Erasmus Mundus

Master Ofiziala
Official Master'sDegree

Hizkuntzaren Azterketa eta Prozesamendua (HAP)

Language Analysis and Processing (LAP)

http://ixa.si.ehu.es/master

Language and Communication Technologies

Universidad del País Vasco / Euskal Herriko Unibertsitatea

MDe
Master eta Doktorego Eskola
Escuela de Máster y Doctorado
Master and Doctoral School

Erasmus Mundus

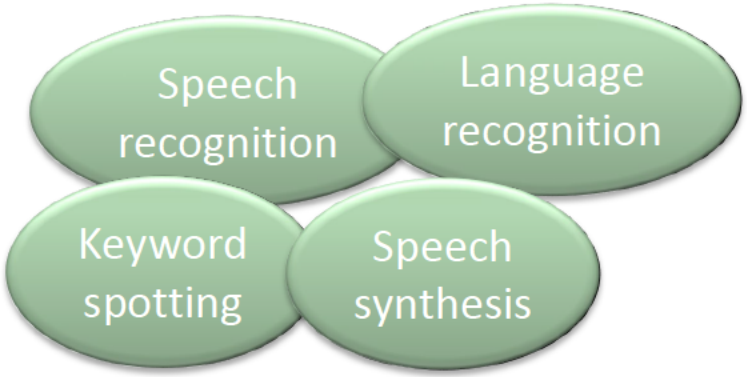
Speech Processing

Introduction: Information in the speech signal

- What kind of information can humans (or machines!) extract from a speech signal?

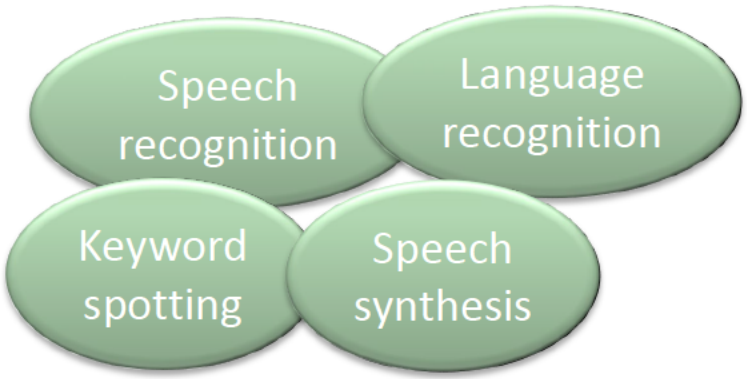
Introduction: Information in the speech signal

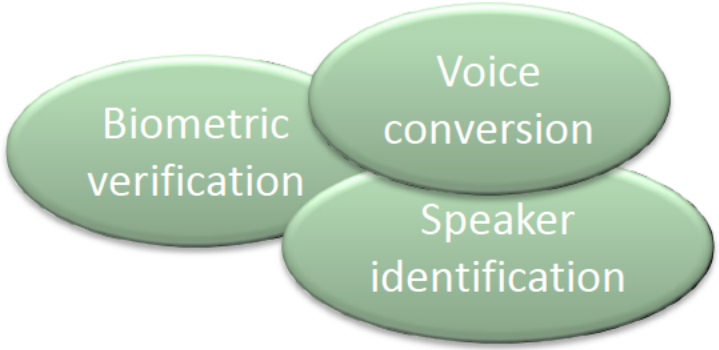
- What kind of information can humans (or machines!) extract from a speech signal?
  - Message or linguistic information



Introduction: Information in the speech signal

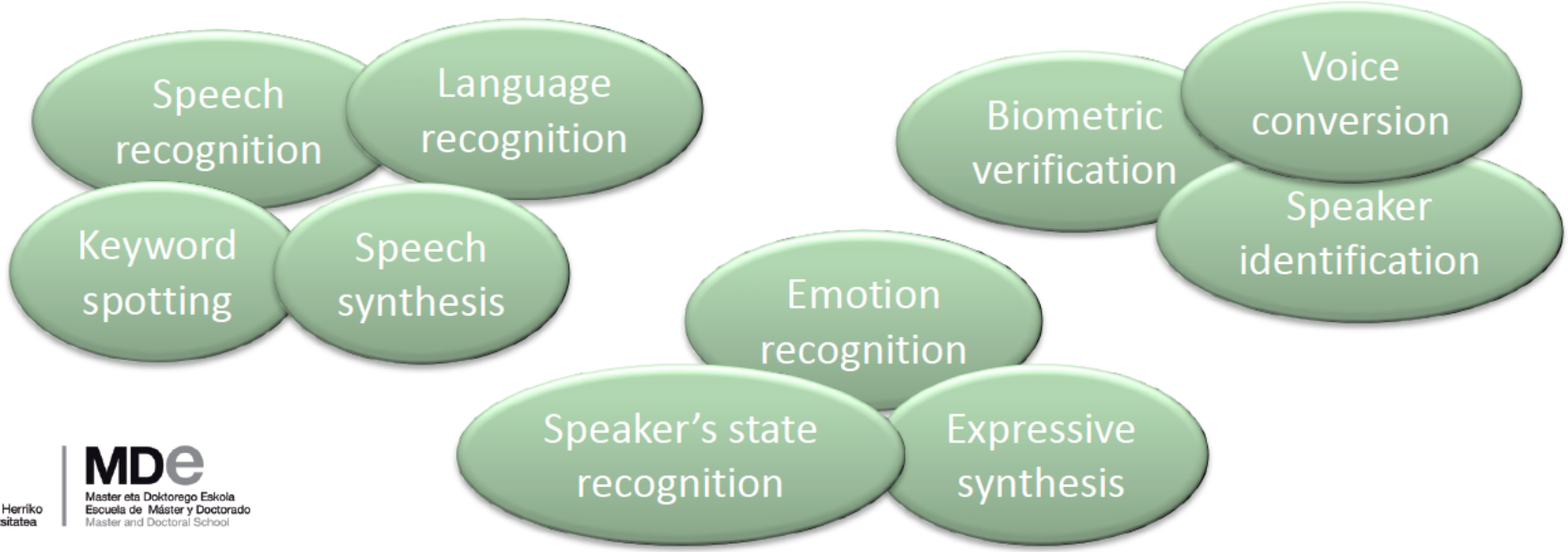
- What kind of information can humans (or machines!) extract from a speech signal?
  - Message or linguistic information
  - Identity of the speaker





Introduction: Information in the speech signal

- What kind of information can humans (or machines!) extract from a speech signal?
  - Message or linguistic information
  - Identity of the speaker
  - Paralinguistic information: emotional state, etc.



Introduction: Information in the speech signal

- What kind of information can humans (or machines!) extract from a speech signal?
  - Message or linguistic information
  - Identity of the speaker
  - Paralinguistic information: emotional state, etc.
- The acoustic realization of this high-level info involves a large variety of measurable low-level features

Syllabus

- Lesson 1: Speech production and perception
- Lesson 2: Basic concepts about signals & systems (Part I & II)
- Lesson 3: Speech signal: representations

Speech production and perception

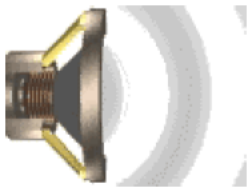
1. Signals and sounds
    1. Basic concepts
    2. Time and Frequency
    3. Sound Pressure Level
2. Speech production
    1. The human phonation system
    2. Voiced and unvoiced sounds
    3. Formants and coarticulation
    4. Nasal sounds
3. Acoustic phonetics
    1. The alphabets
    2. Articulatory mode and place
4. Speech perception
    1. The human auditory system
    2. Pitch perception
    3. Loudness and phons
    4. Masking effect
    5. Temporal and frequency discrimination
    6. Other speech perception effects

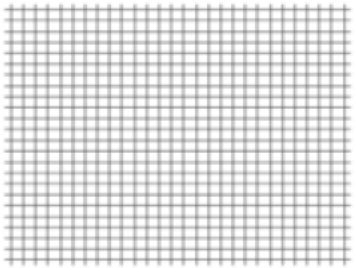
Speech production and perception

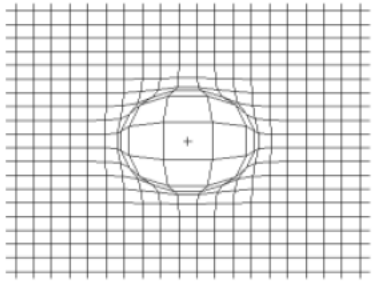
1. Signals and sounds
   1. Basic concepts
   2. Time and Frequency
   3. Sound Pressure Level
2. Speech production
   1. The human phonation system
   2. Voiced and unvoiced sounds
   3. Formants and coarticulation
   4. Nasal sounds
3. Acoustic phonetics
   1. The alphabets
   2. Articulatory mode and place
4. Speech perception
   1. The human auditory system
   2. Pitch perception
   3. Loudness and phons
   4. Masking effect
   5. Temporal and frequency discrimination
   6. Other speech perception effects

Signals and sounds

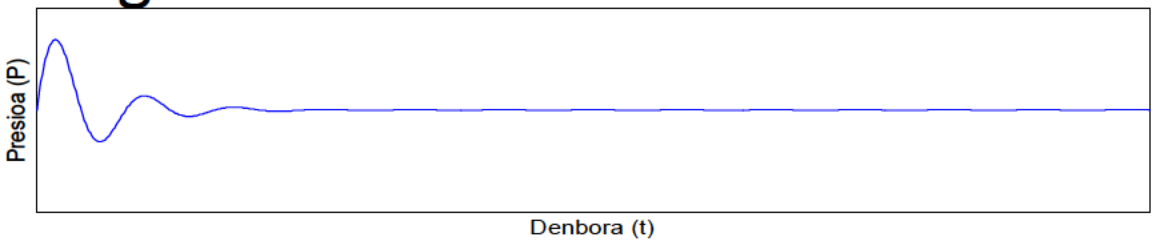
- The sound is the pressure wave generated by a mechanical vibration of some object. The wave is usually propagated by a fluid (air or water).





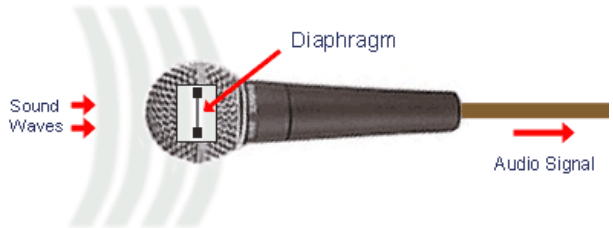


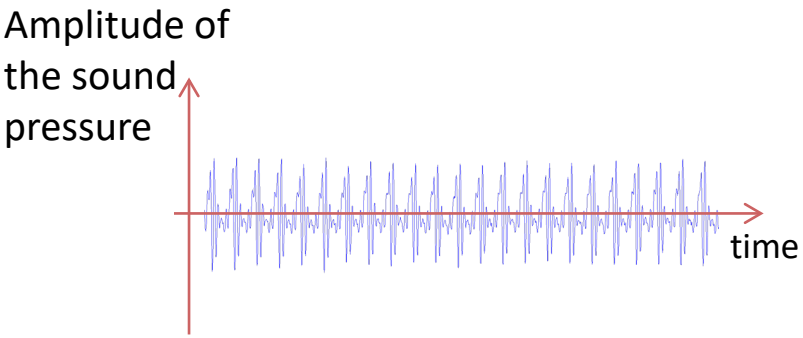
- The measure at every moment of the pressure at a certain point results in a function that represent the sound: a signal.



Signals and sounds

- The microphone converts the mechanical pressure into a electrical signal.





- The electrical signals can be digitized and then stored, transformed or transmitted.

Time and frequency

- The simplest vibration in nature is the sinusoid

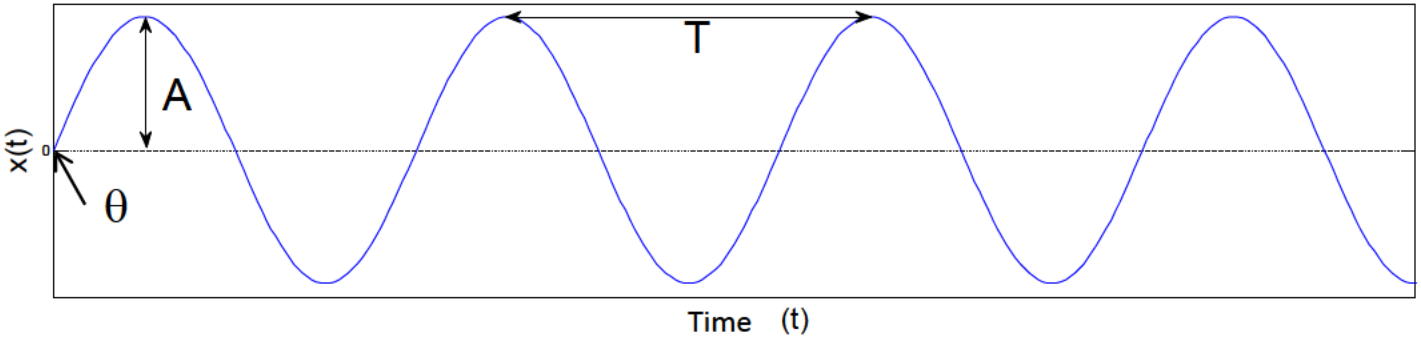
$$s(t) = A\sin(2\pi f_0 t + \theta)$$

A: amplitude

θ: phase

f_0 (Hz): frequency

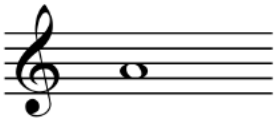
$T=1/f_0$: period

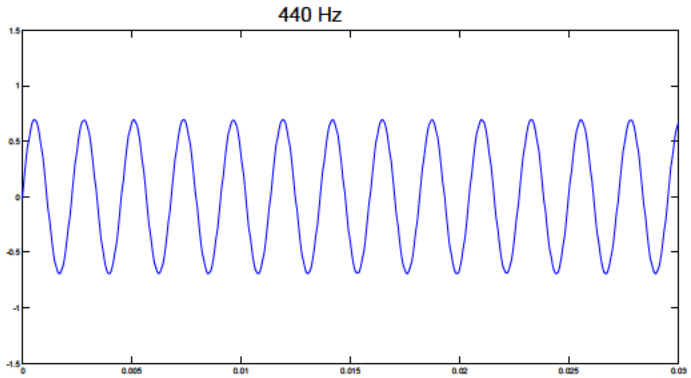


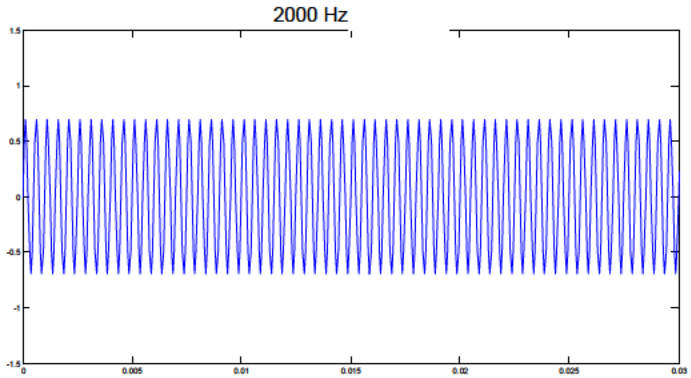
Chapter01, sin_demo1

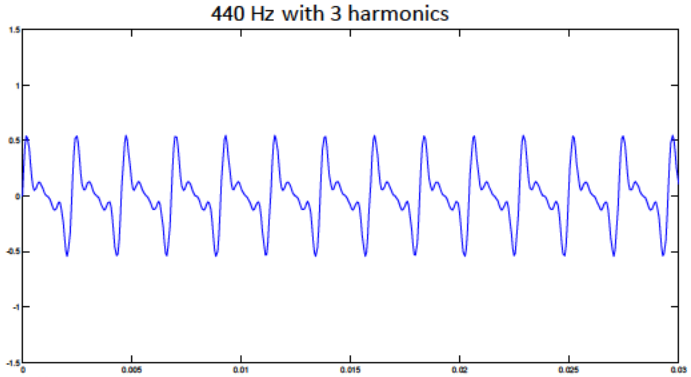
Time and frequency

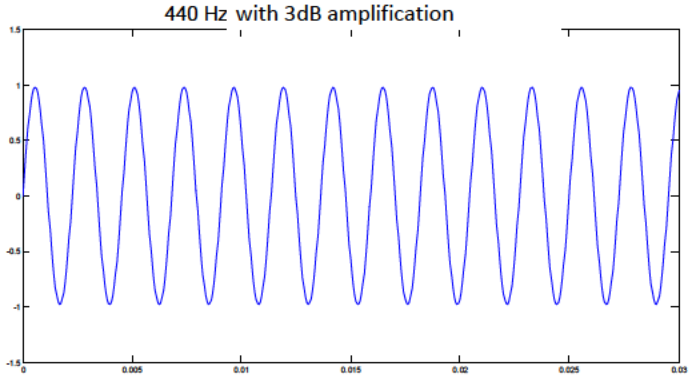
- Relation of sinusoids and sound in time









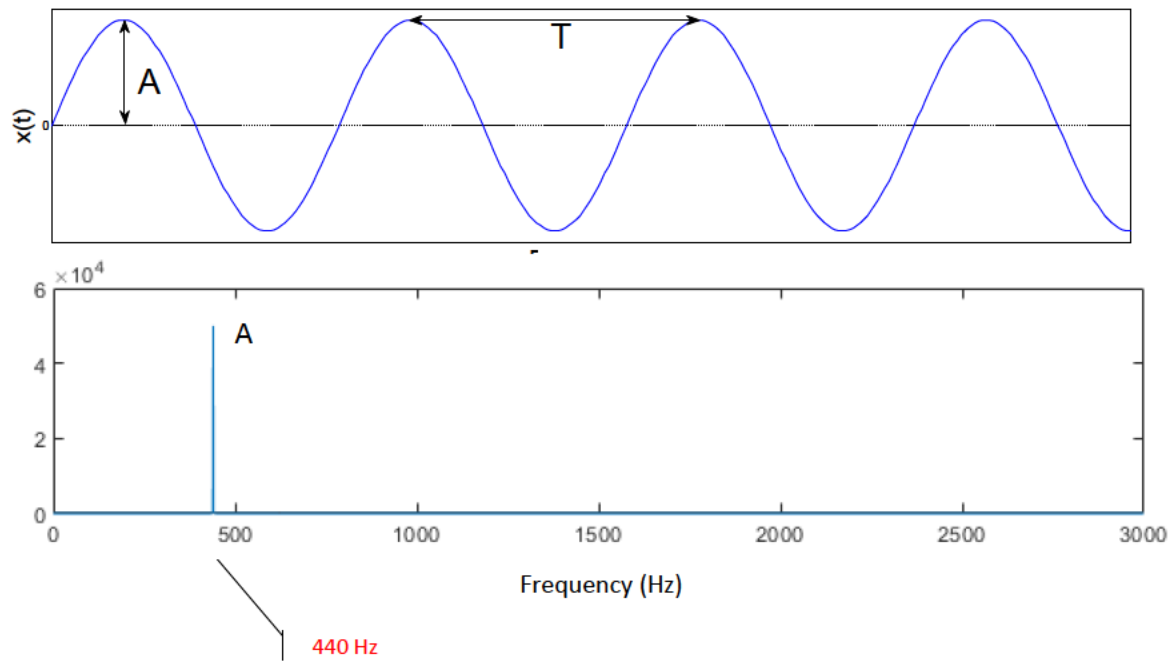


Time and frequency

- **Spectrum:** a representation of the signal using the parameters that define it (Amplitude + Frequency)

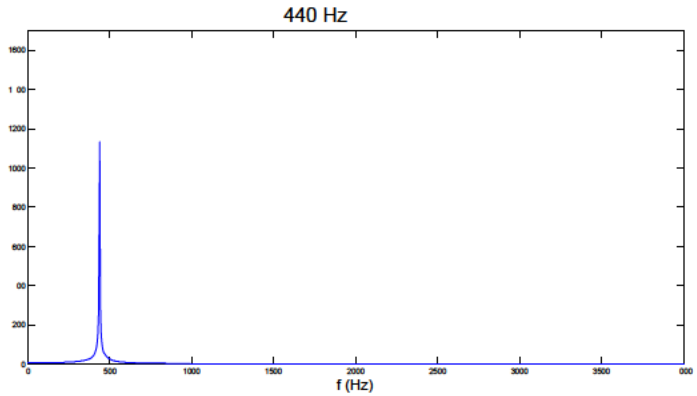
$$s(t) = A\sin(2\pi 440t + 0)$$

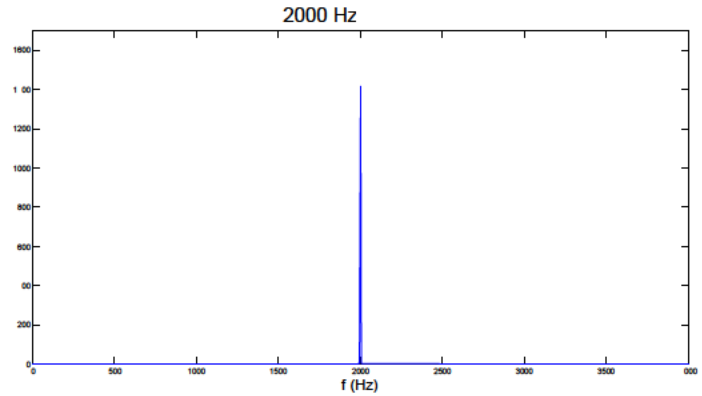
s(t)

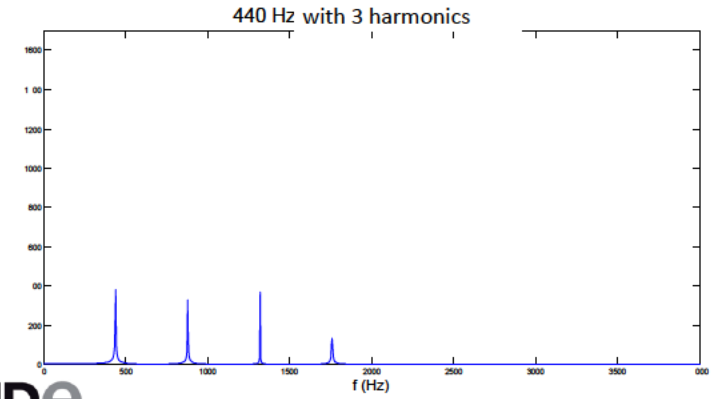


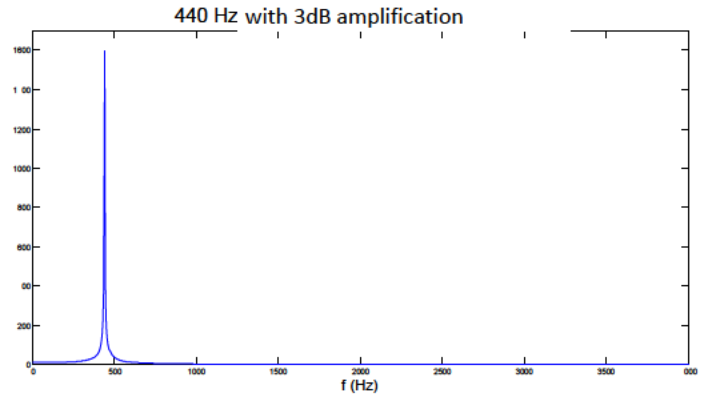
Time and frequency

- Frequency representation of signals: *spectrum*









Sound Pressure Level

- The intensity of the sound is usually measured in logarithmic units dB SPL (Sound Pressure Level) (dB=decibels)

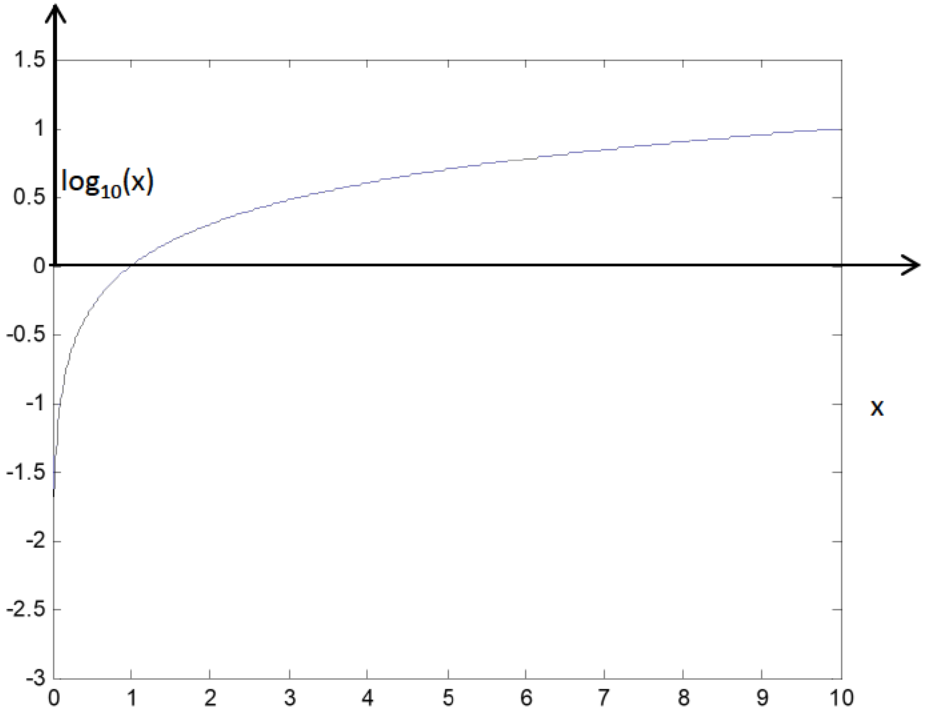
$$PdB(SPL) = 20\log_{10}\left(\frac{P_1}{P_{ref}}\right)$$

  - P_1: the pressure at a certain point (in Pa)
  - P_{ref}: reference level (20µPa at the air)
- It is a relative value, and measures the pressure of the sound at a certain point using a reference value:
  - 0dB SPL – silence threshold
  - 120dB SPL - pain threshold (Risk of instantaneous noise-induced hearing loss)

Sound Pressure Level

P_1 times bigger than P_{ref}	PdB (SPL)
1	0
2	6
10	20
100	40
1000	60
0,1	-20
0,5	-6
0,01	-40

$$P(dB) = 20\log_{10}\left(\frac{P_1}{P_{ref}}\right)$$



Sound Pressure Level

Sound sources	Sound pressure Level L_p dB SPL		Sound pressure(Pa)
Jet aircraft, 50 m away	140	$x10^7$	200
Threshold of pain	130		63.2
Threshold of discomfort	120	$x10^6$	20
Chainsaw, 1 m distance	110		6.3
Disco, 1 m from speaker	100	$x10^5$	2
Diesel truck, 10 m away	90		0.63
Curbside of busy road, 5 m	80	$x10^4$	0.2
Vacuum cleaner, distance 1 m	70		0.063
Conversational speech, 1 m	60	$x10^3$	0.02
Average home	50		0.006 3
Quiet library	40	$x10^2$	0.002
Quiet bedroom at night	30		0.000 63
Background in TV studio	20	$X10^1$	0.000 2
Rustling leaves in the distance	10		0.000 063
Hearing threshold	0	$X10^0$	0.000 02

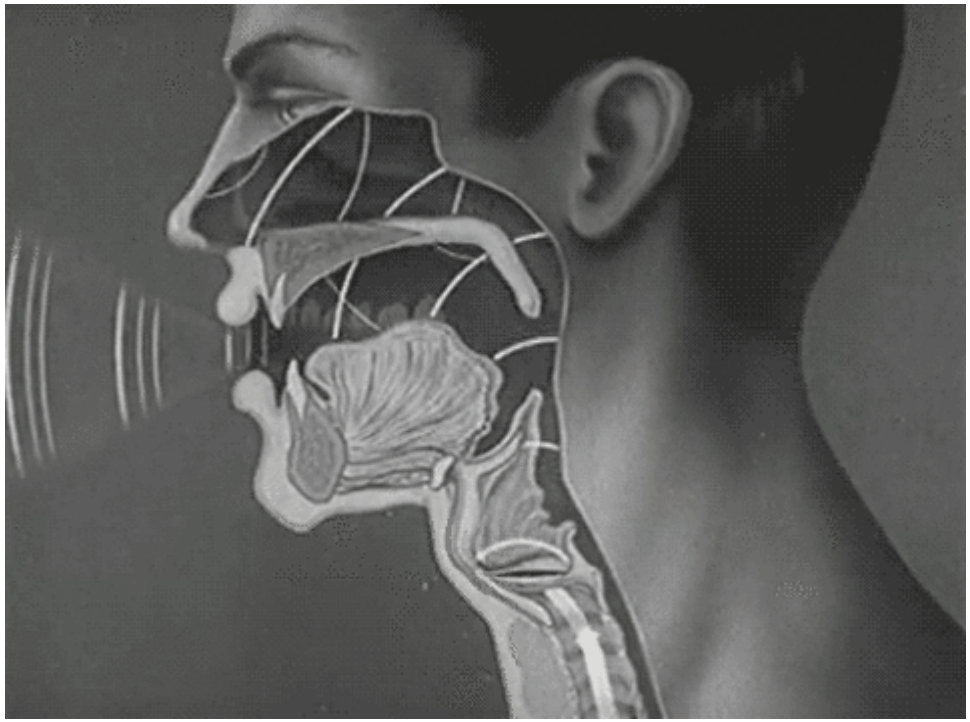
Speech production and perception

- Signals and sounds
  - Basic concepts
  - Time and Frequency
  - Sound Pressure Level
- Speech production
  - The human phonation system
  - Voiced and unvoiced sounds
  - Formants and coarticulation
  - Nasal sounds
- Acoustic phonetics
  - The alphabets
  - Articulatory mode and place
- Speech perception
  - The human auditory system
  - Pitch perception
  - Loudness and phons
  - Masking effect
  - Temporal and frequency discrimination
  - Other speech perception effects

The human phonation system

Speech

The sound wave produced when the air is expelled from the lungs and reaches the air through the vocal tract.

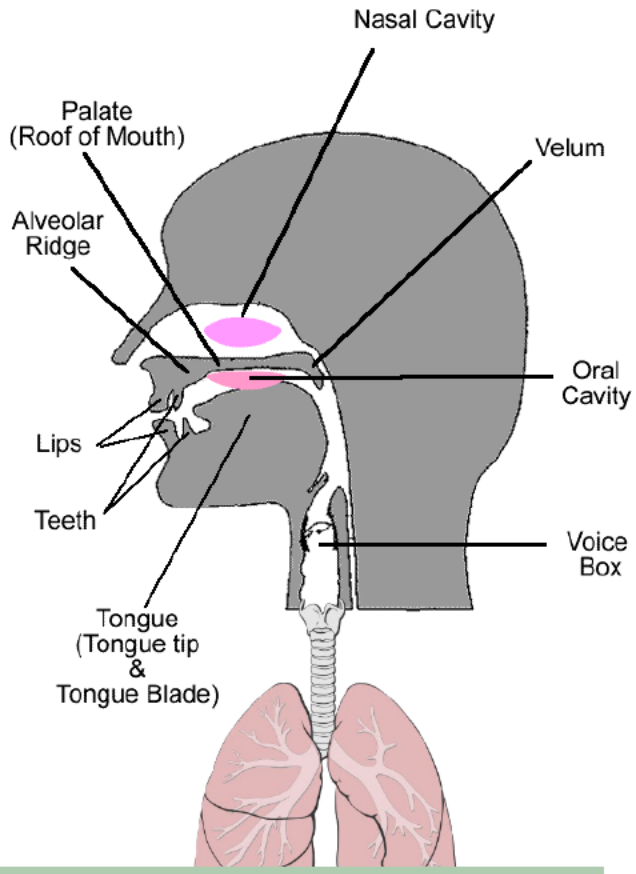


The human phonation system

- Subglottal system (breath organs):
  - Lungs, bronchial tubes, trachea
  - Energy source
- Vocal tract:
  - Glottis (vocal cords)
  - Pharynx
  - Oral cavity
  - Average length :17cm
  - Section: 0 cm-20 cm
  - Nasal cavity

Vocal Tract

Articulators and Places of Articulation

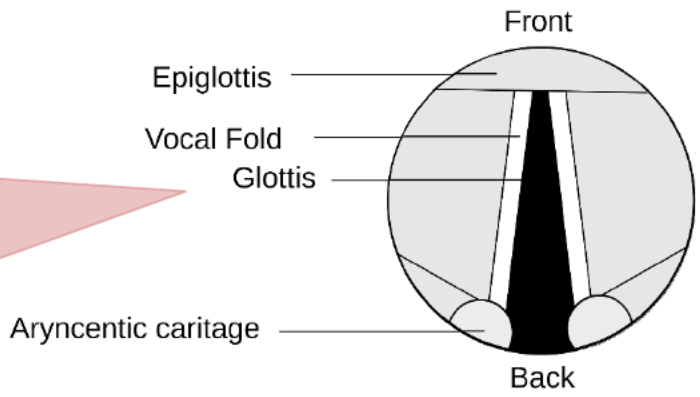


Lungs-simple diagram of lungs and trachea

The human phonation system

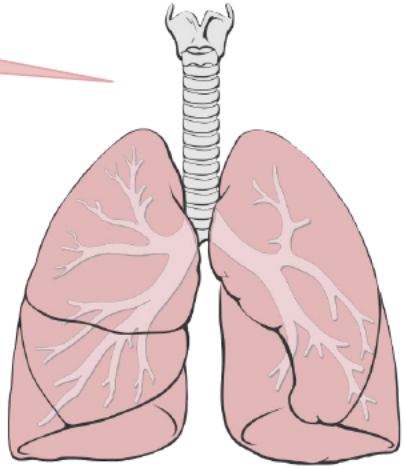
- **Voiced sounds:** The glottis is closed, the vocal cords are tense. The air goes with strength through them and causes the vibration. The airflow will be periodic.
- **Unvoiced sounds:** The glottis is open, the vocal cords are relaxed. The air goes through the glottis without obstacles, and the airflow will be turbulent.

Schematic diagram of the glottis



The airflow picks up speed at the trachea

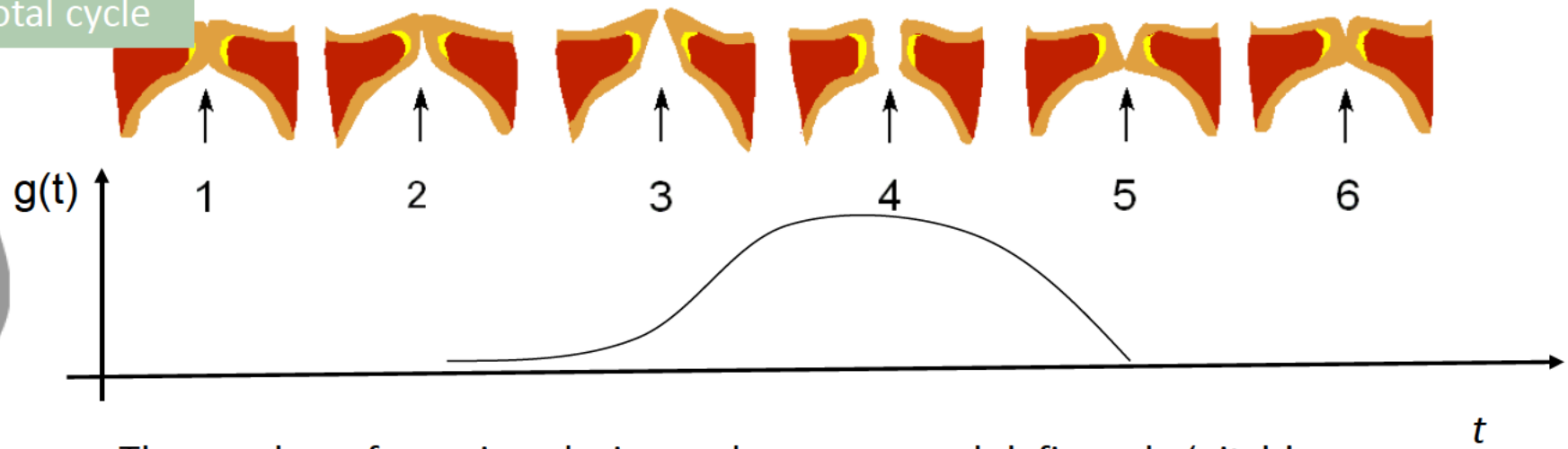
The lungs push the air with strength outwards



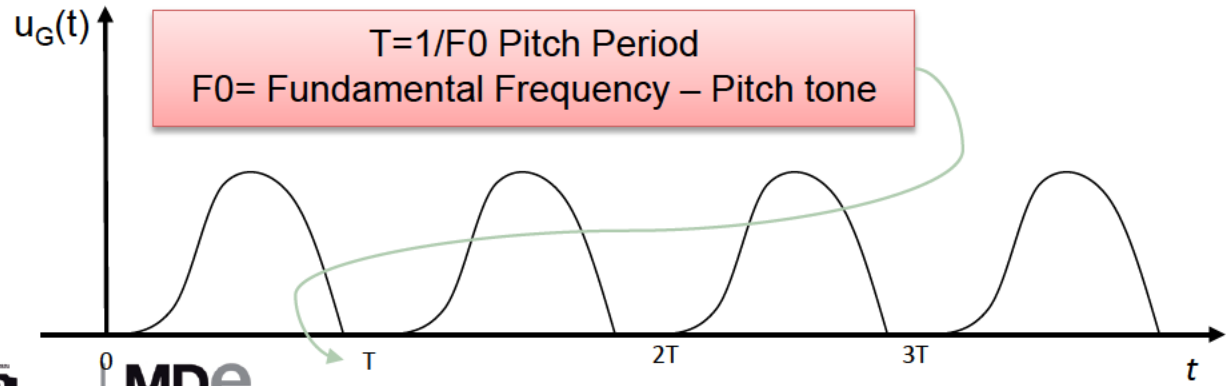
Lungs-simple diagram of lungs and trachea

The human phonation system: the glottal pulse & and the glottal cycle

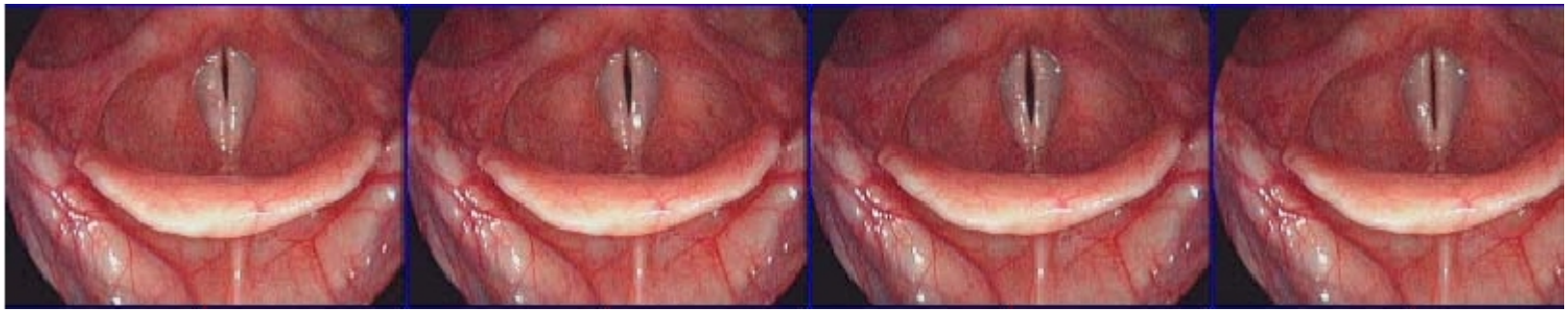
Glotal cycle

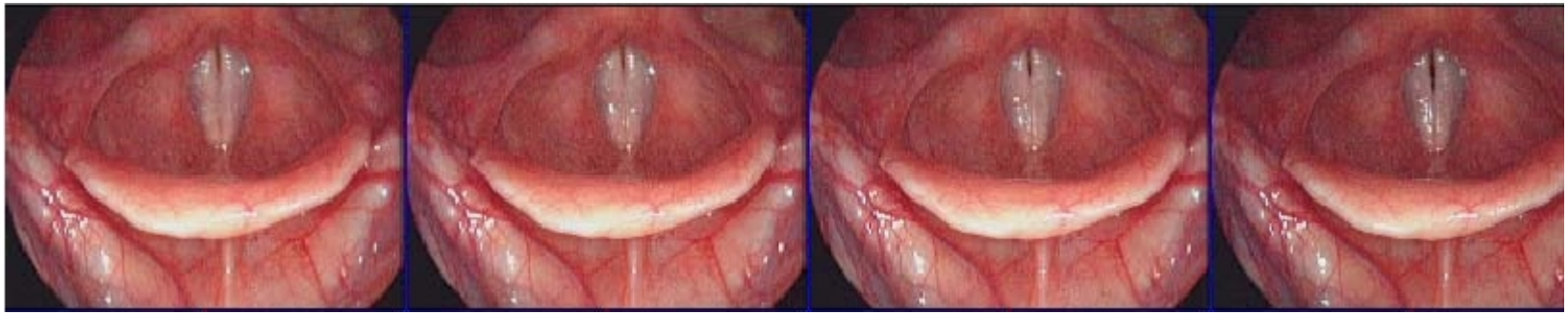


The number of opening-closing cycles per second defines de 'pitch' or fundamental frequency of the voice (F0)



The human phonation system

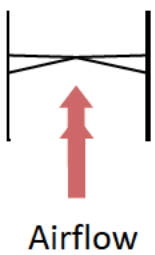


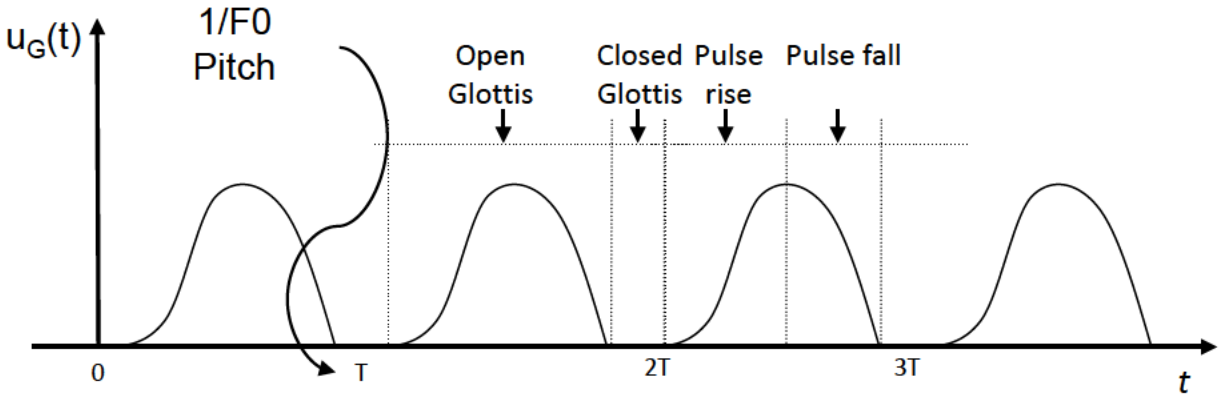


The vocal cords in motion

Voiced and unvoiced sounds: the glottal pulse & and the glottal cycle

- Glottal models:





- Rosenberg Glottal model defines the glottal airflow:
  - *Opening quotient*: pulse duration/pitch period
  - *Speed quotient*: pulse rise/pulse fall

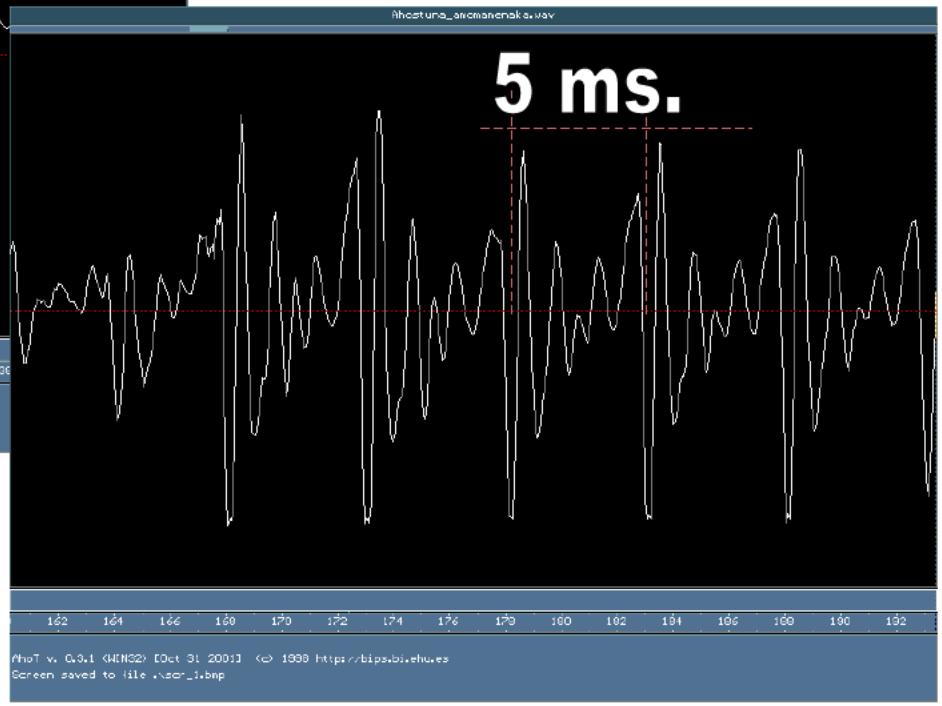
Voiced and unvoiced sounds

- **Pitch, fundamental frequency or F0**:
  - The frequency of the vibration of the vocal cords during the production of voiced sounds
    - Men: larger glottis, longer vocal cords, longer pitch-period, lower pitched voiced
      - Average/typical F0 ~ 110-130Hz
    - Women and children: smaller glottis, shorter vocal cords, higher pitched voices
      - Average/typical F0 ~ 200-230Hz

Voiced and unvoiced sounds: measuring F0

Men

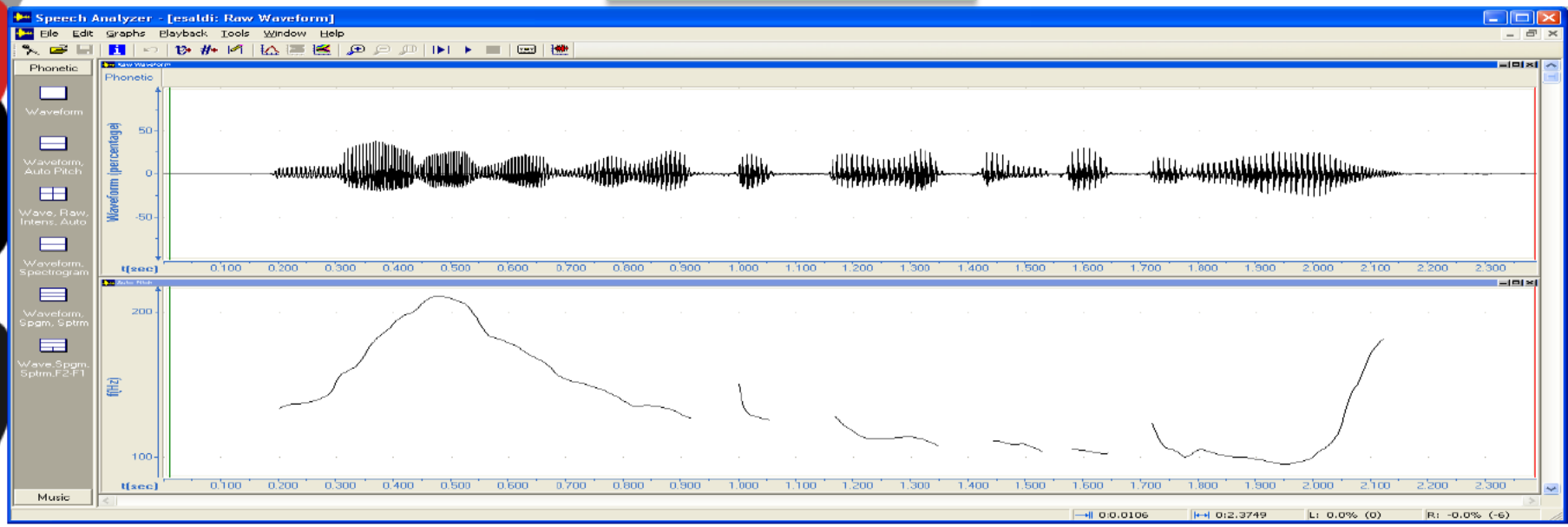
Women



Voiced and unvoiced sounds

- **Pitch, fundamental frequency or F0**:
  - We can represent it as a function of time:

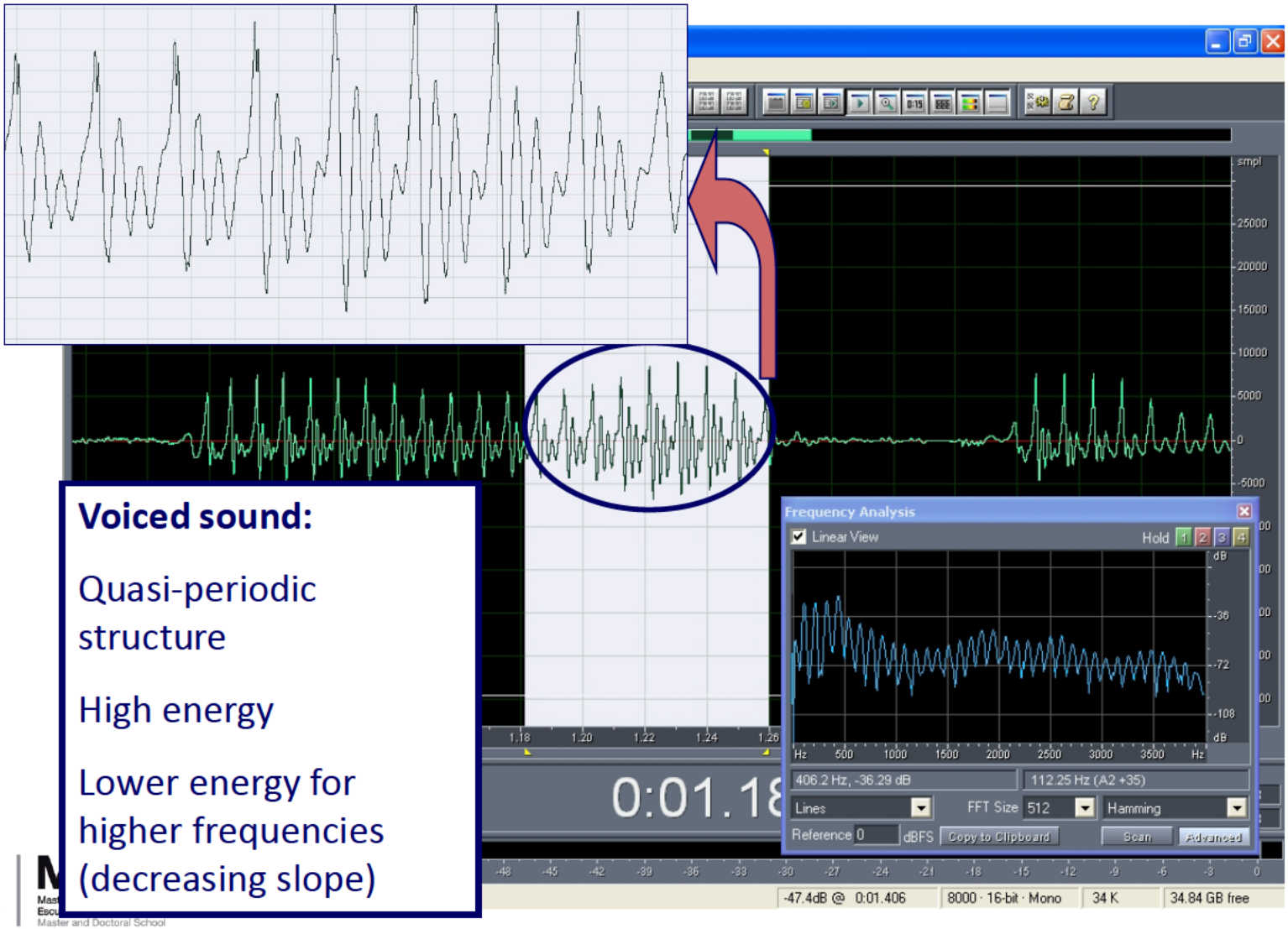
Intonation



Voiced and unvoiced sounds

- **Voiced sounds**:
  - The opening and closing of the vocal cords interrupt periodically the airflow coming from the lungs
  - **The airflow will be *periodic*, with a *Fundamental Frequency of F0*** corresponding to the opening-closing period
- **Unvoiced sounds:**
  - The vocal cords do not vibrate
  - The airflow will find a partial or total obstacle in the vocal tract and will pick up speed with turbulences
- **Vocal and nasal tract:**
  - They behave as resonance cavities
  - The resonance frequencies are called '*Formants*'
  - The values of the formants vary with the dimensions of the vocal and nasal tract, and with the shape of the vocal tract, which can be varied moving the articulators (the tongue, the jaw)

Voiced and unvoiced sounds



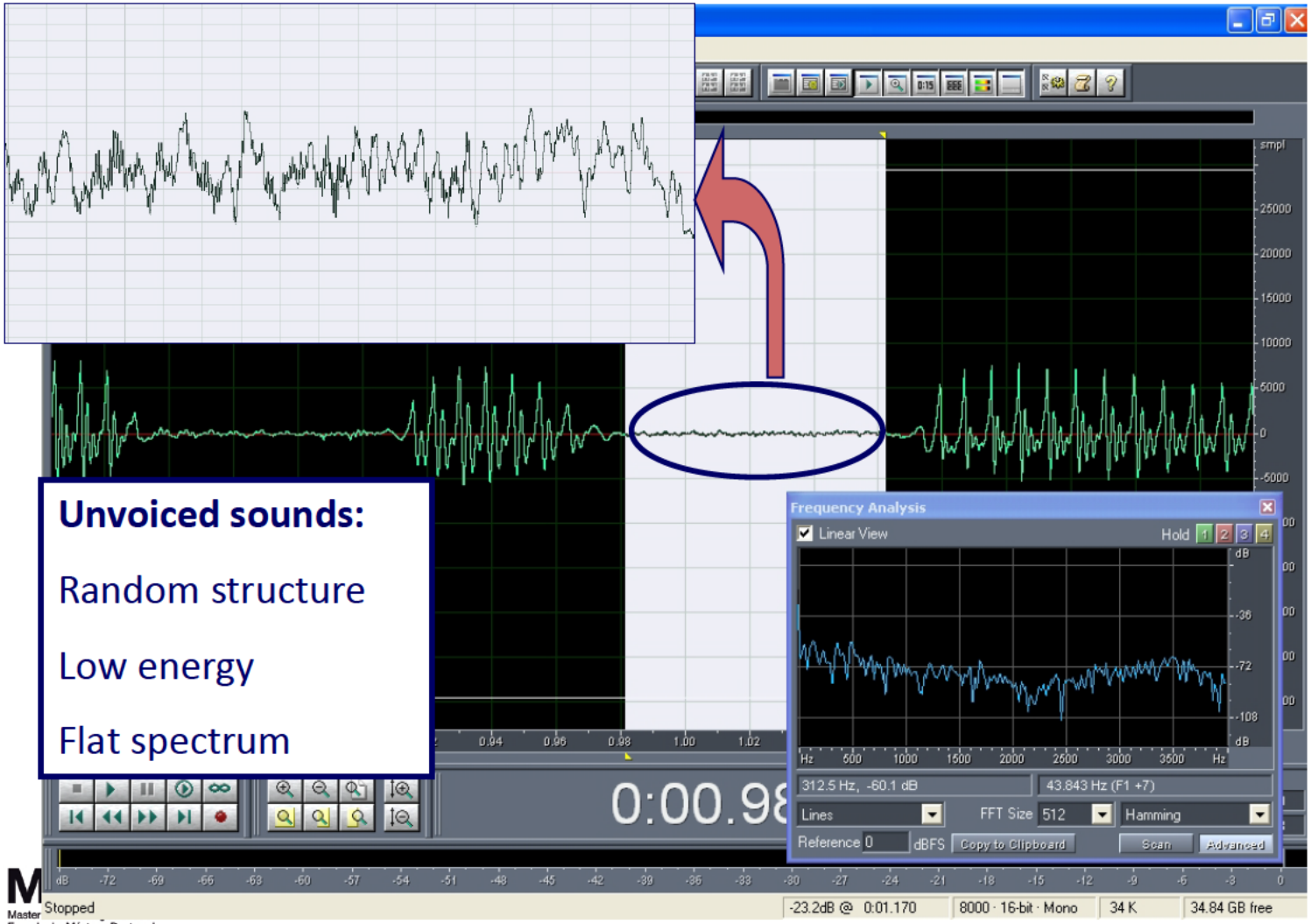
Voiced sound:

Quasi-periodic structure

High energy

Lower energy for higher frequencies (decreasing slope)

Voiced and unvoiced sounds



Unvoiced sounds:

Random structure

Low energy

Flat spectrum

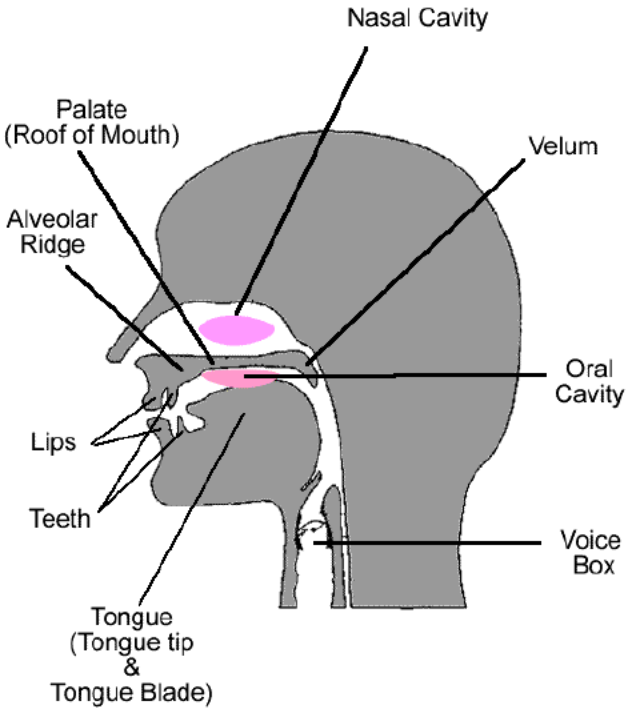
Speech production: formants and coarticulation

- ***Vowels***: The airflow goes from the glottis through the vocal cavity without obstacles. All vowels are voiced.

- ***Consonants***: Some articulator (the tongue, the lips, the teeth) will make an obstacle to the airflow, partially (fricatives, affricates) or totally (plosives).

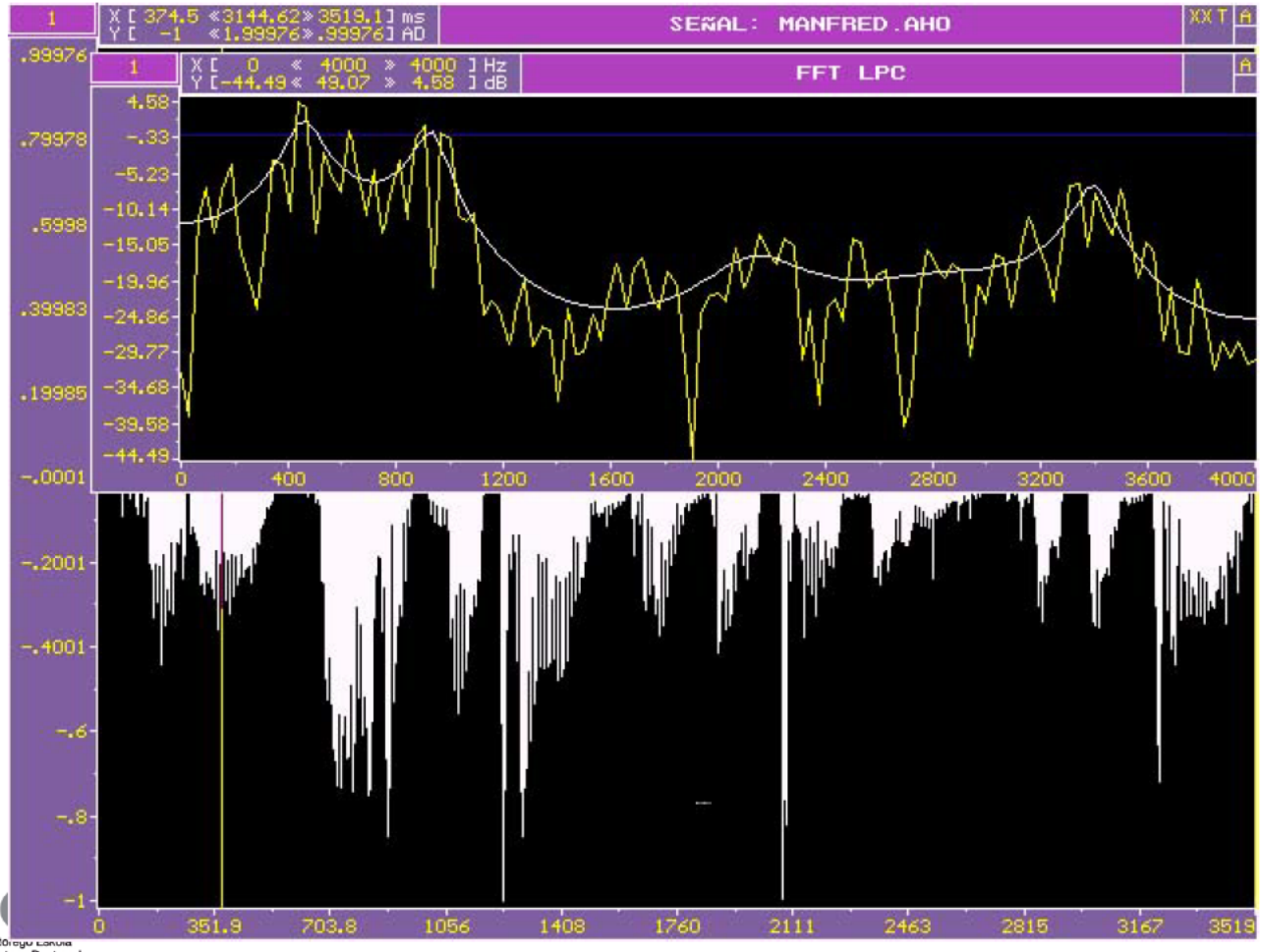
Vocal Tract

Articulators and Places of Articulation



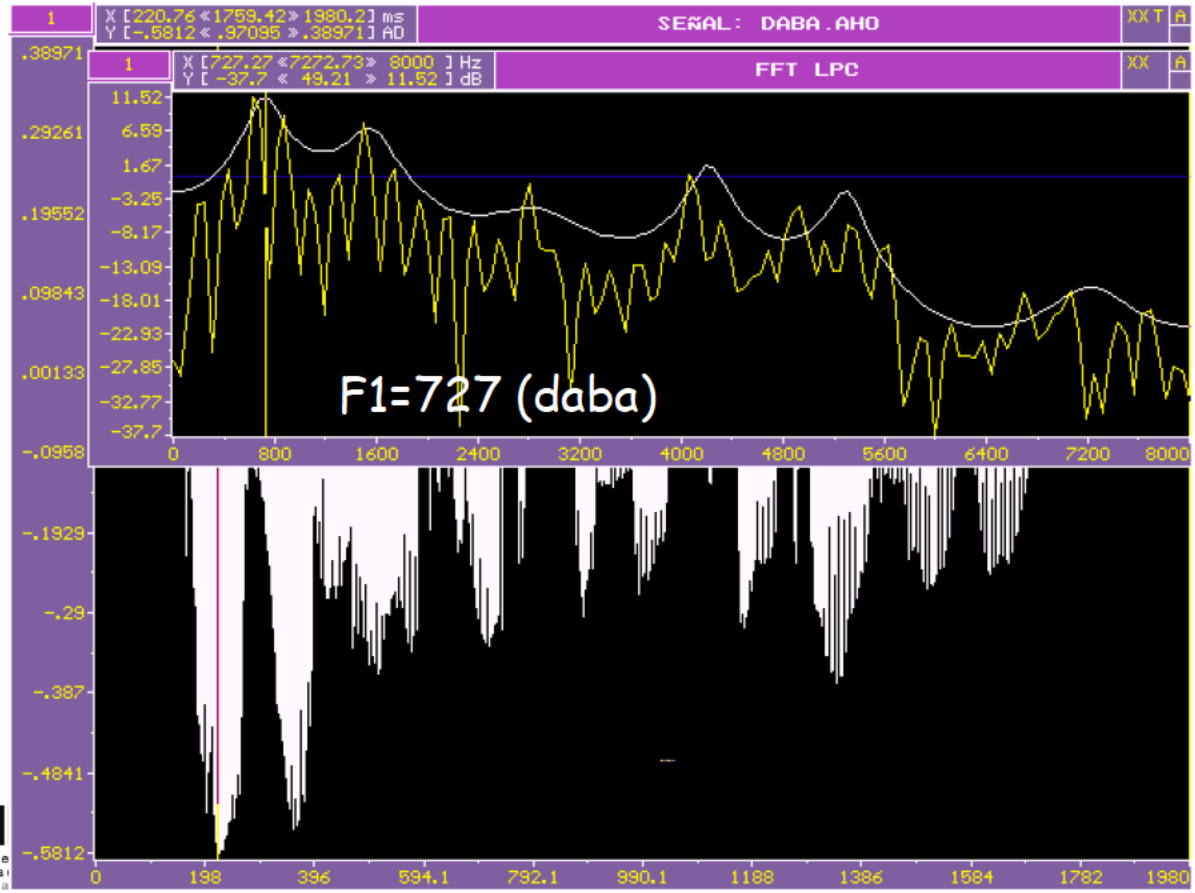
Formants and coarticulation

- Oral resonances: **Formants**



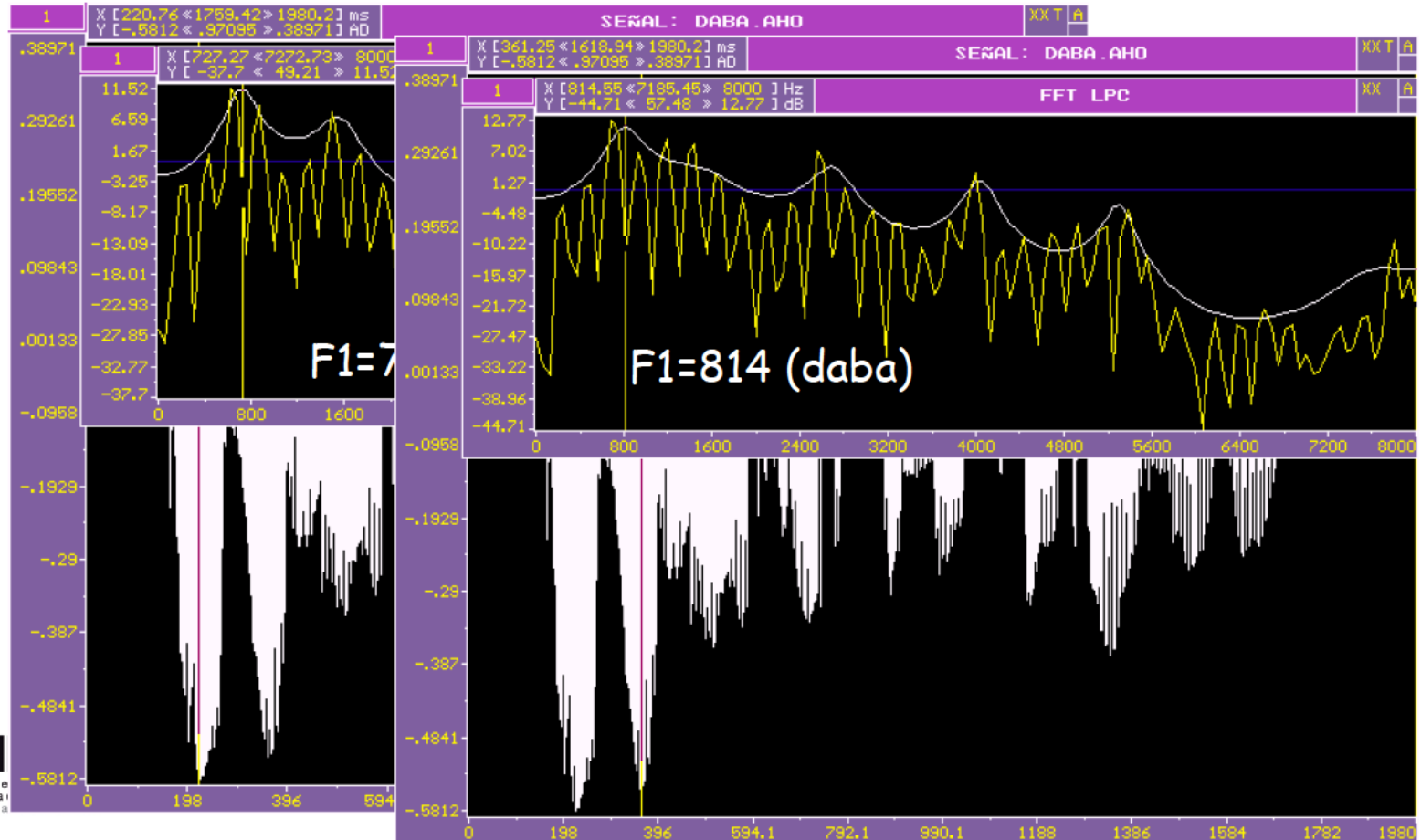
Formants and coarticulation

- Coarticulation: the influence of the surrounding sounds on the actual sound.



Formants and coarticulation

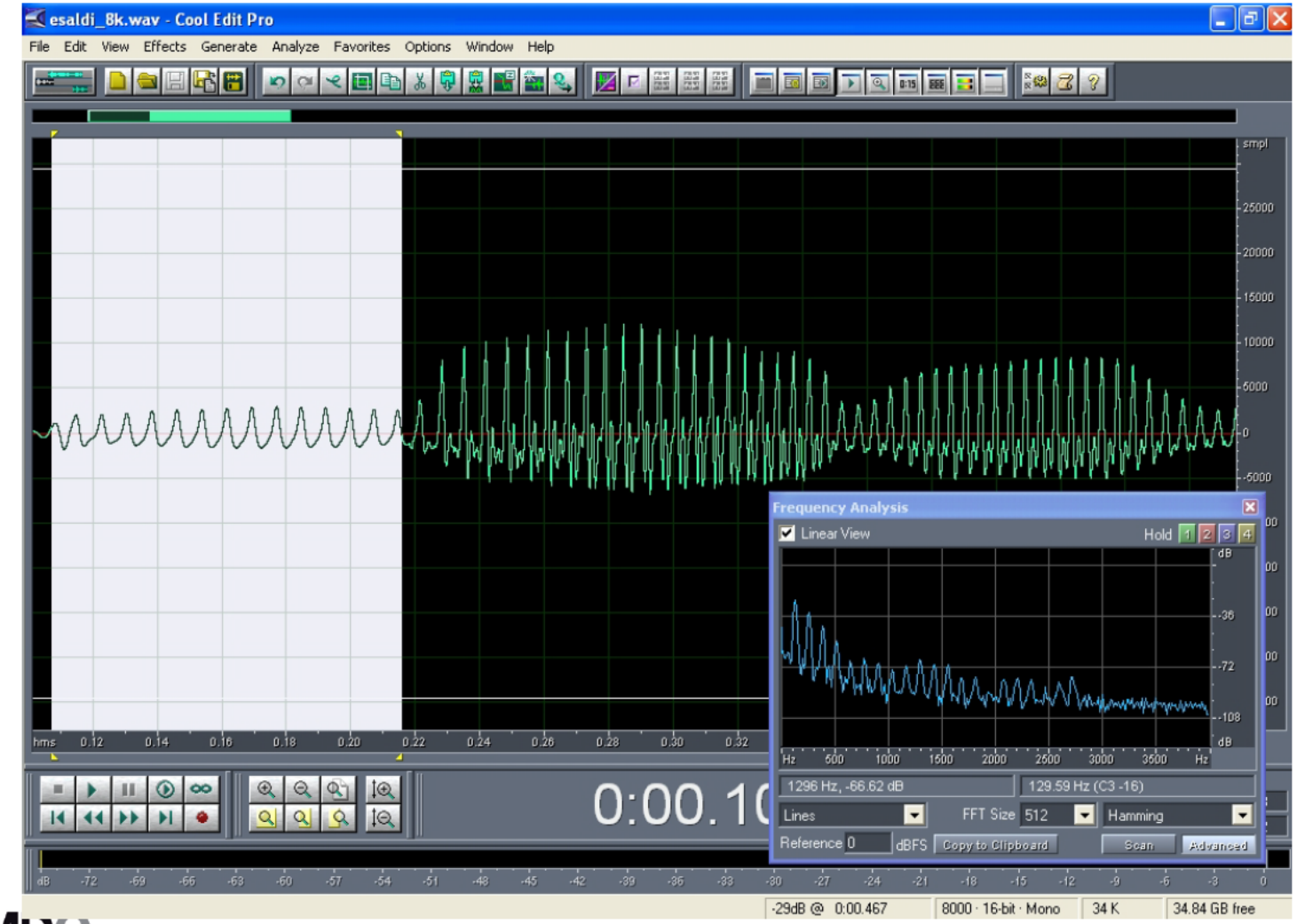
- Coarticulation: the influence of the surrounding sounds on the actual sound.



Nasal sounds

- During the production of the nasal sounds the soft palate goes down and the airflow goes through the nasal cavity
  - Oral and nasal cavities are acoustically coupled
  - The exit from the oral cavity is closed and anti-resonances will appear (the energy at some frequencies gets trapped inside the oral cavity and will not find its way to the exit).
  - Anti-resonance frequency values will define the kind of sound. For example:
    - 750 - 1250 Hz /m/
    - 1450 - 2200 Hz /n/

Nasal sounds



Speech production and perception

- Signals and sounds
  - Basic concepts
  - Time and Frequency
  - Sound Pressure Level
- Speech production
  - The human phonation system
  - Voiced and unvoiced sounds
  - Formants and coarticulation
  - Nasal sounds
- Acoustic phonetics
  - The alphabets
  - Articulatory mode and place
- Speech perception
  - The human auditory system
  - Pitch perception
  - Loudness and phons
  - Masking effect
  - Temporal and frequency discrimination
  - Other speech perception effects

Acoustic phonetics

- It is an area of study in the Phonetic Science
- It focuses the research on the physical acoustic characteristics of the sounds: duration, amplitude, fundamental frequency…
- It also studies the relationships between the physical magnitudes and the linguistic concepts (phonemes, words, sentences).

Acoustic phonetics: the alphabets

- They represent the sounds:
  - *IPA International Phonetic Alphabet*
    - From 1886, last updated in 2005
    - *International Phonetic Association*
  - *SAMPA Speech Assessment Methods Phonetic Alphabet*
    - 1987-1989
    - ASCII characters for basic sounds
  - X-SAMPA Extended -*SAMPA* (with diacritics etc.)

Acoustic phonetics: sounds descriptions

- Sounds are described according to different dimensions:
  - There is or there is not obstruction in the vocal tract: vowels & diphtongs vs consonants
  - Voicing: voiced vs unvoiced
  - Soft palate is used: oral vs nasal
  - Place of articulation: position of the articulator to generate the obstruction (bi-labial, labio-dental…)
  - Manner of articulation: the way the articulators are placed (Stops or plosives, fricatives, affricates…)

Interactive Sagital Section

Acoustic phonetics: manner of articulation

(Basque)

		Closing		
		No	**Yes - Total**	**Yes - Partial**
Voicing	Yes	Vowels /a/ /e/ /i/ /o/ /u/	Nasals /m/ /n/ /J/ Plosives /b/ /d/ /g/	Fricatives /jj/ Liquids /l/ /L/ /r/ /rr/ Laterals (approximants) /B/ /D/ /G/
			Affricates /gj/	
	No		Plosives /p/ /t/ /k/ /c/	Fricatives /f/ /T/ /s/ /s'/ /S/ /x/
			Affricates /tS/ /ts/ /ts'/	

Acoustic phonetics: vowels & articulatory place

- Vowels: Position of the tongue and lips opening

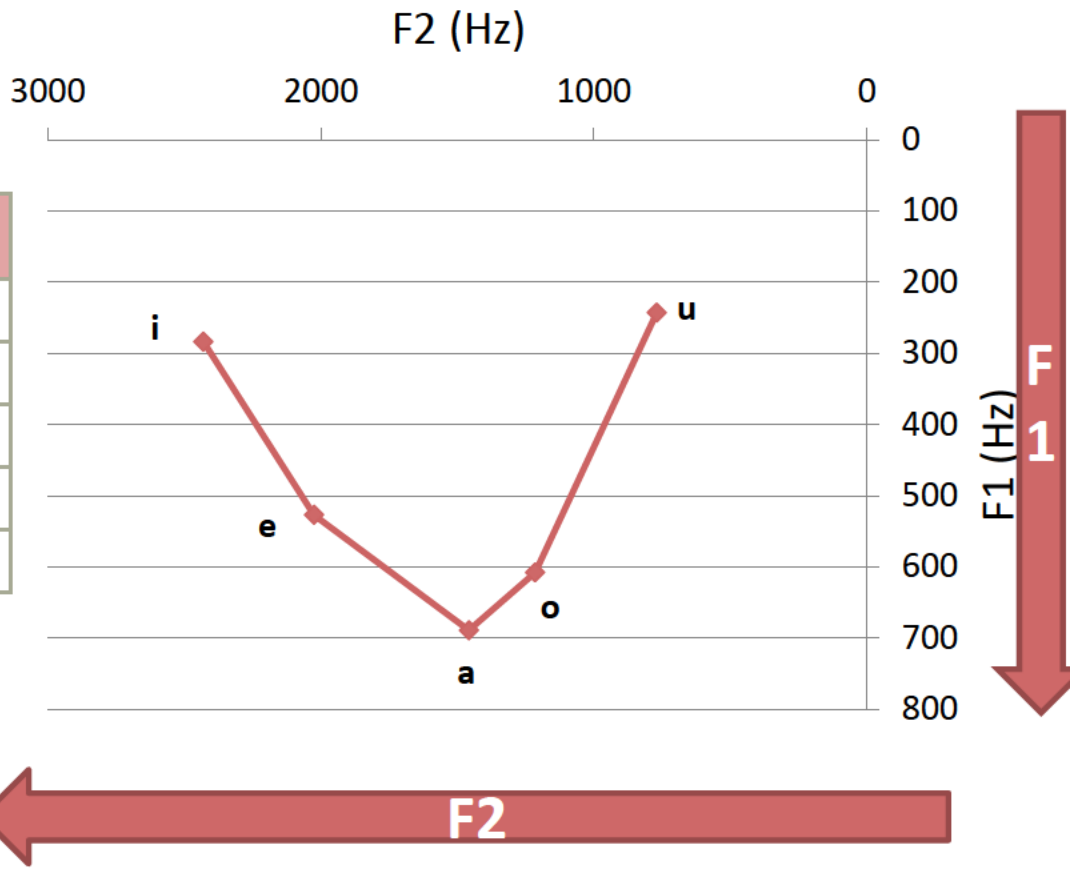
		Position		
		Forward	Middle	backward
Opening	*minimum*	i		u
	medium	e		o
	maximum		a	
		Non-rounded		rounded

F1

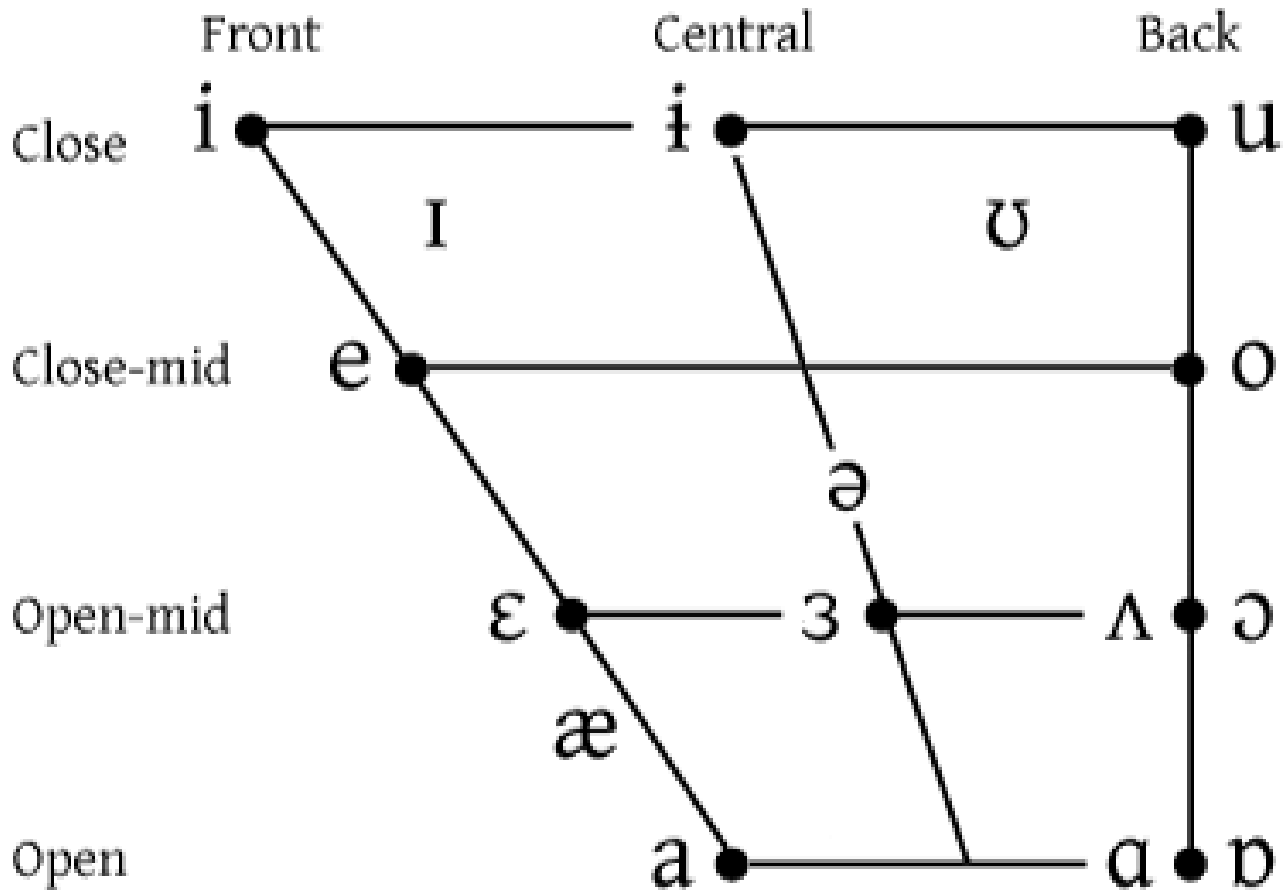
F2

Acoustic phonetics: vowels & articulatory place

Formant (Hz)	F1	F2
/i/	284	2430
/e/	527	2025
/a/	689	1458
/o/	608	1215
/u/	243	770



Acoustic phonetics: vowels & articulatory place



http://www.ipachart.com/

English vowels

Acoustic phonetics: articulatory place

- Consonants: location of the maximum constriction in the vocal tract during production
- Classification of consonant sounds:
  - Bilabials /m/ /p/ /b/
  - Labiodentals /f/
  - Dentals /t/ /d/
  - Interdental /T/
  - Apicals /s/ /n/ /l/ /r/ /rr/
  - Palatals /tS/ /J/ /jj/ /L/
  - Velars /k/ /g/ /x/

http://soundsofspeech.uiowa.edu/resources/spanish/spanish.html
http://soundsofspeech.uiowa.edu/resources/english/english.html

Speech production and perception

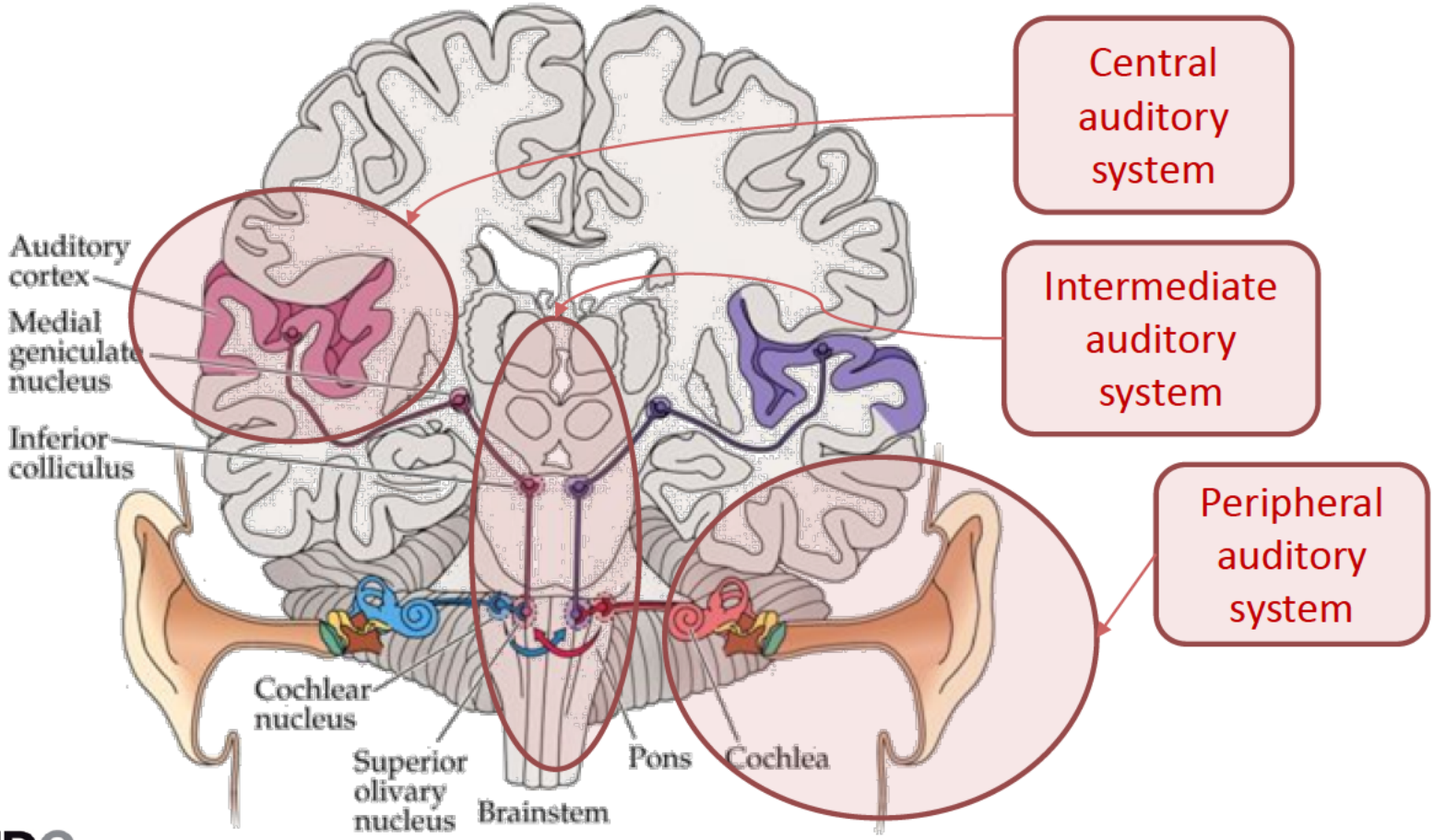
- Signals and sounds
  - Basic concepts
  - Time and Frequency
  - Sound Pressure Level
- Speech production
  - The human phonation system
  - Voiced and unvoiced sounds
  - Formants and coarticulation
  - Nasal sounds
- Acoustic phonetics
  - The alphabets
  - Articulatory mode and place
- Speech perception
  - The human auditory system
  - Pitch perception
  - Loudness and phons
  - Masking effect
  - Temporal and frequency discrimination
  - Other speech perception effects

The human auditory system

- Speech perception is of greatest importance in oral communication, both in human-human and in human-machine communication.
- Components of the auditory system:
  - The **peripheral auditory system**: The ear performs the transduction of the sound as a pressure wave into an electrical signal feeding the nervous system
  - The **intermediate auditory system:** performs binaural processing, sound localization and processing of complex sounds before sending signals to the auditory cortex
  - The **central auditory system**: The encoded sound information travels to the thalamus and from there to the cortex. The sound information will be extracted at different stages in this auditory pathway.

https://blausen.com/en/video/hearing-loss/#

The human auditory system



© 2001 Sinauer Associates, Inc.

The peripheral auditory system

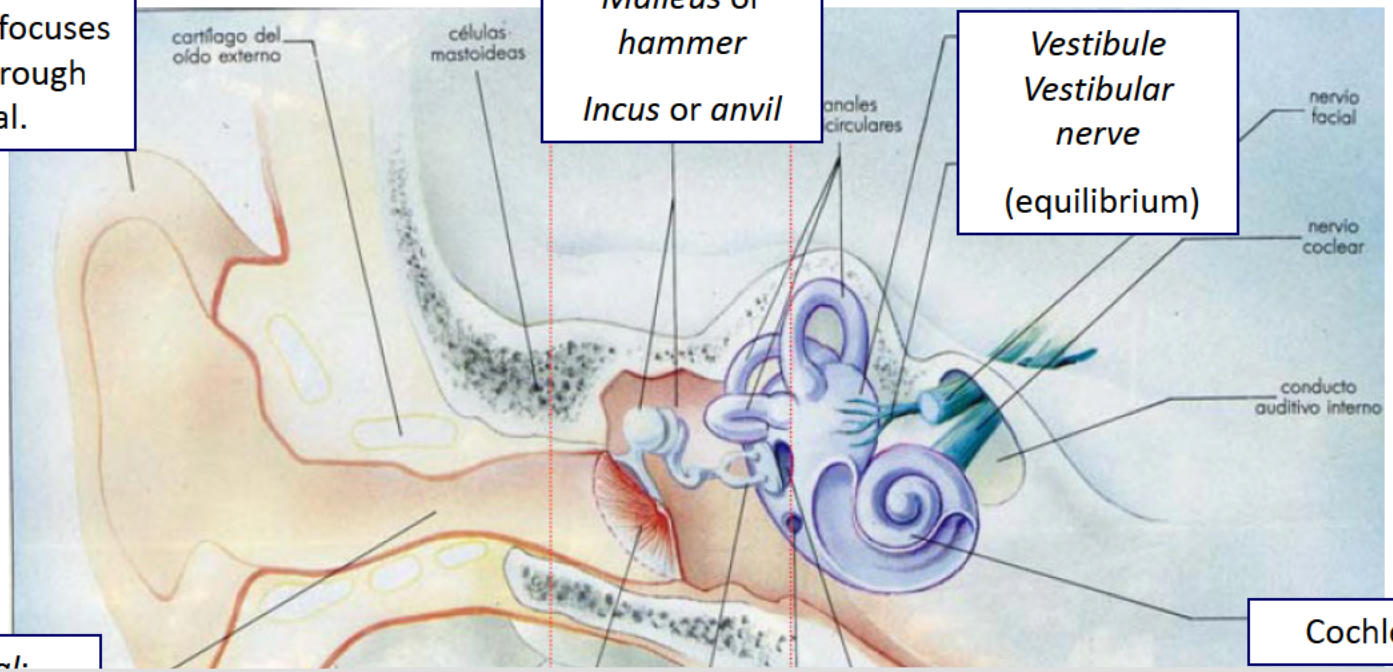
Auricle or *pinna*: focuses sound waves through the ear canal.

Malleus or *hammer*

Incus or *anvil*

Vestibule
Vestibular nerve

(equilibrium)



Cochlea

Ear canal: