Speech sounds and their spectral characterization

Speech Information Systems, Exercise 2

Speech Information Systems – Speech sounds and their spectral characterization

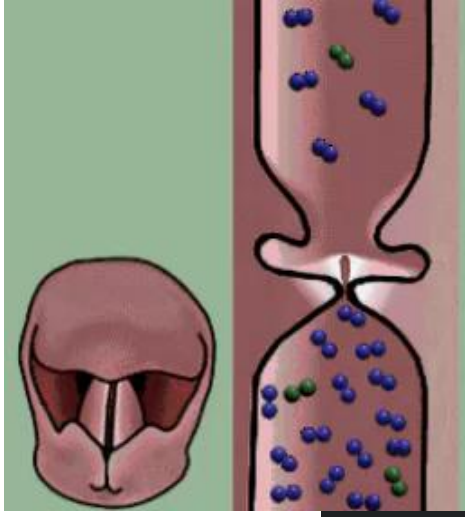
Speech Acoustics Laboratory

Gabor Kiss: kiss.gabor@vik.bme.hu
David Sztahó: sztaho.david@vik.bme.hu

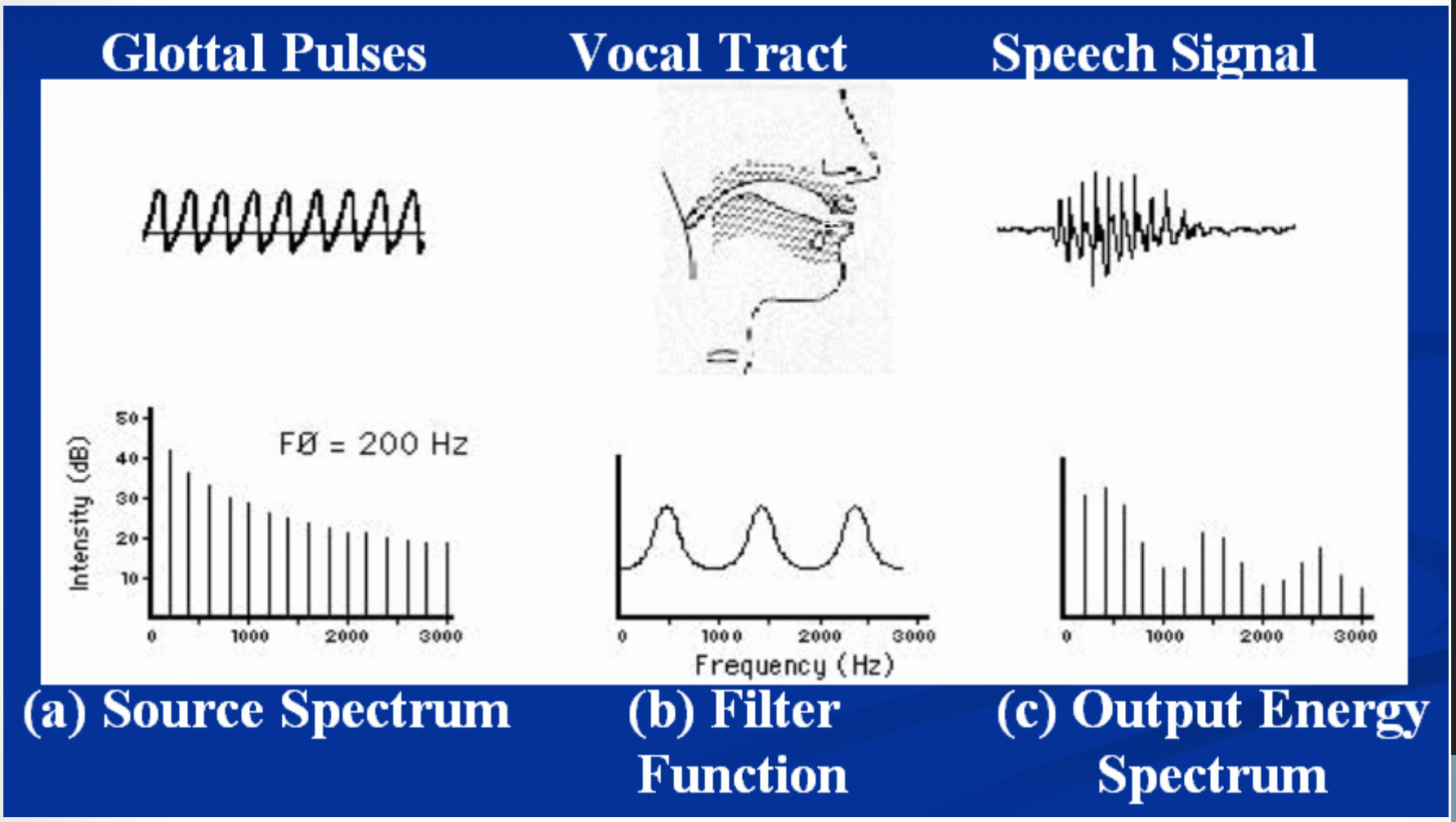
What is this exercise about?

Microsoft Form survey

Task 1



- What is the definition of **fundamental frequency** ?
  - The periodicity of the voiced sound at the vocal cords.
- How would you determine the fundamental frequency?
  - Time domain solution
    - Reading the period duration
  - Frequency domain solution
    - Reading the location of the fundamental frequency on a line spectrum
- What is a **formant**?
  - The resonance frequencies of the cavities of the vocal tract.
- How would you determine the formants F1 and F2?
  - Frequency domain solution
    - Reading the maximum locations of the envelope


Glottal Pulses
Vocal Tract
Speech Signal
FØ = 200 Hz
Intensity (dB)
50
40
30
20
10
0
1000
2000
3000
Frequency (Hz)
(a) Source Spectrum
(b) Filter Function
(c) Output Energy Spectrum

Task 2

- Answer based on what you have learned about the line spectrum.
  a) What does a **spectrum** show regarding speech?
  b) What Fourier method can theoretically determine the line spectrum?
  c) Draw the approximate spectrum of the sound pressure-time function measured in the larynx above the vocal cords when **a woman** pronounces the sound “á".
  d) How many **harmonics** can we count in the diagram of answer c) up to the point of 4000 Hz?
  e) Draw an approximation of the line spectrum of the sound pressure-time function of the sound “á" emitted from the lips (also take what you have learned about the vowel triangle as a basis).
     - Radiation resistance
     - Vowel triangle

a) What does a **spectrum** show regarding speech?
- Does it have a fundamental frequency and harmonics?
  - Yes: voiced sound
- Are there highlights in the spectrum?
  - Yes: formant structure
- Indirect conclusions:
  - Speaker gender (if the underlying tone can be read)
  - Naming a sound from reading the formants (if they can be read)

b) What Fourier method can theoretically determine the line spectrum?
- Fourier series, Fourier transform

$$f(x) \sim \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx)$$

$$a_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos kx \, dx \; (k = 0,1,2 \ldots)$$

$$b_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin kx \, dx \; (k = 0,1,2 \ldots)$$

Sawtooth sign:

$$f(x) = 2 \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\sin kx}{k}$$

c) Draw the approximate spectrum of the sound pressure-time function measured in the larynx above the vocal cords when **a woman** pronounces the sound “á”.

- ‘a' sound: vowel → line spectrum
- Female voice: fundamental frequency range approx. 150-250 Hz
- Harmonic structure: at integer multiples of the fundamental frequency
  - Amplitude reduction: 12 dB/octave
- Draw a line spectrum with a selected fundamental harmonic and harmonics for integer multiples from 150-250 Hz

d) How many **harmonics** can we count in the diagram of answer c) up to the point of 4000 Hz?

- $\frac{4000}{fundamental\ frequency} - 1 = \frac{4000}{200} - 1 = 19$
  (fundamental frequency not included!)

e) Draw an approximation of the line spectrum of the sound pressure-time function of the sound “á” emitted from the lips (also take into account what you have learned about the vowel triangle).

- Draw a spectrum with the following parameters:
  - Line spectrum
  - 150-250 Hz fundamental frequency and associated harmonics
  - Formant structure with two formants at approximately 900 and 1500 Hz
  - Amplitude reduction: 6 dB/octave

Task 3

- Outline the characteristic properties of the spectrum of a sine wave with a frequency of 110Hz and a vowel with the same fundamental frequency and explain the concepts learned in this regard!
  - Sine:
    - Line spectrum
    - One frequency component (110 Hz)
  - Vowel:
    - Line spectrum
    - Fundamental frequency + harmonics
    - Formants
  - Drawing based on parameters
  - Concepts: fundamental frequency, harmonics, formants

Task 4

- Describe the process of acoustic formation of the vowel and the formation of the final spectral image.
  - Steps:
    - The path of air from the lungs, through the vocal cords and the vocal cavity, through the mouth and into the open space
    - Vocal cords: creating sound
    - Vocal cavity
      - Formants, radiation resistance
    - Mouth: passes through almost unhindered

(theory: speech as an acoustic product)

- Illustrate it using any vowel! Use diagrams too!
  - See previous figure

Task 5

- When analyzing a speech fragment, we see a line **spectrum**. What data can we infer?
  - it is a voiced sound
    - Is the spectrum linear? Is there a fundamental frequency visible?
  - It is a vowel
    - Is the spectrum linear? Is there a fundamental? Are there formants?
    - Difficult question!
  - it is a consonant
    - Is the spectrum noisy?
  - the length of a sound
    - Do we see temporal information on a spectrum?
  - age of the speaker
    - How does age affect speech?
  - it is in a declarative or question sentence
    - Modality is determined by: prosody
      - Can you see this?

Task 6

- In a telephone system, the following harmonics were measured: 400Hz, 800Hz, 1000Hz. What could be the fundamental frequency?
  - Important note: Did we ONLY measure these, or could there be harmonics BESIDES THESE?
  - Harmonics are integer multiples of the fundamental frequency.
  - Solution options
    - 150Hz
    - 100Hz
    - 200Hz
    - 400Hz
    - 20Hz
    - 80Hz

Task 7

- A **male** radio announcer from Kossuth Radio is being examined in the laryngology department. He has to pronounce the *sad* word, which are recorded first with a high-quality microphone, and secondly via a landline telephone. The audio recordings are analyzed. Answer the questions asked with justification.
  a) What can be the fundamental frequency of the announcer?
     - Male announcer: average fundamental frequency of about 100 Hz
  b) Draw the pressure-time function of the announcer's voice. Characterize it. 2 cm=10 ms
     - Figure parameters:
       - Drawing a waveform, quasi-sawtooth signal, period: 0.01 s
  c) How many harmonics can we count in the sound “a” in the microphone recording?
     - Harmonics are integer multiples of the fundamental frequency.
     - Depends on the sampling frequency:

$$\#harmonics = \frac{sampling\ frequency/2}{fundamental\ frequency} - 1$$

harmonics can be read up to about 4000 Hz)

d) How many harmonics can we count in the sound “a” on the landline telephone recording?
   - Telephone speech transmission frequency: 300-3400 Hz
   - This may not include the first harmonic!

e) Draw the spectrum measured in the middle of the vowel of the sound sequence recorded with the microphone. Specify the points of the spectrum characteristic of the vowel.
   - See previous task

f) Draw the spectrum measured in the middle of the first sound of the audio sequence recorded with the microphone. Specify the characteristic points of the spectrum.
   - “s":
     - Noise-like spectrum, emphasized frequency range: from approx. 2500 Hz

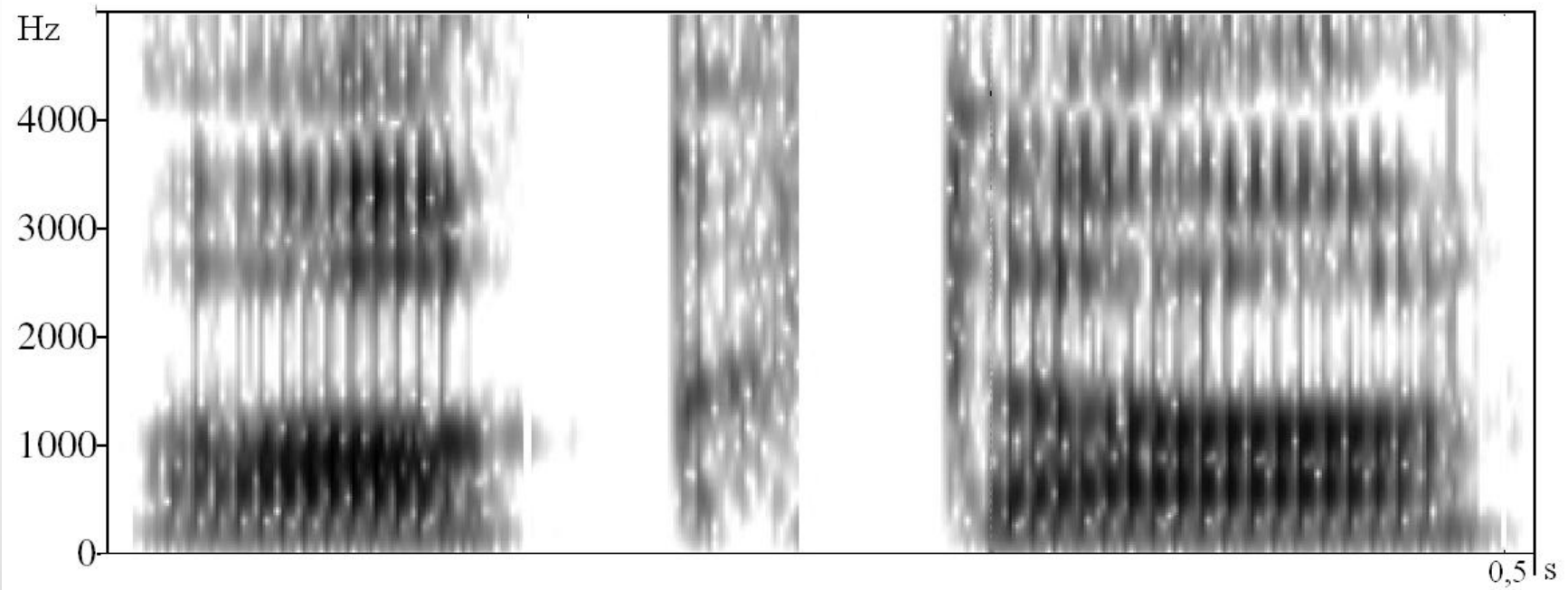
g) Describe with sound symbols what the doctor can hear when listening to the sound sequence obtained from the two types of audio recordings.
   - phone band recording cuts off most of the energy
   - Although the fundamental frequency itself is cut off by the telephone band filter, we will hear the harmonics
   - We will probably understand.
     - Our brain “completes” it based on our learned vocabulary
   - If the context would be confusing, then any sound that has the same formation, differing only in the upper frequencies (e.g. "f") is good.

Task 8

- You need to generate the time function of the word " *kacsa"* from the time function of the word *"kasa"* using a standard audio editing program. How do you complete the task?
  - "s" → "cs"
    - Remains unvoiced
    - ficative sound → affricate sound
    - Duration difference
- Sounds heard according to the duration of friction noise
  - t > 100 ms → fricative sound
  - 50 < t < 100 ms → affricate
  - t < 50 ms → stop sound
  - BUT: the producing location is the same

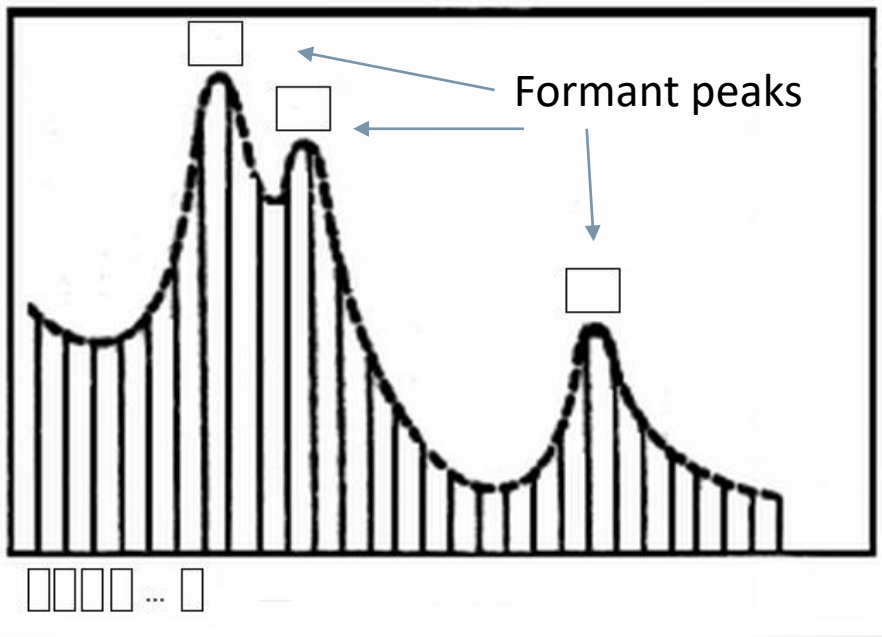
Task 9

- Analyze the meaningful English word shown **in the spectrogram below**.
  - Draw the boundaries of speech sounds. Give reasons.
  - Describe the characteristics of speech sounds based on what you have learned.
  - Try to define the tone sequence. Give reasons.
  - For each sound, specify the type of excitation (in detail) and location.
  - Is the spectrogram a female, male, or child's speech? Explain your



Task 10

- Complete the diagram by writing the appropriate symbols in the boxes. Explain the concepts you have learned in this regard.
- To which type of speech sound does the spectrum shown in the figure belong?
  - Line spectrum
    - Fundamental frequency: sonor
  - Clearly visible formant structure
    - vowel



Task 11

- In what frequency range is the spectrum of the “sz” sound most intense?
  - approx. 30Hz-100Hz
  - approx. 100Hz-300Hz
  - approx. 300Hz-1000Hz
  - approx. 1000Hz-3kHz
  - approx. 3kHz-10kHz
  - approx. 10kHz-20kHz

Praat tutorial

https://www.fon.hum.uva.nl/praat/

Thank you for your attention!

Questions?
Test!

Speech sound and their spectral characeristics

https://forms.office.com/e/CuePTZDiG2