11 AUDITORY PERCEPTION AND COGNITIVE PERFORMANCE

Angélique A. Scharine
Kara D. Cave
Tomasz R. Letowski

Audition

Audition is the act of hearing a sound in response to acoustic waves or mechanical vibrations acting on a body. Sound also may result from direct electrical stimulation of the nervous system. The physical stimuli that are, or may become, the sources of sound are called *auditory stimuli*.

The human response to the presence of auditory stimulus and its basic physical characteristics of sound intensity, frequency, and duration is called *auditory sensation*. The three basic auditory sensations are loudness, pitch, and perceived duration, but there are many others. Auditory sensation forms the basis for discrimination between two or more sounds and may lead to some forms of sound classification (e.g., labeling sounds as pleasant or unpleasant). However, auditory sensation does not involve sound recognition, which requires a higher level of cognitive processing of the auditory stimuli. This higher level processing forms a conceptual interpretation of the auditory stimulus and is referred to as *auditory perception*. Auditory perception involves association with previous experience and depends on the adaptation to the environment and expected utility of the observation. Depending on the level of cognitive processing, auditory perception may involve processes of sound classification, e.g., on speech and non-speech sounds, sound recognition, or sound identification. More complex cognitive processing also may include acts of reasoning, selection, mental synthesis, and concept building involving auditory stimuli but extends beyond the realm of audition.

The study of audition is called *psychoacoustics* (psychological acoustics). Psychoacoustics falls within the domain of cognitive psychophysics, which is the study of the relationship between the physical world and its mental interpretation. Cognitive psychophysics is an interdisciplinary field that integrates classical psychophysics (Fechner, 1860), which deals with the relationships between physical stimuli and sensory response (sensation), and with elements of cognitive psychology, which involve interpretation of acting stimuli (perception). In general, cognitive psychophysics is concerned with how living organisms respond to the surrounding environment (Stevens, 1972b). For the above reasons, Neuhoff (2004) refers to modern psychoacoustics as ecological psychoacoustics.

In general, all content of our experience can be ordered by quantity, quality, relation, and modality (Kant, 1781). These experiences are reflected in perceptual thresholds, various forms of comparative judgments, magnitude estimation, emotional judgments, and scaling. These characteristics define the realm of psychoacoustics and, more generally, psychophysics. Various types of cognitive measurements and methodological issues addressing psychophysical relationships are described in Chapter 15, *Cognitive Factors in Helmet-Mounted Displays,* and are not repeated here. The current chapter presents psychoacoustic relationships and builds upon the information on the anatomy and physiology of the auditory system presented in Chapter 8, *Basic Anatomy of the Hearing System*, and Chapter 9, *Auditory Function*. It describes a variety of auditory cues and metrics that are used to derive an understanding of the surrounding acoustic space and the sound sources operating within its limits. Understanding how a particular sound is likely to be perceived in a particular environment is necessary for the design of effective auditory signals and to minimize the effects of environmental noise and distracters on performance of audio helmet-mounted displays (HMDs). Psychoacoustics provides the basic conceptual framework and measurement tools (thresholds and scales) for the discussion and understanding of these effects.

Sound Pressure and Sound Pressure Level

The main physical quantity that elicits auditory response is time-varying sound pressure. The other quantities are time-varying force (bone conduction hearing) and time-varying (alternating current [AC]) voltage (electric hearing). The unit of sound pressure is the Pascal (Pa), which is equal to a Newton/meter^2 (N/m^2), and the range of sound pressures that can be heard by humans extends from about 10^{-5} Pa to 10^2 Pa. The large range of values needed to describe the full range of audible sound pressure makes the use of Pascals, or other similar linear units, very cumbersome. In addition, human auditory perception is far from linear. Human perception is relative by nature and logarithmic in general, i.e., linear changes in the amount of stimulation cause logarithmic changes in human perception (Emanuel, Letowski and Letowski, 2009). Therefore, sound pressure frequently is expressed in psychoacoustics on a logarithmic scale known as the *decibel scale* from the name of its unit, the *decibel* (dB). The decibel scale has a much smaller range than the sound pressure scale and more accurately represents human reaction to sound. Sound pressure expressed in decibels is called sound pressure level. Sound pressure level (SPL) and sound pressure (p) are related by the equation:

$$SPL\,(\text{dB}) = 20\log\left(\frac{p}{p_o}\right) \quad [\text{dB SPL}] \qquad \text{Equation 11-1}$$

where p_o is the reference sound pressure value and equals 20 x10^{-6} Pa. For example, a sound pressure (p) of 1 Pa corresponds to 94 dB SPL, and the whole range of audible sound pressures extends across about 140 dB SPL. An SPL of 1 dB corresponds to a sound pressure increase of about 1.122 times (~12%).

When dealing with complex continuous sounds, it is frequently more convenient to use energy-related magnitudes such as sound intensity (I) or sound intensity level (SIL) rather than sound pressure and sound pressure level. Such an approach allows one to refrain from the concept of phase that complicates physical analysis and has limited usefulness for many aspects of auditory perception. Sound intensity is the density of sound energy over an area, is expressed in units of Watts/meter^2 (W/m^2), and for a plane traveling sound wave, I ~ p^2. Therefore, the relation between sound pressure level (dB SPL) and sound intensity level (dB SIL) can be written as:

$$SPL\,(\text{dB}) = 20\log\left(\frac{p}{p_o}\right) = 10\log\left(\frac{I}{I_o}\right) = SIL\,(dB) \qquad \text{Equation 11-2}$$

where I_o is the reference sound intensity value of 10^{-12} W/m^2. I_o is the sound intensity produced by the sound pressure equal to p_o. This means that both values refer to the same sound, and the scale of sound pressure level in dB SPL is identical to the scale of sound intensity level in dB SIL. For that reason, the names *sound pressure level* and *sound intensity* are used interchangeably in this chapter and all the graphs labeled sound pressure level (dB SPL) would be identical if the label was replaced by sound intensity level (dB SIL).

Threshold of Hearing

Sensitive hearing is an important listening ability needed for human communication, safety, and sound assessment. An auditory stimulus arriving at the auditory system needs to exceed a certain level of stimulation to cause the temporary changes in the state of the system that result in the sensation and perception of sound. The minimum level of stimulation required to evoke a physiologic response from the auditory system is called the *threshold of hearing* or *detection threshold* and depends on the frequency, duration, and spectral complexity of the stimulus. Thus, the threshold value is the lowest intensity level (e.g., sound pressure level, bone conduction force level, electric voltage level) for which a particular auditory stimulus can be detected. When the sound is

heard in quiet, the detection threshold is called the *absolute detection threshold* (absolute threshold of hearing), and when is presented together with other sounds, the detection threshold is referred to as the *masked detection threshold* (masked threshold of hearing).

The term threshold of hearing implies the existence of a discrete point along the intensity continuum of a particular stimulus above which a person is able to detect the presence of the stimulus. However, an organism's sensitivity to sensory stimuli tends to fluctuate, and several measures of the threshold value must be averaged in order to arrive at an accurate estimation of the threshold. Therefore, the threshold of hearing is usually defined as the sound intensity level at which a listener is capable of detecting the presence of the stimulus in a certain percentage, usually 50%, of cases.

Normal daily variability of the threshold of hearing can be assumed to be 6 dB or less. For example, Delany (1970) and Robinson (1986) used a Bekesy tracing procedure (Brunt, 1985) and supra-aural earphones to assess within-subject variability of the hearing threshold in otologically normal listeners and both reported an average standard deviation in the order of 5 dB at 4000 Hertz (Hz) for repeated measures of the threshold of hearing on the same person. Henry et al. (2001) used insert earphones ER-4B (see Chapter 5, *Audio Head Mounted Displays*) and the ascending method of limits with 1 dB steps and reported average within subject standard deviations between 1.9 dB and 5.3 dB depending on the stimulus frequency.

In this context, it is useful to differentiate between the physiological threshold defined by the inherent physiological abilities of the listener and the cognitive threshold limited by the listener's familiarity with the stimuli, interest in and attention to the task, experience with existing set of circumstances, and the numerous procedural effects in eliciting the listener's response (Seashore, 1899; Letowski, 1985). The cognitive thresholds can be made close to the physiological thresholds by an appropriate amount and type of training (Letowski, 1985; Letowski and Amrein, 2005); but the difference between potential physiological threshold and observed cognitive threshold always has to be taken into account when discussing any specific threshold data.

The need for a statistical approach to the threshold of hearing also exists for normative thresholds for specific populations because people differ in both their overall sensitivity to sound and the shape of the threshold of hearing as a function of frequency. For example, inter-subject (between-subject) standard deviations ranging from 3 dB to 6 dB were reported for low and middle frequencies in a number of studies (Moller and Pedersen, 2004) and the inter-subject data variability tends to increase with stimulus frequency. Thus, due to both individual (intra-subject) and group (inter-subject) variability of the threshold of hearing, human sensitivity to auditory stimulation needs to be defined in terms of a statistical distribution with certain measures of central tendency and dispersion (variability).

Air conduction threshold

The range of frequencies heard by humans through air conduction extends from about 20 Hz to 20 kHz and may possibly include even lower frequencies if stimulus intensities are sufficiently high (Møller and Pedersen, 2004). The average human threshold of hearing, standardized by the International Organization for Standardization (ISO, 2005) for people age 18 to 25 years with normal hearing, varies by as much as 90 dB as a function of the stimulus frequency and is shown in Figure 11-1. The specific threshold values are listed in Table 11-1.

The threshold curves in Figure 11-1 and numbers in Table 11-1 represent average binaural thresholds of hearing in a free sound field and a diffuse sound field. A free sound field is an acoustic field, free of reflections and scattering in which sound waves arrive at the listener's ears from only one direction identified by the position of the sound source relative to the main axis of the listener. A diffuse sound field is a sound field in which the same sound wave arrives at the listener more or less simultaneously from all directions with equal probability and level. The free-field thresholds were measured for pure tones with the subject directly facing the source of sound (frontal incidence). The diffuse-field thresholds were measured for one-third-octave bands of noise. In both cases, the threshold sound pressure level was measured at the position corresponding to the center of the listener's head with the listener absent.

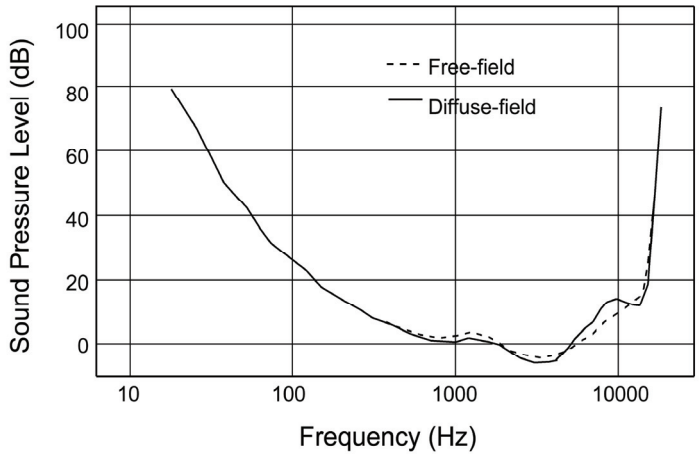

Figure 11-1. Binaural hearing threshold of hearing in a free field (frontal incidence) and in a diffuse field as a function of frequency (adapted from ISO, 2005).

Table 11-1.
Reference thresholds of hearing for free-field listening (frontal incidence) and diffuse-field listening in dB SPL (re: 20 μPa) (ISO, 2005).

Frequency (Hz)	Free-field listening (frontal incidence) (dB SPL)	Diffuse-field listening (dB SPL)
20	78.5	78.5
25	68.7	68.7
31.5	59.5	59.5
40	51.1	51.1
50	44.0	44.0
63	37.5	37.5
80	31.5	31.5
100	26.5	26.5
125	22.1	22.1
160	17.9	17.9
200	14.4	14.4
250	11.4	11.4
315	8.6	8.4
400	6.2	5.8
500	4.4	3.8
630	3.0	2.1
750	2.4	1.2
800	2.2	1.0
1000	2.4	0.8
1250	3.5	1.9
1500	2.4	1.0

Table 11-1. (Cont.)
Reference thresholds of hearing for free-field listening (frontal incidence) and diffuse-field listening in dB SPL (re: 20 μPa) (ISO, 2005).

Frequency (Hz)	Free-field listening (frontal incidence) (dB SPL)	Diffuse-field listening (dB SPL)
1600	1.7	0.5
2000	-1.3	-1.5
2500	-4.2	-3.1
3000	-5.8	-4.0
3150	-6.0	-4.0
4000	-5.4	-3.8
5000	-1.5	-1.8
6000	4.3	1.4
6300	6.0	2.5
8000	12.6	6.8
9000	13.9	8.4
10000	13.9	9.8
11200	13.0	11.5
12500	12.3	14.4
14000	18.4	23.2
16000	40.2	43.7
18000	73.2	--

The binaural thresholds of hearing shown in Figure 11-1 and Table 11-1 are approximately 2 dB lower than the corresponding monaural thresholds if both ears have similar sensitivity (Anderson and Whittle, 1971; Killion, 1978; Moore, 1997). This difference applies to pure tones of various frequencies as well as to speech, music, and other complex stimuli presented under the same monaural and binaural conditions.

Low frequency stimuli are felt as a rumble (Sekuler and Blake, 1994) and require a relatively high sound level to be detected. Their audibility does not vary much among individuals and depends primarily on the mechanical properties of the ear. The audibility of low frequency stimuli improves with increasing frequency at an average rate of about 12 dB/octave and typically decreases with age, from 20 to 70 years of age, by 10 dB or less for frequencies below 500 Hz (ISO, 2000). The audibility of stimuli with frequencies in the upper end of the frequency range varies quite a bit with individuals and decreases substantially with age (Stelmachowicz et al., 1989). The typical changes in the threshold of hearing with age, from 20 to 70 years of age, at frequencies above 8,000 Hz, exceed 60 dB in normally hearing population.

As demonstrated in Figure 11-1, the auditory system is especially sensitive to sound energy in the 1.5 to 6 kHz range with the most sensitive region being in the 3.0 to 4.0 kHz range (Moore, 1997). The high sensitivity of the auditory system in this frequency range results from acoustic resonances of the ear canal and concha described in Chapter 9, *Auditory Function*. The normal hearing threshold level in its most sensitive range is about -10 dB re 2 x 10^{-5} Pa (2 x 10^{-4} μbar)[1] and the amplitude of the tympanic membrane displacement is about 10^{-9} centimeter (cm), i.e., not much larger than the amplitude of the random motion of molecules in solution (Licklider, 1951).

[1] When reporting the relative intensity of a sound, it is important to not only say "dB" but to also add the reference level. This is often written as "dB re 20 μPa" for sounds in air that are measured relative (re) to 20 μPa.

Low hearing sensitivity at low frequencies is primarily due to poor transmission of low frequency energy by the mechanical systems of the outer and middle ears and limited mobility of the outer hair cells in the low frequency range (Moore, Glasberg and Bear, 1997). The presence of the second mechanism is probably due to the high level of low frequency internal body noise, such as that caused by the blood flow, which normally should not be heard.

When describing the threshold of hearing, it is important to consider not only whether the threshold is monaural or binaural but also how the sound is presented and where the level of arriving stimulus is measured. Two specific types of hearing threshold, the *minimum audible field* threshold and *minimum audible pressure* threshold are of special importance. The minimum audible field (MAF) threshold refers to the threshold of hearing for acoustic waves arriving at the ear in a free-field environment from a distal sound source, e.g., a loudspeaker. The minimum audible pressure (MAP) threshold refers to the threshold of hearing from a stimulus arriving from an earphone occluding the ear canal.

The difference between the MAF and MAP thresholds of hearing is illustrated in Figure 11-2. The average difference between both thresholds is in the order of 6 dB to 10 dB and has been sometimes referred to in the literature as the "missing 6 dB." It should be noted that especially large differences between the MAF and MAP thresholds in the 1.5 to 4 kHz frequency region. This difference is the effect of resonance properties of the ear canal and concha.

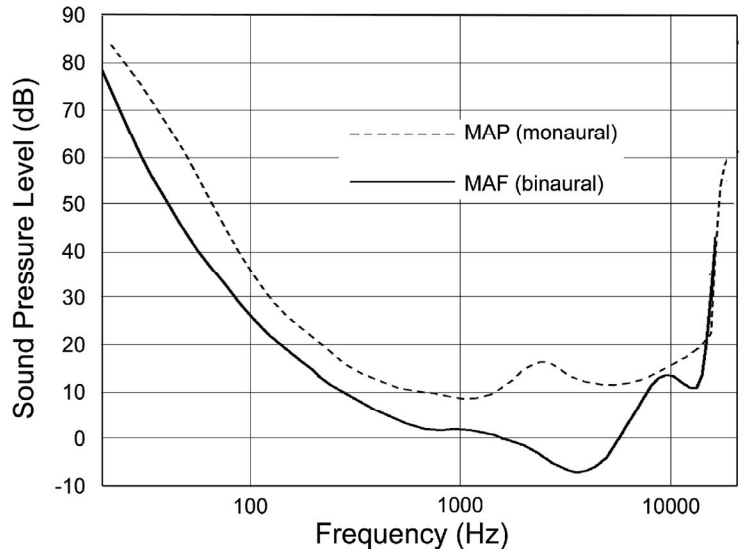

Figure 11-2. Comparison of the MAF (ISO, 2003) and MAP (Killion, 1978) threshold of hearings. The reference points for both measurements are the center of the listener's head with the listener absent (MAF) and a point close to the listener's tympanic membrane (MAP).

The main reason that the MAF and MAP thresholds of hearing, such as shown in Figure 11-2, differ in their values is the differences in the actual point of reference where the sound pressure is measured. In practice, the reference points for the MAF and MAP thresholds do not only differ from each other but they also differ within each threshold category. The typical reference points for MAF threshold are the point at the entrance to the ear canal at the tragus and the point representing the position of the center of the person's head when the person is absent. The MAP measurements are frequently made with a small probe tube microphone where the tip of the probe is located either close to the tympanic membrane, at various points along the ear canal, or in front of the earphone grill. The differences between thresholds obtained for sounds presented in the open field versus those presented through earphones are large due to the reflective properties of the human head and torso and different sound amplification by resonance properties of the ear canal and concha (Chapter 9, *Auditory Function*). In addition, the MAP values for the threshold of hearing are frequently estimated in various acoustic couplers and

ear simulators in which the pressure often bears little resemblance to the actual pressure at the listener's tympanic membrane (Killion, 1978). In such cases, specific "reference equivalent threshold sound pressure levels" (RETSPLs) are established for various earphone-coupler combinations to reference threshold of hearing pressure to voltage value applied to the earphone. The RETSPL values are internationally standardized for several reference earphones including supra-aural earphones (e.g., Telephonics TDH-39 and Beyer DT-48), circumaural earphones (e.g., Sennheiser HD-200), and insert earphones (e.g., Etymotic Research ER-3). Each of the set of RETSPLs is referenced to a specific standardized acoustic coupler and is only valid when the appropriate coupler and appropriate calibration method are used.

Another factor contributing to the difference between MAF and MAP thresholds of hearing is that MAF thresholds are usually determined for binaural listening, while MAP thresholds are usually reported for the monaural condition. In addition, occluding the ear canal by an earphone was reported to cause an amplification of low frequency physiologic noise (e.g., blood flow noise) by the closed cavity of the ear canal and elevation of MAP threshold at low frequencies (Anderson and Whittle, 1971; Block, Killion and Tillman, 2004; Brodgen and Miller, 1947; Killion, 1978). The occluded ear canal also has different resonance modes than the open canal. Similarly, measurements of the MAF threshold in less than ideal (anechoic) conditions may cause room reflections to affect the real threshold values.

The threshold values discussed above have been determined for pure tones and are primarily used for clinical applications. However, for many practical field applications, room or transducer (e.g., audio HMD) frequency response considerations, and special population testing it is important to determine the threshold of hearing for speech signals and other complex acoustic stimuli such as narrow bands of noise and filtered sound effects. One class of such signals is 2% to 5% frequency modulated (FM) tones, called warble tones, which are used commonly in sound field audiometry. They result in the thresholds of hearing similar to the pure-tone thresholds but are less dependent on the acoustic conditions of a room. They are also the signal of choice for high frequency audiometry (Tang and Ltowski, 2007).

In the case of speech signals, there are two thresholds of hearing for speech that are of interest for practical applications: the threshold of speech intelligibility (speech recognition threshold, speech reception threshold) and the threshold of speech audibility (speech awareness threshold, speech detection threshold). The normative speech recognition threshold for English spondee (two-syllable) words is 14.5 dB SPL for binaural listening in a free sound field with the sound presented at 0° incidence (American National Standards Institute [ANSI], 2004). The speech awareness threshold (SAT) is approximately 8 to 9 dB lower (Dobie and van Hemel, 2004; Sammeth et al., 1989).

Auditory thresholds for narrow-band noises have been reported by Garstecki and Bode (1976), Mitrinowicz and Letowski (1966), Sanders and Joey (1970), and others. In all of these studies, the reported thresholds correlated very well with pure tone thresholds and were usually within ±3 dB of each other. Mitrinowicz and Letowski (1966) observed that the relation between the narrow-band noise thresholds and the pure tone thresholds was mitigated by the relation between the width of the noise band and the width of the corresponding critical band (to be discussed further in this chapter). Zarcoff (1958) also reported a good correlation between narrow-band noise thresholds at mid-frequencies and the speech recognition thresholds.

Environmental sound research is still in its beginning stages (Gygi and Shafiro, 2007), and there are few studies reporting thresholds of hearing for various environmental and man-made sounds. Many early reports dealing with environmental sounds are qualitative rather than quantitative in nature and the listening conditions used in these studies are not well described. Yet, they provide much information on human ability to differentiate various sounds, informational and emotional meaning of the sounds, and provide the case for further, more detailed studies. The few quantitative studies report threshold values that vary across more than 30 dB depending on the sound, listener, listening environment, and listening condition. Some of the more important early and quantitative reports related to detection and recognition of environmental sounds include: Abouchacra, Letowski and Gothie (2006); Ballas (1993); Ballas, Dick and Groshek (1987); Ballas and Howard (1987); Ballas and Barnes (1988); Fidell and Bishop (1974); Gygi (2001); Gygi, Kidd and Watson, (2004); Gygi and Shafiro (2007); Price and

Hodge (1976); Price, Kalb and Garinther (1989); and Shafiro and Gygi (2004; 2007). There are also reports on detection and recognition of octave-band filtered environmental sounds for warning signal and auditory icon applications (Myers et al., 1996; Abouchacra and Letowski, 1999).

Bone conduction threshold

Figures 11-1 and 11-2 refer to the air-conducted threshold of hearing. However, sound can also be heard through bone conduction transmission either directly, through contact with a vibrating object or indirectly, by picking up vibrations in the environment (Henry and Letowski, 2007). In real operational environments, bone conducted sound transmission is likely to occur through the use of bone conduction communication devices (direct stimulation) or from very loud sounds in the environment (indirect stimulation). In the former case, the effectiveness of bone conduction depends on the location of the vibrator on the head and the quality of the contact between the vibrator and the skin of the head (McBride, Letowski and Tran, 2005; 2008). In the latter case, bone conducted sounds may be masked by stronger air conducted sounds except for the cases when high-attenuation hearing protection is used (see Chapter 9, *Auditory Function*).

The threshold of hearing for bone conduction is defined as the smallest value of mechanical force (force threshold) or acceleration (acceleration threshold) applied to the skull resulting in an auditory sensation. Table 11-2 lists the force threshold values for bone conduction threshold as given in ANSI) standard S3.6 (ANSI, 2004) for a vibrator placed on the mastoid and on the forehead. Threshold values vary with frequency and are lowest in the 1 to 4 kHz region, similarly as for the air-conduction threshold.

Table 11-2.
Normal monaural force hearing thresholds for bone-conducted sounds at different frequencies for a B-71 vibrator placed on the mastoid and at the forehead (ANSI, 2004).

Frequency (Hz)	Mastoid Location (dB re 1μN)	Forehead Location (dB re 1μN)
250	67.0	79.0
315	64.0	76.5
400	61.0	74.5
500	58.0	72.0
630	52.5	66.0
750	48.5	61.5
800	47.0	59.0
1000	42.5	51.0
1250	39.0	49.0
1500	36.5	47.5
1600	35.5	46.5
2000	31.0	42.5
2500	29.5	41.5
3000	30.0	42.0
3150	31.0	42.5
4000	35.5	43.5
5000	40.0	51.0
6000	40.0	51.0
6300	40.0	50.0
8000	40.0	50.0

Similarly to air conduction thresholds, bone conduction thresholds may be measured using an artificial load, in this case a mechanical load such as an artificial mastoid, in lieu of the human head. In such cases, the bone conduction threshold is referenced by "reference equivalent threshold force levels" (RETFLs), which is the acoustic force needed for threshold sensation when applied to the artificial load.

At present the only well established direct-stimulation bone conduction thresholds are for pure tones. The only other bone conduction thresholds that were published are the thresholds for octave-band filtered sound effects that were published together with the corresponding air conduction thresholds by Abouchacra and Letowski (1999).

In the case of bone conduction stimulation by impinging sound waves, such sound waves need to be 40 to 50 dB more intense than those causing the same sensation though the air conduction pathways. More information on bone conduction mechanisms and the use of bone conduction hearing in audio HMDs is included in Chapter 9, *Auditory Function*, and Chapter 5, *Audio Helmet Mounted Displays,* respectively.

Threshold of Pain

The threshold of hearing is an example of the basic class of perceptual thresholds called the detection thresholds or absolute thresholds, which separate effective from ineffective stimuli. The other type of perceptual threshold is the terminal threshold. The terminal threshold defines the greatest amount of stimulation that can be experienced in a specific manner before it causes another form of reaction. Examples of auditory terminal thresholds are the threshold of discomfort, also referred to as the loudness discomfort level (LDL), and the threshold of pain.

The threshold of discomfort represents the highest sound intensity that is not uncomfortable or annoying to the listener during prolonged listening. According to Gardner (1964), the threshold of discomfort is almost independent of background noise level (in 30 to 70 dB SPL range) and exceeds 80 dB SPL. It depends on the listener, type of sound (tone, speech, noise band), and frequency content of the sound and varies in 80 to 100 dB SPL range for speech signals (Denenberg and Altshuler, 1976; Dirks and Kamm, 1976; Keith, 1977; Morgan et al., 1979). The typical difference between the most comfortable listening level and threshold of discomfort is about 15 dB for pure tones and speech (Dirks and Kamm, 1976) and about 25 dB for noises (Sammeth, Birman and Hecox, 1989).

The threshold of pain represents the highest level of sound that can be heard without producing a pain. The threshold of pain is practically independent of frequency and equals about 130 to 140 dB SPL. The nature of pain, however, varies with frequency. At low frequencies, people experience dull pain and some amount of dizziness, which suggests the excitation of the semicircular canals. At high frequencies, the sensation resembles the stinging of a needle.

There are reports indicating that tones of very low frequencies below 20 Hz, called *infrasound*, can be heard by some people at very high intensity levels exceeding 115 dB (Møller and Pedersen, 2004). In general, however, such tones having sufficiently high sound intensity levels are not heard but immediately felt causing disorientation, pain, and feeling of pressure on the chest (Gavreau, 1966, 1968). A similar situation may exist for high frequency tones exceeding 20 kHz, called *ultrasound*, although there are numerous reports indicating that people can hear ultrasound stimuli if they are applied through bone conduction (Lenhardt et al., 1991).

Area of Hearing

If the energy of an auditory stimulus falls within the sensory limits of the auditory system we can hear a sound, i.e., receive an auditory impression caused by the energy of the signal. The range of audible sound between the threshold of hearing and the threshold of pain displayed in frequency (abscissa) and sound pressure level (ordinate) coordinates is called the area of hearing. The area of hearing, together with smaller areas of music and speech sounds, is shown graphically in Figure 11-3.

The data presented in Table 11-1 and Figures 11-1, 11-2, and 11-3, show that the threshold of hearing changes considerably across the whole range of audible frequencies. Therefore, in some cases it is convenient to describe

actual stimuli in terms of the level above the threshold of hearing rather than sound pressure level. This level is called the hearing level (HL), when referred to the average threshold of hearing for a population, or the sensation level (SL), when referred to the hearing threshold of a specific person. The average hearing threshold level (0 dB HL) for a given population is called in the literature the *reference hearing threshold level* or the *audiometric zero level*. For example, the level of 60 dB SPL in Figure 11-3 would correspond to 25 dB HL and 60 dB HL for a 100 Hz and a 1000 Hz tone, respectively. Keep in mind that these are the approximate values due to the conceptual form of Figure 11-3. The more exact numbers can be found from Figure 11-6 and the relation between dB SPL values and 0 dB HL values may be found is Table 11-1 (air conduction threshold) and Table 11-2 (bone conduction threshold).

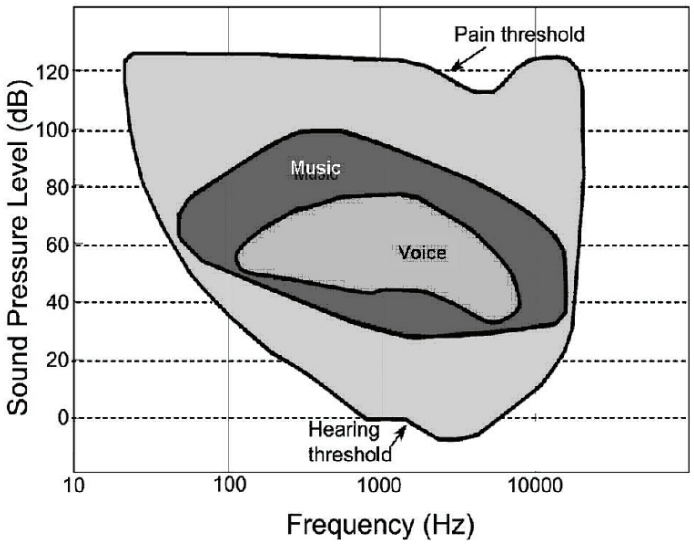

Figure 11-3. Area of hearing (light gray) together with areas of music (black), and speech (dark gray).

As shown in Figure 11-3, the dynamic range of human hearing extends from approximately -10 dB SPL to 130 dB SPL. To make the above numbers more practical, the range of sound intensity levels of the natural and man-made sounds that are generated in the environment is shown in Table 11-3.

The frequency range of hearing shown in Figure 11-3 extends from about 20 Hz (or even less) to 20 kHz. This range is slightly larger than the range of sounds of music but much larger than the range of speech sounds that extends from about 200 Hz to approximately 8 kHz. However, it is important to stress that hearing sensitivity, especially in the high frequency region, declines with age (Galton, 1883; Robinson and Dadson, 1957), exposure to noise, and use of ototoxic drugs, and the standardized threshold levels published in literature refer typically to the average threshold of hearing in young people, i.e., age 18 to 25 years.

The threshold of hearing is not the same for all populations and depends on both gender and ethnicity. Murphy, Themann and Stephenson (2006) evaluated hearing threshold in more than 5000 U.S. adults, age 20 to 69 years, and found: women have on average better hearing than men; non-Hispanic blacks have the best hearing threshold; and non-Hispanic whites have the worst among all ethnic groups evaluated in the study. In addition, Cassidy and Ditty (2001), Corso (1957; 1963) and Murphy and Gates (1999) reported that women of all ages have better hearing than men at frequencies above 2000 Hz, but with aging women have poorer capacity to hear lower frequencies than do men. This means that the pattern of hearing loss with aging for women and men is not the same. The data regarding changes in the threshold of hearing with age can be found in ISO standard 7029 (ISO, 2000).

From an operational point of view it is important to compare the frequency range of human hearing with the frequency ranges of other species. At its high frequency end human hearing extends above 10 kHz, as does the hearing of all other mammals with a few exceptions (e.g., subterranean mammals such as blind mole rat) (Heffner and Heffner, 1993). Birds do not hear sounds higher than 10 kHz and amphibians, fish, and reptiles do not generally hear sounds higher than 5 kHz (Heffner and Heffner, 1998). Dogs hear frequencies up to about 45 kHz and some bats and porpoises can hear sounds beyond 100 kHz. In mammals, smaller head size generally is correlated with better high frequency hearing of the mammal (Masterton, Heffner and Ravizza, 1967). This relationship is important for species survival since small head size produces a smaller acoustic shadow and good high frequency hearing is needed for effective sound localization and effective hunting (Heffner and Heffner, 2003). The importance of high frequency hearing for sound localization has been addressed previously in Chapter 9, *Auditory Function*, and will be further discussed in following sections in this chapter.

Table 11-3.
Sound intensity levels of some environmental and man-made sounds.
Adapted from Emanuel and Letowski (2009).

Sound Level (dB SPL)	Examples of Sounds
0	Quietest 1kHz tone heard by young humans with good hearing. Mosquito at 3 meters (9.8 feet).
10	Human breathing at 3 meters (9.8 feet). Wristwatch ticking at 1 meter (3.28 feet).
20	Rustling of leaves at 3 meters (9.8 feet). Whisper at 2 meters (6.6 feet). Recording studio noise level.
30	Nighttime in a desert. Quiet public library. Grand Canyon at night. Whisper at the ear.
40	Quiet office. Loud whisper. Suburban street (no traffic). Wind in trees.
50	Average office. Classroom. External air conditioning unit at 30 meters (98 feet).
60	Loud office. Conversational speech at 1 meter (3.28 feet). Bird call at 3 meters (9.8 feet).
70	Inside passenger car (65 mph). Garbage disposal at 1 meter (3.28 feet). [1 μbar = 74 dB SPL]
80	Vacuum cleaner at 1 meter (3.28 feet). Noisy urban street. Power lawn mower at 3 meters (9.8 feet).
90	Heavy track (55 mph) at 1 meter. Inside HMMWV (50 mph). [1 pascal = 94 dB SPL]
100	Pneumatic drill at 2 meters (6.6 feet). Chain saw at 1 meter (3.28 feet). Disco music (dancing floor).
110	Symphonic orchestra (tutti; forte fortissimo) at 5 meters (16.4 feet). Inside M1 tank (20 mph).
120	Jet airplane taking off at 100 meters (328 feet). Threshold of pain. [1W/m^2 = 120 dB SPL]
130	Rock band at 1 meter. Civil defense siren at 30 meters (98 feet). [1 mbar = 134 dB SPL]
140	Aircraft carrier flight deck Human eyes begin to vibrate making vision blurry.
150	Jet engine at 30 meters. Formula I race car at 10 meters (32.8 feet).
160	M-16 gunshot at shooter's ear (157 dB SPL). Windows break at about 160 dB SPL
170	Explosion (1 tone TNT) at 100 meters (328 feet). Direct thunder. [1 psi = 170.75 dB SPL]
180	Explosion (0.5 kg TNT) at 5 meters (16.4 feet). Hiroshima atomic bomb explosion at 1.5 kilometers (0.93 miles).
190	Ear drums rupture at about 190 dB SPL. [1 bar = 194 dB SPL]
200	Bomb explosion (25 kg TNT) at 3 meters (9.8 feet). Humans die from sound at 200 dB SPL
210	Space shuttle taking off at 20 meters (65.6 feet). Sonic boom at 100 meters (328 feet).
220	Saturn 5 rocket take off at 10 meters (32.8 feet).

At the low frequency end, human hearing is relatively extensive and very few species (e.g., elephants and cattle) hear lower frequencies than humans. It is noteworthy that the low frequency limit of the mammal hearing

is either lower than 125 Hz or higher than 500 Hz. Only very few species that have been reported to have a low-frequency limit of hearing in 125 to 500 Hz range do not fit this dichotomy (Heffner and Heffner, 2003). This is an important finding because its supports the existence of dual mechanisms of pitch (frequency) perception in mammals (including humans), i.e., place coding and temporal coding (see Chapter 9, *Auditory Function*). It has been argued that temporal coding operates up to less than 300 Hz (Flanagan and Guttman, 1960; Shannon, 1983) while place coding fails to account for good frequency resolution at low frequencies. Thus it seems possible that the mammals that do not hear below 500 Hz may use only place mechanism for pitch coding while the mammals that hear below 125 Hz may be using both place and temporal coding for pitch perception. The list of frequency ranges of selected mammals is given in Table 11-4.

Table 11-4.
Approximate hearing ranges of various species (Fay, 1988; Warfield, 1973).

Species	Low Frequency Limit (Hz)	High Frequency Limit (Hz)
Beluga whale	1,000	120,000
Bat	2,000	110,000
Bullfrog	100	3,000
Cat	45	64,000
Catfish	50	4,000
Chicken	125	2000
Cow	25	35,000
Dog	65	45,000
Elephant	16	12,000
Horse	55	33,000
Owl	125	12,000

Auditory Discrimination

The third, and last, class of perceptual thresholds is the differential thresholds. The differential threshold is the smallest change in the physical stimulus that can be detected by a sensory organ. Such threshold is frequently referred to as a *just noticeable difference* (jnd) or *difference limen* (DL).[2] The size of the differential threshold increases with the size of the stimulus and this relationship is known as Weber's Law, named after Erich Maria Weber who formulated it in 1834 (Weber, 1834). Weber's Law states that the smallest noticeable change in stimulus magnitude (ΔI) is always a constant fraction of the stimulus magnitude (I):

$$\frac{\Delta I}{I} = c = const. \qquad \text{Equation 11-3}$$

where c is a constant called the Weber fraction (Weber, 1834). This expression leads to a logarithmic function describing the dependence of noticeable change in stimulus magnitude on stimulus magnitude.

When Weber's Law is applied to the stimulus intensity, it holds across a large range of intensities except for those intensities close to the threshold of detection. At the low levels of stimulation, the actual differential thresholds are larger than predicted by Weber's Law due to the presence of internal noise. This effect has been

[2] *Difference limen* (as the *jnd*) is the smallest change in stimulation that an observer can detect.

termed the "near miss to Weber's law" (McGill and Goldberg, 1968). Differential thresholds for stimulus characteristics other than stimulus intensity do not demonstrate any notable departure from Weber's Law.

Differential thresholds can be further classified as single-event (step) thresholds and modulation thresholds depending on the nature of the change. Single-event thresholds are the thresholds of detection of a single change in signal property. Modulation thresholds are the thresholds of detection of signal modulation, that is, periodic changes in signal property. If the single-event stimulus change has the form of a step without any interstimulus pause, the differential threshold is generally smaller than the modulation detection thresholds (Letowski, 1982). The tendency of the auditory system to smooth out (ignore) small frequently repeated changes in an auditory stimulus can be, among others, observed in the decrease of audibility of modulation with increasing modulation rate. Since single-event and modulation thresholds correspond to different natural phenomena and may need to be differentiated for some practical applications, it is important to know the specific methodology of the data collection that lead to specific published values of DLs.

After the auditory stimulus is detected, it can be discriminated from other stimuli on the basis of a number of auditory sensations that can be treated as the attributes of an internal image of the stimulus. The three basic auditory sensations are loudness, pitch, and perceived duration. These sensations are highly correlated with the physical properties of sound intensity, sound frequency, and sound duration, respectively, but they are affected by the other two physical properties of sound as well, e.g., loudness does not only depend on sound intensity but also on sound frequency and sound duration. However, when all physical properties of sound except for the property of interest are held constant, the smallest changes in sound intensity, frequency, or duration can be detected using the sensations of loudness, pitch, and perceived duration. These DLs, when obtained, can be used to measure the acuity of the hearing system with respect to a specific physical variable, e.g., intensity resolution, frequency (spectral) resolution, and temporal resolution, or to determine the smallest change in the stimulus that has practical value for signal and system developers.

Intensity discrimination

The two most frequently discussed differential thresholds in psychoacoustics are the differential threshold for sound intensity (intensity DL) and the differential threshold for sound frequency (frequency DL). The DL for sound intensity is the smallest change in sound intensity level that is required to notice a change in sound loudness.

The DL for sound intensity is typically about 0.5 to 1.0 dB within a wide range of intensities (greater than 20 dB above the threshold) and across many types of stimuli (Chochole and Krutel, 1968; Letowski and Rakowski, 1971; Riesz, 1928). This means that Weber's law holds for both simple and complex sounds and applies to both quiet and natural environments. The intensity DL can be as small as 0.2 dB for pure tones in quiet and sound levels exceeding 50 dB SPL (Pollack, 1954) and reaches up to about 3 dB for natural sounds listened to in natural environment. An example of the relation between the DL for sound intensity and the intensity of the pure tone stimulus is shown in Figure 11-4.

The exponential character of the intensity discrimination function can be approximated for pure tones by:

$$\frac{\Delta p}{p} = \frac{1}{4}\sqrt[6]{\frac{p_o}{p}}, \qquad \text{Equation 11-4}$$

where p is sound pressure, Δp is just noticeable increase in sound pressure and p_o is sound pressure at the threshold (Green, 1988).

The intensity DL for pure tones exceeding 50 dB SPL is fairly independent of frequency but increases for low and high frequencies for sound levels lower than 50 dB SPL, especially lower than 20 dB SPL. When the tone is presented in the background of wideband noise, the intensity DL depends on the signal-to-noise ratio (SNR) for

low SNRs but is independent of SNR for SNRs exceeding 20 dB. For SNRs close to 0 dB, the intensity DL is equal about to 6 to 8 dB (Henning and Bleiwas, 1967). Similar values for intensity DL are reported for the threshold of hearing in quiet.

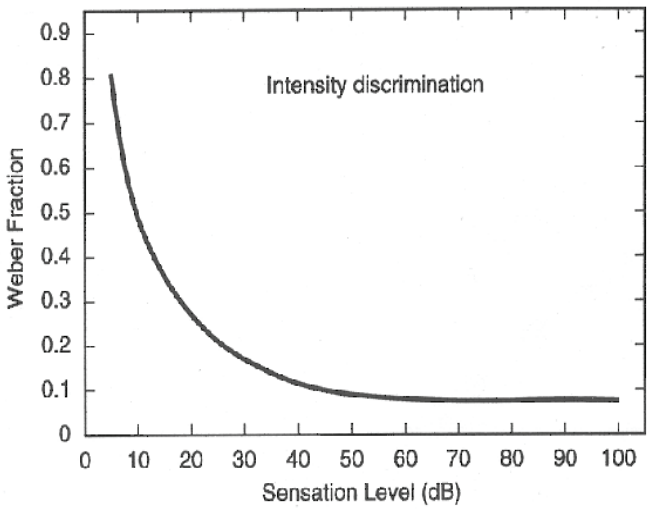

Figure 11-4. Weber fraction for intensity DL as a function of sensation level (i.e., number of decibels above threshold) for a 4 kHz tone. Data from Riesz (1928).

The intensity DL for wideband noises varies from about 0.4 to 0.8 dB depending on the type of noise (Miller, 1947; Pollack, 1951) and rises up to 1 to 3 dB for octave band noises depending on the center frequency of the noise (Small, Bacon and Fozard, 1959).

The Intensity DL depends also on the signal duration. This relationship is exponential and analogous to that of the dependence of stimulus loudness on signal duration (Garner and Miller, 1947b). The intensity DL (in dB) is independent of stimulus duration for durations exceeding 200 millisecond (ms) and increases at a rate of about 3 dB per halving the duration for durations shorter than 200 ms.

Frequency discrimination

The DL for frequency is defined as the minimum detectable change in frequency required detecting a change in pitch. Figure 11-5 presents frequency DL data reported by Wier, Jesteadt and Green. (1977). As can be seen in Figure 11-5 at low frequencies (below 500 Hz) the DL for frequency (in Hz) is relatively independent of frequency and increases logarithmically with frequency at mid and high frequencies. This shape is consistent with Weber's law, i.e., the smallest noticeable change in frequency is a logarithmic function of frequency. For example, the smallest detectable change in frequency is about 1 Hz at 1000 Hz and about 10 Hz at 4000 Hz. In relative terms, the difference threshold at 1000 Hz corresponds to a change of about 0.1% in frequency. However, if expressed in logarithmic units, e.g., *cents* [3] (see the Pitch section later in this chapter), this difference is about 5 cents and remains constant across frequencies.

As shown in Figure 11-5, the frequency DL is dependent on the frequency and intensity of the stimuli being compared. It also depends on the duration and complexity of the stimuli. For tonal stimuli with intensity exceeding 30 dB SPL, average frequency DLs are about 1 to 2 Hz for frequencies below 500 Hz and 0.1 to 0.4%

[3] The *cent* is a logarithmic unit of measure used for musical intervals, often implemented in electronic tuners. Cents are used to measure extremely small intervals or to compare the sizes of comparable intervals in different tuning systems.

for frequencies above 1000 Hz (Koestner and Schenfeld, 1946; König, 1957; Letowski, 1982; Shower and Biddulph, 1931).[4] All these values are typical for average sound intensity levels, and they are the same or slightly smaller for increases in sound intensity up to about 80 dB SL (Wier, Jesteadt and Green, 1977). Similarly, the frequency DL decreases with increasing duration of short auditory stimuli and becomes independent of duration for stimuli exceeding 100 to 200 ms (Grobben 1971; Moore, 1973; Walliser, 1968; 1969). In addition, low frequency sounds need longer duration to be discriminated than high frequency sounds (Liang and Christovich, 1961; Sekey, 1963). Note also a very profound effect of training on frequency discrimination and recognition (Letowski, 1982; Moore, 1973; Smith, 1914).

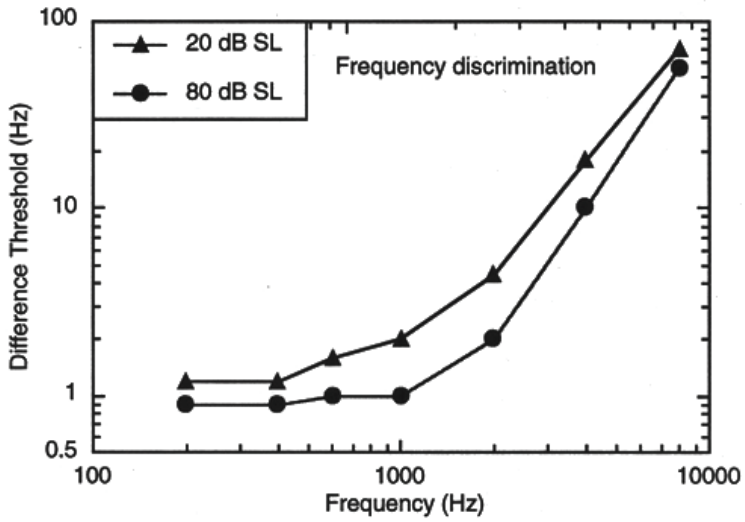

Figure 11-5. Frequency DL as a function of frequency. Data for pure tones presented at 20 and 80 dB SL (i.e. decibels above the threshold of hearing) (adapted from Wier, Jesteadt and Green, 1977).

The frequency DLs for narrow bands of noise are higher that corresponding DLs for pure tones and depend on the bandwidth of noise. According to Michaels (1957), frequency DLs for narrow band noises centered at 800 Hz vary from approximately 3 to 4 Hz for very narrow noises ($\Delta f < 12$ Hz) to more than 6 Hz for a noise band that is 64 Hz wide.

Frequency discrimination for complex tones (fundamental frequency with harmonics) is the same or better than for pure tones (Goldstein, 1973; Henning and Grosberg, 1968). Gockel et al. (2007) reported the frequency DLs as 0.1% and 0.2% for a complex tone and single harmonic, respectively. These values are representative of the frequency DLs found for music notes and vowel sounds, but the actual thresholds vary quite a bit depending on the fundamental frequency of the note and the type of music instrument producing it or the type of voice production, i.e., spectral and temporal envelopes of the sound (Kaernbach and Bering, 2001). However, for practical applications, it can be assumed that frequency DLs for complex tones are approximately constant in the 100 to 5000 Hz range (Wier, Jesteadt and Green, 1977).

Temporal discrimination

Auditory temporal discrimination has various forms and various discrimination thresholds. It refers to the human ability to distinguish between acoustic stimuli or silent intervals of different length, to detect a silent gap in an otherwise continuous stimulus, to resolve between one or two clicks presented in a succession, and to identify

[4] At low frequencies, the DL is constant in Hz; at mid and high frequencies, it is constant in percent (%).

temporal difference and order in the onsets of two overlapping stimuli. The corresponding temporal discrimination measures are called *sound duration DL*, *gap detection threshold*, *temporal resolution*, and *temporal order discrimination*, respectively.

The sound duration DL is the most commonly measured temporal discrimination capability. It depends on sonic content, temporal envelope of sound, and whether it applies to the sound itself or to the pause (gap) between two sounds. Abel (1972) reported duration DLs ranging from approximately 0.4 to 80 ms for stimulus durations of 0.2 and 1000 ms, respectively. In general, the duration DL of uniform (steady-state) sounds follows Weber's Law with a Weber fraction of around 0.1 to 0.2 for time durations greater than about 20 ms (Woodrow, 1951). Sounds with ramped down temporal envelopes are perceived as shorter, and sounds with ramped up temporal envelopes are perceived as longer than those with a uniform envelope (Schlauch, Ries and DiGiovanni, 2001).

A different type of auditory temporal resolution can be assessed by measuring the minimum detectable duration of a gap in a continuous sound. The gap detection is in the order of 2 to 3 ms for tones at moderate and high sound pressure levels (Exner, 1875; Ostroff et al., 2003; Plomp, 1964). Zwicker and Feldtkeller (1967) reported values of 1.5 ms and 5.0 ms for gap detection in tonal signals and noise, respectively. A gap detection threshold exceeding 15 ms is considered abnormal (Keith, Young and McCroskey, 1999). Experiments on gap detection in octave bands of noise have shown that temporal resolution is better at high frequencies than at low frequencies (Shailer and Moore, 1983). At low sound levels, minimum detectable gap duration increases considerably (Florentine and Buus, 1984). If the gap is presented periodically with a frequency $f_{int} \leq 25$ - 40 Hz in a continuous noise, then the noise is heard as a series of separate impulses. However if the frequency f_{int} increases above 25 to 40 Hz, the noise is heard as a continuous noise with a defined pitch corresponding to the frequency of interruptions. The sense of pitch decreases gradually for $f_{int} \geq 250$ Hz and disappears completely for f_{int} above 1000 Hz (Miller and Taylor, 1948).

The minimum detectable gap duration in continuous signals is very similar to the gap duration required to hear two similar clicks as separate events. Such temporal resolution is about 1.5 to 3 ms (Hirsch, 1959; Wallach, Newman and Rosenzweig, 1949), but it may increase to 10 ms for clicks that are greatly dissimilar (Leshowitz, 1971).

Temporal order discrimination requires substantially longer time intervals than temporal resolution or gap detection. The time difference required to determine the order of two sound onsets is to the order of approximately 30 to 60 ms. The actual time depends on gender (shorter for male listeners), age (shorter for young listeners), sound duration and temporal envelope, and whether both stimuli are presented to the same ear or to the opposite ears (dichotic task easier than monotic task) and varies between 20 and 60 ms (Rammsayer and Lustnauer, 1980; Szymaszek, Szelag and Sliwowska, 20006). However, temporal resolution does not seem to be much affected by a hearing loss (Fitzgibbons and Gordon-Salant, 1998). If the sounds overlap in time but have different onset times, they are heard as starting at the different points in time if their onset times differ by more than about 20 ms (Hirsh, 1959; Hirsch and Sherrick, 1961).

It is important to note that perception of the duration of a single acoustic event is affected by the temporal durations of the preceding events as well as by the rhythmic pattern of the events (Fraisse, 1982; Gabrielsson, 1974). For example, the duration of a short pause (gap) between two stimuli (e.g., 250 ms) is underestimated by 25% or more if it is preceded by another shorter pause (Suetomi and Nakajama, 1998). This effect is known as "time shrinking" and is an important element of music perception. It also has been reported that presentation of sound-distracter affects visual time-order perception (Dufour, 1999; McDonald et al., 2005).

Some information about temporal resolution of the auditory system also can be gleaned from data on the auditory detection of amplitude modulation (AM). Viemeister (1979) reported that detection of sinusoidal AM in noise is fairly independent of the modulation rate up to about 50 Hz and gradually decreases beyond this frequency, indicating that fluctuations of noise at higher frequencies are more difficult to detect, i.e., as the modulation rate increases and the time between amplitude peaks of noise becomes shorter, the depth of the modulation must be increased in order for the listener to detect the presence of modulation.

A good example of the practical limits of the auditory system in continuous processing of temporal information is auditory perception of Morse code. Morse code requires discrimination between long (dash) and short (dot) tone pulses separated by short (between symbols) and long (between blocks of symbols) pauses. The specific durations are relative and depend on the individual person, but they are usually in 1:3:1:3 relationships, respectively. Mauk and Buonomano (2004) reported that experts can understand Morse codes at rates of 40 to 80 words per minute (wpm), which for 40 wpm, results in timed events of 30, 90, 30, and 90 ms, respectively.

Cognitive discrimination

The differential thresholds discussed above apply to the smallest change in a single physical dimension that can be measured and assessed using one of the auditory sensations. These thresholds are important for signal and equipment designers and are used to optimize the usability of the products. They are also highly dependent on the overall cognitive capabilities of individual listeners and their familiarity with the situation to be assessed (Deary, Head and Egan, 1989; Helmbold, Troche and Rammasayer, 2005; Smith, 1914; Watson, 1991). However, in many cases the auditory stimuli to be compared differ in more than one physical characteristic, and the differences in their perception cannot be described by loudness, pitch and perceived duration alone. The perceived sounds also may be changing in time as their sources move across space. In such cases, other sensations, such as roughness, sharpness, or spaciousness can be used to differentiate and describe the stimuli of interest. These qualities are part of the domains of timbre and spatial character of sound and will be discussed in later sections of this chapter.

The above approach to evaluating sound events based on the perception of one or more physical dimensions may be quite effective in some situations but will not be sufficient for others. In the latter case, the stimuli can be differentiated at the sensory level using same-different criterion or the differentiation may require higher level processing and cognitive discrimination. For example, an HMD system designer may want to determine if the users will be able to differentiate between the old and new bandwidth of an audio HMD system using a pulsation threshold technique described in Chapter 13, *Auditory Conflicts and Illusions* (Letowski and Smurzynski, 1980). In another study, Nishiguchi et al. (2009) used a paired comparison technique to demonstrate that some listeners are able to discriminate between sounds with and without very high frequency components (f >20 kHz), while the majority of the listeners cannot. Both these studies are examples of the auditory discrimination task. A similar task is involved in differentiating between the sounds of two weapons or two vehicles. However a higher mental processing is required to recognize or identify the specific weapons or vehicles on the basis of their sounds. In an even more complex task performed daily people need to identify large number of speech phonemes in order to communicate by speech. In all these tasks, the listener is required to assign a specific sound to one or more nominal classes of sounds based on the listener's knowledge of the class characteristics. The cognitive processes involved in such decision making are usually described as *classification*, *recognition*, or *identification*.

Loudness

Loudness is an auditory sensation in terms of which sounds may be ordered on a scale extending from soft to loud (ISO, 2006). Loudness depends primarily upon the sound pressure of the stimulus but also depends upon its frequency, temporal envelope, spectral characteristics, and duration (International Electrotechnical Commission [IEC], 1995). Therefore, two sounds that have the same physical intensity (sound pressure, force of vibration) but differ in other physical characteristics may result in different sensations of loudness.

In its common, everyday usage, loudness is a categorical sensation that can be expressed in a number of terms such as *very loud*, *loud*, *soft* and *very soft*. If such a categorical (e.g., Likert) rating scale is used for scientific purposes, it is recommended that the scale has seven steps labeled (ISO, 2006):

Extremely Loud (100) - *Very Loud* (90) - *Loud* (70) -*Medium* (50) -*Soft* (30) - *Very Soft* (10) - *Not Heard* (0)

The numbers in parentheses are numeric values recommended for converting the loudness rating scale into numeric values suitable for averaging several ratings of a single person or a group of judges. In such cases, the minimum number of ratings being averaged should be 20 or higher (ISO, 2006) in order to approximate a Gaussian (normal) distribution in the data set.

Loudness level

Loudness level is a psychoacoustic metric that was developed to determine if sounds that differ in sound pressure (sound intensity) as well as other physical characteristics are equally loud without making a direct comparison for every combination of two of them. The unit of loudness level has been named the *phon*. A sound is said to have a loudness level of N phons if it is equal in loudness to a 1000 Hz tone having a sound pressure (intensity) level of N dB SPL (ANSI, 1994). Thus, a 1-kHz pure tone having a sound pressure level of 60 dB SPL and all other sounds that are equally loud have a loudness level of 60 phons.

The concept of loudness level was introduced primarily to compare the loudness of pure tones of different frequencies. Listeners were given a reference tone of 1 kHz and a test tone of a different frequency and asked to adjust the intensity level test tone until it matched the loudness of the reference tone. Such comparisons lead to the development of equal-loudness contours (iso-loudness curves) (Figure 11-6). The original iso-loudness curves were published by Fletcher and Munson (1933) and became the basis for the current standardized curves approved by the ISO (ISO, 2003). Each curve in Figure 11-6 connects the intensity levels of tones of different frequencies that are equally loud, i.e., the same loudness level in phons. Note that equal-loudness curves flatten gradually with the increase of the sound pressure level. This means that at high intensity levels the ear is less sensitive to fluctuations in intensity as a function of frequency than at low intensity levels.

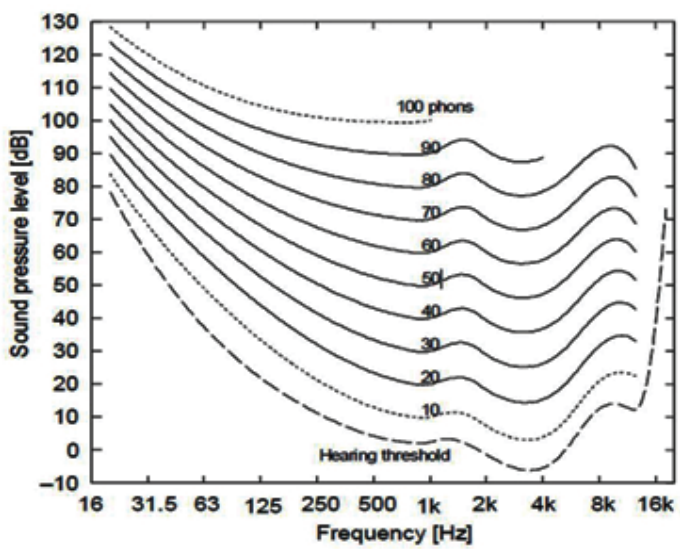

Figure 11-6. Equal-loudness contours for pure tones (adapted from ISO 226, 2003).

The equal-loudness contours for pure tones are not the only equal-loudness contours that have been developed. Similar equal-loudness contours for narrow band noises have been published by Pollack (1952). However, the equal-loudness contours for noises have never gained much popularity and are not widely used.

There are also approximate relationships between the eight formal music dynamic levels and loudness levels for various types of music. An example of such a relationship for symphonic music, based on observations made by Leopold Stokowski in the 1930s, is shown in Table 11-5. The weakness of this relationship is that the music

dynamic levels are relative steps that can be different for each music piece and each music performance, and the relationship shown in Table 11-5 is only an approximation established for an average concert hall performance of symphonic music. The levels for chamber music will be much lower and the levels of rock music much higher (e.g., 140 phons at about 1 meter [3.28 feet] from a loudspeaker).

Table 11-5.
General relationship between music dynamics steps and the loudness levels for a typical concert hall performance of symphonic music (adapted from Slot, 1954).

Dynamic Level	Abbreviation	Loudness Level (phons)
forte fortissimo	Fff	90-100
fortissimo	Ff	80-90
forte	F	70-80
mezzoforte	Mf	60-70
mezzopiano	Pf	50-60
piano	P	40-50
pianissimo	Pp	30-40
piano pianissimo	Ppp	20-30

There also have been some attempts to apply the concept of equal-loudness contours to other perceptual attributes of sound. Fletcher (1934) introduced the concept of pitch level and equal-pitch contours to capture the effect of sound intensity on pitch of sound. Such contours were discussed later by Ward (1954) and Rakowski (1978; 1993). In addition, Thomas (1949) and Guirao and Stevens (1964) attempted to established iso-contours for auditory sensations of volume (Stevens, 1934a) and density (Stevens, 1934b), respectively. All of these attempts were short-lived, and neither triggered any wider interest in scientific community nor found practical applications.

Most comfortable loudness level

Most comfortable loudness (MCL) level has been defined as the listening level selected by the listener to optimize listening pleasure or communication effectiveness. It refers primarily to listening to natural sounds such as music, environmental sounds, and speech. MCL is important for audio HMD design because of the dependence of many perceptual responses on the level (loudness) of incoming stimuli. In almost all practical situations, listening to sound at the MCL results in the best and most consistent human performance. Listening at levels other than the MCL also demands increased attention resources and causes the listener to become fatigued more rapidly. Too high listening levels also may lead to temporary or even permanent hearing loss.

For most listeners the MCL for listening to speech in quiet or low levels of background noise is approximately 60 to 65 dB SPL, which corresponds to the level of normal conversational speech heard at a 1-meter (3.28-foot) distance (Denenberg and Altshuler, 1976; Gardner, 1964; Hochberg, 1975; Kopra and Blosser, 1968; Richards, 1975; Sammeth et al, 1989). This level corresponds roughly to 50 dB HL, which is used in most of the clinical evaluations of speech communication ability. Thus, the MCL of the listener should be the preferred level for speech stimuli delivered through audio HMDs in quiet environments. Speech stimuli also can be presented at both lower and higher levels if they were naturally produced at these levels, and the transmission is intended to truly reproduce the behavior of the talker. For example, natural levels for raised voice (raised speech level), loud speech, and shouting are about 65 to 75 dB SPL, 75 to 85 dB SPL and 85 to 95 dB SPL, respectively (Pearson, Bennett and Fidell, 1977).

One of the most important factors affecting the MCL of a listener for a given listening situation is the level of background noise. Kobayashi et al. (2007) reported that that noise levels up to 40 dB SPL have a negligible effect on the MCL for speech. Above this noise level, the MCL for speech appears to be the level that results in a SNR of approximately 15 dB. However, the fact that conversational speech is at 60 to 65 dB SPL combined with the 15 dB SNR requirement brings the noise levels that are negligible for speech communication to about 50 dB SPL. In addition, at high noise levels, the 15 dB SNR rule cannot be met. Richards (1975) and Beattie and Culibrk (1980) studied MCL levels for speech in noise and concluded that the MCL increases about 7 dB per 10 dB of increase in noise levels, up to about 100 dB SPL.

The MCLs for listening to music are substantially higher than those for speech and depend on the type of music, surrounding acoustics, and type of music instrument. Individual differences in MCLs for music are larger than those for speech and can vary from about 70 to 95 dB SPL.

MCLs are usually expressed as the sound intensity (pressure) level selected by the listener. However, they also can be expressed in phons. When expressed in phons, they become less dependent on the specific sound and are easily transferable to other listening situations. The typical MCL (in phons) for various types of music as calculated by the authors on the basis of several MCL studies (Gabrielsson and Sjögren, 1976; Martin and Grover, 1976; McDermott, 1969; Sone et al., 1994; Suzuki, Sone and Kanasashi, 1982; Staffeldt, 1974; Steinke, 1958) are:

- Symphonic and big-band music: 85 phons
- Solo and chamber music: 75 phons
- Artistic speech and solo singing: 65 phons

The selection of a very high listening level (above 85 phons) frequently makes the listening experience more exciting as opposed to remote (Freyer and Lee, 1980). However, it also makes the perceived sound image less clear due to nonlinear distortions generated in the middle ear and is more tiring (Kameoka and Kuriyagawa, 1966).

One area that requires special attention in respect to MCL is the perceptual assessment of sound, i.e., perceived sound quality (PSQ). Illényi and Korpassy (1981) conducted a number of listening tests of loudspeakers and demonstrated that louder sounds lead to higher ratings of the sound quality of the loudspeaker. This requires very careful loudness balance in PSQ assessment of sounds produced by different sound sources. It is also important for proper PSQ judgments that the sounds need to be reproduced at their natural levels (Gabrielsson and Sjögren, 1976; 1979) or at the ultimate listening levels, if such levels are known (Toole, 1982).

Loudness scale

Sensation of loudness is a perceptual representation of the amount of stimulation and depends primarily on sound intensity. In order to determine the effect of sound intensity on loudness, some type of psychophysical relationship between these two variables needs to be determined. One type of such a relationship is provided by the loudness level that allows comparing loudness of two or more sounds by comparing them to the equivalent loudness of a 1 kHz tone.

However, it does not allow one to determine how much louder one sound is with respect to another one. For example, the fact that one sound has a loudness level of 75 phons and another sound has a loudness level of 83 phones does not make it possible, by itself, to determine how much louder the second sound is. Such a comparison requires the direct representation of both sounds on a quantitative psychophysical loudness scale.

The first attempt to create a quantitative loudness scale was by Fechner (1860), who extended Weber's Law and assumed that the sensation of loudness increases by a constant amount each time the stimulus is increased by one DL. This dependence results in a logarithmic relationship between loudness (L) and sound intensity (I) having interval scale properties and is referred to as Fechner's Law or Weber-Fechner's Law:

$$L = a \times \log(I) + b, \qquad \text{Equation 11-5}$$

where a and b are constants dependent on the type of sound and a particular listener. The unit of loudness on Fechner's loudness scale is 1 DL, and the change of the stimulus intensity by 3 dB results in doubling of loudness. This relationship has been experimentally confirmed at low intensity levels at and slightly above the threshold of hearing where doubling of loudness requires a 2 to 4 dB increase in sound intensity. However, it overestimates the growth of loudness at higher levels. Research by Newman, Volkmann and Stevens (1937), Stevens (1955) and others led to the observation that for moderate and high intensity levels the loudness of a 1-kHz tone doubles when its sound pressure level increases by about 10 dB (Stevens, 1955). Thus, the shape of the loudness scale for the 1 kHz tone has been determined by Stevens to be a power function of the tone sound pressure level described as:

$$L = kI^{0.3} = kp^{0.6}, \qquad \text{Equation 11-6}$$

where L is loudness of sound, I is sound intensity, p is sound pressure, and k is the coefficient of proportionality that accounts for individual differences (Stevens, 1972a). This functional relationship sometimes is referred to in the literature as the Power Law of Loudness.

Since the 1-kHz tone serves as a reference sound for the loudness level, this means that loudness doubles when the loudness level increases by 10 phons. Therefore, in order to determine how much louder one sound is than another, one needs to determine the loudness levels of both sounds and compare them on the loudness scale for the 1-kHz tone.

The unit of loudness expressed by Equation 11-6 is a *sone*, defined as the loudness of a 1-kHz tone having a sound level of 40 dB SPL. Thus, the loudness of 1 sone corresponds to the loudness level of 40 phons. A sound that is N times louder has a loudness of N; and a sound that is N times softer has a loudness of 1/N sones. The relationship between loudness (L) and loudness level (LL) is such that a doubling of L in sones occurs for each increase in LL of 10 phons and can be written as:

$$L = 2^{\frac{LL-40}{10}}. \qquad \text{Equation 11-7}$$

The actual functional relationship between L and LL based on the data collected by Hellman and Zwislocki (1961) and other researchers is shown in Figure 11-7.

The function described by Equations 11-6 and 11-7 is shown in Figure 11-7 as a straight line and it matches experimental data very well for loudness levels of 30 phones or more. The curved portion of the loudness function indicates that at the threshold of hearing loudness grows more rapidly than at higher levels. This growth can be approximated by the modified loudness function equation:

$$L = k(p - p_0)^{0.6}, \qquad \text{Equation 11-8}$$

where p_o is the sound pressure at the threshold (Scharf, 1978). Although physiologic mechanisms behind the growth of the loudness function are not entirely clear, they have been related to both the overall number and timing of neural discharges (Carney, 1994; Fletcher, 1940; Relkin and Doucet, 1997) and the nonlinearities of both cochlear and central parts of the auditory system (Schlauch, DiGiovanni and Ries, 1998; Zeng and Shannon, 1994).

It has to be added that individuals with neurophysiologic hearing loss have elevated thresholds of hearing but still have about the same or even lower threshold of pain. These shifts in thresholds result in a narrower dynamic range of hearing and in a loudness function that has to have a steeper slope than that of normally hearing

individuals. This rapid increase in loudness function associated with neurophysiologic hearing loss is called *recruitment*.

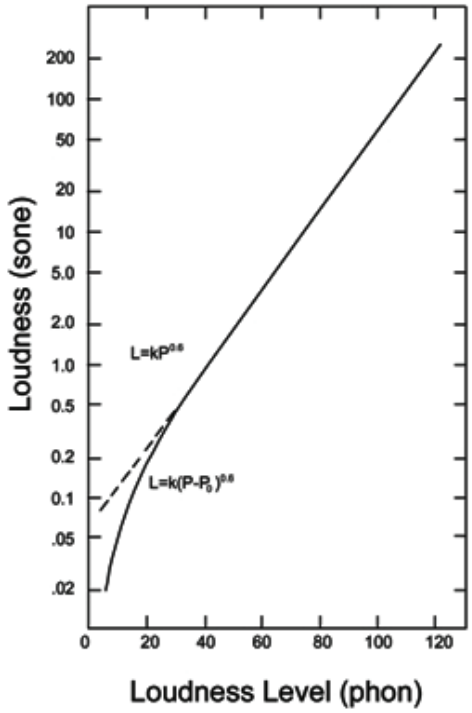

Figure 11-7. Binaural loudness of a 1 kHz tone as a function of loudness level (adapted from Scharf, 1978).

The discussion above assumes one-ear listening and the concept of monaural loudness. There is still a debate in the literature regarding the difference between *monaural* and *binaural* loudness. Marozeau et al. (2006) demonstrated that the difference between monaural and binaural loudness is practically independent of the sound pressure level. However, some researchers (e.g., Fletcher and Munson, 1933; Hellman, 1991; Marks, 1978; Pollack, 1948) reported that an increase in sound loudness due to binaural listening is equivalent to a 3 dB change in sound intensity received monaurally (doubling of sound intensity) while some others concluded that this change is more likely to be in 1.3 to 1.7-dB range (Scharf and Fishken, 1970; Wilby, Florentine, Wagner and Marozeau, 2006; Zwicker and Zwicker, 1991). This summation process seems to parallel an approximate 1.1 times (0.4 dB) binocular visual acuity and a 1.4 times (1.5 dB) contrast sensitivity advantage phenomenon in binocular vision (Rabin, 1995).

Temporal integration

The thresholds of hearing presented in Figures 11-1 and 11-2 were determined using continuous (long) pure tone stimuli; therefore, they are independent of sound duration. The same is true for the loudness functions expressed by Equations 11-6 and 11-8. However, for short sounds, both the threshold of hearing and sound loudness are affected by sound duration. The relationship between stimulus duration and the perceptual effects of the stimulus is referred to in the literature as *temporal integration* or *temporal summation*, and the changes in perceptual effects with stimulus duration have been attributed to temporal summation of excitations in the auditory system (Zwislocki, 1960).

The maximum duration of the stimulus through which the temporal summation effect operates is called *critical duration.* According to many studies, the critical duration for pure tone signals is approximately 200 to 300 ms, although this value depends somewhat on sound frequency (Miskolczy-Foder, 1959; Sanders and Honig, 1967, Zwislocki, 1960). The threshold of hearing is higher for durations shorter than the critical duration and decreases at a rate of about 3 dB per doubling of duration (Zwislocki, 1960). For example, for 100-µsec square-wave clicks presented at the rate of 10 Hz, the threshold of hearing is in the order of 35 dB SPL (Stapells, Picton and Smith, 1982), while the hearing threshold for continuous white noise is near 0 dB SPL. The functional relationship between the threshold of hearing and the stimulus duration for a 1 kHz tone is shown in Figure 11-8.

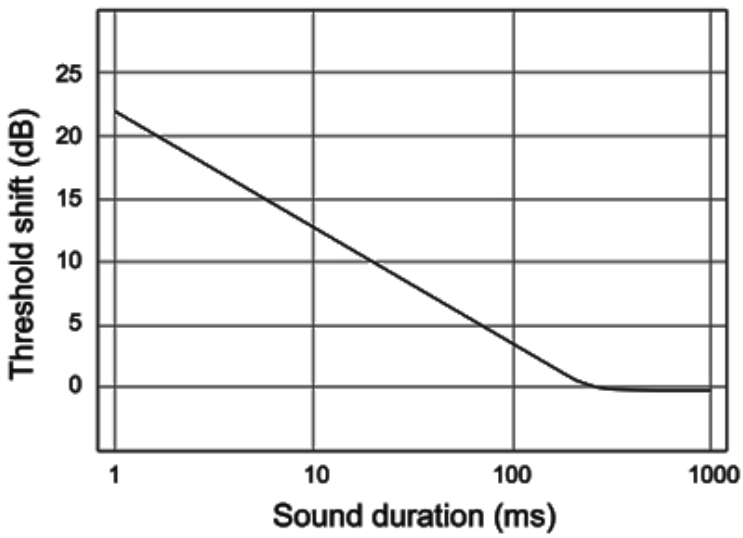

Figure 11-8. The effect of stimulus duration on the threshold of hearing for a 1-kHz tone (adapted from Zwislocki, 1960).

The temporal integration of energy in the auditory system also operates at the above-threshold (suprathreshold) levels, affecting sound loudness. For sounds shorter than critical duration, the loudness of sound increases with sound duration and this relationship can be described as:

$$SIL \times T = \text{constant loudness} \qquad \text{Equation 11-9}$$

where SIL is sound intensity level in dB, and T is stimulus duration (in seconds) (Garner and Miller, 1947). Plomp and Bouman (1959) concluded that loudness is an exponential function of the duration of the sound and depends on the relationship between the stimulus duration and the time constant of the ear (determined to be 50 ms). According to these researchers, tonal stimuli that last for durations of 50 ms and 200 ms produce sensations of loudness that are equal to 62.70% and 99.98% of the loudness produced by a continuous sound, respectively.

The signal does not need to be a single short sound impulse to be affected by the mechanism of temporal integration. Series of clicks or short bursts of noise also are affected by the mechanism of temporal summation. However, bursts of higher repetition rate and shorter duration have been reported to sound louder than the same bursts of longer duration and slower repetition rate (Garner, 1948; Pollack, 1958). This effect may be attributed to an increasing neural firing rate by a group of neurons with increasing number of sound onsets. This increase in the firing rate seems to more than offset the effect of time latency (rest period) in a single neuron firing rate and should result in a decrease in sound loudness (Zwislocki, 1969).

Loudness summation

One important factor affecting loudness of a stimulus is the distribution of sound energy across the auditory frequency range. The loudness of a sound depends on where along the frequency scale the sound energy is located and how concentrated or dispersed is its allocation. Sound energy located in the area of greatest ear sensitivity (the lowest region for a given equal-loudness contour) contributes the most to sound loudness. The distribution of sound energy along the frequency scale affects the manner in which the auditory system integrates spectral components of the stimulus. This process is called *loudness summation* or, more accurately, *spectral integration* of sound energy by the auditory system.

Several algorithms have been proposed to model spectral integration of sound energy process in the development of the sensation of loudness. Some of the algorithms have been proposed by Fletcher and Munson (1933), Beranek et al. (1951), Howes (1971) and Stevens (1956). Further research led to observations that the process of spectral integration of sound is closely associated with the concept of critical bands (discussed later in this chapter). Briefly, if the sound components are located within a narrow frequency band smaller than a single critical band, the total loudness of sound is proportional to the total sound energy contained within the band. If the sound components are separated further apart than a critical band, the sound loudness is the sum of the loudnesses of the individual components. The two modern algorithms of loudness summation based on the general concept of critical band have been developed by Zwicker (Zwicker and Feldtkeller, 1955; Zwicker, 1960; Zwicker and Scharf, 1965) and Moore and Glasberg (Moore and Glasberg, 1996; Moore, Glasberg, and Baer, 1997) (see sections on Critical Bands and Loudness Scale for additional discussion on spectral summation and binaural summation, respectively.)

Auditory adaptation and fatigue

Auditory adaptation, or loudness adaptation, is a gradual decrease in hearing sensitivity during sustained, fixed-level, auditory stimulation. As shown in Figure 11-8, due to the effect of temporal integration, the sensation of loudness increases gradually with sound duration and reaches its terminal value for sounds longer than 200 to 300 ms. However, if the auditory stimulus acts for a prolonged period of time, the sensation of loudness slightly decreases. The decrease in sound loudness is accompanied by some decrease in hearing sensitivity for frequencies outside the frequency range of stimulation (Thwing, 1955).

The amount of adaptation is dependent on the frequency, level and duration of the auditory stimulus and increases with decreasing level of the stimulus and increasing frequency (Scharf, 1983; Tang, Liu and Zeng, 2006). Several early studies indicated strong auditory adaptation at all signal levels (e. g., Hood, 1950), but more recent studies demonstrated that under most listening conditions the auditory adaptation at high intensity levels is relatively minimal (Canévet et al., 1981). For example, Hellman, Miśkiewicz and Scharf (1997) reported that over the period of several minutes, the loudness of a continuous fixed level pure tone can decrease by 70% to 100% at 5 dB SL, 20% at 40 dB SL and stays practically constant at higher SLs. The exceptions are frequencies above 10 kHz, where the auditory adaptation effect is quite strong at both low and high stimulation levels (Miśkiewicz et al., 1992).

The physiologic mechanism responsible for auditory adaptation is still not clear. One possibility is a "restricted excitation pattern" mechanism proposed by Scharf (Miśkiewicz et al., 1992; Scharf, 1983). According to this concept, all low-level stimuli regardless of their frequency and all high-frequency stimuli regardless of their level produce a more restricted excitation pattern along the basilar membrane that is subjected to more adaptation than respective high-level and low-frequency stimuli.

Auditory adaptation needs to be differentiated from auditory fatigue. Fatigue is a loss of sensitivity as a result of auditory stimulation, manifesting itself as a temporary shift in the auditory threshold after termination of the stimulus. It is often referred to as a *temporary threshold shift* (TTS) and appears gradually for sounds exceeding 70 dB SPL. It differs from adaptation in two important ways. First, it is measured after the auditory stimulus has

ended (poststimulatory fatigue); whereas auditory adaptation is measured while the adapting stimulus is still present (peristimulatory adaptation). Second, as a loss of sensitivity (rather than a shift in perception), it is a traumatic response to excessive stimulation by intense auditory stimuli (continuous noise above 85 dB or impulse noise above 140 dB). Exposure to recurring or extreme acoustic trauma can result in permanent hearing loss.

Masking

Sounds very rarely occur in isolation. They are usually heard as signals in the background of other sounds or are themselves a part of the background. The concurrent, or in close succession, presence of two or more sounds causes the audibility of the individual sounds to be adversely affected by the presence of other sounds. This adverse effect is called *masking*. Masking is defined as: (a) a process by which the threshold of hearing for one sound is raised by the presence of another sound and (b) the amount by which the threshold of hearing for one sound is elevated by the presence of another sound (ANSI, 1994).

A *masker* is a sound that affects audibility of another sound, the *target sound* (or *maskee*). More intense sounds mask less intense sounds. Masking effect of a target sound by a masker may be total, making the target sound inaudible, or partial, making it less loud. It should be noted that the masking phenomenon affects not only other sounds but also all individual components of a single complex sound. If the target sound is not completely masked by a given masker, the additional amplification of the masker needed to completely mask the target sound is called the *masking margin* (MM). The concept of the MM applies, among others, to the design of sound masking systems intended to provide privacy and security of acoustic information without creating excessively high noise levels.

The maskers can be of two types: *energetic* maskers, which physically affect the audibility of the target sound, and *informational* maskers, which have masking capabilities due to their similarity to the target sound. In general, both of these masking phenomena may exist together and may be caused by the same stimulus, but they are frequently considered separately due to the difference in the way they affect the audibility and identity of the target sound. Energetic masking is peripheral masking caused by the overlap of the excitation patterns created by the target sound and the masker along the basilar membrane and is considered to be a peripheral type of masking. Informational masking is related to non-energetic characteristics of the masker and may take place even if there is no overlap in the excitation patterns caused by the target stimulus and the masker along the basilar membrane. This type of masking is considered to originate in the central auditory nervous system.

A phenomenon very similar to masking and difficult to differentiate from masking is *perceptual fusion*. The concept of fusion applies mostly to complex sounds that have several qualities that need to be attended separately. In fusion and in masking, the distinct qualities of a target sound, or its partial loudness, are lost, and the physiological mechanisms underlying both phenomena are the same. Thus, both phenomena are most likely two different views of the same physiological process. In the masking approach, the focus of the observation is on the audibility of a single target sound; while in the fusion approach, the focus is on both the masker and the target sound, i.e., whether the masker and the target (masked) sound the same as the masker alone, or not. (Bregman, 1990; Schubert, 1978)

As with most of the auditory phenomena, masking can be monaural or binaural. However, if both the masker and the target sound are delivered to both ears, the target sound audibility is very much the same as in the case of monaural listening assuming that both ears are fairly identical. A common situation is that the masker affects both ears, and the target sound is only available at one of the ears. The reverse situation is also possible and, in such case, may affect localization of the sound source producing the target sound by masking the sound in one of the ears.

In addition, masking can be ipsilateral (masker and target (masked) sound in the same ear) or contralateral (masker and target sound in the opposite ears) (also known as peripheral and central masking, respectively). Ipsilateral masking is much stronger than contralateral masking, but the latter is frequently used to prevent sound

leakage to the opposite ear (e.g., in bone conduction hearing tests). The difference in the effectiveness of both masking modes is in the order of 50 dB.

Energetic masking

The basic form of masking is related to sound energy and its distribution in frequency and time domains. This form of masking is called *energetic masking* (EM). There are two basic forms of energetic masking: *simultaneous* masking and *temporal* masking. Temporal masking is further divided into forward and backward masking. The other types of energetic masking discussed in the psychoacoustic literature, such as an overshoot masking, are just combinations of the two basic forms of energetic masking.

Simultaneous masking

Simultaneous masking is masking caused by a masker that is present throughout and possibly beyond the duration of the target sound. It is the most effective form of energetic masking. The amount of masking is dependent on the sound intensity of the masker and its spectral proximity to the target sound. Therefore, this form of masking is sometimes also referred to as spectral masking.

When the masker contains sufficient energy in the frequency region of the target sound, the masked threshold increases about 10 dB for every 10 dB increase in masker. Such relation between masker and masked threshold of the target sound can be observed when a pure tone is masked by wideband noise (Hawkins and Stevens, 1950; Zwicker and Fastl, 1999). This situation is shown in Figure 11-9.

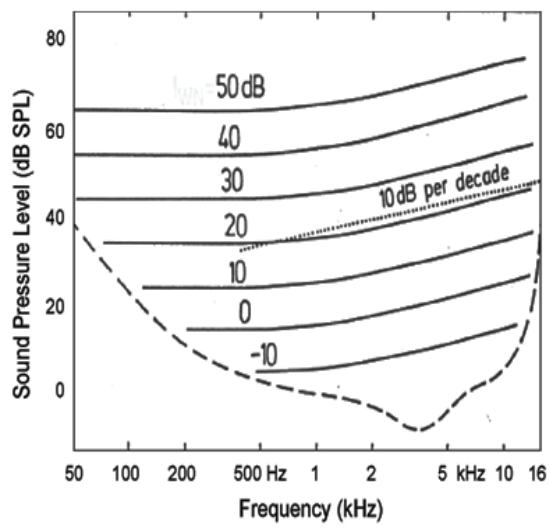

Figure 11-9. Detection thresholds of pure tones masked by white noise as a function of the frequency. Horizontal lines show masked thresholds for noise density levels from -10 to 50 dB and their relation to the threshold of hearing in quiet (dashed line) (adapted from Zwicker and Fastl, 1999).

The noise spectrum (spectral density) levels listed in Figure 11-9 indicate the density per Hz of white noise stimulus used as the masker. The masked threshold curves produced by white noise are fairly independent of frequency up to about 500 Hz and then increase with frequency at a rate of approximately 3 dB/octave (10 dB/decade). The equally-masking noise, which has constant density per Hz up to 500 Hz and then constant density per octave, i.e., density per Hz decreasing at a rate 3 dB/octave, would result at higher frequencies in practically frequency-independent masked threshold curves being parallel to the frequency axis. Other noises will result in quite different masking contours.

Masking produced by a continuous stationary noise is the simplest and most common form of energetic masking. The most common broadband noises that can be used as maskers in audio HMD testing (depending on the field application) are listed in Table 11-6. White noise and pink noise – noise that has the same power per relative (Δf/f) bandwidth – together with the equally-masking noise are frequently used as maskers in laboratory studies because they are well defined mathematically, and their effects on the audibility of the individual frequency components in the target sound are relatively easy to quantify. In addition, white noise and pink noise represent two important classes of real world maskers, e.g., thermal noise (e.g., heat noise, power generator noise, fan noise) and environmental noise (1/f noise).

Table 11-6.
Common wideband noises used for research purposes (adapted from Internet webpage the *Colors of Noise* [http://en.Wikipedia.org/wiki/Colors_of_noise.html]).

Noise Name	Description	Comments
Black Noise	No noise	Silence
Blue Noise	Noise that has a frequency spectrum envelope that changes proportionally to frequency. Blue noise has a spectral power density that increases by 3 dB per octave.	
Brown Noise	Noise with frequency spectrum envelope that changes proportionally $1/f^2$. Brown noise has a spectral power density that decreases by 6 dB per octave.	This name refers to Brownian motion that has these specific properties; also called Red Noise
Equally Masking Noise	Noise that equally masks tones of all frequencies	Also called Gray Noise
Pink Noise	Noise with frequency spectrum envelope that changes proportionally $1/f^2$. Pink noise has a spectral power density that decreases by 6 dB per octave.	
Purple Noise	Noise that has a frequency spectrum envelope that changes proportionally to f^2. Purple noise has a spectral power density that increases by 6 dB per octave.	Also called Violet Noise
White Noise	Noise that has flat frequency spectrum envelope. White noise has a constant spectral power density per Hz.	Acoustic analog of white light

Masking situations where a pure tone is masked by a narrow band of noise or another pure tone are shown in Figure 11-10. When the masking stimulus is a narrow band of noise the elevation of the threshold of hearing for pure tone target sounds is the greatest about the centre frequency of the noise. The masked threshold gradually and smoothly decreases for both low and high frequency target tones.

When both masker and the target sound are pure tones and have similar frequencies, they create beats (Egan and Hake, 1950; Wegel and Lane, 1924). Beats are periodic changes in sound intensity resembling amplitude modulation. When beats appear, they are heard as a tone with basic frequency f_o, which is the mean frequency of the two beating frequencies of the masker f_2 and target f_1:

$$f_o = (f_1 + f_2)/2 \qquad \text{Equation 11-10}$$

and the frequency of (f_{beats}) is equal to the difference between the frequencies of the masker and target:

$$f_{beats} = f_2 - f_1 \ . \qquad \text{Equation 11-11}$$

The presence of beats makes it easier for the listener to detect the target tone, even when the masker level is relatively high. These situations are shown in Figure 11-10 as dips in the masking curve around the frequency of the masker (400 Hz) and its harmonics (800 and 12000 Hz). The presence of beatings at the harmonic frequencies of the masker reveals the existence of nonlinear processes in the ear and the presence of the aural harmonics in the processed sound.

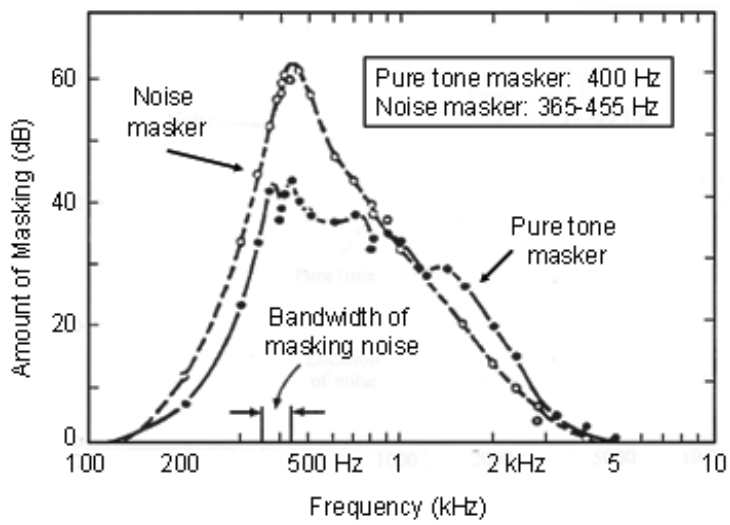

Figure 11-10. Masking effects of a 400 Hz pure tone and a narrow band of noise centered at 400 Hz (adapted from Egan and Hake, 1950).

The shape of the masked thresholds in Figure 11-10 shows that masking effect extends further in the high frequency region than in the low frequency region. In other words the upward spread of masking is much greater than the downward spread of masking, and this disproportional growth increases with the increase in the intensity of the masker. This situation is shown in Figure 11-11. The presence of the upper spread of masking also means that low frequency stimuli mask better high frequency stimuli better than the reverse.

In general, masking varies as a function of the frequency content of the masker. The closer the masker and target sound are on the frequency scale, the greater the masking. Thus, a narrowband noise centered on the frequency of a pure tone will have the greatest masking effect on that pure tone. As the bandwidth of the narrowband masker increases, its masking effectiveness increases until its bandwidth exceeds the limits of the critical band (see the later section on Critical Bands). However, further increase of the bandwidth of noise beyond the width of the critical band does not increase the masking power of the noise (Fletcher, 1940; Hamilton, 1957; Greenwood, 1961a,b). This can be explained by the fact that noise energy within the critical band prevents detection of the target sound because both the target sound and the masker are being passed to the same auditory system filter (auditory channel). However, noise energy outside of the critical band has no effect on detection of target sound because they pass through different filters. In addition, Buus et al. (1986) reported a 6 to 7 dB difference between the thresholds of detection for a pure tone (220, 1110, or 3850 Hz) and for an 18-tone complex tone[5] of uniform intensity when both were being masked by the same 64 dB SPL equally masking noise. The complex tone was detected easier. This finding indicates that simultaneous presence of signal energy in several critical bands aids signal detection.

[5] A *complex tone* is a sound consisting of several, usually harmonically related, pure tones.

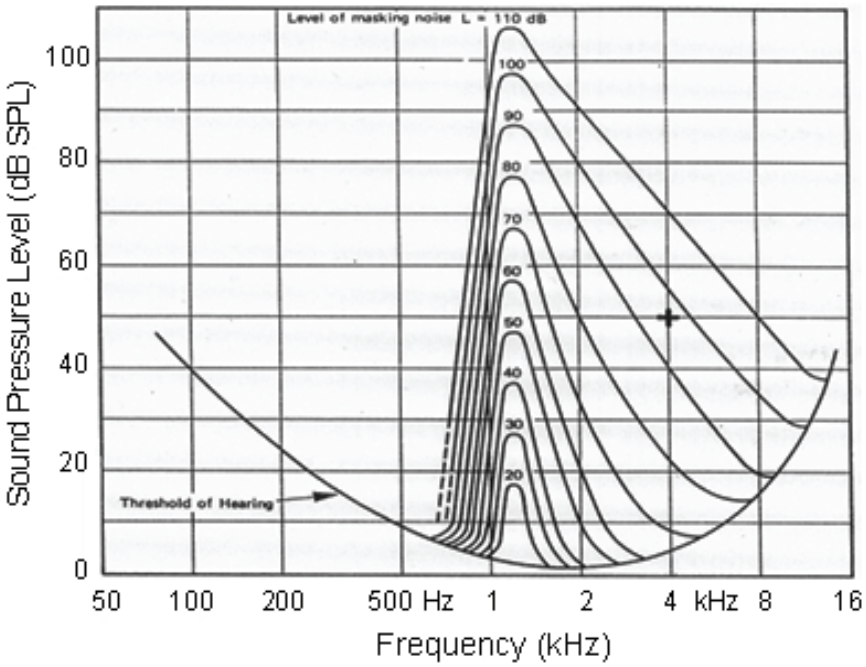

Figure 11-11. Masking effect of a narrow band of noise centered at 1200 Hz. The level of masking noise is shown next to each masked threshold (adapted from Zwicker and Feldtkeller, 1967).

Temporal masking

Masking caused by sounds that are not simultaneous with the target sound is called *temporal masking*. When two sounds arrive at the listener in short succession, the listener may hear only one sound event due to limited temporal resolution of the hearing system. However, if one of the two sounds has much higher sound intensity than the other, the listener still may hear only the more intense sound, even if the time difference between the sounds is above the temporal resolution limit of the hearing system.

There are two forms of temporal masking: forward (post-stimulatory) masking and backward (pre-stimulatory) masking. Forward masking appears when a short target sound is played after the end of the masker sound. If the time difference between the offsets of masker and target sound is very short, the sensory trace left by the masker decreases hearing sensitivity to the target stimulus resulting in its masking. The level of forward masking is dependent on the intensity of the masker, spectral similarity between the masker and target sound, and the time difference between the offsets of both sounds. Masking decreases as the intensity of the masker decreases, the separation between the sounds increases, and the time difference between the two offsets increases. In general, the increase of the wide band noise masker by 10 dB causes the increase of the detection threshold for immediate following tone by about 3 dB. Little masking occurs for times longer than 200 ms (Fastl, 1976; Jesteadt, Bacon and Lehman, 1982). The time difference between the offset of a masker and the onset of the target sound is inappropriate as a variable describing forward masking because the listener may still detect the target sound by hearing its end.

It has been demonstrated that the level of forward masking exerted by one tone on the subsequent tone can be decreased if additional tone is added to the masker in the region outside of the critical band of the target tone (Houtgast, 1974; Shannon, 1976). A similar but smaller effect can be observed in simultaneous masking (Fastl and Bechly, 1983). This phenomenon has been labeled *spectral unmasking* (Shannon, 1976) and is probably a result of physiological suppression of the excitatory response to the first tone by the addition of another tone (Sachs and Kiang, 1968).

Backward masking appears when the target sound is presented just before the masker. As with forward masking, the amount of backward masking is dependent on the intensity of the masker, spectral similarity between the masker and target sound, and the time difference between the offsets of both sounds. However, the time interval between the onsets of both sounds during which backward masking takes place rarely exceeds 25 ms. Although a large number of studies have been published on backward masking, the physiologic basis of this phenomenon is still largely unknown. Moore (1997) observed that, contrary to forward masking, the amount of backward masking decreases substantially with listener's experience and argued that backward masking may result from some kind of "confusion" between the target sound and the masker. The observed effects of forward and backward masking are illustrated in Figure 11-12.

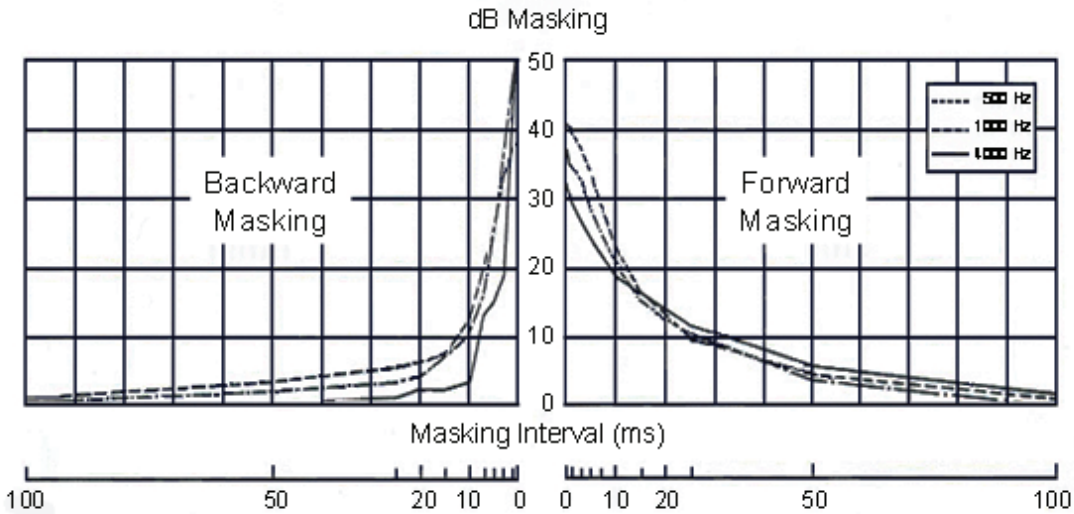

Figure 11-12. The relationship between the amount of backward (left panel) and forward (right panel) masking and time interval between the masker and target sound (adapted from Elliott, 1962).

Temporal masking, and especially forward masking, plays an important role in auditory perception because it degrades temporal cues in perceived stimuli. For example, in speech perception, a strong vowel may mask a weak consonant following or preceding the vowel. In addition, if speech communication takes place in a sound field, a strong reflection from a nearby wall may mask subsequent weak sound arriving along the direct pathway.

Informational masking

Informational masking (IM) is the amount of masking of one stimulus by another that cannot be explained by the presence of energetic masking. In other words, informational masking is the masking caused by the characteristics of the masker other than its energy. The amount of informational masking can be determined as a difference between the overall masking level and the masking level due to the energetic masking only. For example, *multitalker noise* (MTN), also known as speech babble, can serve as both energetic and informational maskers of a speech target, whereas a random (or frozen) white noise with speech spectrum envelope of the actual MTN can serve as an approximation of pure energetic masker.

Two main causes of informational masking are similarity between the masker and the target sound and the variability (uncertainty) of the masker (Durlach et al., 2003). The concept of informational masking originated in 1970s and was initially associated with the effect of spectro-temporal variability (uncertainty) of the masker or target on detection of the target stimulus (Dirks and Bower, 1968; Pollack, 1995; Watson, Kelly and Wroton, 1976). This concept later was expanded to include similarity between the masker and the target sound and spatial uncertainty regarding the location of the masker (Durlach et al., 2003). It has been demonstrated that the decrease

in the degree of similarity between the target sound and the masker reduces substantially the amount of informational masking affecting the target sound (Kidd et al., 1994; Micheyl et al., 2000).

The reason that MTN is such an effective informational masker of speech is its overall similarity to the target speech. However, its actual effectiveness depends on the number of voices constituting the MTN, gender of the talkers, synchrony and rate of speech of the MTN voices, and the overall similarity of speech patterns of the MTN and the target speech. For example, masking effectiveness of an MTN increases with the number of voices, reaches its plateau for about 10 voices and then declines. Conversely, the content of the spoken messages, being a positive, neutral, or negative content, does not seem to have bearing on masking effectiveness of a MTN (Letowski et al., 1993; Letowski et al., 2001).

Informational masking due to masker uncertainty may be a result of either spectro-temporal uncertainty, spatial uncertainty, or both. Random variations in a masker spectrum outside of the protected zone located in close vicinity of the target stimulus have been reported to cause as much as 20 to 40 dB of additional masking. Numerous studies demonstrating the presence of additional masking caused by spectro-temporal uncertainty have been cited by Durlach et al. (2005). However, this increase reflects the joint effect of masker-target similarity and masker uncertainty. It can be argued that masker-target similarity is still the main cause of the masking increase shown in the reported studies. For example, Durlach et al. (2003) observed that masker-target similarity seems to greatly increase the effect of masker uncertainty on its masking effectiveness. Lufti (1990) analyzed a number of masking studies with naturally varying masking noise in each masking trial and concluded that the amount of informational masking in these studies was about 22% of the overall masking. The effect of spectro-temporal variability of the masker on the overall amount of masking also has been shown by Pfafflin and Matthews (1966), Pfafflin (1968), Lufti (1986) and others who compared effectiveness of natural random noise with that of the fixed (frozen) noise played in each masking trial.

Similarly, it has been shown that uncertainty of the spatial position of the masker can reduce speech intelligibility of the speech target (Kidd et al., 2007) or detection of the nonspeech target (Fan, Streeter and Durlach, 2008). Evidence of informational spatial masking in speech-on-speech masking situations can be found in frequent errors in substituting target words with words contained in the masking message. However, as compared to the effects of masker-target similarity or even spectral uncertainty, the effect of spatial uncertainty is very small. It can be argued that spectro-temporal and spatial uncertainties of the masker cause uncertainty about the target sound template or distract the attention of the listener, drawing it away from the target sound (Best et al., 2005; Conway, Cowan and Bunting, 2001).

It is also important to note that the amount of informational masking caused by a specific masker-target relationship is highly dependent on the listener, and that the inter-subject differences in respect to informational masking are very large (Durlach et al., 2003). Equally important is that the effectiveness of informational masking increases with age even in people with otologically normal hearing. Rajan and Cainer (2008) reported that with aging, independent of any hearing loss, older individuals (age 60 or greater) performed as well as younger individuals in speech recognition in an energetic masking background but performed much poorer in the presence of informational maskers. The authors attributed this difference to an age-related increase in competing-signal interference in the processing of auditory and phonetic cues.

Critical Bands

The concept of *critical band* is central to understanding the mechanisms of sound processing in the auditory system. This concept was introduced by Fletcher (Fletcher and Munson, 1933, 1937; Fletcher, 1948, 1940) to account for filtering actions of the human auditory system. Fletcher et al. studied loudness summation and masking of tones by various wideband noises and observed that only noise energy contained in a certain frequency band centered on the frequency of the pure tone contributes to the tone masking. They also noticed that the loudness of the tones separated by the width of this band is additive, while the loudness of the tones within this bandwidth is not. They called this bandwidth the critical band.

Fletcher (1940) originally assumed that to mask a tone, the total power of the masking noise has to be equal to the power of the tone and defined the critical band as a bandwidth of noise having power equal to the power of the tone:

$$P = N \times CB \,, \qquad \text{Equation 11-12}$$

where P is the power of a tone, N is the noise spectrum (noise spectrum density) level, and CB is the bandwidth of the noise that contributes to the masking effect, i.e., the critical band width. This concept is shown graphically in Figure 11-13.

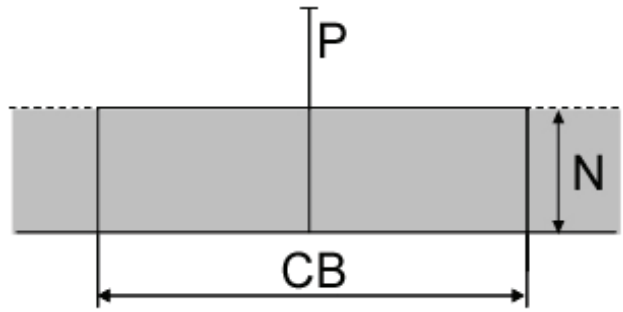

Figure 11-13. Fletcher's concept of the critical band. P – power of the tone (dB), N – noise spectrum level (dB), CB – critical band (Hz).

Equation 11-12 also can be written as:

$$CB = \frac{P}{N} \qquad \text{Equation 11-13}$$

And, after taking the logarithm of both sides and multiplying it by 10:

$$10 \log CB = 10 \log \frac{P}{N} = CB(dB) = CR \qquad \text{Equation 11-14}$$

where CB(dB) is a critical band expressed in dB, which is currently called the *critical ratio* (CR).

Critical ratio specifies the number of dB by which the power of the tone needs to exceed the noise spectrum level in order for the tone to be detected. For example, according to Fletcher's concept of critical bands, for a tone with a frequency of 1000 Hz, CB equals 65 Hz , and CR equals 18.1 dB.

Fletcher's concept of the critical bands was revised in 1950s by Zwicker when it was determined that in order to make a tone inaudible, the power of the masking noise needs to be about 2.5 times (4 dB) greater than the power of the masked tone (Zwicker, 1952;1961; Zwicker, Flottorp and Stevens, 1957). This finding extended the width of the critical bands by a factor of approximately 2.5. The new width of the critical bands also was confirmed in experiments on the threshold of hearing (Gässler, 1954; Hamilton, 1957; Zwicker and Feldtkeller, 1955) and loudness (Gässler, 1954; Zwicker, 1952) of complex sounds. For example, the relationship between the threshold of hearing at 1100 Hz and the bandwidth of the auditory stimulus reported by Gässler (1954) is shown in Figure 11-14. Gässler measured the threshold of hearing for a multi-tone complex composed of from 1 to 40 equal-amplitude pure tones evenly spaced 10 or 20 Hz apart. As the tones were added sequentially to the complex, the overall sound pressure at the threshold of hearing remained constant up to some defined width of the bandwidth. When the tones were added beyond this width, the overall sound pressure needed to elicit a threshold sensation increased with a slope of 3 dB per doubling of the signal bandwidth outside of the critical band. When

additional components were added symmetrically on both sides of the critical band, the threshold increased at a rate of 1.5 dB per doubling of the signal bandwidth outside of the critical band.(Spiegel, 1979). These findings are consistent with the predictions of an energy-detector model of the auditory system.

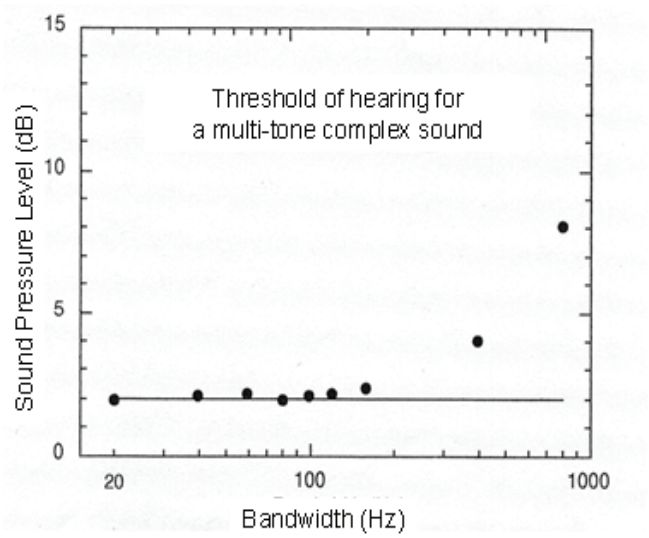

Figure 11-14. Threshold of hearing for a multi-tone complex as a function of bandwidth. Data are shown for tones added every 20 Hz below 1100 Hz. Continuous line shows the threshold of hearing for a single 1100 Hz tone (adapted from Gässler, 1954).

Similarly, the loudness of the complex auditory stimulus, with sound energy contained within a single critical band, is independent of the distribution of sound energy within the band. The effects of critical band on the loudness of a narrowband noise with a bandwidth changing from very narrow one to one that is wider than a critical band is shown in Figure 11-15.

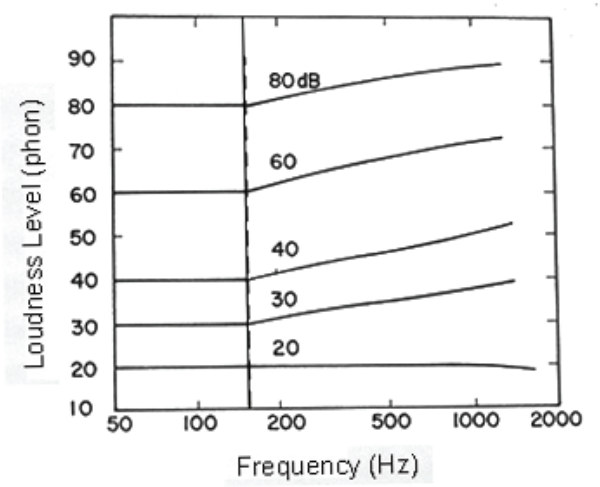

Figure 11-15. Loudness level of a narrow band of noise as a function of noise bandwidth. Numbers on the curves indicate the overall sound intensity level of the band (adapted from Scharf, 1978).

There is also very little effect on loudness due to the number of spectral components contained within a single critical band as long as the total energy of the complex remains unchanged. For example, several researchers have

reported no difference in the loudness of two-tone complexes, four-tone complexes, and a broadband stimulus for stimuli contained within the same critical band (Zwicker and Feldtkeller, 1955, Feldtkeller and Zwicker, 1956; Zwicker et al., 1957, Scharf, 1959). Others have found a slightly higher loudness of a broadband noise in comparison to the loudness of the tonal stimuli, particularly at loudness levels near 65 phons (Florentine, Buus and Bonding, 1978). The overall loudness of two tones that are separated by less than 20 Hz is affected by the audible changes in sound intensity caused by beats and is dependent on the phase relationship between the tones (Zwicker, Flottorp and Stevens, 1957). More information about auditory system sensitivity to phase is included in the later section Phase and Polarity.

The size of Zwicker's critical bands (*Frequenzgruppen*) is about 100 Hz for frequencies below 500 Hz and increases with frequency *f* at about the 0.2*f* rate; this relationship is shown in Figure 11-16. Thus, the bandwidth of the critical band above 500 Hz can be roughly approximated by the bandwidth of 1/4 octave filters ($\Delta f = 0.18f$) with the same center frequency.

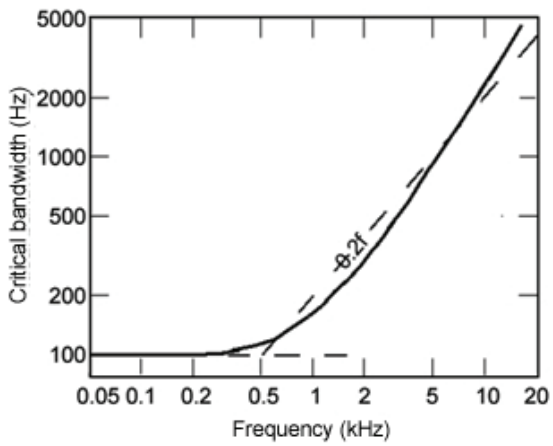

Figure 11-16. Critical bandwidth as a function of frequency (adapted from Zwicker and Fastl, 1999).

Since von Békésy's studies of basilar membrane and its tonotopic organization in 1920s and 1930s (Békésy, 1960), the term critical band is used also to denote regions of the basilar membrane that respond to stimulation by a sine wave input. Zwicker (1952) observed that when subsequent 24 CBs are placed back-to-back they cover almost the whole rage of hearing (0 to 15,500 Hz) and can be represented conveniently along the basilar membrane. He also demonstrated that a CB takes a relatively constant length of 1.3 mm along the basilar membrane and can be used as a unit of frequency along the basilar membrane. It also corresponds to about 1300 neurons in the cochlea (Zwislocki, 1965). This unit has been named the *bark* in honor of German physicist Heinrich Barkhausen who initiated perceptual measurements of loudness (Zwicker, 1961). The bark scale extends from 1 bark to 24 barks, and its functional relationship with frequency is shown in Figure 11-17.

The relationship between CB (in Hz) and the specific frequency *f* of the tonal stimulus, i.e., the center of the CB, as well as the distance *x* (in mm) of the point of maximal excitation on the basilar membrane from the oval window, can be calculated using a formula proposed by Greenwood (1961b):

$$CB = 22.9\,(0.006046\,f + 1) = 22.9 \times 10^{0.06x} \qquad \text{Equation 11-15}$$

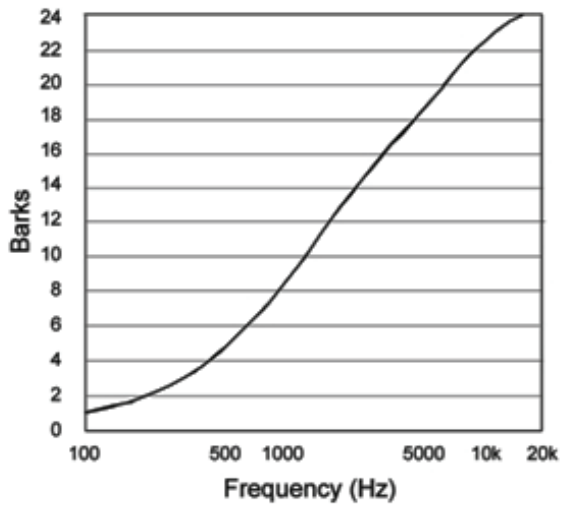

Figure 11-17. Critical band rate or barks as a function of frequency (adapted from Zwicker and Terhardt, 1980).

This relationship between the bark scale and the frequency scale shown in Figure 11-17 can be expressed mathematically as (Zwicker and Terhardt, 1980):

$$z = [13\arctan(0.76f) + 3.5\arctan\left(\frac{f}{56.25}\right)^2], \quad \text{Equation 11-16}$$

where z is the distance along the basilar membrane in barks, and f is frequency of the stimulus in kHz. Other formulae to calculate the width of CB in Hz and in barks for specific frequencies have been published by Zwicker and Terhardt (1980) and Traunmüller (1990). This equation can be reformulated to calculate stimulus frequency for a known location on the bark scale and expressed as (Lubman, 1992):

$$f = \{[(\frac{e^{0.219z}}{352} + 0.1)z] - [0.032e^{[-0.15(z-5)^2]}]\} \quad \text{Equation 11-17}$$

Barks are used frequently in modeling and simulations as an input to models of pitch perception, masking, and loudness summation, and noise hazard. The widths and lower and upper limits of critical bands for the 24 steps of the bark scale are listed in Table 11-7.

It is still unclear what the shape of the critical band filters is and whether it depends on sound intensity (e.g., Fletcher and Munson, 1937 (Figure 17); French and Steinberg, 1947 (Figure 8); Glasberg and Moore, 1990; Greenwood, 1961a,b). As with each mechanistic entity, such a filter has to have skirts with finite slopes. However, for many practical applications, it is convenient to assume that critical bands are brick-wall filters[6] with rectangular shapes. In their revision of Zwicker's loudness model, Moore and Glasberg (Glasberg and Moore, 1990; Moore and Glasberg, 1983; 1996; Moore, Glasberg and Baer, 1997) derived such a filter shape for critical bands in order to better account for the shape of the equal-loudness contours in low frequency range and the loudness of partially masked sounds. In their model of loudness summation Moore and Glasberg introduced the concept of the *equivalent rectangular bandwidth* (ERB) as a replacement for the critical band (bark) scale. The ERB is the bandwidth of a rectangular filter that has the same peak transmission as the auditory filter for that

[6] *Brick-wall filter* is an informal term for an idealized electronic filter, having full transmission in the pass band, complete attenuation in the stop band, and an abrupt transition(s) between the two bands.

frequency and passes the same total power for a white noise input (Moore, 1997). Its bandwidth varies as a function of frequency as:

$$ERB = 24.7(4.37f + 1) \qquad \text{Equation 11-18}$$

where f is the center frequency of the ERB filter. The function in Equation 11-18 has the same shape as the function (Equation 11-17) proposed by Greenwood for CBs and differs only in respect to constant values. A comparison of the critical bandwidths and the ERBs is shown in Figure 11-18.

Table 11-7.
Critical bands corresponding to 24 steps of the bark scale (adapted from Zwicker and Feldtkeller, 1967).

Bark band	Lower limit Frequency (Hz)	Center Frequency (Hz)	Bandwidth Δf (Hz)	Upper limit Frequency (Hz)
1	20	50	80	100
2	100	150	100	200
3	200	250	100	300
4	300	350	100	400
5	400	450	110	510
6	510	570	120	630
7	630	700	140	770
8	770	840	150	920
9	920	1000	160	1080
10	1080	1170	190	1270
11	1270	1370	210	1480
12	1480	1600	240	1720
13	1720	1850	280	2000
14	2000	2150	320	2320
15	2320	2500	380	2700
16	2700	2900	450	3150
17	3150	3400	50	4700
18	4700	4000	700	4400
19	4400	4800	900	5300
20	5300	5800	1100	6400
21	6400	7000	1300	7700
22	7700	8500	1800	9500
23	9500	10500	2500	12000
24	12000	13500	3500	15500

The shape of the critical bands (CBs) function in Figure 11-18 ERB is the same as in Figure 11-16 and the shape of ERB function is described by Equation 11-18. In some cases it is also convenient to think about ERBs as units of the frequency scale analogous to barks. The functional relationship between the number of ERBs, and the specific frequency is given by:

$$E = 21.4 \times \log(4.37f + 1), \qquad \text{Equation 11-19}$$

where E is the number of ERBs, and f is frequency in kHz. The constants of integration have been chosen to make E=0 where f=0 (Glasberg and Moore, 1990).

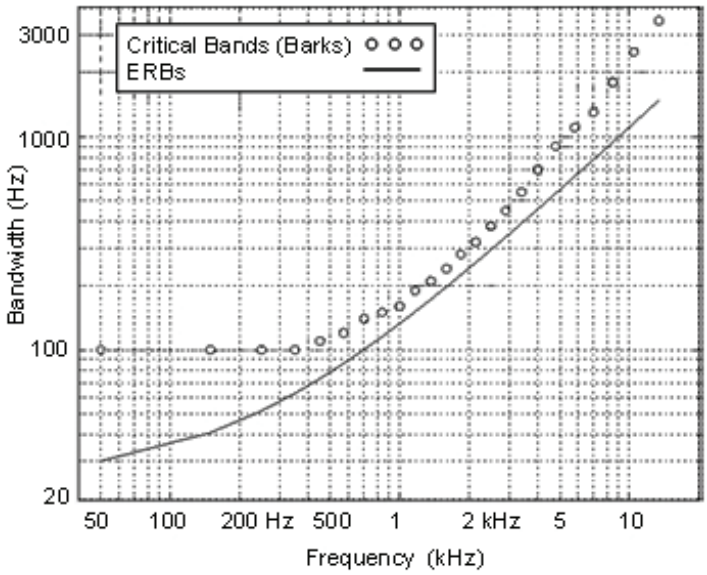

Figure 11-18. Critical bandwidth (Bark scale) and equivalent rectangular bandwidth (ERB) as a function of frequency (adapted from Smith and Abel, 1999).

Pitch

Pitch is the perceptual correlate of frequency. It is a sensation that the sound has a specific physical frequency. According to the formal ANSI standard definition of pitch, it is an auditory sensation that can be ordered on a scale extending from low to high (ANSI, 1994). Thus, low frequency pure tones are heard as being low in pitch, and high frequency pure tones are heard as being high in pitch. However, most sounds that occur are not pure tones, and yet many of them, but not all, have an identifiable pitch. Thus, pitch and frequency are not related in a simple one-to-one manner but depend on other physical properties of the stimulus, such as spectral complexity and intensity. Pitch is also a much more complex sensation than the sensations of loudness or perceived duration and actually has a multidimensional character

The sensations of pitch and rhythm are the foundation of music, which is a deliberate rhythmic sequence of sounds that differ in their pitch and/or timbre. The concept of pitch in music is closely related to the concepts of the musical scale and music intervals. The musical scale is a succession of selected notes (frequencies) arranged in ascending or descending order. The scale represents a specific music system and includes all the notes that are allowed to be used in this system, e.g., pentatonic system (5 notes in an octave), diatonic system (7 notes), or chromatic system (12 notes). The *key* or *tonic* of the scale is the first tone in the scale, and all subsequent tones are defined by simple ratio multiples of the tonic, e.g., 2:1 (octave), 3:2 (Major 5th), 3:4 (Major 4th), 5:4 (Major 3rd), or 6:5 (minor 3rd). The frequency ratios (pitch differences) within a given scale are referred to as intervals, and in many traditional music systems, they are not the exact multiples of each other. For example, in the diatonic scale, there are two unequal whole tone intervals 9:8 (major whole tone) and 10:9 (minor whole tone). Thus, because of these strict ratio relationship requirements, a musical instrument tuned to a particular key (tonic) would require retuning if one changed the key up or down a step. For instruments like the piano, or its earlier cousins, the clavier or the harpsichord, this was an onerous task. In the early 18th century, it became common for Western music to be written using an equally-tempered scale in which each octave is divided into 12 steps (semitones) (Helmholtz, 1863). These semitones are further divided into *cents*. Each semitone is 100 cents, and an octave is 1200 cents. The advantage of the equally-tempered scale is that any key can be played without changing the tuning of the instrument, and any song using the Western musical system can be written out using this notation.

The lower the frequency ratio, i.e., the smaller the numbers describing this ratio, the more similar in the character are the two notes separated by the interval (Galilei, 1638). The smallest possible frequency ratio is the ratio 2:1=2, which is called an octave. The octave has a special meaning in music because all sounds that are separated by one or more octaves fuse together very well and are sometimes very hard to differentiate from one another (Shepard, 1964). All other music intervals, such as semitone, tone, major third, or perfect fifth, are well defined within an octave and have the same sonic quality, called tone chroma, when repeated in other octaves. This octave equivalence led to the naming convention used in Western music, such that the notes (frequencies) that are an octave apart are named with the same letter (e.g., C, D, E) or syllable (e.g., do, re, mi) (Justus and Bharucha, 2002).

The concepts of music scale and octave similarity (tone chroma) led to the recognition of the two-dimensional character of pitch: pitch height and pitch class. Pitch height is the pitch defined in the ANSI standard (used at the beginning of this section). It is a continuous dimension logarithmically related to stimulus frequency. Therefore, a sound at 440 Hz (A4) is perceived as being equidistant from both 220 Hz (A3) and 880 Hz (A5). As all other auditory sensations, pitch height depends also to some degree on other basic physical parameters of sound, e.g., sound intensity and duration.

Pitch class, or tone chroma, is a dimension arranging music intervals from the smallest to the largest within a single octave. So, a middle C in the Western music system is lower in pitch height than the C one octave above it, but they occupy the same position on the pitch class scale. This terminology captures the circular nature of pitch that is a foundation of most of Western music. Both pitch dimensions, pitch height and pitch class, can be combined together in one helical representation of pitch.

One of the most remarkable properties of the human auditory system is its ability to extract pitch from complex tones. If a group of pure tones, equally spaced in frequency are presented together, a pitch corresponding to the common frequency distance between the individual components will be heard. For example, if the pure tones with frequencies of 700, 800, and 900 Hz are presented together, the result is a complex sound with an underlying pitch corresponding to that of a 100 Hz tone. Since there is no physical energy at the frequency of 100 Hz in the complex, such a pitch sensation is called *residual pitch* or *virtual pitch* (Schouten 1940; Schouten, Ritsma and Cardozo, 1961). Licklider (1954) demonstrated that both the *plac*e (spectral) pitch and the *residual* (virtual) pitch have the same properties and cannot be auditorally differentiated. In a harmonic tone, such as descried above, the residual pitch is often described as pitch corresponding to a *missing fundamental* frequency. The sensation of residual pitch is the main evidence that the auditory system must be able to code frequency based on its periodicity (see Chapter 9, *Auditory Function*). It also invalidates the so-called Ohm's Acoustic Law, which states that "Each tone of different pitch in a complex sound originates from the objective existence of that frequency in the Fourier analysis of the acoustic wave pattern."

Note, also, that a listener may listen to a complex sound in two different ways: analytically and synthetically. When listening analytically, the listener is focused on individual components of the sound and may hear their individual pitches. When listening synthetically, or holistically, the listener perceives the sound as a whole and pays attention only to the fundamental (or residual) pitch (Smoorenburg, 1970). So, the listener who listens analytically to a complex tone with a missing fundamental may not immediately recognize its residual pitch.

In reality, the dimensions of pitch height and pitch class (tone chroma) are not the only two dimensions of pitch. Another dimension of pitch is the pitch strength. Pitch height and pitch class are sufficient to describe the relationship between pure tones but not between complex natural, synthetic, and speech sounds. Sounds are usually composed of a number of frequency components that may be in harmonic or inharmonic relationships. It is the relationship of these components that determines whether a sound is tonal – i.e., carries the pitch of its fundamental frequency, or atonal. Most, although not all, of the music sounds are presumed to be tonal, however, outside of the realm of music many sounds contain frequencies that are not in harmonic relations. There are also musical instruments that produce sounds with inharmonic overtones (partials). The degree to which a specific sound has an identifiable pitch is called its *pitch strength* (Rakowski, 1977). Fastl and Stoll (1979) asked listeners to complete a magnitude estimation task for a number of test sounds, including pure tones, low-pass complex

tones, complex tones, narrow-band noise and various other kinds of modulated or filtered noises. The general findings were that sounds with an orderly pattern of harmonics had the strongest pitch, as well as those containing a narrow band of frequencies. The pitch strength ranking of some test sounds investigated by Fastl and Stoll (1979) is shown in Table 11-8. The sounds with a more random and/or broadband spectral content have only a faint or no pitch strength. Shofner and Selas (2002) summarized their findings by stating that pitch strength depends primarily on the fine structure of the waveform and secondarily on the stimulus envelope. The relative perceptual salience of pitch in tonal complexes can be also estimated using an algorithm developed by Terhardt et al. (1982). In the case of residual pitch, the pitch strength decreases with an increase of the average frequency of the tonal complex and is the strongest for harmonics in the region of the third, fourth, and fifth harmonic (Boer, de, 1956; Ritsma, 1967; Ritsma and Bilsen, 1970).

Table 11-8.
Pitch strength rankings for 11 test sounds as obtained by Fastl and Stoll (1979).

Pitch Strength Ranking	Test Sound
1	Pure tone
2	Complex tone: -3 dB/octave low pass
3	Complex tone: -3 dB/octave
4	Narrow-band noise: Δf = 10Hz
5	AM tone: m=1
6	Complex tone: Band Pass
7	Band-pass noise: 96 dB/octave
8	Low-pass noise: 192 dB/octave
9	Comb-filtered noise: d=40 dB
10	AM noise: m=1
11	High-pass noise: 192 dB/octave

The Western music tonal system has influenced heavily the human concept of the pitch height scale, which is based on the logarithmic scaling of frequency perceived as pitch. Octave intervals are said to have the same pitch class (tonal chroma) and serve as equal steps of the music scale. However, it does not mean that they are perceptually equal although they are frequently treated that way. In order to answer this question Stevens, Volkmann and Newman (1937) constructed perceptual scale of pitch asking listeners to adjust a pure tone stimulus until it sounded half as high as a comparison stimulus (ratio scaling). They also proposed the *mel* (from the word "melody") as a unit of the pitch scale. The mel has been defined as 1/1000th of the pitch of a 1000 Hz tone presented at 40 dB HL. Thus, the pitch of a 1000 Hz tone at 40 dB HL is equal to 1000 mels and equal numeric distances on the pitch scale were defined as equal perceptual distances although the developed scale should be treated more like a ratio scale. Later, Stevens and Volkmann (1960) conducted a similar study asking the listeners to divide frequency range from 200 to 6500 Hz into four equal intervals (interval scale). This new pitch scale used the same reference point of 1000 mels at 1000 Hz as the previous scale but was truly an interval scale. Due to the difference in testing methodology, the scales ware not identical, and the new scale was compressed heavily above 1000 Hz in comparison to the old scale. The relationship between pitch and frequency arrived at by Stevens and Volkmann (1940) is:

$$m = 1127 \times \ln\left(1 + \frac{f}{700}\right), \qquad \text{Equation 11-20}$$

where m is pitch in mels, and f is the frequency in Hz.

The pitch scale based on mels differs from the music scale and has been criticized for this reason. The musical objections to the pitch scale are that it is counterintuitive and counterproductive for different octaves to have different perceptual size (Hartmann, 1997). An explanation of confusion between pitch doubling and octave similarity might be found in the complex tones that are generated by musical instruments, which commonly consist of frequency components having a harmonic relationship to each other. These same harmonic relationships are the basis of the scales upon which musical structure is formed. Further, music pieces consisting of multiple voices depend on these same harmonic relationships to create consonance and dissonance. Thus, the entire structure of Western music depends on the mathematical relationships of frequency components of the complex tones within it. It does not depend on whether or not those pitches are perceived as being equidistant. In reality, an octave from 50 to 100 Hz sounds perceptually smaller than the octave from 2000 to 4000 Hz, and the frequency of 1300 Hz is reported by several investigators as having a half of the pitch of frequency of 8000 Hz (Zwicker and Fastl, 1999). These observations support the concept that pitch has actually two separate dimensions: pitch height (measured in mels) and pitch class (measured in music intervals).

In addition to two pitch scales developed by Stevens and his colleagues, Zwicker and Fastl (1999) constructed a third scale, also based on mels, using ratio scaling and a reference point of 125 mels set at 125 Hz. The scale extends form 0 to 2400 mels for the frequency region from 20 Hz to about 16 kHz and is shown in Figure 11-19.

Below about 500 Hz, the mel scale and the Hz scale are roughly equivalent, and the standard tuning frequency of 440 Hz has pitch of 440 mels. Above 500 Hz, larger and larger intervals are considered to be equivalent. As a result, four octaves on the frequency scale above 500 Hz correspond to about two octaves on the pitch mel scale. For example, the frequency of 8000 Hz has pitch equal to 2100 mels, while the frequency of 1300 Hz has pitch of 1050 mels. This relationship agrees well with the earlier experimental finding that a tone of 1300 Hz has half of the pitch of a 8000 Hz tone.

The pitch scale developed by Zwicker and his colleagues is highly correlated with the bark scale and with the distribution of excitation patterns along the basilar membrane. One can note the remarkable similarity between the bark scale (Figure 11-17) and the mel scale shown in Figure 11-19. This similarity indicates that the mel scale is practically parallel to the bark scale, and the unit of 1 bark corresponds to 100 mels. Both scales also are related to the ERB scale (Moore and Glasberg, 1983) and to the distance along the basilar membrane (Greenwood, 1961; 1990). It must be stressed, however, that all these scales have been developed for pure tones and do not directly apply to speech and music. Their major role is to help to understand frequency decoding and pitch encoding by the auditory system.

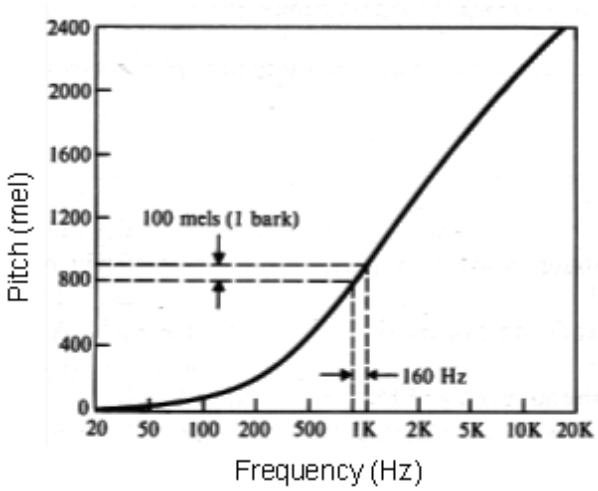

Figure 11-19. The relationship of the mel scale to frequency (adapted from Wightman and Green, 1974).

The relationships between frequency, pitch, critical bands, and the distance along the basilar membrane are shown together in Figure 11-20.

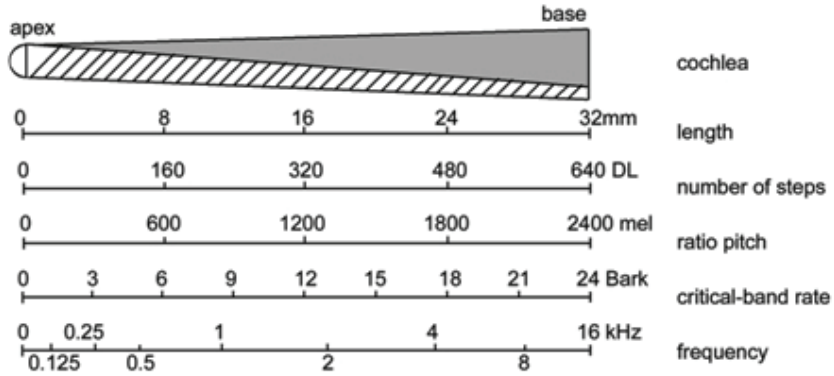

Figure 11-20. Similarity between various psychophysical scales and distribution of neurons along basilar membrane. Note that the scales of the length of the basilar membrane, numbers of DL steps, ratio pitch, and barks are linearly related while scale while frequency is not (adapted from Zwicker and Fastl, 1999).

One of the important concepts in music and everyday sound perception is the concept of consonance and dissonance. Music intervals, chords, or any combination of frequencies may be pleasant or unpleasant to the listener. The pleasant sounds are called *consonant sounds* and those that are unpleasant are called *dissonant sounds*. Dissonance occurs if the frequency separation between the individual tones of the sound is smaller than a critical band with its maximum for tones separation equal about ¼ of the critical band (Plomp and Levelt, 1965). This separation corresponds to about 20 Hz for lower and about 4% for higher sound frequencies. Helmholtz (1863) attributed the perception of dissonance to the sensation of beats or the roughness of sound, and Stumpf (1911) attributed it to perception of sound fusion, i.e., to ability of two sounds to assume a new identity, independent of their individual identities, when heard together (see section on Masking). Roughness and the dissonance, according to Helmholtz, are more likely represented in the auditory cortex by neural responses phase-locked to the amplitude-modulated temporal envelope of complex sound (Fishman et al., (2001).

In addition to the relationship to frequency discussed above, the pitch of a sound is dependent on its intensity and duration. Stevens (1935) and Gullick (1971) demonstrated that for middle frequencies (1 to 2 kHz in Stevens' and 2.5 kHz in Gullick's case), the pitch of a pure tone is independent of the stimulus intensity. However, for tones of higher frequencies, increased sound intensity produces an increase in pitch. Conversely, for tones of lower frequencies, increased sound intensity produces a decrease in pitch. Gullick (1971) reported that both shifts are similar for frequencies equidistant form the reference frequency of 2.5 kHz tone if expressed in terms of frequency DLs but not Hz. For example, a change in sound intensity of 40 dB resulted in similar but opposite shifts by 7 DLs for tones of 700 and 7000 Hz. For music, the effect of sound intensity on sound pitch is much smaller than for pure tones and is of the order of 17 cents for 30 dB change in sound intensity (Rossing, 1989). The direction of the change depends on the dominant components of the sound (Terhardt, 1979).

The effects of sound intensity on perceived pitch reported by Stevens (1936) and Gullick (1971) were measured by presenting two static tones of different frequencies and intensities and asking the listener to adjust the frequency or intensity of one of the tones until they seemed equal in pitch. However, sounds also can shift dynamically in both the frequency and intensity like in the case, for example, of the Doppler effect (Doppler shift). The Doppler effect is the change in the frequency of the arriving at the listener sound produced by a moving sound source. As the sound source approaches the listener, the compressions and rarefaction of the produced sound wave become compressed, making the frequency of the sound reaching the listener's ears higher

than the frequency of the actually emitted sound. As the sound source passes the listener and moves away, the distances between the compressions and rarefactions of the sound wave became stretched, making the frequency of the sound reaching the listener's ears lower that the frequency of the sound produced by the departing sound source. So, if a sound source emitting a sound of frequency f_o is traveling at a constant velocity directly toward the listener, the sound that reaches the listener's ears has higher frequency than the frequency of the sound produced by the sound source but the difference between both frequencies is constant until the sound source reaches the listener's position. As the sound source passes the listener, the frequency of the propagating sound will drop suddenly. As a result, as the sound source moves away from the listener, the frequency of the sound that reaches the listener's ears is lower than that of the emitted sound. During the same time, the intensity of sound arriving at the listener's ear will gradually rise as the sound source is approaching the listener and gradually fall as it sound source moves away. A common example given of this effect is that of the sound of a passing vehicle using a classical siren.

Neuhoff and McBeath (1996) studied the effect of the Doppler shift on the pitch perceived by the listeners and found that the majority of the listeners reported that a Doppler shift consists of a rising (sound source is approaching the listener) and then falling (sound source is moving away from the listener) pitch shift. They then presented listeners with 27 Doppler shifted tones, created using three frequencies (220, 932 and 2093 Hz), three levels of spectral complexity (sinusoid, square wave, tone complex), and three velocities (10, 15 and 20 meters/second [m/s]). Listeners tracked the pitch using a pitch wheel. Listeners reported hearing a rise in pitch on 70% of all trials. The only conditions where a rising frequency was not reported were those of the lowest frequency pure tone. The probability of reporting a rise in pitch increased as a function of frequency and spectral complexity. The contour of the reported rise in the perceived pitch occurred synchronously with the rise in intensity suggesting that listeners perceived the rising sound intensity as an upward shift in pitch. This was true even at the two lowest frequencies, i.e., the frequencies where there should be no shift or downward shift in frequency according to Stevens (1955) and Gullick (1971). This situation is shown in Figure 11-21. Note also that if the sound source is traveling slightly at the angle to the listener's position, there is a slight relative decrease in the velocity of the sound source as it gets closer to the listener. However, this change will result in be a small decrease and not increase in sound frequency arriving at the listener's ears.

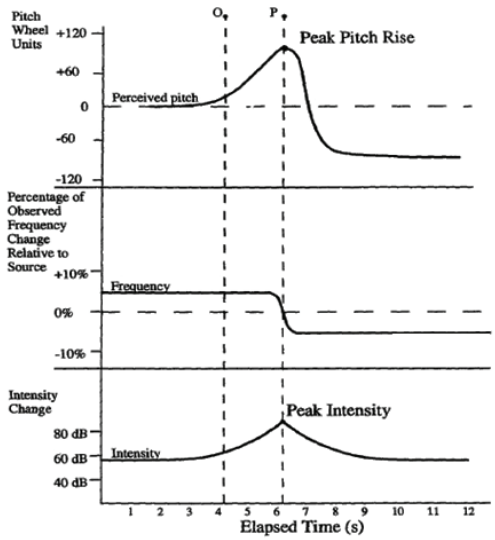

Figure 11-21. Schematic representation of the stimuli used in Neuhoff and McBeath's (1996).study. The bottom frame shows the intensity of the sound at the listener's ears. The middle frame shows the frequency at the listener's ears. The top frame shows the perceived pitch as listeners reported it using a pitch wheel (used with permission from Neuhoff and McBeath, 1996).

To test the hypothesis that the reported effect was due to dynamically changing sound intensity, Neuhoff and McBeath (1996) then asked the listeners to select the higher pitch tone of pairs of static tones consisting of a loud, lower frequency tone and a soft, higher frequency tone. For static tones, listeners accurately judged pitch, suggesting that the dynamic changes in both the intensity and frequency of the Doppler shifted tones are responsible for their perceptual interaction. Neuhoff's data suggest that pitch and loudness are perceived integrally (i.e., changes in one dimension can be perceived as changes in the other), a finding supported later by other research (Grau and Kemler-Nelson, 1988: Scharine, 2002). From a practical standpoint, the interrelationship of two perceptual dimensions underscores the complexity of pitch scaling and suggests that signal designers must exercise caution in using frequency as the basis for presenting dynamically changing information as its perception can be easily influenced by secondary factors. These situations are discussed further in Chapter 14, *Auditory Conflict and Illusions*.

The minimal duration of a pure tone needed to develop the full sensation of pitch depends primarily on the frequency of the stimulus and to a smaller degree on its intensity (Doughty and Garner, 1947). In general, for frequencies below 1000 Hz, a tone needs about 6 to 10 periods (cycles) to develop a sense of tonality, the so-called click-pitch sensation. For frequencies above 1000 Hz, the minimal duration needed to develop a click-pitch sensation is about 10 ms (Gullick, Gescheider and Frisona, 1989). The strength of pitch of short tonal and harmonic stimuli increases gradually up to about 100 to 250 ms (Bürck, Kotowski and Lichte, 1936; Moore, 1973; Turnbull, 1944). For unresolved complex tones, i.e., the tones consisting of only high order harmonics, pitch perception depends primarily on the repetition rate of the sound envelope and sound duration (White and Plack, 2003).

Phase and Polarity

The perception of phase and polarity has been a long-debated topic in audition. In general, numerous studies have shown that people are sensitive to neither absolute nor relative phase difference between various components of the periodic stimulus if the components are separated in their frequencies by more than one critical band. Hartman (1997) observed that phase difference between two pure tones separated by more than one critical band is irrelevant to audition because there is no single auditory neuron that responds to both tones. For example, changes in the phase relationship between the fundamental frequency and its lower resolved harmonics (separate by more than one critical band) have no audible effect, despite the fact that these changes greatly affect the temporal properties of the signal waveform. The fact that people are in general insensitive to the phase of the signal supports the general concept that the auditory system is a power (energy) detector rather than a pressure detector (Howes, 1971).

However, if two frequency components, e.g., harmonics, fall into the same critical band and their difference in frequency is rather small, the changes in their phase relationship are audible and affect both pitch value and pitch clarity (Lundeen and Small, 1984; Moore, 1977; Moore and Peters, 1992), There are also reports that for tone-on-tone modulation the AM is easier to detect than the frequency modulation (FM), if in both cases the carrier frequency and modulation frequency differ by less than a half of the critical band and the FM modulation index is less than 1 (Dau, 1997; Zwicker, 1952; Zwicker, Flottorp and Stevens, 1957; Schorer, 1986). In such cases, the spectra of modulated signals differ only by one sideband shifted in phase by 180°; this is the case of very low modulation rates. For higher modulation rates, where the frequency components are separated by more than one critical band, the detectability is the same. This view about the importance of the critical band for detecting phase differences is challenged by the results of some other studies, where the researchers demonstrated that the listeners were able to hear phase changes even if the frequency components were separated by more than one critical band (Lamore, 1975; Patterson, 1987).

Several authors report that humans cannot detect short term phase reversal of the stimulus (Warren and Wrightson, 1981; Sakaguchi, Arai and Murahara, 2000) or that their threshold of hearing is different for rarefaction or condensation clicks (Stapells, Picton and Smith, 1982). However, there are also reports that short

clicks may be heard differently depending on their polarity. In addition, there are reports indicating differences in auditory brainstem responses (ABRs) to sequences of condensation and rarefaction clicks (Berlin et al., 1998).

Timbre

Auditory image and timbre

Physical sounds stimulating the auditory system generate *auditory images* in our perceptual space (Letowski and Makowski, 1977; Letowski, 1989). McAdams (1984) defined an auditory image as a "psychological representation of a sound exhibiting an internal coherence in its acoustic behavior." A single auditory image can be analyzed perceptually by a listener focusing attention on the individual sensations or details of the image.

Auditory images are commonly described in terms of loudness, pitch, perceived duration, spatial character (spaciousness), and timbre (Letowski, 1992). The first three dimensions are perceptual reflections of basic physical properties of simple sounds, i.e., sound intensity, frequency, and duration, and have been discussed above. Timbre and spaciousness are multidimensional characteristics carrying information about the sound source and its acoustic environment, respectively.

Timbre has been defined by the ANSI as that attribute of an auditory image "in terms of which a listener can judge that two sounds, similarly presented and having the same loudness and pitch are dissimilar" (ANSI, 1994; Moore, 1997). A footnote to the definition explains that the term 'similarly presented' refers foremost to sound duration and spatial presentation. In a similar definition listed by Plomp (1970) loudness and pitch are supplemented by perceived duration.

In other words, timbre is the characteristic other than loudness, pitch, and perceived duration that makes two sounds perceptually different. Unfortunately, such a definition of timbre is not very useful in practical applications since it tells what timbre is *not,* rather than what timbre *is*. It also makes it unclear whether or not loudness, pitch, and perceived durations are the dimensions of timbre (Letowski, 1989). Therefore, in addition to the standardized, theoretical definition of timbre, many authors introduce another working definition of timbre. According to this definition, timbre is the perceptual property of sound that reflects unique properties of the sound and its sound source. This definition of timbre focuses on the perceptual representation of a specific pattern of temporal and spectral characteristics of a sound resulting from the specific operational principles of the sound source and facilitates identification and storage of auditory images. For example, Roederer (1974) defined timbre as "the mechanism by means of which information is extracted from the auditory signal in such a way as to make it suitable for: (1) storage in the memory with an adequate label of identification and (2) comparison with previously stored and identified information." The basic sensations of loudness, pitch, and auditory duration usually do not convey information about sound source behavior, so the above definition does not seem to contradict the formal definition of timbre. In addition, timbre defined in this way is less restrictive and allows for the differences in loudness, pitch, and auditory duration to be taken into account by the listener in assessing timbre, if needed. In other words, the sounds may differ in any or in all three of those characteristics and still have distinct differences in timbre. It also clarifies the requirement of equal loudness, pitch, and perceived duration in the standardized definition of timbre as an attempt "to bring other dimensions into focus" (Gabrielsson and Sjögren, 1979).

Timbre and pattern recognition

Two physical factors are commonly mentioned as physical correlates of timbre: the spectral envelope and the temporal envelope of the sound. Two complex tones creating the same sensations of pitch and loudness can differ greatly in their spectral content and temporal envelope. An example of such a difference is shown in Figure 11-22, which compares spectral properties of the sounds of guitar, bassoon, and alto saxophone having the same loudness and pitch. The sound of the guitar has a very dense pattern of harmonics, even in the upper range of

frequencies, while the sound of the alto saxophone has very little energy above about 5 kHz. It is this spectral pattern that helps one to hear and recognize differences among musical instruments and talkers.

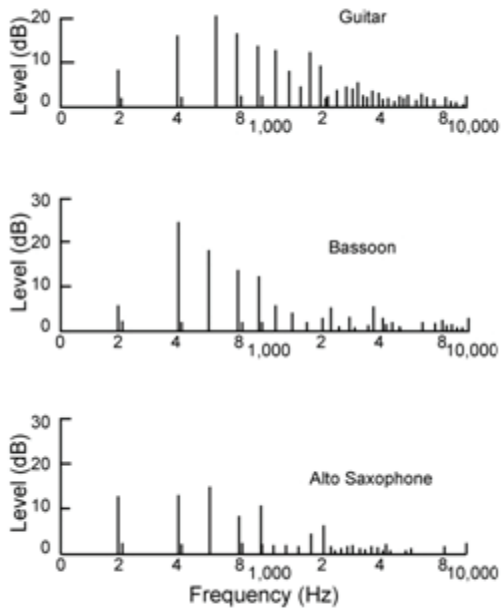

Figure 11-22. Line spectra of three instruments playing a tone with the same pitch and loudness (adapted from Olson, 1967).

It is perceptually easy to differentiate continuous sounds that differ in their spectral pattern. However, it is the temporal envelope of sound that is the main property of sound leading to sound source identification. For example, there are reports indicating that it takes at least 60 ms to recognize the timbre of continuous sound after its onset. Thus, although the differences between stationary sounds can be heard and may lead to general sound source classification (recognition), they are usually not sufficient for sound source identification. To account for this deficiency, stationary timbre is frequently referred to as sound color (noises) or tone color (periodic sounds).

Intensity changes occurring over time form the temporal envelope of a sound. In general, the temporal envelope of an isolated sound includes three distinct portions – the onset (rise), steady state (sustain), and offset (decay). In Figure 11-23, panels (a) and (b), two sound waveforms of a violin tone resulting from the vibration of a violin string actuated by (a) plucking and (b) bowing are shown. Note that the onset is quite abrupt for the plucked tone, and gradual for the bowed tone. Further, the offset begins quite early for the plucked tone; there is very little steady state content. Speech also can be described as a series of spectral changes that create the different phonemes. In Figure 11-23, panels (c) and (d), the waveforms of two different consonant-vowel syllables /*ba*/ and /*wa*/ are shown; they differ only in their initial consonant. The consonant /*b*/ is a voiced stop created by the closing of the vocal tract that produces an abrupt onset similar to that of the plucked violin. The consonant /*w*/ is an approximant, a sound produced by an incomplete constriction of the vocal tract. Although both syllables have nearly the same pitch due to a common fundamental frequency, the other peaks in the spectral content (formants) shift as the utterance shifts from the initial consonant to the vowel, and this causes a timbre difference between the two syllables.

The two examples of different spectral and temporal patterns that result in timbre differences are an indication that timbre is an important perceptual cue in pattern recognition and the dominant cue in differentiating between various music instruments. They also can be used to explain the importance of timbre judgments for sounds that

differ in loudness or pitch. The sounds of the same music instrument played in different registers of the instrument may have different timbre, but since these differences in timbre are expected for sounds of different pitch played on this instrument, they are recognized as coming from the same sound source.

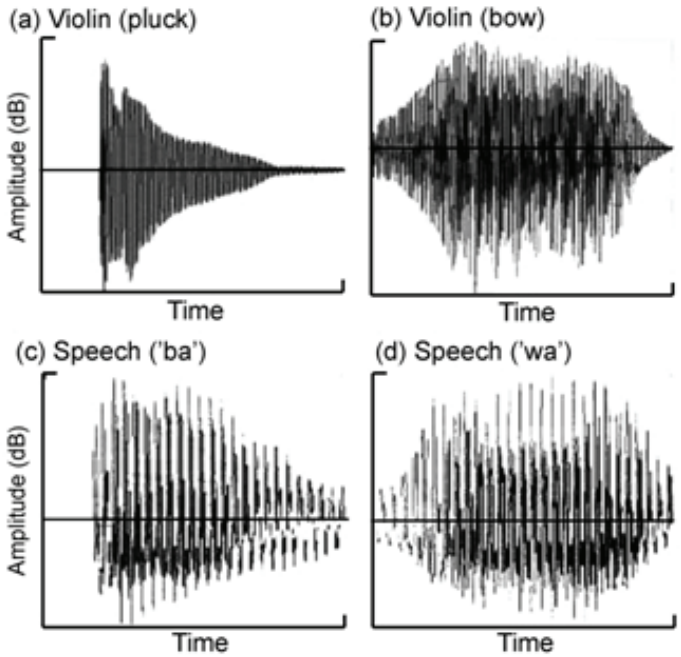

Figure 11-23. Two examples of a violin tone produced by plucking (a) or bowing (b) and of speech (c-d) (see text for details).

Timbre dimensions

Timbre is a sensation of the sound pattern (structure), which together with spaciousness reflecting the environment surroundings of the sound source and the listener, forms an auditory image of the acoustic reality outside of the listener. However, as a complex multidimensional sensation, it cannot be described well by a global assessment alone.

It is very important to realize that the multidimensional character of timbre is not a combination of a small number of well-defined subservient sensations but rather a rich language that consists of a myriad of terms with overlapping or even redundant meaning and that has a number of dialects used by engineers, musicians, psychologists, journalists, and other professional groups. Therefore, due to the richness of timbre terminology, it is necessary to identify and define some basic dimensions of timbre in order to establish universally accepted although limited timbre terminology needed for scientific and human communication purposes.

Many theoretical and experimental studies have been devoted to the identification of the dominant sensations that constitute timbre. The majority of studies used either factor analysis (FA) techniques applied to ratings made on the numerous semantic differential scales or multidimensional scaling (MDS) techniques applied to similarity judgments (Letowski, 1995). The common goal of these studies was to establish a set of meaningful descriptive adjective-based scales permitting quantitative description of timbre changes. For example, Stevens and Davis (1938) and Lichte (1941) investigated timbre dimensionality by using a semantic differential method and identified the following sensations as the main dimension of timbre: loudness, pitch (pitch height), volume,

density, brightness (spectral balance), vocality (vowel-likeness), and tonality (strength of pitch). The spatial character (spaciousness) of sound usually was not addressed in the semantic differential and similar studies, with the exception of studies dealing with sound reproduction systems and stereophonic music recording (Eisler, 1966; Gabrielsson and Sjögren, 1979) or the sound character of concert halls (Hawkes and Douglas, 1971). Therefore, although the majority of the proposed systems are limited to the timbre dimensions, there are some systems that are applicable to the overall sound image. Another complicating factor is that in many of these systems sound character (timbre) and sound quality (pleasantness) criteria were mixed together resulting in poorly designed systems. Some examples of the semi-orthogonal linear systems of bi-polar timbre or auditory image dimensions proposed by various authors are listed in Tables 11-9 to 11-12 (Letowski, 1995). The tables list the proposed dimensions and the adjectives defining both ends of the bi-polar scales.

None of the systems listed in Tables 11-9 to 11-12 seem to fully capture the dominant aspects of either timbre or auditory image, but they are listed here as examples of systems available in the literature.

One attempt to identify timbre dimensions involved the division of the spectral range into 18 one-third octave bands, assessing loudness of each of these bands, and defining timbre as a perceptual spectrum of a sound. Another attempt involved creating several perceptual dimensions based on combinations of on-third octave bands and applying them to a specific class of sounds, e.g., vowel sounds (Plomp, Pols and van der Geer, 1967; Pols, van der Kamp and Plomp, 1969; Plomp, 1970, 1976). Such approaches led to several advances in signal processing techniques, but they did not enhance our knowledge of timbre dimensions.

Table 11-9.
The system of timbre dimensions proposed by Bismarck (1974a, 1974b) for the assessment of complex tones.

Dull	Sharpness	Sharp
Compact	Density	Scattered
Empty	Fullness	Full
Colorless	Coloration	Colorful

Table 11-10.
The system of timbre (sound quality) criteria proposed by Yamashita et al. (1990) for the assessment of automotive noises.

Pleasant	Annoyance	Annoying
Weak	Powerfulness	Powerful
Dull	Sharpness	Sharp

Table 11-11.
The system of timbre dimensions developed by Ptratt and Doak (1976).

Dull	Sharpness	Brilliant
Cold	Warmth	Warm
Pure	Richness	Rich

Table 11-12.
The system of auditory image dimensions proposed by Gabrielsson and Sjögren (1979) and Gabrielsson and Lindström (1985) for the assessment of audio systems.

Dull	Sharpness	Sharp
Unclear	Clarity	Clear
Distant	Nearness	Near
Closed	Spaciousness	Open
Dull	Brightness	Bright
Soft	Loudness	Loud
Thin	Fullness	Full
Absent	Disturbance	Present

There were also several attempts, e.g., Solomon (1958), to divide the entire frequency range into a number of bands and assign timbre dimensions to sounds characterized by the dominant energy in each individual band. An example of such a system based on octave bands proposed by Letowski and Miskiewicz (1995) is shown in Table 11-13.

In addition to one-level semi-orthogonal systems of timbre dimensions, there were some attempts to create hierarchical systems in which auditory image or timbre was gradually divided into more and more detailed descriptors forming separate layers of auditory image dimensions (Clark, 1987; Steinke, 1958; Szlifirski and Letowski, 1981). An example of this type of system for two-dimensional auditory images, called MURAL, proposed by Letowski (1989), is shown in Figure 11-24.

Table 11-13.
A system of timbre dimensions for description of stationary sounds (Letowski and Miskiewicz, 1995).

Center frequency of the octave band (Hz)	Timbre Dimension
63	Boom
125	Rumble
250	Powerfulness
500	Hollowness
1000	Nasality
2000	Presence
4000	Sharpness
8000	Brilliance
16000	Rustle

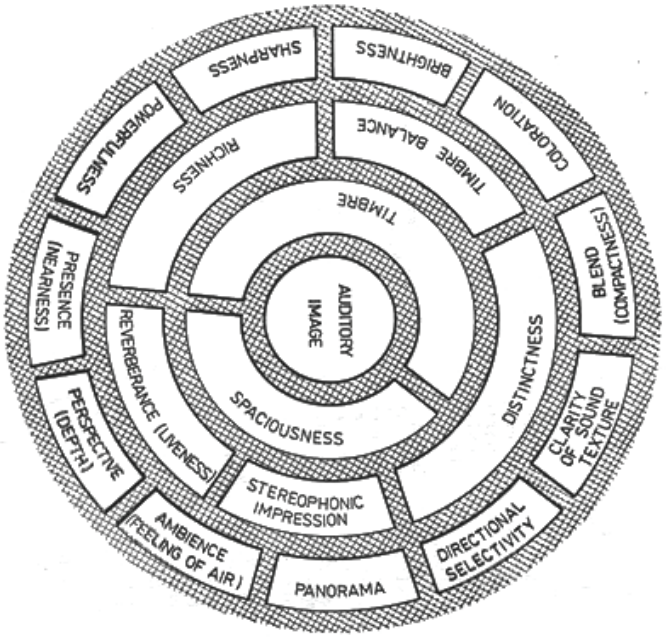

Figure 11-24. **MU**ltilevel audito**R**y **A**ssessment **L**anguage (MURAL) for timbre and sound quality assessment (Letowski, 1989).

Sound Quality

It should be recognized that the effects of auditory stimulation involve not only quantitative judgment of sensations and the subsequent perception of the acting stimulus, but also the emotional judgment of the stimulus' aesthetic value (beauty) and the assessment of the degree of the listener's satisfaction (utility). These types of judgments are together called the sound quality judgments.

Sound quality can be broadly defined as a set of properties of a given sound that determines the capability of the sound or its source to fulfill a particular function or need. As defined above, the sound quality may be either objective (technical) or subjective (perceptual). If the sound quality fulfills a perceptual need, it is sometimes called *perceived sound quality* (PSQ) to clearly identify its origin.

Letowski (1989) described PSQ as the emotional aspect of the overall auditory image. One auditory image is not better than another, they are just different. However, one auditory image may fit better a particular need than another or be closer to the desired standard than another. This underlines the basic difference between the *sound character*, expressed in terms of auditory image, timbre, spaciousness, and a multitude of other auditory sensations (e.g., roughness, breathiness, or ambience) and the *sound quality*. Sound character is expressed on scales from *more* to *less*, while sound quality is expressed on scales from *better* to *worse*.

There are two fundamental forms of sound quality assessment: global assessment and parametric assessment. Global assessment of sound quality can be made according to one of the three basic aspects of quality:

- *Fidelity* (accuracy), which reflects similarity of a given auditory image to a specific auditory standard or to another auditory image,
- *Naturalness*, which reflects an agreement of a given auditory image with general expectations of the listener, and
- *Pleasantness*, which reflects the degree of the listener's satisfaction with a given auditory image.

All these aspects of sound quality also may be expressed in terms of the opposite end of the respective perceptual scale, i.e., inaccuracy (instead of fidelity), awkwardness (instead of naturalness), and annoyance or unpleasantness (instead of pleasantness). Focusing on the positive or negative end of the scale allows the ability to differentiate better small differences between stimuli occupying a particular end of the scale, but it does not change the general order (ranking) of the stimuli quality (Letowski and Dreisbach, 1992).

Note that while an auditory image or timbre cannot be assessed globally on a more-less sound character scale, sound quality, as expressed above, can be assessed globally on a better-worse quality scale. Whether the assessment has a form of fidelity, naturalness, or pleasantness depends entirely on the application of such judgment. Audio HMD assessment may be performed with either of these criteria, however, in most practical cases, it will be done in the form of a fidelity assessment of transmitted speech (speech intelligibility), spatial auditory image (localization accuracy), or sound source identification (signature identification).

While the value of the global assessment of sound quality should not be underestimated, such assessment does not provide information about specific aspects of the auditory stimulus. Recall the multidimensional character of the auditory image and timbre discussed above. The multidimensional character of sound requires multidimensional (parametric) assessment of its sound quality in such processes as audio equipment design or selection. In order to conduct parametric assessment of sound quality, one of the timbre or auditory image dimensions subsystems discussed in the section on timbre dimensions, or any other arbitrary or experimental selection of auditory dimensions, can be used. The data collection process may have two forms: (1) classical assessment of timbre and spaciousness of a number of subservient more-less sound character scales or (2) on the same system of scales converted to better-worse sound quality scales. In the first case, the users (listeners) make classical psychophysical judgments, and the designers or researchers interpret the data as subject matter experts (SMEs) and make the decision whether more or less is good or bad. In the second case, the users themselves make these decisions.

An example of a dedicated system of objective criteria for parametric sound quality assessment is the system of metrics proposed by Zwicker and his coworkers. This system consists of one global assessment scale (sound pleasantness or annoyance) and five subordinate dimension scales: loudness, sharpness, fluctuation strength, roughness, and tonality (Aures, 1984, 1985; Bismarck, 1974b; Terhardt, Stoll and Seewann, 1982; Zwicker and Fastl, 1999). Although all the above criteria have the same names as the perceptual dimensions of the auditory image they are only certain approximations of the perceptual dimensions and are frequently referred to as, for example, calculated roughness rather than roughness to stress their objective character. Calculated loudness, sharpness, fluctuation strength, roughness, and tonality are briefly described in Table 11-14.

Various implementations of listed above calculated sound quality metrics are available in many major sound analysis software packages (e.g., PULSE by Bruel and Kjaer, dBFA32 by 01dB, Artemis and SQLab II by HEAD Acoustics, DATS by Prosig). Other objective metrics of sound quality proposed in literature include, among others, booming (sound level in 22.4 to 224 Hz band) and impulsiveness (crest factor). It may also be helpful to indicate that there are two basic methods to calculate the tonality (pitch strength) of sound used in sound analysis systems. The first method uses the concept of *tone-to-noise ratio*, defined as the ratio of the power of the tone of interest to the power of the critical band centered on that tone (excluding the tone power) (ANSI, 2005). Usually the tone is audible at tone-to-noise ratios above approximately -4 dB. Recall that noise within the critical band is masking the tone, so this is more a measure of the effective SNR than a measure of tonality. The second method uses the concept of *prominence ratio*, defined as the ratio of the power in the critical band centered on the tone of interest to the mean power of the two adjacent critical bands (ANSI, 2005). According to this metric, a tone is prominent if this ratio is above 7 dB. Neither of these metrics says much about whether the sound is perceived as a coherent tone, but rather whether a noisy

Table 11-14.
System of objective sound quality metrics developed by Zwicker and his coworkers (Zwicker and Fastl, 1999)

Dimension	Definition	Comments
Loudness	Perceptual impression of the intensity of sound.	See section on loudness. The unit of loudness is *sone*
Sharpness	Sensation caused by acoustic energy concentrated in a narrow band around relatively high center frequency of sound; perceptual metric related to the spectral center of gravity.	The unit of sharpness is *acum* (Latin for sharp). One acum is defined as the sharpness of a 1 kHz narrowband (one critical band wide) noise at 60 dB SPL.
Roughness	Perceptual impression created by amplitude and frequency modulations in sound at high modulation rates, above about 20 Hz. Roughness notably decreases for modulation frequencies higher than about 50 Hz (Terhardt, 1974).	The unit of roughness is *asper* (Latin for rough). One asper is defined as the roughness of 1 kHz tone at 60 dB SPL that is 100% modulated at 70 Hz. (Aures, 1985)
Fluctuation strength	Perceptual impression created by amplitude and frequency modulations in sound at low modulation rates, up to about 20 Hz. The greatest amount of fluctuation strength is perceived at modulation frequency of 4 Hz.	The unit of modulation strength is *vacil (*Latin for vacillate). One vacil is defined as the fluctuation strength of a 60 dB SPL, 1 kHz tone 100% amplitude modulated at 4 Hz.
Tonality	Degree to which a sound has a distinct pitch; strength of pitch.	See section on pitch. The unit of tonality is *tu* (tonality unit). One tu is defined as tonality of 1 kHz tone at 60 dB SPL
Annoyance	Combination of sharpness, fluctuation strength, roughness, and loudness.	Global assessment of sound quality The unit of calculated (unbiased) annoyance is *au*.
Pleasantness	Combination of roughness, sharpness, tonality and loudness.	Global assessment of sound quality.

Perceived Duration

Acoustic events may appear perceptually shorter or longer than the actual physical events. This phenomenon is generally described as *time distortion* or time warping, and the amount of time assigned by a person to a specific physical event is called *perceived duration*. In the case of long lasting events, exceeding several seconds, perceived duration is primarily dependent on emotional state, expectations, and activity of a person and is very difficult to generalize. The only rule that can be generally applied is that pleasant events appear to last shorter (time contraction) and unpleasant events longer (time dilation) than their actual physical durations.

In the case of very short acoustic events, humans have a general tendency to overestimate sound duration, and the amount of overestimation seems to be inversely proportional to the actual duration of the sound. When the sound duration exceeds about 200 to 300 ms and is less than several seconds, perceptual duration is very close to physical duration and generally assumed to be identical (Zwicker and Fastl, 1999).

One important condition where perceived duration differs greatly from the physical duration is perception of short silent intervals. Again, if the intervals are longer than several hundred milliseconds (500 to 1000 ms) but shorter than a few seconds the perceived duration and physical duration are about the same. Similarly, for shorter pauses, the pause duration is overestimated. However, short pauses seem to last as much as 2 to 4 times longer than short sound bursts of the same duration (Zwicker and Fastl, 1999). The higher the frequency of the sound, the greater this perceptual difference. This perceptual difference has direct impact on music perception (rhythm perception) as well as the design of periodic, fast-rate changing, signals for industrial and military applications.

Time Error

The duration of the gap between two stimuli is not only an object of detection itself, but it also moderates the effect that separated stimuli may have on each other. When the stimuli are separated by a very short period of time, they are subjected to the effects of temporal masking. If they are far apart, their comparison is affected by the decaying memory trace of the first stimulus. These phenomena are not unique to audition but occur throughout human perception.

The error in sensory judgment resulting from sequential presentation of stimuli is referred to as the *time error* (TE) or sequential error. The TE was originally observed and described by Fechner (1860) and has been studied extensively for more than a century. The type and size of TE depends on the duration of the gap between the stimuli, duration of stimuli, and the property being judged (Hellström, 1977). In the case of short time gaps when the TE is primarily a result of a forward masking, the TE is positive (+TE). In the case of long time gaps, when the time error is due to decaying memory trace of the first stimulus, the TE is negative (-TE). The duration of the time interval between the stimuli when +TE changes into –TE has been of great interest to psychologists because such stimulus separation seems to eliminate the need for consideration of TE in comparative studies.

Köhler (1923) investigated the effect of temporal gap on comparative judgment of loudness and observed -TE for gap of 1.5 seconds and +TE for gaps of 6 and 12 seconds. Based on these observations, he concluded that the optimum time gap for pair comparison of loudness should be about 3.0 seconds. The results of later studies by Needham (1935) and Pollack (1954) shortened this time to about 1.0 to 1.5 seconds.

According to Stevens (1956, 1957), the TE in pitch comparison should be very small or not present due to the relative (metathetic; associated with a quality) character of pitch as opposed to loudness that has an absolute (prothetic; associated with the quantity) character. Small TE values for pitch also were reported by Köenig (1957) who observed that the optimum gap duration for pitch comparisons were the same as for loudness comparison. Other studies concluded that as long as the temporal gap is within 0.3 and 6.0 seconds, the effect of TE on pitch perception seems negligible (Jaroszewski and Rakowski, 1976; Koester, 1945; Massaro, 1975; Postman, 1946; Truman and Wever, 1928). A similar conclusion was reached for the comparative judgment of auditory brightness, a timbre dimension very close in its character to pitch, by Letowski and Smurzyński (1980). Note that these gap durations are about the same as the silent intervals, which durations are perceived without substantial time distortions.

Unlike the rather wide range of time gaps that can be used for pitch and brightness comparisons, successive presentation of complex sounds for sound quality assessment may require gaps similar to those used for loudness comparisons (Letowski, 1974). Qualitative assessment of complex, usually time-varying sounds, seems to require shorter temporal gaps, as the listener tends be biased toward the "preferential" treatment toward the second stimulus (Choo, 1954; Brighouse and Koh, 1950; Koh, 1962, 1967; Saunders, 1962). In addition, regardless of whether the judgment is quantitative or qualitative, longer temporal gaps between signals lead to larger variability in listener judgments (Bindra, Williams and Wise, 1965; Shanefield, 1980).

Speech Perception

Speech is a system of sounds produced by the vocal system that allow human-to-human communication. Simple sounds, called phonemes, are combined together in more complex structures (strings) to convey thoughts, feelings, needs, and perceptions. Small structures are combined in larger and larger structures, called syllables, words, phrases, sentences, and stories, respectively, depending on the complexity of the intended message. Each spoken language has a certain limited number of phonemes that form the basis of speech communication and has a practically infinite number of higher order structures that can be constructed with these phonemes.

Liberman et al. (1967) stated that the perception of speech is different from the perception of all other complex sounds because it is mentally tied up to the process of speech production. However, if the speech sounds are unfamiliar to the listener (e.g., listening to an unknown foreign language), the speech looses its special character caused by coupling between speech perception and speech production of the listener; such sounds should be treated as non speech sounds.

Speech production

Speech sounds can be spoken or sung, especially the voiced phonemes. Depending on the range of frequencies that the singer can produce, singer voices are typically classified as soprano, mezzo-soprano, and alto (female voices) and tenor, baritone, and bass (male voices), starting from the highest through the lowest. In addition to spoken and sung speech sounds, human vocal production includes whistling, crying, murmuring, tongue clicking, grunting, purring, kissing sounds and laughing.

The process of speech production is called articulation and involves the lungs, larynx and vocal folds, and vocal tract. The vocal tract is the air tube that begins at the mouth's opening and ends at the vocal folds with branches off to the nasal cavity. In the process of speech production the stream of air controlled by the lungs and vocal folds is processed by the set of three articulators located in the mouth cavity – tongue, teeth, and lips – and becomes a string of speech sounds, i.e., phonemes. The process of combining phonemes into larger structures, i.e., the process of chaining the phonemes together into strings, is called *coarticulation*.

Two basic classes of phonemes are vowels and consonants, which can be divided further in many subclasses depending on the form and degree of activation of the vocal folds and mouth articulators. The vowels are usually classified based on the tongue position and lips openness. The consonants are classified on the basis of their voicing (voiced and unvoiced), place and manner of their production. All vowels and voiced consonants are the results – but not solely – of the acoustic filtering by the vocal tract of the saw tooth-like periodic waveform generated by vocal folds in a process of phonation. The momentary positions of speech articulators during the process of phonation divide the vocal tract into a series of resonance tubes and cavities that produce local concentrations of energy in the spectrum of output signal. These concentrations are called formants, and their relative positions on the frequency scale identify individual vowels. Vowels are very important to speech production, but it is the consonants, i.e., the very movement of articulators, which make the speech rich in meanings and contexts.

In addition to the factors discussed above, the emotional and ecological conditions during speech production lead to various forms and levels of speech: a soft whisper (30 to 40 dB SPL), a voiced whisper (40 to 55 dB SPL), conversational speech (55 to 65 dB SPL), raised speech (65 to 75 dB SPL), loud speech (77 to 85 dB SPL), and shouting (85 to 95 dB SPL). These values correspond to the sound pressure levels at about 1 meter (3.28 feet) from the talker's lips. Directly at the lips, these values are much higher. A list of selected basic factors affecting speech production is presented in Table 11-15.

The Lombard effect (Lombard, 1911) is a phenomenon in which a talker alters his or her voice in noisy environments. Generally, there is an increase in vowel duration and voice intensity (Summers et al, 1988; Junqua, 1996). In addition, Letowski, Frank and Caravella (1993) reported changes in the fundamental frequency of the voice (male voices) and spectral envelope of the long term spectrum (female voices). These changes to the speech

produced in noise are most likely caused by the talker's attempt to improve audibility of the sidetone (i.e., audibility of the talker's own voice) and result in improved speech intelligibility (Lane and Tranel, 1971; Letowski, Frank and Caravella, 1993). There observations are in agreement with the reports that signal processing techniques that replicate the Lombard effect improve the intelligibility of speech in a noise environment (Chi and Oh, 1996). However, the human tendency to alter speech in this way is largely automatic (Pick et al., 1989), and individuals has no control over the Lombard effect. The existence of the Lombard effect also affects the accuracy of speech recognition software (Junqua, 1993). Because of this, the presence of the Lombard effect is worth considering when designing audio HMDs that will be used in conjunction with speech recognition software in noisy environments.

Table 11-15.
Basic factors affecting talker's speech production.

Factors Affecting Speech Production
Fundamental frequency of the voice
Language (primary vs. secondary)
Articulation and coarticulation
Breathing (emotions)
Vocal effort (whisper to shout)
Auditory feedback (sidetone)
Ambient noise (Lombard effect)
Hearing loss of the talker

Speech communication

Speech communication refers to the processes associated with the production and perception of sounds used in spoken language. Humans are able to understand speech produced by an infinite variety of voices in an infinite variety of combinations. However, individuals differ in their hearing ability and language proficiency, environments are noisy or unpredictable, and equipment supporting speech communication may be noisy or problematic.

The highest level of speech understanding is referred to as speech comprehension. Speech comprehension is a function of environmental conditions, the communication channel and its capacity, and peripheral hearing ability and higher order cognitive factors of the listener. Speech comprehension can only be approximated, and the process is time consuming and tedious. A few such tests have been developed, but are not commonly used.

Speech recognition (SR) is a lower level of speech understanding. SR is the human ability to understand speech. It is measured by the percent of correctly recognized speech items (phonemes, syllables, words, phrases, or sentences). The result can be expressed as percent correct responses for the whole speech test (speech recognition score), as a speech intensity level for which a person is able to recognize 50% of the speech items (speech recognition threshold), or as a speech level at which an individual is able to recognize 50% of the test items as speech (speech detection level).

The two lowest levels of speech understanding are speech discrimination and speech detection. Speech discrimination tests are used very rarely, and they are intended to measure the degree to which a person is able to hear the differences between speech items, even if they are meaningless. One practical application of speech discriminations tests is prediction of potential problems in acquiring a second language. Various pairs of

phonemes or syllables in a new language are played to a person before the language training begins to determine which sounds would be the most difficult for this person to differentiate and, because of it, clearly produce.

The *speech detection threshold* (SDT), frequently referred to as the *speech awareness threshold* (SAT), has been introduced above in the section about the air conduction threshold. This metric is used mainly to determine minimum required levels of masking stimulus to mask speech, e.g., in open office situation. SDTs in quiet and in noise are used also for testing human ability to hear speech for those who does not speak a given language and in lieu of more time consuming tonal audiometric tests to roughly assess hearing threshold of a person in a given environment.

Various speech communication terms used in speech communication testing are shown in Figure 11-25. Speech recognition is a measure of a person's ability to hear speech. Speech articulation is a measure of the clarity of speech production. Speech transmission is the measure of the effect of a communication channel on the clarity of speech. These three basic elements of speech communication assessment may be combined in different configurations resulting in speech intelligibility encompassing speech articulation and transmission or speech audibility encompassing speech transmission and recognition.

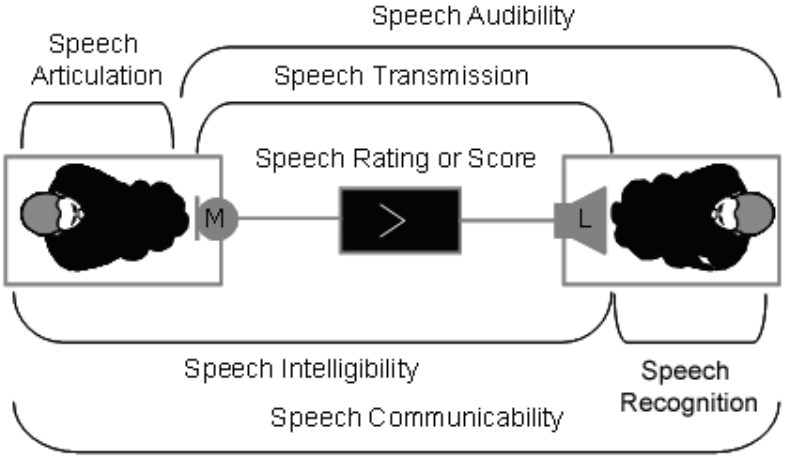

Figure 11-25. Speech communication terminology used in the assessment of the effects of various elements of the speech transmission chain on speech communication.

For example, *speech intelligibility* (SI) is the understanding of speech in a particular environment by expert listeners with normal hearing. SI testing is used to quantify the operational conditions of a speech communication environment in order to determine whether there are problems that threaten the transmission of spoken information. Speech intelligibility is affected by imperfect speech production and by properties of the communication channel between the talker and the listener, including environmental conditions surrounding both the talker and the listener. It varies as a function of SNR, reverberation time, rate of speech and other factors. It is measured usually as a word recognition score for a given transmission system or environment but it can be applied to sentences and connected speech as well.

In many cases, neither the talker's characteristics, environmental conditions, nor the listener's characteristics are ideal, and it is necessary to capture human ability to communicate under these conditions. Such speech tests are referred to in this chapter as speech communicability tests (Figure 11-25). Note, however, that regardless of the specific part of the communication chain being assessed, the same physical speech tests may be used for data collection. There is a very large selection of speech tests that differ in their redundancy, complexity, and vocabulary and result in fairly different test scores for the same auditory environment. Therefore, it is important that speech communication data are reported together with the name of the speech test used for the data collection

and the name of speech communication measure to document what was actually measured and how. Unfortunately, there is a general lack of terminological discipline among the people developing speech tests and conducting speech assessments, and the described terms are frequently misused.

There are a number of perceptual tests of speech recognition, intelligibility, or communicability including perceptual tasks of recognition, discrimination, and detection of speech. In addition, speech intelligibility (clarity) can be also rated on the scale from 0% to 100%. This test procedure is called a *speech intelligibility rating* (SIR) test and may be applied not only to intelligibility testing but also to the other forms of speech assessment shown in Figure 11-25. It is a fast data collection procedure that provides data highly correlated with measures requiring much more effort and time consuming scoring procedures.

Speech perception and environment

The primary environmental effects on speech communication are those of noise and reverberation. Good understanding of speech requires high SNRs in the order of 15 to 30 dB. Smaller SNRs lead to reduced speech intelligibility scores. The exact SNR level required to achieve minimal required speech intelligibility depends on the speech material, type of noise, the acoustic environment, and the listeners themselves.

As long as the SNR is sufficiently high to allow enough of the speech signal to be heard in the noise, the absolute level of the noise has a minimal effect on speech understanding as long as the noise levels are below 85 dB SPL. Conversely, speech understanding depends to a large degree on the type of background noise. As discussed earlier in the section on masking, steady-state broadband noise causes primarily energetic type of masking. As long as a sufficient proportion of the speech energy is audible, speech is heard and understood. However, random, unpredictable noise, or noise where the temporal and spectral characteristics are similar to that of speech, can add informational masking to the energetic masking. Therefore, the most efficient masker of speech is other speech, such as a multitalker noise including a moderate number of voices. An example of functional relationship between speech recognition score and SNR is shown in Figure 11-26.

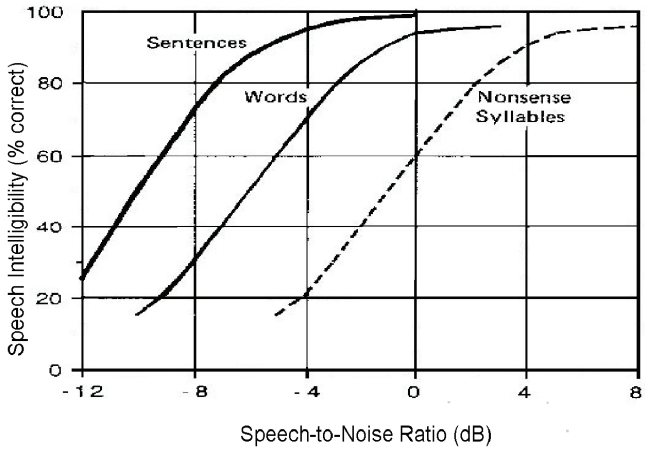

Figure 11-26. The effect of speech-to-noise ratio (SNR) on intelligibility of nonsense syllables, words, and sentences. Adapted from Levitt and Webster (1997, Figure 16.3).

An acoustic environment can also add reverberation to the speech signal. Reverberation consists of multiple reflections of the sound that mask the direct sound. In the case of speech, the masking is simultaneous and temporal. Thus, it adds noise to and alters the spectro-temporal envelope of the original speech signal. Some very large spaces have a sound decay time (i.e., reverberation times) to the order of 5 seconds and higher, especially at low frequencies, which can make normal speech communication in these spaces virtually impossible.

The spatial configuration of the speech source and the noises source(s) can also have an effect on intelligibility. If the talker and the noise are located in the same location, the listener can only use spectral and temporal characteristics of the two signals to parse the two signals. This is true also if the listener only has a monaural signal (e.g., phone and radio). If the two signals are separated in space, binaural cues can be used to separate the two locations, and the listener can selectively attend to the target speech signal.

All of these environmental characteristics affect speech intelligibility and can often be quantified and described to some degree. However, cognitive features of the speech message also affect comprehension to some degree as well. The most notable of these is known as the "cocktail party phenomenon" where the listener embedded in "party noise" can clearly hear his or her own name when spoken, even though other speech may not be audible (for a more thorough discussion of cocktail party effect, see Bronkhorst, 2000).

Speech recognition

The term *speech recognition* (SR) is often confusing because it has two related by separate meanings. In its narrow sense, it is a metric that provides information about individual's ability to hear speech as shown in Figure 11-25. In its broadest meaning it is the score on any speech test regardless of the specific type of communication assessment. For example, one can use a speech recognition test to assess speech articulation, speech recognition, speech transmission, or speech communicatibility. It is this second, broader, meaning of the *speech recognition* term that we use throughout the rest of this chapter.

The *SR score* and *speech recognition threshold* (SRT) are two basic metrics of speech recognition. They are used to characterize SR ability of an individual listener under specific test conditions but in practice they are also dependent on speech material, the talker's voice, and many procedural factors. However, as long as these test elements are kept constant, any speech tests can be used as a relative measure of human capabilities. It is the predictive value of the speech test for the specific operational conditions that makes various test more or less appropriate. It is important to recognize that all speech tests data are limited by the degree to which selected speech material is representative of the speech vocabulary and speech structures used in the operational environment for which performance is being predicted.

The SR score is simply the percentage of speech material understood by the listener. The ANSI standard S3.5 – 1997 (revised in 2007) (ANSI, 1997) gives speech recognition scores for a number of commonly accepted perceptual speech recognition measures and compares them to objective measures described later in this section

Basic test conditions and test material for SRT testing are addressed by ANSI standard S3.6, "Specification for Audiometers" (ANSI, 2004). As the speech test complexity gets lower and the background noise gets quieter, the SRT decreases. A number of other factors also affect SRT level. First, individuals differ in their hearing sensitivity. Hearing loss due to trauma and age typically occurs in the range of frequencies containing information about consonants. Second, there will be an effect on scores due to whether the speech material consists of syllables, words or sentences, as there is more disambiguating information for longer speech items. Third, the size of the speech vocabulary and the type of speech information used can affect both the SRT level and the SR score. If the operational vocabulary is relatively small and the items within the set phonemically distinct, scores will remain high and SRT levels low even under poor listening conditions. Further, if the items are presented as a closed set (e.g., multiple choice and limited options), performance will be higher than if the items are presented as an open set. Fourth, the quality of the speech presentation will have an effect on recognition performance. If speech material is presented over high-fidelity audio display equipment, scores will be higher than if it were distorted by poor quality radio-type transmissions with low bandwidth. Finally, the spatial arrangement of the speech source relative to that noise also will affect the degree of masking.

Speech perception tests

If speech transmission is to be characterized in terms of performance in a specific environment, either perceptual or objective, microphone-based predictive speech tests can be used. Perceptual measures entail the presentation of speech material at one or more fixed intensity levels to a group of listeners. Performance is given as the average percent correct recognition. Objective measures predict intelligibility by calculating an index based on a recorded sample of ambient environment. Perceptual measures are limited by the speech materials used, the talkers presenting the materials, and the listener sample involved in the study. Figure 11-27 graphs the relationship of scores obtained in a range of common perceptual tests to an objective measure of speech intelligibility called the *articulation index* (AI). The AI will be discussed below, but note that for a particular AI value, performances on the perceptual measures vary widely.

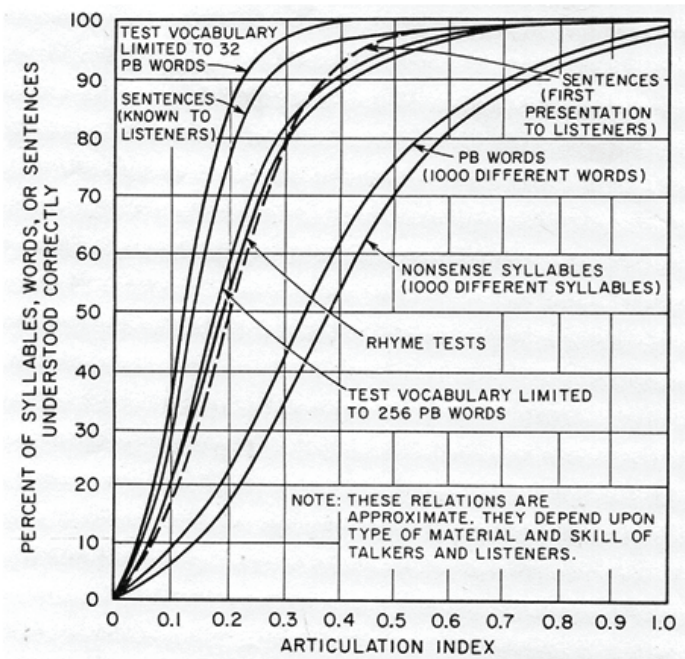

Figure 11-27. Relationship between various perceptual measures of speech intelligibility and articulation index (AI) (after ANSI S3.5 [1997]).

Speech intelligibility performance generally improves as the material becomes more complex and contains more contextual information and higher degree of redundancy (Egan, 1948; Hochhaus and Antes, 1973; Hudgins et al., 1947; Miller, Heise and Lichten, 1951; Rosenzweig and Postman, 1957). Representative tests used in architectural acoustics, communications, and audiology are included in Table 11-16. The tests are classified according to the speech material used in the test and whether the number of alternative answers to the test item was finite (closed set test) or infinite (open set text).

The tests listed in Table 11-16 differ not only by the speech units used for testing and by the open or closed set of possible answers but may also differ in the way they are administered. For example, they may differ by the presence or absence of carrier phrases (word tests), monotone or natural presentation (spondee, phrase, and sentence tests), recorded or live voice administration, and many other technical elements. Therefore, it is important that in a comparative evaluation not only that the same test is used for all audio HMD systems or listening conditions under comparison, but also that it is administered in the same manner.

Table 11-16.
A listing of speech tests using speech material of various degrees of speech complexity.

Speech Unit	Name of Test	Citation
digits and words (open set)	SDS Speech Discrimination Scale	Gaeth, 1970
Oo, ah, ee, sh, M, sss (open set)	Five Sound Test	Ling, 1978
phonemes (closed set)	Test of Phoneme Identities	Murray et al. 2000
phonemes (closed set)	Keating-Manis Phoneme Deletion Test	Keating andManis, 1988
nonsense syllables (closed set)	CUNY-NST Nonsense Syllable Test	Resnick et al., 1975
nonsense syllables (closed set)	DFD Distinctive Feature Difference Test	Feeney and Franks, 1982
words (open set)	PB-50 Phonetically Balance (PB) Word Lists	Egan, 1948
words (open set)	PBK-50 PB Kindergarten Lists	Haskins, 1949
words (open set)	W-22 CV, VC, and CVC Words Test	Hirsh, 1952
words (open set)	CNC Consonant-Nucleus -Consonant Test	Lehiste and Peterson, 1959
words (open set)	NU-6 Northwestern University Lists	Tillman et al., 1963
words (open set)	ABL Test	Boothroyd, 1967
words (open set)	HFCDT High Frequency Consonant Discrimination Test	Gardner, 1971
words (open set)	SPRINT Speech Recognition in Noise Test	Cord et al., 1992
words (closed set)	RT Rhyme Test	Fairbanks, 1958
words (closed set)	Fry Lists	Fry, 1961
words (closed set)	MRT Modified Rhyme Test	House et al., 1965
words (closed set)	DRT Diagnostic Rhyme Test	Voiers, 1975; 1983
words (closed set)	DIP Discrimination by Identification of Pictures Test	Siegenthaler and Haspiel, 1966
words (closed set)	RMC Rhyming Minimal Contrasts Test	Griffiths, 1967
words (closed set)	MCDT Multiple Choice Discrimination Test	Schultz and Schubert, 1969
words (closed set)	WIPI Word Intelligibility by Picture Identification Test	Ross and Lerman, 1970
words (closed set)	OUCRT Oklahoma University Closed Response Test	Pederson and Studebaker, 1972
words (closed set)	CCT California Consonant Test	Owens and Schubert, 1977
words (closed set)	Perception of Words and Word Patterns	Erber and Witt, 1977
words (closed set)	NU-Chips Children's Perception of Speech	Katz and Elliott, 1978
words (closed set)	DFDT Distinctive Features Discrimination Test	McPherson and Pang-Ching, 1979

Table 11-16. (Cont.)
A listing of speech tests using speech material of various degrees of speech complexity.

Speech Unit	Name of Test	Citation
words (closed set)	ANT Auditory Numbers Test	Erber, 1980
words (closed set)	Picture Identification Task	Wilson and Antablin, 1980
spondees	Psycho Acoustic Laboratory (PAL) List #9	Hudgins et al., 1947
spondees	CID W-1 and W-2 (Spondaic Word Lists)	Hirsh et al., 1952
sound effects (closed set)	SERT Sound Effects Recognition Test	Finitzo-Hieber et al., 1980
sentences (open set)	Sentence Tests for Deaf People	Fry and Kerridge, 1939
sentences (open set)	CID Central Institute for the Deaf Test	Silverman and Hirsh, 1955
sentences (open set)	Fry Revised Sentence Test	Fry, 1961
sentences (open set)	SPIN Speech Perception In Noise	Kalikow et al., 1977
sentences (open set)	R-SPIN Revised Speech Perception In Noise	Bilger, 1984
sentences (open set)	BKB (Bamford-Kowal-Bench)	Bench et al., 1979
sentences (open set)	HINT Hearing in Noise Test	Nilsson et al., 1994
sentences (closed set)	KSUDT Kent State University Discrimination Test	Berger, 1969
sentences (closed set)	SSI Synthetic Sentence Identification Test	Speaks and Jerger, 1965
sentences (closed set)	PSI Pediatric Speech Intelligibility Test	Jerger et al., 1980
sentences (closed set)	CRM Coordinate Response Measure (AFRL)	(Brungart, 2001)
sentences (closed set)	NSMRL Tri-Test of Intelligibility	Sergeant et al., 1981
sentences (closed set)	DSI Dichotic Sentence Identification Test	Fifer et al., 1983
phrases (close set) 2 word-3 syllable phrases	CAT Callsign Acquisition Test	Rao and Letowski, 2004; Blue, Ntuen and Letowski, 2004
passages of speech (9-10 sentences)	Connected Speech Test (CST)	Cox et al., 1987

Tests that differ in units of speech and sets of available responses differ also in the test difficulty. Usually, open set tests are more difficult then closed set tests, and tests using meaningless syllables or sentences are more difficult than tests using meaningful items. For example, percent correct scores on nonsense syllables may not exceed 70% correct even at very high SNRs (Miller, Heise and Lichten, 1951). In contrast, scores on words in meaningful sentences may reach 100%, even at moderate SNR, and scores on digits may reach this limit at SNRs as low as 6 dB. The context existing in a meaningful sentence provides information about what kinds of words would be probable at a given place in the sentence, effectively limiting the listener's choices. Thus, even if a particular word is partially masked, enough information is available for the listener to fill in the missing information. In the case of the digits, the listener is limited to 10 or even less available numerical digits and has a high probability of guessing correctly even if a part of the digit is masked or distorted.

As discussed previously in this section, speech understanding depends, among other things on the talker. A trained talker such as radio announcer who speaks clearly will be more intelligible than a talker using normal conversational speech. Clear speech has been found to be slower, both the phonetic components and the pauses between words are drawn out more (Picheny, Durlach and Braida, 1985, 1986, 1989; Uchanski et al., 1996). Usually, clear speech is used for test materials; however, most speech in operational environments is conversational and will not be as intelligible. If testing is done using recordings made of a trained speaker using clear speech, it will probably overestimate performance in most settings. Further, there is a large difference between the intelligibility of different talkers (Black, 1957; Hood and Poole, 1980; Bond and Moore, 1994). For example, a female voice is generally more intelligible than a male voice (Bradlow, Torretta and Pisoni, 1996). Therefore, it is important to use several talkers in validating communication effectiveness of audio HMDs. The current ANSI S3.2-1989 standard for speech intelligibility testing requires that the number of talkers is at least equal the number of listeners.

It is important to recognize that the training and hearing sensitivity of the listener also affect speech intelligibility performance. Trained listeners who are familiar with the test and the speech material to be tested will perform best and have the most reliable scores (Hood and Poole, 1980). Listeners who have impaired hearing will perform differently than normal hearing counterparts, even if the average intensity levels are above threshold (Ching, Dillon and Byrne, 1998). It needs to be stressed that hearing loss is common in those working in high noise environments. Therefore, a measure of the speech intelligibility of a particular environment obtained using normal hearing listeners and professional talkers may overestimate performance by operators in that environment.

Speech intelligibility criteria used by the U.S. Army are listed in he MIL-STD-1472F (Department of Defense, 1999) and presented as Table 11-17. The criteria list the Phonetically Balanced (PB) Word List[7] and Modified Rhyme Test (MRT)[8] perceptual test scores and calculated value of the AI. The criteria are intended for voice communication over a transmission system such as radio and intercom and apply to audio HMDs. The criteria listed in the second row of Table 11-17 are acceptable for normal operational environments. AI should only be used to evaluation of natural speech but not synthetic speech, because some key acoustic features of speech are not present in synthetic speech.

All perceptual speech intelligibility measures discussed above and the speech tests listed in Table 11-16 are influenced by a number of factors that limit applicability of their scores to the operational environments in which they were obtained. Their results may be generalized to other similar environments but not to environments that are very different from the one that was selected for testing. Further, perceptual studies are costly in terms of time

---

[7] In the Phonetically Balanced Word Lists, the monosyllabic test words are chosen so that they approximate the relative frequency of phoneme occurrence in each language (Goldstein, 1995).

[8] The modified Rhyme Test is a word list for statistical intelligibility testing that uses 50 six-word lists of rhyming or similar-sounding monosyllabic English words. Each word is constructed from a consonant-vowel-consonant sound sequence, and the six words in each list differ only in the initial or final consonant sound. Listeners are shown a six-word list and then asked to identify which of the six is spoken by the talker.

and the number of persons required when obtaining speech intelligibility performance data. Therefore, it is sometimes preferable to estimate the effect of the specific environment on speech intelligibility on the basis of physical measurements of the operational acoustic environment. Such measures do not eliminate the need for final assessment of speech intelligibility using perceptual speech intelligibility tests; however, they are fast and convenient measures for comparing various environments and for making numerous initial predictions regarding speech intelligibility.

Table 11-17.
Speech intelligibility criteria for voice communication systems recommended by MIL-STD 1472F (1999).

Communication Requirement	Score		
	PB	MRT	AI
Exceptionally high intelligibility; separate syllables understood	90%	97%	0.7
Normal acceptable intelligibility; about 98% of sentences correctly heard; single digits understood	75%	91%	0.5
Minimally acceptable intelligibility; limited standardized phrases understood; about 90% sentences correctly heard (not acceptable for operational equipment)	43%	75%	0.3

Speech intelligibility index (SII)

Since speech intelligibility is a function of the SNR and acoustic characteristics of the environment, speech intelligibility in a given environment may be estimated on the basis of some physical data collected in this environment. Such estimations cannot replace completely perceptual tests described in the above section; however, they are much faster and cheaper to conduct, and they have some predictive value.

The standard objective measure of speech intelligibility used in the U.S. is the *speech intelligibility index* (SSI) described in the ANSI S3.5-1997 (R2007) standard. There are two specific speech intelligibility indexes recommended by and described in this standard. The first is a revised version of the AI, and the second is the *speech transmission index* (STI).

AI is a measure of speech intelligibility originally proposed by French and Steinberg (1947) and Beranek (1947) and further developed by Kryter (1962a, 1962b; 1965). The AI concept is based on the relationship between the "standard" speech spectrum and the spectrum of the background noise. The noise spectrum is measured in several frequency bands across the frequency range from 160 Hz to 6300 Hz, which was determined to be critical to the understanding of speech. The AI is calculated by combining the SNRs of all bands weighted by coefficients indicating each band's relative contribution to speech intelligibility. The overall intelligibility then is expressed on a scale from 0 to 1. Several methods of dividing the speech spectrum into frequency bands are suggested in ANSI S3.5-1997 (R2007), including the twenty-band, one-third octave band, and an octave band method. Each method uses different weighting factors representing the corresponding band's overall contribution to speech intelligibility. The AI gives an index value that represents the proportion of speech information that is above the noise – not the percentage of speech items that will be recognized. The ANSI S3.5-1997 (R2007) standard provides data relating AI values to speech intelligibility scores obtained for several common perceptual tests (e.g., nonsense syllables, rhyme tests, PB words, sentences, limited vocabularies and known sentences). These relationships are shown in Figure 11-27. The AI version described in the ANSI standard also can be applied to determine the effects of hearing loss on intelligibility. The calculation procedure treats the hearing threshold in the same way as ambient noise and intelligibility is calculated given as the percentage of speech information that is above the hearing threshold.

One of the drawbacks to AI is that it does not account for temporal distortion (e.g., echoes, reverberation, and automatic gain control), and non-linear distortion (e.g., system overload, quantization noise) affecting the speech. To account for these effects Steeneken (1992) and Steeneken and Houtgast (1980, 1999) developed the *speech transmission index* (STI) based on the concept of the modulation transfer function (MTF). The authors assumed that the intelligibility of a transmitted speech is related to the preservation of the original temporal pattern of speech and created a test signal that represents speech as a noise 100% modulated with several modulation frequencies. This modulated test signal is broadcast from a loudspeaker at the talker's location. A microphone is placed at the receiving end of the communication system to capture the broadcasted signal along with effects of reverberation and background noise present in the environment. The residual depth of modulation of the received signal is compared with that of the test signal in a number of frequency bands. Reductions in the modulation depth are associated with loss of intelligibility. These reductions constitute, in part, the *effective* SNR and are calculated in seven relevant frequency bands from 125 Hz to 8 kHz. These weighted values then are combined into a single index having a value between 0 and 1. As with the AI, intelligibility performance on a number of common perceptual measures is given for a number of STI values.

Both the AI and the STI have been implemented in psychoacoustic software programs and commercial room acoustics measurement devices that can be used to measure intelligibility in any operational environment. Although STI accounts fairly well for temporal and nonlinear effects, translation of index values to percent correct scores is only approximate, as in the AI case. Further, neither AI nor STI can account for the spatial arrangement of the sound sources in the operational environment, and they estimate speech intelligibility for the "worst-case scenario," when speech and noise are arriving from the same location in space.

Both AI and STI are based on measurements taken from a single microphone. Although there have been recent efforts to account for binaural effects (Wijngaarden and Drullman, 2008), to date no official binaural version of these tests exist. Many reports have shown an advantage of binaural listening for speech recognition. Two factors seem to contribute to this advantage. First, binaural listening allows the listener to utilize the *better ear advantage*, i.e., the listener can attend to the signal with his/her better ear or with the ear where the SNR is highest and ignore information in the less favorably positioned second ear (Brungart and Simpson, 2002; Culling, Hawley and Litovsky, 2004). Second, the listener can use spatial localization cues (described later in this chapter) to separate the speech information from the noise and can attend to the spatial location of the speech (Hawley, Litovsky and Culling, 2004; Kopčo and Shinn-Cunningham, 2008).

Despite the large number of measures of speech intelligibility, both perceptual and objective, none of these yet are truly able to measure the degree to which speech communication occurs. Beyond recognizing the phonemes and syllables that make up words and sentences, speech communication requires higher order comprehension of the thoughts and ideas that are transmitted along with them. Sometimes communication occurs that is not explicitly contained in the speech. Pragmatic information contained in our schematic knowledge about the world and the meaning of certain words in combination with certain patterns of events is not easily measured by either perceptual or objective speech measures. Nor can these measures fully predict which information will be attended to or processed by the listener or how this will change as a function of the contextual environment. Therefore, the speech tests and intelligibility indexes are best used for the comparison of elements in the channels affecting communication, i.e., they can give relative information (better/worse) but not absolute information about actual speech intelligibility.

Speech quality

The concepts of timbre and sound quality are used mainly in reference to music, sound effects, and virtual auditory environments. Assessment of speech is almost entirely based on speech intelligibility. Speech quality – in its sound quality meaning – is assessed much less frequently. It is also a confusing term because it has a second connotation that is similar to speech intelligibility.

Speech quality in its first, traditional meaning refers to the pleasantness of talker's voice and naturalness of speech flow. To stress this fact, such speech quality is sometimes referred to as vocal quality or voice quality, although such usage is not consistent. A limitation of speech quality defined as above is that it difficult to assess it for speech of poor intelligibility. Similarly, changes in speech quality can only be reliably assessed if they are not accompanied by large changes in speech intelligibility. If they are, speech quality changes are buried deep below speech intelligibility changes and are hard, if even possible, to be evaluated. In such cases the parallel judgments of speech intelligibility and speech quality frequently result in highly correlated scores, especially if the changes in speech intelligibility are fairly large (McBride, Letowski and Tran, 2008; Studebaker and Sherbecoe, 1988; Tran, Letowski, and McBride, 2008). However, when the compared samples of speech have the same, or very similar, and relatively high speech intelligibility, these samples may still greatly differ in speech quality and these differences may be reliably assessed by the listeners. In general, the higher speech quality, the greater satisfaction and listening comfort of the listener. The listener may prefer even a little less intelligible speech if the benefit in speech quality is large. For example, changes of low frequency limit of the channel transmitting speech signal have small if any effect on speech intelligibility but can greatly affect speech quality. Therefore, once the speech is sufficiently intelligible, it is important to assess and maximize its speech quality. If speech quality is relatively low, it may cause listener's annoyance and affect listening comfort and long-term performance of the listener.

The second meaning of speech quality is more technical and encompasses all aspects of transmitted speech. Its function is to represent overall audio quality of transmitted speech. The fact that scores for speech intelligibility and speech quality are highly correlated for imperfect speech led to the concept that speech quality really incorporates speech intelligibility and may be used as a sole criterion for assessment of speech transmission. Such connotation of speech quality is primarily used in telephony, speech synthesis, and digital communication. It encompasses natural and digital (lost packets, properties of speech codecs) causes of degraded speech intelligibility, presence of noise in the transmission channel, transmission reflections (echoes), and channel cross-talk. It is usually assessed by the listeners' ratings on a 5-step quality scale leading to a score called the *mean opinion score* (MOS). The standardized MOS scale used in evaluation of both audio and video transmission quality is shown in Table 11-18.

Table 11-18.
Mean opinion score (MOS) scale used in assessment of audio quality of transmitted speech (ITU-T, 1996).

MOS	Quality rating	Impairment rating
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

Auditory Spatial Perception

Spatial perception is the awareness of environment, its boundaries, and internal elements. It allows an observer to determine directions, distances to and between, sizes, and shapes (spatial orientation) as well as realize utilitarian value or emotional impact of the whole environment or its specific part (space quality). If this awareness is focused on acoustic properties of the environment and their auditory image, such spatial perception is called *auditory spatial perception*. Auditory spatial orientation is mostly involved with directions, distances, and characteristic sound reflections that provide information about the size of the environment and materials of its boundaries. The utilitarian and emotional aspects of auditory spatial perception are reflected in the listener satisfaction with the perceived spaciousness of the environment, that is, in perceived spatial sound quality.

Auditory spatial orientation

Auditory spatial orientation is one of the critical abilities of living organisms. In the case of humans, the sense of balance located in the vestibular portion of the inner ear provides information about the position of the human body in reference to the force of gravity. Senses of vision and hearing, and to much lesser extent, the sense of smell, provide information regarding positions of other objects in space in relation to the position of the body. While vision is the primary human sense in providing information about the surrounding world that can be seen, hearing system is the main source of spatial orientation allowing humans to locate objects in space, even if they cannot be seen.

Anatomic structures and physiologic processes of the auditory system have been described in Chapter 8, *Basic Anatomy of the Hearing System,* and Chapter 9, *Auditory Function.* In general, human ability to perceive spatial sound and localize sound sources in space is based on the presence of two auditory sensors: the ears and the presence and elaborate shape of human pinnae.

Several acoustic cues are used by humans for auditory orientation in space. The importance of the specific cues depends on the type of surrounding environment and the specific characteristics of the sound sources present in this environment Thus, in order to understand the mechanics of the spatial auditory perception, it is necessary to outline the primary elements of the space leading to spatial orientation (Scharine and Letowski, 2005). These elements are:

- *Azimuth* – the angle at which the specific sound source is situated in the horizontal plane or the angular spread of the sound sources of interest in the horizontal plane (horizontal spread or panorama; see Figure 11-24),
- *Elevation* (zenith) – the angle at which the specific sound source is situated in the vertical plane or the angular spread of the sound sources of interest in the vertical plane (vertical spread),
- *Distance* – the separation of the listener from the specific sound source or the separation between two sound sources situated in the same direction (perspective or depth; see Figure 11-24), and
- *Volume* - the size and the shape of the acoustic environment in which the observer is situated.

Azimuth, elevation, and distance represent polar coordinates of any point of interest in a Cartesian space having its origin anchored at the listener's location, and the volume is a global measure of the extent of the space that affects the listener. The set of polar coordinates is shown in Figure 11-28. The awareness of these four elements of space leads to auditory perception of surrounding space. This perception encompasses sensations (perceptions) of directions in horizontal and vertical plane, recognition of auditory distance and auditory depth, and sensation (perception) of ambience (perceived size of space) that together allow us to navigate through the space and feel its spaciousness (see Figure 11-24). They also allow us to distinguish between the specific locations of various sound sources and to describe their relative positions in space. Some of these abilities are the direct result of different auditory stimuli acting on each of the ears of the listener, whereas others result from single ear stimulation. In the latter case, the presence of two ears improves auditory performance, but the perceptual response is not the result of differential processing of two ears' inputs by the auditory system.

Binaural hearing

The human ability to hear a sound with two ears is called *binaural hearing*. If the same sound is received by both ears such auditory stimulation is called the *diotic presentation* and has been described in Chapter 5, *Audio Helmet Mounted Displays*. Diotic presentation of a stimulus results in the lower binaural threshold of hearing and the higher binaural loudness than the respective monaural (single ear) responses. This process is called *binaural summation* or *binaural advantage* and has been discussed previously in this chapter. Note, however, that when a

target sound is presented in noise, the same masked threshold in observed in both the monaural and binaural listening conditions assuming that both the target sound and the noise are the same in both ears (Moore, 1989).

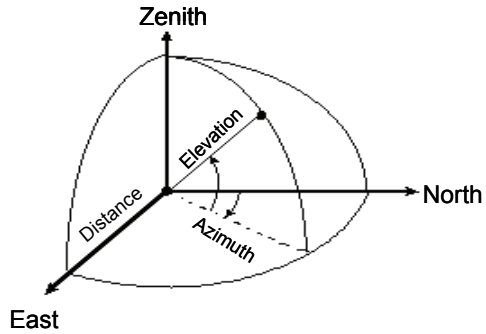

Figure 11-28. Azimuth, elevation, and distance in polar coordinates.

If the same sound is received by both ears but the ears differ in their properties, the binaural advantage is typically less. In addition, the ear disparity may lead to difficulties in pitch perception, called *binaural displacusis* (Van den Brink, 1974; Ward, 1963). The binaural diplacusis is the difference in pitch sensation in the left and right ear in response to the same pure tone stimulus. This difference leads to difficulty in pitch judgments of pure tone stimuli, but it is washed out and not perceived for complex stimuli.

If the sounds received by two ears are not the same in each ear, they can be treated by the auditory system as two independent sounds that contralaterally mask each other or as two slightly different representations of the same stimulus, resulting in a single fused auditory image appearing to exist inside or outside of the listener's head. The actual perceptual response depends on the character and extent of the differences between the sounds, i.e., the degree of correlation between the left and right ear stimuli. Note that the ears being some distance apart allows even the same original sound arriving at the left and right ear to differ to some degree in its spectral content and temporal envelope. Perception of such different but highly correlated stimuli is the basis for sound source localization in space. In contrast, when the left and right ear stimuli are not or poorly correlated with each other, such stimulation is called the *dichotic presentation* (described in Chapter 5, *Audio Helmet Mounted Displays*).

One of the most intriguing phenomena of binaural listening is the *binaural masking level difference* (binaural MLD or BMLD). The binaural MLD is the decrease in the masked threshold of hearing under some binaural listening conditions in comparison to the monaural listening condition. This phenomenon can be observed when a person listens binaurally to a target sound masked by a wideband noise but either target sound or noise differs in phase between the ears.

As mentioned earlier, the binaural masked threshold of hearing is the same as the monaural masked threshold of hearing if both the target sound and the masking noise are identical in both ears. However, when the phase of either the target sound or noise is reversed 180° in phase in one of the ears, the audibility of the target sounds markedly improves (Noffsinger, Martinez and Schaefer, 1985). Even more surprisingly, the monaural masked threshold of hearing improves when the same noise is added to the opposite ear. The improvement is in the order of 9 dB, which is larger than the approximate 6-dB improvement observed when the target sound, rather than the masking noise is added to the opposite ear (Moore, 1989). When both the masking noise and the target sound are added to the opposite ear, the masked threshold increases and becomes again the same as in the monaural listening condition.

The binaural MLD phenomenon was originally reported by Licklider (1948) for speech recognition in noise and by Hirsh (1948) for detection of pure tone signals in noise. Licklider (1948) reported that speech recognition

through earphones in a noisy environment was greatly improved when the wires leading to one of the earphones were reversed. Hirsh (1948) and others (e.g., Durlach and Colburn, 1978; Egan, 1965; Roush and Tait, 1984; Schoeny and Carhart, 1971) reported thereafter that when continuous tone and noise are presented in phase in both ears (S_oN_o condition) or are reversed in phase in both ears ($S_\pi N_\pi$ condition), the masked detection threshold for the tone is the same as for the monaural condition. However, when either the tone or noise are reversed in phase, the $S_\pi N_o$ condition or $S_o N_\pi$ condition, respectively, the detection threshold for the tone improves dramatically and the improvement is as large as 10 to 15 dB. The masking noise can be either a wideband noise or, in the case of a pure tone target sound, a narrowband noise centered on the frequency of the pure tone target sound. The improvement is the greatest for low frequency tones in 100 to 500 Hz range and decreases to 3 dB or less for stimulus frequencies above about 1500 Hz. If the phase shifts are smaller than 180°, a similar, although smaller, binaural MLD effect has been observed. In general, the larger the phase shift the larger the size of the binaural MLD effect. The exact physiologic mechanism of the binaural MLD is still unknown although some MLD results can be explained by the equalization-cancellation (EC) mechanism proposed by Durlach (1963). According to Colburn (1977), the binaural MLD effect also can be accounted for by the response patterns at the outputs of the coincidence detectors in the medial superior olivary (MSO) nucleus (e.g., Colburn, 1977).

Masking by noise aside, if the same sound is received by both ears, the sound source is perceived as located in the median plane of the listener. If the sounds differ in their time of arrival and/or intensity, the sound source is perceived as being located at a certain azimuth angle to the left or to the right of the median plane but not at the median plane. This effect is called *lateralization*, interpreted as "to the left" or "to the right" from the median plane but does not necessarily imply any specific location.

Note that in the case of binaural reception of auditory stimuli a sound source can be perceived as located outside the head (e.g., in natural environments or during loudspeaker-based sound reproduction) or inside the head (e.g., earphone-based sound reproduction). In the former case, the sound source location can be identified relatively precisely in both horizontal and vertical plane. When the sound source is located inside the head, it can only be crudely located on a shallow imaginary arc connecting left and right ear, and the perceived deviation of the auditory image location from the median plane can only be judged as partial of full (toward one of the ears) lateralization. Thus, spatial phenomena inside-the-head is referred to as lateralization while the term localization is reserved for spatial phenomena outside of the head.

Lateralization and binaural cues

Spatial perception of sound is based on two main sets of cues: binaural cues and monaural cues. Binaural cues result from the differences in stimuli received by two ears of the listener and are basic cues facilitating sound lateralization. The binaural cues were described first in 1907 by Lord Rayleigh as foundation of what is often called the Lord Rayleigh's *duplex theory*. There are two binaural cues that facilitate sound localization in the horizontal plane: (a) *interaural level difference* (ILD), also referred to as *interaural intensity difference* (IID), and (b) *interaural time differences* (ITD) or *interaural phase difference* (IPD). Both cues are shown in Figure 11-29.

The IID refers to the difference in the intensity of sound arriving at the two ears caused by the baffling effect of the head. The head is casting an "acoustic shadow" on the ear farther away from the sound source, decreasing the intensity of sound entering that ear. The effectiveness of the baffling effect of the head depends on the relative size of the head (d) and the sound wavelength (λ). The larger the difference between d and λ ($d >> \lambda$), the stronger the baffling effect. Thus, at low frequencies, where the dimensions of the human head are small in comparison to the wavelengths of the sound waves, sound waves diffract around the head, and the difference in sound intensity at the left and right ear is small if any. However, at high frequencies, the intensity differences caused by the acoustic shadow of the head are large and provide effective localization cues. For example, when the sound source is situated in front of one ear of the listener, the IID (or ILD) can be as large as 8 dB at 1 kHz and 30 dB at 10 kHz (Steinberg and Snow, 1934). The effect of sound frequency on the size of the ILD cue is shown in Figure 11-30.

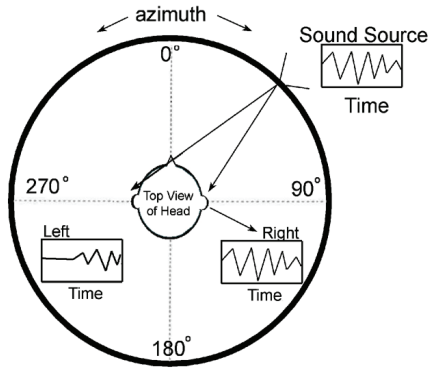

Figure 11-29. The interaural time difference (ITD) and interaural level differences (ILD) created by a sound arriving from 45° azimuth angle. Note that the sound arrives earlier and has more energy at the right ear.

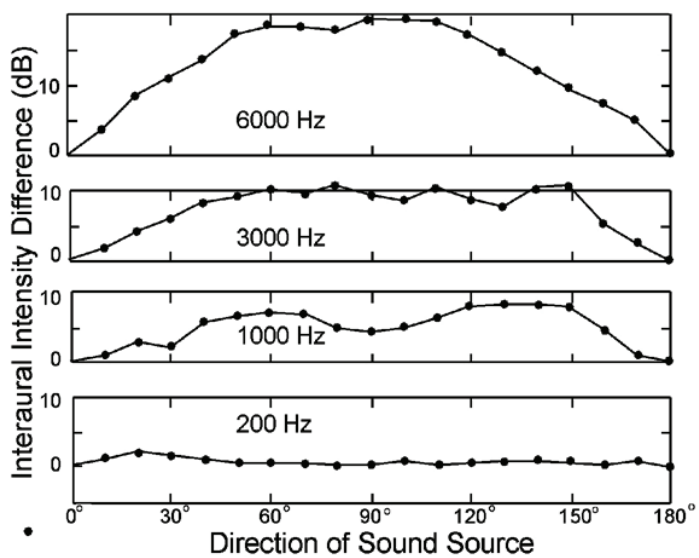

Figure 11-30. Interaural level differences (IID/ILD) for four pure tone signals; 220 Hz, 1000 Hz, 3000 Hz, and 6000 Hz. At 200 Hz, there is no shadowing effect due the sound diffraction around the head (adapted from Feddersen et al., 1957).

The ITD refers to the difference in the time of arrival of the sound wave at the two ears. If a sound source is located in the median plane of the head, there is no difference in the time of sound arrival to the left and right ear. However, if a sound is presented from the side of the head or any other angle off the median plane, the sound reaching the further away ear arrives with a certain time delay. Assuming that the human head can be approximated by a sphere, the resulting time difference can be calculated from the equation:

$$\Delta t = \frac{r}{c}(\alpha + \sin\alpha), \qquad \text{Equation 11-21}$$

where Δt is the ITD in seconds, r is the radius of the sphere (human head) in meters, c is the speed of sound in m/s, and α is the angle (azimuth) of incoming sound in radians (Scharine and Letowski, 2005). The dependence of the ITD on the angular position of the sound source is shown in Figure 11-31. The maximum possible ITD occurs when the sound source is located on the imaginary line connecting both ears of the listener and is dependent on

the size of the listener's head, the speed of sound, and to some extent on the distance of the sound source from the listener's head. For example, for a head with the diameter d = 20 cm and a sound wave velocity c = 340 m/s, the maximum achievable ITD is about 0.8 ms. For a given head size, larger ITDs indicate more lateral and smaller ITDs less lateral sound source locations. The smallest perceived ITD is to order of 0.02 to 0.03 ms and is being detected when the sound is arriving from a 0° angle, i.e., from a sound source directly in front of the listener. This difference corresponds to the shift in the horizontal position of the sound source by about a 2° to 3° angle.

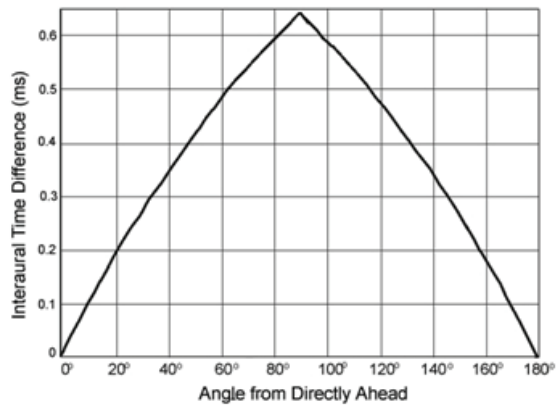

Figure 11-31. Interaural time differences plotted as a function of azimuth (adapted from Feddersen et al., 1957).

The ITD cue works well at lower frequencies but it fails at high frequencies. First, the phase information becomes ambiguous above approximately 1200 to 1500 Hz depending on the size of the head. At this frequency the length of one period of the sine wave corresponds to the maximum time delay of sound traveling around the head. This means that at this and higher frequencies ITD may be larger than duration of a single period of the waveform making time delays ambiguous. This ambiguity is shown in Figure 11-32. Note that the second waveform in the last pair of waveforms is delayed by the whole period regarding the previous waveform, but both of them arrive at the same phase.

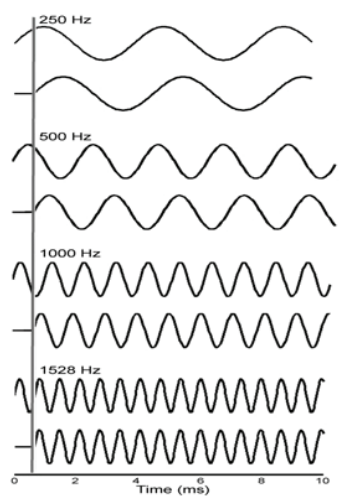

Figure 11-32. Comparison of the interaural phase relationship of various sinusoids.

Second, each auditory neuron fires in synchrony with a particular phase in the auditory waveform. This effect is called *phase locking.* The frequency of neuron firing is limited to about 4 to 5 kHz (Rose et al., 1968; Palmer and Russell, 1986), and this limit is because phase timing variability becomes large with respect to the length of the frequency cycle. This means that any phase synchrony in neuron firing is lost for frequencies above 4 to 5 kHz.

Note that the high frequency limitations discussed above are actually the interaural phase difference (IPD) limitations and apply only to the continuous stimuli. However, in the case of clicks, onset transients, and similar non-periodic sounds, the time difference shorter than 0.6 to 0.8 ms can be used to guide sound localization since this is the difference between two single temporal events that are not repeated periodically (Leakey, Sayers and Cherry, 1958; Henning, 1974).

Both of the above mechanisms limit the use of ITD to localization of only low and mid frequency sounds. However, high frequency sound can be localized using ILDs. The complimentary character of the ITD and ILD cues became the foundation of the Lord Rayleigh's (1907) duplex theory, which states that the lateral position of a sound source in the space is determined by the combination of both cues, ILDs at high frequencies and ITDs at low frequencies. A consequence of the duplex theory is that sounds containing frequencies between 1 to 4 kHz should be difficult to lateralize accurately because neither the ITD nor ILD cue is strong enough in this frequency region. Later studies have largely confirmed this theoretical assumption (Stevens and Newman, 1936; Wightman and Kistler, 1992). However, it should be cautioned that these facts only hold true for pure tones. Most sounds are composed of multiple frequencies and can be lateralized using a combination of both cues for their lower and higher components. Thus, pulses of wideband noise containing both the low-and high-frequency energy are the easiest stimuli to localize (Hartmann and Rakerd, 1989) and are the preferred stimuli for directional beacons (Tran, Letowski and Abouchacra, 2000).

Localization and monaural cues

Binaural cues allow effective left-right lateralization, but they have two major limitations. First, binaural cues do not differentiate between sound arriving from the front or the rear of the head. Relative symmetry between front and back of the head results in confusion as to whether the sound is arriving for example, from the 10° or 170° direction. Some binaural differentiation is possible because the head is not cylindrical and the ears locations are not exactly symmetrical, but the front-back localization is not improved much by binaural cues.

Second, binaural cues do not provide any information about sound source elevation. In fact, if one assumes a spherical head, there is a conical region, called a *cone of confusion*, for which a given set of binaural cues is the same. This means that all sound sources located on the surface of a given cone of confusion generate identical binaural cues. As a result, the relative locations of these sound sources cannot be differentiated by the binaural cues alone (Oldfield and Parker, 1986). The concept of a cone of confusion is shown in Figure 11-33. Numerous studies have demonstrated that the cone of confusion is the source of localization errors in both the vertical and the front-back directions (e.g., Oldfield and Parker, 1984; Makous and Middlebrooks, 1990).

The differences between various sound-source locations within a cone of confusion are resolved by the presence of the spectral localization cues, called also the *monaural cues* since they do not require two ears to operate. Monaural cues are the primary cues allowing sound source localization in the vertical place and along the front-back axis.

Monaural cues are directionally dependent spectral changes that occur when sound is reflected from the folds of the pinnae and the shoulders of the listener. These reflections create peaks and notches in the spectrum of the arriving auditory stimulus, changing the spectral content of the waveform arriving at the tympanic membrane. This effect is described in Chapter 9, *Auditory Function*, and shown in two different ways in Figures 9-2 and 11-34. The locations of peaks and notches in the sound spectrum of the arriving auditory stimulus change as a function of the angle of incidence, thus providing information that can be used to distinguish the front from the rear hemisphere (Musicant and Butler, 1985) and between various elevations (Batteau, 1967; Hebrank and Wright, 1974).

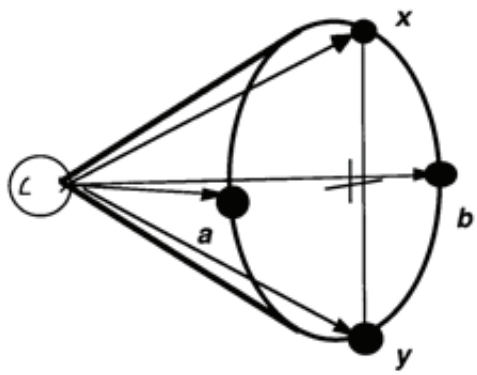

Figure 11-33. The concept of the *Cone of Confusion*. The cone represents a region for which interaural level and phase cues would be the same if a spherical head is assumed (after Mills, 1972).

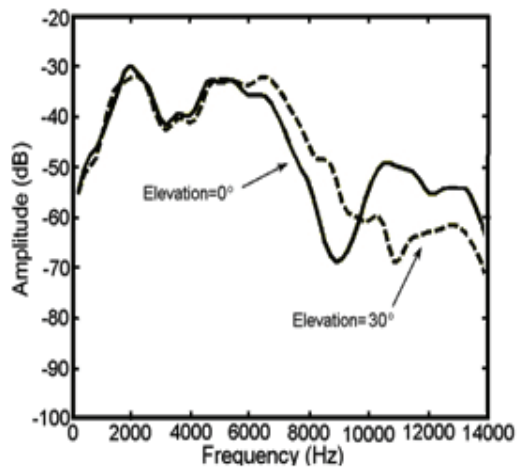

Figure 11-34. Two head-related transfer functions measured for 0° and 30°. This figure illustrates how the frequency notch changes as a function of angular position.

Passive filtering of sound by the concave surfaces and ridges of the pinna is the dominant monaural cue used in sound localization. Gardner and Gardner (1973) observed that localization performance for sound stimuli located on the medial sagittal plane got progressively worse as the pinnae cavities were filled in by using silicon fillers custom-made for each listener. They also observed that the precision of sound source localization in the vertical plane was the best for wideband noises and for narrowband noises with center frequencies in 8 to 10 kHz region. The filtering effect of the shoulders is weaker than that of the concha and pinna ridges, but it is also important for sound localization since it operates in slightly lower frequency range than the others.

Despite their name – monaural cues – these cues are duplicated by the simultaneous monitoring of the sound source location by both ears of the listener. Any asymmetry in the vertical or front-back location of the ears on the surface of the head provides important enhancement of the listener's ability to localize sounds along the respective directions. For some species, like owls, ear asymmetry is the main localization cue in the vertical direction. It also has been mentioned that monaural cues do not only operate in the vertical plane and along front-

back axis, but they also operate together with the binaural cues along the left-right axis and enhance human localization precision in the horizontal plane.

The effects of both binaural and monaural cues can be captured by the placement of very small probe microphones in the ear canal of the listener. A sound is presented from various angular locations at some distance from the human head, and the directional effects of the human body and head are captured by the microphone recordings. The difference between the spectrum of the original auditory stimulus and the spectrum of the auditory stimulus recorded in the ear canal is called the *head-related transfer function* (HRTF). The HRTF varies as a function of the angle of incidence of the arriving auditory stimulus and, for small distances between the head and the sound source, also as a function of the distance (Brungart, 1999).

The HRTFs can be recorded for a selection of azimuths and elevations relative to the orientation of the listener's head and in the form of impulse responses convolved with any arbitrary sound to provide arbitrary spatial information about the sound source location. This technique is used to create externalized spatial locations of the sound sources when the auditory stimuli are presented through the earphones. Such spatial reproduction of sound through earphones is often referred to as *3-D audio* when referring to auditory display systems. Additional information about practical applications of the HRTFs may be found in Chapter 5, *Audio Helmet-Mounted Displays.*

It needs to be stressed that the monaural cues are relative cues. Unless a listener is familiar with the original signal and surrounding space, there is no invariant reference to be used to determine what notches and peaks related to sound source location are present in the arriving auditory stimulus. Therefore, sound localization ability, especially in the vertical plane and along the front-back axis, improves with experience and familiarization with both the stimuli and environment (Plenge, 1971). This is also the reason that some authors consider auditory memory as another auditory directional cue (Plenge and Brunschen, 1971). For example, if a listener is familiar with somebody's voice, this familiarity may help the listener to differentiate whether the talker is located in front or behind the listener. The lack of familiarity with specific auditory stimuli is reported frequently in the literature as the secondary reason for front-back confusions and poor localization in vertical plane.

There is one more potential reason for the poor front-back and vertical discrimination of sound source locations. Blauert (2001) observed that narrowband stimuli presented within the medial sagittal plane have the tendency to be associated with the front, rear, or overhead, regardless of the actual position of the sound source. Since this tendency is the same for sounds located in the specific frequency bands, Blauert called these bands the *directional bands*. The concept of the bands is shown in Figure 11-35. In general, the listeners have the tendency to localize stimuli in 125 to 500 Hz and 2 to 6 kHz bands as coming from the front, stimuli in 500 to 2000 Hz and 10 to14 kHz bands as coming from the back, and stimuli in 6 to 10 kHz band as coming from the top if they do not have any other environmental cues. The bands apply to tones and narrowband stimuli but under some condition they may also affect localization of more complex stimuli.

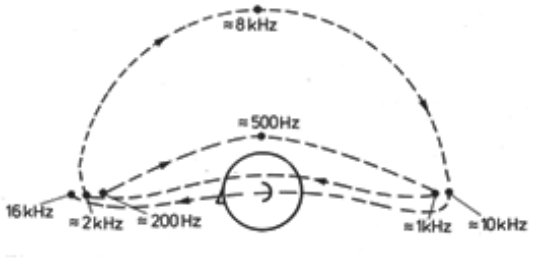

Figure 11-35. The listener's tendency to localize narrowband noises as coming from the front, top, or back if the sound is presented the same number of times from each of these direction (after Blauert, 2001 [Figure 2.6]).

The last but very effective cue that is extremely important in real world environments is that provided by head movement. Even if a sound is unfamiliar, the auditory system can gain disambiguating information if the listener moves his or her head while the sound is present (Perrott and Musicant, 1981; Thurlow, Mangels and Runge, 1967). Small movements of the head from the left to the right will quickly clarify whether the sound is in the front or rear hemisphere. Vertical movements give salient elevation information (Wallach, 1940). Assuming that the sound is long enough to allow for movement, most of the shortcomings of binaural and monaural cues can be overcome (Hirsh, 1971).

All previous discussion has centered on sounds emitted from the stationary sound sources. There are two general measures of sound localization ability that apply to stationary sound sources: *localization acuity* and *localization accuracy*. Localization acuity is a person's ability to discriminate whether the sound source changed its position or not. It is usually described as the *minimum audible angle* (MAA), which is the DL of directional perception. Localization accuracy is a person's ability to localize the sound source in space. It is usually characterized by the standard error (or other measure of dispersion) in the direction recognition task. However, in the real world environments a large proportion of sound sources is not stationary but is moving at various directions and various speeds. Human localization precision of such sound sources is usually measured as the *minimum audible movement angle* (MAMA) for specific direction and speed of the moving sound source (Perrott and Musicant, 1977; 1981). The MAMA is usually larger than the MAA, but they characterize different auditory abilities of the listener. In terms of absolute sound localization, a fast sound source movement makes it more difficult to identify the momentary position of sound source.

The polar characteristic representing a listener's ability to localize sounds in the horizontal plane is frequently referred to as the directional characteristic of the human head. Such a characteristic usually is measured with a narrowband signal for a selection of azimuth angles or by using a turntable and an automatic Bekesy tracing technique (Zera, Boehm and Letowski, 1982). The data usually are displayed as a family of polar patterns obtained for signals of various frequencies in a manner similar to polar patterns of microphones or loudspeakers. Another method of displaying directional characteristics of the listener's head is shown in Figure 11-36 where localization precision for a selected stimulus is shown in numeric form on the imaginary circle around the listener's head.

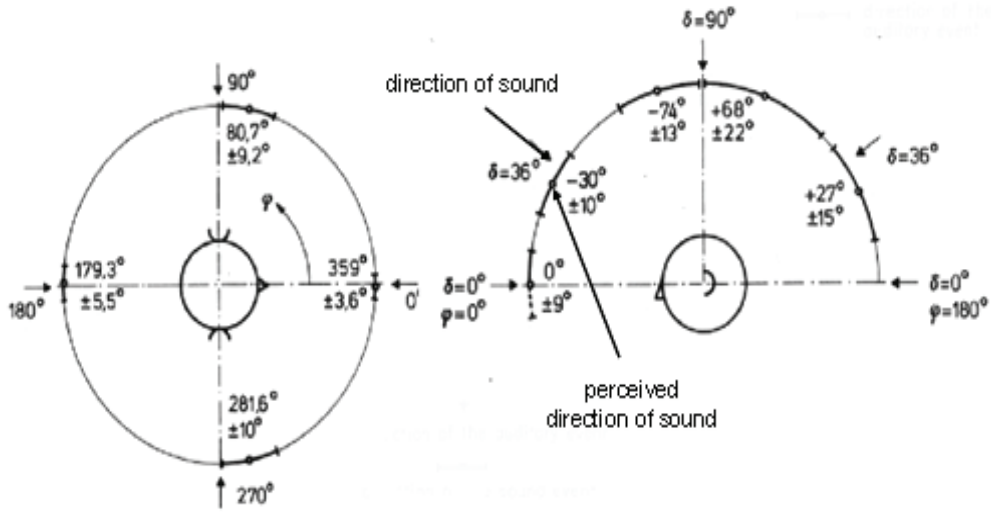

Figure 11-36. Localization uncertainty in the horizontal plane (left panel) and vertical plane (right panel) of white noise pulses of 100 ms duration presented at 70 dB phon level (adapted from Blauert, 2001 [Figures 2.2 and 2.5]).

The precedence effect

Much of the research reported here was conducted in free field environments, e.g., in open outdoor spaces or in anechoic chambers. However, such environments are rare, and it is more common to hear sounds in rooms, near buildings, and in other locations where there are reflective surfaces. The effect of such environments is that one or more reflected sounds arrive at the listener's ears shortly after the original sound. Most of the time, the listener is unaware of this reflected sound, and to some extent continues to be able to localize the sound accurately in spite of the presence of reflected sounds. The mechanism by which this occurs is called the precedence effect.

The *precedence effect* is the phenomenon that the perception of the second of two successively received similar sounds is suppressed if the second sound is delayed up to about 30 to 40 ms, and its intensity does not exceed the intensity of the first sound by more than 10 dB. The precedence effect was discovered originally by Wallach, Newman and Rosenzweig (1949), who conducted a series of experiments testing the effect of a delayed copy of the sound on localization by presenting pairs of clicks over headphones. They observed that if the clicks were less than 5 ms apart, they are fused and are perceived as a single sound image located in the center of the head. However, if the clicks were 5 to 30 ms apart only the first of the two clicks is heard.

The existence of the precedence effect explains the ability of the auditory system to determine the actual position of the sound source without being confused by early sound reflections. If the locations of the two arriving sounds differ, the perceived location of the fused click image is largely determined by the location of the first click. The suppression of the location information carried by the second sounds is known also as he *Haas effect* (1951), after Haas who rediscovered this effect in 1951, and as the *law of the first wavefront*.

In a typical precedence effect demonstration, the same sound is emitted by two loudspeakers separated in space. If the sound coming from one of the loudspeakers is delayed by less than about 1 ms, the fused image is localized somewhere between the locations of both loudspeakers in agreement with the ITD mechanism described previously. Such phantom sound source location resulting from perception of two separate sounds is called *summing location* and the process is called *summing localization* (Blauert, 1999). The range of time in which summing localization operates is approximately equivalent to maximum ITD for a given listener.

When the time delay of the lagging sound is between 1 and 5 ms, the sound appears to be coming from only the lead loudspeaker, but its timbre and depth perception change. If the time delay of the second sound is more than 5 ms but less than 30 ms, depending on the specific environment, only the first sound is heard, and the second sound has no effect. Obviously, if the lagging sound is more than 10 to 15 dB more intense than the leading sound, only the second sound and its direction are heard (Moore, (1997). If the time delay is longer than about 30 ms and the sounds are very similar, the second sound is heard as an echo of the first sound. If the sounds are very different, two separate sounds are heard.

To some degree, the two sounds can be different and the precedence effect may still occur (Divenyi, 1992), but similarity increases the effect. It also has been shown that the precedence effect can take some time to build up (Freyman, Clifton and Litovsky, 1991). The authors described an experiment in which a train of leading clicks with simulated echo clicks delayed by 8 ms was presented. At first, two clicks were clearly heard, but after a few repetitions, the two clicks fused. This fusion is disrupted if the acoustical conditions are changed (Clifton, 1987) For example, Clifton showed that if a train of lead-lag click pairs is presented and then the locations of the leading and lagging click are swapped, fusion is disrupted temporarily and the clicks are once again heard separately. After a few more presentations, they fuse again (see Chapter 13, *Auditory Conflicts and Illusions*, for a more complete description of the Clifton effect and a related effect, the Franssen effect). This can be compared with becoming adapted to a particular acoustic environment. After a few seconds, one begins to ignore the acoustic effects of the room. If the room were suddenly to become drastically altered (an improbable event), the echoes suddenly would become more apparent, only to fade away afterwards.

Auditory distance perception

Auditory distance is the distance between the listener and a sound source, determined on the basis of available auditory cues. If the perception involves an estimation of the distance between two sound sources located along the same imaginary line passing through the head of the listener, such distance is called *auditory depth.* In both cases there are no absolute cues for distance perception; however, there are several relative ones, which combined with non-auditory information, allow individuals to determine the distance from the sound source to the listener. These cues, called distance cues or range cues, depend on the specific environment but in general include: (a) sound loudness, (b) spectral changes, (c) space reverberance (liveness), and (c) motion parallax.

The primary auditory distance cue is *sound loudness*. For familiar sounds, one can compare the loudness of the perceived sound with the knowledge about the natural loudness and intensity of its source (Mershon and King, 1975). In the case of the prerecorded sounds, a critical requirement is that the prerecorded and the reproduced sounds have the same loudness (Brungart and Scott, 2001). Still, the distance to an unfamiliar sound source (or a sound source that may have various sound loudnesses at the source) is difficult to estimate using the loudness cue.

The loudness cue is the most obvious distance estimation cue in the outdoor environments. According to the inverse square law of sound propagation in open space, sound intensity decreases by 6 dB per doubling of the distance from the sound source. However, this rule only holds true for free-field environments. In enclosed spaces, wall reflections reduce this intensity drop associated with the distance from the sound source and at some *critical distance*, which is a function of the sound source distance from the listener and the reflecting walls, obviate the cue altogether.

The second cue is *sound timbre*. Low frequency components are less likely to be obstructed by objects and meteorological conditions than high frequency components of a sound. High frequency components are attenuated by humidity and transmitting matter and absorbed by nearby surfaces. Consequently, distant sounds will have relatively more low frequency energy than the same sounds radiated from proximal (nearby) sound sources and result in different sound timbre. Unfortunately, this cue also requires knowledge of the original sound source in order to be used effectively utilized (Little, Mershon and Cox, 1992; McGregor, Horn, and Todd, 1985).

The third cue, *reverberance* or *liveness*, is a major cue for distance perception in closed spaces or in situations that produce an echo. If the sound source is located close to the listener, the direct-to-reverberant sound ratio is high and sound clarity (e.g., speech intelligibility) is high as well. If the distance is large, the sound becomes less clear and its sound source location less certain (Mershon et al., 1989; Nielsen, 1993). For a given listening space, there may be multiple sound sources each with their own direct-to-reverberant sound intensity ratio. A listener familiar with that space can use this as a source of distance information. However, the specific ratio of these two energies depends on the directivity of the sound source and the location of the listener relative to the sound source and reflective surfaces of the space (Mershon and King. 1975).

Finally, when the listener moves (translates) the head, the change in the azimuth toward the sound source is distance dependent. This cue is called the *motion parallax cue*. If the sound source is located far away from the listener even a relatively large displacement of the head position results only in very small change in the azimuth and relatively small and slow imaginary movement of the sound source. Conversely, if the sound source is located nearby, even a small movement of the head causes a large change in the azimuth that results in a larger imaginary movement of the sound source and in a potential change in the loudness of the sound.

Despite the four auditory cues listed above, the auditory distance estimation is relatively poor, especially if the estimate is expressed in absolute numbers as opposed to a comparative judgment of two distances in space. In general, people underestimate the distance to the sound source in an exponential manner -- the larger the distance to the sound source the larger relative error (Fluitt, Letowski and Mermagen, 2003).

Space perception (Spaciousness)

Spaciousness is a catch-all term describing the human impression made by a surrounding acoustic space. Spaciousness embraces such sensations as ambience (impression that sound is coming from many directions), reverberance or liveness (impression of how much sound is reflected from the walls), warmth (impression of the spectral balance/imbalance in the reflected sound energy), intimacy, panorama, perspective, and many others. Similarly to the timbre domain, some of spaciousness-related terms are highly correlated and there are many that do not have an established meaning.

There are different types of acoustic spaces and therefore different forms of spaciousness. They include such spatial concepts as battlefield, serenity, and soundscape. Each of these terms brings with it a connotation of a specific sonic environment, and often related to it an emotional underpinning reflected in perceived spatial sound quality. For example, Krause (2008) divides all soundscapes into biophony (spaces dominated by sounds produced by non-human living organisms), geophony (spaces dominated by non-biological sounds of nature), and anthrophony (spaces dominated my man-made sounds). Every sound within a particular space brings with it information about the space as well as certain expectations regarding its natural fit within this environment.

In its general meaning, spaciousness is a sensation parallel to the sensation of timbre described previously. It is a perceptual reflection of the size and the character of the area over which a particular sound can be heard and perceived. Therefore, it is a perceptual characteristic of a soundstage, that is, a sound source operating within a particular environment. It needs both the sound and the space to exist. One important element of spaciousness is the size of a personal space in voice communication. Personal space can be generally defined as the area surrounding an individual that the individual considers as personal territory in any human-to-human interaction (Hall, 1966). A personal space is usually highly variable and depends on the personal traits of the individual and the social and cultural upbringing. For example, in the Nordic cultures the radius of personal space is generally larger than in the Southern cultures.

The same general comment about the variability of a personal space applies to auditory personal space defining the minimum acceptable distance between two unrelated people who communicate by voice. However, the radius of the auditory personal space seems to vary between individuals and is less than the radii of social space, aggression space, or work space. It generally is assumed that the distance of one meter (3.28 feet) defines a typical conversational situation and serves as a good estimate of the radius of the auditory personal space.

The concept of the auditory personal space is important for audio communication and this space needs to be preserved in creating phantom sound sources representing real people communicating through audio channels with real or virtual people. If the perceived auditory distance to another talker in an audio channel is quite different from 1 to 1.5 m, such voice communication will distract the operator, increase the workload, and increase the overall level of anxiety. Obviously, the above recommendation has only a general characteristic, and there are specific situations that the communication distance has to be different.

Hearing Loss

Hearing loss is a decreased ability to perceive sound due to an abnormality in the hearing mechanism. According to the American Speech-Language-Hearing Association (ASHA), 28.6 million people in the United States are living with hearing loss (ASHA, 2006). Hearing loss can affect not only the individual's ability to hear the surrounding environment but also the clarity of speech perception. The functional effects of hearing loss can vary greatly according to the age of onset, period of onset, degree, configuration, etiology, and the individual's communication environment and needs.

Three main types of hearing loss are labeled as: conductive, sensorineural, and mixed hearing loss. With conductive hearing losses, either the outer ear, middle ear or both are affected. Sensorineural hearing refers to loss that originates in the cochlea, the auditory nerve, or in any portion of the central auditory nervous system

(CANS). A mixed hearing loss is a combination of a sensorineural and a conductive hearing loss, and therefore, can involve many combinations and/or portions of the ear.

The disorders of the outer ear that can lead to a conductive hearing loss can be caused by congenital anomalies or acquired ones. Examples of congenital anomalies include narrowing of the ear canal known as stenosis, an absence of the ear canal known as atresia, a partially formed pinna known as microtia, or an absence of the pinna known as anotia. Examples of acquired anomalies include impacted ear wax or a foreign body in the ear canal. Otitis externa, also known as "swimmer's ear", can cause stenosis of the canal, and therefore, lead to a conductive hearing loss. Most conductive hearing losses within the outer ear partially or completely block the transmission of acoustic energy into the middle ear and they are treatable. The resonance of the pinna and the outer ear canal is between 2 to 7 kHz, which enhances the ability of the listener to localize and perceive their acoustic space (Rappaport and Provencal, 2002). Recalling that higher frequency sounds have shorter wavelengths, and therefore, can more easily be deflected, anomalies in the outer can impede the listener's ability to localize.

Abnormalities in the middle ear causing conductive hearing losses can also be congenital or acquired. Serous otitis media, or inflammation of the middle ear fluid, can cause conductive hearing losses and are temporary in 90% of the cases and usually resolve without treatment within 3 months (Rappaport and Provencal, 2002). Perforations on the tympanic membrane can lead to minimal or maximal hearing loss depending on where the perforation occurs on the membrane. An abnormality in the ossicles caused by chronic ear infections can lead to ossicular erosion, creating a maximum conductive hearing loss of about 60 dB. Conductive hearing losses only affect up to about 60 dB since bone conduction hearing takes place beyond that level. Bone conduction hearing occurs when the sound intensity is strong enough to bypass the middle ear and move the bones in the skull, which moves the cochlear fluids in the inner ear. This stimulates hair cells, which leads to the perception of an auditory signal. Like outer ear conductive hearing losses, losses caused by middle-ear abnormalities are usually treatable.

Sensorineural hearing losses can be caused by one of hundreds of syndromes, a single genetic anomaly, perinatal infection, drugs that are toxic to the auditory system, tumors, idiopathic disease process, aging, or from noise exposure. ASHA reports that 10 million of the 28 million with hearing loss are due to noise exposure (ASHA, 2006). Sensorineural hearing loss is characterized by irreversible damage that distorts auditory perception. Generally, "sensory" hearing loss refers to an abnormality in the cochlea and "neural" refers to an abnormality beyond the cochlea, meaning in the VIIIth nerve or beyond.

Sensorineural loss can be either permanent or temporary. Temporary hearing loss, called also *temporary threshold shift* (TTS), is usually noise-induced and may last from under an hour to several days, and the degree and duration of loss depends upon the duration, intensity, and frequency of the noise exposure (Feuerstein, 2002). Excessive exposure to sounds energy in the frequency range from 2000 to 6000 Hz is most likely to cause permanent changes in hearing. Permanent hearing loss (permanent threshold shift [PTS]), is the residual loss from a noise-induced temporary hearing loss or aging process and results from damaged hair cells. Usually it is a slow process and the individual may not perceive any change in hearing for long time. If the hearing loss is due to acoustic trauma (sudden exposure to very intense noise), the PTS typically will plateau by within the first eight hours. With conductive hearing loss, once sound level is increased to compensate for the PTS, the signal is audible and clear. With sensorineural hearing loss, at frequencies where the PTS occurs, even once audibility is restored, some level of sound distortion usually persists.

The mechanism for cochlear damage may come from a variety of sources including: interruption of blood flow due to ischemic damage, mechanical injury to the hair cells due to the shearing force of the traveling wave, hair cell toxicity caused by a disruption of ionic balances in the cochlear fluids, or hair cell toxicity from an over-active metabolic process caused by an immune response (Lonsbury-Martin and Martin 1993). Age-related hearing loss is most commonly seen in the 7th decade of life, with the basal region of the cochlea being most affected (Weinstein, 2000; Willott, 1991). Both noise-induced and age-related hearing loss (presbycusis) are characterized by high-frequency loss of varying degrees. Therefore, presbycusis, or age related hearing loss, is often difficult to tease out from noise-induced hearing loss. However, age has been positively correlated with a worsening in pure tone thresholds. As women age, on average their thresholds from 3000 Hz and above worsen to a mild hearing

loss, whereas men's threshold from 2000 Hz and above worsen to a mild to moderate hearing loss (Weinstein, 2000). Given that background noise is generally low-frequency, this type of noise can exacerbate a high-frequency hearing loss since it can mask the portion of the signal that the listener is able to hear in quiet.

As previously stated, mixed hearing loss is a combination of both sensorineural and conductive hearing loss. Otosclerosis is an example of mixed hearing loss that is caused by a disease process on the ossicular chain, most commonly affecting the footplate of the stapes. Although the cause is unknown, the disease process usually softens the bone, and the bone then hardens and can become fixed to the oval window. Since the resonant frequency of the ossicular chain is around 2 kHz, the hearing loss usually presents as a conductive hearing loss with a sensorineural component at 2 kHz. Another example of a mixed hearing loss could be due to severe acoustic trauma. Trauma could cause a tympanic membrane perforation as well as a noise-induced hearing loss. Mixed losses can be caused by a variety of pathologies or combination of pathologies that can vary greatly in severity.

Hearing loss is described by the degree and type, and configuration. The basic metric used to assess degree of hearing loss is the *pure tone average* (PTA) calculated as an average hearing threshold across a specific range of frequencies. The frequencies considered most important to speech perception are 500, 1000, and 2000 Hz. A loss at these frequencies can more adversely affect speech perception than one that occurs at 3000 Hz and above. Therefore, PTA is most commonly calculated as the average value of the threshold of hearing at 500, 1000, and 2000 Hz expressed in dB. Sometimes the average includes different combination of frequencies, which in such cases should be clearly stated. If they are not, the 500, 1000, and 2000 Hz average needs to be assumed.

The degree of hearing loss based on standard PTA calculation is separated into seven categories listed in Table 11-19. Since this range of frequencies is the most important for speech recognition such defined PTA should be numerically close to the speech reception threshold, which is usually within 5 dB of each other.

Table 11-19.
Classification of hearing loss (Harrell, 2002).

Extent of Hearing Loss (dB HL)	Degree of Hearing Loss
-10 to 15	Normal
16 to 25	Slight
26 to 40	Mild
41 to 55	Moderate
56 to 70	Moderately-severe
71 to 90	Severe
>90	Profound

The PTA metric is generally a monaural metric and defines hearing loss for each ear separately. A symmetric hearing loss is assumed when the PTAs calculated across 500, 1000, and 2000 Hz, and frequently also 3000 Hz or 4000 Hz frequencies for the left and right ear are within 10 dB of each other. Binaural hearing loss may be assessed by the PTA by calculating an arithmetic average of the PTAs obtained separately for left and right ears or by using a better threshold level in either ear at 500, 1000, and 2000 Hz. Each approach leads to a slightly different result but there is yet no strict standard accepted how to calculate the bilateral PTA.

Configuration refers to the hearing thresholds relative to one another. For example, a flat hearing loss means that less than a 5 dB average change exists among octaves. Other categories include gradually sloping, sharply sloping, precipitously sloping, rising, and notch, but their descriptions are beyond the scope of this book.

The U.S. Army classifies hearing loss into four fitness-for-duty categories, H1-H4 (Army Regulation 40-501 [Department of the Army, 2008]). An H-1 designation means that no limitations exist based on the Warfighter's hearing. The determination of fitness for duty is based on threshold levels, measured in dB HL at 500, 1000, 2000, and 4000 Hz. For an H-1 designation, neither ear can have an average threshold at 500, 1000, and 2000 Hz

greater than 25 dB HL, with no individual level greater than 30dB. Thresholds at 4000 Hz cannot exceed 45 dB HL. For a list a specific requirements see AR 40-501. An example of a hearing test is given in Figure 11-37. In general, hearing profiles are intended to influence the Warfighter's occupation specialty to ensure that no further loss results to due duty and that no harm results due to the hearing loss.

The Department of Defense and the Occupational Health and Safety Association (OHSA) require Warfighters to have hearing threshold tests prior to hazardous noise exposure. For Warfighters routinely exposed to noise hazards, annual pure tone threshold tests are required (AR 40-501). Not only does the annual hearing test help determine the need for further audiology testing to evaluate fitness for duty status, but also monitors any change in hearing that may occur. Specifically, these hearing tests document if any significant changes that occur at 2000, 3000, and 4000 kHz. A *significant threshold shift* (STS) is defined as a 10 dB or more average shift at the aforementioned frequencies. A STS can be consistent with noise-induced hearing loss and can alert the Warfighter and his/her command to needed improvements in compliance with hearing conservation measures (i.e., hearing protection devices). Permanent noise-induced hearing loss is a pervasive hazard in the military but it is preventable.

REFERENCE AUDIOGRAM

1. ZIP CODE/APO/FPO/PAS

(This form is subject to the Privacy Act of 1974 - use Blanket PAS - DD Form 2005)

AUDIOMETRY

1 — 1 - REFERENCE ESTABLISHED PRIOR TO INITIAL DUTY IN HAZARDOUS NOISE AREAS; 2 - REFERENCE ESTABLISHED FOLLOWING EXPOSURE IN NOISE DUTIES; 3 - REFERENCE RE-ESTABLISHED AFTER FOLLOW-UP PROGRAM

16. AUDIOMETRIC DATA RE: ANSI S3.6 - 1996	LEFT						RIGHT					
	500	1000	2000	3000	4000	6000	500	1000	2000	3000	4000	6000
17. DATE OF AUDIOGRAM 11-OCT-2006	-5	0	5	65	70	45	5	0	0	25	60	45

18. MEETS REFERRAL CRITERIA	19. MILITARY TIME OF DAY *(Optional)*	20. HOURS SINCE LAST NOISE EXPOSURE	21. EAR, NOSE, AND THROAT PROBLEM AT TIME OF TEST
2 — 1 - NO, 2 - YES	13:55:50	14	2 — 1 - NO, 2 - YES, 3 - UNKNOWN

Figure 11-37. Example of a hearing test from a Warfighter with an H-3 hearing test.

References

Abel, S.M. (1972). Discrimination of temporal gaps. *Journal of the Acoustical Society of America,* 52, 519-524.

Abouchacra, K., and Letowski, T. (1999). Comparison of air-conduction and bone-conduction hearing thresholds for pure tones and octave-band filtered sound effects. *Journal of the American Academy of Audiology,* 10, 422-428.

Abouchacra, K., Letowski, T., and Gothie, J. (2007). Detection and recognition of natural sounds. *Archives of Acoustics,* 32, 603-616.

American Speech-Language-Hearing Association. (2006). The Prevalence and Incidence of Hearing Loss in Adults. Retrieved October 1, 2006 from: http://www.asha.org/public/hearing/disorders/prevalence_adults.htm

Anderson, C.M.B., and Whittle, L.S. (1971). Physiological noise and the missing 6 dB. *Acustica,* 24, 261-272.

American National Standards Institute. (1989). Method for measuring the intelligibility of speech over communication systems. ANSI S3.2-1989. New York: American National Standards Institute (ANSI).

American National Standards Institute. (1994). Acoustical terminology. ANSI S1.1-1994 (R2004). New York: American National Standards Institute (ANSI).

American National Standards Institute. (1997). Methods for calculation of the speech intelligibility index. ANSI S3.2-1997 (R2007). New York: American National Standards Institute (ANSI).

American National Standards Institute. (2004). Specifications for Audiometers. ANSI S3.6-2004. New York: American National Standards Institute (ANSI).

American National Standards Institute. (2005). Measurement of sound pressure levels in air. ANSI S1.13-2005. New York: American National Standards Institute (ANSI).

American National Standards Institute. (2007). Procedure for the computation of loudness of steady sounds. ANSI S3.7-2007. New York: American National Standards Institute (ANSI).

Aures, W. (1984). Berechnungsverfahren für den Wohlklang beliebiger Schallsignale, ein Beitrag zur gehörbezogenen Schallanalyse, Doctoral dissertation; Technische Univesität München.

Aures, W. (1985). Ein Berechnungsverfahren der Rauhigkeit. *Acustica*, 58, 268-281.

Ballas, J.M., and Howard, J.H. (1987). Interpreting the language of environmental sounds. *Environment and Behavior,* 19*,* 91-114.

Ballas, J.M., Dick, K.N., and Groshek, M.R. (1987). Failure to identify "identifiable" sounds. *Proceedings of the 31st Annual Meeting of the Human Factors Society* (p. 144-146). New York: Human Factors Society.

Ballas, J.M., and Barnes, M.E. (1988). Everyday sound perception and aging. *Proceedings of the 32nd Annual Meeting of the Human Factors Society* (p. 194-197). Anaheim, CA: Human Factors Society.

Ballas, J.M. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology: Human Perception and Performance,* 19*,* 250-267.

Batteau, D.W. (1967). The role of the pinna in human localization. *Proceedings of the Royal Society B.* 168*,* 158-180.

Beattie, R.C., and Culibrk, J. (1980). Effect of competing message on the speech comfortable loudness level for two instructional sets. *Ear and Hearing,* 1*,* 242-248.

Bench J., Kowal A., and Bamford J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology,* 13*,* 108-12.

Beranek, L.L. (1947). The design of speech communication systems. *Proceedings of the Institute of Radio Engineers*, 35, 880-890.

Beranek, L.L., Marshall, J.L., Cudworth, A., and Peterson, A.P.G. (1951). Calculation and measurement of the loudness of sound. *Journal of the Acoustical Society of America*, 23, 161-169.

Berger, K., (1969). A speech discrimination task using multiple-choice key words in sentences. *Journal of Auditory Research*, 9*,* 247-262.

Berlin, C.I., Bordelon, J., St. John, P., Wilensky, D., Hurley A., Kluka, E., and Hood, L.J. (1998). Reversing click polarity may uncover auditory neuropathy in infants. *Ear and Hearing,* 19*,* 37-47.

Best, V., Ozmeral, E., Gallun, F.J., Sen, K., and Shinn-Cunningham, B.G. (2005). Spatial unmasking of birdsong in human listeners: Energetic and informational factors. *Journal of the Acoustical Society of America,* 118*,* 3766-3773.

Bilger, R.C., Nuetzel, J.M., Rabinowitz, W.M., and Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research,* 27*,* 32-48.

Bindra, D., Williams, J.A., and Wise, J.S. (1965). Judgment of sameness and difference: Experiments in decision time. *Science,* 150*,* 1625-1627.

Bismarck, G. von. (1974a). Timbre of steady sounds, A factorial investigation of its verbal attributes. *Acustica*, 30, 146-158.

Bismarck, G. von. (1974b). Sharpness as an attribute of the timbre of steady sounds. *Acustica*, 30, 159-171.

Black, J.W. (1957). Multiple choice intelligibility tests. *Journal of Speech and Hearing Disorders,* 22(2), 213-235.

Blauert, J. (2001). *Spatial hearing: The psychophysics of human sound localization (3rd Ed.).* Cambridge, MA: MIT Press.

Block, G.B., Killion, M.C., and Tillman, T.W. (2004). The "Missing 6 dB" of Tillman, Johnson, and Olsen was found – 30 years ago. *Seminars in Hearing,* 25*,* 7-16.

Blue, M., Ntuen, C., and Letowski, T. (2004). Speech Intelligibility of the Callsign Acquisition Test in a Quiet Environment. *International Journal of Occupational Safety and Ergonomics,* 10*,* 179-189.

Boer, E. de. (1956). On the residue in hearing. Doctoral dissertation. Amsterdam: University of Amsterdam.

Boothroyd, A. (1968). Statistical theory of the speech discrimination score. *Journal of the Acoustical Society of America,* 43*,* 362-367.

Bond, Z.S., and Moore, T.J. (1994). A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech Communication,* 14(4), 325-337.

Bradlow, A.R., Torretta, G.M., and Pisoni, D.B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication,* 20(3-4), 255-272.

Bregman, A.S. (1990). *Auditory Scene Analysis*. Cambridge (MA): MIT Press.

Brighouse, G., and Koh, S.D. (1950). The time error in aesthetic judgment. American Psychologist, 5, 317.

Brodgen, W.J., and Miller, G.A. (1947). Physiological noise generated under earphone cushions. *Journal of the Acoustical Society of America,* 19*,* 620-623.

Bronkhorst, A. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple talker conditions. *Acustica,* 86*,* 117-126.

Brungart, D.S. (1999). Auditory parallax effects in the HRTF for nearby sources. Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. New Paltz, NY: IEEE

Brungart, D.S. (2001). Evaluation of speech intelligibility with the coordinate response measure. *Journal of the Acoustical Society of America,* 109*,* 2276-2279.

Brungart, D.S., and Simpson, B.D. (2002). The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal. *Journal of the Acoustical Society of America,* 112*,* 664-676.

Brungart, D.S., and Scott, K.R. (2001). The effects of production and presentation level on the auditory distance perception of speech. *Journal of the Acoustical Society of America,* 110*,* 425-440.

Brunt, M. (1985). Bekesy audiometry and loudness balance testing. In: Katz, J. (Ed.), *Handbook of Clinical Audiology*. Baltimore, MD: Williams and Wilkins.

Bürck, W., Kotowski, P., and Lichte, H. (1936). Frequenzspektrum und Tonnerkennen. *Annalen der Physik,* 25*,* 433-449.

Buus, S., Schorer, E., Florentine, M., and Zwicker, E. (1986). Decision rules in detection of simple and complex tones. *Journal of the Acoustical Society of America,* 80*,* 1646-1657.

Canévet, G., Germain, R., Marchioni, A., and Scharf, B. (1981). Adaptation de sonie. *Acustica,* 49*,* 239-244.

Carney, L.H. (1994). Spatiotemporal encoding of sound level: Models for normal encoding and recruitment of loudness. *Hearing Research,* 76*,* 31-44.

Cassidy, J., and Ditty, K. (2001). Gender differences among newborns on a transient otoacoustic emissions test for hearing. *Journal of Music Therapy,* 37*,* 28-35.

Chi, S.-M., and Oh, Y.-H. (1996). *Lombard effect compensation and noise suppression for noisy Lombard speech recognition.* Paper presented at *The Fourth International Conference on Spoken Language Processing*, Philadelphia.

Ching, T.Y.C., Dillon, H., and Byrne, D. (1998). Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification. *The Journal of the Acoustical Society of America,* 103(2), 1128-1140.

Choo, T.H. (1954). A study of time error in aesthetic judgment of auditory stimuli. Studies in Psychology EWHA Women's University, 1, 67-83.

Chocholle, R., and Krutel, J. (1968). Les seuils auditifs différentiels d'intensité en fonction de la durée des stimuli. *Comptes Rendus de la Société de Biologie,* 162*,* 848-885.

Clark, D. (1987). Listening test technology for automotive sound systems. SAE 1987 International Conference and Exhibition, Paper 870145. Detroit, MI: Society of Automotive Engineers (SAE).

Clifton, R.K. (1987) Breakdown of echo suppression in the precedence effect. *Journal of the Acoustical Society of America*, 82, 1834-1835.

Colburn, H.S. (1977). Theory of binaural interaction based on auditory-nerve data. II : Detection of tones in noise. *Journal of the Acoustical Society of America,* 61*,* 525-533.

Conway, A.R.A., Cowan, N., and Bunting, M.F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin and Review,* 8*,* 331-335.

Cord, M.T., Walden, B.E., and Atack, R.M. (1992). Speech Recognition in Noise Test (SPRINT) for H-3 Profiles. Unpublished report. (Available from authors, Army Audiology and Speech Center, Walter Reed Army Medical Center, Washington, D.C. 20307-5001)

Corso, J. (19573). Age and sex differences in thresholds. *Journal of the Acoustical Society of America,* 31*,* 498-507.

Corso, J.H. (1963). Aging and auditory thresholds in men and women. *Archives of Environmental Health,* 6, 350-356.

Cox, R.M., Alexander, G.C., and Gilmore, C. (1987). Development of the Connected Speech Test (CST). *Ear and Hearing,* 8*,* 119S-126S.

Culling, J.F., Hawley, M.L., and Litovsky, R.Y. (2004). The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources. *Journal of the Acoustical Society of America,* 116*,* 1057-1065.

Dau, T. (1996). Modeling auditory processing of amplitude modulation. Ph.D. dissertation. Oldenburg (Germany): Carl von Ossietzky Universität Oldenburg.

Deary, I.J., Head, B., and Egan, V. (1989). Auditory inspection time, intelligence, and pitch discrimination. *Intelligence,* 13*,* 135-147.

Delany, M.E. (1970). *On the stability of auditory threshold. NPL Aero Report Ac44*. London: National Physical Laboratory.

Denenberg, L.J., and Altshuler, M.W. (1976). The clinical relationship between acoustic reflex and loudness perception. *Ear and Hearing,* 2*,* 79-82.

Department of Defense. (1999). Department of Defense design criteria standard: Human engineering. MIL-STD-1472F. Washington, DC: U.S. Department of Defense.

Department of the Army. (2008). Army Regulation (AR 40-501 - 2008). Standards of Medical Fitness.

Dirks, D., and Bower, D. (1968). Masking effect of speech competing messages. *Journal of Speech and Hearing Research,* 12*,* 229-245.

Dirks, D., and Kamm, C. (1976). Psychometric functions for loudness discomfort and most comfortable loudness level. *Journal of Speech, Language, and Hearing Research,* 19*,* 613-627

Divenyi, P.L. (1992). Binaural suppression of nonechoes. *Journal of the Acoustical Society of America*, 91, 1078-1084.

Dobie, R.A., and Van Hemel, S. (2004). *Hearing Loss.* Washington, D.C.: National Academies Press.

Doughty, J.M., and Garner, W.R. (1947) Pitch characteristics of short tones. II: Pitch as a function of duration. *Journal of Experimental Psychology,* 38*,* 478-494.

Dufour, A. (1999). Importance if atttentional mechanisms in audiovisual links. *Experimental Brain Research*, 126, 215-222.

Durlach, N. (1963) Equalization and cancellation theory of binaural masking level differences. *Journal of the Acoustical Society of America*, 35*,* 1206-1218.

Durlach, N., and Colburn, H.S. (1978). Binaural phenomena. In: Carterette, E.C., and Freidman, M.P. (Eds.), *Handbook of Perception*. New York: Academic Press.

Durlach, N.I., Mason, C.R., Kidd, G. Jr., Arbogast, T., Colburn, H.S., and Shinn-Cunningham, B.G. (2003). Note on informational masking (L). *Journal of the Acoustical Society of America*, 113*,* 2984-2987.

Durlach, N.I., Mason, C.R., Gallun, F.J., Shinn-Cunningham, B.G., Colburn, H.S., and Kidd, G., Jr. (2005). Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers. *Journal of the Acoustical Society of America,* 118*,* 2482-2497.

Egan, J.P. (1948). Articulation testing methods. *Laryngoscope,* 58(9), 955-991.

Egan, J.P. (1965). Masking-level differences as a function of interaural disparity in intensity of signal and of noise. *Journal of the Acoustical Society of America,* 38*,* 1043-1049.

Egan, J.P., and Hake, H.W. (1950). O the masking pattern of simple auditory stimulus. *Journal of the Acoustical Society of America,* 22*,* 622-630.

Eisler, H. (1966). Measurements of perceived acoustic quality of sound-reproducing systems by means of factor analysis. *Journal of the Acoustical Society of America,* 39*,* 484-492.

Elliott, L.L. (1962). Backward and forward masking of probe tomes of different frequencies. *Journal of the Acoustical Society of America,* 34*,* 1116-1117.

Elliott, L.L., and Katz, D.R. (1980). *Northwestern University children's perception of speech (NU-CHIPS)*. St. Louis, MO: Auditec of St. Louis.

Emanuel, D., Letowski, S., and Letowski, T. (2009). The decibel. In: Emanuel, D., and Letowski, T., *Hearing Science*. Baltimore, MD: Lippincott, Williams, and Wilkins.

Erber, N.P. (1980). Use of the auditory numbers test to evaluate speech perception abilities of hearing-impaired children. *Journal of Speech and Hearing Disorders,* 45*,* 527-532.

Erber, N.P., and Witt, L. H. (1977). Effects of stimulus intensity on speech perception by deaf children. *Journal of Speech and Hearing Disorders,* 42*,* 271-278.

Exner, S. (1875). Experimentelle Untersuchung der einfachsten psychischen Processe. *Pflüngers Archiv für de gesamte Physiologie,* 11*,* 403-432.

Fairbanks, G. (1958). Test of phonemic differentiation: The Rhyme Test. *Journal of the Acoustical Society of America,* 30*,* 596-600.

Fan, W.L., Streeter, T.M., and Durlach, N.I. (2008). Effect of spatial uncertainty of masker on masked detection for nonspeech stimuli (L). *Journal of the Acoustical Society of America,* 124*,* 36-39.

Fastl, H. (1976). Temporal masking effects: I. Broad band noise masker, *Acustica,* 35*,* 287-302.

Fastl, H., and Bechly, M. (1983). Suppression in simultaneous masking. *Journal of the Acoustical Society of America,* 74, 754-757.

Fastl, H., and Stoll, G. (1979). Scaling of pitch strength. *Hearing Research,* 1*,* 293-301.

Fay, D.D. (1988). *Hearing in Vertebrates: A Psychophysics Databook*. Winnetka (IL): Hill-Fay Associates.

Fechner, G.T. (1860). *Elemente der Psychophysik.* Leipzig (Germany) Breitkopf und Härtel. [Elements of Psychophysics (Vol. 1, Adler, H.E., translatior). New York: Holt, Rinehart, and Winston (1966)].

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

Feeney, M.P., and Franks, J.R. (1982). Test-retest reliability of a distinctive feature difference test for hearing aid evaluation. *Ear and Hearing,* 3*,* 59-65.

Feldtkeller, J. and Zwicker, E. (1956). *Das Ohr als Nachrichtenempfanger.* Stuttgart, Germany: Hirzel.

Feuerstein, J. (2002). Occupational hearing conservation. In: Katz, J., Brukard, R., and Medwetsky, L. (Eds.), *Handbook of Clinical Audiology* (5th Ed.). Baltimore: Lippincott Williams and Wilkins.

Fidell, S., and Bishop, D.E. (1974). Prediction of acoustic detectability Technical Report 11949. Warren (MI): U.S. Army Tank Automotive Command.

Fifer, R.C., Jerger, J.F., Berlin, C.I., Tobey, E.A., and Campbell, J.C. (1983). Development of a dichotic sentence identification test for hearing-impaired adults. *Ear and Hearing,* 4*,* 300-305.

Finitzo-Hieber T., Gerling, I.J., Matkin, N.D., and Cherow-Skalka, E. (1980). A sound effects recognition test for the pediatric audiological evaluation. *Ear and Hearing,* 1*,* 271-276.

Fishman, Y., Bolkov, I.O., Noh, M.D., Garell, P.C., Bakken, H., Arezzo, J.C., Howard, M.A., and Steinschneider, M. (2001). Consonance and dissonance of musical chords: natural correlates in auditory cortex of monkeys and humans. *Journal of Neurophysiology,* 86*,* 2761-1788.

Fitzgibons, P.J., and Gordon-Salant, S. (1998). Auditory temporal order perception in younger and older adults. *Journal of Speech and Hearing Science*, 41, 1052-1060.

Flanagan, J.L., and Guttman, N. (1960). On the pitch of periodic pulses. *Journal of the Acoustical Society of America,* 32*,* 1308-1319.

Fletcher, H. (1934). Loudness, pitch, and the timbre of musical tones and their relation to the intensity, the frequency, and the overtone structure. *Journal of the Acoustical Society of America,* 6, 59-69.

Fletcher, H. (1938). Loudness, masking and their relation to the hearing process and the problem of noise measurement. *Journal of the Acoustical Society of America*, 9, 275-293.

Fletcher, H. (1940). Auditory patterns. *Reviews of Modern Physics,* 12*,* 47-65.

Fletcher, H., and Munson, A. (1933). Loudness, its definition, measurement and calculation. *Journal of the Acoustical Society of America,* 5*,* 82-108.

Fletcher, H., and Munson, W.A. (1937). Relation between loudness and masking. *Journal of the Acoustical Society of America,* 19*,* 90-119.

Florentine, M., and Buus, S. (1984). Temporal gaps detection in sensorineural and simulated hearing impairment. *Journal of Speech and Hearing Research,* 27*,* 449-455.

Fuitt, K., Letowski, T., and Mermagen, T. (2003). Auditory performance in the open sound field. *Journal of the Acoustical Society of America,* 113*,* 2286.

Fraisse, P. (1982). Rhythm and tempo. In: Deutsch, D. (Ed.), *The Psychology of Music*. New York: Academic Press.

French, N.R., and Steinberg, J.C. (1947). Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America,* 19*,* 90-119.

Freyman, R.L., Clifton, R.K., and Litovsky, R.Y. (1991). Dynamic processes in the precedence effect. *Journal of the Acoustical Society of America,* 90*,* 874-884.

Fry, D.B. (1961). Word and sentence tests for use in speech audiometry. *Lancet,* 2*,* 197-199.

Fry, D.B., and Kerridge, P.M.T. (1939). *Tests for the hearing of speech by deaf people*. London: H.K. Lewis and Co.

Fryer, P.A., and Lee, R. (1980). Absolute listening tests – further progress. The 65th Audio Engineering Convention, preprint A-2. London: Audio Engineering Society (AES).

Gabrielsson, A. (1974). Performance of rhythm patterns. *Scandinavian Journal of Psychology,* 15*,* 63-72.

Gabrielsson, A., and Lindström, B. (1985). Perceived sound quality of high fidelity loudspeakers. *Journal of Audio Engineering Society,* 33, 33-53.

Gabrielsson, A., and Sjögren H. (1976). Preferred listening levels and perceived sound quality at different sound levels in "high-fidelity" sound reproduction. Report TA-82. Stockholm (Sweden): Karolinska Institutet.

Gabrielsson, A., and Sjögren H. (1979). Perceived sound quality of sound-reproducing systems. *Journal of the Acoustical Society of America*, 65, 1919-1933.

Gaeth, J.H. (1970). *A scale for testing speech discrimination: Final report, project no. RD-2277-S*. Washington: Division of Research and Demonstration Grants, U.S. Social and Rehabilitation Service.

Galilei, G. (1638). Dialogues concerning two new sciences. Translated by H. Crewe and A. de Salvio. New York: McGraw-Hill, 1963]

Galton, F. (1883). *Inquires into human faculty and its development*. London: J.M. Dent and Sons.

Gardner, H.J. (1971). Application of a high-frequency consonant discrimination word list in hearing-aid evaluation. *Journal of Speech and Hearing Disorders,* 36*,* 354-355.

Gardner, M.B. (1964). Effect of noise on listening levels in conference telephony. *Journal of the Acoustical Society of America,* 36, 2354-2362.

Gardner, M.B. (1964). Study of noise + other factors in conference telephony. *Journal of the Acoustical Society of America,* 36, 1036.

Gardner, M.B., and Gardner, R.S. (1973). Problem of localization in the median plane: effect of pinnae cavity occlusion, *Journal of the Acoustic Society of America,* 53, 400-408.

Garner, W.R., and Miller, G.A. (1947a). The masked threshold for pure tones as a function of duration. *Journal of the Acoustical Society of America,* 37*,* 293-303.

Garner, W.R., and Miller, G.A. (1947b). Differential sensitivity to intensity as a function of the duration of the comparison tone. *Journal of Experimental Psychology*, 34, 450-463.

Garstecki, D.C., and Bode, D.L. (1976). Aided and unaided narrow band noise thresholds in listeners with sensorineural hearing impairment. *Ear and Hearing*, 1, 258-262.

Gavrenau, V. (1966). Infra sons: Générateurs, détecteurs, propriétés physiques, effects biologiques. *Acustica*, 17, 1-10.

Gavrenau, V. (1968). Infrasound. *Science Journal*, 4, 33.

Gässler, G. (1954). Über die Hörschwelle für Schallereignisse mit verschieden breitem Frequentzspektrum. *Acustica,* 4*,* 408-414.

Glasberg, B.R., and Moore, B.C.J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research,* 47*,* 103-138.

Gockel, H., Moore, B.C.J., Carlyon, R.P., and Plack, C.J. (2007). Effect of duration on the frequency discrimination of individual partials in a complex tone and on the discrimination of fundamental frequency. *Journal of the Acoustical Society of America,* 121*,* 373-382.

Goldstein, J.L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Acoustical Society of America,* 54*,* 1496-1516.

Goldstein, M. (1995). Classification of methods used for assessment of text-to-speech systems according to the demands placed on the listener. *Speech Communication*, 16, 225-244.

Grau, J.W., and Kemler Nelson, D.G. (1988). The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General,* 117*,* 347-370.

Green, D.M. (1988). *Profile Analysis: Auditory Intensity Discrimination*. Oxford: Oxford University Press.

Greenwood, D.D. (1961a). Auditory masking and the critical band. *Journal of the Acoustical Society of America,* 33*,* 484-502.

Greenwood, D.D. (1961b) Critical bandwidth and frequency coordinates of the basilar membrane. *Journal of the Acoustical Society of America,* 33*,* 1344-1356.

Greenwood, D.D. (1990). A cochlear frequency-position function for several species – 29 years later. *Journal of the Acoustical Society of America,* 87*,* 2592-2605.

Griffiths, J.D. (1967). Rhyming minimal contrasts: a simplified diagnostic articulation test. *Journal of the Acoustical Society of America,* 42*,* 236-241.

Groben, L.M. (1971). Appreciation of short tones. *Proceedings of the 7th ICA Congress on Acoustics, Paper 19-H-6*. Budapest (Hungary): International Commission on Acoustics (ICA).

Gullick, W.L. (1971). *Hearing: Physiology and Psychophysics*. New York: Oxford University Press.

Gullick, W.L., Gescheider, G.A., and Frisina, R.D. (1989). *Hearing*. New York: Oxford University Press.

Guirao, M., and Stevens, S.S. (1964). Measurement of auditory density. *Journal of the Acoustical Society of America,* 36*,* 1176-1182.

Gygi, B., Kidd, G.R., and Watson, C.S. (2004). Spectral-temporal factors in the identification of environmental sounds. *Journal of the Acoustical Society of America,* 115*,* 1252-1265.

Gygi, B. (2001). Factors in the identification of environmental sounds. Ph.D. dissertation, Department of Psychology, Indiana University.

Gygi, B., and Shafiro, V. (2007). General functions and specific applications of environmental sound research. *Frontiers in Bioscience,* 12*,* 3152-3166.

Gygi, B., Kidd, G.R., and Watson, C.S. (2007). Similarity and categorization of environmental sounds. *Perception and Psychophysics,* 69*,* 839-855.

Haas, H. (1951). Über den Einfluss eines Einfachechos an die Hörsamkeit von Sprache. *Acustica,* 1, 49-58.

Hall, E.T. (1966). *Then Hidden Dimension*. New York: Anchor Books.

Hamilton, P.M. (1957). Noise masked threshold as a function of tonal duration and masking tome band width. *Journal of the Acoustical Society of America,* 29*,* 506-511.

Harrell, R. (2002). Pure tone evaluation. In: Katz, J., Burkard, R., and Medwetsky, L. (Eds.), *Handbook of Clinical Audiology* (5th Ed.). Baltimore, MD: Lippincott Williams and Wilkins.

Hartmann, W.M. (1997). *Signals, Sound, and Sensation*. Woodbury, NY: American Institute of Physics.

Hartmann, W.M., and Rakerd, B. (1989). Localization of sound in rooms. IV: The Franssen effect. *The Journal of the Acoustical Society of America.* 86*,* 1366-1373.

Haskins, H.L. (1949). A phonetically balanced test of speech discrimination for children. Unpublished master's thesis, Northwestern University, Evanston, IL.

Hawkes, R.J., and Douglas, H. (1971). Subjective acoustic experience in concert auditoria. *Acustica,* 24*,* 235-240.

Hawkins, J.E., and Stevens, S.S. (1950). The masking of pure tones and of speech by white noise. *Journal of the Acoustical Society of America,* 22*,* 6-13.

Hawley, M.L., Litovsky, R.Y., and Culling, J.F. (2004). The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. *Journal of the Acoustical Society of America,* 115*,* 833-843.

Hebrank, J., and Wright, D. (1974). Spectral cues used in the localization of sound sources on the median plane, *Journal of the Acoustical Society of America,* 56*,* 1829-1834.

Heffner, H.E., and Heffner, R.S. (1998). Hearing. In: Greenberg, G., and Haraway, M.M. (Eds.), *A Handbook of Comparative Psychology*. New York: Garland.

Heffner, H.E., and Heffner, R.S. (2003). Audition. In: Davis, S. (Ed.), *Handbook of Research Methods in Experimental Psychology* (pp. 413-440). Hoboken, NJ: Blackwell Publishing.

Heffner, R.S., and Heffner, H.E. (1993). Degenerate hearing and sound localization in naked mole rats (Heterocephalus glaber) with an overview of central auditory structures. *Journal of Comparative Neurology,* 3321*,* 418-433.

Hellman, R. (1991). Loudness measurements by magnitude scaling: Implication for intensity coding. In: Bolanowski, S.J., and Gescheider, G.A. (Eds.), *Ratio Scaling of Psychological Magnitude*. Hillsdale, NJ: Erlbaum.

Hellman, R., and Zwislocki, J. (1961). Some factors affecting the estimation of loudness. *Journal of the Acoustical Society of America,* 60*,* 672-679.

Hellman, R., Miśkiewicz, A., and Scharf, B. (1997). Loudness adaptation and excitation patterns: Effects of frequency and level. *Journal of the Acoustical Society of America,* 101*,* 2176-2185.

Hellström, A. (1977). Time errors are perceptual. *Psychological Research,* 39*,* 345-388.

Helmholtz, H von. (1863). *Der Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. Braunschweig (Germany):Fredrich Viewwieg und Sohn. [On the Sensations of Tone as a Physiological Bases for the Theory of Music (Edited and translated by A. J. Ellis). New York: Dover, 1954].

Henning, G.B. (1974). Lateralization and binaural masking-level difference. *Journal of the Acoustical Society of America,* 55(6), 1259-1262.

Henning, G.B., and Bleiwas, S.L. (1967). Amplitude discrimination in noise. *Journal of the Acoustical Society of America,* 41*,* 1365-1366.

Henning, G.B., and Grosberg, S.L. (1968). Effect of harmonic components on frequency discrimination. *Journal of the Acoustical Society of America,* 44*,* 1386-1389.

Henry, J.A., Flick, C.L., Gilberty, A., Ellingson, R.M., and Fausti, S.A. (2001). Reliability of hearing thresholds: Computer-automated testing with ER-4B Canal Phone™ earphones. *Journal of Rehabilitation Research and Development,* 38*,* 1-18.

Henry, P., and Letowski, T. (2007). Bone conduction: Anatomy, physiology and communication. Aberdeen Proving Ground, MD: U.S. Army Research Laboratory. ARL Technical Report ARL-TR-4138.

Hirsh, I. (1948). The influence of interaural phase on interaural summation and inhibition. *Journal of the Acoustical Society of America,* 20*,* 536-544.

Hirsh, I. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America,* 31*,* 759-767.

Hirsh, I.J. (1971). Masking of speech and auditory localization. *Audiology,* 10, 110-114.

Hirsh, I.J., Davis, H., Silverman, S.R., Reynolds, E.G., Eldert, E., and Benson, R.W. (1952). Development of materials for speech audiometry. *Journal of Speech and Hearing Disorders,* 17*,* 321-337.

Hirsh. I., and Sherrick, C.E., Jr. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology,* 62*,* 423-432.

Hochberg, I. (1975). Most comfortable listening for the loudness and intelligibility of speech. *Audiology,* 14, 27-33.

Hood, J.C. (1950). Studies on auditory fatigue and adaptation, *Acta Otolaryngologica*, (Suppl.) 92, 1-57.

Hood, J.D., and Poole, J.P. (1980). Influence of the speaker and other factors affecting speech-intelligibility *Audiology*, 19(5), 434-455.

Hochhaus, L., and Antes, J.R. (1973). Speech identification and knowing that you know. *Perception and Psychophysics,* 13(1A), 131-132.

House, A.S., Williams, C.E., Hecker, M.H.L., and Kryter, K.D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. *Journal of the Acoustical Society of America,* 37*,* 158-166.

Houtgast, T. (1974). Lateral suppression and loudness reduction of a tone in noise. *Acustica,* 30*,* 215-221.

Howes, W.L. (1971). Loudness determined by power summation. *Acustica,* 25*,* 343-248.

Hudgins, C.V., Hawkins, J.E., Karlin, J.E., and Stevens, S.S. (1947). The development of recorded auditory tests for measuring hearing loss for speech. *Laryngoscope,* 57(1), 57-89.

Illényi, A., and Korpassy, P. (1981). Correlation between loudness and quality of stereophonic loudspeakers. *Acustica,* 49*,* 334-336.

International Electrotechnical Commission. (1995). IEC 50801, International Electrotechnical Vocabulary – Chapter 801: Acoustics and Electroacoustics. Geneva: International Electrotechnical Commission.

International Organization for Standardization. (1975). ISO 532, Acoustics – Method for calculating loudness level. Geneva: International Organization for Standardization (ISO).

International Organization for Standardization. (2000). ISO 7029, Acoustics – Statistical distribution of hearing thresholds as a function of age. Geneva: International Organization for Standardization (ISO).

International Organization for Standardization. (2003). ISO 226 (R2003), Acoustics - Normal equal-loudness level contours. Geneva: International Organization for Standardization (ISO).

International Organization for Standardization**.** (2005). ISO 389-7, Acoustics - Reference zero for the calibration of audiometric equipment - Part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions. Geneva: International Organization for Standardization (ISO).

International Organization for Standardization. (2006). ISO 16832, Acoustics - Loudness scaling by means of categories. Geneva: International Organization for Standardization (ISO).

International Telecommunication Union - Telecommunication. (1996). ITU-T Recommendation P.800. Methods for subjective determination of transmission quality. Geneva: International Telecommunication Union – Telecommunication (ITU-T).

Jaroszewski, A., and Rakowski, A. (1976). Pitch shifts in post-stimulatory masking. *Acustica,* 34*,* 220-223.

Jerger, S., Lewis, S., Hawkins, J. and Jerger, J. (1980). Pediatric speech intelligibility test. I. Generation of test materials. *International Journal of Pediatric Otorhinolaryngology,* 2*,* 217-30.

Jesteadt, W., Bacon, S.P., and Lehman, J.R. (1982). Forward masking as a function of frequency, masker level, and signal delay. *Journal of the Acoustical Society of America,* 71*,* 950-962.

Junqua, J. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers, *Journal of the Acoustical Society of America,* 93*,* 510-524.

Junqua, J. (1996). The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex, *Speech Communication,* 20*,* 13-22.

Justus, T.C., and Bharucha, J.J. (2002). Music perception and cognition. In: Yantis, S., and Pashler, H. (Eds.), *Stevens' Handbook of Experimental Psychology*, *Sensation and Perception*, I, 453-492.

Kaernbach, C., and Bering, C. (2001). Exploring the temporal mechanism involved in the pitch of unresolved harmonics. *Journal of the Acoustical Society of America,* 110*,* 1039-1048.

Kalikow, D.N., Stevens, K.N., and Elliott, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America,* 61*,* 1337-1351.

Kant, I. (1781). Kritik der reinen Vernunft. Riga: J.F. Hartknoch [Critique of Pure Reason. English translation: Werner Pluhar. Indianopolis: Hackett, 1996].

Katz, D.R., Elliott, L.L. (1978). Development of a new children's speech discrimination test. Paper presented at the 1978 *Annual American Speech-Language-Hearing Association Convention*. Chicago, IL.

Keating, P.A., and Manis, F. (11998). The Keating-Manis phoneme deletion test. *UCLA Working Papers in Phonetics,* 96*,* 162-165.

Keith, R.W. (1977). An evaluation of predicting hearing loss from the acoustic reflex. Archives of Otolaryngology, 103, 419-424.

Keith, R.W., Young, M., and McCroskey, R. (1999). A brief introduction to the Auditory Fusion Test – Revised. *Educational Audiology Review*, 16, 2.

Kidd, G. Jr., Mason, C.R., Deliwala, P.S., Woods, W.S., and Colburn, H.S. (1994). Reducing informational masking by sound segregation. *Journal of the Acoustical Society of America*, 95*,* 3475-3480.

Kidd, G. Jr., Mason, C.R., Richards, V.M., Gallun, F.J., and Durlach, N.I. (2007). Informational masking. In: Yost, W. (Ed.), *Springer Handbook of Auditory Research*, 29, 143-190. New York: Springer Verlag.

Killion, M.C. (1978). Revised estimate of minimum audible pressure: Where is the "missing 6 dB"? *Journal of the Acoustical Society of America,* 63*,* 1501-1508.

Kobayashi, M., Morimoto, M., Sato, H., and Sato, H. (2007). Optimum speech level to minimize listening difficulty in public spaces. *Journal of the Acoustical Society of America,* 121*,* 251-256.

Koester, T. (1945). Time-error and sensitivity in pitch and loudness discrimination as a function of time interval and stimulus level. *Archives of Psychology,* 297*,* 69.

Koester, T., and Schoenfeld W.N. (1946). The effect of context upon judgment of pitch differences. *Journal of Experimental Psychology,* 36*,* 417-430.

Koh, S.D. (1962). The Fechnerian time error and Pratt's time error in affective judgments of musical excerpts. *American Psychologist*, 17*,* 369.

Koh, S.D. (1967). Time-error in comparison of preferences for musical excerpts. *American Journal of Psychology,* 80*,* 171.

Kohler, W. (1923). Zur theorie des sukzessivvergleichs und der zeitfehler. *Psychologische Forshung,* 4*,* 115-175.

König, E. (1957). Effect of time on pitch discrimination thresholds under several psychophysical procedures. *Journal of the Acoustical Society of America,* 29*,* 606-612.

Kopčo, N., and Shinn-Cunningham, B.G. (2008). Influences of modulation and spatial separation on detection of a masked broadband target. Journal *of the Acoustical Society of America,* 124*,* 2236-2250.

Kopra, L.L., and Blosser. D. (1968). Effects of method of measurement on most comfortable loudness level for speech. *Journal of Speech and Hearing Research,* 11*,* 497-508.

Krause, B. (2008). Interview: From rock'n'roll to the sounds of nature. *New Scientist,* 2664, 42-43

Kryter, K.D. (1962a). Methods for the calculation and use of the Articulation Index, *Journal of the Acoustical Society of America*, 34, 1689-1697.

Kryter, K.D. (1962b). Validation of the articulation index. *Journal of the Acoustical Society of America,* 34*,* 1698-1702.

Kryter, K.D. (1965). Some comparisons between rhyme and PB-word intelligibility test. *Journal of the Acoustical Society of America,* 37*,* 1146.

Kuriyagawa, M., and Kameoka, A. (1966). PSE tracing method for subjective harmonics measurement and monaural phase effect on timbre. *Journal of the Acoustical Society of America,* 39, 1263-1263 (A).

Lamore, J.J. (1975). Perception of two-tone octave complexes. *Acustica*, 34*,* 1-14.

Lane, H.L., and Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research,* 14*,* 677-709.

Leakey, D.M., Sayers, B.M., Cherry, C. (1958). Binaural fusion of low-frequency and high-frequency sounds. *Journal of the Acoustical Society of America,* 30(3), 222-223.

Lenhardt, M., Skellett, R., Wang, P., and Clarke, A. (1991). Human ultrasonic speech perception. *Science*, 253, 82-85.

Leshowitz, B. (1971). Measurement of the two-click threshold. *Journal of the Acoustical Society of America,* 49*,* 462-466.

Letowski, T. (1982). A note on the differential limen for frequency differentiation, *Journal of Sound and Vibration*, 85, 579-583.

Letowski, T. (1984). Auditory assessment of acoustic signals and audio equipment. *Chopin Academy Scientific Series,* 8, 1-401. (Monograph).

Letowski, T. (1985). Development of technical listening skills: Timbre Solfeggio. *Journal of Audio Engineering Society,* 33*,* 240-244.

Letowski, T. (1989). Sound quality assessment: Concepts and criteria. Paper D-8 presented at the 87th Audio Engineering Convention (Preprint # 2825). New York: Audio Engineering Society (AES).

Letowski, T. (1992). Timbre, tone color, and sound quality: Concepts and definitions. *Archives of Acoustics*, 17, 17-31.

Letowski, T. (1994). Guidelines for conducting listening tests on sound quality. Proceedings of the Noise-Con 1994 Conference. Fort Lauderdale, FL: International Institute of Sound and Vibration.

Letowski, T. (1995). Sound quality scales and systems. *Proceedings of the VI Symposium on Tonmeistering and Sound Engineering*. Warszawa, Poland: Polskie Radio S.A.

Letowski, T., and Amrein, K. (2005). The effects of auditory training on the listener's ability to detect and recognize sounds in noise. *Proceedings of the 12th International Congress on Sound and Vibration (ICSV).* Lisbon, Portugal: International Institute of Acoustics and Vibration (IIAV).

Letowski, T., and Dreisbach L. (1992). Pleasantness and unpleasantness of speech. Proceedings of the 11th International AES Conference, pp. 159-165. Portland (OR): Audio Engineering Society (AES).

Letowski, T., Frank T., and Caravella J. (1993). Acoustical properties of speech produced in noise. *Ear and Hearing,* 14(5), 332-338.

Letowski, T., and Makowski, W. (1977). Właściwości obrazów słuchowych [Properties of auditory images]. *Technika Radia i TV*, 1, 16-21.

Letowski, T., and Miskiewicz, A. (1995). Development of technical listening skills for sound quality assessment. Proceedings of the Inter-Noise 1995. Newport Beach, CA: Inter-Noise.

Letowski, T., and Rakowski A. (1971). The DL for loudness of musical sounds. *Proceedings of the VII ICA Congress*, 19-H-5, 325-328. Budapest, Hungary: International Commission on Acoustics (ICA).

Letowski, T., and Smurzyński, J. (1980). Time error in perception of sound brightness. *Archives of Acoustics,* 5, 143-146.

Letowski, T., and Smurzynski, J. (1983). Evaluation of electroacoustic devices by equivalent scale methods. *Archives of Acoustics,* 8, 133-154.

Letowski, T., Tornatore, A., Clark, J., and MacBeth, B. (1993). Word discrimination in various multitalker noises. *Journal of the Acoustical Society of America,* 94(4/Pt.2), 1778

Letowski, T., Tornatore, A., MacBeth, B., Salter, S., and Smeal, M. (2001). Acoustic and perceptual properties of several multitalker noises. Proceedings of the 8th International Congress on Sound and Vibration, p. 2467-2474 [Hong Kong: July 2-6, 2001], Hong Kong: The Hong Kong Polytechnic University.

Levitt, H., and Webster, J.C. (1997). Effects of noise and reverberation on speech. In: Harris, C.M. (Ed.), *Handbook of Acoustical Measurements and Noise Control*, New York: Acoustical Society of America.

Liang, C.A., and Christvich, L.A. (1961). Frequency difference limens as a function of tonal duration. *Soviet Physical Acoustics,* 6*,* 75-80.

Liberman, A., Coper, F.S., Shankweiler, D., and Studert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review,* 74*,* 431-461.

Lichte, W.H. (1941). Attributes of complex tones. *Journal of Experimental Psychology,* 28, 455-481.

Licklider, J.C.R. (1948). The influence of interaural phase relations upon the masking of speech by white noise. *Journal of the Acoustical Society of America,* 20*,* 150-159

Licklider, J.C.R. (1951). Basic correlates of the auditory stimulus. In: Stevens, S.S. (Ed.), *Handbook of Experimental Psychology*. New York: Wiley and Sons.

Licklider, J.C.R. (1954). "Periodicity" pitch and "place" pitch. *Journal of the Acoustical Society of America*, 26, 945.

Ling, D., and Ling, A. (1978). Aural habilitation: The verbal foundations of learning in hearing-impaired children. Washington, DC: A.G. Bell.

Little, A.D., Mershon, D.H., and Cox, P.H. (1992). Spectral content as a cue to perceived auditory distance. *Perception,* 21, 405-416.

Lombard, E. (1911). Le signe de l'élévation del la voix. *Annales del Maladies de l'Oreille et du Larynx,* 37, 101-119.

Lonsbury-Martin, B., and Martin, K. (1993). Auditory dysfunction from excessive sound stimulation. In: Cummings, C., and Fredrickson, J. (Eds.), *Otolaryngology-Head and Neck Surgery*, 4, 2885-2990. St. Louis: Mosby Year Book.

Lubman, D. (1992). Objective metrics for characterizing automotive interior sound quality. *Proceedings of the Inter-Noise 1992 Conference*, 139, 1067. Toronto: Institute of Noise Control Engineering.

Lufti, R.A. (1986). Two- versus four-tone masking, revisited. *Journal of the Acoustical Society of America,* 80*,* 422-428.

Lufti, R.A. (1990). How much masking is informational masking? *Journal of the Acoustical Society of America,* 88, 2607-2610

Lundeen, C., and Small, A.M. (1984). Te influence of temporal cues on the strength of periodicity pitch. *Journal of the Acoustical Society of America,* 75, 1578-1587.

Makous, J.C., and Middlebrooks, J.C. (1990). Two-dimensional sound localization by human listeners, *Journal of the Acoustical Society of America,* 87*,* 2188-2200.

Marks, L.E. (1978). Binaural summation of the loudness of pure tones. *Journal of the Acoustical Society of America,* 64*,* 107-113.

Martin, M.C., and Grover, B.C. (1976). Current hearing aid performance and the requirements of hearing impaired persons. *Journal of the Audio Engineering Society,* 24*,* 177-181.

Marozeau, J., Epstein, M., Florentine, M., and Daley, B. (2006). The test of the Binaural Equal-Loudness-Ratio hypothesis for tones. *Journal of the Acoustical Society of America,* 120*,* 3870-3877.

Massaro, D. (1975). Perceptual processing time in audition. *Journal of the Acoustical Society of America*, 57*,* S5 (abstract).

Masterton, B., Heffner, H.E., and Ravizza, R. (1969). The evolution of human hearing. *Journal of the Acoustical Society of America,* 45*,* 966-985.

Mauk, M.D., and Buonomano, D.V. (2004). The neural basis of temporal processing. *Annual Review of Neuroscience*, 27, 307-340.

McAdams, S. (1984). Spectral fusion, spectral parsing, and the formation of auditory images. Technical Report STAN-M-22. Stanford, CA: Stanford University.

McBride, M.E, Letowski, T., and Tran, P. (2005). Search for the optimum vibrator location for bone conduction communication. Proceedings of the HFES 49th Annual Meeting, CD ROM, Orlando (FL): September 26-30, 2005.

McBride, M., Letowski, T., and Tran, P. (2008). Bone conduction reception: Head sensitivity mapping. *Ergonomics*, 55, 702-718.

McBride, M., Letowski, T., and Tran, P. (2008). Head mapping: Search for an optimum bone microphone placement. *Proceedings of the 42nd Annual HFES Meeting*. New York (NY): Human Factors Society (HFS).

McDermott, J.B. (1969). Multidimensional analyses of circuit quality judgments. *Journal of the Acoustical Society of America,* 45*,* 774-781.

McDonald, J.J., Teder-Sälejärvi, W.A., DiRusso, F., and Hillyear, S.A. (2005).Neural basis of auditory-induced shifts in visual time-order perception. *Nature Neuroscience (Online)*, 1-6.

McGill, W.J., and Goldberg, J.P. (1968). Pure-Tone Intensity Discrimination and Energy Detection. *The Journal of the Acoustical Society of America,* 44*,* 576-581.

McGregor, P., Horn, A.G., and Todd, M.A. (1985). Are familiar sounds ranged more accurately? *Perceptual and Motor Skills,* 61, 1082.

McPherson, D.F., and Pang-Ching, G.K. (1979). Development of a distinctive feature discrimination test. *Journal of Auditory Research,* 19*,* 235-246.

Mershon, D.H., Ballenger, W.L., Little, A.D., McMurtry, P.L., and Buchanan, J.L. (1989). Effects of room reflectance and background noise on perceived auditory distance. *Perception,* 18, 403-416.

Mershon, D.H., and King, L.E. (1975). Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception and Psychophysics,* 18, 409-415.

Michaels, R.M. (1957). Frequency difference limens for narrow bands of noise. *Journal of the Acoustical Society of America,* 29*,* 520-522

Micheyl, C., Arthaud, P., Reinhart, C., and Collet, P. (2000). Informational masking in normal-hearing and hearing-impaired listeners. Acta Otolarngologica*,* 120*,* 242-246.

Miller, G.A. (1947). Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *Journal of the Acoustical Society of America,* 19, 609-619.

Miller, G.A., and Taylor, G. (1948). The perception of repeated bursts of noise. *Journal of the Acoustical Society of America,* 20*,* 171-182.

Miller, G.A., Heise, G.A., and Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology,* 41(5), 329-335.

Mills, A.W. (1972). Auditory localization. In: Tobias, J.V. (Ed.), *Foundations of Modern Auditory Theory*, II, 301-348. New York: Academic Press.

Miskolczy-Foder, F. (1959). Relation between loudness and duration of tone pulses. *Journal of the Acoustical Society of America,* 31*,* 1128-1134.

Miśkiewicz, A., Scharf, B., Hellman, R., and Meiselman, C. (1992). Loudness adaptation at high frequencies. *Journal of the Acoustical Society of America,* 94*,* 1281-1286.

Mitrinowicz, M.A., and Letowski, T. (1966). Auditory threshold measurements with one-third octave noise bands. *Archiwum Akustyki*, 1, 103-112.

Møller, H., and Pedersen, C.S. Hearing at low and infrasonic frequencies. (2004). *Noise Health,* 6*,* 37-57.

Moore, B.C.J. (1973). Frequency difference limes for short-duration tones. *Journal of the Acoustical Society of America,* 54*,* 610-619.

Moore, B.C.J. (1977). Effects of relative phase of the component on the pitch of three-component complex tones. In: Evans, E.F., and Wilson, J.P. (Eds.), *Psychophysics and Physiology of Hearing*. London: Academic Press.

Moore, B.C.J. (1997). *An introduction to the psychology of hearing* (4th Ed.). San Diego, CA: Academic Press.

Moore, B.C.J., and Glasberg, B.R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America,* 88*,* 750-753.

Moore, B.C.J., and Glasberg, B.R. (1996). Revision of Zwicker's Loudness Model. *Acta Acustica*, 82*,* 335-345.

Moore, B.C.J., and Glasberg, B.R., and Baer, T. (1997). A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Acoustical Society of America,* 45*,* 224-240.

Moore, B.C.J., and Peters, R.W. (1992). Pitch discrimination and phase sensitivity i9n young and elderly subjects and its relationship to frequency selectivity. *Journal of the Acoustical Society of America,* 91*,* 2881-2893.

Morgan, D.E., Dirks, D., Bower, D., and Kamm, C. (1979). Loudness discomfort level and acoustic reflex threshold. *Journal of Speech and Hearing Research,* 22*,* 849-861.

Murphy M.P., and Gates, G.A (1997). Hearing loss: Does gender play a role? *Medscape Women's Health eJournal,* 2*,* 5.

Murphy, W., Themann, C., and Stephenson, M. (2006). Hearing levels in U.S. adults aged 20-69 Years – National Health and Nutrition Examination Survey (NHANES) 1999-2004. *Journal of the Acoustical Society of America,* 119*,* 3268-3268.

Murray, B.A., Smith, K.A., and Murray, G.G. (2000). The test of phoneme identities: Predicting alphabetic insight in prealphabetic readers. *Journal of Literacy Research,* 32, 421-447.

Musicant, A.D., and Butler, R.A. (1985). Influence of monaural spectral cues on binaural localization, *Journal of the Acoustical Society of America,* 77*,* 202-208.

Myers, L., Letowski, T., Abouchacra, K., Haas, E., and Kalb, J. (1996). Detection and recognition of filtered sound effects. *Journal of the American Academy of Audiology,* 7*,* 346-357.

Needham, J.G. (1935). Contrast effect in judgment of auditory intensities. *Journal of Experimental Psychology,* 18*,* 214-226.

Neuhoff, J.G. (2004). Ecological psychoacoustics: Introduction and history. In: J.G. Neuhoff, J.G. (Ed.), *Ecological Psychoacoustics*. Amsterdam: Elsevier.

Neuhoff, J.G., and McBeath, M.K. (1996). The Doppler illusion: The influence of dynamic intensity change on perceived pitch. *Journal of Experimental Psychology-Human Perception and Performance,* 22*,* 970-985.

Nielsen, S.H. (1993). Auditory distance perception in different rooms. *Journal of the Audio Engineering Society,* 41*,* 755-770.

Nilsson, M., Soli, S.D., and Sullivan, J.A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America,* 95*,* 1085-1099.

Nofsinger, D., Martinez, C.D., and Schaefer, A.B. (1985). Puretone techniques in evaluation pf central auditory function. In: Katz, J. (Ed.), *Handbook of Clinical Audiology*, 18, 337-354. Baltimore, MD: Williams and Wilkins.

Nye, P.W., and Gaitenby, J.H. (1974). The intelligibility of synthetic monosyllabic words in short, syntactically normal sentences. Haskins Labs Status Report on Speech Research, 37/38, 169-190.

Oldfield, S.R., and Parker, S.P.A. (1984). Acuity of sound localisation: A topography of auditory space. I. Normal hearing conditions, *Perception*, 13*,* 581-600.

Olson, H.F. (1967). *Music, Physics and Engineering*. Dover: New York

Owens, E., and Schubert, E.D. (1977). Development of the California Consonant Test. *Journal of Speech and Hearing Research,* 20*,* 463-474.

Ostroff, J.M., McDonald, K.L., Schneider, B.S., and Alain, C. (2003). Aging and the processing of sound duration in human auditory cortex. *Hearing Research,* 181*,* 1-7

Palmer, A.R., and Russell, I.J. (1986). Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells, *Hearing Research,* 24*,* 1-15.

Patterson, R.D. (1987). A pulse ribbon model of monaural phase perception. *Journal of the Acoustical Society of America,* 82*,* 1560-1586.

Pearson, K.S., Bennett, R.L., and Fidell, S. (1977). Speech levels in various noise environments. Report No. EPA-600/1-77-025; Washington, DC: U.S. Environmental Protection Agency (EPA).

Pederson, O.T., and Studebaker, G.A. (1972). A new minimal contrasts closed-response-set speech test. *Journal of Auditory Research,* 12*,* 187-195.

Perrott, D.R., and Musicant, A.D. (1977). Minimum auditory movement angle: Binaural localization of moving sound sources. *The Journal of the Acoustical Society of America,* 62*,* 1463-1466.

Perrott, D.R., and Musicant, A.D. (1981). Dynamic minimum audible angle: Binaural spatial acuity with moving sound sources, *Journal of Auditory Research,* 21*,* 287-295.

Peterson, G.E., and Lehiste I. (1962). Revised CNC lists for auditory tests. *Journal of Speech and Hearing Disorders,* 27, 62-70.

Pfafflin, S.M. (1968). Detection of auditory signals in restricted sets of reproducible noise. *Journal of the Acoustical Society of America,* 43*,* 487-490.

Pfafflin, S.M., and Matthews, M.V. (1966). Detection of auditory signals in reproducible noise. *Journal of the Acoustical Society of America,* 39*,* 340-345.

Picheny, M.A., Durlach, N.I., and Braida, L.D. (1985). Speaking clearly for the hard of hearing. 1. Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research,* 28(1), 96-103.

Picheny, M.A., Durlach, N.I., and Braida, L.D. (1986). Speaking clearly for the hard of hearing. 2. Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research,* 29(4), 434-446.

Picheny, M.A., Durlach, N.I., and Braida, L.D. (1989). Speaking clearly for the hard of hearing. 3. An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech and Hearing Research,* 32(3), 600-603.

Pick, J.H.L., Siegel, G.M., Fox, P.W., Garber, S.R., and Kearney, J.K. (1989). Inhibiting the Lombard effect. *The Journal of the Acoustical Society of America,* 85*,* 894-900.

Pickles, J. (1988). *An introduction to the physiology of hearing.* San Diego: Academic Press Inc.

Plenge, G. (1982). Über das Problem der Im-Kopf-Localisation. *Acustica,* 26, 241-252.

Plenge, G., and Brunschen, G. (1971) A priori knowledge of the signal when determining the direction of speech in the median plane. Proceedings of the 7th International Congress on Acoustics, (ICA), Paper19-H-10. Budapest, Hungary: ICA.

Plomp, R. (1964). Rate of decay of auditory sensation. *Journal of the Acoustical Society of Americ*a, 36, 277-282.

Plomp, R. (1970). Timbre as a multidimensional attribute of complex tones, In: Plomp, R., and Smoorenburg, G.F. (Eds.), *Frequency Analysis and Periodicity Detection in Hearing,* Leiden: Sijthoff.

Plomp, R. (1976). *Aspects of Tone Sensation: A Psychophysical Study*. London: Academic.

Plomp, R., and Bouman, M.A. (1959). Relation between hearing threshold and duration for tone pulses. *Journal of the Acoustical Society of America,* 31*,* 749-758.

Plomp, R., and Levelt, W.J.M. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America,* 38*,* 548-560.

Plomp, R., Pols, L.C.W., and van der Geer, J.P. (1967). Dimensional analysis of vowel tones, *Journal of the Acoustical Society of America,* 41*,* 707-712.

Pollack, I. (1948). Monaural and binaural threshold sensitivity for tones and for white noise. *Journal of the Acoustical Society of America*, 20, 52-57.

Pollack, I. (1951). Sensitivity to differences in intensity between repeated bursts of noise. *Journal of the Acoustical Society of America*, 23, 650-653.

Pollack, I. (1952). Loudness of bands of noise. *Journal of the Acoustical Society of America*, 24, 533-538.

Pollack, I. (1954). Intensity discrimination thresholds under several psychophysical procedures. *Journal of the Acoustical Society of America*, 26, 1056-1059.

Pollack, I. (1975). Auditory informational masking. *Journal of the Acoustical Society of America,* 57(Suppl.1), S5.

Pols, L.C.W., van der Kamp, L.J.T. and Plomp, R. (1969). Perceptual and physical space of vowel sounds, *Journal of the Acoustical Society of America,* 46, 458-467.

Postman, L. (1946). The time-error in auditory perception. *American Journal of Psychology*, 53, 193-219.

Pratt, R., and Doak, P. (1976). A subjective rating scale for timbre. *Journal of Sound and Vibrations,* 45, 317-321.

Price, G.R., and Hodge, D.C. (1976). Combat sound detection: Monaural listening in quiet. Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratory. HEL Technical Memorandum HEL-TM-35-76.

Price, G.R., Kalb, J.T., and Garinther, G.R. (1989). Toward a measure of auditory handicap in the Army. *Annals of Otology, Rhinology and Laryngology,* 98, 42-52.

Rabin, J. (1995). Two eyes are better than one: binocular enhancement in the contrast domain. *Ophthalmic and Physiological Optics*, 15(1), 45-48.

Rajan, R., and Cainer, K.E. (2008). Ageing without hearing loss or cognitive impairment causes a decrease in speech intelligibility only in informational maskers. *Neuroscience,* 154*,* 784-795.

Rakowski, A. (1977). Measurements of pitch. *Catgut Acoustical Society Newsletter, 27,* 9-11.

Rakowski, A. (1978). *Kategorialna percepcja wysokości dźwięku w muzyce* [Categorical perception of pitch in music]. Warsaw, Poland: Państwowa Wyższa Szkoła Muzyczna.

Rakowski, A. (1978). Kategoriala percepcja wysokosci dzwieku w muzyce. Habilitation dissertation. Warszawa, Poland: Panstwowa Wyzsza Szkola Muzyczna.

Rammsayer, T., and Lustnauer, S. (1989). Sex differences in time perception. *Perceptual and Motor Skills*, 68, 195-198.

Rao, M., and Letowski, T. (2004). Callsign Acquisition Test (CAT): Speech intelligibility in noise. *Ear and Hearing,* 27*,* 120-128.

Rappaport, J., and Provencal, C. (2002). Neurotology for audiologists. In: Katz, J., Brukard, R., and Medwetsky, L. (Eds.), *Handbook of Clinical Audiology* (5th Ed.). Baltimore: Lippincott Williams and Wilkins.

Rayleigh, L. (1907). On our perception of sound direction, *Philosophical Magazine,* 13*,* 214-232.

Relkin, E.M., and Doucet, J.R. (1997). Is loudness simply proportional to the auditory nerve spike count? *Journal of the Acoustical Society of America,* 101*,* 2735-2740.

Resnick, S.B., Dubno, J.R., Hoffnung, S., and Levitt, H. (1975). Phoneme errors on a nonsense syllable test. *Journal of the Acoustical Society of America,* 58*,* S114.

Richards, A.M. (1975). Most comfortable loudness for pure tones and speech in the presence of the masking noise. *Journal of the Speech and Hearing Research,* 18*,* 498-505.

Riesz, R.R. (1928). Differential intensity sensitivity of the ear. *Physics Review,* 31*,* 867-875.

Ritsma, R.J. (1967). Frequencies dominant in the perception of the pitch of complex sounds. *Journal of the Acoustical Society of America,* 42*,* 191-198.

Ritsma, R.J., and Bilsen, F.A. (1970). Spectral regions dominant in the perception of repetition pitch. *Acustica,* 23*,* 334-340.

Robinson, D.W. (1986). Sources of variability in normal hearing sensitivity. *Proceedings of the 12th International Congress on Acoustics (paper B11-1)*. Toronto: ICA.

Robinson, D.W., and Dadson, R.S. (1957). Equal-loudness relations and threshold of hearing for pure tones. *Journal of the Acoustical Society of America,* 29*,* 1284-1288.

Roederer, J.G. (1974). *Introduction to the Physics and Psychophysics of Music*. New York: Springer-Verlag.

Rose, J.E., Brugge, J.F., Anderson, D.J. and Hind, J.E. (1968). Patterns of activity in single auditory nerve fibres of the squirrel monkey. In: de Reuck, A.V.S., and Knight, J. (Eds.), *Hearing Mechanisms in Vertebrates,* London: Churchill.

Ross, M., and Lerman, J. (1970). A picture identification test for hearing-impaired children. *Journal of Speech and Hearing Research,* 13*,* 44-53.

Rossing, T. (1989). *The Science of Sound*. Reading, MA: Addison-Wesley Publishing Company.

Rosenzweig, M.R., and Postman, L. (1957). Intelligibility as a function of frequency of usage. *Journal of Experimental Psychology,* 54(6), 412-422.

Roush, J., and Tait, C.A. (1984). Binaural fusion, masking level difference, and auditory brainstem response in children with language-learning disabilities. *Ear and Hearing*, 5, 37-41.

Sachs, M.B., and Kiang, N.Y.S. (1967). Two-tone inhibition in auditory-nerve fibers. *Journal of the Acoustical Society of America,* 43, 1120-1128.

Sakaguchi, S., Arai, T., and Murahara, Y. (2000). The effect of polarity inversion of speech on human perception and data hiding as an application. Proceedings of the IEEE 2000 International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2000), 2, 917-920. Istanbul (Turkey): ICASSP.

Sammeth, C.A., Birman, M., and Hecox, K.E. (1989). Variability of most comfortable and uncomfortable loudness levels to speech stimuli in the hearing impaired. *Ear and Hearing,* 10*,* 94-100.

Sanders, J.W., and Honig, E.A. (1967). Brief tone audiometry: Results in normal and impaired ears. *Archives of Otolaryngology,* 85*,* 640-647.

Sanders, J.W., and Josey, A.F. (1970). Narrow-band noise audiometry for hard-to-test patients. *Journal of Speech and Hearing Research*, 13, 74-81.

Saunders, F.A. (1962). Violins old and new – an experimental study. *Sound*, 1, 7-15.

Scharf, B. (1983). Loudness adaptation. In: Tobias, J.V., and Schubert, E.D. (Eds.), *Hearing Research and Theory,* 2, 1-56. New York: Academic Press.

Scharine, A.A. (2002). *Auditory scene analysis: The role of positive correlation of dynamic changes in intensity and frequency.* Unpublished Dissertation. Tempe, AZ: Arizona State University.

Scharine, A.A., and Letowski, T. (2005). Factors affecting auditory localization and situational awareness in the urban battlefield. ARL Technical Report ARL-TR-3474. Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Scharf, B. (1978). Loudness. In: Carterette, E.C., and Friedman, M.P. (Eds.), *Handbook of Perceptio*n, 4, 187-242. New York: Academic Press

Scharf, B., and Fishken, D. (1970). Binaural summation of loudness: Reconsidered. *Journal of Experimental Psychology,* 86, 374-379.

Schlauch, R.S., DiGiovanni, J.J., and Ries, D.T. (1998). Basilar membrane nonlinearity and loudness. *Journal of the Acoustical Society of America,* 103, 2010-2020.

Schlauch, R.S., Ries, D.T., and DiGiovanni, J.J. (2001). Duration discrimination and subjective duration for ramped and damped sounds. *Journal of the Acoustical Society of America,* 109*,* 2880-2887.

Schoeny, Z.G., and Carhart, R. (1971) Effects of unilateral Ménières disease on masking-level difference. *Journal of the Acoustical Society of America,* 50, 1143-1150.

Schorer, E. (1986). Critical modulation frequency based on detection of AM versus FM tones. *Journal of the Acoustical Society of America,* 79, 1054-1057.

Schouten, J.F. (1940). The residue and the mechanism of hearing. *Proceedings of the Koninklije Nederlandse Akademie van Wetenschappen,* 43, 991-999.

Schouten, J.R., Ritsma, R.J., and Cardozo, B.L. (1961). Pitch of the residue. *Journal of the Acoustical Society of America,* 34, 1418-1424,

Schubert, E.D. (1978). History of research on hearing. In: Carterette, E., and Friedman, M. (Eds.), *Handbook of Perception*, 4, 40-80. New York: Academic Press.

Schultz, M.C., and Schubert, E.D. (1969). A multiple choice discrimination test. *The Laryngoscope,* 79*,* 382.

Seashore, C.E. (1899). Some psychological statistics. *University of Iowa Studies in Psychology*, 2, 1-84.

Sekey, A. (1963). Short-term auditory frequency discrimination. *Journal of the Acoustical Society of America,* 35*,* 682-690.

Sekuler, R., and Blake, R. (1994). *Perception* (3rd Ed.). New York: McGraw Hill.

Sergeant, L., Atkinson, J.E., and Lacroix, P.G. (1979). *The NSMRL tri-word test of intelligibility*. Springfield, VA: National Technical Information Service.

Shafiro, V., and Gygi, B. (2004). How to select stimuli for environmental sound research and where to find them. *Behavioral Research Methods, Instrumentation, and Computers,* 36, 590-598.

Shailer, M.J., and Moore, B.C.J. (1983). Gap detection as a function of frequency, bandwidth, and level. *Journal of the Acoustical Society of America*, 74, 567-473.

Shanefield, D. (1980). The great ego crunchers: Equalized, double-blind tests. *High Fidelity,* 45*,* 57-61.

Shannon, R.V. (1976). Two-tone unmasking and suppression in a forward-masking situation. *Journal of the Acoustical Society of America,* 59, 1460-1470.

Shannon, R.V. (1983). Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics. *Hearing Research,* 11, 157-189.

Shofner, W.P., and Selas, G. (2002). Pitch strength and Stevens' power law. *Perception and Psychophysics,* 64*,* 437-450.

Shower, E.G., and Biddulph, R. (1931). Differential pitch sensitivity. *Journal of the Acoustical Society of America,* 3, 275-287.

Siegenthaler, B.M., and Haspiel, G.S. (1966). *Development of two standardized measures of hearing for speech by children*. University Park, PA: Pennsylvania State University.

Silverman, S.R., and Hirsh, I.J. (1955). Problems related to the use of speech in clinical audiometry. *The Annals of Otology, Rhinology, and Laryngology,* 64, 1234-1244.

Slot, G. (1954). *Audio Quality*. Eindhoven, Holland: Philips.

Small, A.M., Bacon, W.E., and Fozard, J.L. (1959). Intensive differential threshold for octave=band noise. *Journal of the Acoustical Society of America,* 31, 508-520.

Smith, F.O. (1914). The effect of training on pitch discrimination. *Psychological Monographs*, 16(69), 67-103.

Smith, J.O., and Abel, J.S. (1999). Bark and ERB bilinear transforms. *IEEE Transactions on Speech and Audio Processing,* 7, 697-708.

Smoorenburg, G.F. (1970). Pitch perception of two-frequency stimuli. *Journal of the Acoustical Society of America,* 48, 924-942.

Solomon, L.N. (1958). Semantic approach to the perception of complex sounds. *Journal of the Acoustical Society of America,* 30, 421-425.

Somerville, T., and Gilford, C.L.S. (1959). Acoustics of large orchestral studios and concert halls. *Journal of the Audio Engineering Society,* 7, 160-172.

Sone, T., Suzuki, Y., Ozawa, K., and Asano, F. (1994).Information of loudness in aural communication. *Interdisciplinary Information Sciences,* 1, 51-66.

Speaks, C., and Jerger, J. (1965). Method for measurement of speech identification. *Journal of the Acoustical Society of America,* 37, 1205.

Spiegel, M.F. (1979). The range of spectral integration. *Journal of the Acoustical Society of America*, 66, 1356-1363.

Staffeldt, H. (1974). Correlation between subjective and objective data for quality loudspeakers*. Journal of the Audio Engineering Society,* 22, 402-415

Stapells, D.R., Picton, T.W., and Smith, A.D. (1982). Normal hearing thresholds for clicks. *The Journal of the Acoustical Society of America,* 72, 74-79.

Steeneken, H.J.M. (1992). On measuring and predicting speech intelligibility. Unpublished Dissertation, University of Amsterdam. Amsterdam (Holland): Institute of Phonetic Sciences,

Steeneken, H.J.M., and Houtgast, T. (1980). A physical method for measuring speech-transmission quality. *Journal of the Acoustical Society of America,* 67, 318-326.

Steeneken, H.J.M., and Houtgast, T. (1999). Mutual dependence of octave-band weights in predicting speech intelligibility. *Speech Communication,* 28, 109-123.

Steinberg, J.C., and Snow, W.B. (1934). Auditory Perspective - Physical Factors. Proceedings of the Symposium on Wire Transmission of Symphonic Music and its Reproduction in Auditory Perspective. *Electrical Engineering*, 1, 12-17.

Steinke, G. (1958). Subjektive Bewertung der Übertragungsqualität. *Technische Mitteilungen aus dem Betriebslaboratorium für Rundfunk und Fernsehen,* 13, 27-30.

Stelmachowicz, P.G., Beauchaine, K.A., Kalberer, A., and Jesteadt, W. (1989). Normative thresholds in the 8- to 20-kHz range as a function of age. *Journal of the Acoustical Society of America,* 86, 1384-1391.

Stevens, S.S. (1934a). The volume and intensity of tones. *American Journal of Psychology*, 46, 397-408.

Stevens, S.S. (1934b). Tonal density. *Journal of Experimental Psychology*, 17, 585-592.

Stevens, S.S. (1935). The relation of pitch to intensity. *Journal of the Acoustical Society of America,* 6, 150-154.

Stevens, S.S. (1955). The measurement of loudness. *Journal of the Acoustical Society of America,* 27, 815-829.

Stevens, S.S. (1956). Calculation of the loudness of complex noise. *Journal of the Acoustical Society of America,* 28*,* 807-832.

Stevens, S.S. (1957). On the psychophysical law. *Psychological Review*, 64, 153-181.

Stevens, S.S. (1972). Perceived level of noise by Mark VII and decibels. *Journal of the Acoustical Society of America*, 51, 575-601.

Stevens, S.S., and Davis H. (1938). *Hearing.* New York: Wiley and Sons.

Stevens, S.S., and Newman, E.B. (1936). The localization of actual sources of sound, *American Journal of Psychology,* 48, 297-306.

Stevens, S.S., and Volkmann, J. (1940). The relation of pitch to frequency: A revised scale. *The American Journal of Psychology,* 8*,* 329-353.

Stevens, S.S., Volkmann, J., and Newman, E.B. (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America,* 8, 185-191.

Studebaker, G.A., and Sherbecoe, R.L. (1988). Magnitude estimations of the intelligibility and quality of speech in noise. *Ear and Hearing*, 9, 259-267.

Stumpf, C. (1911). Konsonanz und Konkordanz. *Beitrage zur Akustik und Musikwissenschaft,* 6*,* 116-150.

Suetomi, D., and Nakajama, Y. (1998). How stable is time-shrinking? *Journal of Music Perception and Cognition,* 4, 19-25.

Summers, W.V., Pisoni, D.B., Bernacki, R.H., Pedlow, R.I., and Stokes, M.A. (1988). Effects of noise on speech production: acoustic and perceptual analyses. *The Journal of the Acoustical Society of America,* 84*,* 917-928.

Suzuki, Y., Sone, T., and Kanasashi, K. (1982). The optimum level of music listened in the presence of noise. *Journal of the Acoustical Society of Japan (E),* 3, 55-65.

Szlifirski, K., and Letowski, T. (1981). System podstawowych kryteriów oceny nagrań dźwiękowych [A system of criteria for the quality evaluation of sound recordings]. *Przegląd Techniki Radia i TV,* 1, 36-42.

Szymaszek, A., Szelag, E., and Sliwowska, M. (2006). Auditory perception of temporal order in humans: The effect of age, gender, listener practice and stimulus presentation mode. *Neuroscience Letters*, 403, 190-194.

Tang, Q., Liu, S., and Zeng, F-G. (2006). Loudness adaptation in acoustic and electric hearing. *Journal of the Association for Research in Otolaryngology,* 7, 59-70.

Tang H., and Letowski T. (2007). High-frequency hearing threshold measurements using insert earphones and warble signals. *Proceedings of the 2007 Industrial Engineering Research Conference.* Nashville, TN: Institute of Industrial Engineers (IIE).

Terhardt, E. (1974). On the perception of periodic sound fluctuations (roughness). *Acustica,* 30, 201-213.

Terhardt, E. (1979). Calculating virtual pitch. *Hearing Research,* 1, 155-182.

Terhardt, E., Stroll, G., and Seewann, M. (1982). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America,* 71, 679-688.

Thomas, G.J. (1949). Equal-volume judgment of tones. *American Journal of Psychology,* 62*,* 182-201.

Thwing, E.J. (1955). Spread of perstimulatory fatigue of the pure tone to neighboring frequencies. *Journal of the Acoustical Society of America,* 27, 741-748.

Thurlow, W.R., Mangels, J.W., and Runge, P.S. (1967). Head movements during sound localization, *Journal of the Acoustical Society of America,* 42, 489-493.

Tillman, T.W., Carhart, R.T., and Wilber, L.A. (1963). *A test for speech discrimination composed of CNC monosyllabic words (N.U. Auditory Test No. 4)*. Brooks Air Force Base, TX: USAF School of Aerospace Medicine, Aerospace Medical Division (AFSC).

Toole, F.E. (1982), Listening tests- turning opinion into fact. *Journal of the Acoustical Society of America,* 30, 431-445.

Tran T., Letowski T., and Abouchacra K. (2000). Evaluation of acoustic beacon characteristics for navigation tasks. *Ergonomics*, 43(6), 807-827.

Tran, P., Letowski, T., and McBride, M. (2008). Bone conduction microphone: Head sensitivity mapping for speech intelligibility and sound quality. *Proceedings of the 2008 International Conference on Audio, Language and Image Processing (ICALIP 2008)*, 107-111. Shanghai (China): Institute of Electrical and Electronic Engineers (IEEE).

Truman, S.R., and Wever, E.G. (1928). The judgment of pitch as a function of the series. *University of California Publications in Psychology,* 3, 215-223.

Turnbull, W.W. (1944). Pitch discrimination as a function of tonal duration. *Journal of Experimental Psychology,* 34, 302-316.

Uchanski, R.M., Choi, S.S., Braida, L.D., Reed, C.M., and Durlach, N.I. (1996). Speaking clearly for the hard of hearing .4. Further studies of the role of speaking rate. *Journal of Speech and Hearing Research,* 39(3), 494-509.

van den Brink, G. (1974). Monotic and dichotic pitch matchings with complex sounds. In: Zwicker, E., and Terhardt, E., (Eds.), *Models in Hearing*. Berlin, Germany: Springer-Verlag.

van Wijngaarden, S.J. and Drullman, R. (2008). Binaural intelligibility prediction based on the speech transmission index. *Journal of the Acoustical Society of America,* 123, 4514-4523.

Viemeister, N. (1979). Temporal modulation transfer function based upon modulation thresholds. *Journal of the Acoustical Society of America,* 66, 1364-1380.

Voiers, W.D. (1983). Evaluating processed speech using the Diagnostic Rhyme Test. *Speech Technology,* 1, 338-352.

Voiers, W.D., Sharpley, A. and Hehmsoth, C. (1975). Research on diagnostic evaluation of speech intelligibility. Bedford, MA: U.S. Air Force Cambridge Research Laboratories. Research Report AFCRL-72-0694.

Wallach, H. (1940). The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology,* 27, 339-368.

Wallach, H., Newman, E.B., and Rosenzweig, M.R. (1949). Precedence effect in sound localization. *American Journal of Psychology,* 62, 315-336.

Walliser, K. (1968). Zusammenwirkung von Hüllkurven-periode und Toneheit bei der Bildung der Periodentonhöhe. Doctoral dissertation. München (Germany): München Technische Hochschule.

Walliser, K. (1969). Über die Abhängigkeiten der Tonhöhenempfindung von Sinustönen vom Schallpegel von überlagertem drosselndem Störschall und von der Darbietungsdauer. *Acustica,* 21, 211-221.

Ward, W.D. (1954). Subjective musical pitch. *Journal of the Acoustical Society of America,* 26, 369-380.

Ward, W.D. (1963). Diplacusis and auditory theory. *Journal of the Acoustical Society of America,* 35, 1746-1747.

Warfield, D. (1973). The study of hearing in animals. In: Gay, W. (Ed.), *Methods of Animal Experimentation, IV*. London: Academic Press.

Warren, R.M., and Wrightson, J.M. (1981). Stimuli producing conflicting temporal and spectral cues to frequency. *Journal of the Acoustical Society of America,* 70, 1020-1024.

Watson, B.U. (1991). Some relationship between intelligence and auditory discrimination. *Journal of Speech and Hearing Research,* 34, 621-627.

Watson, C.S., Kelly, W.J., and Wronton, H.W. (1976). Factors in the discrimination of tonal patterns, II: Selective attention and learning under various levels of stimulus uncertainty. *Journal of the Acoustical Society of America,* 60, 1176-1186.

Weber, E.H. (1834). *De Pulsu, Resorpitone, Auditu et Tactu: Annotationes Anatomicae et Physiologica*e. Leipzig, Germany: Koehlor.

Webster, F.A. (1951). The influence of interaural phase on masked thresholds, I: The role of the interaural time-deviation. *Journal of the Acoustical Society of America,* 23, 452-462.

Wegel, R.L., and Lane, C.E. (1924). The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear. *Physical Review,* 23, 266-285.

Weinstein, B. (2000). *Geriatric Audiology.* New York: Thieme.

Whilby, S., Florentine, M., Wagner, E., and Marozeau, J. (2006). Monaural and binaural loudness of 5- and 200-ms tones in normal and impaired hearing. *Journal of the Acoustical Society of America,* 119, 3931-3939.

White, L.J., and Plack, C.J. (2003). Factors affecting the duration effect in pitch perception for unresolved complex tones. *Journal of the Acoustical Society of America,* 114, 3309-3316.

Wier, C.C., Jesteadt, W., and Green, D.M. (1977). Frequency discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America,* 61, 178-184.

Wightman, F.L., and Green, D.M. (1974). The perception of pitch. *American Scientist,* 62, 208-215.

Wightman, F.L., and Kistler, D.J. (1992). The dominant role of low-frequency interaural time differences in sound localization, *Journal of the Acoustical Society of America,* 91, 1648-1661.

Willott, J. (1991). *Aging and the Auditory System.* San Diego: Singular Publishing Group.

Wilson, R.H., and Antablin, J. K. (1980). A picture identification task as an estimate of the word-recognition performance of nonverbal adults. *Journal of Speech and Hearing Disorders,* 45, 223-238.

Woodrow, H. (1951). Time perception. In: Stevens, S.S. (Ed.) *Handbook of Experimental Psychology*, Chapter 32. New York: Wiley and Sons.

Yamashita, T., Ishii, Y., Nakamura, M., and Kitamura, O. (1990). A method for evaluating interior sound quality of automobiles. Paper presented at the 119th Acoustical Society of America (ASA) Conference, paper PP7. State College, PA: Acoustical Society of America (ASA).

Zarcoff, M. (1958). Correlation of filtered, band-limited white noise and the speech reception threshold. *Archives of Otolaryngology - Head and Neck Surgery*, 63, 372-381.

Zeng, F-G., and Shannon, R.V. (1994). Loudness-coding mechanisms inferred from electric simulation of the human auditory system. *Science,* 264, 564-566.

Zera J., Boehm T., and Letowski T. (1982). Application of an automatic audiometer in the measurement of the directional characteristic of hearing. *Archives of Acoustics* 7, 197-205.

Zwicker, E. (1952). Die Grenzen der Hörbarkeit der Amplituden-modulation und Frequenz-modulation eines Tones. *Acustica*, 3, 125-133.

Zwicker, E. (1960). Ein Verfahren zur Berechnung der Lautstärke. *Acustica,* 10, 304-308.

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *Journal of the Acoustical Society of America,* 33*,* 248-248.

Zwicker, E., and Fastl, H. (1999). *Psycho-acoustics: Facts and Models* (2nd Ed.). Berlin: Springer.

Zwicker, E., and Feldtkeller, R. (1955). Über die Lautstärke von gleichformigen Geräuschen. *Acoustica* 5, 303-316.

Zwicker, E., and Feldtkeller, J. (1967). *Das Ohr als Nachrichtenempfanger* (2nd Ed.). Stuttgart (Germany): Hirzel Verlag.

Zwicker, E., Flottrop, G., and Stevens, S.S. (1957). Critical band width in loudness summation. *Journal of the Acoustical Society of America*, 29, 548-557.

Zwicker, E., and Scharf, B. (1965). A model of loudness summation. *Psychological Review,* 72, 3-26.

Zwicker, E., and Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bandwidth as a function of frequency, *Journal of the Acoustical Society of America*, 68, 1523-1525.

Zwicker, E., and Zwicker, U.T. (1991). Dependence of binaural loudness summation on interaural level differences, spectral distribution, and temporal distribution. *Journal of the Acoustical Society of America,* 89, 756-764.

Zwislocki, J. (1960). Theory of temporal auditory summation. *Journal of the Acoustical Society of America,* 32, 1046-1060.

Zwislocki, J.J. (1965). Analysis of some auditory characteristics. In: Luce, R.D., Bush, B.R., and Galanter, E. (Eds.), *Handbook of Mathematical Psychology*, III. New York: Wiley and Sons.