of the stereocilia opens ion channels and cause an influx of positively charged ions from the endolymph to the hair cell body. The inner hair cells react to this influx of ions and change of electrical potential by generating a neural impulse. When the basilar membrane moves in the opposite direction, the ion channels close and the cell return to its resting potential.

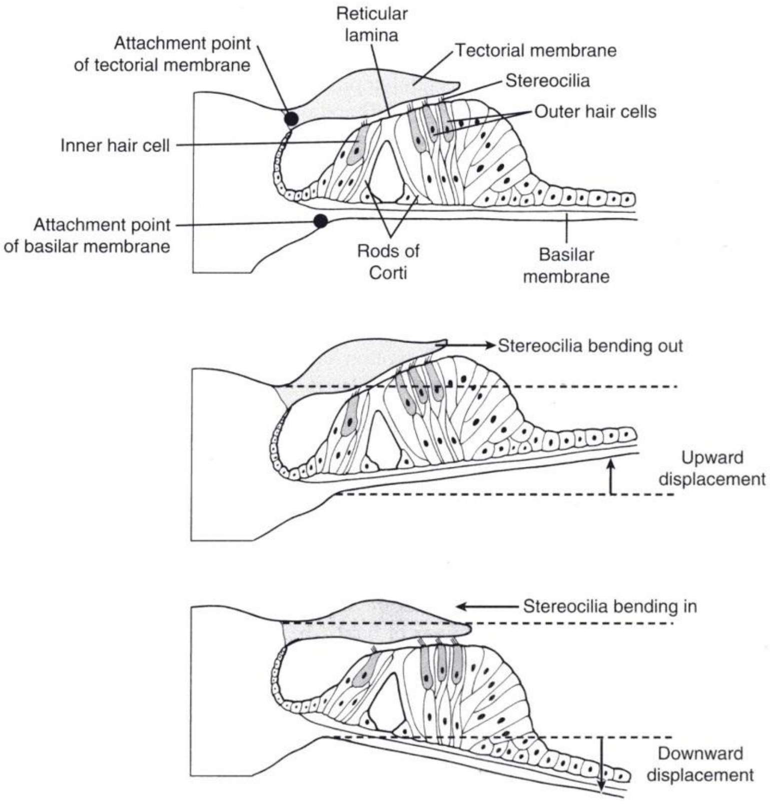

Fig. 1.7. Bending of the hair cells caused by upward and downward displacement of the basilar membrane. From Emanuel and Letowski (2009).

The outer hairs react to the influx of ions and change of electrical potential by expansion and compression of their length, an effect known as *electromotility*. The inner hair cells are part of a physiological mechanism that amplifies and sharpens the response of the organ of Corti to acoustic signals. This mechanism, called the *cochlear amplifier*, is controlled by neural impulses transmitted from the brain through the efferent pathway system.

1.3. Sound analysis in the cochlea

Detection and coding of signal frequency

The manner in which the basilar membrane responds to the movement of the oval window surface is the key to understanding the mechanism of sound spectrum analysis in the cochlea and the principles of encoding the vibration frequency of the membrane into electrical neural impulses. The mechanic properties of the basilar membrane vary along its length, from the base to the apex: at the base the membrane is narrower and stiffer compared to the part at the apex. As a result of this difference the position of maximum deflection of the membrane differs along its length, depending on the frequency of stimulation applied to the oval window.

Figure 1.8 shows the envelopes of vibrations produced during stimulation of the cochlea by pure tones with various frequencies. The place at which the deflection of the membrane reaches its maximum along the membrane's length gradually shifts from the base to the apex as the stimulation frequency decreases.

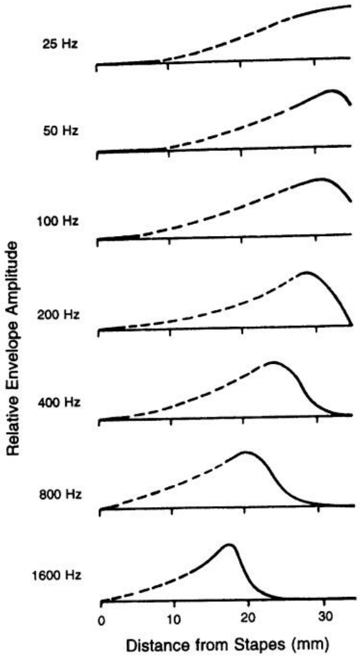

Fig. 1.8. Amplitude envelopes of vibration of the basilar membrane for tones of various frequencies. From Moore (2013).

The place at which stimulation at a given frequency produces maximum deflection of the basilar membrane in called the *characteristic frequency* of the membrane. The signal filtering mechanism, based on the properties of the basilar membrane is supported by electrical filtering of the neural impulses triggered by the basilar membrane vibrations. The auditory nerve fibers distributed along the basilar membrane are frequency selective and are most sensitive at the frequency corresponding to the characteristic frequency of the place at which they are connected with the hair cells on the basilar membrane. The frequency selectivity of auditory fibers sharpens the signal filtering capability of the basilar membrane.

Theories of hearing

The mechanism of detection and encoding the frequency of stimulation in the cochlea is explained by two theories, known as the *theories of hearing*. One of those theories is called the *place theory* and the other one the *periodicity* or the *temporal theory*. The mapping of stimulation frequency to place of maximum deflection, described above, is the main postulate of the place theory. The place theory further assumes that the information on the place of basilar membrane deflection, representing the frequency of stimulation, is conveyed to the brain owing to the *tonotopical organization* of the cochlear neural pathways and the auditory cortex. Owing to such an organization of neural transmission, the information from the cochlea is projected to different parts of the auditory cortex, depending on the place from which it originates on the basilar membrane.

The other theory of frequency coding, the periodicity theory, assumes that the information about stimulation frequency is conveyed by the temporal patterns of neural spikes transmitted to the brain. The basic evidence on which this theory is based in the effect of phase locking. Phase locking means that electrical spikes are triggered in a neural fiber not on every cycle of periodic stimulation but they always occur at the same phase of the waveform. The time intervals between the spikes convey therefore the information about the frequency of stimulation.

The place and temporal theories were presented in the literature as alternative, competing explanations of the mechanism of frequency coding in the cochlea. At present most authors agree that both theories do not exclude each other but describe complementary elements of the complex process of the spectral analysis of acoustic signals in the cochlea.

The auditory filter bank

A large body of evidence from psychoacoustic experimental studies demonstrates that the inner ear processes acoustic signals in a way as if it contained a bank of overlapping band-pass filters. The auditory filter bands are referred to as *critical bands*.

The concept of the *auditory filter bank* was introduced to explain the results of experiments in which the listeners detected a pure tone in the presence of masking noise with increasing bandwidth. The noise band had a constant spectrum level and was centered at the frequency of the tone detected in the experiment. A typical outcome of such an experiment is shown in Fig. 1.9. When the masker bandwidth increases, the detection threshold for the tone is gradually elevated, but only until the noise band reaches a certain width. Further increase of the noise bandwidth does not change the threshold for the tone.

According to the assumptions of the auditory filter bank model the tone is masked only by the spectral components of the masker that pass through the auditory filter (critical band) centered at the tone's frequency. The masker spectral components remote from the tone frequency pass through other filter channels and have no effect on the detection threshold.

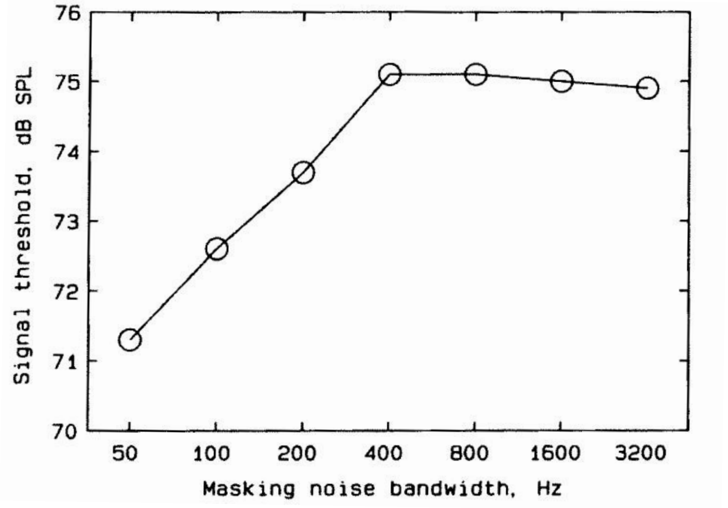

Fig. 1.9. Detection threshold for a 2-kHz pure tone as a function of the bandwidth of a noise masker centered at the tone's frequency. From Moore (2013).

The auditory filter bank model is supported by evidence from experiments concerned with a variety of auditory phenomena. For example, the loudness of noise with increasing bandwidth and constant spectral level remains does not change until the noise bandwidth is less that the critical bandwidth at the center frequency of the noise band. When the noise band is broadened beyond the critical bandwidth the spectral components falling into other bands of the filter bank contribute to the sensation of loudness and the overall loudness of noise increases. (Fig. 1.9).

The bandwidth of the auditory filter, called the *critical bandwidth*, varies depending on the filter's center frequency. Early estimations made in the 1950s suggested that the critical band has a constant width of about 100 Hz up to 500 Hz and is 15-20% of the center frequency at higher frequencies. More recent estimations have demonstrated that the auditory filter bandwidth increases in the whole range of audible frequencies and are between 11 and 17% of the center frequency. The new estimation has been termed the equivalent rectangular bandwidth, ERB.

Coding of sound intensity

When the sound intensity increases the travelling wave produced by the movement of the oval window has a larger amplitude and broader extent along the basilar membrane. The broader extent of the travelling wave results in a larger number of hair cells excited by the stimulus and a larger number of auditory nerve fibers in which firings are triggered. The group of simultaneously firing neurons, i.e., the number of neurons and their location along the basilar membrane, is one element of the neural code that conveys the information about the sound intensity.

Another element of the code is based on the rate of neural firings produced in response to the stimulus. In the absence of stimulation each neuron produces a number of spontaneous firings. The rate of those firings, that is the number of firings per second, is called the *spontaneous firing rate*. The lowest stimulus intensity producing a noticeable change in the firing rate is called the *threshold of the neuron*. When the intensity of a stimulating tone increases above the neuron's threshold the firing rate increases to a level called the *saturation point* and further increase of the tone level above that point does not change the neuron's firing rate. The range between the lowest stimulus intensity producing neural firing and stimulus intensity at the saturation point is called the *dynamic range* of a neuron. The dynamic range of most auditory neural fibers varies from about 25 to about 40 dB.

The auditory nerve fibers may be divided into three types as to their dynamic range of firing: (1) the *high spontaneous fibers* which are more sensitive and saturate at lower stimulus intensities, (2) *low spontaneous fibers*, ale less sensitive and having a higher saturation point, (3) *medium spontaneous fibers*. The three types of neurons, combined with the intensity coding mechanism based on the extent of excitation, make it possible to encode a much larger range of sound intensities than that corresponding to the dynamic range of a single neuron.

1.4. Range of hearing

The range of hearing describes the ranges of frequencies and sound pressure levels of audible tones. The normal range of human hearing is shown schematically in Fig. 1.10. The curve at the bottom part of the graph indicates the general shape of the *threshold of hearing*, the lowest sound pressure level at which a sound may be heard at a given frequency. The human threshold hearing at different frequencies varies over a large sound pressure range. The human auditory system is much more sensitive to tones in the range between 2000 and 5000 Hz than to tones at frequencies below and above that range.

The threshold of hearing varies among individuals. The upper frequency range of hearing decreases as part of natural ageing process during life. The change in the frequency range of hearing is accompanied by an elevation of the hearing threshold which means that the auditory system becomes less sensitive to sounds.

The curve at the top of the graph shows a rough estimate of the *threshold of pain*. The threshold of pain is the boundary sound pressure level beyond which an acoustic wave is no longer perceived as sound and gives rise to non-auditory sensations of tickle and pain.

The range of hearing may also be reduced due to various kind of hearing disorders. A common factor causing a permanent elevation of the threshold of hearing and reduction of the frequency range of hearing is prolonged exposure to high-intensity sounds.

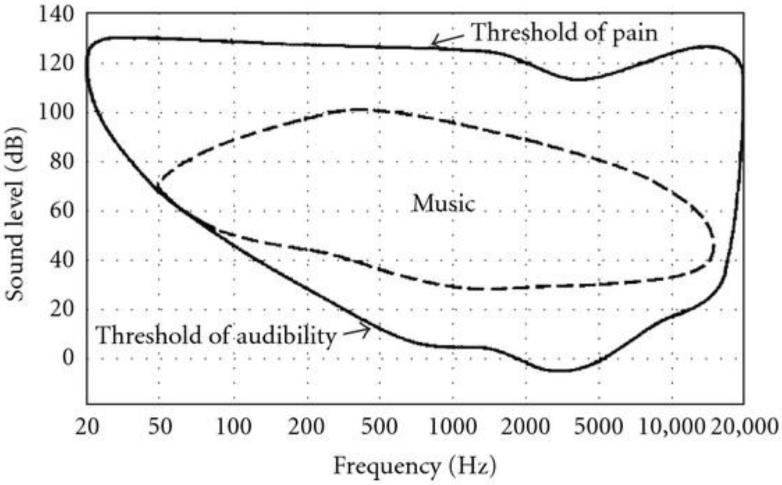

Fig. 1.10. The ranges of tone frequencies and sound pressure levels within which the human auditory system responds to sounds. Adapted from Rossing et al. (2002).

Review questions

1. What are the main parts of the peripheral auditory system called?
2. What is the main function of the external ear?
3. What is the function of the bones in the middle ear?

4. What is the function of the acoustic reflex?
5. What does the term *bone conduction* refer to?
6. What are the main parts of the inner ear?
7. How is the tone frequency detected and coded in the inner ear?
8. What does the concept of the auditory filter refer to?
9. How does the threshold of hearing change with tone frequency?

Recommended further reading

Gulick, W. L., Gescheider, G. A., Frisina, R. D. (1989). *Hearing: Physiological Acoustics, Neural Coding, and Psychoacoustics.* New York – Oxford: Oxford University Press.

Moore, B. C. J. (2013). *An Introduction to the Psychology of Hearing,* sixth edition, Chapter 1, pp. 1–56. Leiden-Boston: Brill.

2. Listening

2.1. The difference between hearing and listening

In Chapter 1 we have presented a general overview of the sensory effects underlying the process of transduction of acoustic waves received by the ear into electric impulses transmitted through the auditory neural pathways to the brain. The present chapter discusses another kind of phenomena involved in the perception of sound, known as cognitive processes. The term *cognitive processes* refers to a variety of mental and behavioral phenomena that operate together in the reception of sensory stimuli from the environment. In the case of the sense of hearing cognitive processes are involved in the capturing of sounds from the environment, memorizing sounds, extracting and processing the information they convey, making conclusions about the environment from the perceived sounds, and translating all those operations into decisions and actions.

The sensory and cognitive auditory processes are distinguished by two terms commonly used in everyday language: hearing and listening. The term *hearing* is usually referred to the sensory part of sound reception, a passive process occurring without any effort of the listener. *Listening* is an active process involving a variety of mental operations conducted in a person's mind. The two kind of auditory processes interact in the reception of sounds, but when we hear a sound, we do not always listen to it.

2.2. Listening tasks

Scientific studies of the perceptual phenomena functioning within the process of hearing have typically a form of psychoacoustic experiments in which the participants are presented with special test sounds and are required to evaluate a given aspect of the auditory sensation or to perform a certain kind of task in response to the sounds. The task performed by the listeners depends on the aim of the experiment. The listeners may be asked, for example, to indicate whether they are able to hear a sound, hear a difference between two or more sounds, or to identify the sound source.

Figure 2.1 shows a classification of basic auditory tasks encountered in psychoacoustic experiments. The classification has a two-level structure; its upper level comprises two categories termed differentiation and categorization.

Differentiation may consist in detection or discrimination of sounds. In a detection task the listener is asked to indicate whether or not he or she is able to hear a sound in a trial or to indicate the observation interval in which the sound is presented. The listener's ability to detect the sound is usually expressed as the detection threshold determined in the experiment. Detection threshold is the lowest sound pressure level at which a sound can be heard in given conditions of presentation. In psychoacoustics detection threshold is also called the hearing threshold or the threshold of hearing.

The purpose of discrimination tasks is to determine the listener's ability to differentiate sounds that are nearly similar. In such tasks the listeners may be asked to indicate whether the sounds presented in a trial are identical or different or to report which sound is more intense in a given perceptual characteristic, for example, has a higher pitch or is louder. Such experiments lead to the determination of the discrimination threshold. Discrimination threshold is the smallest noticeable change in a given physical attribute of sound required for a listener to notice a difference between sounds. Discrimination threshold is also called the difference limen (DL) or the just noticeable difference (jnd).

Categorization, the other of the two main categories of auditory tasks listed in Fig.1, comprises two sub-categories, termed identification and scaling. Identification is the task of assigning nominal labels to sounds. In such a task the listener may be asked, for example, to identify the sound source or a given property of the sound source. Scaling of auditory consists in assigning numbers to sounds in such a way that they reflect the amount of a given perceived attribute of sound. In a scaling task the listener may be asked to judge the amount of a given attribute directly, by assigning numerical values to it or indirectly, by specifying in which of the sounds presented in a trial the studied attribute is more intense. In the latter type of judgment the listener may be asked to indicate, for instance, which sound is louder, or has a higher pitch.

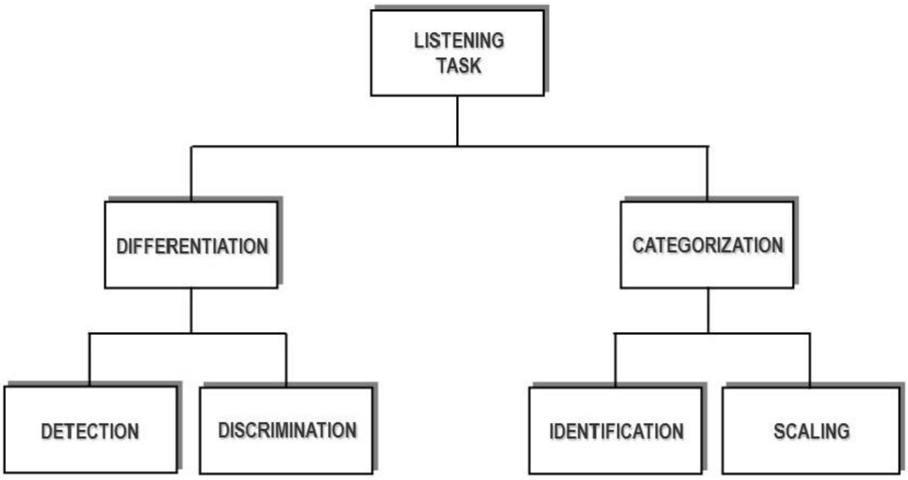

Fig. 2.1. Basic listening tasks in psychoacoustic studies. Adapted from Letowski (1984).

2.3. Cognitive processes in audition

The process of cognition comprises a wide variety of interrelated mental and behavioral phenomena. Most researchers agree that the basic parts of this process are the following:

- memory,
- attention,
- perception,
- thinking,
- problem solving,
- the use of language,
- creativity.

Cognitive processes are studied in relation to various sensory modalities and behavioral effects in psychology, psychophysics, and also in other scientific disciplines. Within the scope of this textbook we shall discuss the basic operation principles of memory and attention. Those processes, often considered the core of cognition, are crucial for an understanding of the phenomena governing the perception of sound within the context of multimedia applications.

2.4. Memory

Memory is a general term given to mental processes whereby information is encoded and stored in the brain for future retrieval. Memory is thought to involve three main stages of information processing:

- encoding – receiving, processing, and combining information,
- storage – creation of information records in memory,
- retrieval - getting back the stored information.

The nature of memory structures and processes has been extensively studied in cognitive psychology. The most widely known model describing the functioning of memory is schematically shown in Fig. 2.2. The model, developed by Atkinson and Shiffrin, postulates that there are three basic types of memory called the *sensory register* (*sensory memory*), *short-term memory* (*short-term store*), and *long-term memory* (*long-term store*).

Sensory register is an ultra-short-term mechanism that holds information received from the sensory organs for less than about 500 milliseconds after stimulation. It acts as a buffer for stimuli received through the senses of vision, hearing, smell, taste, and touch. The sensory register for visual information is called iconic memory, the register for auditory information – echoic memory, and the register for information acquired by touch – haptic memory. Information passes from the sensory register into the short-term memory. The transfer of information to the short-term memory is mediated by the process of attention which filters the stimuli to that which are needed at a given instance.

Short-term memory, also known as working memory, allows recall of information for a period of about 15 seconds to a minute. Duration of short-term memory may be increased through rehearsal, that is repeating the stimulus in mind several times.

The store capacity of short-term memory for a given sensory modality is 7±2 items. The short-term memory capacity may be increased through chunking, that is grouping of a number of items after which they can be remembered as a single item. For example, to facilitate the memorizing of a nine-digit telephone number in Poland we may chunk the digits into four groups: the area code (two digits), the exchange code (three digits), and two two-digit chunks representing the subscriber's number.

Long-term memory (long-term store), the final stage of the model shown in Fig. 2.2, is a permanent memory store and its duration and capacity are practically unlimited.

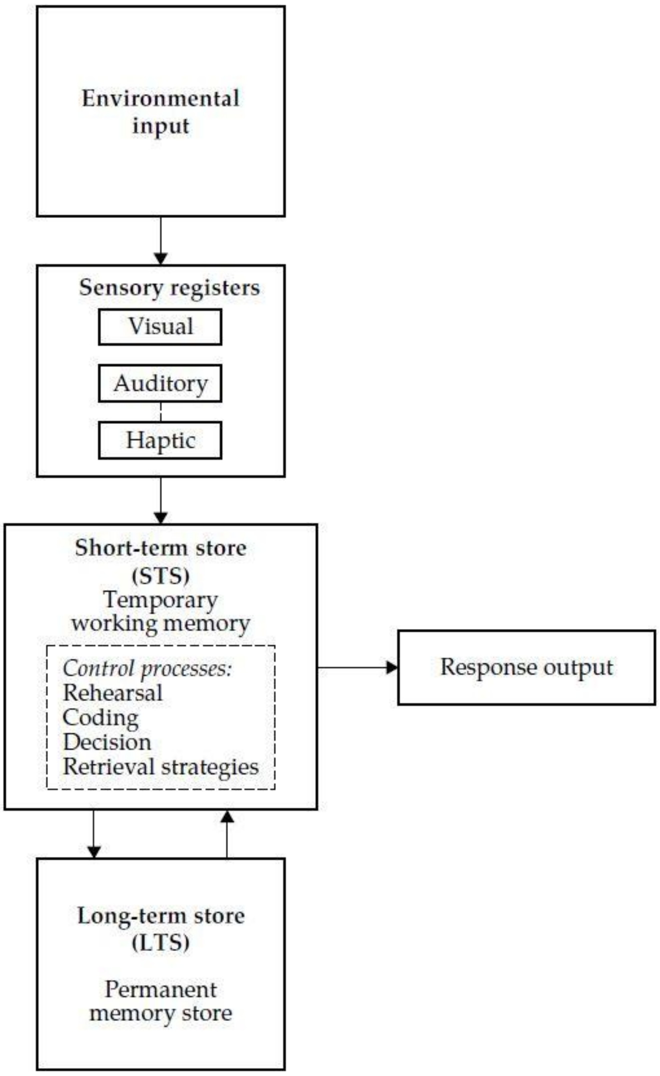

Fig. 2.2. The model of human memory developed by Atkinson and Shiffrin. From Baddeley (2004).

2.5. Auditory attention

Basic functions of attention

Attention is the ability of selectively directing one's mind to a given part or element of information while ignoring other information that might be perceived. Attention may be either conscious or unconscious; one may voluntarily attend to an object or stimulus, but a sudden appearance of another intense stimulus may attract attention to it without any conscious will or action of the perceiver.

Attention has been extensively studied in psychology, mainly in relation to the sense of vision. Although the processes of visual and auditory attention have in many aspects similar nature, they differ in an important feature: attending to a visual stimulus requires from a person to turn the eyes to a given location in space whereas auditory attention is independent of the position of the listener's ears and operates within a 360° range around the listener. Owing to this capability auditory attention plays a very important role in the spatial orientation in the environment and also serves the function of a very efficient safety warning system.

The scope of scientific explorations of human attention encompasses a large variety of perceptual and behavioral phenomena. Researchers in psychology usually distinguish the following general aspects of attention:

selectivity – the act of focusing on a particular object in the perceptual field while simultaneously ignoring other perceivable stimuli,

perceptual search – scanning of the perceptual field to select its parts for further processing,

vigilance – the ability to maintain prolonged, concentrated attention,

shifting of attention (**switching between tasks**) – the ability to switch attention between multiple, simultaneously performed tasks.

Coctail party effect

The most widely known in the literature example of the operation of auditory attention has been termed the *coctail party* effect. Coctail party effect is the ability to focus attention on a particular talker while ignoring the voices of other persons within a jumble of conversations. The studies of this effect were inspired by communication problems commonly faced in air traffic control towers in early 1950s. At that time, air traffic controllers listened to messages received from the pilots over a loudspeaker and to communicate with the pilots they had to separate each message from other intermixed voices and background noise.

The coctail party effect was first described by Colin Cherry, a British cognitive scientist who introduced the technique of dichotic listening for its investigation. In a dichotic task the listener is presented simultaneously with two different messages played separately to the left and to the right ear, through a pair of earphones and has to tell the content of the message he/she is instructed to attend to.

Experiments conducted with the use of dichotic listening have demonstrated that listeners can easily tell the content of the auditory message they are asked to attend to, but are unable to recall the unattended messages. For example, in one of Cherry's experiments the listeners were asked to repeat aloud the words presented in one ear while other words were simultaneously presented to the other ear. At the beginning of the task the words presented in both ears were spoken in English but at some point the language has changed from English to German in the unattended ear. In conditions of dichotic speech presentation the listeners did not realize the change of language as their attention was focused on the sounds heard in only one ear.

Selectivity of auditory attention

Early studies of auditory attention speech Most studies concerned with auditory attention were concerned with the selectivity consisted in the measurement of signal detection and signal discrimination in conditions when the listener is focused to a particular perceived characteristic of sound – pitch, loudness, or the sound's spatial locus. The most widely studied aspect of attention was its dependence on the pitch of a sound. Experiments have demonstrated that the detection threshold is usually considerably higher when the listener does not know at what pitch the test sound is to be presented than when the listener does know the pitch.

Figure 2.1 shows an example of the effect of directing the listener's attention to a particular pitch in a tone detection task. In each trial the listener first heard a cue tone the frequency of which was varied in various conditions in the experiment. The cue tone was followed by a 1-kHz probe tone presented in one of two observation intervals indicated by a visual signal on a computer screen. The listener had to tell in which interval he or she heard the probe tone. The level of the probe tone was the same in all conditions, chosen such as to obtain about 90% of correct answers when the cue was presented at the probe tone's frequency.

The graph in Fig. 2.1 shows the percentage of correct answers as a function of the frequency of the cue tone. The data indicate that detection threshold increases as the frequency of the cue tone moves away from the probe tone's frequency.

Review questions

1. What is the difference between hearing and listening?
2. What listening tasks are called detection, discrimination, identification, and scaling?
3. What is the duration of information storage in the sensory register?
4. How can we increase the duration of short-term memory?
5. What memory process is called chunking?
6. Give an example of auditory attention selectivity in everyday life.

Recommended further reading

Eysenck, M. W., Keane, M. T. (2015). *Cognitive psychology*, seventh edition, chapters 5-7. Hove, UK: Psychology Press.

Scharf, B. (1998). Auditory attention: The psychoacoustical approach. In: H. Pashler (Ed.) *Attention*, pp. 15–60. Hove, UK: Psychology Press.

3. Perceived characteristics of sound

3.1. Auditory objects, events and images

Acoustic waves that arrive at the listener's ears are produced by different sources and are combined into one sound field in the environment. To identify individual sound sources and the properties of their acoustic environment the auditory system must be capable of extracting the information about the physical characteristics of the waves produced by each source from the cumulative stimulation of the ear.

The result of the physical stimulation of the organ of hearing is referred to as the *sensation of sound*. The noun sensation stresses that sound is understood in this case as an auditory percept evoked in the listener's mind, not as stimulation by an acoustic wave which is a physical phenomenon. We The sensation of sound may also be referred to with the use of other terms, such as auditory sensation, auditory object, auditory event, or auditory image.

In everyday language, *object* means something perceptible by one or more sensory senses, usually a material thing perceived by vision or touch. The word object is often used by analogy in reference to various kind of auditory percepts. The term *auditory object* may denote a single sound source, a group of sound sources perceived as a distinct entity, or a single element of a complex structure of sounds, for example, a melodic sequence forming a separate part in a polyphonic texture in music.

The concept of auditory object differs in its core from visual objects. Auditory objects persist through time, but they do not have a material form. To emphasize this difference some researchers have argued that sounds are *auditory events*. We hear auditory events produced by material sound sources (objects) and those kind of events are bearers of audible features.

The overall mental image evoked by the sounds perceived in a given time interval is also called the *auditory image*. Auditory images are also termed *auditory scenes*, especially in studies of auditory scene analysis aimed at exploring the perceptual mechanisms of extracting individual auditory objects from a complexity of sound waves captured by the listener's ears.

3.2. Perceived dimensions of sound

In everyday life, listening is usually aimed at the following objectives: a) to obtain auditory information needed for situational awareness in the environment, b) to communicate with other people by means of speech and with the use of other kinds of acoustic signals, c) to experience artistic impressions. In such practical instances of listening the main function of the auditory system is to identify the sound sources and their location in space, identify elementary and complex structures of speech or other communication acoustic signals and to receive the artistic messages expressed in the language of a given genre of acoustic art.

We may also listen to sounds in a different fashion, analytically, to determine their component perceived characteristics, aside from identification of the sound source and the information a sound might convey. Such a kind of analytical auditory assessment is made, for example, for the purpose of sound quality assessment of musical instruments, sound recordings and of sound reproduction equipment.

The sensation of sound is multidimensional therefore the listeners use various evaluation criteria and verbal terms to describe the sensation they perceive. The choice of descriptive terms depends on various factors, such as the kind of sounds being evaluated, cultural background of the listeners, their experience in sound quality evaluation, and the purpose of evaluation.

There have been proposed several systems of auditory perceptual dimensions intended for various applications. Figure 3.1 shows a model of an intuitive system commonly used in music and in everyday listening situations. According to this model sounds are described in terms of loudness, pitch, perceived duration, timbre, and spaciousness. The model assumes that loudness, pitch and perceived duration are unidimensional attributes whereas timbre and spaciousness are multidimensional and comprise a number of component dimensions. In further sections of this chapter we shall discuss the dependence of loudness, pitch, perceived duration, and timbre on the physical characteristics of sound. Spatial properties of auditory images will be discussed in the next chapter.

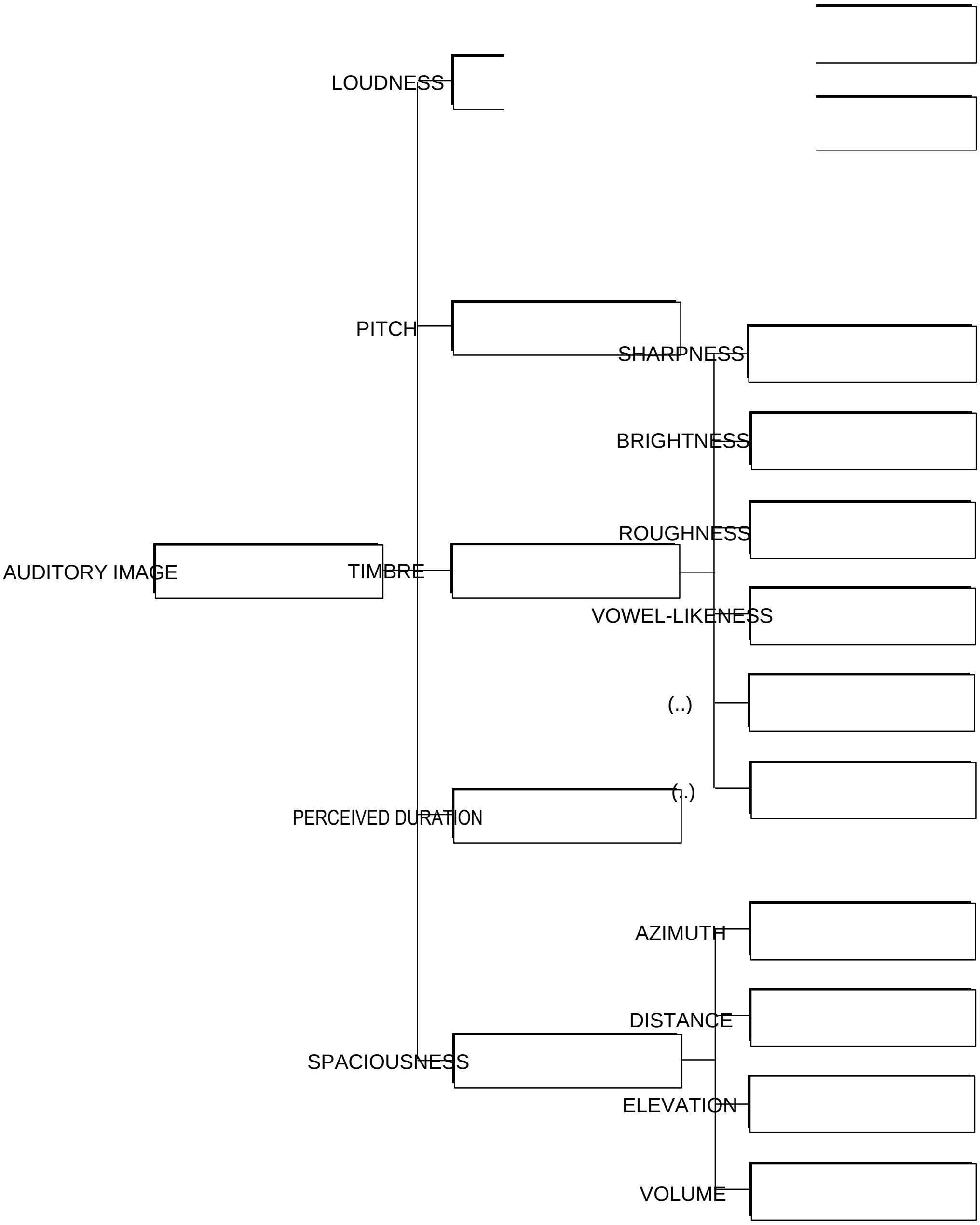

Fig. 3.1. A system of perceptual dimensions commonly used for the description of sound character. Adapted from Miśkiewicz (2002).

3.3. Loudness

Loudness of pure tones

The sensitivity of the ear varies with the frequency of the sound therefore tones of different frequencies, presented at the same sound pressure level, differ in loudness. Figure 3.2 presents the equal-loudness contours for pure tones. The contours show the change in sound pressure level required to maintain equal loudness at frequencies plotted on the abscissa. The bottom curve, drawn with broken line, represents the threshold of hearing.

Loudness of a sound may be quantified by specifying its *loudness level*. The loudness level for a sound is determined by subjective comparison to the loudness of a 1-kHz pure tone. The unit for loudness level is called the *phon*. Loudness level in phons is equal in number to the sound pressure level in decibels of a 1-kHz pure tone perceived to be as loud as the sound being measured.

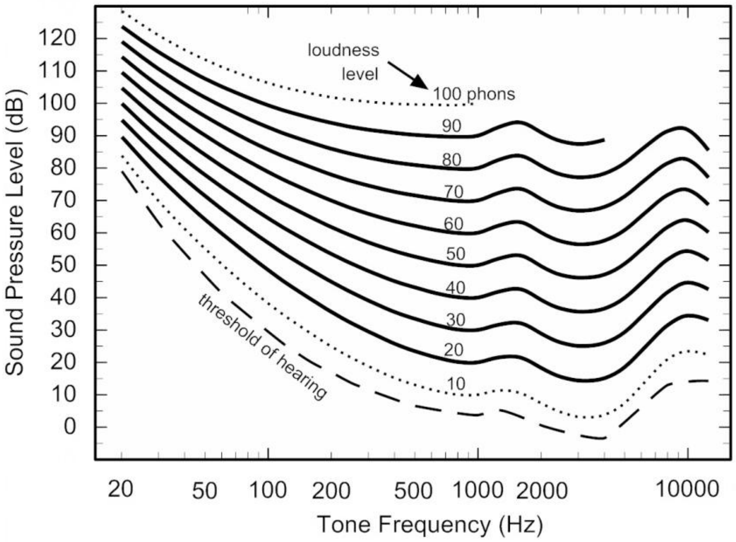

Fig. 3.2. Equal-loudness contours (ISO, 2003).

A noteworthy effect seen in Fig. 3.2. is that the equal-loudness contours are not parallel to one another which means that the loudness of pure tones does not grow as a function of sound pressure level at the same rate at all frequencies. A change by 20 dB in sound pressure level changes the loudness level by 20 phons at 1 kHz and by nearly 40 phons in the frequency region close to the lower bound of hearing.

Another scale used for quantifying loudness in psychoacoustics is the *sone scale*. The scale unit, one sone, has been defined as the loudness of a 1-kHz pure tone presented at a sound pressure level of 40 dB SPL. The sone scale is a ratio-type measurement scale which means that a sound *n* times louder than 1 sone is assigned a loudness value of *n* sones.

The straight line in Fig. 3.3. shows the loudness of a 1-kHz pure tone as a function of sound pressure level. The graph represents sound pressure and loudness in log-log ccordinates: the abscissa is sound pressure level, a logarithmic measure of sound pressure, and the loudness values are plotted on a logarithmic ordinate axis. The function shown in Fig. 3.3 indicates that the loudness of a 1-kHz pure tone doubles for every 10-dB increase in sound pressure level.

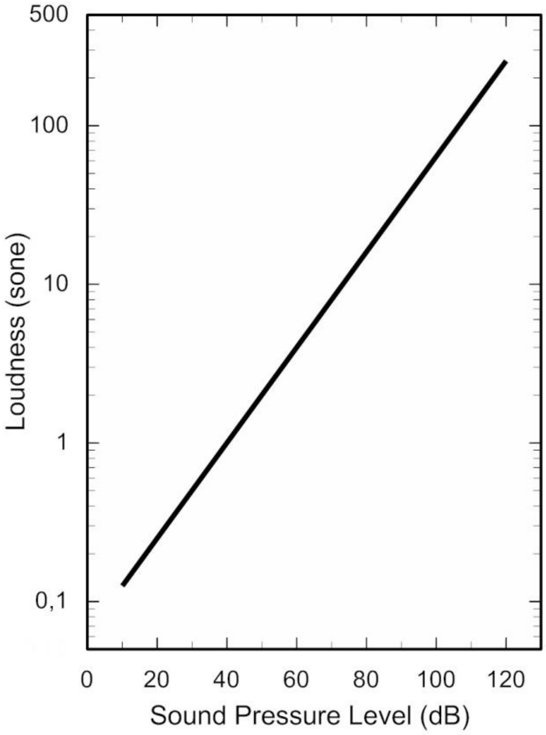

Fig. 3.3. A function showing the relation between the sound pressure level of a 1-kHz pure and its loudness in sones.

Loudness of complex tones and noise: the effect of bandwidth

The loudness of a complex sound (complex tone or noise) increases with the sound's bandwidth although the sound pressure level is held constant. This effect is called *spectral loudness integration* or *spectral loudness summation*. Figure 3.4 shows how the loudness level of a noise band centered at 1 kHz changes as a function of bandwidth. Up to a bandwidth of 160 Hz, corresponding to the so called critical bandwidth, loudness is independent of bandwidth. When the bandwidth of noise is further broadened, beyond the width of the critical band, loudness increases due to the influence of spectral loudness summation.

A similar effect of bandwidth on loudness is also observed in the perception of sounds with line spectra, comprising a number of discrete partials. Loudness of complex tones increases with increasing frequency separation of their spectral components when the tone's bandwidth exceeds the critical bandwidth.

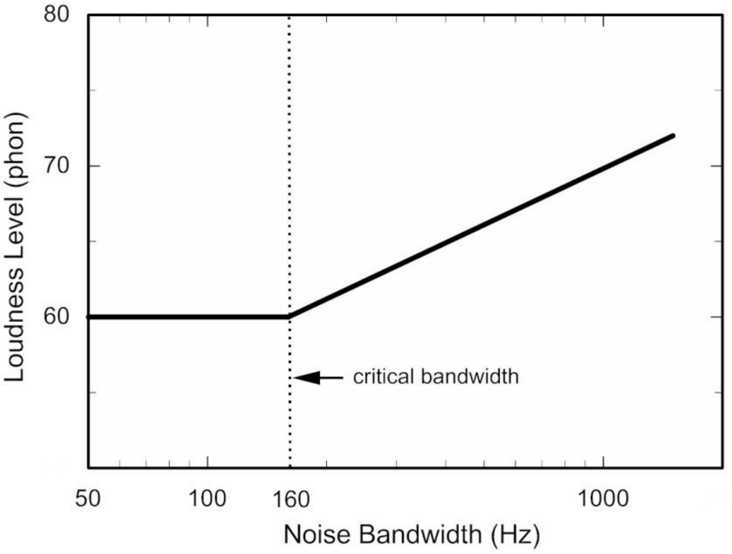

Fig. 3.4. Loudness level as a function of the bandwidth of noise centered at 1-kHz. Adapted from Fastl and Zwicker (2007).

Dependence of loudness upon tone duration

The auditory system integrates energy over time up to about 200 ms. This effect, called *temporal integration* or *temporal summation* has an influence both on the detection threshold and loudness. Detection threshold decreases a rate of about 3 dB per doubling the duration as a tone is lengthened to about 200 ms. A similar effect of temporal integration has an influence on loudness. When a sound is shortened below 200 ms its sound pressure level must be increased to maintain constant loudness.

Loudness under masking

Sounds are usually heard in the presence of other sounds. A sound that affects the audibility of another sound is called a *masker* and the effect of such a sound is called *masking*. Masking is a process by which the threshold of hearing of one sound is elevated by the presence of another sound. Masking may be complete or partial. *Complete masking* occurs when a sound becomes inaudible; *partial masking* raises the detection threshold for the sound and reduces its loudness.

Masking may occur when a sound and a masker are presented together or when they are separated in time. When the sound and the masker are presented at the same time the effect of masker is called *simultaneous masking*. When the sound follows the termination of the masker the masking effect is termed *forward masking*. When the sound is presented before the masker is turned on the masking effect is called *backward masking*.

Figure 3.5 illustrates the effect of a simultaneous masker on the loudness of a 1-kHz pure tone. The graph shows the loudness in sones as a function of the tone's sound pressure level in three conditions: (1) when the tone is heard in quiet, (2) when the tone is heard in the presence of a noise masker and the detection threshold for the tone is raised by 40 dB, (3) when the tone is heard in the presence of a more intense noise masker that raises the detection threshold by 60 dB. The loudness functions plotted on the graph indicate that the reduction in loudness produced by the masker is much greater for low than for high tone levels. At the highest levels shown on the abscissa the presence of the masker affects very little, if at all, the loudness of the tone.

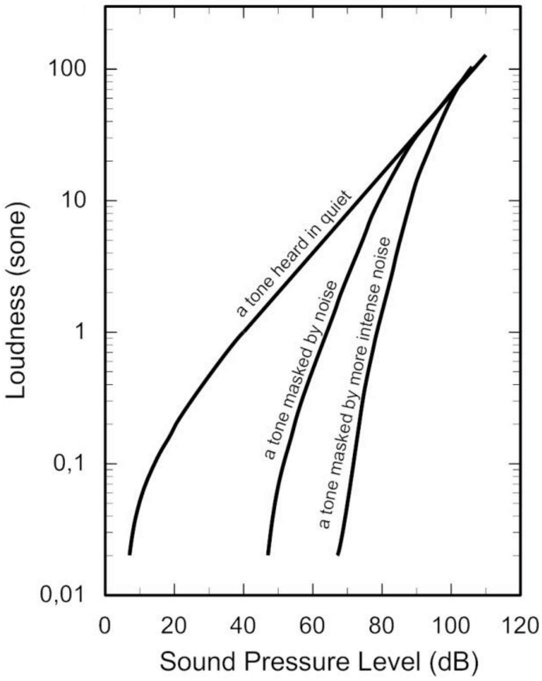

Fig. 3.5. Loudness of a 1-kHz pure tone in quiet and in the presence of a noise masker sufficient to elevate the detection threshold for the tone by 40 and 60 dB. From Gulick et al. (1989).

Binaural loudness summation

A sound presented to both ears is louder than the same sound presented to one ear only. This effect is called *binaural loudness summation*. A typical way of investigating binaural loudness summation is by having a subject listen to the same sounds played back through one and through two earphones. Such experiment have demonstrated that the level of a sound presented to only one ear must be increased by 6-8 dB to equalize this sound in loudness with the same sound heard with two ears. Such a magnitude of binaural summation was observed in experiments conducted with the use of test tones presented in the laboratory, with no picture of the sound source. More recent studies have demonstrated that binaural loudness summation is much smaller in natural listening conditions when the sound source is visible to the listener.

3.4. Pitch

Definition of pitch

Pitch is that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high. Pitch is a particularly important attribute of sound in music as it is an essential element for melody and harmony.

Pitch depends primary on the tone frequency: an increase in frequency of a pure tone or a harmonic complex tone usually results in an increase in pitch. If a sound contains several component tones with various frequencies we may hear a cluster of pitches or a single pitch produced by the whole of the sound's components.

Dependence of pitch upon the tone frequency: the mel scale

Several researchers have attempted to determine a psychophysical scale for pitch that would represent the relationship between the frequency and the pitch of a pure tone. The classical result is the mel scale shown in Fig. 3.6. The reference point is defined on this scale by assigning a pitch value of 1000 mels to a 1-kHz tone.

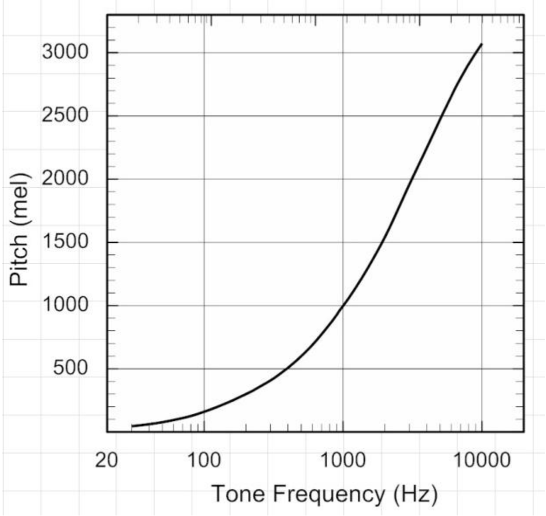

Fig. 3.6. Pitch in mels as a function of the frequency of a pure tone. Adapted from Stevens (1975).

Although the mel scale has been widely cited in the literature its validity is questionable. The scale has been criticized because it contradicts the main principle of the musical scale which is based on an assumption that equal pitch distances (musical intervals) are represented by the same frequency ratio, regardless of the location of the interval on the frequency scale. The interval sizes calculated from the mel scale are unequal when an interval is transposed to different frequency regions. Another objection as to the validity of the mel scale is that different methods used for pitch scaling yield highly differing results so it is difficult to verify the scale's validity.

The influence of sound pressure level on pitch

The pitch of pure tones changes slightly with sound pressure level and the influence of level depends on the tone frequency. With increasing level high-frequency tones, (at frequencies above 2 kHz) increase whereas tones at frequencies below 1 kHz decrease in pitch. In the middle frequency range (1-2 kHz) a change in level does not change the pitch of a tone. The effect of sound pressure level on pitch is small and has no practical importance in music.

Pitch and tone duration

Very short tones are perceived as clicks and have no identifiable pitch. As the tones lengthen the clicks take on a more tonal character and produce a sense of pitch which becomes more salient when the tone is lengthened, up to a duration of several hundreds of milliseconds.

Pitch of complex tones

When a complex tone is composed of harmonically related partials its pitch corresponds to the frequency of the fundamental. This pitch is also heard when the fundamental is not present in the sound spectrum. An example of such an effect in pitch perception are low-pitched sounds heard from a tiny loudspeaker of a notebook computer or from a small portable radio. The pitch of such sounds corresponds to a frequency being below the loudspeaker's frequency response. This kind of pitch is referred to as *virtual pitch*, *residue pitch*, or *periodicity pitch*.

When the component tones of a sound are not harmonic the sensation of virtual pitch may also be heard but in such a case it is determined in the auditory system by some more complex strategy. Usually the perceived virtual pitch is derived from a series of nearly harmonic partials and corresponds to the largest near-common factor in the series of frequencies.

Pitch strength

Pitch may be heard with different precision. A sound may have a distinct, salient pitch or a weak indistinct pitch which may be difficult to define. *Pitch strength*, also called *pitch saliency*, represents the precision with which the pitch of a tone is heard. Pure tones and harmonic complex tones generally produce relatively strong pitch sensation whereas the pitches of sounds with continuous spectra are much weaker. An exception to this general rule are narrow-band noises which produce a clear, salient pitch.

The difference in pitch strength between various sounds becomes clearly apparent when we try to adjust the pitch of a pure tone to the pitches of sounds with various types of spectra. When a sound produces a strong pitch sensation such a pitch matching is easy and the dispersion of several adjustments made by different persons is small. When the pitch sensation is weak the dispersion of pitch matchings considerably increases.

3.5. Perceived duration

Perceived duration not always corresponds exactly to the physical duration of an acoustic event. Listeners have a tendency to overestimate the duration of very brief sounds and the duration of intervals of silence, when such acoustic events are shorter than about 200 ms. Another discrepancy between the physical and perceived duration is that a short interval of silence usually appears perceptually longer than a tone of the same duration.

Some researchers have also reported an influence of pitch on perceived duration. They observed that listeners tend to judge a sound with a higher frequency as slightly longer than a sound of the same physical duration, with a lower frequency.

3.6. Timbre

Timbre is usually defined as that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar. Another definition states that timbre is the perceptual property of sound that reflects the unique properties of the sound and its source. Differences is timbre enable the listener to distinguish between the same notes played on different instruments. Timbre is also called *tone color* or *tone quality*.

Timbre is a multidimensional attribute so there is no single scale along which we could compare the timbres of different sounds. Numerous attempts were made to uncover the dominant auditory perceptual dimensions that constitute timbre. The goal of such kind of studies was to determine a set of semantic differential scales that enable to describe timbre in a meaningful way. Differential scales represent the verbal terms, usually adjectives, used for timbre description.

The findings published in the literature indicate that the vocabulary used for timbre description is very rich and consists of an indefinite number of terms. It is practically impossible to determine an exact number of timbre dimensions as many of the verbal timbre descriptors have overlapping meaning are not always referred to the same properties the auditory sensation by different people. Table 3.1 lists some examples of semantic differential scales used for timbre description. The two edge table columns show the anchors of each scale defined by a pair of bipolar adjectives.

Table 3.1. Example timbre dimensions compiled from the literature.

lower scale end	dimension	upper scale end
dull	**sharpness**	sharp
smooth	**roughness**	rough
dark	**brightness**	bright
empty	**fullness**	full
colorless	**coloration**	colorful
cold	**warmth**	warm
pure	**richness**	rich
dirty	**cleanness**	clean

Timbre depends upon several physical variables which may be broadly divided into two classes: the spectral envelope and the temporal envelope of sound. The variables related to the spectral envelope include the type of spectrum (line, continuous), frequency content, harmonicity and inharmonicity of partials, formant frequencies, formant bandwidth and amplitudes and other factors constituting the sound's spectral profile. Variables related to the temporal envelope of sound include the time course of attack and decay, and the presence of modulations and other kind of temporal variations of the sound.

Review questions

1. What are the basic perceived attributes of sound?
2. What scales are used for specifying the magnitude of loudness?
3. On what physical characteristics of sound does loudness depend?
4 What effect is called spectral loudness summation?
5. What phenomenon is referred to as virtual pitch?
6. What characteristic of the pitch sensation is called the pitch strength?
7. Give examples of timbre dimensions.
8. What physical characteristics of sound have the most pronounced influence on timbre?

Recommended further reading

Howard, D. M., Angus, J. (2006). *Acoustics and Psychoacoustics.* Boston: Focal Press.

Moore, B. C. J. (2013). *An Introduction to the Psychology of Hearing,* sixth edition. Leiden-Boston: Brill.

Yost, W. A. (2007). *Fundamentals of Hearing.* Burlington, MA: Academic Press.

4. Perception of auditory space

4.1. Auditory space

Auditory space is the perceived image of the distribution of sound sources in the surrounding space and the image of interactions of the sound sources with the environment. The auditory space image results from the listener's abilities to identify the direction from which a sound is coming from, the distance of the sound source, the volume and size of the sound source, and from the capability of identifying the spatial characteristics of the environment in which the sounds are heard. Auditory space may be natural, perceived in real acoustic environment, or virtual, created by means of electroacoustic reproduction of sound.

Human sense of sound localization is based primarily on *interaural cues* – acoustic differences between the sounds reaching the two ears. Interaural cues are complemented by monaural cues produced by the head and the pinna and by cues obtained from the movements of the head.

The directions of sound sources in space are usually defined relative to the position of the listener's head. The perceived spatial position of a sound is referred to as *sound localization*. However, if the listener uses headphones, the auditory image is located inside the head rather than in the surrounding space. The apparent location of the sound source in the listener's head, in headphone listening, is called *lateralization*.

Figure 4.1. shows a coordinate system used for defining the positions of sounds relative to the head. The position of each sound is defined by three coordinates:

- *azimuth* – the angle at which the apparent sound source is situated in the horizontal plane,
- *elevation* – the angle at which the apparent sound source is situated in the vertical plane,
- *distance* – the distance of the sound source from the listener's head.

Another important attribute of the auditory space is *volume*. Volume is the perceived size and shape of the acoustic natural or virtual environment in which the sounds are heard.

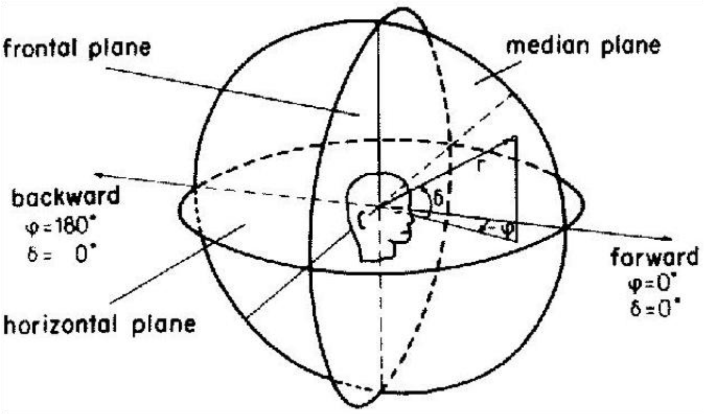

Fig. 4.1. A coordinate system used to define the apparent position of sounds relative to the listener's head. From Blauert (1983),

4.2. Localization in the horizontal plane

The azimuth of the sound source

Figure 4.2. shows schematically an acoustic wave reaching the listener's ears from a source located to one side of the head. The sound arriving at the farther ear is delayed and less intense than the sound received

by the nearer ear. This difference results in two acoustic cues as to the azimuth of the sound source: an interaural intensity difference (IID) and an interaural time difference (ITD). Interaural intensity difference is called interaural level difference (ILD), when specified in decibels. Owing to the nature of acoustic wave propagation the two cues are not equally effective at all frequencies.

The ILD is caused by the baffling effect of the head. The effectiveness of the baffle and the amount of ILD depend on the relative size of the head and the sound wavelength.

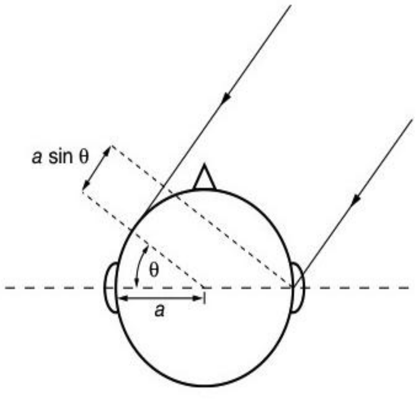

Fig. 4.2. A schematic showing an acoustic wave reaching the listener's ears from a source located to the right side of the head. From Moore (2013).

Figure 4.3 shows the IID measured for a pure tone as a function of sound source azimuth at nine frequencies. The data plotted on the graph indicate that ILD increases with the tone frequency. The wavelength of low-frequency sounds is long compared with the size of the head so the acoustic waves bend around the head, due to diffraction, which results in very small ILD. At high frequencies, where the wavelength is much shorter, very little diffraction occurs and the head produces an effective acoustic shadow resulting in relatively large ILD.

The ITD can be easily calculated from the path difference between the two ears (Fig. 4.2) and the speed of sound propagation in air. The distance between two ears is on average about 23 cm. ITD thus ranges from 0 for a sound source located straight ahead to about 690 μs for a 90° azimuth.

For a periodic sound two cues based on ITD are possible. One cue is related to time of sound arrival (interaural delay of the temporal envelope) and the other one to the sound's phase difference between two ears. Experiments have demonstrated that phase cues are effective up to a frequency of about 1.6 kHz.

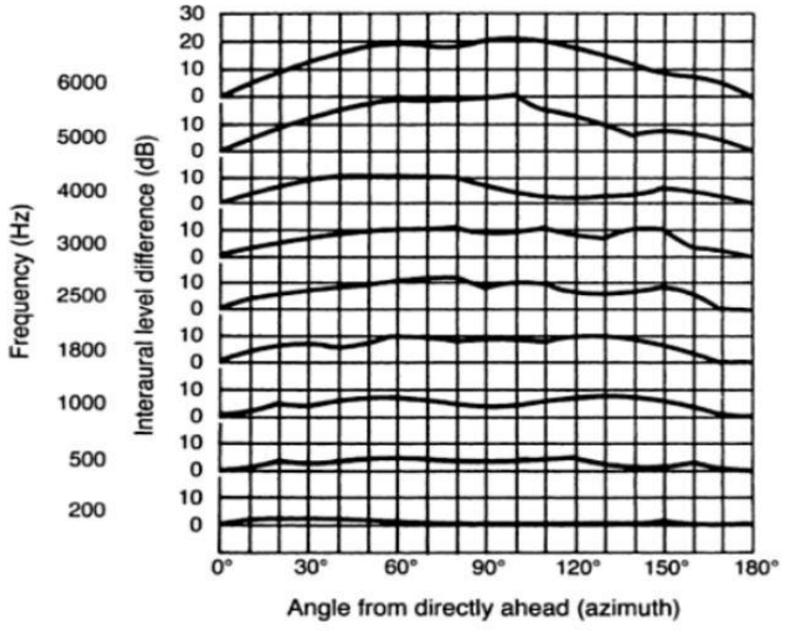

Fig. 4.3. Intensity level differences (ILD) as a function of azimuth, for pure tones at different frequencies. From Feddersen et al. (1957).

The mechanism of detecting the direction from which the sound arrives is based on two kinds of cues: ILD and ITD. ILD plays a dominant role at frequencies above about 1.5 kHz where the amount of ILD is large. ITD is the dominant mechanisms in the frequency region below 1.5 kHz. The theory postulating that sound localization is based on ILD at high frequencies and ITD at low frequencies is called the *duplex theory* of sound localization.

Distance and depth

Distance is a term used to specify how far away from the listener a sound source appears to be. *Depth* describes the overall distance from the front to the back of an auditory image or auditory scene.

A number of cues appear to contribute to the perception of auditory distance. The primary cue is loudness. Loudness decreases as the distance of the sound source from the listener is increased and is the most effective cue for the estimation of distance in outdoor environments.

The second cue is timbre. A change in the distance of a sound source from the listener has an influence on the frequency content of the sound spectrum and, in consequence, on timbre. High frequency spectral components are more damped by the air and humidity than low frequency components. The sounds arriving from distant sources have therefore relatively less high-frequency energy than the same sounds produced by more proximal sources so the sound appears more muffled to the listener.

A cue for estimating the distance in closed spaces is the proportion of direct to reverberant sound. When the sound source is close to the listener the direct-to-reverberant sound ratio is high and the sound is heard more clearly when compared to the sounds arriving from more distant sources.

Movements of the listener’s head provide a cue called *motion parallax*. If the sound source is located far away from the listener the movements of the head result only in small changes in the azimuth of the sound source. If the sound source is close to the listener the movements of the head result in much larger changes in the sound’s azimuth.

4.3. Localization in the vertical plane

The main cue upon which we identify the vertical position of sound sources is the pattern of peaks and valleys in the spectral envelope of a broadband sound caused by the characteristics of the pinna and the ear canal. The spectral changes caused by the movements of the sound source are described by the head-related transfer function (HRTF).

4.4. The precedence effect

In environments with sound-reflecting surfaces, the acoustic waves produced by a given sound source reach the listener’s ears via a number of different paths: via the direct path which corresponds to the shortest distance between the sound source and the listener, and after numerous reflections from the surfaces in the environment. The successive reflections reach the listener with different delays. In spite of this, we perceive the sound as a homogenous acoustic event and usually do not hear the reflections.

If the successive sounds are fused the location of the sound is determined by the location of the first sound. This phenomenon is called the *precedence effect*, the *Haas effect*, or the *law of the first wavefront*. The delayed sound is fused if its level does not exceed the level of the primary sound by more than 10 dB and the delay is not more than about 20 ms. For longer delays, an echo is heard.

Figure 4.4 shows a stimulus sequence used in a classical experiment to investigate the precedence effect. The stimuli were brief clicks presented to the left and to the right ear through headphones. The first pair of clicks arrived at the ears with a small interaural delay. In the second pair the delay was increased to simulate a room reflection of the acoustic wave. The sequence of clicks was perceived as a single sound the location of which was determined by the interaural delay of the leading pair.

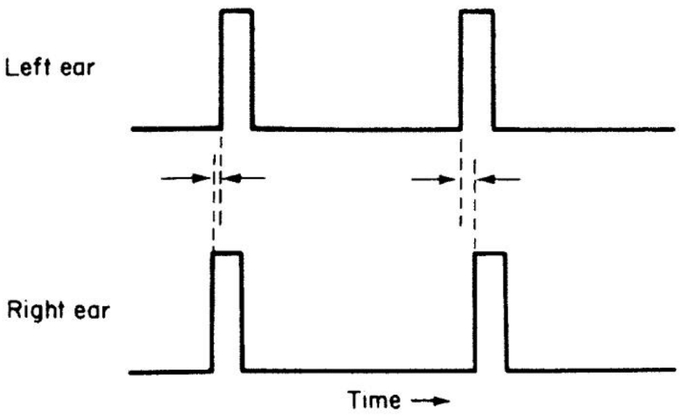

Fig. 4.4. A sequence of clicks used to investigate the precedence effect. From Moore (2013).

4.5. The influence of vision on sound localization

Vision dominates audition with respect to sound source localization. This phenomenon is called the *capture effect*. The capture effect manifests itself in a tendency to perceive sounds as coming from the location of the picture of the sound source despite that the sound actually comes from an unseen source located in a different place. Typical conditions in which the capture effect may be demonstrated are shown in Fig. 4.5.

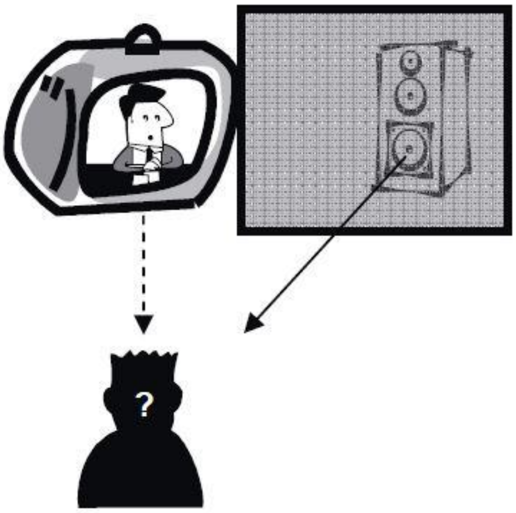

Fig. 4.5. Schematic experiment demonstrating the capture effect. From Ghirardelli and Scharine (2009).

Review questions

1. What cues enable the listener to identify the azimuth of the sound source in the horizontal plane?
2. What does the duplex theory of sound localization state?
3. What cues enable the listener to identify the vertical position of the sound source?
4. Upon what cues is the distance of the sound source from the listener identified?
5 What phenomenon is called the precedence effect?
6. What phenomenon is called the capture effect?

Recommended further reading

Blauert, J. (1996). *Spatial Hearing: The Psychophysics of Human Sound Localization.* Cambridge, MA: MIT Press.

Ghirardelli, T. G., Scharine, A. A. (2009). Auditory-Visual Interactions. In: *Helmet-Mounted Displays: Sensation, Perception, and Cognition Issues,*

C. E. Rasch, M. B. Russo, T. R. Letowski, E. T. Schmeisser (Eds). Aberdeen, MD: U. S. Army Research Laboratory.

Letowski, T. R., Letowski, S. T. (2012). *Auditory Spatial Perception: Auditory Localization.* U.S. Army Research Laboratory, Technical Report 6016.

Sources of figures

Baddeley, A. D. B. (2009). The psychology of memory. In: *The Essential Handbook of Memory Disorders for Clinicians,* A. D. Baddeley., M. D. Kopelman, B. A. Wilson (Eds). Chichester: John Wiley & Sons.

Blauert, J. (1983). *Spatial Hearing: The Psychophysics of Human Sound Localization.* Cambridge, MA: MIT Press.

Emanuel, D. C., Letowski, T. (2009). *Hearing Science.* Philadelphia: Wolters Kluwer.

Fastl, H., Zwicker, E. (2007). *Psychoacoustics: Facts and Models,* third edition. Berlin: Springer.

Feddersen W. E., Sandel T. T., Teas D. C., Jeffress L. A. (1957) Localization of high-frequency tones. *Journal of the Acoustical Society of America* 29, pp. 988–991.

Ghirardelli, T. G., Scharine, A. A. (2009). Auditory-Visual Interactions. In: *Helmet-Mounted Displays: Sensation, Perception, and Cognition Issues,* C. E. Rasch, M. B. Russo, T. R. Letowski, E. T. Schmeisser (Eds). Aberdeen, MD: U. S. Army Research Laboratory.

Gulick, W. L., Gescheider, G. A., Frisina, R. D. (1989*). Hearing: Physiological Acoustics, Neural Coding, and Psychoacoustics.* New York – Oxford: Oxford University Press.

Harrison, R.V., Hunter-Duvar I.M. (1988). An anatomical tour of the cochlea. In: *Physiology of the Ear*, A.F. Jahn, J. Santos-Sacchi (Eds). New York: Raven Press, pp. 159–171.

ISO (2003). *Normal equal-loudness level contours* -ISO 226:2003. Geneva: International Organization for Standardization (ISO)

Letowski. T. (1984). *Auditory Assessment of Auditory Signals and Audio Equipment.* Warsaw: AMFC (in Polish).

Miśkiewicz, A. (2002). *Pitch, Loudness, and Timbre: A Study of Perceived Dimensions of Sound.* Warsaw: AMFC (in Polish).

Moore, B. C. J. (2013). *An Introduction to the Psychology of Hearing,* sixth edition. Leiden-Boston: Brill.

Rossing., T. D., Moore, F. R., Wheeler, P. A. (2002). The Science of Sound. San Francisco: Addison Wesley.

Stevens. S. S. (1975). *Psychophysics: Introduction to its Perceptual. Neural, and Social Aspects.* New York: Wiley&Sons.

Internet web sites

http://bio1152.nicerweb.com/Locked/media

droualb.faculty.mjc.edu

www.ssc.education.ed.ac.uk