Hearing

The nervous system's cognitive response to sound stimuli is known as psychoacoustics: it is partly acoustics and partly psychology. Hearing is a feature resulting from our physiology that we tend to take for granted: something we just experience as part of life. It is only when we begin to work with sound recording that we become aware of some of the subtleties present in our auditory system. For example, phenomena like masking, in which only the louder of two sounds close together in frequency is perceived, are attributable to the behavior of our auditory physiology. A complete study of the function of the auditory system is exceedingly complex and beyond the scope of this discussion, however we should understand some of the features of the system that affect directly how we perceive sounds, especially when they are critical to the processes employed in the recording of sound.

Our auditory system is incredibly sensitive, allowing perception over many orders of magnitude in both amplitude and frequency. The quietest sound we can hear involves air motion on the order of the diameter of a hydrogen atom! We can discriminate tiny changes in sound timbres and accurately determine where in space a sound originates. We experience sound through our sensory organs and nervous system, so our perception cannot be divorced from these mechanisms and their characteristics influence what we hear. The pattern of air pressure vibrations striking our ears provides the input, but just as the structure of a room alters sound waves as they pass through it, so the apparatus of hearing changes the information we ultimately receive in our brains' auditory processing areas. We have the advantage of an adaptive brain, one that learns how to process the sensory inputs we receive from the cochlea and incorporate the inherent imperfections of our own auditory system. Nevertheless, we must process sound information through our hearing organs and our perception includes distortions: alterations in the way sounds are transmitted and converted to neuronal signals and in the way our brain interprets these inputs and renders for us what we experience as hearing.

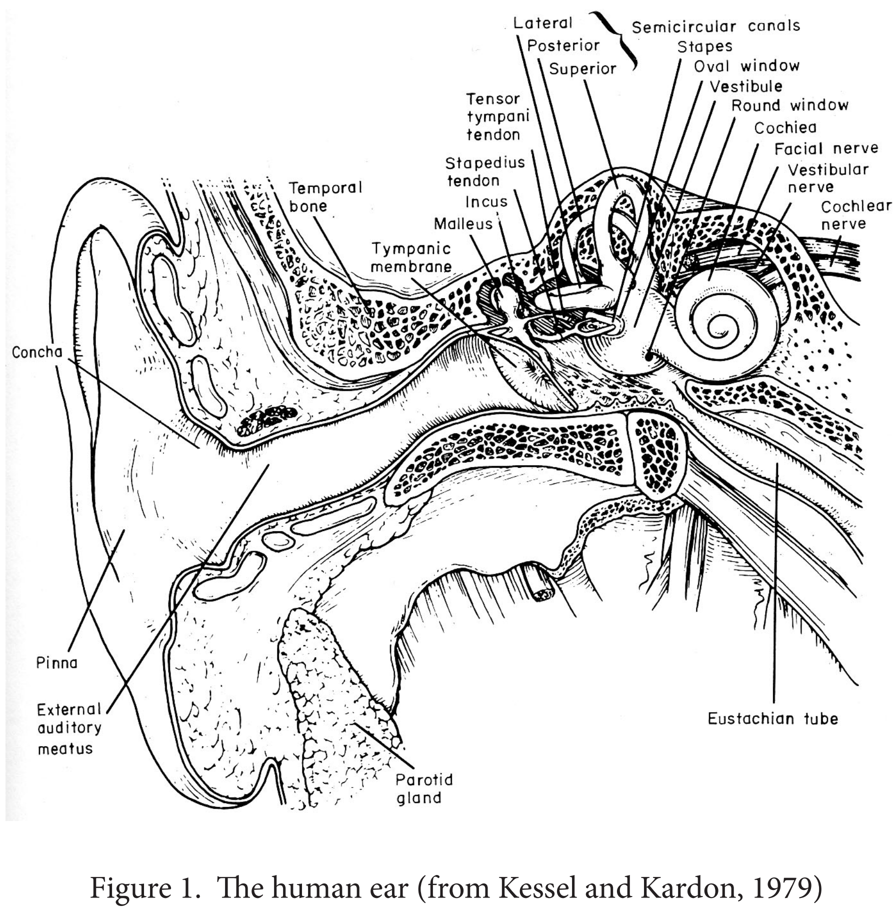

Figure 1. The human ear (from Kessel and Kardon, 1979)

We often assume that what we perceive as pitch is exactly equivalent to the actual vibratory frequency of the sound wave and that what we perceive as loudness is directly proportional to the amplitude of the sound wave pressure variations. In fact, the operation of our auditory system deviates somewhat from these ideals and we must factor these deviations into our understanding of the process of hearing. The first stage in the process of hearing, handled by the outer ear (pinna) and ear canal, distorts the incoming pressure wave. The ridges of the ear reflect particular frequencies from certain directions in order to create an interference pattern that can be used to extract information about the elevation from which a sound originates. Front-to-rear discrimination also depends in part on the shadowing of rear-originating sounds by the pinna. The external auditory meatus, or auditory canal, the guide conducting the sound wave to the eardrum (tympanic membrane), is a resonant tube that further alters the frequency balance of the sound wave. The resonant frequency falls in the same range as the peak in our sensitivity: around 4 kHz, and creates a maximum boost of about 15 dB (See Figure 2). There is significant variability in the exact resonant frequency of the outer ear due to the variety of ear shapes and sizes.

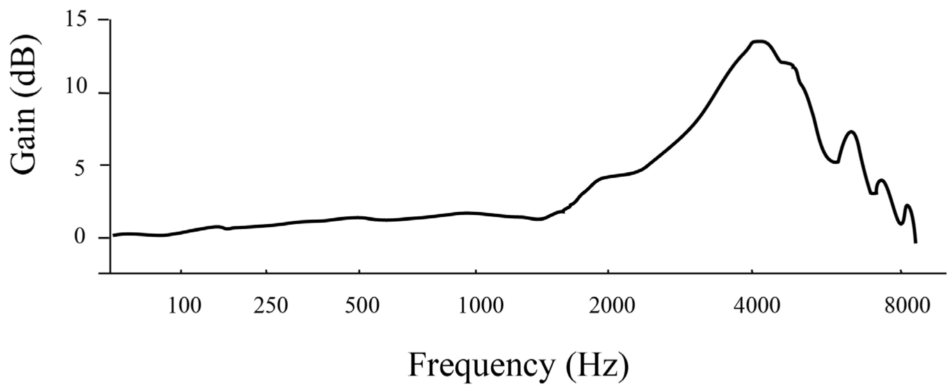

Figure 2: Resonance of ear canal.

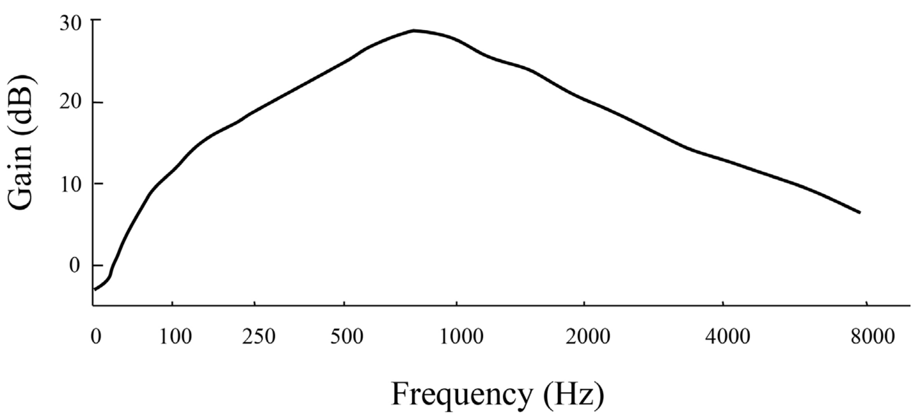

Figure 3: Frequency response of middle ear.

The tympanic membrane, or eardrum, is a flat conical membrane. It is open to the auditory canal on one side and in contact with a set of three tiny middle ear bones, the ossicles, on the other. The pressure on the outside is determined by the sound wave and the static atmospheric pressure. The static pressure in the middle ear is equilibrated through the Eustachian tube to the throat while the sound vibrations are conducted mechanically through the bones to the cochlea. The mechanical characteristics of the middle ear allow for active control of the transmission efficiency between the eardrum and the cochlea. The bones of the middle ear convert the motion of air striking the tympanic membrane into vibrations of the fluid in the cochlea, a process that involves mechanical advantage derived from lever action to provide forces capable of driving the liquid. Tiny muscles and ligaments connecting and suspending the bones can contract, pulling the bones away from their attachments to the eardrum and cochlea. This allows for adjusting the sensitivity of the hearing process, an action that can shift the amplitude range of hearing based on the loudness of incoming signals. This acoustic reflex acts as a compressor with a characteristic time course, producing attack times of about 40 milliseconds and release times of about 135 milliseconds. Loud sounds can also directly affect the cochlea's sensitivity to stimulation. This process can have a long recovery time, producing the phenomenon called threshold shift. After prolonged exposure to loud sounds, the sensitivity of the system can be reduced for hours afterwards, a fact often observed after hours of playing or mixing loud music.

Once the sound vibrations reach the cochlea, they are converted from solid to liquid medium vibrations. An important function of the bones of the middle ear is to amplify mechanically the vibrations in preparation for transfer to a liquid medium. Since liquids are denser than gases and less dense than solids like bone, we encounter a potential problem when converting the energy in one medium to energy in another: the systems require different amounts of force to drive them. The bones act to focus the vibrations of the eardrum and deliver them efficiently to the cochlea as well as to protect the cochlea from too much input. They act as an impedance converter, efficiently coupling the low-impedance air pressures with higher-impedance liquid pressures inside the cochlea. The middle ear produces a further gain of more than 20 dB at around 1000 Hz (See Figure 3).

The cochlea is a dual-purpose structure: it converts mechanical vibrations into neuronal electrical signals and it separates the frequency content of the incoming sound into discrete frequency bands. It functions like a spectrum analyzer, a device that breaks sounds or other signals into their component sine wave elements. It is, however, an imperfect analyzer and there is some overlap between the bands, so that strong signals in one band slightly stimulate adjacent ones, creating harmonic distortion. It is up to the brain to sort out the raw data from the cochlea and produce the sensation we call hearing.

The cochlea performs the separation of different frequencies by providing an array of sensing cells, the inner hair cells, which are mechanically stimulated by the movement of a membrane to which they are connected. The membrane is caused to vibrate by the fluid filling the chamber, which is in turn caused to vibrate by the bones of the middle ear. The cochlear cells are distributed along the length of the coiled structure and each area resonates at a particular frequency moving from high to low frequency as one proceeds along the length of the membrane holding the cells. Due to its placement, each cell responds to vibrations of a specific range of frequencies and each is connected to a nerve that sends signals to the brain. The delicate hair cells are susceptible to damage from excessive excitation. Overexposure to loud sound can destroy these cells, leading to diminished sensitivity as the individual frequency detectors are lost. Although early research indicates some regeneration might be possible, the condition should be considered irreversible and avoided at all cost. Reducing exposure to loud sound should be a primary consideration of everyone working in audio.

Since the action of the cochlea depends on the location along its length that vibrates with a particular frequency of sound input, multiple sounds close together in frequency interact. Louder sounds tend to "drown out" lower amplitude sounds at the same frequency, a phenomenon known as masking. This masking behavior has profound inplications for mixing as we will see. This principle allows some types of noise reduction processing and methods of audio data compression like mp3s to work. It also complicates mixing two sounds with similar frequency content. Dynamic range processors like compressors depend on altering masking for their effectiveness.

Exactly how our brains process auditory inputs and create for us the conscious awareness of hearing sound is still within the realm of mystery. Although research has elucidated much of the structure of the neuronal pathways and processing centers in the brain, the explanation of exactly how we perceive sound is still unsatisfying. Fortunately, for the understanding required to become adept at sound recording we need only appreciate the operational characteristics of the system as it applies to how we perceive sound.

There are critical features of our auditory system that we must consider to appreciate which characteristics of sounds are necessary to preserve in our recordings. In order to preserve the cues we use in localizing the positions of sound sources in space, we need to understand how we tell where in space sounds originate. In order to make accurate sound recordings we need to be aware of how we perceive the amplitude and frequency information in sound waves. To convincingly manipulate sounds in the studio, we need to know how sounds behave in our environment.

Any sound originating in space reaches us through our two ears. Since they are separated by several inches, sounds not originating directly ahead or behind reach them at slightly different times. Further, they strike the closer ear with slightly more energy, making that side sound louder. By using the relative time-of-arrival and loudness cues, we determine where in space a sound originated. We can make use of these cues to fool the ear, as we might when mixing sounds in the studio, placing sounds in different apparent positions in a mix. Preserving these cues is critical to making stereo recordings that capture the realism we desire. Sound striking one ear first is a very strong cue to the position of a sound source. By delaying an element panned to one side in mix by a few milliseconds, it can take many decibels of gain to make that sound seem as loud as the same element panned un-delayed to the opposite side. The pan controls built into most mixers use only the apparent loudness cue to pan signals from left to right, ignoring the time-of-arrival difference.

Our perception of timbre depends on the frequency response behavior of our auditory system. Because we do not perceive all frequencies to be equally loud at the same measured sound pressure level, discussing loudness demands a measure of loudness that takes the sensitivity to frequency into account. A unit known as the phon is used in loudness comparisons; it is the amplitude (sound pressure level in decibels) of a 1kHz sine wave tone judged to have the same loudness as the signal in question. A similar unit, the sone, is used in comparisons of frequency.

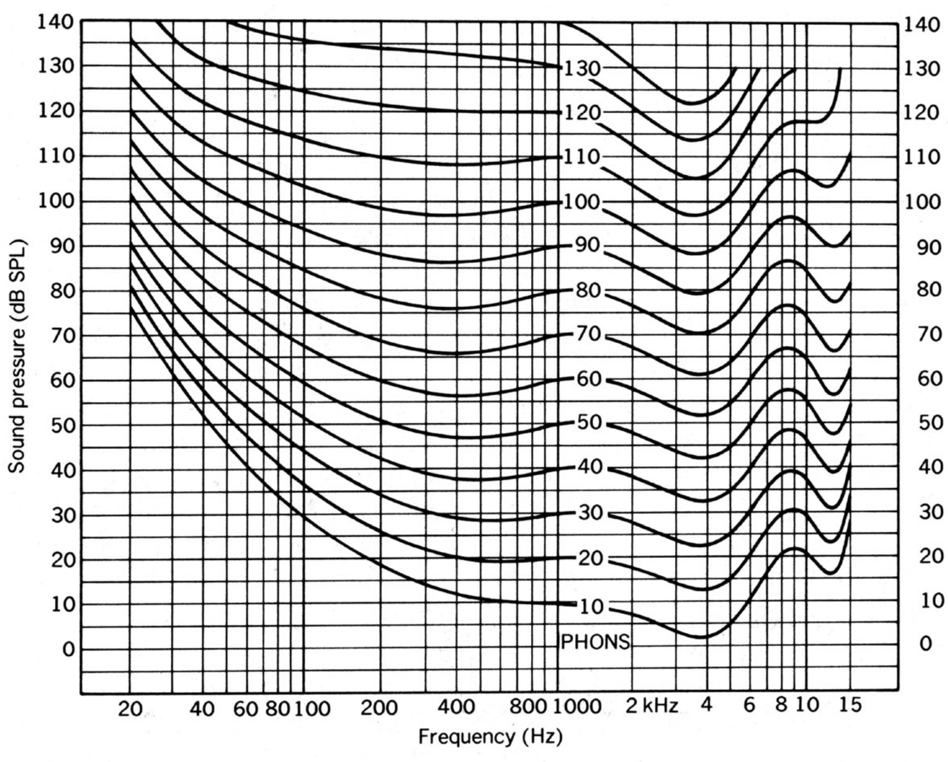

Fig. 4 Equal loudness curves (From Robinson and Dadson, 1956)

Due to the physical characteristics of the auditory system, we do not perceive all frequencies as equally loud for the same sound pressure level. Further, the variation in perceived loudness as a function of frequency changes with the loudness level. The curves of equal loudness, often called Fletcher-Munson curves after the original researchers findings (Fletcher and Munson, 1933), show the effect of sound level on the spectral sensitivity of our auditory system. These measurements vary significantly among individuals and are average values. While Fletcher and Munson used headphones and pure tones in their work, the graph above is the work of Robinson and Dadson, who used loudspeakers to reproduce the pure tones in an anechoic room. The two sets of curves differ, but both show a peak of sensitivity at around 4 kHz, near the resonant frequency of the auditory canal, and a significant decrease in sensitivity at extreme frequencies, especially low frequencies. This phenomenon has important implications for sound mixing: the level at which we listen affects our perception of the balance of frequencies present in the signal.

References:

Tissues and Organs: A Test-Atlas of Scanning Electron Microscopy, Kessel, R. G. and R. H. Kardon, W. H. Freeman and Company, 1979

Robinson, D. W. and R. S. Dadson, "A re-determination of the equal loudness relations for pure tones", Brit. J. App. Phys.,7, 166-181, 1956

Fletcher, H. and W. A. and Munson, "Loudness, its definition, measurement and calculation", J. Acoust. Soc. Amer., 5, 82-108, 1933

Suggested reading:

An Introduction to the Physiology of Hearing, James O. Pickles, Academic Press Limited, 1988. ISBN 0-12-554754-4

Hearing: Physiological Acoustics, Neural Coding, and Psychoacoustics, Gulick, W. L., G. A. Gescheider, and R. D. Frisina, Oxford University Press, 1989. ISBN 0-19-504307-3