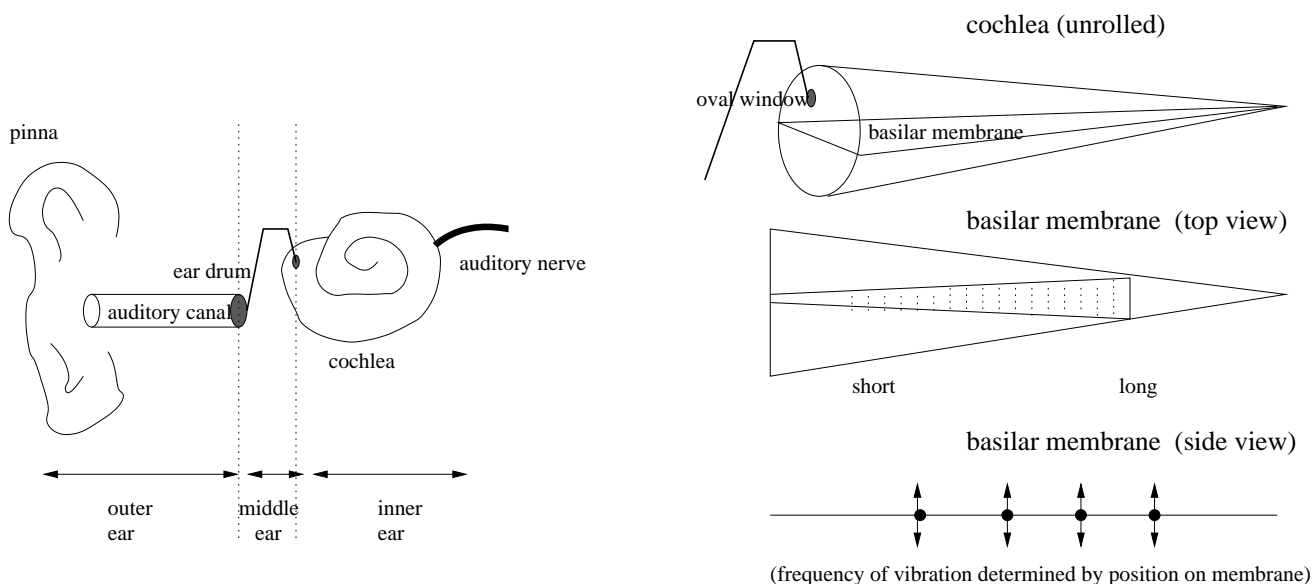


I will present a simplified model of MP3 audio compression in the next few lectures. This compression scheme is based fundamentally on properties of human hearing. In today's lecture, I will discuss the basics of the anatomy. Next lecture, I will cover some of the psychology (perception).

I will not examine you on the fine details of the material that I am presenting today, but I may ask about the general ideas. The general ideas are necessary for you to appreciate what I present next class. That in turn will help you understand MP3.

Human ear

Let's review some of the basics of human hearing. We think of our ears as the two fleshy appendages on the side of the head. These appendages are called the *pinnae* (one pinna, two pinnae). Each pinna contains a hole which leads to a tube-like cavity called the *auditory canal*. At the end of this canal is the *ear drum* (*tympanic membrane*) which vibrates in response to air pressure variations. The ear drum marks the boundary between the *outer ear* and the *middle ear*.



The middle ear is an air-filled cavity behind the ear drum. This cavity is connected to various other cavities in the head, i.e. mouth and nasal cavities. These connections make the middle ear susceptible to infections, especially in children.) The middle ear contains a chain of three small bones (*ossicles*). One end of the chain is attached to the ear drum and the other end is attached to the *cochlea* which is part of the *inner ear*. The chain of ossicles is near-rigid and acts as a lever. It transfers large amplitude oscillations of the eardrum to small amplitude (but large pressure) oscillations to the base of the cochlea. The large pressure is applied to the *oval window* at the base of the cochlea. See figure above.

The cochlea is a fluid-filled snail-shaped organ which contains the nerve cells that encode the pressure changes. If we would unwind the cochlea¹, it would have a cone-like shape: the thick end is the *base* and the thin end is the *apex*. The interior of the cochlea is partitioned into two parts by

¹We cannot unwind it because the shell is hard.

a large long triangular membrane. This *basilar membrane* contains the nerve cells. The nerve cells lie on an inverted triangle of elastic fibres that reach across the membrane. Counterintuitively, the fibres are shorter at the base of the cochlea and longer at the apex of the cochlea (see figure above).

The mechanical properties of the basilar membrane are not straightforward. But the basic story is that different positions along the length of the basilar membrane resonate to different temporal frequency components of the sound wave – that is, the different frequencies k that we have been discussing in the context of the discrete cosine transform. The basilar membrane resonates to low temporal frequencies (long wavelengths) at the far end (apex of cochlea) where the fibres are longest, and it resonates to high temporal frequencies (short wavelengths) at the near end (base of cochlea) where the fibres are shortest.

For this mechanical vibration to be sensed by the brain, it must be encoded somehow by nerve cells and the code must be sent to the brain. There are a large number of nerve cells (tens of thousands) distributed along the basilar membrane, from the base to the apex. Each of these nerve cells has a long cable-like fibre that connects it to the brain. These fibres are all bundled together, forming the auditory nerve.² In particular, each nerve cell encodes the amplitude of the vibrations of the part of the basilar membrane where it is located. The code is a sequence of electronic pulses³. Since each position on the basilar membrane resonates to a particular range of sound frequencies, each neuron encodes the amplitude of vibrations at those frequencies. Over a small interval of time, these amplitudes correspond roughly to the values in the spectrogram discussed last class.

To summarize, we should think of the spectrograms discussed last class as an encoding of the various frequency components of the sound over various intervals of time. This analysis of the sound into different frequency components corresponds roughly to the analysis of the sound that is performed in the ear. Different positions on the basilar membrane of the inner ear resonate to particular frequency components of the sound. Finally, these mechanical vibrations are converted by neurons on the basilar membrane into pulses that are sent to the brain.

²Similarly, the nerve cells in the retina of the eye have fibres that are bundled together to form the *optic nerve* and that carry the coded images from the eye to the brain.

³called "action potentials"