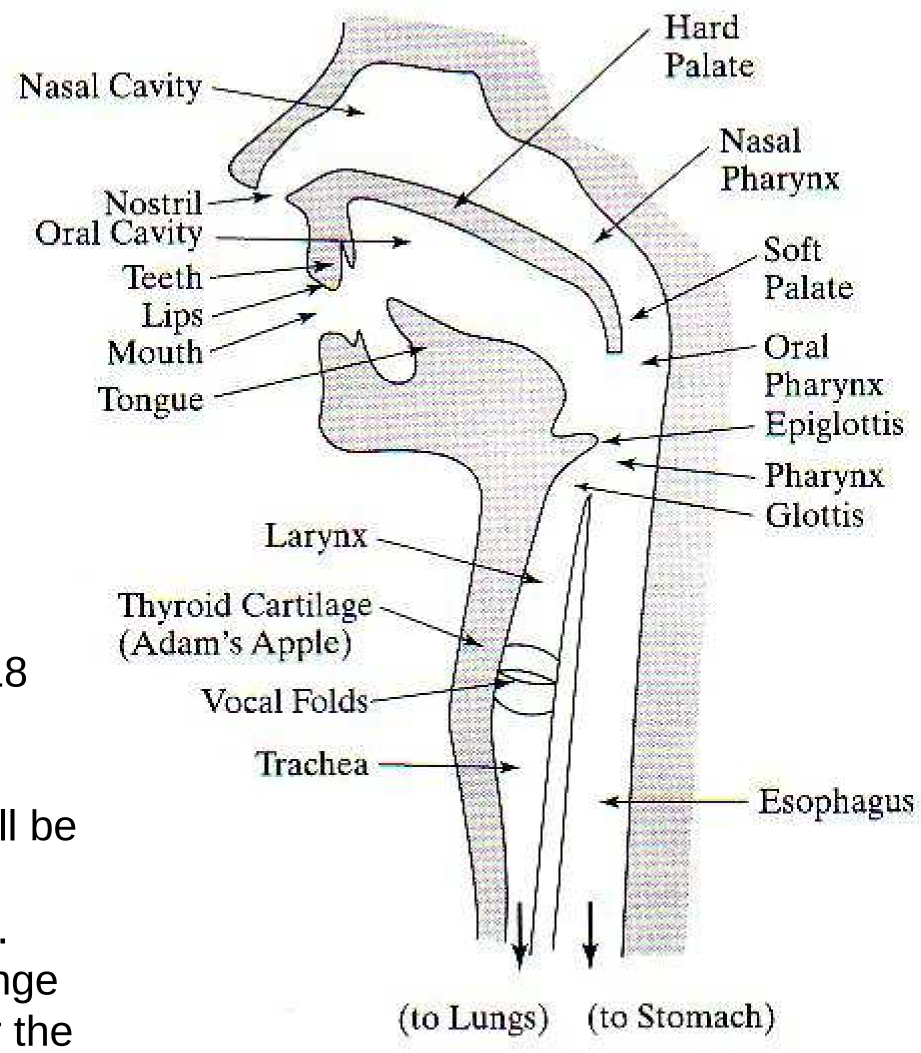
6 – Speech and singing

Speech and singing are examples of how the resonator shapes the sound out of the input broadband signal. In this case the resonator is a complicated system composed of the throat, the mouth, and the nasal cavities that is on the way of the initial signal produced by the vocal folds to the outside. Some professional singers say that the resonances inside the torso are important as well, and it is very important to properly control the diaphragm etc.



The length of the way from the vocal folds to the opening of the mouth of adults is between 17 and 18 cm. If we approximate this cavity (of a complicated form!) by a tube with one closed and one open end (node-antinode boundary conditions), then there will be the fundamental and odd overtones (that is, resonances) at approximately 500, 1500, 2500, etc. Hz. Changing the shape of the mouth one can change the resonance frequencies – almost by 1000 Hz for the fundamental that indicates that the approximation by a simple tube is oversimplified. These resonances are formants and they should not be confused with the fundamental and harmonics produced by vocal folds.

Singing and speech

The difference between speech and singing is the following. In the case of singing, the vocal folds vibrate producing the fundamental frequency and many harmonics. The tension of the vocal folds controlled by the muscles defines the frequency of the oscillations (similarly to the Mersenne‘s law), whereas the air throughput defines the amplitude of the oscillations. The oscillations of the vocal folds and thus the air blowing through them are <u>pulse trains</u>. The sound produced by the vocal folds is called <u>glottal wave</u>. Then the resonator shapes this sound.

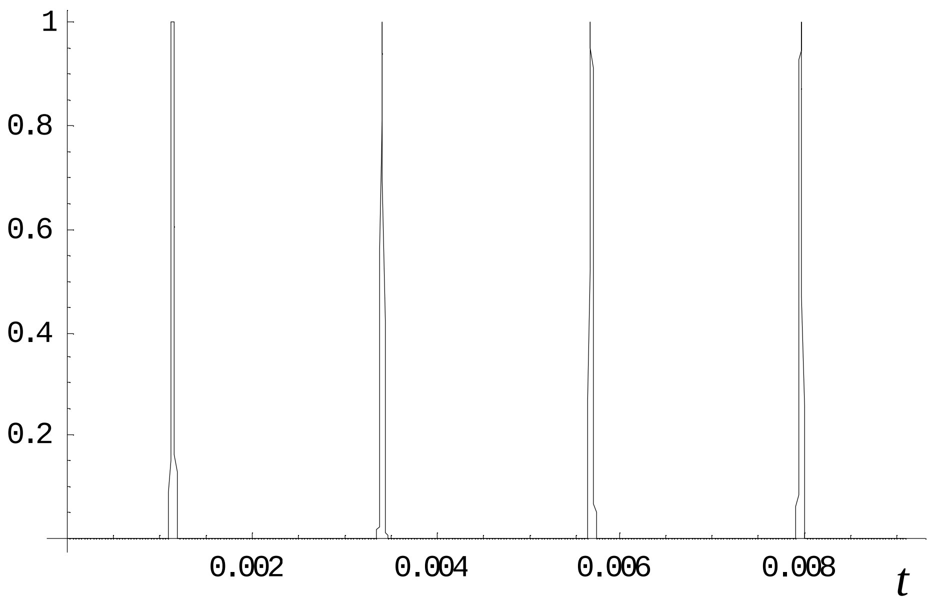
In the speech, where elementary sounds, wovels and consonants, are rapidly following one after the other, both vocal folds and the resonator do not have time to readjust from one frequency to the other, and also the speaker does not attempt to precisely control the frequency of the speech. As a result speech is a <u>transient</u> process with no well-defined frequency like in singing. Instead of the spikes in the Fourier spectrum at the fundamental frequency and overtones in singing, the Fourier spectrum of the speech has only broadened maxima responsible for the general height of the voice and for the <u>intonation</u>.

If the person is hoarse, vocal folds do not vibrate at all but the air blowing through them produces a kind of noisy, hissing sound. Still the resonator can shape a recognizable (however low-quality) speech out of this noise.

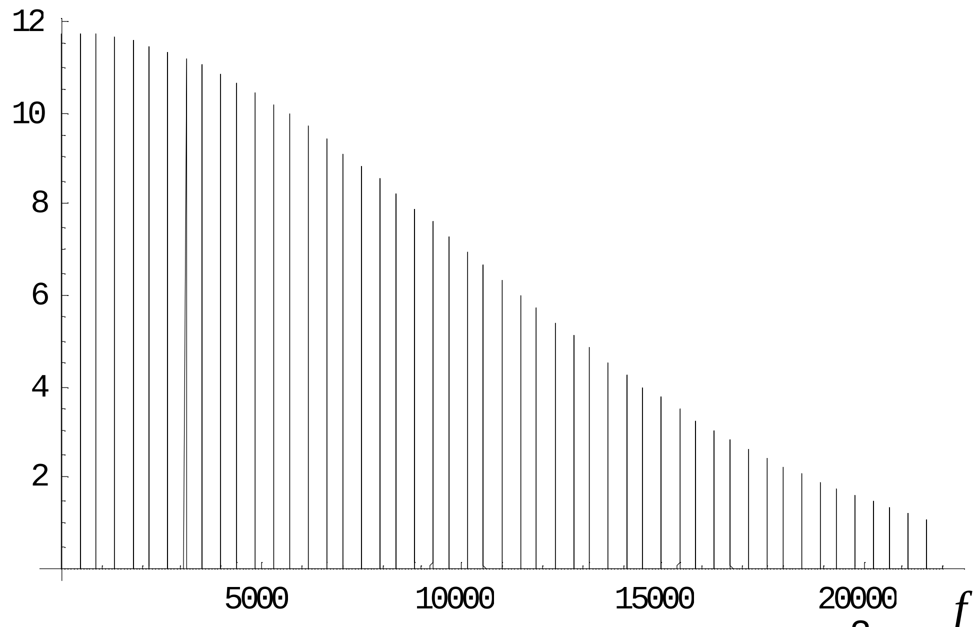
Glottal wave

Formation of the glottal wave (in fact, oscillation, not a wave!) is based on the Bernoulli's law. When the air (or any fluid) flows through a constriction, its speed increases and, as a consequence, the pressure in the constriction decreases. (This is exactly the physical law that makes airplanes fly!) The vocal folds create a constriction for the air, and the decrease of the pressure can become so large that the folds close for a very short moment, stopping the air flow. After that the pressure very quickly recoveres, the folds open, and everything repeats periodically. The resulting glottal wave is a pulse train that sounds like a buzz. It is a broadband signal with amplitudes of the overtones decreasing very slowly.

Glottal wave with the frequency 440 Hz

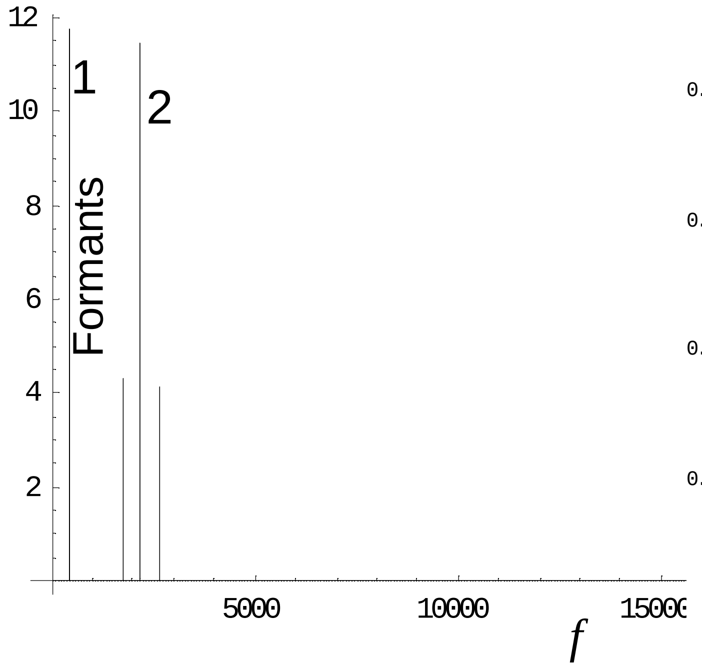


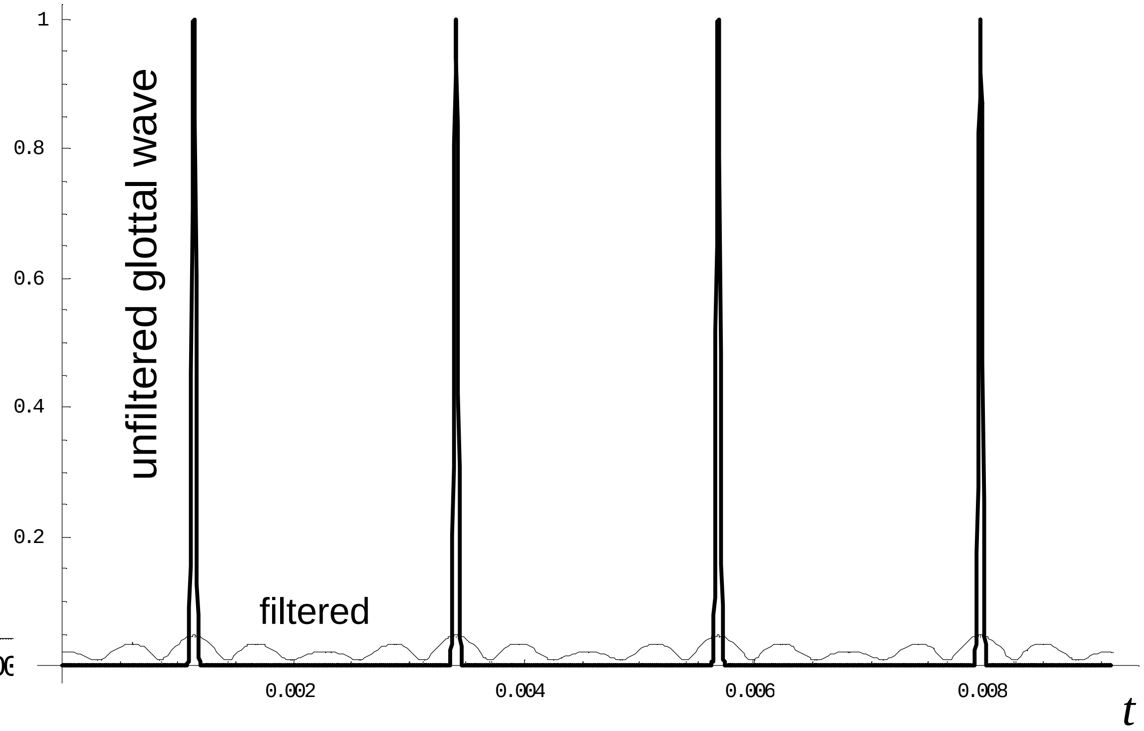
Its Fourier spectrum



Filtering of the glottal wave by the resonator

Although real vowels are shaped by 3 or 4 formants, one can sinthesize simplified vowels *u, o, a, e, i* (Latin phonetic notations) using only two formants. For *u*, the two formants are mostly close to each other while for *i* they are mostly separated. Below is the Fourier spectrum of the glottal wave above filtered with two formants corresponding to *a* and the resulting wave form (filtered) obtained by the inverse Fourier transformation.





m

Spectrograms of different vowels