4.1 Anatomy and Physiology of the Ear

The ear can be divided into three gross sections shown in Figure 4.1.

- Outer Ear
- Middle Ear
- Inner Ear

Each section plays an important and unique part in decomposing and translating acoustical waves into electrochemical impulse signals that are used by the brain.

![Cross-Section of the Human Ear](image)

Figure 4.1: Cross-Section of the Human Ear [W04]

4.2 The Outer Ear

The main function of the outer ear is to direct and focus acoustic vibrations to the eardrum. The parts of the outer ears are:
• Pinna
  This is the external flap of cartilage surrounding the entrance to the ear. The shape of the pinna causes a
  resonance effect that will alter the amplitude of the pressure wave at different frequencies. Since this spectrum
  shaping changes based on the sounds origin, the pinna also helps with sound localization.

• Auditory Canal
  The auditory canal is a complex cavity roughly 3cm long. It acts as a resonator that further shapes the spectrum.
  Specifically, this resonator amplifies the spectrum between 2kHz and 5kHz. This is an important range for
  speech recognition. Figure 4.2 shows the overall transfer function for both the pinna and auditory canal. The
  auditory canal combined with the pinna is known as the meatus.

• Tympanic Membrane (eardrum)
  The function of the tympanic membrane is to collect air vibrations at the end of the auditory canal and convert
  them to mechanical movement in the middle ear. It is an incredibly sensitive instrument with an operating range
  of more than 100dB. In other words the maximum sound pressure level (spl) the eardrum can detect is more
  than 10000000000 times the minimum!

![Figure 4.2: Transfer Function of the External Ear [W04]](image)

### 4.3 The Middle Ear

The middle ear is a 2 cm² space between the tympanic membrane and the cochlea. It consists of three bones (the
malleus, incus, and stapes) surrounded by air. This is shown in Figure 4.3. Together these bones are known as the
ossicles. The function of the middle ear is to correct for an impedance mismatch between the air in the outer ear and
the fluid in the inner ear. It do this by collecting energy in the relatively large tympanic membrane and focusing it on
the relatively small oval window that sits at the base of the stapes. Like any medium, the middle ear will also slightly
alter the spectrum. Its transfer function is shown in Figure 4.4.

One thing that can affect the performance of the ossicles is the air pressure in the middle ear. To perform optimally the
tympanic membrane needs the air pressure to be equal on both the outside and inside. In order to equalize this pressure
we have an eustachian tube that connects the inner ear to the back of the mouth, allowing the pressure in the middle
ear to equalize with the ambient pressure. This is what happens when your ears “pop” when changing elevation.

Another thing that can decrease the transfer of energy in the middle ear are the muscles surrounding the ossicles.
When these muscles tighten they dampen the amount of sound that is passed through the middle ear. This can happen
both consciously and unconsciously to prevent intense sounds from damaging the inner ear. This often happens after a loud sound, since loud sounds often follow loud sounds. The military often takes advantage of this when firing large cannons. They will set off an initial pre-charge to ‘prep’ peoples middle ears for the following larger sound.

4.4 The Inner Ear

- Cochlea
  
The most important part of the inner ear is the cochlea, shown in Figure 4.5. It is a coiled tube about 35mm long. The tube is filled with a lymphatic fluid and is divided lengthwise by the basilar membrane and organ of corti. The stapes connects to the cochlea at its base through the oval window. Vibrations from the stapes travel through the lymphatic fluid to the apex of the cochlea and then back down the other side of the partition to the round window. This induces movement of the partition. A lengthwise cross-section of the cochlea is shown in

![Figure 4.3: The Middle Ear [W04]](image1)

![Figure 4.4: Transfer Function of the Middle Ear [P88]](image2)
Figure 4.6. The manner in which the partition moves plays an important part in how the sound is encoded into neural impulses.

Figure 4.5: The Inner Ear [W04]

Figure 4.6: The Cochlea Unrolled [P88]

- **Organ of Corti**

  Figure 4.7 shows a cross-section of the cochlea. Along the left side there are a series of nerves that are fed into the middle of the cochlea. This area is known as the organ of corti and is where transduction (conversion of physical movement to neural impulses) occurs. Figure 4.8 show a closer view of the organ of corti. It sits atop the basilar membrane and below the tectorial membrane. As the basilar membrane moves up and down the tectorial membrane sheers across the organ of corti. This causes the cilia that sit above the hair cells to bend. This results in the firing of the nerves attached to the hair cells. It is important to note that at different points along the cochlea the organ of corti may be registering different levels of vibration. All told, there are about 30,000 sensory hair cells measuring the exact movement of the cochlea.
4.5 Encoding of Sound

It is very difficult to obtain a precise measurement of how the cochlea moves. It is surrounded by bone; as a result any instrument use for measurement is bound to affect the normal operation of the ear. Therefore, there is no consensus as to exactly how the inner ear decomposes acoustical waves. Two common theories are the place theory and the timing theory. Both believe the ear contains a series of band pass filters that perform something similar to a short-time fourier transform. These filters are thought to have a constant Q, where Q is defined as the center frequency divided by its bandwidth.

- Place Theory
  As sound travels down the cochlea, a traveling wave is induced along the basilar membrane. Measurements have shown that the peak of the traveling wave occurs at different positions along the basilar membrane depending
upon the frequency of the sound. For high frequencies the peak is near the base of the cochlea, and for low frequencies it is near the apex. This effect is shown in Figures 4.9 and 4.10. The characteristic frequency (CF) for a position is the frequency that causes the most movement at that position. The place theory states that each position acts like a bandpass filter, only recording sounds around its CF. This is how we distinguish between frequencies. The problem with this theory is that it is weak at encoding information at less than 1kHz, when in fact the ear is very good at encoding this information. This discrepancy is due to the fact that the traveling wave envelope is very wide at low frequencies.

Figure 4.9: Traveling wave in the cochlea [Y00]

Figure 4.10: The upper plot show the traveling wave envelope for different frequencies. The lower plot show the filter responses for different positions [W04]
• Time Theory

To compensate for problems in the place theory the time theory was introduced. The time theory states that the traveling wave in the cochlea moves at the frequency of the sound being played. This motion causes the neurons in the cochlea to fire at that frequency. The result being that the time waveform is encoded as neural firing. The issue with this theory is the relatively slow neural recovery time. A neuron needs about 1ms of recovery time after it has fired before it can fire again. This means that any single neuron cannot encode frequencies above 1kHz. In order to increase this number the brain is thought to sum the responses of a number of neurons. Now when some neurons are in recovery others will be firing and higher frequencies can now be encoded. This is shown in Figure 4.11.

![Pure Tone Signal and Neural Firing Patterns](image)

Figure 4.11: Summed response of neural firings [W04]

The current consensus is that the brain uses both the time and place encoding in sound perception. Under 1kHz timing is probably used because it is most efficient in that range, between 1 and 5kHz the brain may use a combination and above 5kHz place is used because it is most efficient in that range. The loss of timing around 5kHz could explain why we can’t hear musical tones above that frequency.

The Mel-scale frequency warping has been developed to simulate place and time encoding. This warping simulates a constant Q filter bank and has helped improved speech recognition.

References


