physics,
:mains at
obvious
it about
1 wave is
:en have
and fre-
ely satis-

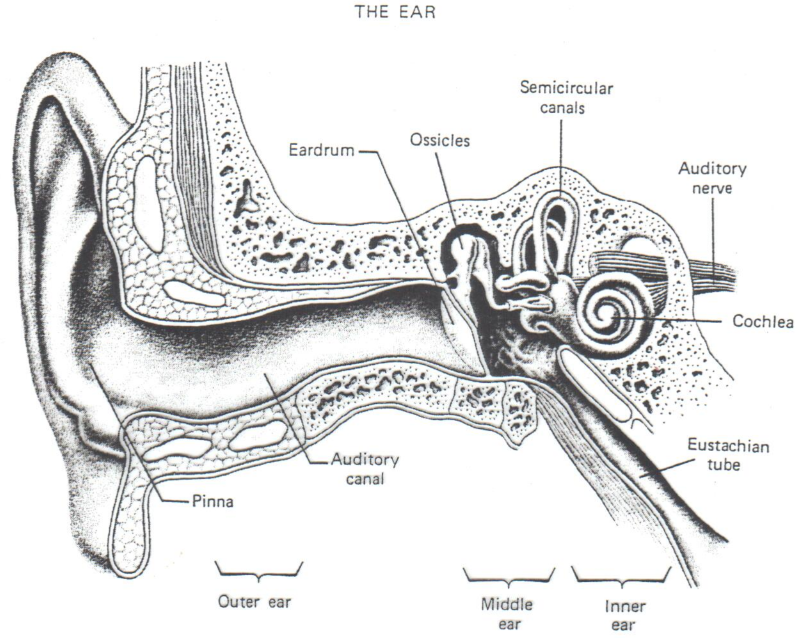
cceeding
in which
: shall be
it are re-
ɔ analyze
ıat enter

frequen-
to 20,000
resenting
e. Sound
)0 are
:.

ıic ⟶
High *f*

cies rang-
piano, A_0,
seven oc-
quency of
ontra bas-
cal instru-
later) that
es of har-

e between
', at atmo-
(The mid-
·.) When a
ıons in the
stablished
ın: A com-



The anatomical structure of the ear can be conveniently divided into three parts: the outer ear, the middle ear, and the inner ear. The outer ear consists of the external portion called the pinna, the auditory canal that is approximately 3 cm long, and the membrane at the inner end of the auditory canal called the eardrum. The middle ear begins just inside the eardrum and consists of a chain of three bones called the ossicles: the hammer, the anvil, and the stirrup. Opening into the middle ear from the throat is the Eustachian tube that permits equal pressures to be maintained on each side of the eardrum.

The stirrup links the anvil (on the middle-ear side) to the round window, which is the beginning of the inner ear. The inner ear is a liquid-filled, coiled-up cavity called the cochlea. If the cochlea were uncoiled, it would have a length of about 3.5 cm. Dividing the cochlea along its length is the basilar membrane. Hairlike cells line the basilar membrane, and these hair cells are "activated" in the perception process. (From Fig. 71, "Auditory Mechanism," in *Anatomy for Speech and Hearing,* 2nd ed., by John M. Palmer. Copyright © 1972 by John M. Palmer. By permission of Harper & Row, Publishers.)

pression forces the eardrum *in* and a rarefraction forces the eardrum *out*. The vibrations of the eardrum are transmitted by the bone structure of the middle ear to the inner ear where the stimuli are processed mechanically and neurologically. When these vibrations have a frequency in the range $15 \leq f \leq 20{,}000$, nerve impulses are initiated in the inner ear that

When we listen to a pure tone whose frequency and amplitude can be changed at will, we verify a correspondence between *pitch* and *frequency* and between *loudness* and *amplitude*. One has a fairly good idea on how the ear's primary frequency and amplitude detection mechanism works for pure sounds. In this chapter we only consider pitch.

The simple harmonic oscillations of the eardrum are transmitted by a chain of three tiny bones in the middle ear called "hammer," "anvil," and "stirrup" (or, in more erudite parlance, malleus, incus, and stapes, which means exactly the same thing in Latin) to the entrance ("oval window") of the inner ear proper (Fig. 2.6). The marble-sized *cochlea* is a tunnel spiralling as a snail shell through the human temporal bone. This cavity, shown in Fig. 2.6(b) in a highly simplified, stretched out version, is partitioned into two channels, the scala vestibuli and scala timpani, filled with an incompressible fluid, the *perilymph* (a direct filtrate of the cerebrospinal fluid). Both channels behave as one hydrodynamic system, because they are connected at the far end, or apex, by a small hole in the partition called *helicotrema*; the lower section is sealed off with an elastic membrane at the "round window" (Fig. 2.6(b)). The partition separating both scalae is in itself a highly structured duct of triangular cross section (also called scala media; Fig. 2.7(a)), filled with another fluid, the *endolymph*. Its boundaries are the *basilar membrane* which holds the sensory organ proper (*organ of Corti*), Reissner's membrane, which serves to separate endolymph from perilymph, and the rigid lateral wall of the cochlea.

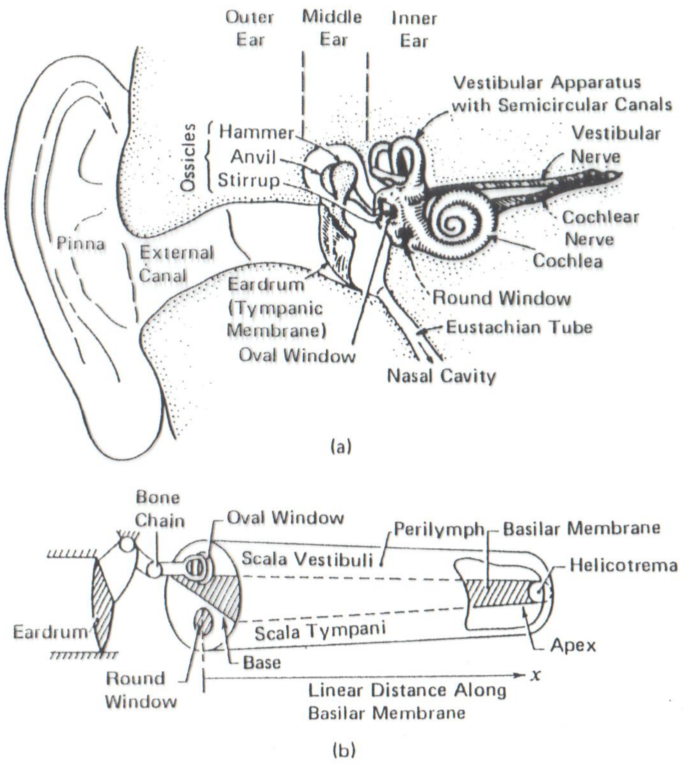


FIGURE 2.6 (a) Schematic view of the ear (Flanagan, 1972; Fig. 4) (not in scale); (b) the cochlea shown stretched out (highly simplified).

The elasticity of the basilar membrane determines the cochlea's basic hydromechanical properties. In the human adult the membrane is about 34 mm long from the base (the input end) to the apex; because of its gradual change in width and thickness there is a 10,000-fold decrease in stiffness from base to apex, which gives the basilar membrane its fundamental frequency analyzing function. Vibrations transmitted by the bone chain to the oval window are converted into pressure oscillations of the perilymph fluid in the scala vestibuli. The ensuing pressure differences across the cochlear partition between the two scalae flex the basilar membrane up and down setting it into motion like a waving flag; as this wave travels toward the apex, its amplitude increases to a maximum at a given place, which depends on the input frequency, and then dies down very quickly toward the apex. About 16,000 receptor units, called *hair*

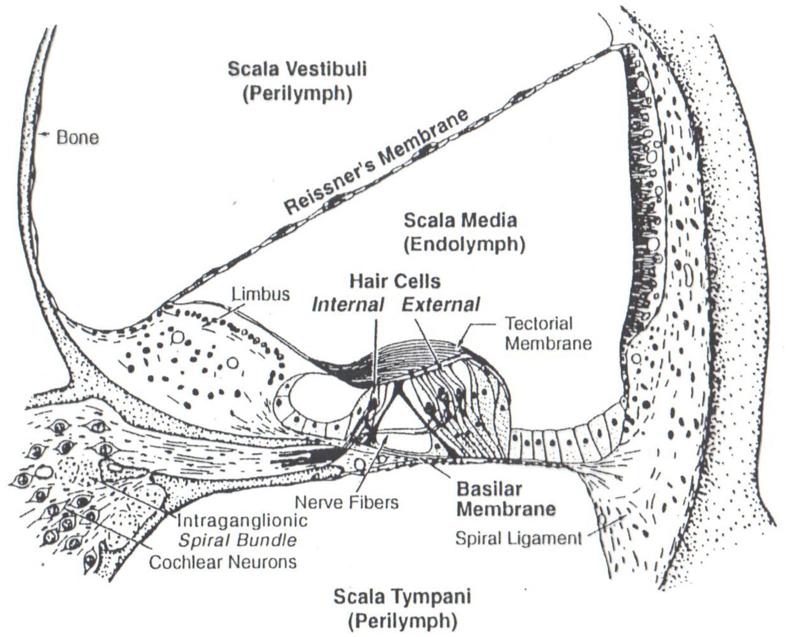


FIGURE 2.7 (a) Cross section of the organ of Corti. (After Davis, 1962.)

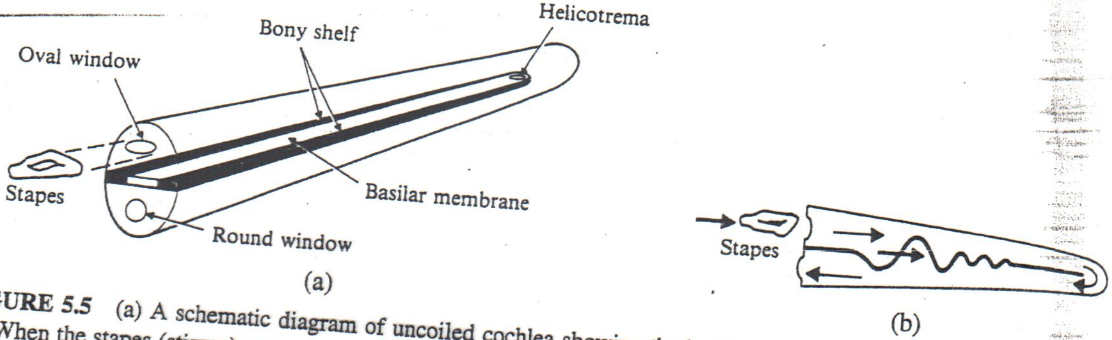


FIGURE 5.5 (a) A schematic diagram of uncoiled cochlea showing the basilar membrane and oval and round windows. (b) When the stapes (stirrup) presses against the oval window, a pressure pulse propagates through the cochlear fluid toward the round window, causing ripples to occur in the basilar membrane.

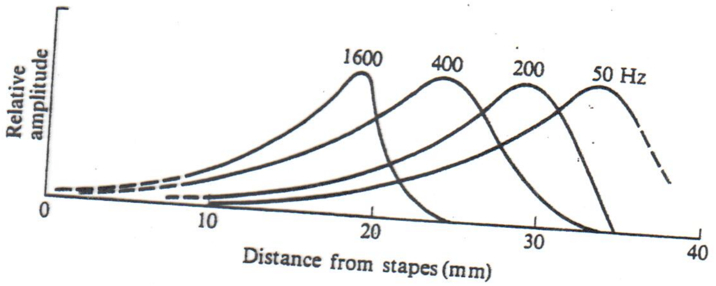


FIGURE 5.6 Basilar membrane displacement amplitude as a function of distance for several different frequencies. (After von Békésy 1960.)

of the sense of pitch is determined in the central nervous system, where the data from the auditory nerve are processed.

The conversion of the mechanical vibrations of the basilar membrane into electrical impulses in the auditory nerve is accomplished in the organ of Corti. When the basilar membrane vibrates, the "hairs" of the hair cells are bent, thus generating nerve impulses that travel to the brain. The impulse rate on the auditory nerve depends on both the intensity and the frequency of the sound.

The overall hearing mechanism is illustrated in Fig. 5.7. Sound waves propagate through the ear canal, excite the eardrum, and cause mechanical vibrations in the middle ear. The

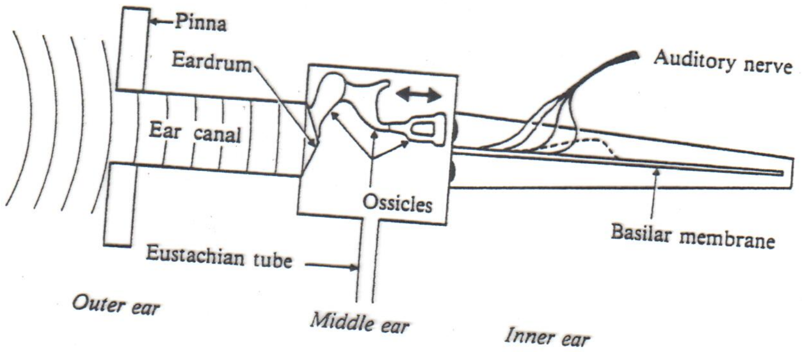


FIGURE 5.7 A schematic representation of the ear, illustrating the overall hearing mechanism. Sound waves in the outer ear cause mechanical vibrations in the middle ear, and eventually nerve impulses that travel to the brain to be interpreted as sound.

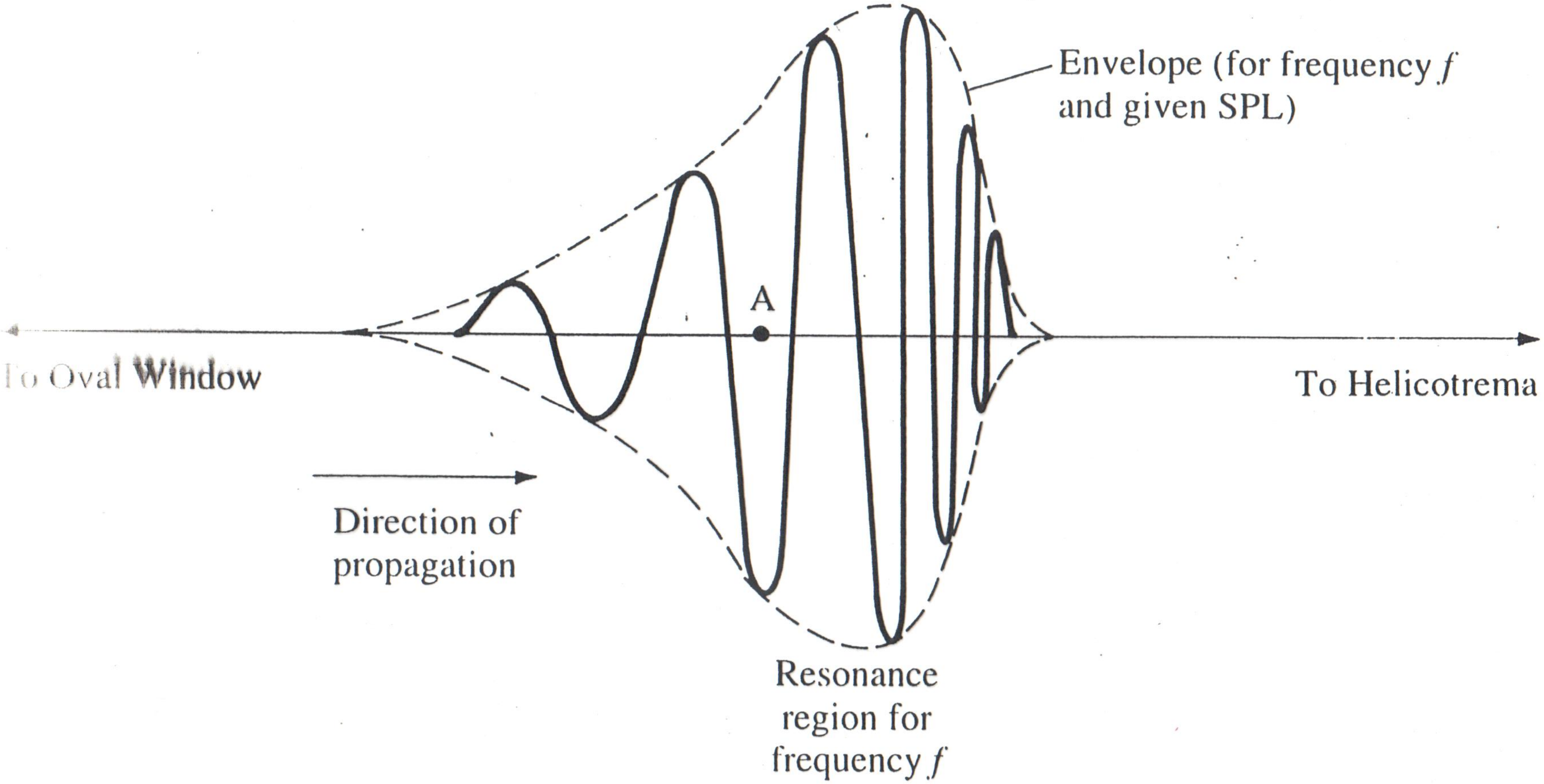


FIGURE 3.5 Sketch of a traveling wave along the basilar membrane, generated by a single-frequency tone. Full curve: snapshot of transverse displacements of the membrane (not in scale!). Picture in your mind this curve traveling within the broken lines toward the right and slowing down as its amplitude dies down on the right. Broken curve: amplitude envelope (which remains fixed unless the frequency and/or amplitude of the tone change).

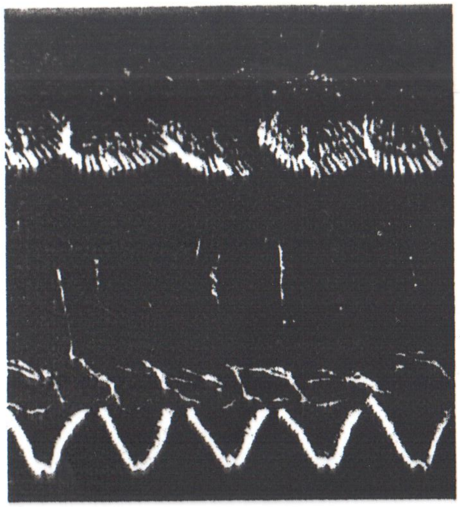


FIGURE 2.7 (b) Scanning electron micrograph (Bredberg et al., 1970) of the stereocilia of inner row (*top*) and outer row (*bottom*—only one of three rows in shown) hair cells on the basilar membrane of a guinea pig. (These animals, as well as chinchillas and cats have peripheral acoustic systems very similar to humans and are the laboratory animals most frequently used in hearing research.)

cells, arranged in one "inner" row and three "outer" rows along the basilar membrane, pick up the motions of the latter and impart signals to the nerve cells, or *neurons*, that are in contact with them. The name "hair cell" comes from the fact that at its top there is a bundle of 20–300 tiny processes called *stereocilia* (Fig. 2.7(b)) protruding into the endolymphatic fluid, whose deflection triggers a chain of electrochemical processes in the hair cell and its surroundings that culminate in the generation of electrical signals in the acoustic nerve. The *tectorial membrane* is a gelatinous tissue suspended in the endolymph above the organ of Corti (Fig. 2.7(a)), into which the cilia of outer hair cells are inserted; it plays a key role in stimulating and receiving the motile action of the latter. We shall return to the cochlear function in much more detail in Secs. 2.8 and 3.6.

The remarkable fact is that for a *pure tone* of given frequency, the maximum basilar membrane oscillations occur only in a limited region of the membrane, *whose position depends on the frequency of the tone*. In other words, for each frequency there is a region of maximum stimulation, or "resonance region," on the basilar membrane. The lower the frequency of the tone, the closer to the apex (Fig. 2.6(b)) lies the region of activated hair cells (where the membrane is most flexible). The higher the frequency, the closer to the entrance (oval window) it is located (where the membrane is stiffest). *The spatial position x along the basilar membrane* (Fig. 2.6(b)) *of the responding hair cells and associated neurons determines the primary sensation of pitch* (also called spectral pitch). A change in frequency of the pure tone causes a shift of the position of the activated region; this shift is then interpreted as a change in pitch. We say that the primary information on tone frequency is "coded" by the sensorial organ of the basilar membrane in the form of *spatial location* of the activated neurons. Depending on which group of neural fibers is activated, the pitch will appear to us as low or high.

Figure 2.8 shows how the position x (measured from the base, Fig. 2.6(b)) of the region of maximum sensitivity varies with the frequency of a pure, sinusoidal tone, for an average adult person (von Békésy, 1960). Several important conclusions can be drawn. First of all, note that the musically most important range of frequencies (approximately 20–4000 Hz) covers roughly two-thirds of the extension of the basilar membrane (12–35 mm from the base). The large remaining portion of the frequency scale (4000–16,000 Hz, not shown beyond 5000 Hz in Fig. 2.8) is squeezed into the remaining one-third. Second, notice the significant fact that whenever the frequency of a tone is doubled, that is, the pitch jumps one

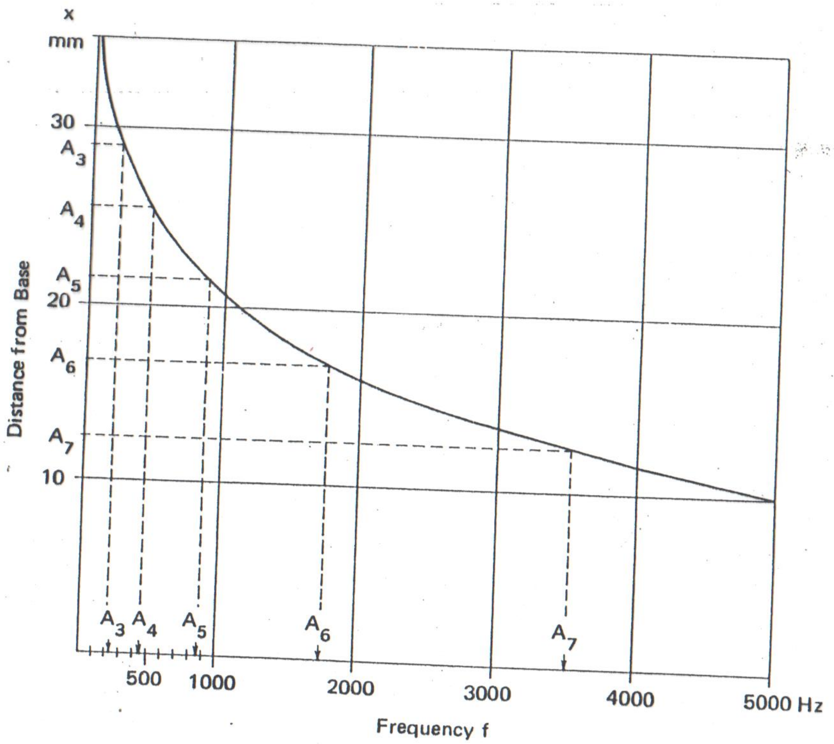


FIGURE 2.8 Position of the resonance maximum on the basilar membrane (after von Békésy, 1960) for a pure tone of frequency *f* (linear scales).

ésy, can be seen and

es he developed for
roscope with micro-
silar membrane. The
a suspension of pow-
e suspended powder,
cochlea was excited

amusing story is told
the Budapest zoo. He
recover the elephant's
ble also in the basilar

vo parts: that done in
the auditory nervous
rst transforming it int
in Figs. 5.5 and 5.6,
ted by the auditory ne
ory nervous system.
nerve, to pick up the e
m the cochlea to the b
each spike correspon
e. The spikes are foun

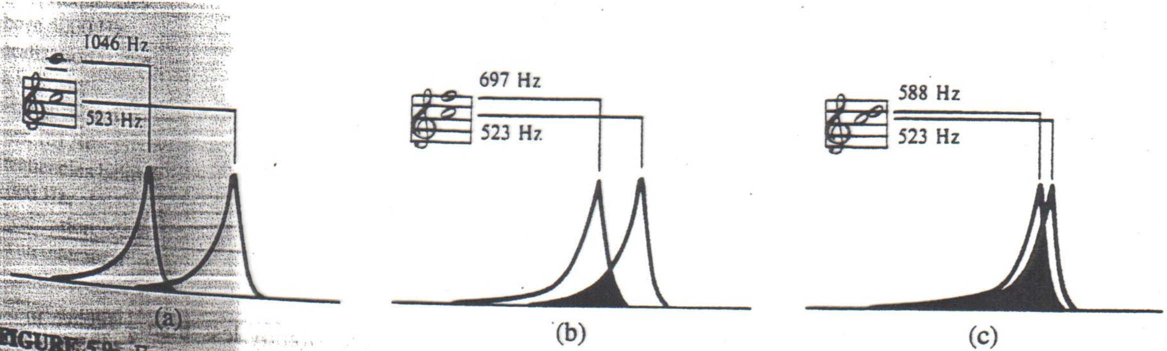
---

be closely correlated to the mechanical vibration pattern on the basilar membrane up to frequencies of about 4000 or 5000 Hz.

Each auditory nerve fiber responds over a certain range of frequency and sound pressure. Each nerve fiber has a characteristic frequency (CF) at which it has maximum sensitivity. Fibers with a high CF show a rapid rolloff in sensitivity above their CF but a long "tail" below it. A 90-dB stimulus at 500 Hz, for example, causes spikes to appear on all six fibers. By sophisticated techniques such as probing with laser light (Khanna and Leonard 1982) and using the Mössbauer effect (Johnstone and Boyle 1967), it has been found that basilar membrane displacements in live animals show a much sharper frequency response than those of Fig. 5.6 in the cochlea of a dead animal. Rhode and Robles (1974) found that within several hours after death, the basilar membrane response decreases 10–15 dB, the frequency of maximum response shifts downward, and the response curve broadens. In fact, the mechanical frequency response of the basilar membrane in live cochleas is quite comparable to the tuning curves observed in nerve fibers. There is some evidence for sharpening of neural tuning curves further along the neurological pathway, however.

If we were to observe the spikes on a nerve fiber when the stimulus is a tone of a single frequency, we would note that the time between spikes almost always corresponds to one or two or more periods of the tone. Although the nerve fiber does not fire at the peak of every vibration cycle in the basilar membrane, it rarely fires at any other time. The situation is a little more complicated when the stimulus is a complex tone, but still we find that the pattern of spikes on the auditory nerve carries accurate information about the frequency spectrum of the stimulus tone.

Consider a stimulus consisting of the pure tones C_4 (523 Hz) and C_5 (1046 Hz), spaced one octave apart. Their neural tuning curves (or frequency response curves) shown in Fig. 5.9(a) show very little overlap, so very few hair cells respond to both frequencies. Processing of the one component in the brain is only slightly affected by the presence of the other one.



(a) (b) (c)

FIGURE 5.9 Frequency response curves for pairs of pure tones. As the interval between them decreases, their response curves show increasing overlap.

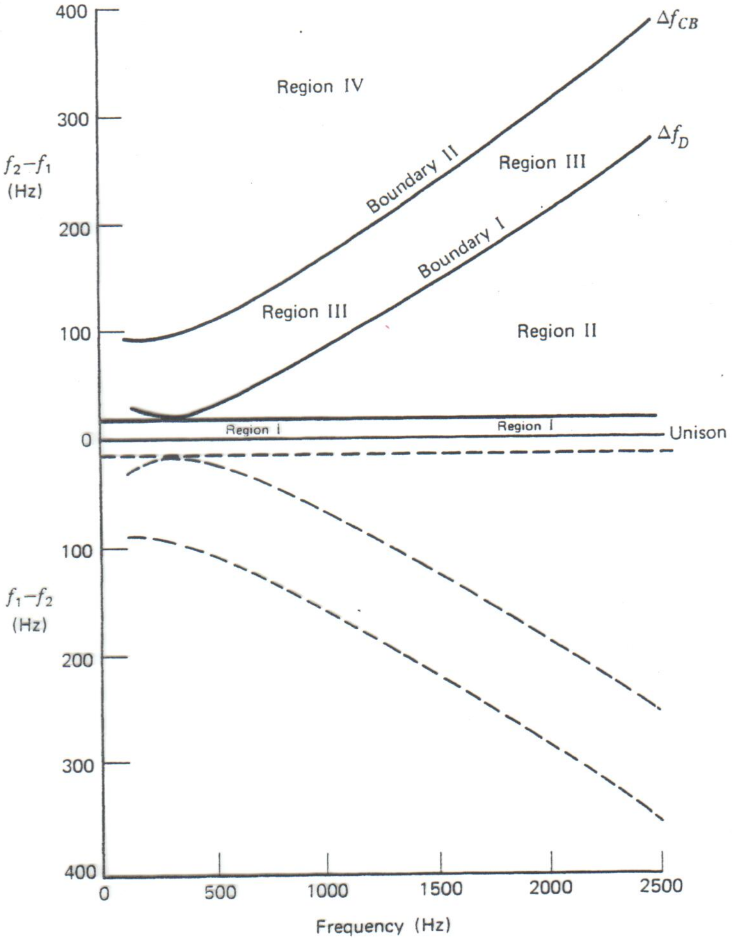


Figure 7.1
A graphical summary of the perceptual features of two pure tones heard together.

for them to be heard separately. Figure 7.1 shows all the regions discussed thus far as well as the boundary giving Δf_D.

Notice from the diagram that Δf_D changes with frequency. When f_1—the fixed frequency—equals 500 Hz, Δf_D is approximately equal to 30 Hz. Thus, f_2 must equal 530 Hz before f_1 and f_2 are heard as two separate tones. If f_2 equals 524 Hz, f_1 and f_2 will be perceived as a harsh fused tone having a frequency of 512 Hz. On the other hand, if f_1 = 2000 Hz, then $\Delta f_D = 200$ Hz; therefore, f_2 must equal 2200 Hz if f_1 and f_2 are to be heard individually. **FFT 3**

Below Boundary I, two tones are perceived as a fused tone. At or above Boundary I, both tones are heard separately. Roughness still persists.

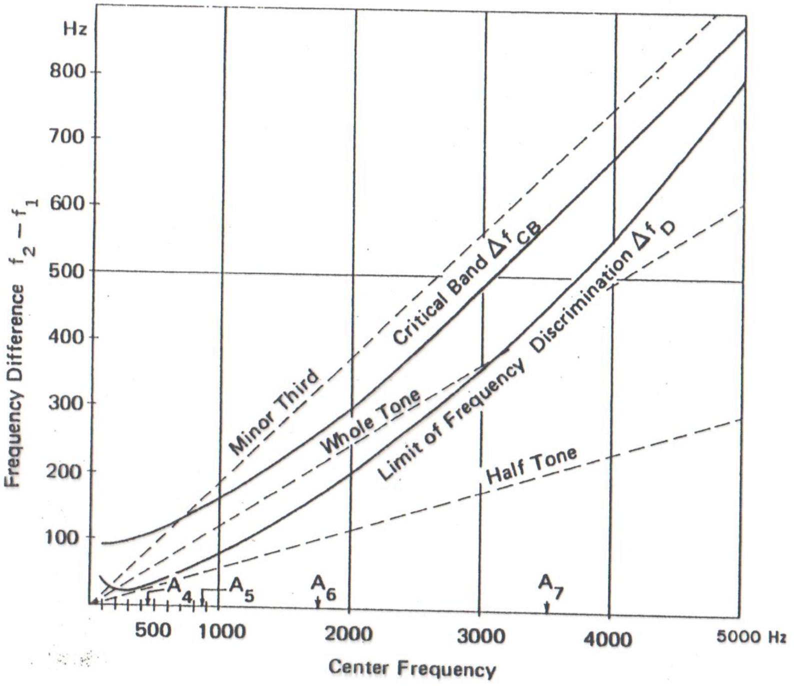


FIGURE 2.13 Critical bandwidth Δf_{CB} (after Zwicker, Flottorp, and Stevens, 1957) and limit of frequency discrimination δf_D (Plomp, 1964) as a function of the center frequency of a two-tone stimulus (linear scales). The frequency difference corresponding to three musical intervals is shown for comparison.

discrimination and critical band are larger than a minor third (and even a major third). This is why thirds in general are not used in the deep bass register!

Compare Fig. 2.13 with Fig. 2.9: the limit for frequency discrimination Δf_D is roughly 30 times larger than the JND for frequency resolution. In other words, we can detect very minute frequency changes of one *single* pure tone, but it takes an appreciable frequency difference between *two* pure tones sounding simultaneously, to hear out each component separately.[8]

What are the implications of these results for the theory of hearing? The existence of a finite limit for tone discrimination is an indication that *the activated region on the basilar membrane corresponding to a pure tone must have a finite spatial extension*. Otherwise, if it were perfectly "sharp," two superposed tones would always be heard as two separate tones as long as their frequencies differed from each other—no matter how small that difference—and no beat sensation would ever arise. Actually, the fact that the roughness sensation persists even beyond the discrimination limit, is an indication that the two activated regions still overlap or interact to a certain degree, at least until the critical band frequency difference is reached. An illustrative experiment is the following: feeding each one of the two tones f_1 and f_2 *dichotically* into a different ear, the primary beat or roughness sensation disappears at once, both tones can be discriminated even if the frequency difference is way below Δf_D, and their combined effect sounds smooth at all times! The moment we switch back to a monaural input, the beats or roughness come back. Of course, what happens in the dichotic case is that there is only *one* activated region on each basilar membrane with no chance for overlapping signals in the cochlea;[9] hence no beats or roughness.

At this stage the reader may wonder: if the region activated on the basilar membrane by *one* pure tone of *one* frequency is spatially spread, covering a certain finite range Δx along the membrane, how come we hear only *one* pitch and not a whole "smear" over all those pitches that would correspond to the different positions within Δx that have been activated? Unfortunately, we must defer the answer to later sections (e.g., Sec. 3.6). Let us just anticipate here that a so-called "*sharpening*" process takes place in which the activity collected along the whole region Δx is "focused" or "funneled" into a much more limited number of responding neurons.

The beat phenomenon plays an important role in music. Whenever beats occur, they are processed by the brain giving us sensations that may range from displeasing or irritating to pleasing or soothing, depending on the beat frequency and the musical circumstances under which they occur. The peculiar, displeasing sound of an instrument out-or-tune with

---

[8] There is an equivalent experiment that can be performed with the sense of *touch* to point out the difference between "resolution" and "discrimination." Ask somebody to touch the skin of your underarm for about one second on a fixed point with a pointed pencil *while you look away*. Then ask the person to repeat this at gradually displaced positions. It will require a certain small but finite minimum distance before you can tell that the position of touch has changed—this is the JND for localization of a single touch sensation, or "touch resolution." Now ask the person to use two pencils and determine how far both touching points must be from each other before you can identify *two* touch sensations. This is the minimum distance for "touch discrimination," which turns out to be considerably larger than the JND. Both touch resolution and discrimination vary along the different parts of the body. The equivalence between touch and hearing experiments is not at all casual: the basilar membrane is, from the point of view of biological evolution, a piece of epithelial tissue (skin) with an enormously magnified touch sensitivity! This analogy has been profusely used by von Békésy (1960) in his superb experiments.

[9] There is, however, an overlap of *neural signals* in the upper stages of the neural pathway, giving rise to "second order" effects, to be discussed in Secs. 2.6–2.9.