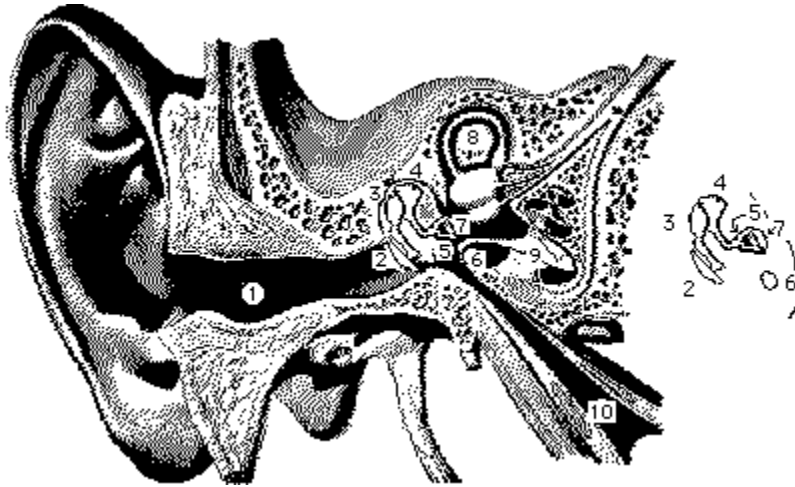


HEARING AND PERCEPTION

The operation of the ear has two facets: the behavior of the mechanical apparatus and the neurological processing of the information acquired. The mechanics of hearing are straightforward and well understood, but the action of the brain in interpreting sounds is still a matter of dispute among researchers.



[Click Here](#) to get a narrative description of the ear drawings.

Fig. 1 Parts of the ear

- | | |
|-------------------|------------------------|
| 1. Auditory canal | 6. Round window |
| 2. Ear drum | 7. Oval window |
| 3. Hammer | 8. Semicircular canals |
| 4. Anvil | 9. Cochlea |
| 5. Stirrup | 10. Eustachian tube |

THE EAR MECHANISM

The ear contains three sections, the outer, middle, and inner ears. The outer ear consists of the lobe and ear canal, structures which serve to protect the more delicate parts inside.

The outer boundary of the middle ear is the eardrum, a thin membrane which vibrates in sympathy with any entering sound. The motion of the eardrum is transferred across the middle ear via three small bones named the hammer, anvil, and stirrup. These bones are supported by muscles which normally allow free motion but can tighten up and inhibit the bones' action when the sound gets too loud. The leverages of these bones are such that rather small motions of the ear drum are very efficiently transmitted.

The boundary of the inner ear is the oval window, another thin membrane which is almost totally covered by the end of the stirrup. The inner ear is not a chamber like the middle ear, but consists of several tubes which wind in various ways within the skull. Most of these tubes, the ones called the semicircular canals, are part of our orientation apparatus. (They contain fine particles of dust-the location of the dust tells us which way is up.) The tube involved in the hearing process is wound tightly like a snail shell and is called the cochlea.

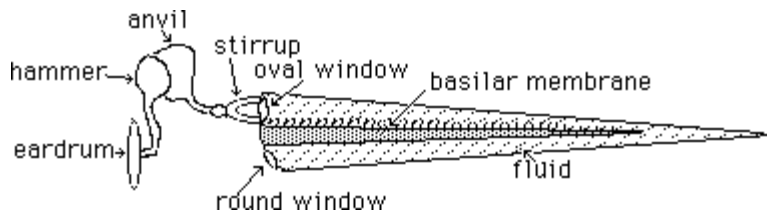


Fig 2. Schematic of the ear

This is a diagram of the ear with the cochlea unwound. The cochlea is filled with fluid and is divided in two the long way by the basilar membrane. The basilar membrane is supported by the sides of the cochlea but is not tightly stretched. Sound introduced into the cochlea via the oval window flexes the basilar membrane and sets up traveling waves along its length. The taper of the membrane is such that these traveling waves are not of even amplitude the entire distance, but grow in amplitude to a certain point and then quickly fade out. The point of maximum amplitude depends on the frequency of the sound wave.

The basilar membrane is covered with tiny hairs, and each hair follicle is connected to a bundle of nerves. Motion of the basilar membrane bends the hairs which in turn excite the associated nerve fibers. These fibers carry the sound information to the brain. This information has two components. First, even though a single nerve cell cannot react fast enough to follow audio frequencies, enough cells are involved that the aggregate of all the firing patterns is a fair replica of the waveform. Second, and probably most importantly, the location of the hair cells associated with the firing nerves is highly correlated with the frequency of the sound. A complex sound will produce a series of active loci along the basilar membrane that accurately matches the spectral plot of the sound.

The amplitude of a sound determines how many nerves associated with the appropriate location fire, and to a slight extent the rate of firing. The main effect is that a loud sound excites nerves along a fairly wide region of the basilar membrane, whereas a soft one excites only a few nerves at each locus.

PERCEPTION

The mechanical process described so far is only the beginning of our perception of sounds. The mechanisms of sound interpretation are poorly understood, in fact is not yet clear whether all people interpret sounds in the same way. Until recently, there has been no way to trace the wiring of the brain, no way to apply simple stimuli and see which parts of the nervous system respond, at least not in any detail. The only research method available was to have people listen to sounds and describe what they heard. The variability of listening skills and the imprecision of the language combined to make psycho-acoustics a rather frustrating field of study. Some of the newest research tools show promise of improving the situation, so research that is happening now will likely clear up several of the mysteries.

The current best guess as to the neural operation of hearing goes like this:

We have seen that sound of a particular waveform and frequency sets up a characteristic pattern of active locations on the basilar membranes. (We might assume that the brain deals with these patterns in the same way it deals with visual patterns on the retina.) If a pattern is repeated enough we learn to recognize that pattern as belonging to a certain sound, much as we learn a particular visual pattern belongs to a certain face. (This learning is accomplished most easily during the early years of life.) The absolute position of the pattern is not very important, it is the pattern itself that is learned. We do possess an ability to interpret the location of the pattern to some degree, but that ability is quite variable from one person to the next. (It is not clear whether that ability is innate or learned.) What use the brain makes of the fact that the aggregate firing of the nerves more or less approximates the waveform of the sound is not known. The processing of impulse sounds (which do not last long enough to set up basilar patterns)

is also not well explored.

INTERPRETATION OF SOUNDS

Most studies in psycho-acoustics deal with the sensitivity and accuracy of hearing. This data was intended for use in medicine and telecommunications, so it reflects the abilities of the average untrained listener. It seems to be traditional to weed out musicians from such studies, so the capabilities of trained ears are not documented. I suspect such capabilities are much better than that suggested by the classic studies.

LOUDNESS

The ear can respond to a remarkable range of sound amplitude. (Amplitude corresponds to the quality known as loudness.) The ratio between the threshold of pain and the threshold of sensation is on the order of 130 **dB**, or ten trillion to one. The judgment of relative sounds is more or less logarithmic, such that a tenfold increase in sound power is described as "twice as loud". The just noticeable difference in loudness varies from 3 dB at the threshold of hearing to an impressive 0.5 dB for loud sounds.

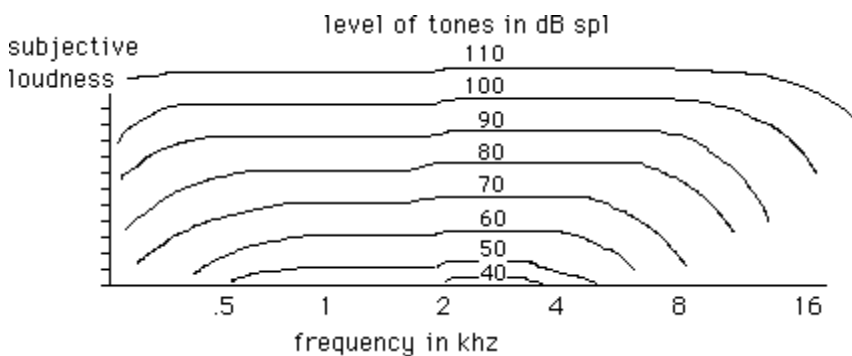


Fig. 3 Perceived loudness of sounds

The sensation of loudness is affected by the frequency of the sound. A series of tests using sine waves produces the curves shown. At the low end of the frequency range of hearing, the ear becomes less sensitive to soft sounds, although the pain threshold as well as judgments of relatively loud sounds are not affected much. Sounds of intermediate softness show some but not all of the sensitivity loss indicated for the threshold of hearing. At high frequencies the change in the sensitivity is more abrupt, with sensation ceasing entirely around 20 kHz. The threshold of pain increases in the top octave also.

The ability to make loudness judgments is compromised for sounds of less than 200ms duration. Below that limit, the loudness is affected by the length of the sound; shorter is softer. Durations longer than 200ms do not affect loudness judgment, beyond the fact that we tend to stop paying attention to long unchanging tones.

MASKING

The threshold of hearing for a particular tone can be raised by the presence of another noise or another tone. White noise reduces the loudness of all tones, regardless of absolute level. If the bandwidth of the masking noise is reduced, the effect of masking loud tones is reduced, but the threshold of hearing for those tones remains high. If the masking sound is narrow band noise or a tone, masking depends on the frequency relationship of the masked and masking tones. At low loudness levels, a band of noise will mask tones of higher frequency than the noise more than those of lower frequency. At high levels, a band of noise will also mask tones of lower frequency than itself.

PITCH

People's ability to judge pitch is quite variable. (Pitch is the quality of sound associated with frequency.) Most subjects studied could match pitches very well, usually getting the frequencies of two sine waves within 3%. (Musicians can match frequencies to 1%, or should be able to.) Better results are obtained if the stimuli are similar complex tones, which makes sense since there are more active points along the basilar membrane to give clues. Dissimilar complex tones are apparently fairly difficult to match for pitch (judging from experience with ear training students; I haven't seen any studies on the matter to compare them with sine tone results).

Judgment of relative pitch intervals is extremely variable. The notion of the two to one frequency ratio for the octave is probably learned, although it is easily learned given access to a musical instrument. An untrained subject, asked to set the frequency of a tone to twice that of a reference, is quite likely to set them a twelfth or two octaves apart or find some arbitrary and inconsistent ratio. The tendency to land on "proper" intervals increases if complex tones are used instead of sine tones. Trained musicians often produce octaves slightly wider than two to one, although the practical aspects of their instrument strongly influence their sense of interval. (As a bassoonist who has played the same instrument for twenty years, I have a very strong tendency to place G below middle C a bit high.)

Identification of intervals is even more variable, even among musicians. It does appear to be trainable, suggesting it is a learned ability. Identification of exact pitches is so rare that it has not been properly studied, but there is some anecdotal evidence (such as its relatively more common occurrence among people blind from birth) suggesting it is somehow learned also.

The amplitude of sound does not have a strong effect on the perception of pitch. Such effects seem to hold only for sine tones. At low loudness levels pitch recognition of pure tones becomes difficult, and at high levels increasing loudness seems to shift low and middle register pitches down and high register pitches up.

The assignment of the quality of possessing pitch in the first place depends on the duration and spectral content of the sound. If a sound is shorter than 200ms or so, pitch assignment becomes difficult with decreasing length until a sound of 50ms or less can only be described as a pop. Sounds with waveforms fitting the harmonic pattern are clearly heard as pitched, even if the frequencies are offset by some additive factor. As the spectral plot deviates from the harmonic model, the sense of pitch is reduced, although even noise retains some sense of being high or low.

TIMBRE

Recognition of sounds that are similar in aspects other than pitch and loudness is not well studied, but it is an ability that everyone seems to share. We do know that timbre identification depends strongly on two things, waveform of the steady part of the tone, and the way the spectrum changes with time, particularly at the onset or attack. This ability is probably built on pattern matching, a process that is well documented with vision. Once we have learned to identify a particular timbre, recognition is possible even if the pitch is changed or if parts of the spectrum are filtered out. (We are good enough at this that we can tell the pitch of low sounds when played through a sound system that does not reproduce the fundamentals.)

LOCALIZATION

We are also able to perceive the direction of a sound source with some accuracy. Left and right location is determined by perception of the difference of arrival time or difference in phase of sounds at each ear. If there are more than two arrivals, as in a reverberant environment, we choose the direction of the first sound to arrive, even if later ones are louder. Localization is most accurate with high frequency sounds

with sharp attacks.

Height information is provided by the shape of our ears. If a sound of fairly high frequency arrives from the front, a small amount of energy is reflected from the back edge of the ear lobe. This reflection is out of phase for one specific frequency, so a notch is produced in the spectrum. The elongated shape of the lobe causes the notch frequency to vary with the vertical angle of incidence, and we can interpret that effect as height. Height detection is not good for sounds originating to the side or back, or lacking high frequency content.

Peter Elsea 1996

