CSC3160 - Fundamentals of Speech and Language Processing

Lecture 4: Introduction of speech production

Zhizheng Wu

Outline

- Information in human speech
- Speech production
- Source filter model
- Timbre
- Prosody

Text version of a speech

Trask: Sir, you are out of order!

Slade: Outta order? I'll show you outta order! You don't know what outta order is, Mr. Trask! I'd show you but I'm too old; I'm too tired; I'm too fuckin' blind. If I were the man I was five years ago I'd take a FLAME-THROWER to this place! Outta order. Who the hell you think you're talkin' to? I've been around, you know? There was a time I could see. And I have seen boys like these, younger than these, their arms torn out, their legs ripped off. But there isn't nothin' like the sight of an amputated spirit; there is no prosthetic for that. You think you're merely sendin' this splendid foot-soldier back home to Oregon with his tail between his legs, but I say you are executin' his SOUL!! And why?! Because he's not a Baird man! Baird men, ya hurt this boy, you're going to be Baird Bums, the lot of ya. And Harry, Jimmy, Trent, wherever you are out there, FUCK YOU, too!


Spoken version

Ways to say mom

Text version

Mom
妈妈

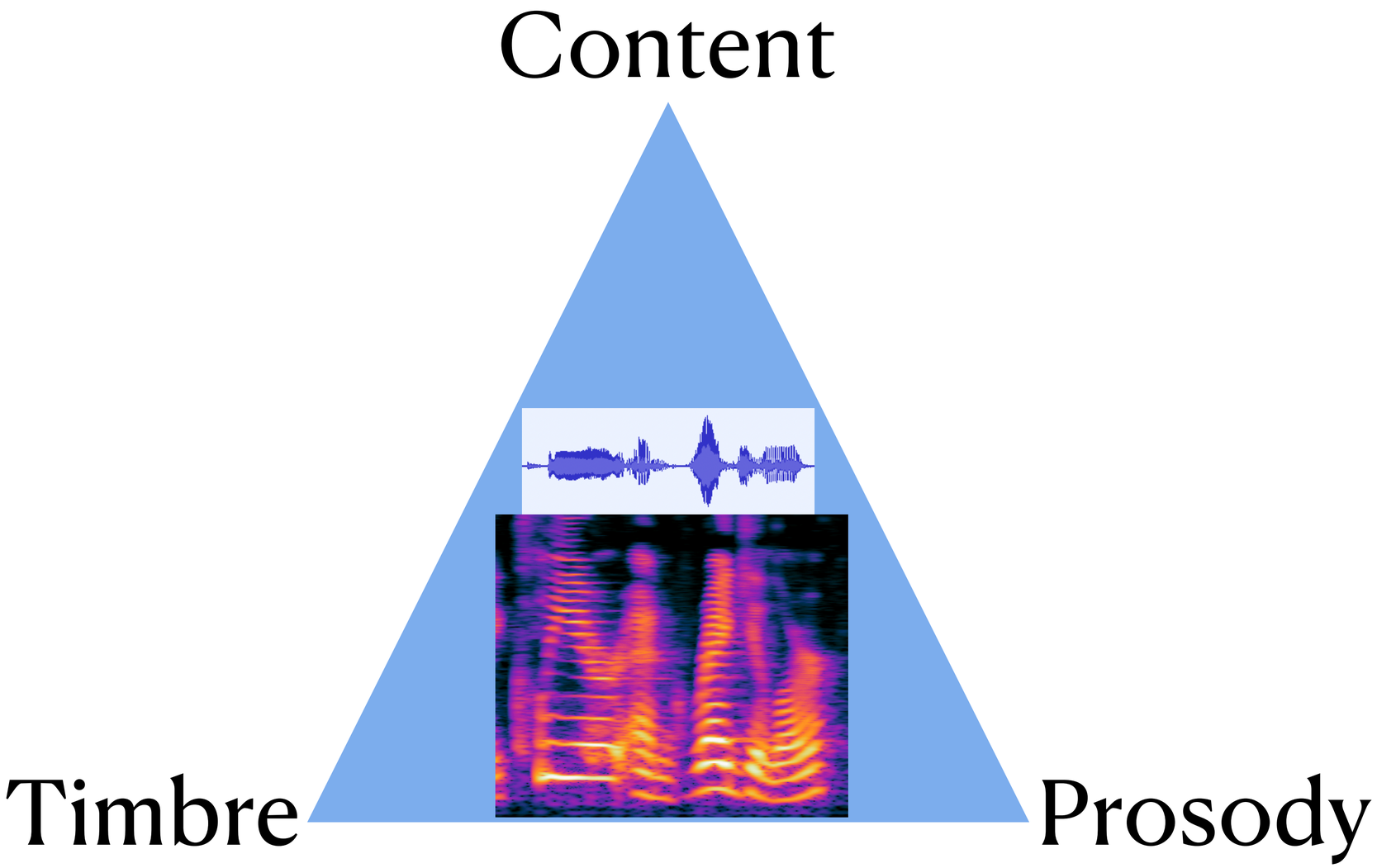
Spoken version

Content

Identity

Emotion

Age, etc


Content
Timbre
Prosody

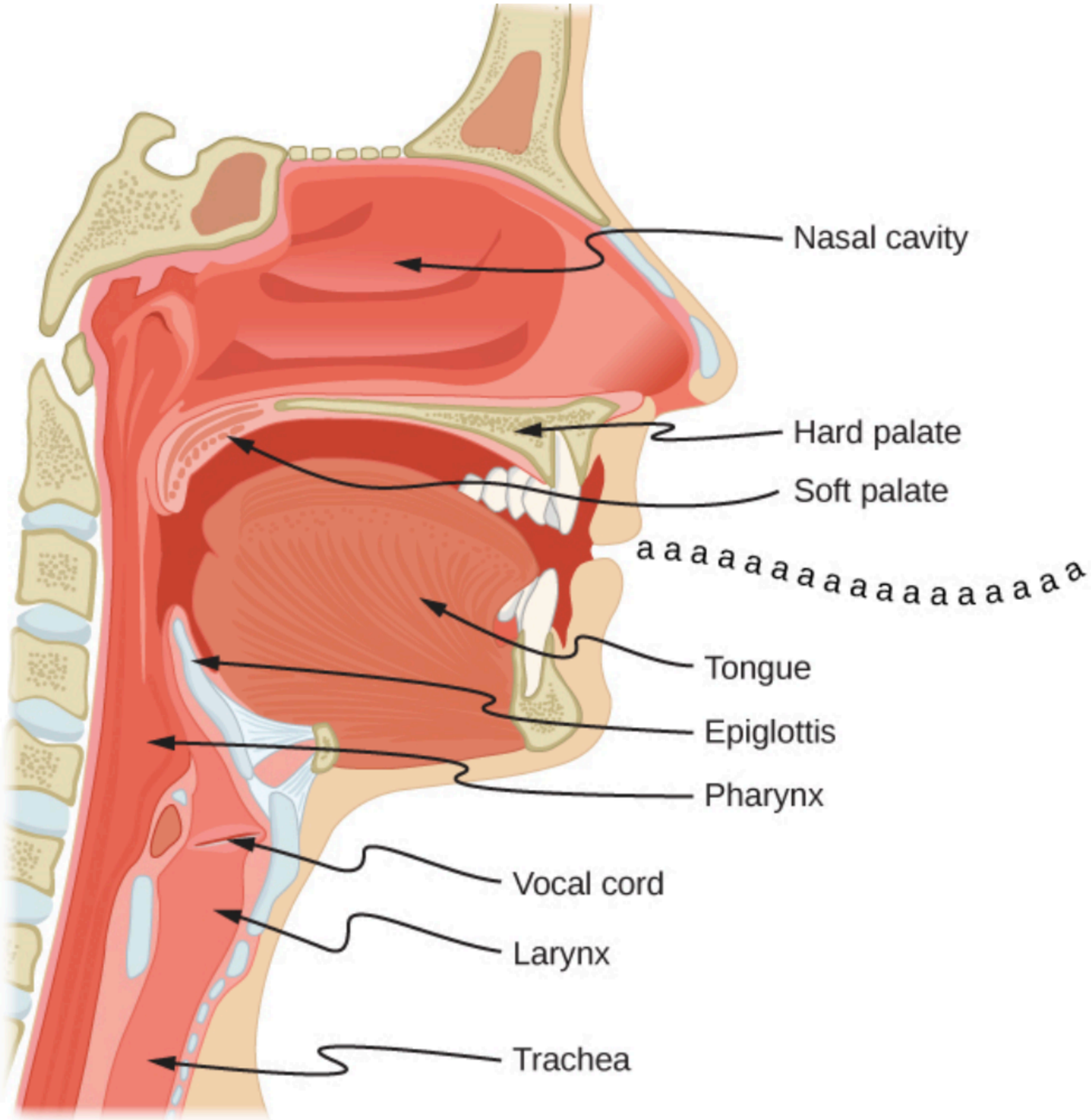
Sound: Speech

- Speech requires
  - a source of sound waves (**vibrations**)
  - a means of shaping those vibrations into **words**

Speech production

- When a person has the urge or intention to speak, their brain forms a sentence with the intended meaning

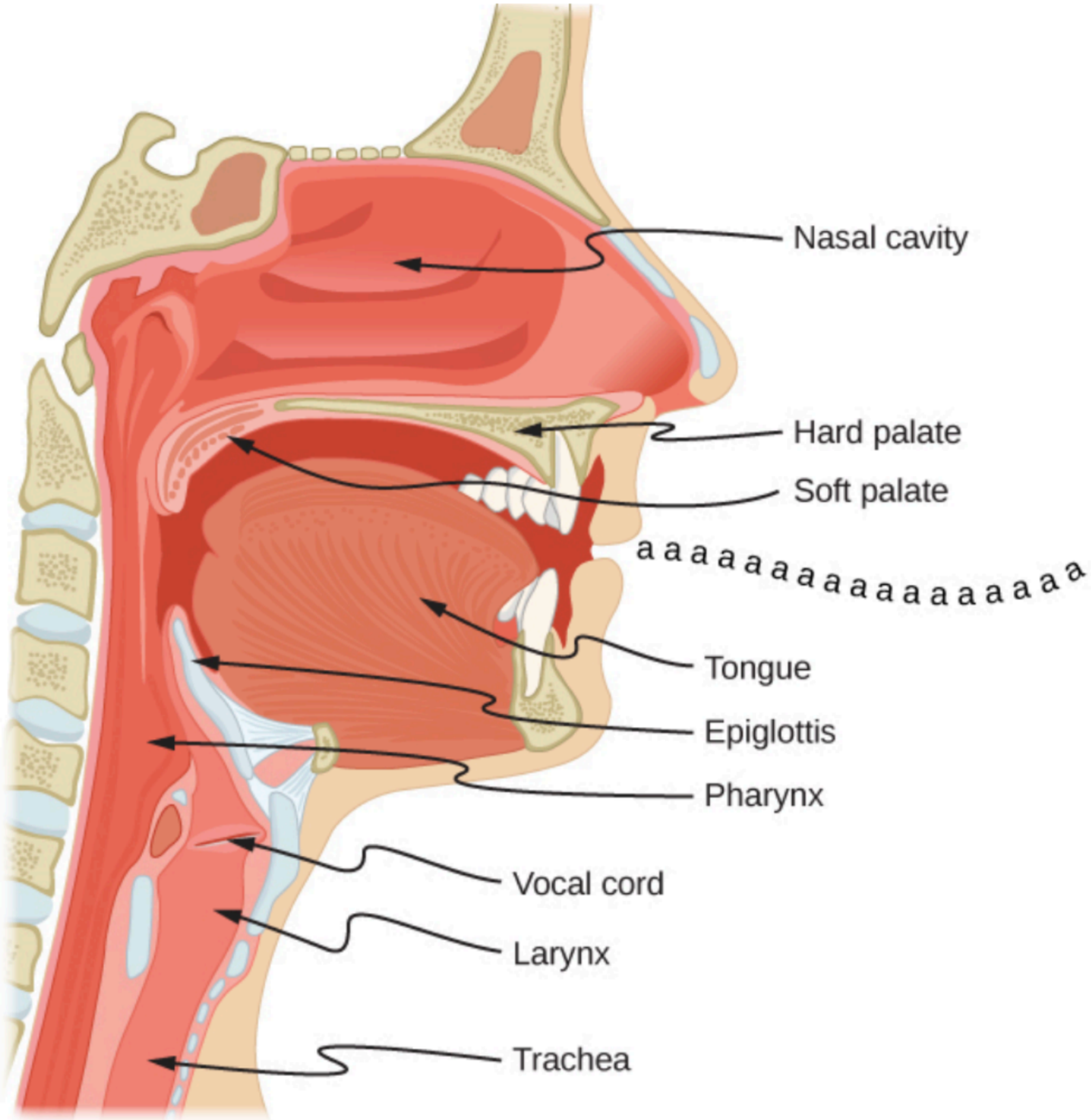
- Maps the sequence of words into physiological movements required to produce the corresponding sequence of speech sounds



Speech production

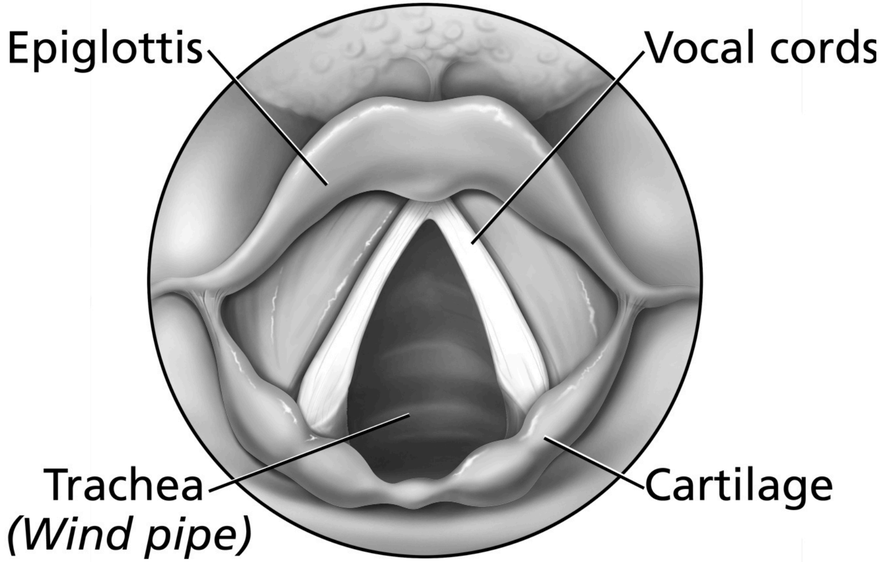
- Contracting the lungs, pushing out air from the lungs
  - Airflow in itself is not audible as a sound

- Obstruct airflow to obtain an oscillation or turbulence
  - Oscillations are primarily produced when the vocal folds are tensioned appropriately
  - Sounds without oscillations in the vocal folds are known as unvoiced sounds



Vocal fold/vocal cord

- Glottis
  - The opening between the vocal folds (the empty space between the vocal folds)
- Subglottal area
  - the airspace between the vocal folds and the lungs

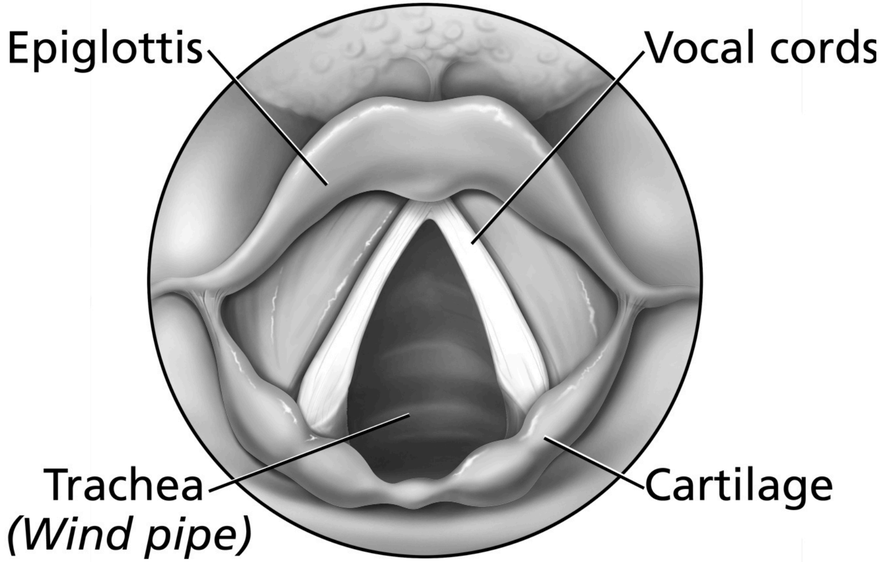


National Cancer Institute

Frequency: Vocal fold oscillation

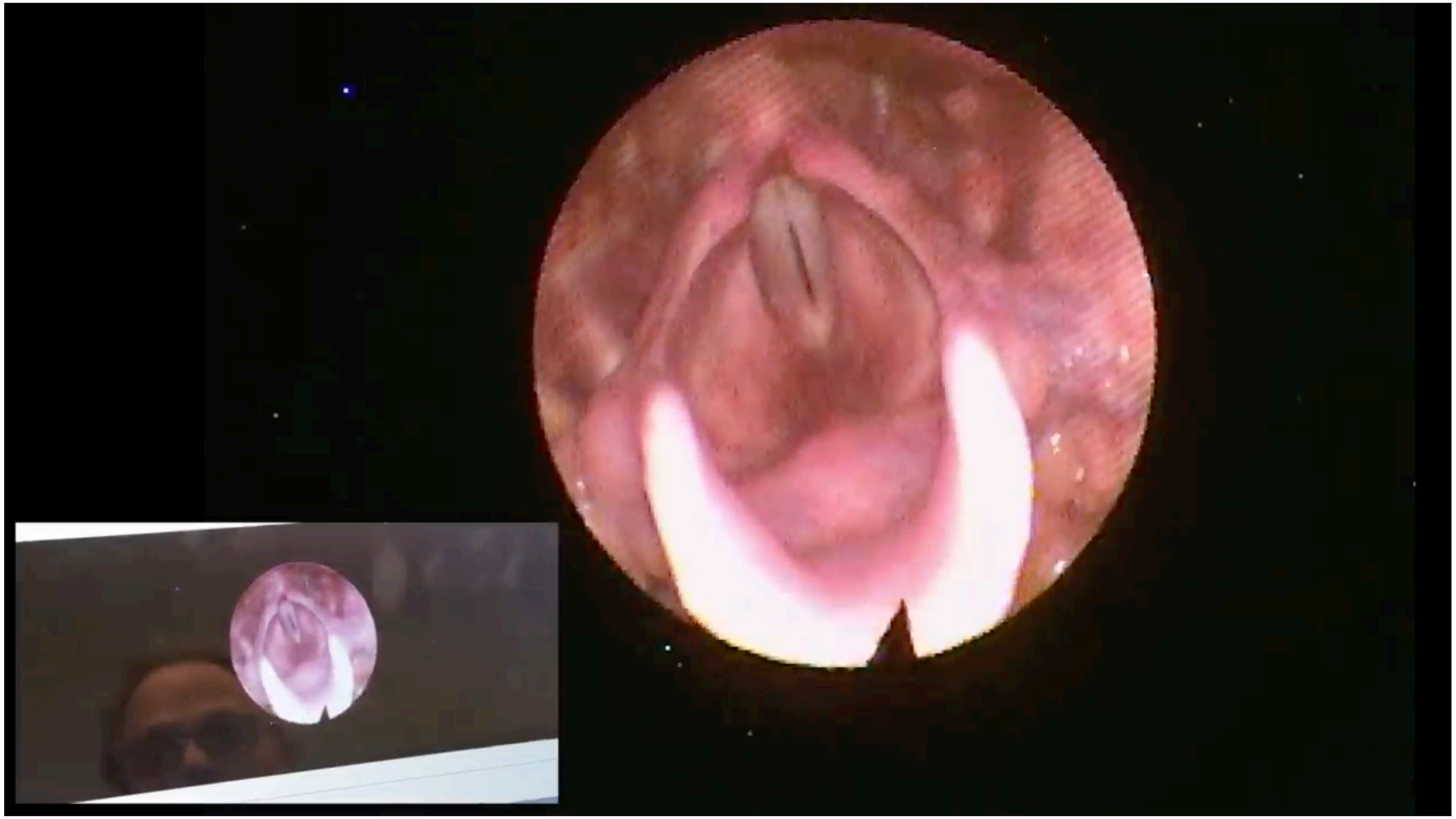
- The frequency of vocal folds oscillation depending on three main components
  - amount of lengthwise tension in the vocal folds
  - pressure differential above and below the vocal folds
  - length and mass of the vocal folds

- Pressure and tension can be intentionally changed to cause a change in frequency
- The length and mass of the vocal folds are in turn correlated with overall body size of the speaker
  - children and females have on average a higher pitch than male speakers



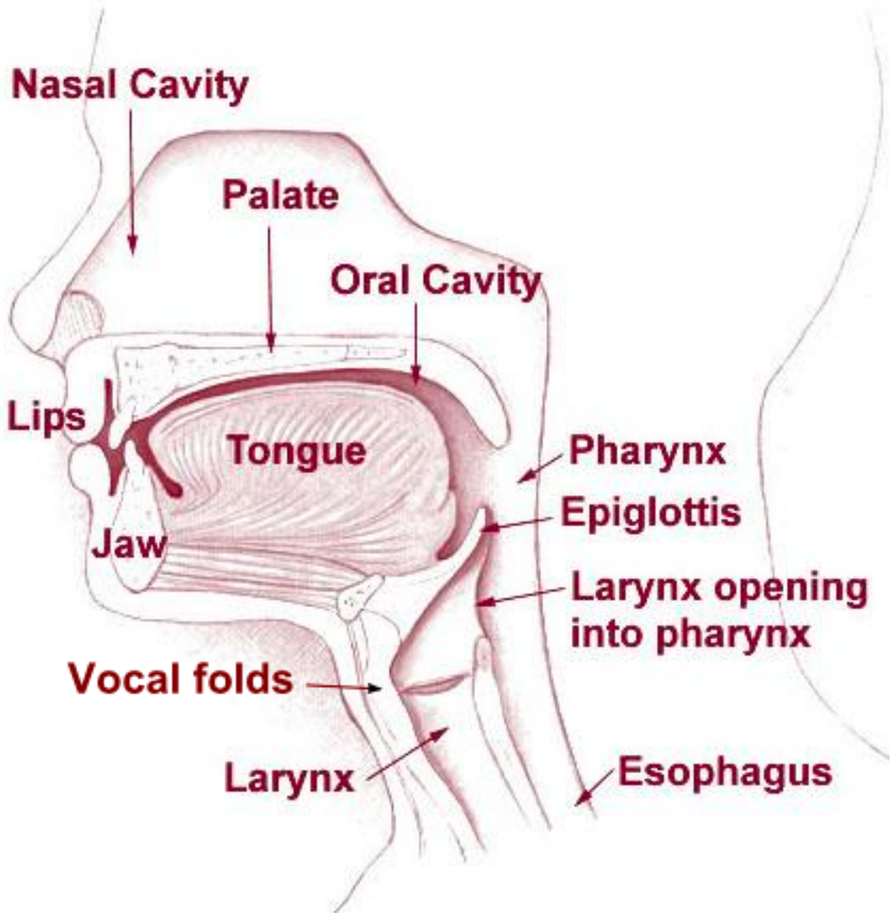
National Cancer Institute

Vocal fold



Vocal tract

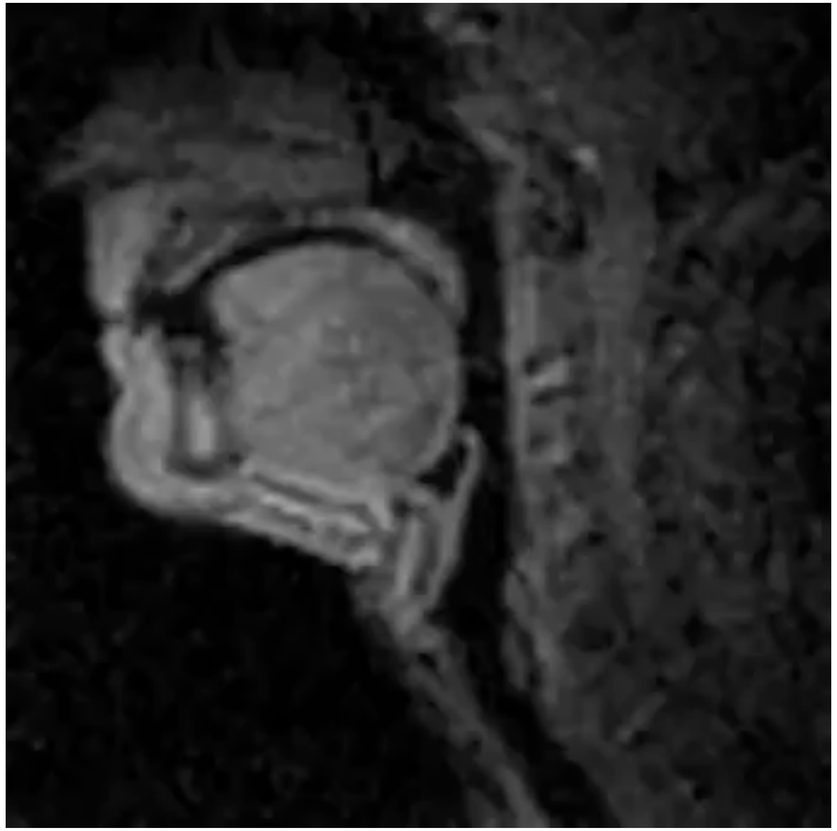
- Including the larynx, pharynx and oral cavities
- Have a great effect on the timbre of the sound
- Vocal tract doesn’t change frequency, but change the modify the air flow for different sounds

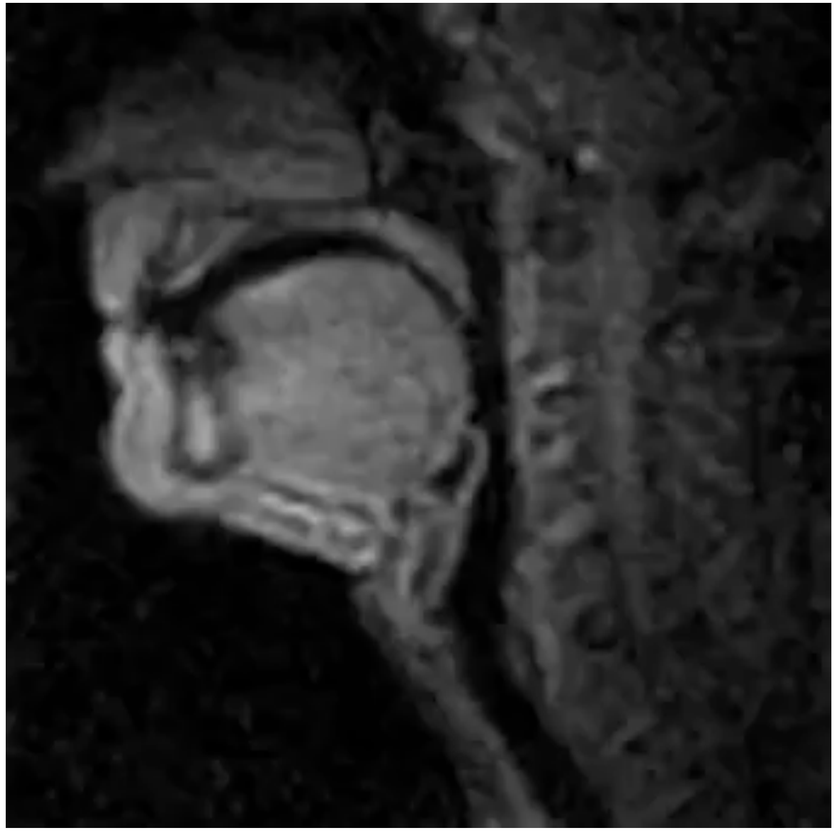


Vocal tract

- Different shapes of vocal tract result in different vowels

Head





Had

Video credit: https://sail.usc.edu/~lgoldste/General_Phonetics/

Resonance

- A resonant frequency is a natural frequency of vibration determined by the physical parameters of the vibrating object.

Vocal tract & resonance

- The shape of the vocal tract determines the resonances
- The shape is determined by a multitude of components, in particular by the position of the jaw, lips and tongue

- The resonances are easily modified by the speaker and perceived by the listener

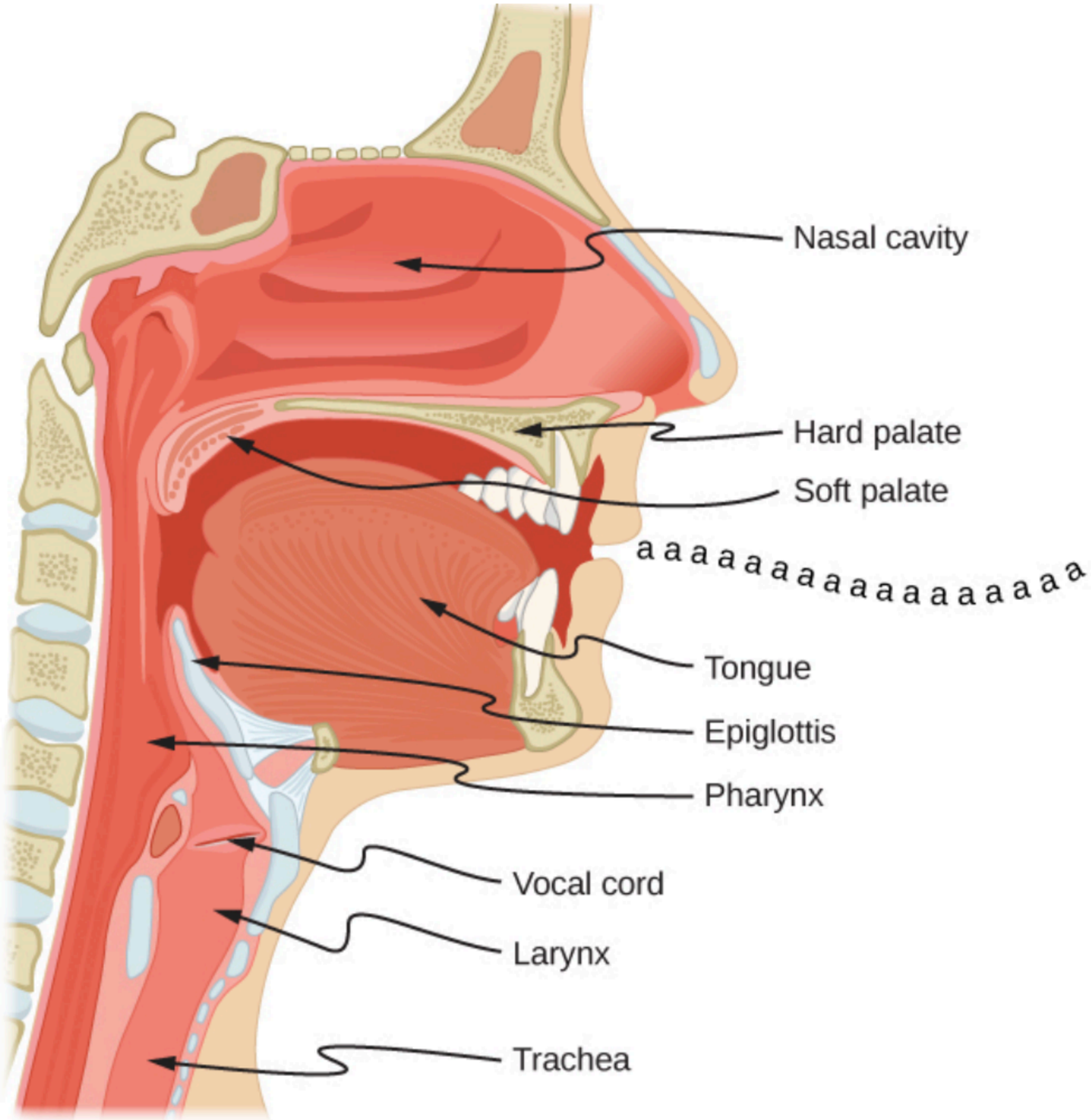
Vocal tract: Consonant sound

- In consonant sounds, there is a partial or full obstruction at some part of the vocal tract

  - Fricative consonants are characterized by a narrow gap between the tongue and front/top of the mouth
  - Plosives, the airflow in the vocal tract is fully temporarily obstructed

Speech production

- Larynx
- Vocal tract
- Brain
- Etc



Speech disorders

https://youtu.be/Wv9V8wnCOGI?t=316

Laryngectomy: removal of voice box

Source-filter model

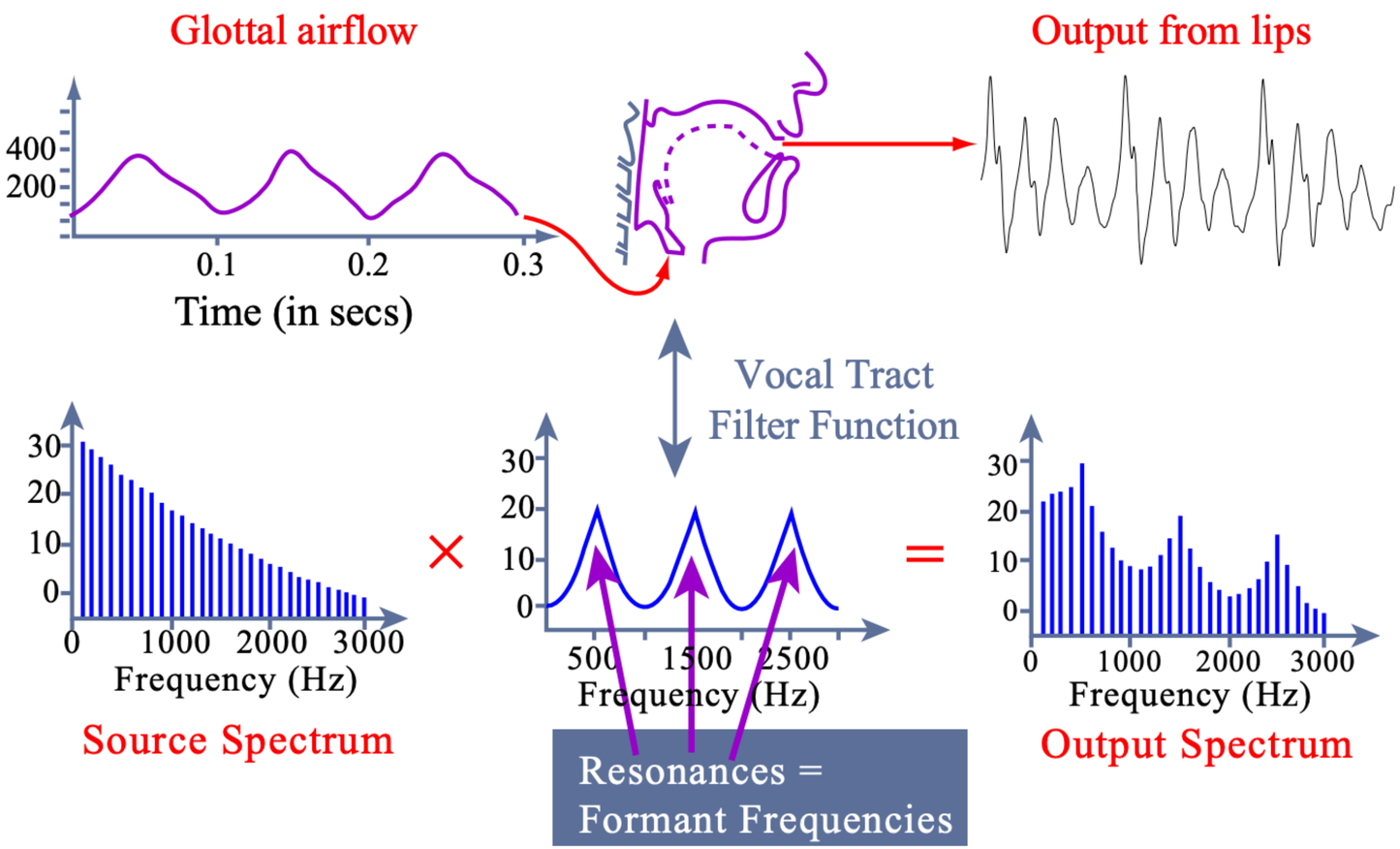
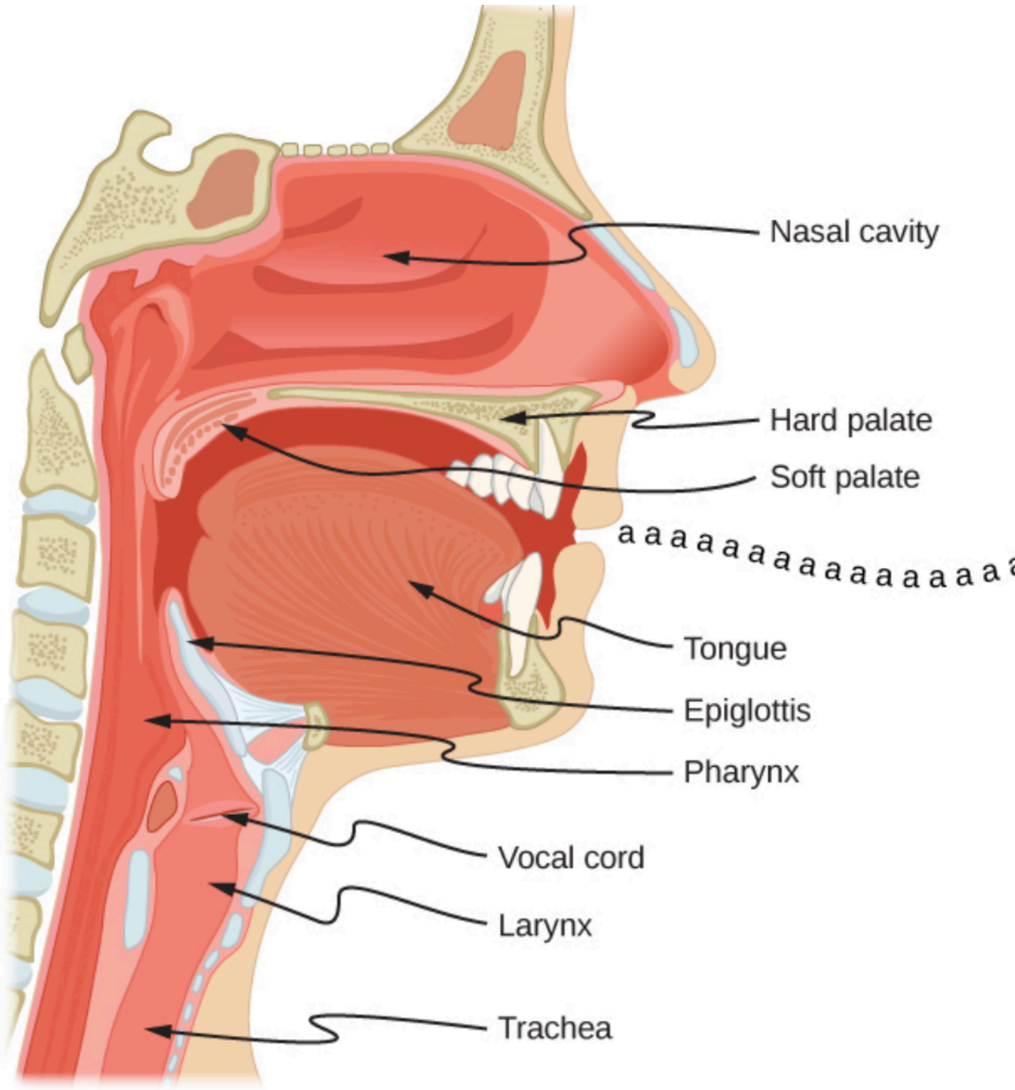


Image credit: MIT OCW

Speech production: source-filter model

- Source-filter model
  - Source produces an initial sound
  - Vocal tract filter modifies it
- Source
  - An input of acoustic energy into the speech production system
- Vocal tract filter
  - Articulators: tongue, teeth, lips, velum etc



https://www.youtube.com/watch?v=DcNMCB-Gsn8

https://www.youtube.com/watch?v=n4Y4EQaw5oU

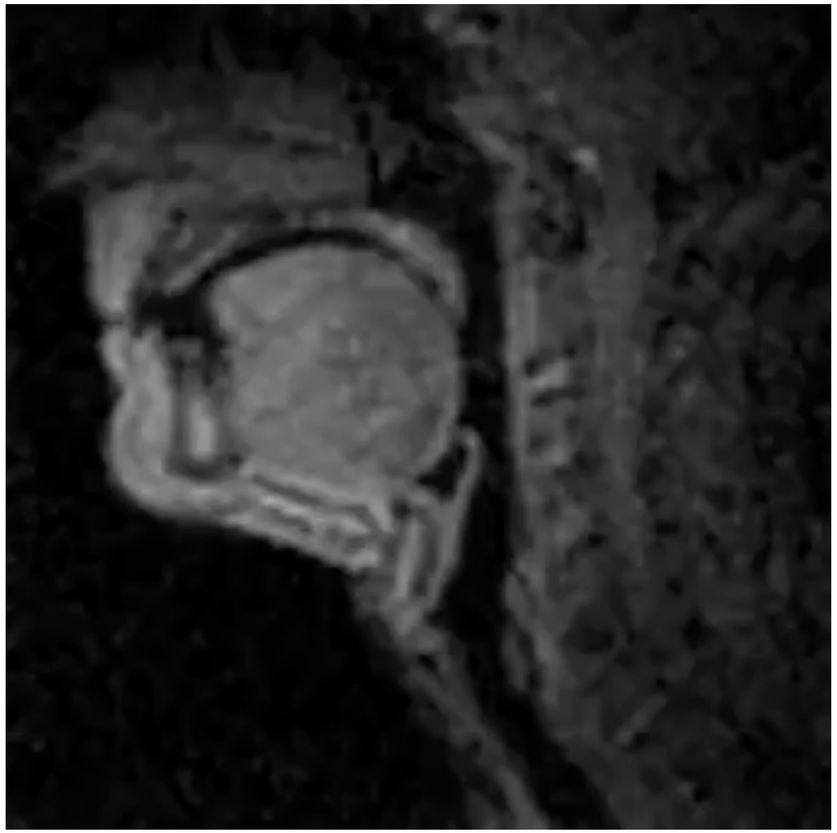
Source

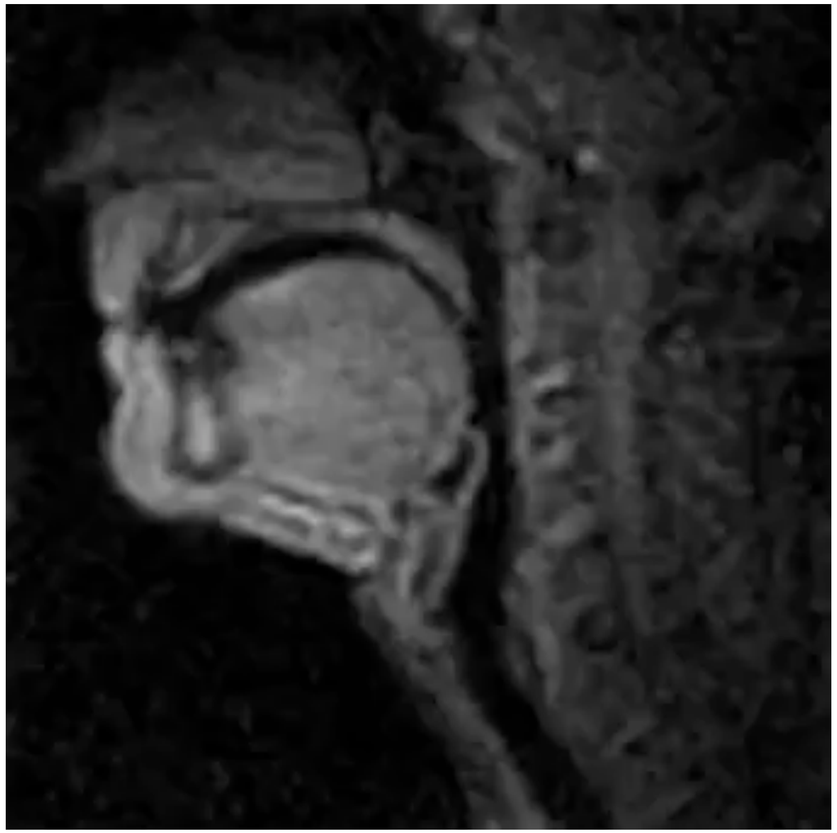
- Voicing source: Vocal folds vibrating
  - A periodic source produced by modulation of the airflow from the lungs by the vocal folds
    - The **vocal folds** are muscular folds located in the **larynx**
  - If the vocal folds are close together, then air pressure from the lungs can cause them to vibrate periodically, generating voicing.
- Unvoicing source: vocals fold holds close but not vibrating

Filter

- The vocal tract acts as a filter, modifying the source waveform
- The sound wave at some distance from the speaker is the result of filtering the source with the vocal tract filter, plus the radiation characteristics of the lips/nose.

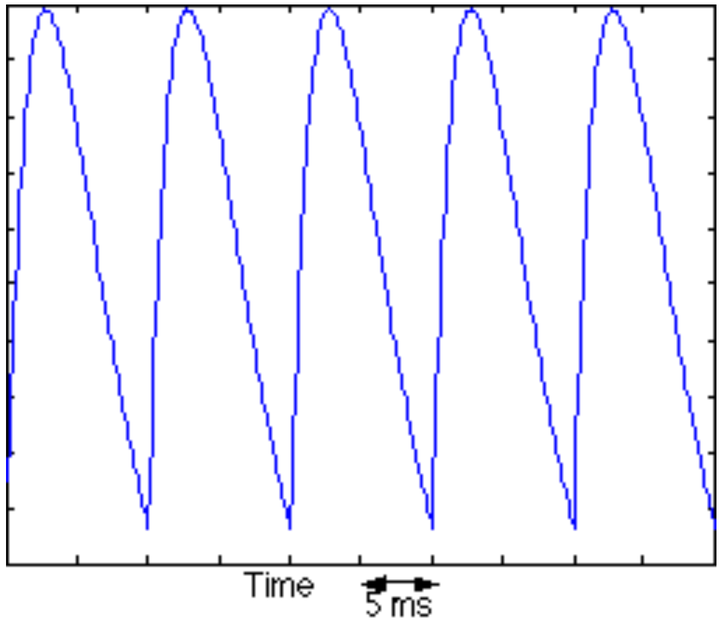
Head

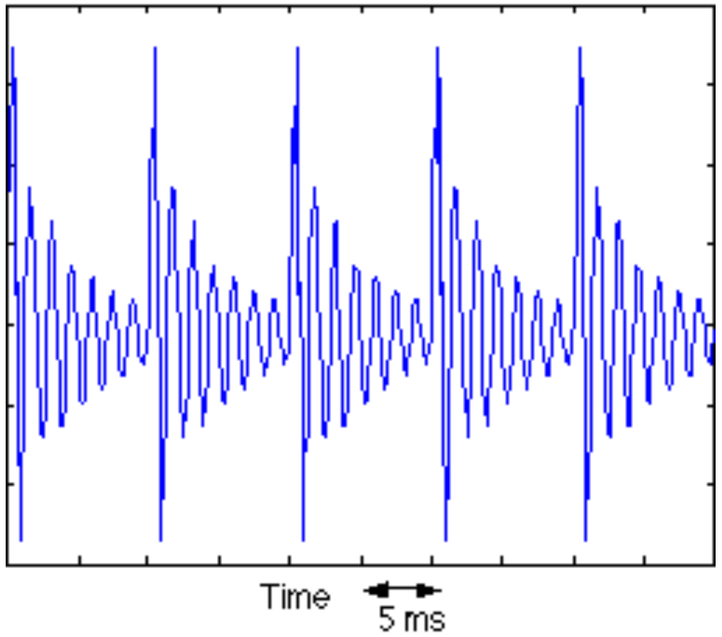




Had

Video credit: https://sail.usc.edu/~lgoldste/General_Phonetics/





Source

Filter

https://sail.usc.edu/~lgoldste/General_Phonetics/Source_Filter/MATLAB_demo/source-filter.html

Source-filter in time domain

- Convolution in time domain

$$s[t] = e[t] + \sum_{k=1}^{K} a[k]s[t-k]$$

Source

Filter

Colab example: https://colab.research.google.com/drive/18jmXe1OcdbRknGx27dh-OCeaOjafv_rt?usp=sharing

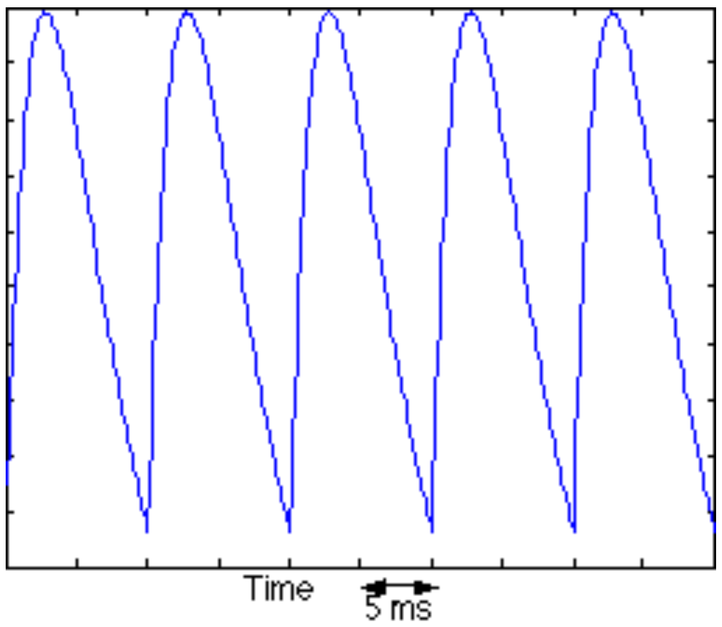
Source-filter: Multiplication in frequency domain

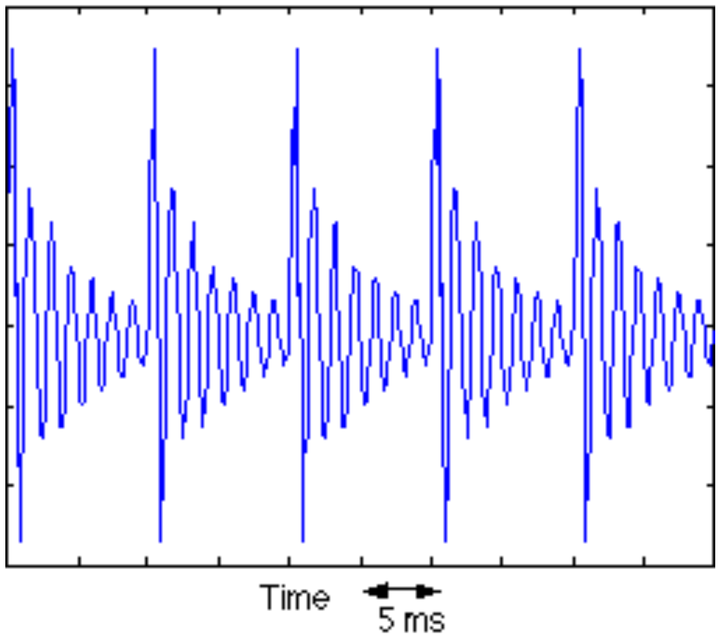
- Convolution in time domain equivalent to multiplication in frequency domain

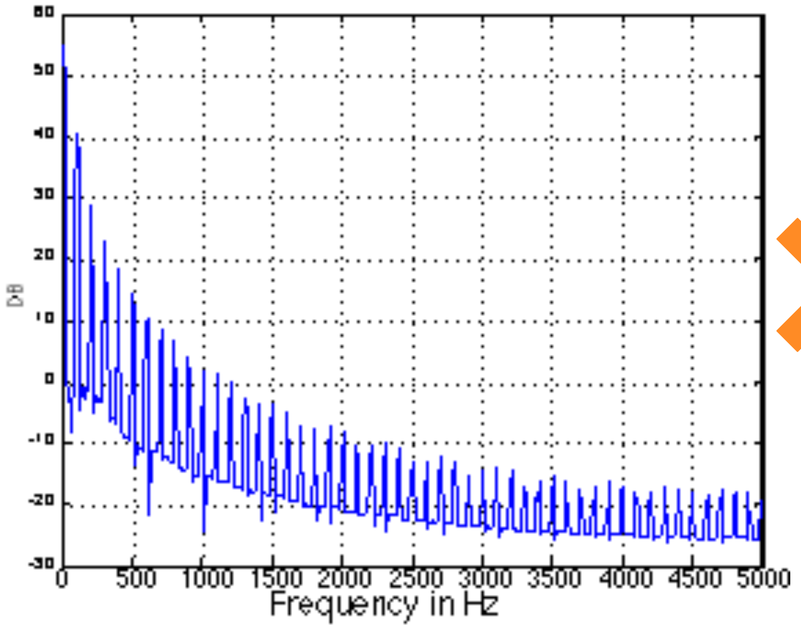
$$S(m) = H(m) \cdot X(m)$$

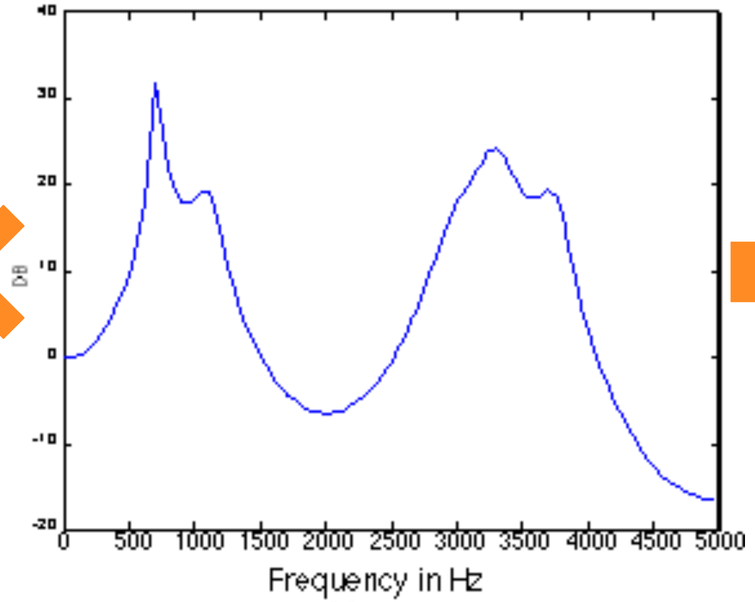
Filter

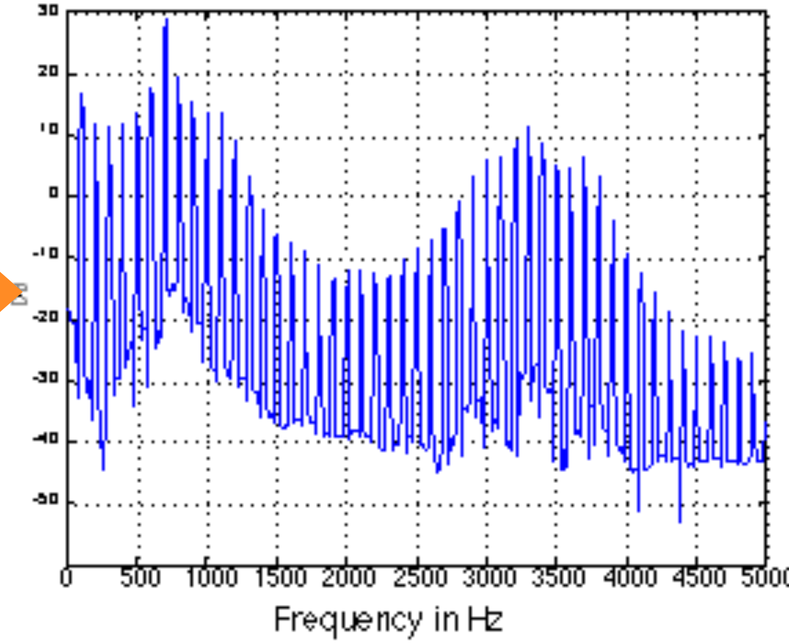
Source spectrum







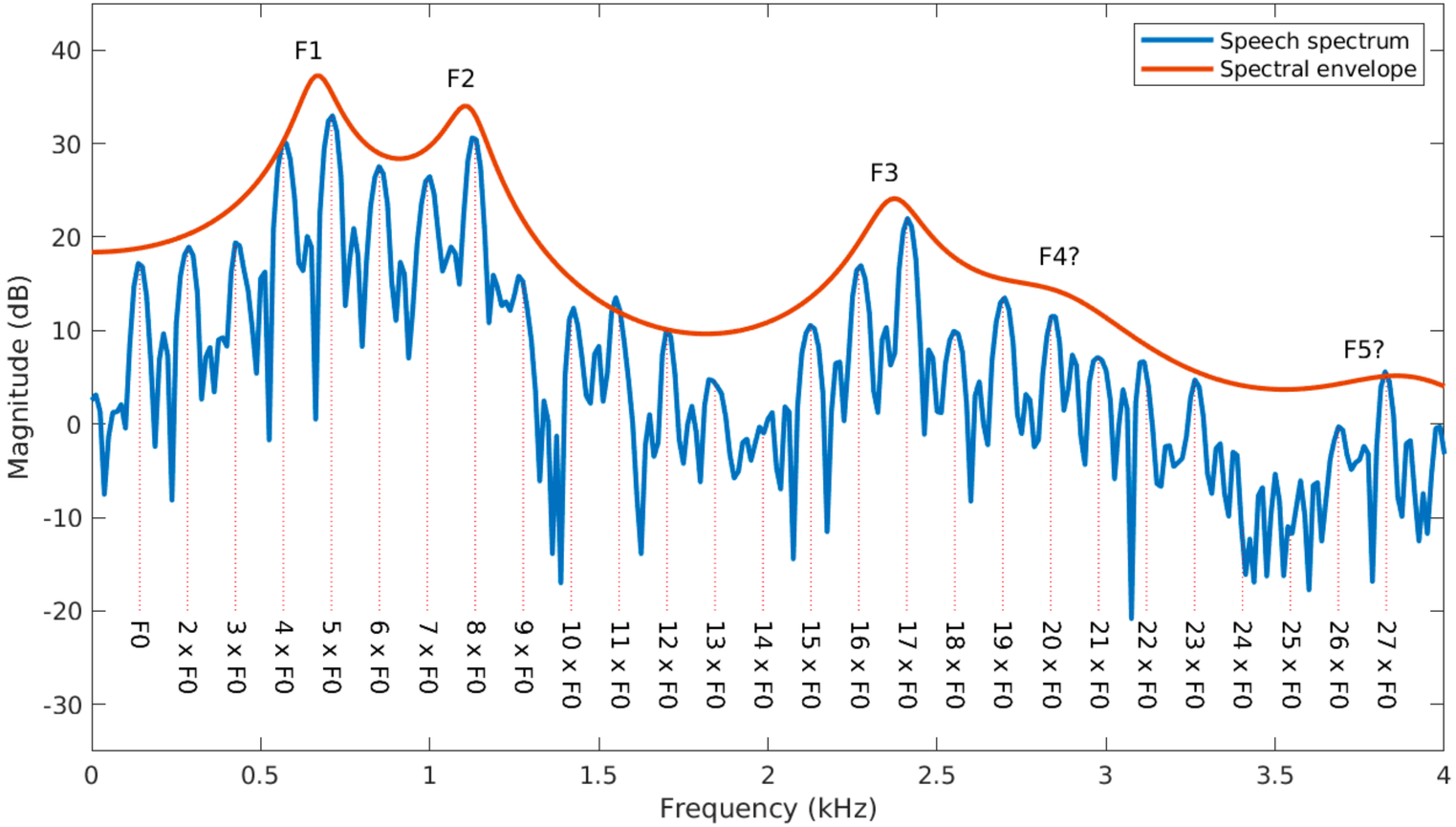




https://sail.usc.edu/~lgoldste/General_Phonetics/Source_Filter/MATLAB_demo/source-filter.html

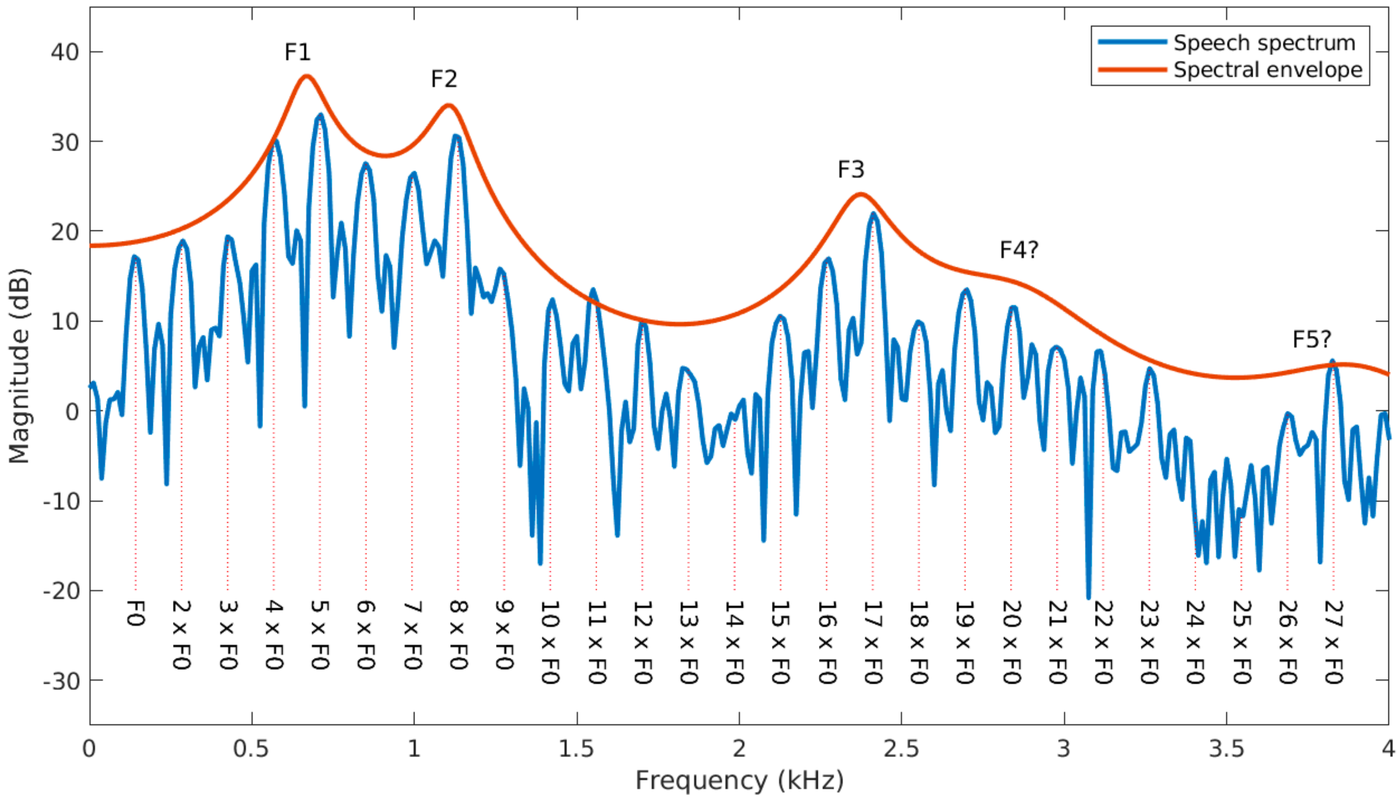
Resonance

- The resonances of the vocal tract are called **formants**
- The resonances are easily modified by the speaker and perceived by the listener



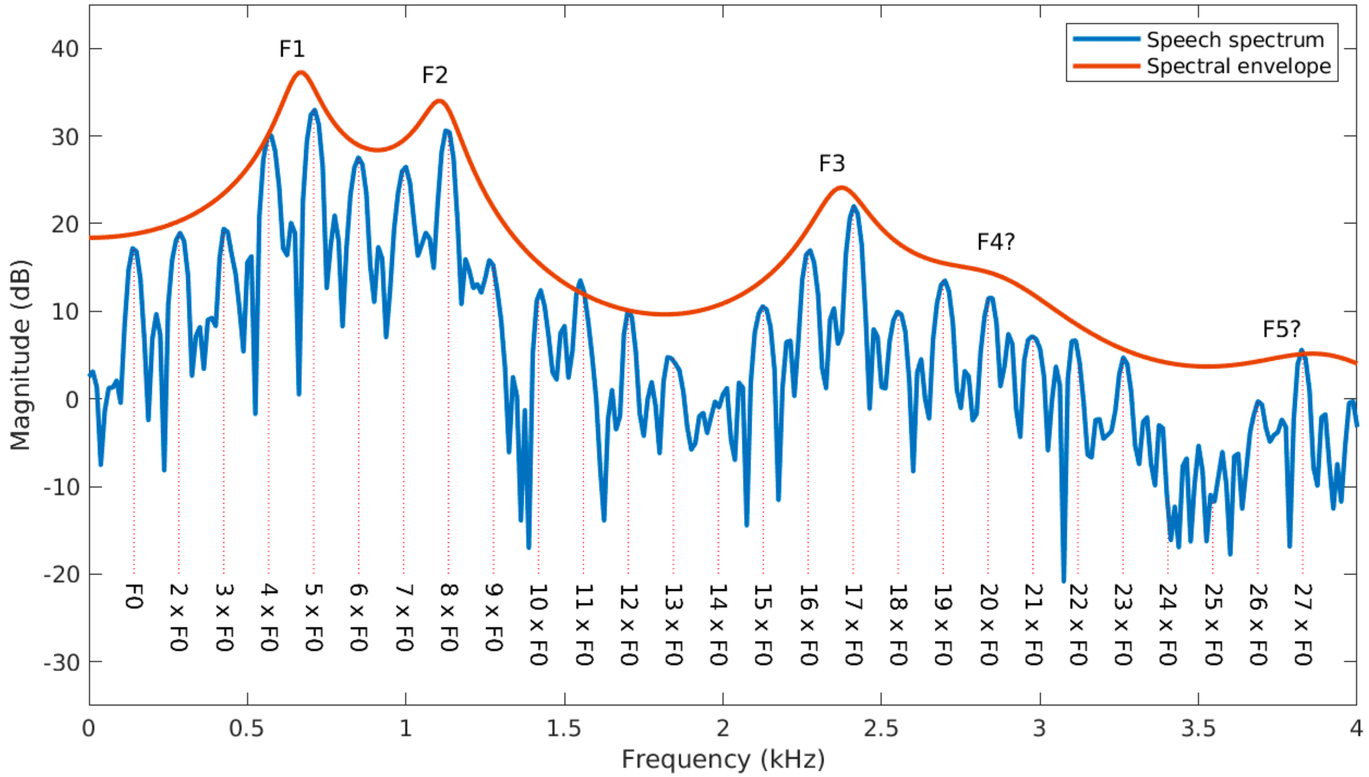
Resonance

- the acoustic features which differentiate *vowels* from each other are the frequencies of the resonances in the vocal tract, corresponding to specific *places* of articulation primarily in terms of tongue position
- The spectrum of a voiced speech have the structure of a harmonic signal



Frequency vs pitch

- Frequency of the vocal folds refers to the actual physical phenomenon
- Resonances in the vocal tract can emphasize harmonics of the fundamental frequency such that the harmonics are louder than the fundamental
- The perceived pitch is then the frequency of the harmonic instead of the fundamental



Independence of source and filter

- Source
  - Fundamental frequency (F0) is driven by the frequency of vocal fold vibrations
  - Harmonics are multiples of F0
- Filter
  - Resonances are driven by the shape of the vocal tract (physical property)
  - Formants are peaks in the spectral envelope that correspond to resonances (acoustic property)
- Independence of source and filter
  - You can change F0 without changing the vowel you are saying: harmonics change, formants stay the same

Timbre

- The characteristic quality of a sound, independent of pitch and loudness

- Spectral envelope and its time variation can represent timbre

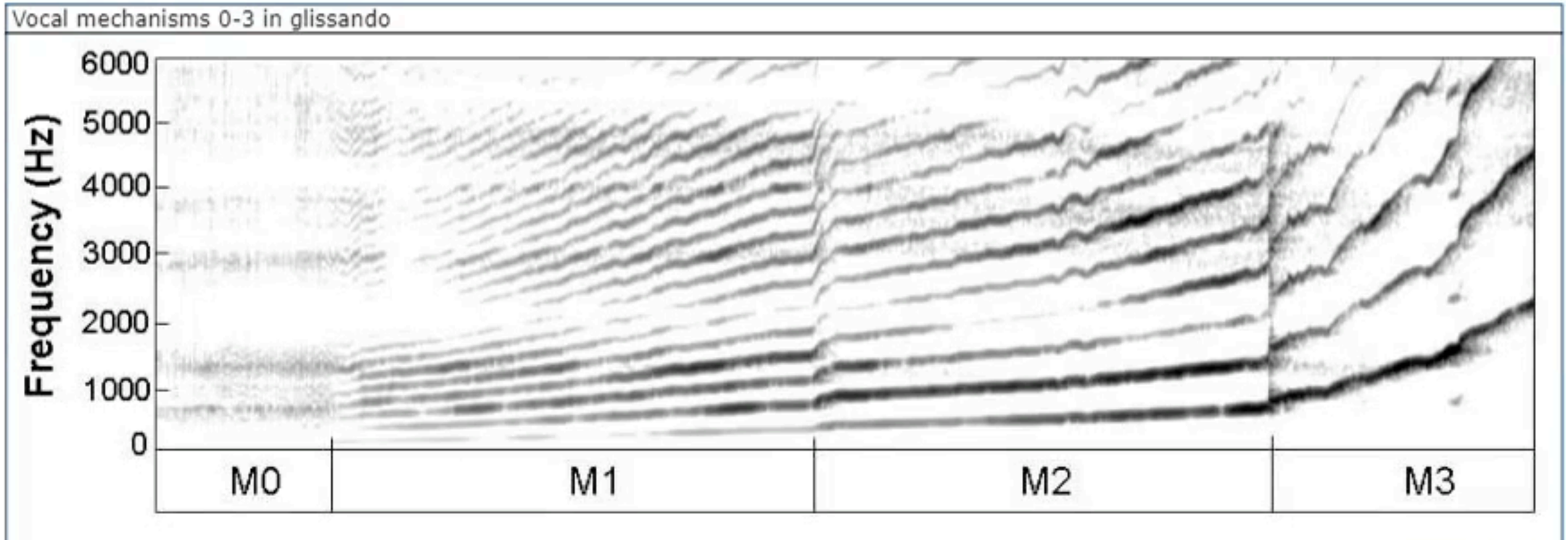
- The independence of source and filter explains
  - why vowels of the same timbre can be produced on different pitches
  - why vowels of the same pitch can have different timbres

Prosody: Melody of speech

- Same word can have different prosody
- Prosody includes pitch, duration, stress

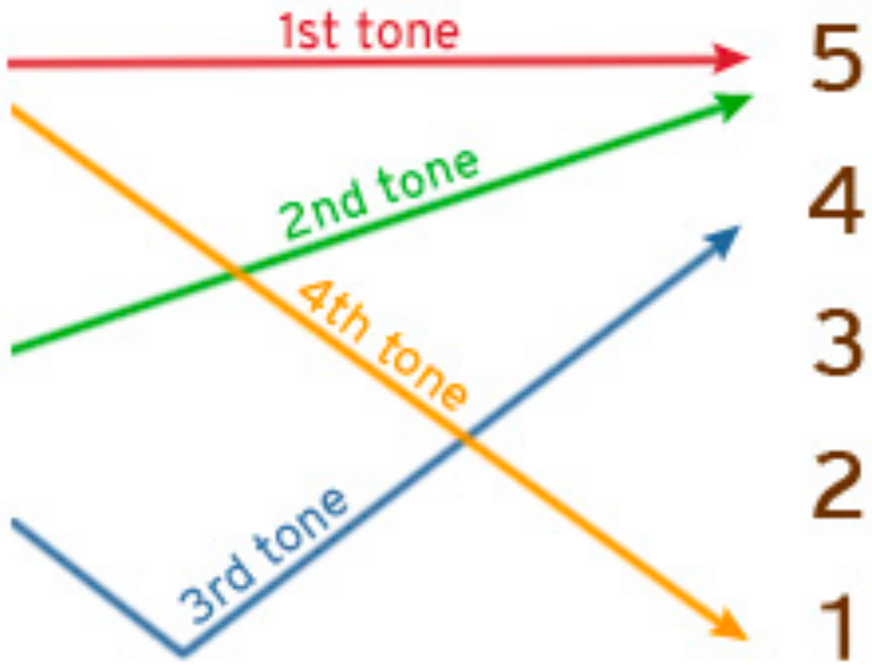
Perceived Pitch

- Fundamental frequency (F0)
  - the lowest frequency of a periodic waveform
  - F0 is driven by the frequency of vocal fold vibrations, not vocal tract resonances
- In a speech segment, F0 is semi-continuous

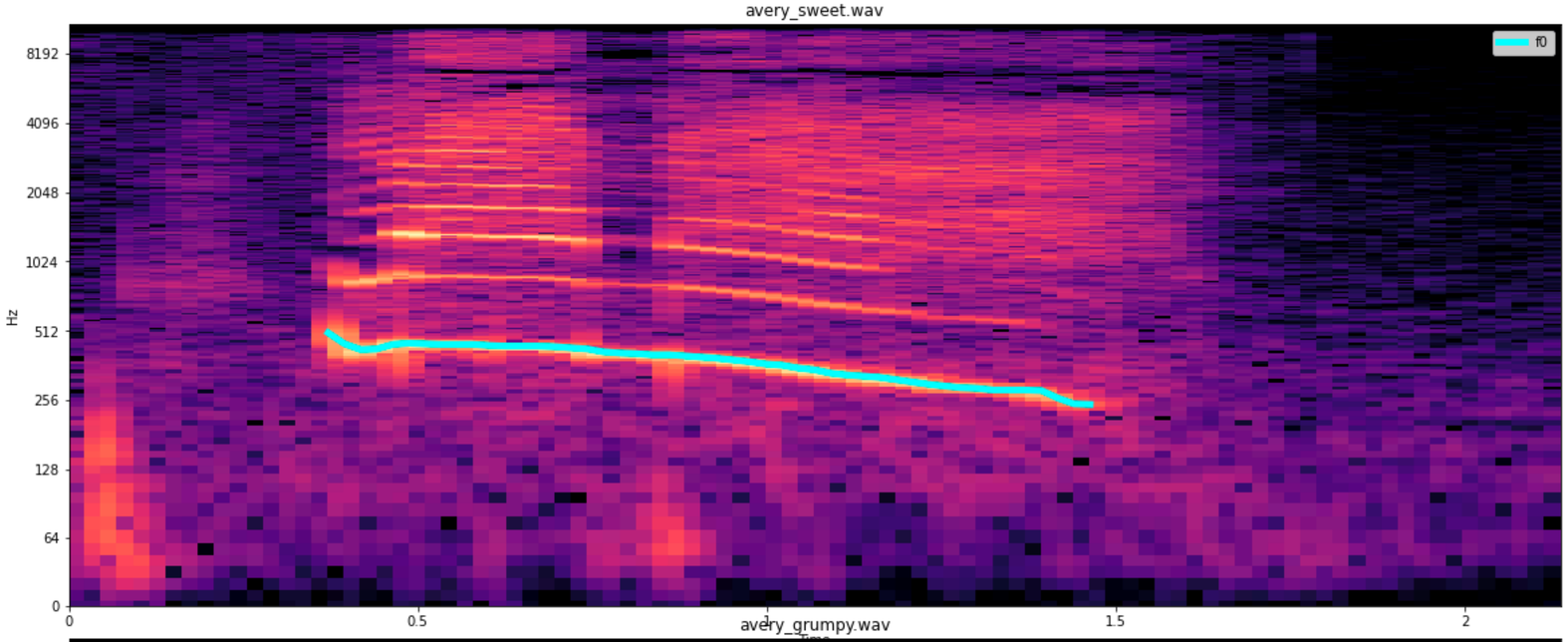


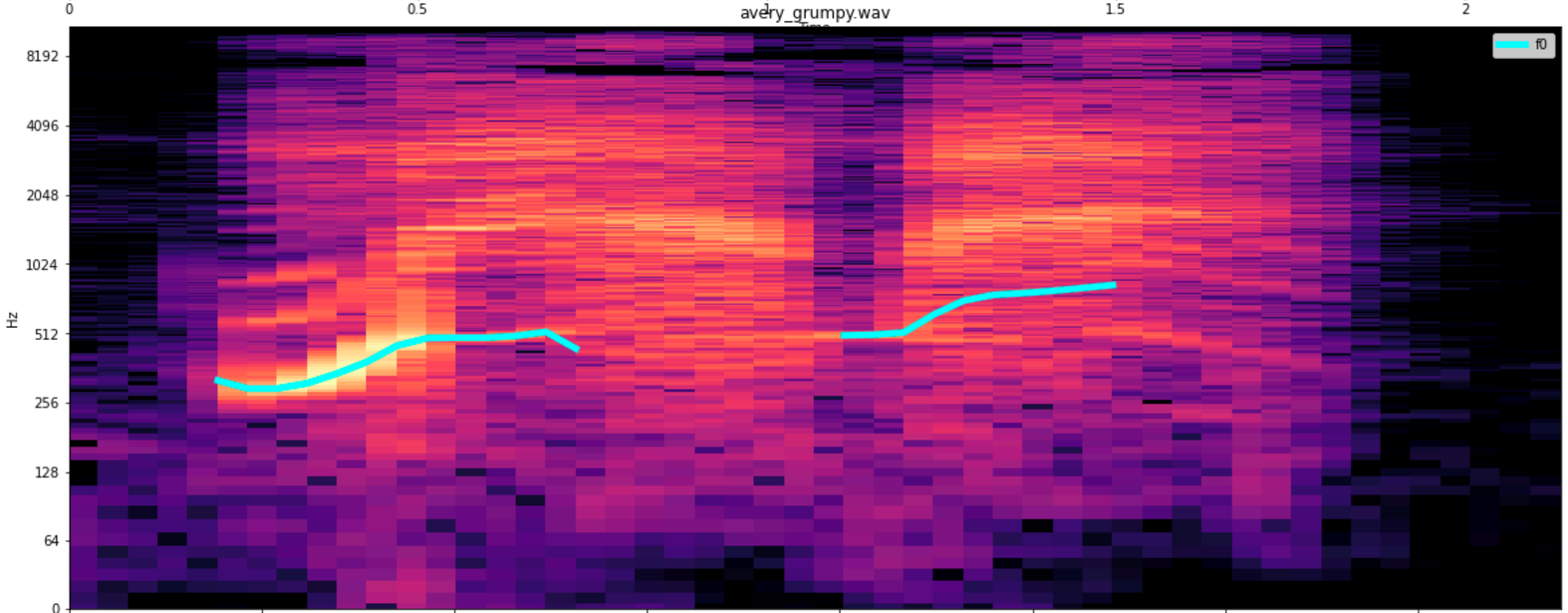
Tone

- Tonal language: different tonal inflections will convey different meanings



Intonation


avery_sweet.wav
f0
8192
4096
2048
1024
512
256
128
64
0
Hz
0
0.5
1
1.5
2


avery_grumpy.wav
f0
8192
4096
2048
1024
512
256
128
64
0
Hz

Summary

- Speech production
- Source-filter model
  - Independence of source and filter
    - vowels of the same timbre can have different pitches
    - vowels of the same pitches can have different timbre

Readings

- Chapter 2.2: Speech production and acoustic properties
  - https://speechprocessingbook.aalto.fi/Introduction/Speech_production_and_acoustic_properties.html