- 13 -
Psychoacoustics

The perception of sound

Psychoacoustics is the study of the perception of sound. The perception of sound is determined by the human auditory system and to be able to extract spectral and temporal information, it requires a substantial amount of signal processing. In many respects the ear is a remarkably accurate sensor and yet in others it is notably insensitive. Sorting out these differences is a monumental task.

Psychoacoustics is a difficult field of study. Understanding hearing is like trying to understand a loudspeaker without the ability to look at or test one directly – where we can only infer what is going on from a limited amount of data. More recently, direct testing of the ear has been possible, but not all of this data has yet made it into the academic literature and even less so the fundamental theories. As such, there remains much that we do not know.

This chapter will provide a basic overview that is pertinent to the understanding of the spectral and temporal analysis of the ear. First, we will discuss the anatomy and physiology of the ear, and then we will discuss critical concepts that relate to the perception of sound. We will describe what we know and what we think we know at this point in time. Our primary topics of interest will include masking, loudness perception, and binaural hearing.

13.1 Anatomy and Physiology

The peripheral auditory system is divided into three parts: the outer, the middle, and the inner ear. Fig. 13-1 shows the cross-section of what is called the peripheral auditory system. The outer ear consists of the pinna and the ear canal. The middle ear is an air-filled cavity that consists of three bones: the malleus, the incus and the stapes, commonly known as the ossicles. One end of the ossicular chain is attached to the tympanic membrane (eardrum), which separates the outer and middle ear and the other end (footplate of the stapes) is attached to the oval window, which separates the middle and the inner ear. The ossicles are critical in matching the low impedance[1] of air in the middles ear to the much higher impedance of the fluid-filled inner-ear. As sound travels through the middle ear, the ossicles provides about 30 dB of gain along its transmission path to the inner ear. This is primarily due to the areal ratio between the tympanic membrane and the

1. See Pickles, *An Introduction to the Physiology of Hearing*

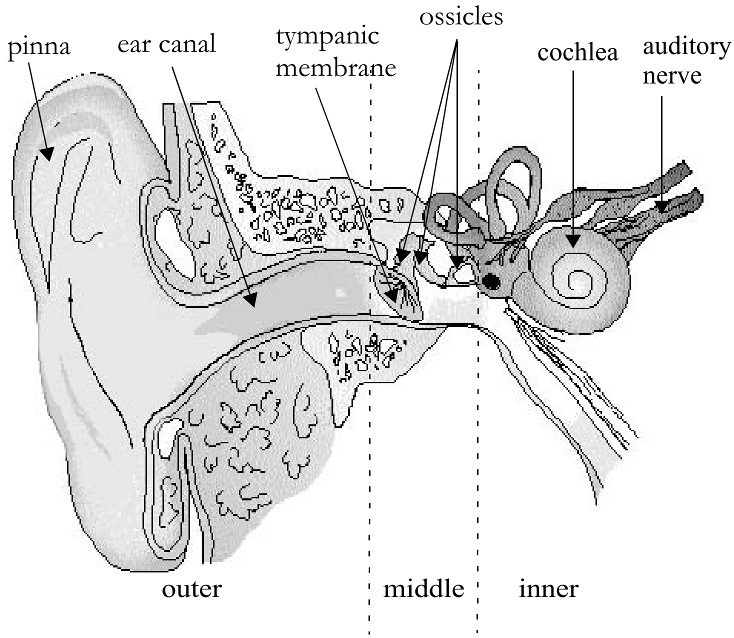

Figure 13-1 - Cross section of the peripheral auditory system

oval window which creates a mechanical lever much akin to or earlier use of an Acoustic Lever.

The parts of the inner ear that are important to auditory processes include the cochlea and the auditory nerve. The cochlear is a snail-shaped structure that is divided longitudinally into three chambers: the scala vestibuli, the scala media and the scala tympani. The Reissner's membrane separates the scala vestibuli and the scala media. The basilar membrane separates the scala media and the scala tympani. On the basilar membrane sits the organ of Corti, which houses the hearing sensory cells. Fig.13-2 is a cross-section of the organ of Corti. There are two types of sensory cells: outer hair cells and inner hair cells. As the footplate of the stapes oscillates back and forth in the oval window (OW), a traveling wave is generated along the basilar membrane which initiates the excitation of the inner and outer hair cells. The hair cells are tuned both mechanically and electrically and transfer particular frequencies of sound into activity in the organ of Corti. The auditory system provides us with the sensory capacity to analyze complex sounds, such as speech and music.

13.2 Theories of Hearing

Research in psychoacoustics and its development are closely associated with the theories of hearing. We will briefly introduce the basics of the hearing theories. Traditional theories of hearing are divided into two main categories: place

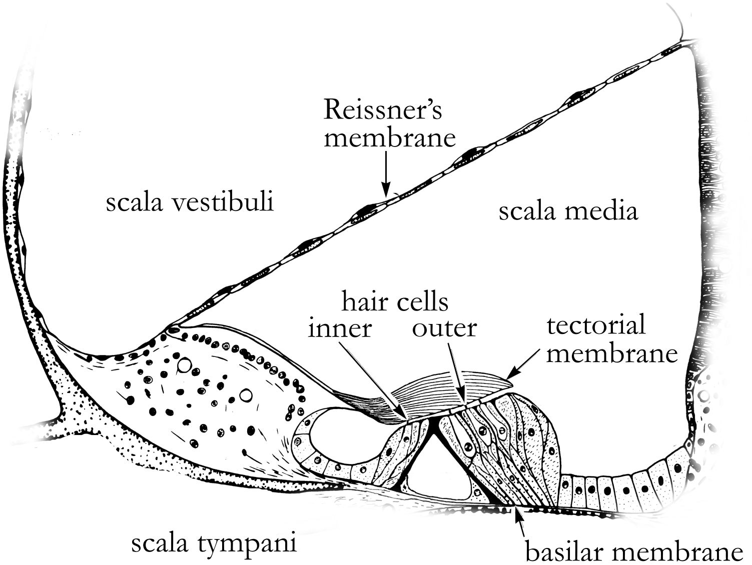

Figure 13-2 - Cross section of the organ of Corti

theory and frequency theory. Place theory suggests that the basilar membrane is frequency sensitive, and it is at the cochlea where the primary signal analysis take place. The frequency theory suggests the opposite. Frequency theory proposes that the cochlea is not sensitive to frequency, and its primary function is to transmit information to the auditory nerve and on to the central auditory nervous system where the analysis of the signal occurs. Frequency theory focuses on the temporal and periodicity characteristics of the signal. Based on these two fundamental concepts place-frequency theory has evolved, which combines the two to better explain the hearing mechanism.

Frequency theory

Frequency theory is based on the assumption that the auditory nerve fibers are capable of firing at a very high rate in order to transmit the information. For example, in order to transmit a 500 Hz signal, the fiber has to fire 500 times a second. This theory is adequate at explaining our perception of low frequencies, however, above about 1 kHz it becomes untenable due to the refractory period of the neuron (the recharge rate). The refractory period is the time a cell takes to re-establish its polarization in order to fire again and is typically 1 ms long (which correspond to 1 kHz). Hence, above 1 kHz, the neurons are unable to fire sychronously with the input signal.

A modification of the frequency theory is the volley principle, which suggests that groups of fibers combine their information together to represent a frequency in the auditory nerve. The composite image reflects the periodicity characteristic of the signal and can explain our perception of signals as high as 5 kHz.

Place theory

Place theory is considered to be the most well established theory to date. Helmholtz resonance theory suggests that the basilar membrane is partitioned into different segments and each segment resonates in response to a certain band of frequencies. These segments are differentiated because of the varying tension that occurs along the length of the basilar membrane. Hence, the frequency is tuned by its place on the cochlea. As a complex signal reaches the cochlea, a sort of Fourier analysis is performed. The different components of a complex signal excites the cochlea to a maximum amplitude of vibration at different points depending on the frequency components of the signal. Another prominent place theory was introduced by Nobel laureate Georg von Békésy, known as traveling wave theory[2]. Békésy showed that there is a widening of the cochlear partition from the base to the apex which results in a stiffness variation along the membrane. Fig. 13-3 shows the schematic of the uncoiled snail-shaped cochlea. The

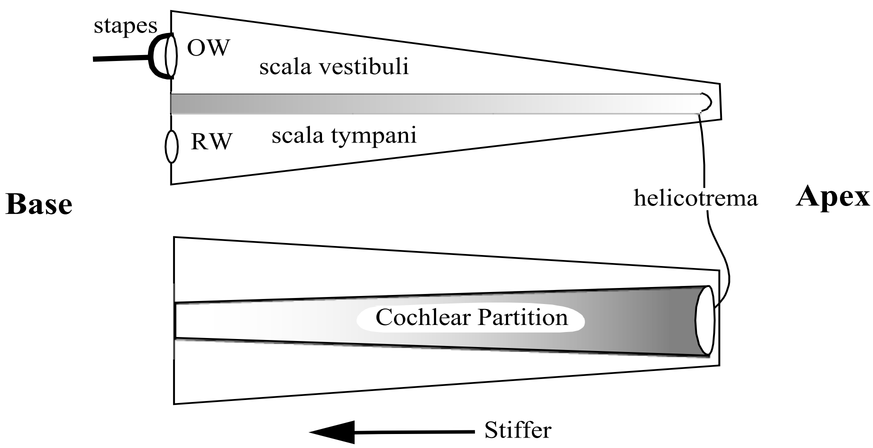

Figure 13-3 - Schematic of an uncoiled cochlea

stiffness gradient of the cochlear partition of the basilar membrane decreases as the partition widens from base to apex. The partition is stiff and narrow at the base, and is responsive to high-frequency signals; as the partition extends towards the apex, it systematically widens and becomes more flaccid and is therefor more responsive to low-frequency signals. Fig. 13-4 displays a simulation of a traveling wave on the basilar membrane at successive instances in time (dotted lines), at different frequencies. We can see in this figure that the waves travel down the

2. See von Békésy, *Experiments in Hearing*.

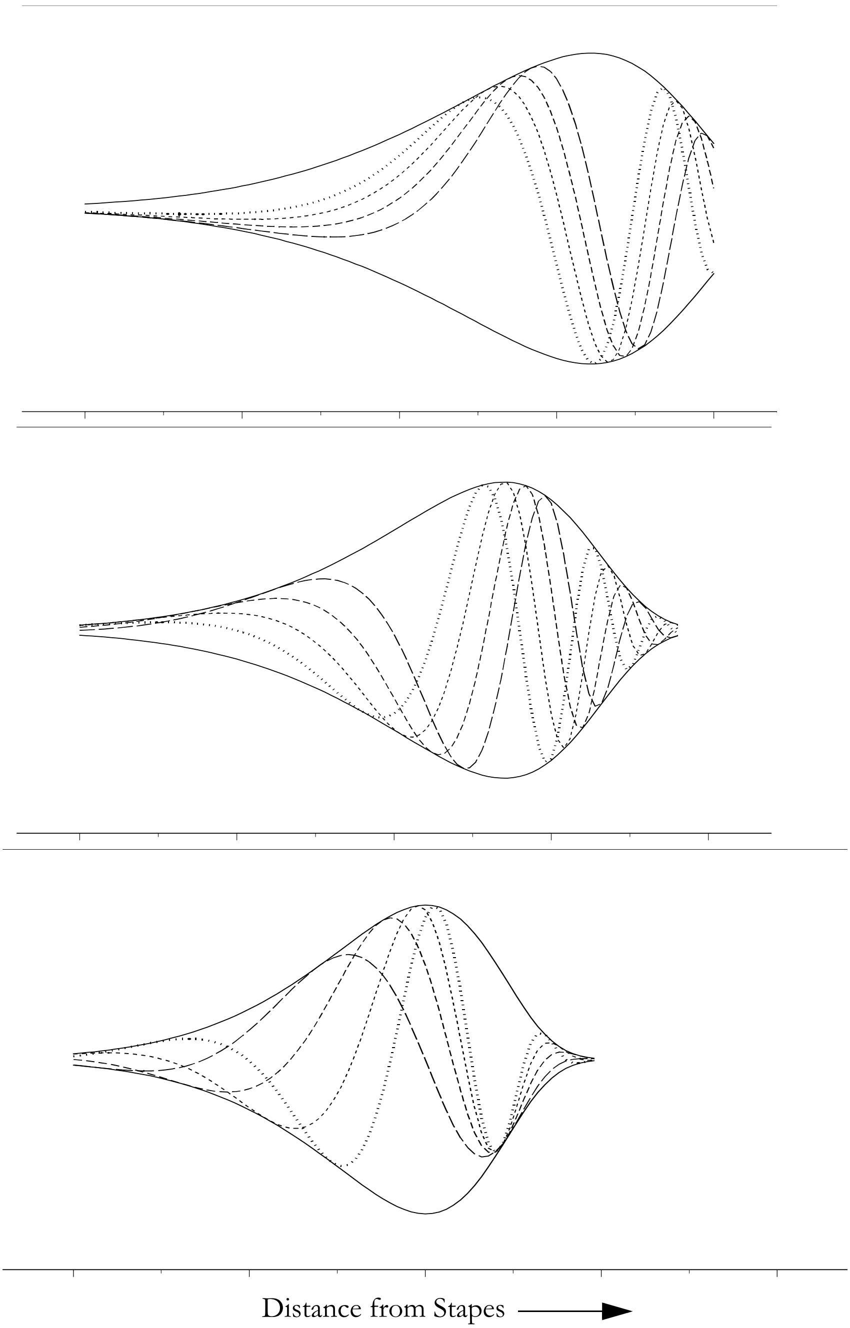

Figure 13-4 - Graphic representation of traveling waves at different frequencies: 100 Hz (top), 400 Hz (middle), 800 Hz (bottom) membrane and reach their point of maximum displacement at different locations. The resulting curves trace out an envelope for the traveling wave shown as the solid line. Note that the displacement builds gradually until it reaches a maximum which in a gentle slope towards the basal end (high frequency). It then decays rapidly beyond that point, resulting in a steep apical slope (towards the low frequency). Fig. 13-5 displays von Békésy's experimental results for the traveling wave pattern as it moves from the base to the apex. Note that the maximum displacement of a low frequency signal is towards the apex (away from the stapes) and that the maximum displacement of a high frequency signal is towards the base of the membrane (close to the stapes). Note also that as a low frequency signal travels, it vibrates the membrane at the basal end reaching its maximum displacement at the apical end. A high frequency signal has its maximum displacement at the basal end – the apical portion of the membrane is not disturbed. This effect is a critical factor in understanding the masking effect that we will discuss in subsequent sections.

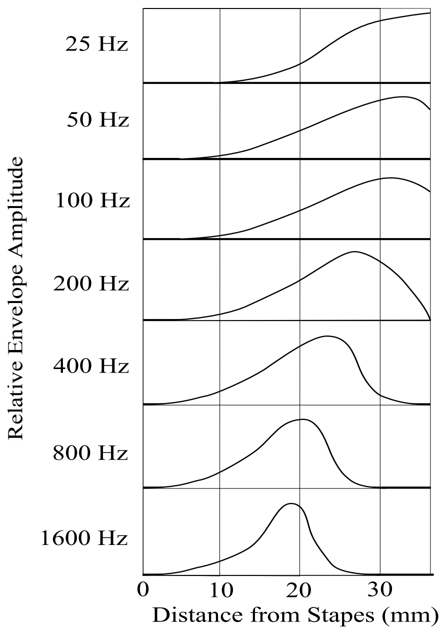

Figure 13-5 - Von Békésy's experimental results for cochlear traveling wave

13.3 Psychoacoustical Measurement Protocols

It is well recognized that our psychological response to a stimulus is not a direct reflection of the physical stimulus. The primary goal of psychoacosutics research is to construct a relationship between the two. In order to be successful, the first goal is to have well designed, controlled experiments, so that the outcome a measurement is conclusive. Experimental protocols in psychoacoustical research usually require the use of double-blind, randomized presentations to reduce the error associated with self deception and bias. Furthermore, it is critical for an experiment to avoid any ambiguities when presenting a stimulus, or when obtaining a response. The stimulus and the response need to be clearly specified, while certain aspects of the stimulus are manipulated. The subjects task must be predetermined so that unambiguous responses can be obtained and interpreted. These criteria can be fulfilled quite easily with today's computer-driven technology. To do psychoacoustical experiments without these strict controls is akin to palm reading: one has to accept the results on faith.

Frequency and temporal resolution are critical in our perception of sound. But like any other sensory system, there are limitations to our auditory system. Frequency and temporal selectivity refers to the ability of a subject to discriminate spectral and sequential component differences in a complex signal. This phenomenon has been studied extensively in psychoacoustical research by means of masking experiments, measuring either detection threshold or differential threshold. Gap detection is a protocols commonly used to study temporal resolution.

Detection threshold

Detection threshold represents the softest sound pressure necessary for the listener to detect the presence of a signal. Detection threshold for a sinusoid varies with the duration of the signal. As the duration of the sinusoid increases, the power of the signal increases, hence the decrease of the detection threshold. Our auditory system integrates energy across a temporal window. The total power of the signal is integrated as a function of time. This phenomenon of temporal integration is true for tonal signal that ranges about 10 to 300 ms depending on the frequency. The threshold does not continue to decrease as the signal duration reaches beyond 300 ms and as the duration decrease below 10 ms, the spread of the spectral content of the signal makes detection threshold invalid because of its transient quality.

Differential threshold

Differential threshold represents the smallest difference that the listener requires to discriminate between two conditions (a.k.a. just noticeable difference – JND or difference limen – DL). Differential sensitivity is described either by the absolute DL (ΔS), or a relative DL ($\Delta S/S$). According to Weber's law, the relative DL in intensity is a constant, regardless of the stimulus level. However, studies have shown that Weber's fraction only holds for limited ranges in both intensity and frequency – 10-40 dB SL (where SL is the sensation level, the level above threshold) for intensity and frequencies between 500-2 kHz. Furthermore, Δf (the differential frequency limen) remains constant up to about 1 kHz at various intensity levels and then gradually increases as the frequency increases. At low frequencies and low sensation levels, the Weber fraction is about 0.002 for the mid frequencies at 40 dBSL. Our auditory system is sensitive to frequency changes as low as 1 Hz at some frequencies.

Gap detection

The basic principle underlying this protocol is the ability of a listener to detect a brief temporal gap in an otherwise steady-state signal. The results reflect a listener's temporal acuity. Care must be taken to ensure that the listener can only cue to temporal information with no assistance from spectral or intensity information. Current researchers have successfully utilized digital filter techniques to minimize the frequency smearing that takes place when gating a signal. Studies have reported that observers cannot distinguish between a continuous sound burst or two sequential sound bursts separated by a brief gap of silence, if the gap is shorter than 3 ms. Furthermore, the gap detection threshold has been found to be dependent upon external variables such as the level and bandwidth of the signal. There are conflicting results as to the role of the frequency of the signal, with a number of studies demonstrating a frequency effect, and others suggesting no effect of frequency.

13.4 Masking

Masking is an important phenomenon in the understanding of psychoacoustics, hence it is a crucial component of this chapter. As we stated earlier masking is the dominate effect that allows for the perceptual coding of sound signals. Masking refers to the change in the sensitivity of a stimulus signal (probe) in the presence of a interfering signal (masker). The amount of interference is denoted by the shift in threshold of the probe.

An example of a basic masking experiment procedure is in order. First, the test stimulus is presented and its threshold is obtained. This unmasked threshold defines a baseline. Second, the masker is introduced and the threshold of the test stimulus is re-assessed in the presence of the masker. The difference in the threshold measures reflects the amount of masking which has occurred. Masking effects can be measured by simultaneous or non-simultaneous masking. Simultaneous masking implies that the probe and the masker are present at the same time. Non-simultaneous masking includes: forward masking when the masker is presented prior to the probe; and backward masking when the masker is presented after the probe. The characteristics of masking phenomenon[3] are usually presented as psychophysical tuning curves or masking patterns.

3. See Yost, *Fundamentals of Hearing*.

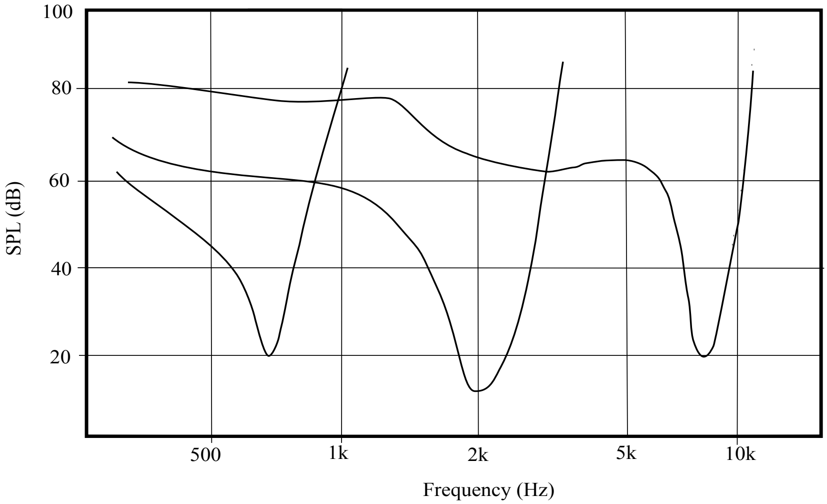

Figure 13-6 - Graphic representation of psychophysical tuning curves at different frequencies

Psychophysical tuning curves

Psychophysical tuning curves (PTC) are obtained by presenting a probe tone at fixed frequency and level (usually 10-15 dB above threshold) and varying the frequency and level of the masker. A plot of the intensity of the masker that just masks the probe tone at different frequency is the PTC. Fig. 13-6 displays the PTC at different frequencies. PTC's obtained with simultaneous and non-simultaneous masking techniques broaden as the probe level increases from 5 to 60dB SPL.

Masking pattern

Masking pattern is observed by measuring the threshold for a signal of variable frequency and level in the presence of a masker at fixed frequency and level. Masking patterns are dependent on the intensity level of the masker because of the nonlinear nature of the cochlea. At a low intensity levels, the pattern is quite symmetrical, but as the intensity increases the pattern becomes increasingly asymmetrical, which results in additional masking towards high frequencies (see Fig. 13-7). Generally speaking, the most effective masking occurs when the probe and the masker frequency are in close proximity. Low frequency maskers are more effective than high frequency maskers, and the amount of masking increases markedly as the masker intensity increases. This is the basis for statements that we made in Sec.10.5 on page 237.

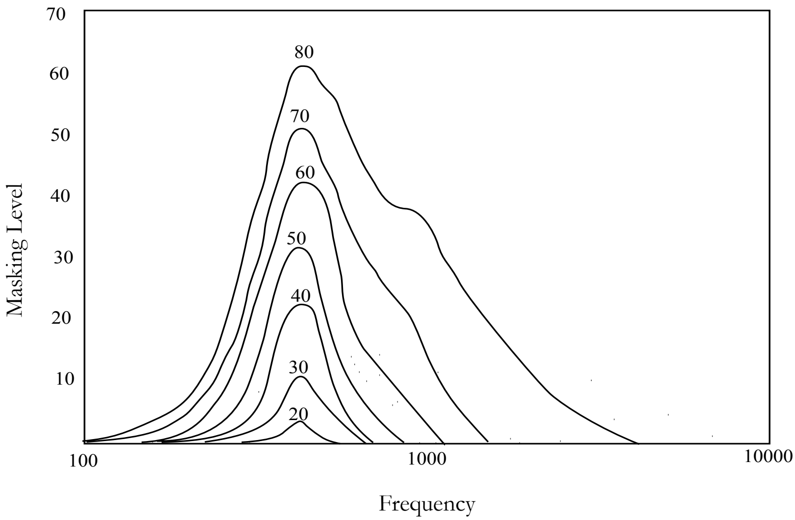

Figure 13-7 - Graphic representation of masking patterns of a fixed probe frequency. Each contour represents a different probe intensity

The masking phenomenon can be logically explained with traveling wave theory. Fig. 13-8 shows a schematic illustrating the masking effect. Panel **I** illustrates the traveling wave envelopes on the cochlea for two signals. Signal A (dotted line) depicts a high frequency tone, and Signal B (solid line) depicts a low frequency tone. The x-axis represents the distance from the stapes. Panel **II** is a mirror image of Panel **I**, and the x-axis now represents frequency. Panel **II** illustrates the masking patterns of the two identical signals as shown in Panel A. The top panels depict two signals of low intensity level. There is no overlapping of the two signals and therefore there is only minimal masking between the two signals at this intensity level. The middle panels show an increase in intensity level for signal A. As indicated, there is a greater, but not significant, overlap of the two signals. This implies that as the high frequency signal increases in intensity, it will not mask out the low frequency tone of substantially lower intensity. The bottom panel shows an increase in intensity for signal B. As revealed in panel I and II, the high frequency signal is almost completely engulfed by the low frequency signal. This implies that:

- low frequencies are very effective maskers of high frequencies and this effect increases substantially with signal levels
- high frequencies are not effective as maskers for signals at lower frequencies which holds true for a wide range of signal levels
- low frequency maskers are effective over a wider range of frequencies than high frequency maskers

These results are extremely important aspects in the perception of sound.

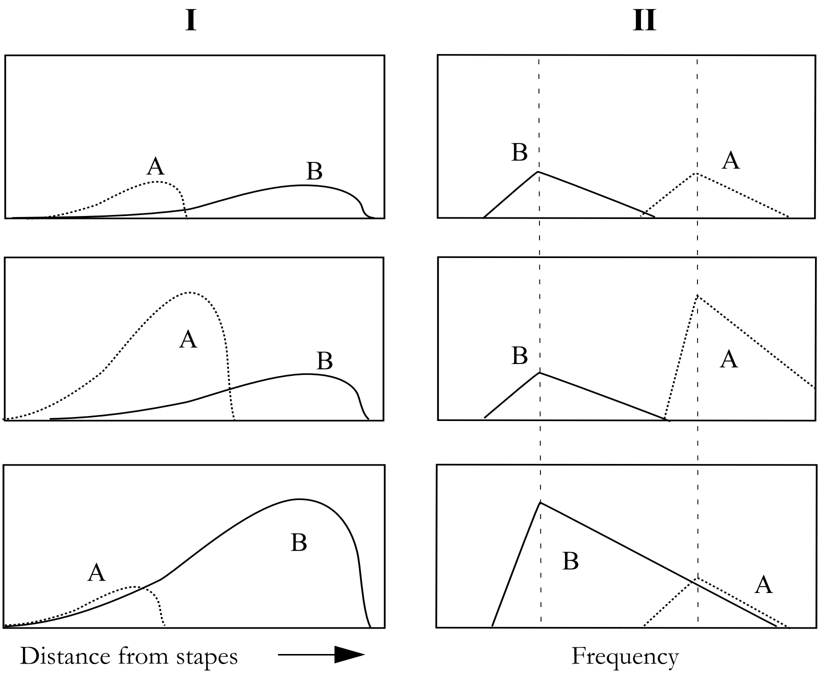

Figure 13-8 - Schematic demonstration of masking

13.5 Loudness

Loudness is an attribute of the auditory sensation where sounds are scaled on a continuum from soft to loud. Stevens & Davis[4] proposed the sone as a unit of loudness, and suggested a 1 kHz tone at 40 dB above threshold in quiet as a 1 sone reference for this scale. Both the unit and the reference are now universally accepted and internationally standardized. They also introduced the log-log coordinate plot for loudness-growth functions, which is now the conventional method used to present such functions.

According to Stevens' Power Law, the loudness-growth function takes the form

$$L_x = k\, I_x^p \tag{13.5.1}$$

L_x=Loudness sensation associated with stimulus x (in sones)

I_x=stimulus intensity

p=an exponent that estimates the slope of the function

k=an arbitrary fitting constant which varies with the scaling units

4. See Stevens, *Hearing, Its Psychology and Physiology.*

For loudness, p is approximately 0.3 for sound intensity, and 0.6 for sound pressure. The power function suggests that the stimulus-sensation relationship is linear across the intensity range when plotted on log-log coordinates.

Studies have demonstrated that the power-law model describes the loudness function at moderate and high sound-pressure levels (>30dB sensation level). Near threshold, however, loudness decreases more rapidly than that suggested by the power-law model. A modification to the classic power-law model[5], to better describe the loudness of near-threshold sounds, takes the form

$$L_x = k\left(I_x - I_t\right)^p \tag{13.5.2}$$

I_t=intensity of threshold

This modification takes into account the quiet threshold by subtracting the threshold intensity (I_t) from the stimulus intensity (I_x) prior to compression. This function suggests a linear correction to the power-law model and only deviates from the uncorrected power-law function until it approaches threshold (within 5-10dB).

In another attempt to better account for the loudness of near-threshold stimuli, a "power-group transformation" was recommended, which modeled the loudness function in two segments: a steep segment representing intensities within 30dB of threshold; and a flatter segment at higher intensities. Each segment was described by its own power-law equation, with separate values of p. Consequently, this modification increased the model complexity and required at least three free parameters (two values of p and the intensity at which they change).

An alternate modification of the original power-law model was introduced as another means of accounting for the loudness of near-threshold stimuli. This modification takes the form:

$$L_x = k\left(I_x^p - I_t^p\right) \tag{13.5.3}$$

p = varies from 0 to 1 and defines the slope of the function at high intensities

It applies a nonlinear transformation to both the stimulus intensity and the threshold intensity. This function, referred to subsequently as the modified power-law (MPL), provides a good description of loudness growth across the entire range of stimulus intensities, from threshold to high intensity.

Other modifications have also been suggested and have resulted in reasonably good approximations of the observed data on loudness growth across a wide range of stimulus intensities. However, they require additional free parameters than those previously discussed, and have not been applied frequently in academic literature.

---

5. See Scharf, "Loudness", *Handbook of Perception.*

Factors affecting the loudness function

The various power-law models described above relate loudness to a single stimulus characteristic: stimulus intensity. Loudness, however, depends on several stimulus factors, other than intensity, such as frequency, bandwidth, and duration. The presence of hearing loss can also affect the loudness-growth function. We will review each of these factors.

Hearing threshold varies across frequency. In the low frequencies, hearing threshold decreases from 45 dB SPL to 7 dB SPL as frequency increases from 125 Hz to 1 kHz and then increases again above 6 kHz. Steeper loudness-growth functions occur at low frequencies than those observed at 1 kHz. The effect of signal frequency on loudness varies with sound level. At low levels (< 20 dB SPL), loudness is greater in the mid frequencies than either the low or high frequencies. At high intensities (100 dB SPL), loudness varies little with frequency.

Loudness of a complex sound depends on both the spectral distribution of its components and the frequency range between the highest and the lowest component (ΔF). Loudness does not increase until ΔF reaches a minimum value, called the critical band, which varies with the center frequency of the complex stimulus. Within a critical band, loudness of a tone complex depends solely upon the overall sound-pressure level of the complex. Beyond the critical band, loudness increases with bandwidth, even when keeping the overall intensity constant. This is usually referred to as loudness summation. Loudness summation across critical bands has its greatest effect at moderate levels, and is lesser at either extreme. The loudness level of a complex signal does not increase with ΔF when the signal is within 10-15 dB SL.

For tonal complexes, loudness is greatest when the components are evenly spaced with respect to critical bands[6]. Also, loudness is greatest when the components are all equal in amplitude, and loudness seems to be constant regardless of the number of components in the complex, as long as DF and the overall sound pressure level are kept constant.

In general, for sounds shorter than a "critical duration", the auditory system integrates energy over time. The critical duration varies over a wide range, depending on the nature of the stimulus, but most values have been within 150-200 ms. Frequency appears to have a minimal effect on critical duration. Studies of the loudness-growth functions of white noise as a function of duration have shown that signal duration has no effect on the exponents of the power function.

Loudness recruitment refers to an abnormally rapid growth of loudness with sound intensity in ears with sensorineural hearing loss. It is most commonly observed when there are defects in the sensory cells within the cochlea. The slope of the loudness-growth function increases with the severity of the hearing loss. A study of loudness summation in impaired ears with an average hearing loss of 65 dB revealed a much wider critical band than normal. Loudness did not

---

6. See Zwicker, "Critical bandwidth in loudness summation", *J. Acoust. Soc. Am.*

change with ΔF when ΔF was increased up to seven times the normal-sized critical band.

Regarding the mechanisms underlying loudness recruitment in impaired ears, abnormally steep rate-intensity functions are observed for auditory nerve fibers from ears with cochlear impairment. The dynamic range, defined as the difference between the neuron's threshold and its saturation intensity, was between 5-20 dB SPL in impaired ears, in contrast to dynamic ranges of 10--50 dB SPL in normal ears. Recruitment may be due to the loss of neurons that respond to low intensity, while little or no loss of sensitivity occurs in neurons that detect more intense stimuli. As a result, loudness grows at an abnormally rapid rate in the impaired ear from near-threshold levels, where sensitive, low-threshold neurons are defective, to high intensities, where the operation of high-threshold neurons, including those tuned to other frequencies, remains normal. Recruitment is a consequence of an abrupt increase in the firing rate of the collective auditory fibers as stimulus intensity increases. In this model, the spread of excitation to stimulate adjacent nerve fibers at high intensities is critical.

Models of loudness

Two basic types of loudness model have been described in the literature. In one type, the loudness is calculated simply from the overall intensity of the stimulus. The original power-law model of Stevens, and subsequent variations, are examples of this approach. As noted, however, changes in the bandwidth of a stimulus can affect the loudness of that sound independent of the overall intensity. Consequently, other models based on excitation patterns have been developed in an attempt to provide a more complete accounting of loudness.

Modified Power-law Model

The modified power-law model[7] Eq. (13.5.3), originally used as a model of masking additivity in normal and hearing-impaired listeners, has been extended to describe the growth and summation of loudness for normal and hearing-impaired listeners. Previous studies on masking additivity and loudness suggest that the best-fit value for p is between 0.1 to 0.4 for both normal-hearing and hearing-impaired listeners. Note that instead of utilizing the excitation-pattern of a stimulus, this model considers only the overall intensity of the stimulus, together with the threshold intensity for that stimulus, in calculating the loudness.

Excitation-pattern Model

A series of critical bandwidth studies in loudness summation suggests that an underlying loudness pattern can be estimated from the psychoacoustical excita-

---

7. See Humes, "Models of the effects of threshold on loudness growth and summation", *J. Acoust. Soc. Am.*

tion patterns[8] derived from masking patterns. First, the model transforms a masking pattern, plotted as masked threshold (dB SPL) as a function of signal frequency, to an excitation pattern plotted as the excitation level (*LE*) in dB versus critical-band rate in Barks. A Bark (named after the German acoustician Barkhausen) is a unit of tonality that corresponds numerically to the critical band after the entire frequency range has been partitioned into 24 contiguous critical bands. The excitation-pattern model also takes into account the middle-ear transmission characteristics and applies a correction factor to the frequency range above 2kHz. According to the model, the extent of upward spread of excitation depends on the stimulus intensity. However, the spread of excitation to low frequencies remains constant at all intensity levels.

Once the excitation pattern has been determined for a particular input spectrum, the next step is to generate a specific-loudness pattern. This is accomplished by converting the excitation level (*LE*) in dB SPL, at each critical band, to a specific loudness (N') in Sones/Bark. The transform takes this form

$$N' = 0.08\left(10^{\frac{LE_0}{10}}\right)^{0.23}\left[\left\{0.50\left(1.00+10^{\frac{LE-LE_0}{10}}\right)\right\}^{0.23}-1.00\right] \qquad (13.5.4)$$

LE_0 = *the excitation level associated with quiet threshold.*

The model states that the total loudness of a stimulus, in sones, corresponds to the area under the specific-loudness pattern for that stimulus. That is, the loudness, in sones, is the integral of the specific-loudness pattern, in Sones/Bark, across the Bark scale. Note that the total loudness can be of a single tone or of overlapping components across the Bark scale. The highest amplitude at each Bark unit is taken to calculate for the total loudness. Therefore, this model also accounts for the effects of partial masking by assuming that additional loudness occurs only when the excitation pattern of the stimuli exceed that of the background noise. The excitation pattern of a tone spread over a greater area on the high-frequency slope than on the low-frequency slope, the model suggests that at low and moderate level of the noise band, the loudness of the tone is reduced more by a band of noise located above, rather than below its frequency.

The excitation-pattern model of loudness has been successfully applied to individuals with conductive impairment and simulated hearing loss at moderate to high intensity range. Furthermore, an extension of the model has been applied successfully to intensity discrimination.

Masking additivity and loudness summation

Temporal effects in simultaneous masking are studied by measuring the amount of masking created by a brief signal presented at various temporal positions within a masker. For noise maskers and tonal signals, there is generally a

8. See Zwicker, *Psychoacoustics-Facts and Models.*

small temporal effect. Under nonoverlapping conditions, the combination of two equally effective maskers can result in 10-15dB more masking than either masker alone. It is well known that two equally effective broadband noises or two spectrally overlapping narrow-band maskers of the same level will produce a threshold in combination which is only 3dB greater than that produced by either masker alone. This 3dB increase is consistent with linear power summation of masker energy within critical bands centered at the signal frequency. It is important to note that rules governing the additivity of masking are different for spectrally overlapping and nonoveralapping simultaneous maskers. Linear additivity of masking appears to be restricted to conditions involving spectral and temporal overlap of the two maskers. A study reported by Humes and Lee[9] suggested that when two maskers overlap within the critical band centered at the signal frequency, linear additivity of masking results. Conversely, when there is no such overlap, then nonlinear additivity of masking occurs. This explanation of the additivity of simultaneous masking is also consistent with the literature on the monaural additivity or summation of loudness. When two equal-intensity tones fall within the same critical band, their combined loudness is equivalent to that resulting from an increase in the intensity of either tone by 3dB. In other words, linear power summation occurs for components within a single critical band. On the other hand, for two tones falling in separate critical bands, the loudness adds, rather than the intensity. Thus if two equally loud tones separated by at least one critical band are presented together, the combined loudness is twice that of either individual component. A two fold increase in loudness would correspond to increasing the intensity of either component presented individually by about 10dB. This modified power law model can also be used in a excitation framework.

Studies have shown that when two simultaneous maskers overlap in the spectral domain, the additivity of masking can be described by a linear model. However, when the maskers are separated spectrally, a nonlinear additivity of masking is apparent. The linear effect can be demonstrated by the classic linear energy summation model, which predicts a 3dB increment in masked threshold when two equated maskers are presented simultaneously. The nonlinear additivity is perhaps best described by a modified power law model.

13.6 Binaural Hearing

In everyday situations, we listen with two ears. It is well documented that the difference limen decreases with binaural hearing because of loudness summation. Perfect binaural summation provides a 3dB increase in loudness perception. This advantage is often interpreted as doubling the power, that is, $10\log(2)=3$dB.

9. See Humes, "Two Experiments On The Spectral Boundary Conditions For Nonlinear Additivity Of Simultaneous Masking", *J. Acoust. Soc. Am.*

Generally speaking, binaural hearing is pertinent to our understanding of localization.

In brief, we can localize a sound source because we listen with two ears. We detect the origin of a sound source by comparing the auditory information we obtain between our ears. Duplex theory of localization proposes that low frequencies signals are detected by interaural time differences (ITD), and high frequencies signals are detected by interaural intensity differences (IID). The distinction of low versus high frequency is referenced base on the dimension of the head. A dividing frequency is approximately 1.5 kHz for a head diameter of approximately 8.5 in – the range above is considered as high frequency and the range below is as low frequency.

ITD refers to the difference in time for the signal to reach the left and right ears. This time difference, together with the interaural phase difference (IPD) provides pertinent low frequency localization information. Without the critical IPD information, temporal confusion would occur and hinder our ability to localize sound source. Assume that a signal approaches from the left side (90 degrees) in the horizontal plane. The stimulus will arrive at the left ear before it reaches the right ear, hence creating a time difference between the two ears. As the signal moves from the left (90 degrees) to the front (0 degrees), or to the back (180 degrees), the ITD is indistinguishable, that is, the signal arrives at both ears at the same time if presented at either location. Hence, it is difficult for us to differentiate sound sources in the median plane based solely on ITD. In a reverberant room, studies have demonstrated that our ears can systematically identify the first wavefront that arrives at our ears and localize the sound source – rejecting the confusion caused by the many reflected secondary signals. This phenomenon is known as the precedence effect or the law of the first wavefront.

IID refers to the difference in signal level between the two ears. In Fig. 13-9 we have shown a simulation of the Head Related Transfer Functions (HRTF) which we calculated in Chap. 11. Due to the finite number of modes in this calculation the results are only valid up to the dotted line shown in this figure. The figure clearly shows that for frequencies above 1.5 kHz, there is a significant head shadow effect which results in an increase in IID. Zero degrees in this plot refers to the normal axis to one of the ears. At low frequencies, this phenomenon disappears because of diffraction of the wave around the head – diminishing IID.

We can see that localization at low frequencies is dependent on the ITD and the IID at high frequencies. At very low frequencies the IPD is so small that good localization becomes impossible.

13.7 Summary

In this chapter, we summarized a few of the main psychoacoustics concepts critical to the understanding of this book. First, we reviewed the anatomy and physiology of the auditory system mostly to give a physical understanding for the importance of traveling wave theory and excitation pattern. The dominant effect,

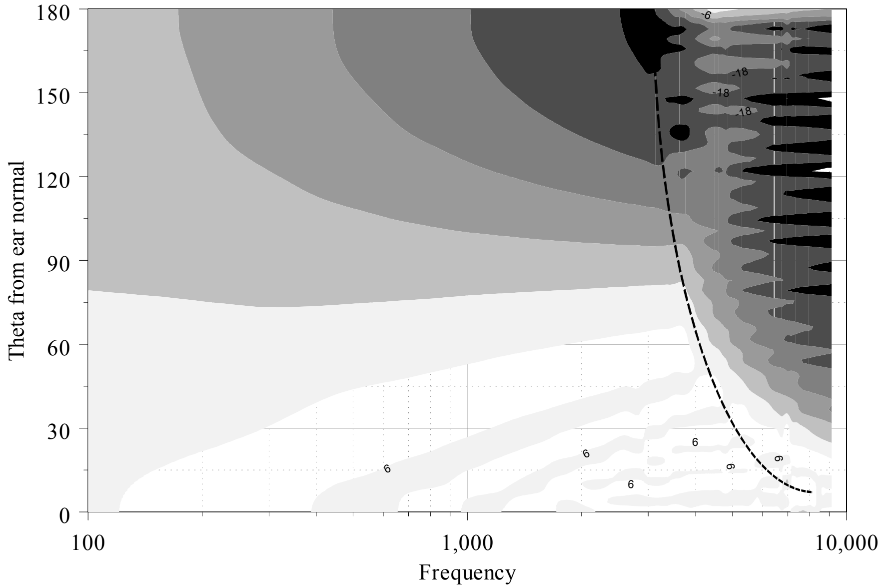

Figure 13-9 - Spherical model of head shadow effect in IID (result valid to dotted line)

as regards our subject matter, is masking. Masking has a clear foundation in the physical aspects of the function of the ear. Aspects of loudness perception, temporal and spectral masking effect, and binaural hearing were also reviewed. For a more comprehensive understanding of each of the topics, we recommend Moore (1986) and Zwicker and Fastl (1987).