CHAPTER 2 HEARING AND PSYCHOACOUSTICS

WITH LIDIA LEE

I would like to lead off the specific audio discussions with a description of the audio receptor—the ear. I believe it is always a good idea to understand some basics of what these receptors are capable of, but equally or more importantly, what they are not capable of, before discussing design details of systems that utilize these receptors as their target.

2.1 The Ear

In this chapter, the key concepts in psychoacoustics that are critical to later chapters will be discussed. Understanding of psychoacoustics requires a basic understanding of the theory of hearing. Hearing is an incredibly complex multidimensional system which is composed of acoustical, mechanical, hydrodynamic and neurological subsystems all acting basically in series. Each one of these subsystems affects the perception of sound, but some more than others. Figure 2-1 shows a schematic drawing of the ear where three prominent subsystems called the **outer ear**, the **middle ear** and the **inner ear** are seen.

Outer, Middle, Inner Ear the three fundamental subdivisions of the ear.

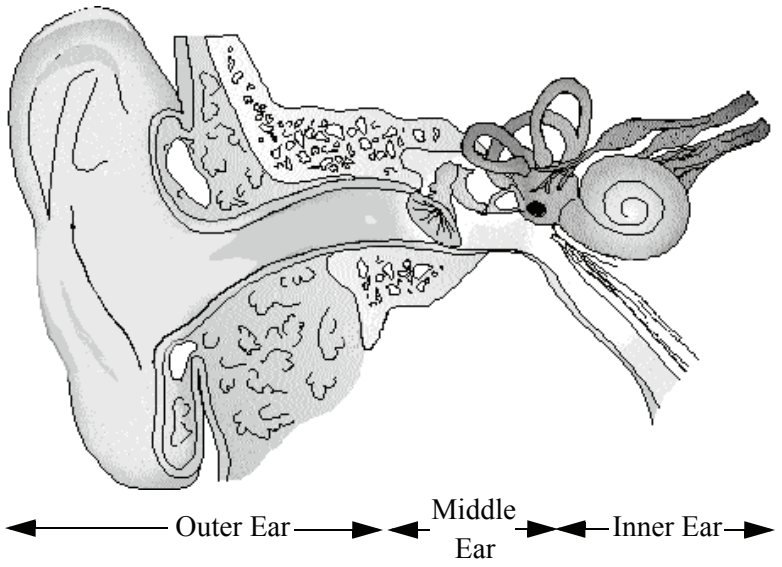

Figure 2-1. Schematic drawing of the ear system with three subsystems

2.1.a The Outer Ear

Binaural
the use of two ears.

ITD
the time difference between sound arrivals at the ears.

IID
the intensity difference between the sound at the ears.

The effect of the **outer ear** is primarily an acoustical one. Combined with the head, the outer ear forms a spatial acoustic filter which acts on sounds coming from different directions in different ways. The **binaural** hearing system (mostly responsible for sound localization) is composed of two ears, the signals coming from these two receivers and the signal processing that takes place in the brain. Sound localization is a function of two basic characteristics of the ears. The first is the time delay from one ear to the other which is called the **interaural time difference (ITD)**. There is also a level difference between the ears due to the head shadow effect which results in an **interaural intensity difference (IID)**. The brain acts on these two signal parameters to determine the location of a sound source, and it does so in slightly different ways depending on the frequency.

The Duplex Theory of localization proposes that low frequency signals are detected by ITD, and high frequency signals are detected by IID. The distinction of low versus high frequency is determined by the dimension of the head where the dividing frequency is approximately 1 kHz for a head

diameter of approximately 8.5" —the range above is considered as high frequency and the range below is considered as low frequency.

IPD
the phase difference between the sound arrival at the ears.

The ITD together with the **interaural phase difference (IPD)** provides pertinent low frequency localization information. Without the critical IPD information, temporal confusion would occur and hinder the ability to localize low frequency sound sources. Assume that a signal approaches from the left side (270°) 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 (270°) to the front (0°), or to the back (180°), 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 localize sound sources at many locations based solely on ITD. In a reverberant room, studies have demonstrated that a person's ears can systematically identify the first wavefront that arrives at the 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. It should be noted, however, that nothing is implied about tone color changes or increased source location confusion that may occur with these reflections, only that the sources location is principally controlled by the first arrival. This later aspect of the precedence effect is often misunderstood.

The IID refers to the difference in signal level between the two ears. It is created by the diffraction effects of the human head at mid frequencies and the pinna effects at high frequencies. Sound localization is dependent on the ITD at low frequencies and the IID at high frequencies. At low frequencies, there are only small differences between the signals at the two ears, no matter where the sound source is located, which results in small ITD value and poor localization at low frequencies.

Since the two ears lie in the horizontal plane, the interaural time and intensity differences for sound sources in this plane are at a maximum. For sources in the vertical plane the ITD and IID virtually vanish. This means that the sense of localization is greatest in the horizontal plane (just as we might have expected of nature since the world lies mostly in this plane).

Thus, it can be seen that the acoustic filters of the outer ear accounts for much of one's localization capabilities, but the brain does do the final determination.

2.1.b The Middle Ear

Cochlea
a spiral organ lined with receptors, fluid filed, which is the main detector of sound.

The next subsystem in the chain is primarily mechanical consisting of the ear drum and small bones connected to another small membrane that is attached to the fluid-filled **cochlea**. The principle function of the middle ear is to transform the small acoustic pressure vibrations in the ear canal into a fluid motion in the cochlea. The bone structure of the middle ear can also act as a slow reaction time sound compressor that results from changes in the flexibility of the system with sound level. The three bones in the middle ear (the smallest in the body) are held in place by muscles and constitute a mechanical lever. At high sound levels these muscles tighten, thereby stiffening the lever. This effect is used as a protection device against sustained loud sounds, and results in the occurrence of the temporary threshold shift that we have all experienced after being subjected to loud sounds for a sustained period. Unfortunately, this protective mechanism cannot react fast enough to protect from sharp impulsive sounds like gun shots, etc. This is why these sounds are the most damaging to the hearing system. No doubt, the lack of naturally occurring impulsive sounds is the reason for this lack of protection. Virtually all impulsive sounds in the environment are man made (thunder being one notable exception), and, unfortunately, they are abundant in the current environment.

2.1.c The Inner Ear

The inner ear, which is principally hydrodynamic in its action, is by far the most complex system that will be discussed in this text. Even today its function is not completely understood. (Just imagine how difficult testing this system must be! It is extremely difficult to test directly on live subjects and its function deteriorates within minutes of death.)

Basilar membrane
a membrane in the cochlea which contains the principle hair cells for sound detection.

The inner ear starts with the **oval window** at one end of the cochlea which is excited by the middle ear's mechanical vibrations. This excitation creates a wave that propagates into the fluid-filed cochlea. The cochlea consists of a long tapered tube which is divided in half by a membrane and coiled up in a helix. The dividing membrane is called the **basilar membrane** and its motion triggers small hair cells embedded in it which activate nerve impulses that travel to the brain. The exact action of this motion and the nature of how the neurons fire is far too complex to get into in detail here;

however, an overview of the general theory of how wave motion on the cochlea is translated into the perception of sound will be discussed.

2.2 Theories of Hearing

2.2.a Place Theory

Place Theory is considered to be the most well established to date. This theory suggests that the cochlea is partitioned into different segments along its length. There are thousands of these segments lying along the length of the cochlea, which, because of its constantly changing size and shape, forms a sort of place-tuned resonant system. Each segment thus responds to a certain limited band of frequencies. As a complex signal reaches the cochlea, a form of mechanical frequency analysis takes place as the waveform travels along the cochlea. This analysis is similar to a Fourier analysis, however significant difference exist. It is thus erroneous to think of the cochlea as a Fourier Analyzer. The different components of a complex signal excite the cochlea to a maximum amplitude of vibration at different points along the cochlea depending on the frequency of these components of the signal. The cochlea senses higher frequencies at its input end and lower ones at the opposite end. Thus, the high frequency hair cells are excited by all frequencies—a reason why one tends to loose the high frequency hair cells first, which explains the domination of high frequency hearing loss in the general population.

2.2.b Frequency Theory

Frequency Theory is based on the assumption that the auditory nerve fibers fire at a rate which is synchronous with the input signal, transmitting this information to the brain. For example, a 100 Hz signal is detected by nerve firings at a rate of 100 times a second. The intensity of the sound is coded by the number of impulses in each volley. This theory is adequate at explaining the perception of low frequencies, however, above about 1 kHz it becomes untenable due to a nerves finite recharge rate. The recharge rate is the time it takes an auditory nerve cell to re-establish its polarization in order to fire again and is typically about 1 ms long (which corresponds to

1 kHz). Hence, above 1 kHz, the neurons are simply unable to fire synchronously with the input signal. They cannot keep up, and Place Theory becomes the dominate perception mechanism.

A look at Figure 2-2 shows some interesting results of these two theories. This figure has several curves, each of which represents an equal perceived loudness set by their value at 1 kHz. Above 1 kHz, we see a basically flat response which Place Theory would suggest. There are of course various resonances in this response due to the acoustic resonances of the outer ear structure. The dip at about 3 kHz is caused by the ear-canal resonance and a second resonance occurs at about 14 kHz.

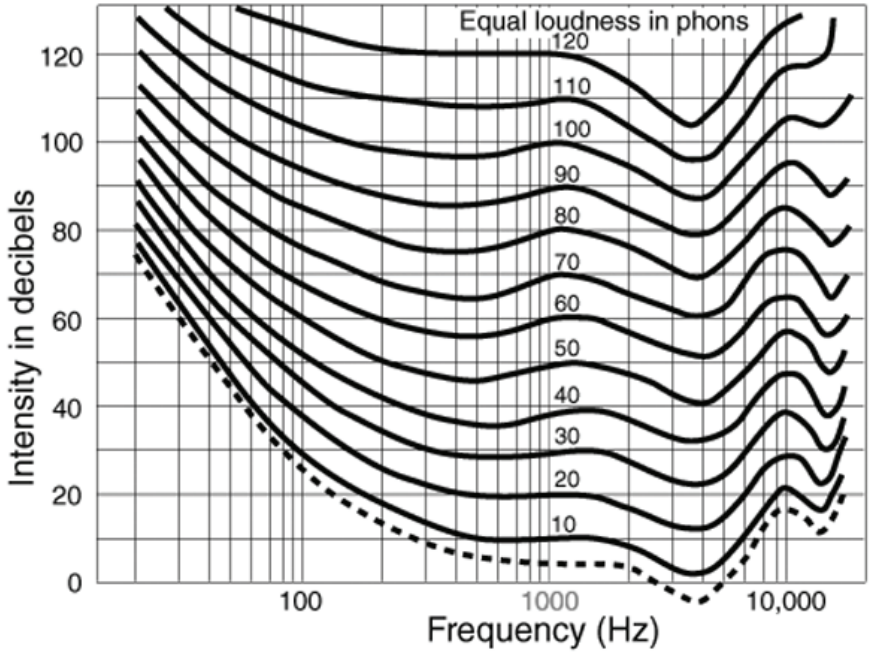

Figure 2-2. Contours of equal perceived loudness.

Below about 500 Hz, there is a direct result of Frequency Theory where the perception of level is strongly dependent on the frequency and the input level, but less so in the later case. This is easy to explain with Frequency Theory. Consider that; as frequency drops below about 500 Hz, there are less and less nerve firing (they are synchronous with the frequency) and since the loudness is perceived by the number of impulses, the perception decreases with frequency (a rising curve in Fig. 2-2). Further, as the level

increases there are more nerve impulses per volley (because of the higher hair cell excitation). Thus, the curve flattens out at higher levels of low frequency simply because the perception of loudness is proportional to the number of nerve firings in any given period.

2.2.c Place-Frequency Theory

The prevailing theory of hearing is called **Place-Frequency theory**. This theory combines the theories of Place and Frequency into one. At low frequencies the auditory system does an actual time-domain sampling of the signal (Frequency Theory), while at higher frequencies the location of maximum amplitude on the cochlea is the primary detector (Place Theory).

2.2.d Masking

The next concept in hearing that will be discussed is best visualized by considering the Place Theory description of wave motion on the cochlea as shown in Figure 2-3 on the next page. This figure is a graphic representation of how a wave travels along the basilar membrane inside of the cochlea. The solid line represents the envelope of the wave motion and the dotted lines are instantaneous wave amplitudes. It is readily apparent that for frequencies above a few hundred Hertz, there is a strong place or location component to the amplitude of excitation. At low frequencies, this place aspect is becoming less apparent and the frequency theory component of hearing takes over as the dominant factor in perception. Note that the frequency "place" on the cochlea moves towards the input end at high frequencies of excitation. In this figure, the frequency place is the opposite of what we are used to seeing where high frequencies are to the right.

The most important aspects of these curves are the tails in the envelope that extend both above and below the peak. The lagging tail, towards higher frequencies, always has a greater extension than the leading tail, to lower frequencies. Acoustic excitation that lies "inside" these tails is said to be masked. These frequencies are masked simply because the ear cannot perceive signals which lie within one of these tails. Masking theory is probably the single most important concept in psychoacoustics as it applies to audio. It will show in the next chapter that masking accounts for most of the enormous data reductions that are now possible with **perceptual audio coding** techniques such as MP3 or WMA (see Chapter 3), and it also has a strong

Perceptual Coding
the use of perception as a basis for data reduction techniques.

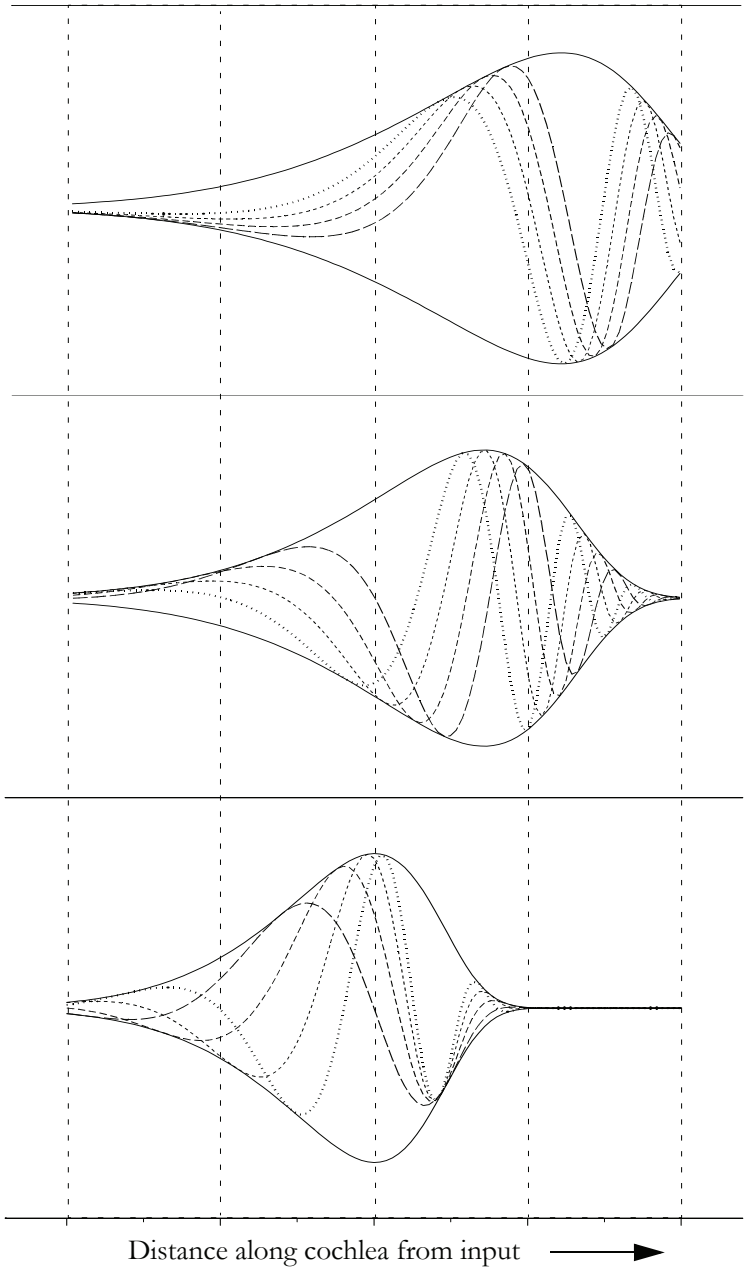

Figure 2-3. Graphic representation of cochlea traveling waves at different frequencies: 100 Hz (top), 400 Hz (middle), 800 Hz (bottom)

effect on the perception of distortion. Distortion products cannot be heard when they are masked.

Masking predominantly affects higher frequencies of the original sequence and the term "upward spread of masking" is usually used. Masking is also level dependent, which is not apparent from Figure 2-3. Figure 2-4 shows measurements of the masking curves for various signal levels of a 500 Hz tone (Denoted as the numbers on the curves 20–80 dB). A signal that lies below the pertinent masking curve in this plot would not be audible. Note that the masking spreads upward in frequency to a greater extent at higher levels and that the masking curves get wider at higher levels. This effect also impacts the perception of distortion since distortion by-products tend to be upward in frequency and will be masked to a great extent.

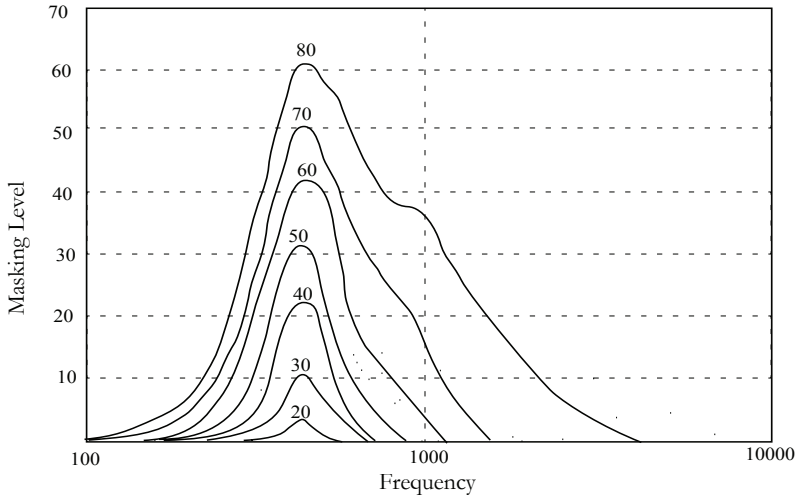

Figure 2-4. Graphic representation of masking patterns of a fixed probe frequency. Each contour represents a different probe intensity.

For a simple single-tone excitation, it has been shown that the perceived loudness is a complicated function of the frequency and level of the excitation. More importantly, the perceived loudness of a tone in the presence of other tones can vary from normal to completely inaudible depending on the level and location of the other tones. The net result is that the perception of a complex pattern of sound excitation can be difficult to analyze and understand. It should come as no surprise that simple acoustic measurements of sound levels are a long way from giving an accurate description of how a sound system is perceived.

Since the waveform on the cochlea takes a finite amount of time to build up and decay, there will also be a temporal masking effect, but this effect is not as great as frequency masking. More important, in the temporal domain, the ear has a finite integration time within which discrete temporal events are fused together into a single excitation pattern. This later effect is of fundamental importance in small room acoustics.

2.2.e Critical Bands

A direct result of the finite width of the wave-packet motion on the cochlea, as shown in Figure 2-3, is that the ear has a finite capability to resolve signals in the frequency domain. The extended width of the wave packet simply does not allow for a precise resolution of signals with components that are close together in frequency. As a system, the ear acts as if it had a multiplicity of bandpass filters operating in parallel. The bandwidth of these filters is called the **Critical Band**. It must be stressed that the critical band filters themselves are not actually fixed in location, rather they are more like swept filters. It is convenient in discussions about psychoacoustics to consider the ear as having a frequency domain resolution that is equivalent to a set of about 30 bandpass filters covering the auditory spectrum.

Critical band
the effective bandwidth of an auditory filter.

Fused
the effect of two sounds arriving within a small time window being detected as a single sound.

The critical band is defined as the frequency bandwidth within which different frequency components of a signal integrate together—the term **fused** is sometimes used. Individual tones within a critical band are not perceived individually but as a single tone at a single level. In other words, two tones within a critical band will not have an additive loudness characteristic as two tones separated by more than a critical band will. The bandwidth of these critical band filters is frequency dependent as shown in Figure 2-5.

The critical-band concept is important because it implies that those portions of a signal that lie within a common critical band can be treated as a single composite signal. Each critical bands composite signal then interacts with the other critical bands of auditory filters as simple composite signals and not as complex signals. This concept will be shown to be necessary for the audio compression techniques described in the next chapter.

It is important to realize that even though we have a finite critical band resolution in the frequency domain, we can still resolve the frequency of a signal within a critical band to a resolution of about 5% of the bandwidth of

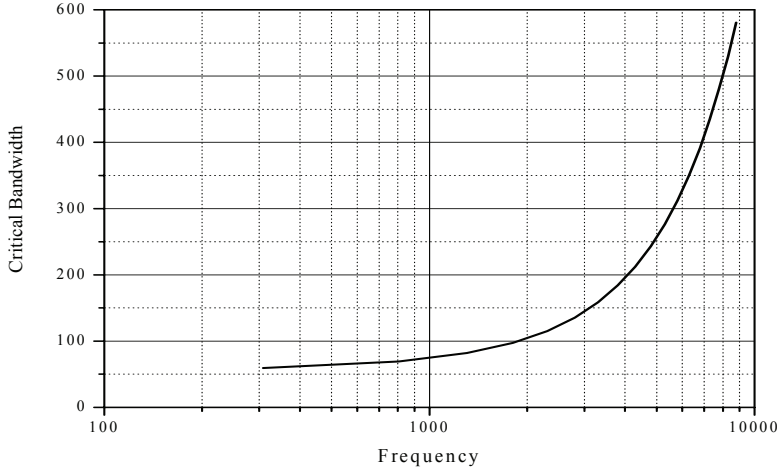

Figure 2-5. Critical bandwidth as a function of frequency.

the critical band. Thus, the critical band concept does not alter the perception of pitch and perfect pitch is quite possible. The critical band concept is only useful when talking about complex signals with a multiplicity of tones present. When two or more of these tones lie in a critical band, they must be treated differently than multiple tones that lie outside of a critical band. This is a particular instance where the Fourier analysis concept fails for the hearing mechanism.

2.3 Summary

To design optimum AV systems, one has to know what the target resolutions of human receptors are—in real terms—for it is unwise to spend time or money chasing after criteria which are not perceivable in the end product.

There are several main features of the human auditory system that you should have learned from this chapter.

- The ear has good spatial resolution in the horizontal plane at mid to high frequencies, but at low frequencies its spatial resolution is poor.
- The ear has a spatial resolution in the vertical plane that is poor because of the equal distance from each ear to points in this plane.
- The ear masks smaller signal components in the presence of larger signal components. This effect acts greater on signal components whose frequency is above the main signal component (the masker) than below it. The overall effect increases with the level of the masker.
- The ear also masks smaller signal in time making a small signal which follows a larger one in time inaudible.
- The ear has a finite frequency resolution which can be approximated as a bank of bandpass filters where the width of each filter is known as the critical band. A critical band's width varies with frequency.
- The ear fuses complex sounds that lie within a critical band into a single perceived tone.

These features will be used in the following chapters to define and quantify numerous aspects of signal compression, audio system sound quality and room acoustics.