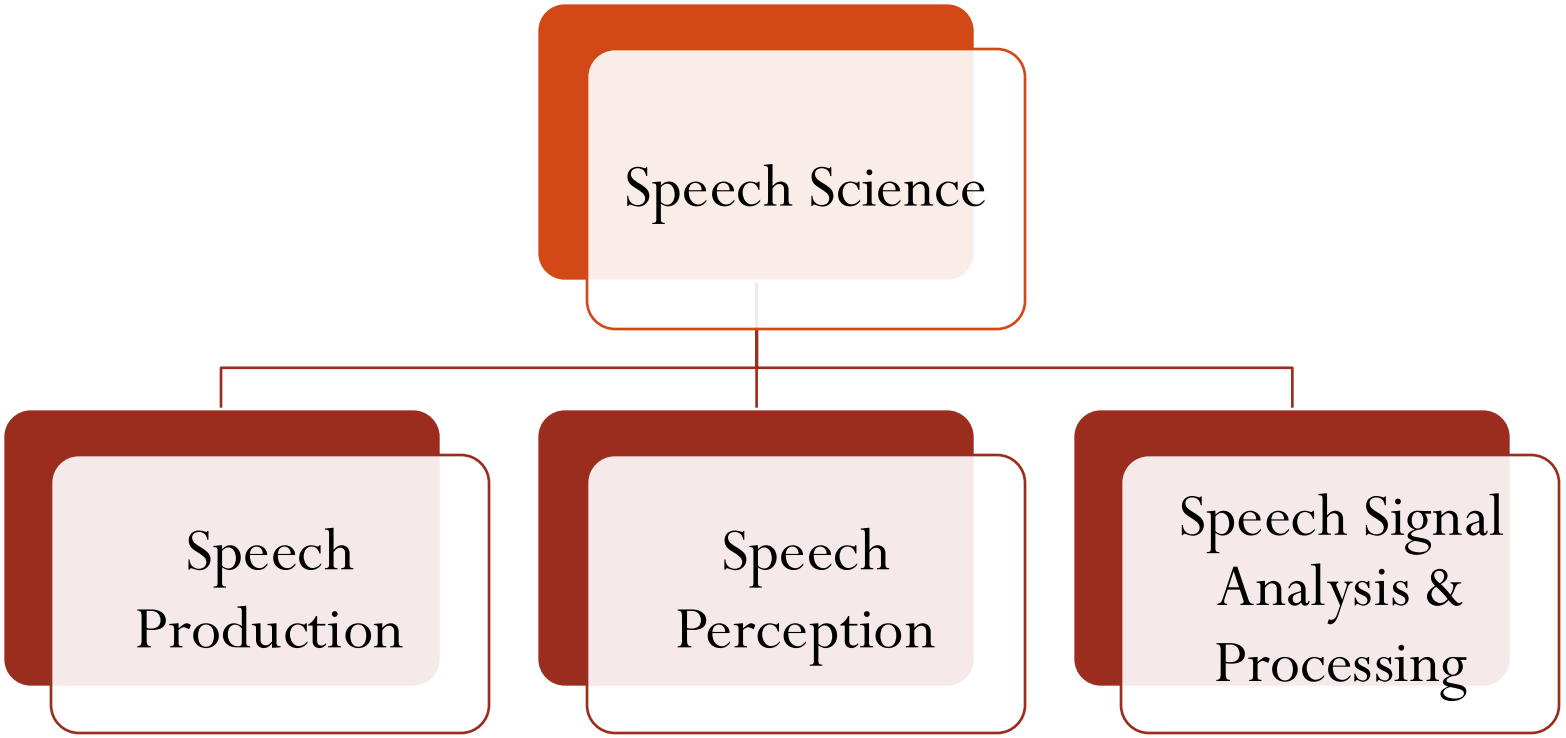
Speech Science: Articulation and Acoustics

CS578
Winter Term, 2025-26
CSD UOC

Invited Lecture
Dr Anna Sfakianaki
Assistant Professor of Phonetics/Phonology
University of Ioannina
asfakianaki@uoi.gr

Speech Science

- Speech Science is the experimental study of **speech communication**.

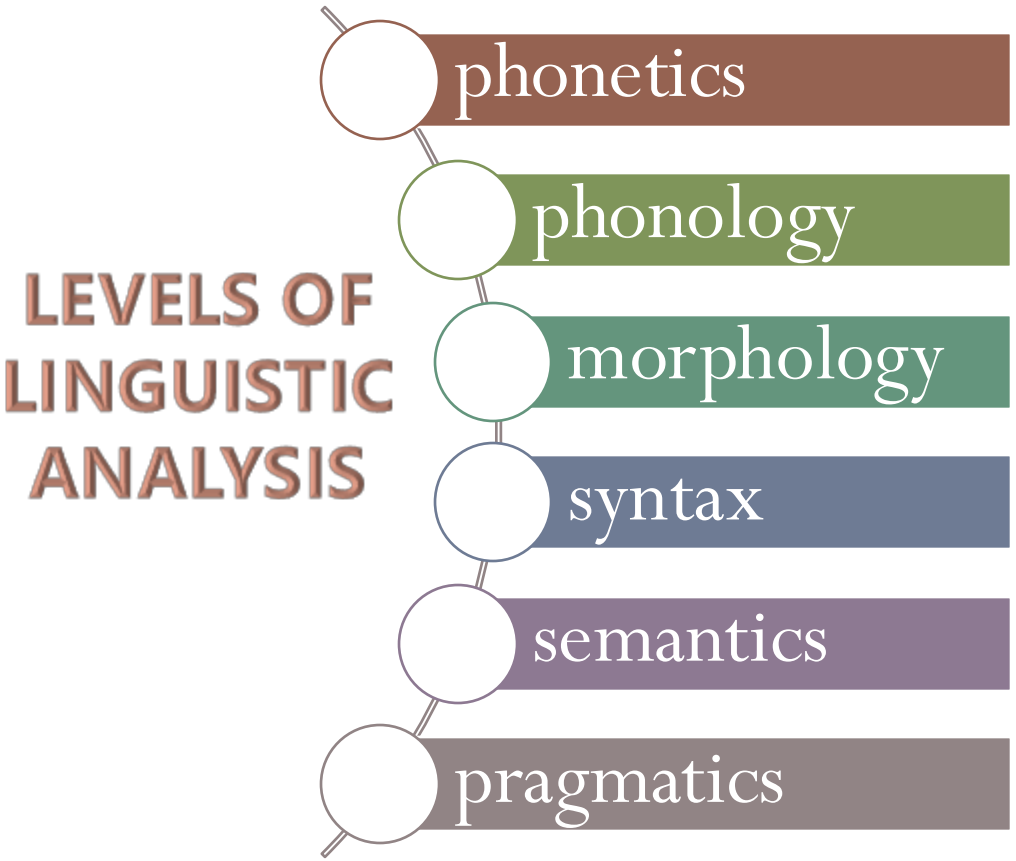


Speech Science and Phonetics

- Speech Science has its origins in **Phonetics**
  - **Phonetics** is the branch of linguistics that studies the sounds of speech.

Field work: **Peter Ladefoged**
Language: **Toda**
Kiawiarh Village, South India, 22/01/2006
https://linguistics.ucla.edu/people/ladefoge/Remember/Index.htm

Linguistics



production and use of **speech sounds** in communication; articulatory and acoustic properties of speech

relations of **speech sounds** within the linguistic system

word formation and word alteration for sentence construction

grammar rules for **sentence** construction

meaning of words, utterances, sentences

influence of **context** on utterance **meaning**

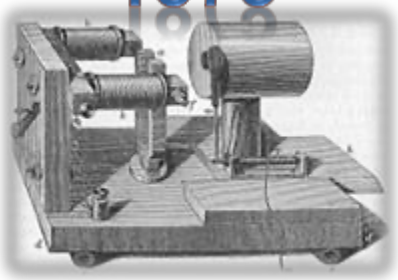
Speech Science and Phonetics

- Speech Science has its origins in **Phonetics**
  - **Phonetics** is the branch of linguistics that studies the sounds of speech.
  - The **sounds of speech** are the pieces of the linguistic code used to communicate meaning.

PHONETICS + TECHNOLOGY = SPEECH SCIENCE

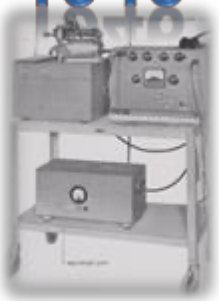
Adding Technology to Phonetics →
empirical investigation of speech production and perception

1875



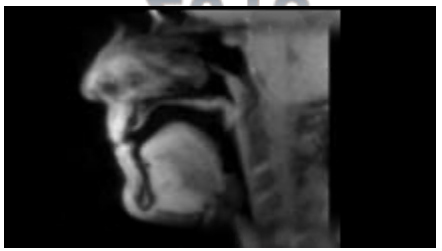
Helmholtz Resonator

1946



Sound Spectrograph
Bell Laboratories

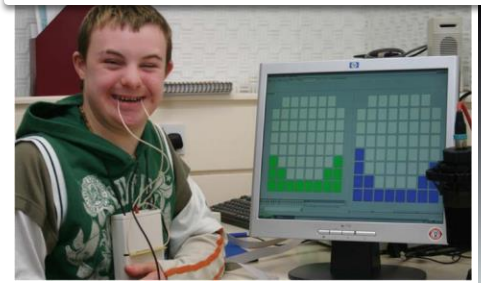
2018



real-time MRI
Max Planck Institute

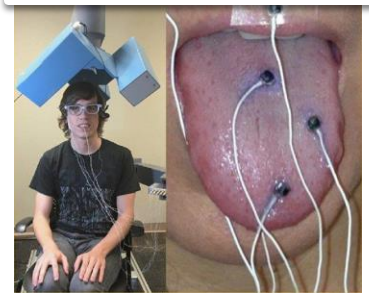
SPEECH SCIENCE
The instrumental study of speech

EPG-electropalatography

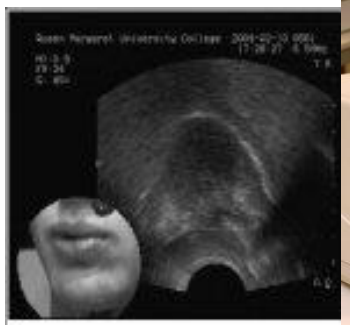


MRI

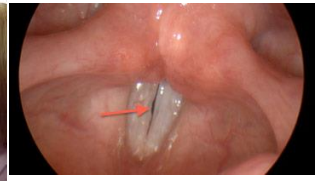
EMA-electromagnetic articulography



UTI -Ultrasound



Stroboscopy



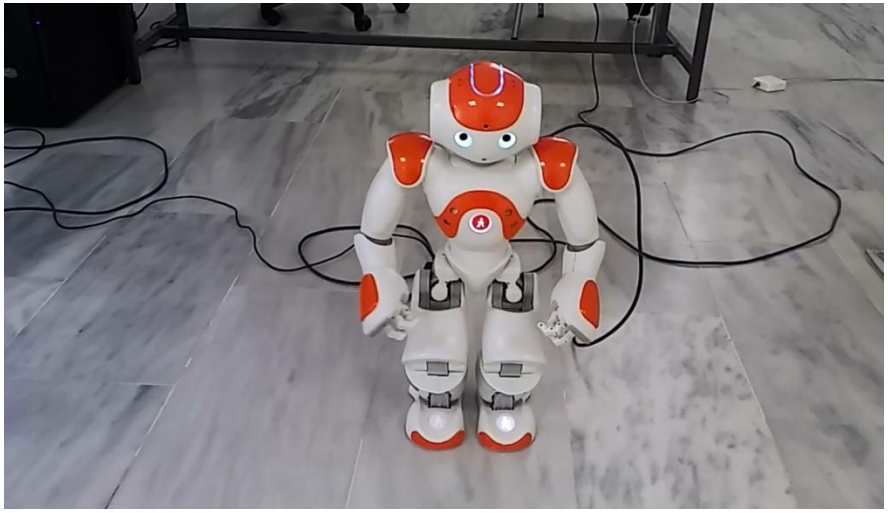
Questions posed by Speech Science

- How is speech planned and executed by the vocal system?
- How do the acoustic properties of sounds relate to their articulation?
- How and why do speech sounds vary from one context to another?
- How do listeners recover the linguistic code from auditory sensations?
- How do infants learn to produce and perceive speech?
- How and why do speech sounds vary between speakers?
- How and why do speech sounds vary across speaking styles or emotions?

Speech Science Applications

- Core Applications
  - Speech recognition
  - Speech synthesis
  - Speaker recognition

Nao Robot, FORTH

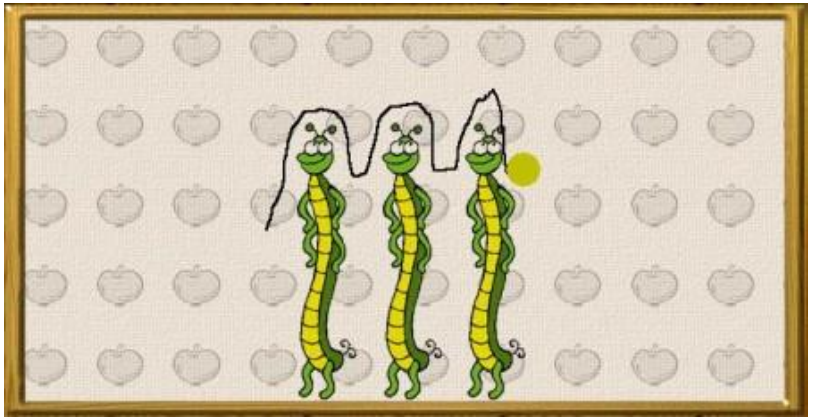


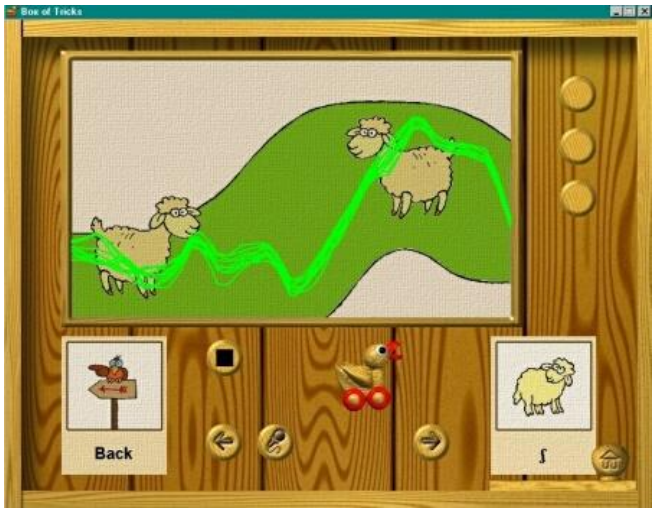
- Other applications
  - Forensic speaker comparison
  - Language pronunciation teaching
  - Assessment and therapy for disorders of speech and hearing
  - Monitoring of well-being and mood

SpeakGreek, AUTh

SPECO (SPEech COrrector)

- EU project (1998-2001)
- Visual display of acoustic information for children in need of assistance with various aspects of speech production
- Developed in 4 languages
- Multi-speaker database
- Commercial product (RCS)





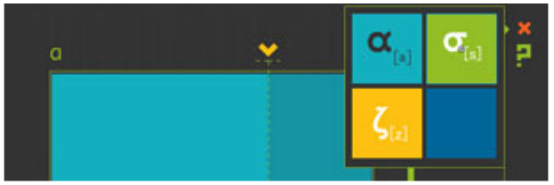
SpeakGreek

ΜΙΛΩ ΕΛΛΗΝΙΚΑ
SpeakGreek
http://speakgreek.web.auth.gr
ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

https://www.enl.auth.gr/speakgreek/index.html

- Free online pronunciation training tool for learners of Greek as a foreign/second language and for people with speech and hearing disorders
- Database of 60 speakers (men, women, children)

Voice Training



It contains applications which train users to produce sounds with appropriate voicing, to sustain sounds for as long as possible, to control the intensity and pitch of their voice.

Listen And Learn

It contains applications which train users to perceive and identify correctly the Greek vowels and consonants in syllables, words, word pairs, and sentences.

Say And Learn

It contains applications which train users to produce correctly the Greek vowels and consonants in isolation, in syllables, words, word pairs, and sentences. It also trains users to produce the appropriate melody of Greek in statements, questions, and sentences with different focus.

SpeakGreek –Phonetic Library

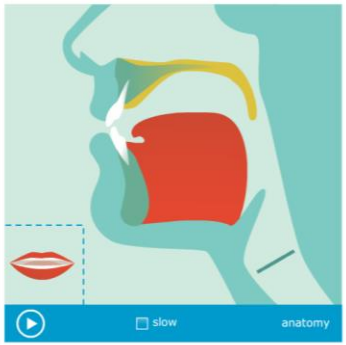
[i]

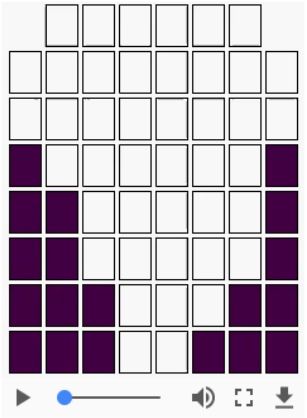
Male Female Child

+ EPG & Ultrasound

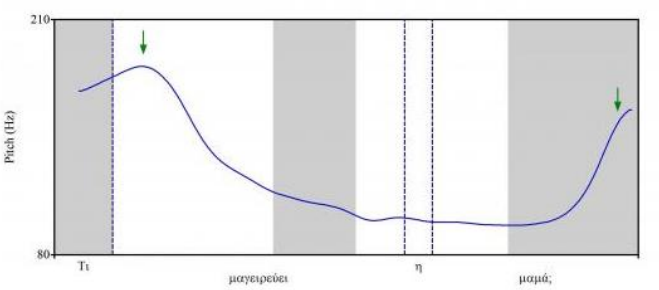
Examples:
- [i]
- ήχοι
- φίλη
- φίδι
- είδηση
- ησυχία

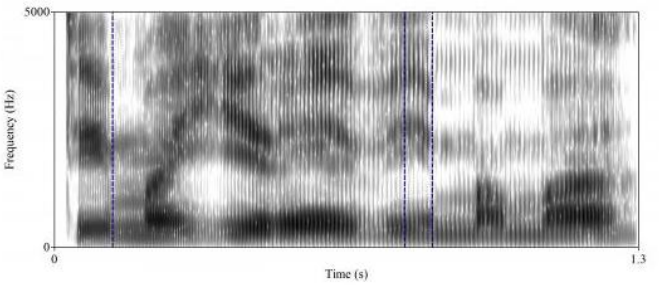
/i/: close front unrounded
The tongue front is raised towards the hard palate. The tongue is in advanced ar
The sides of the tongue
lips are spread. The soft
vibrate.

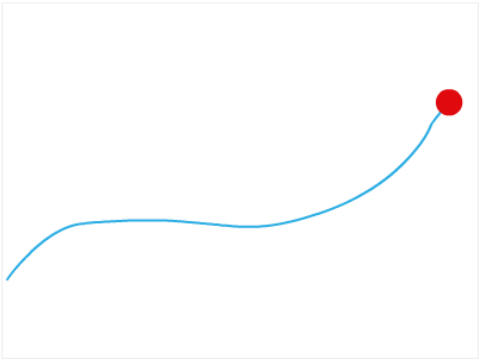




intonation



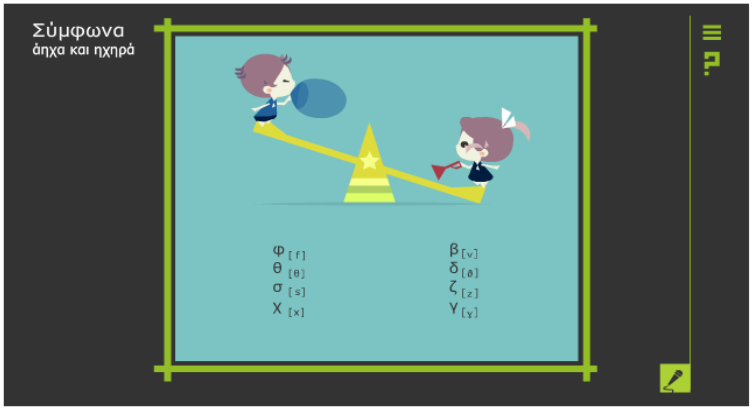




SpeakGreek –Voice Training

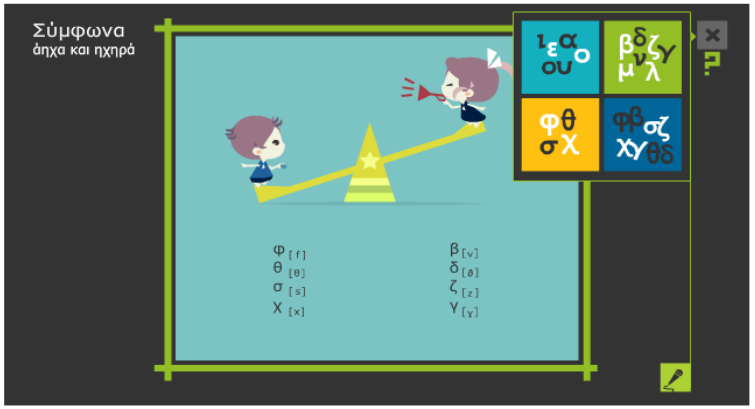
Μαθαίνω να παράγω φωνή

Πίσω στις εφαρμογές



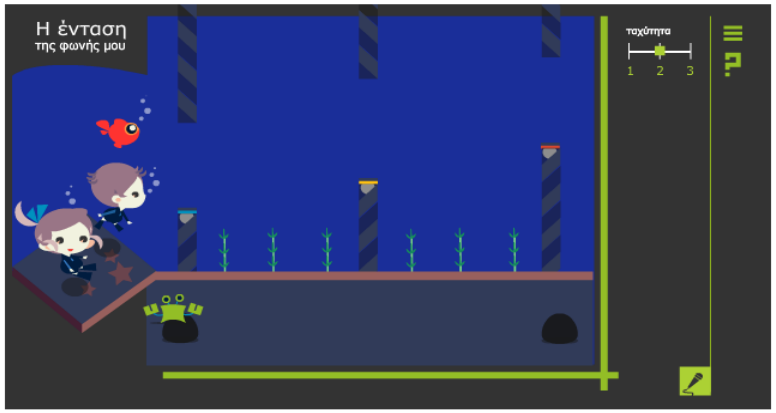
Μαθαίνω να παράγω φωνή

Πίσω στις εφαρμογές



Μαθαίνω να ελέγχω την ένταση της φωνής μου

Πίσω στις εφαρμογές



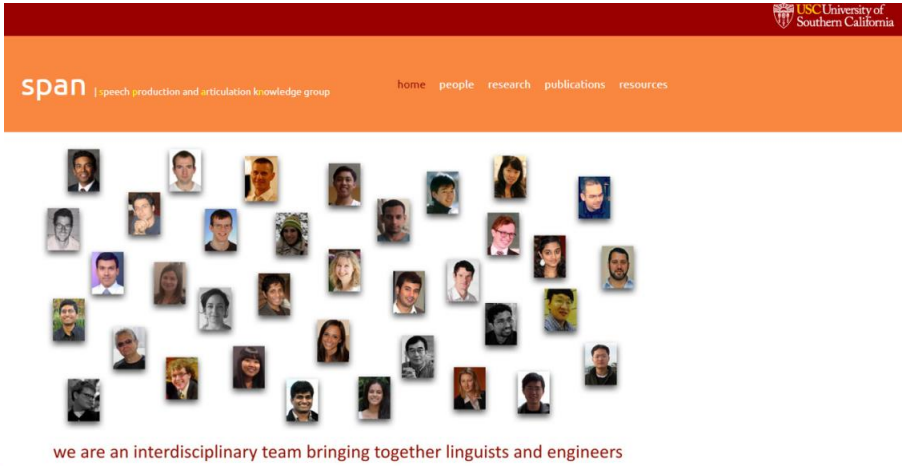
Μαθαίνω να ελέγχω την ένταση της φωνής μου

Πίσω στις εφαρμογές

Interdisciplinary Research

- **Phoneticians/Linguists + Engineers**
- **ENRICH**: Speech modifications/enhancements for easier cognitive processing
  http://www.enrich-etn.eu/
- **SPAN** (**S**peech **P**roduction and **A**rticulation K**n**owledge Group) University of Southern California
  http://sail.usc.edu/span/index.html

https://youtu.be/_2W52Y3IE_Y



ENRICH EUROPEAN TRAINING NETWORK

Fruitful interdisciplinary cooperation

SSPL

CSD, UoC

Articulation of Vowels & Consonants

Speech Production

https://www.youtube.com/watch?v=osvE5Op1VzM&t=9s

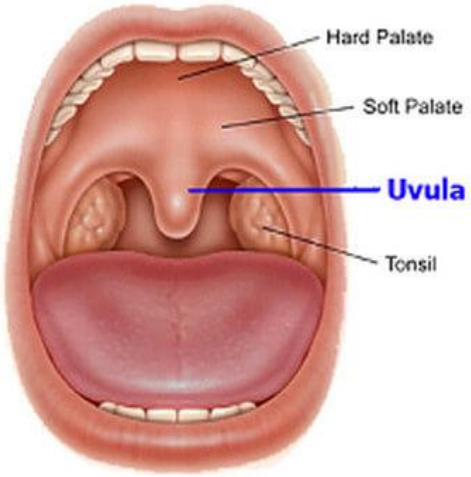
Speech Production

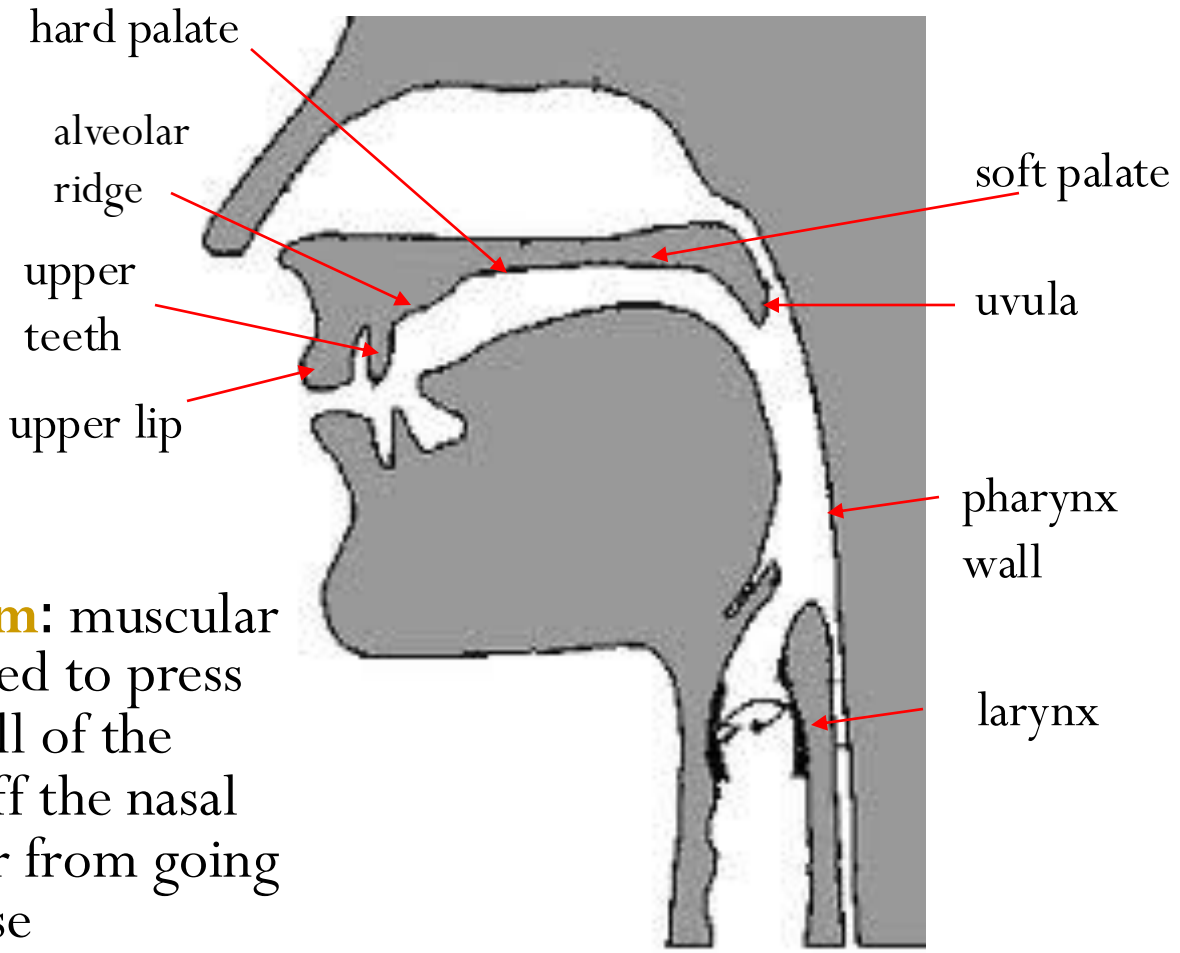
https://www.youtube.com/watch?v=SVKR3ESdAk8

Real-time MRI span, USC University of Southern California

- The tongue and lips move rapidly from one position to another.
- The actions of the tongue are among the **fastest** and the most **precise** physical movements that people make.

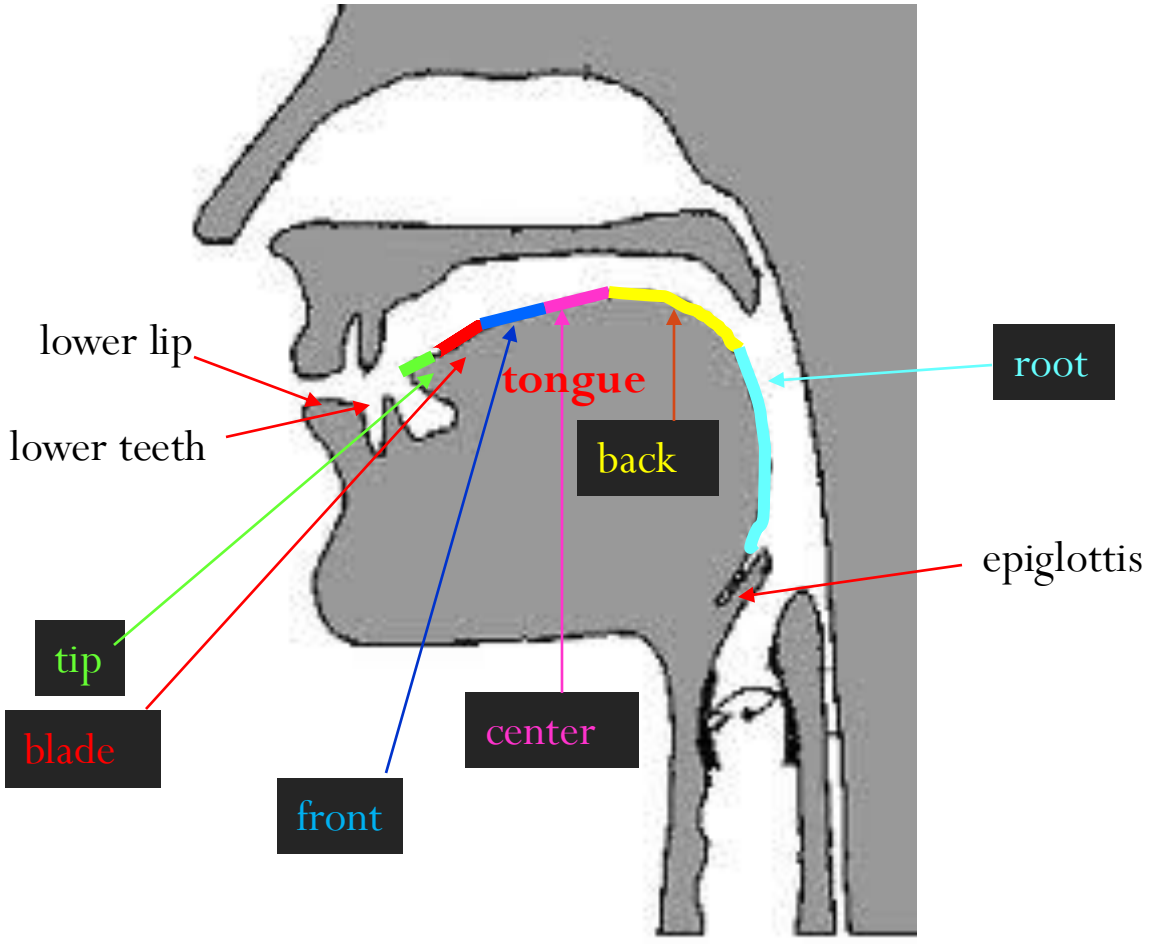
Upper surface articulators



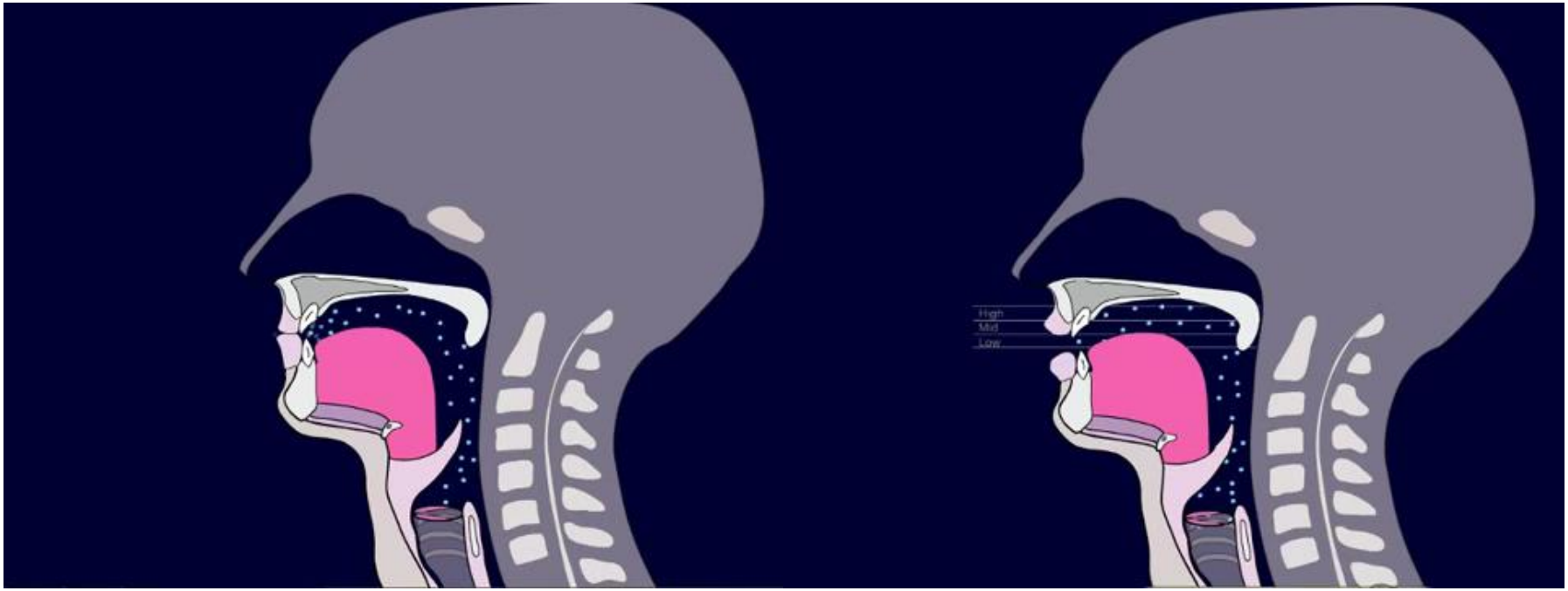


- **Soft palate/velum**: muscular flap that can be raised to press against the back wall of the pharynx and shut off the nasal tract, preventing air from going out through the nose (**velic closure**).

Lower surface articulators

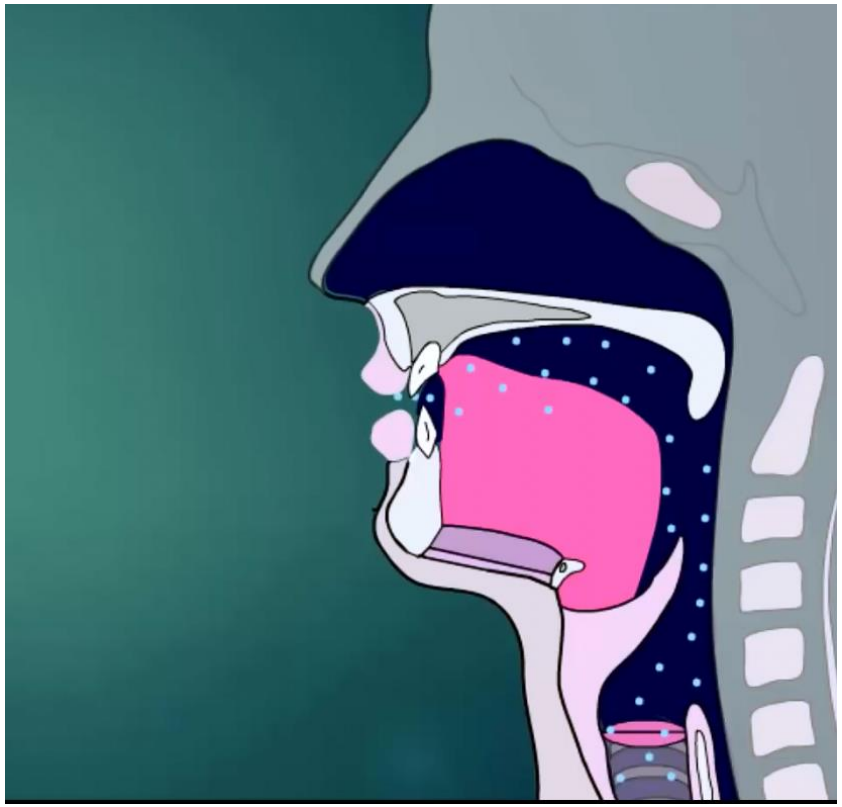


Consonants - Vowels



Consonants

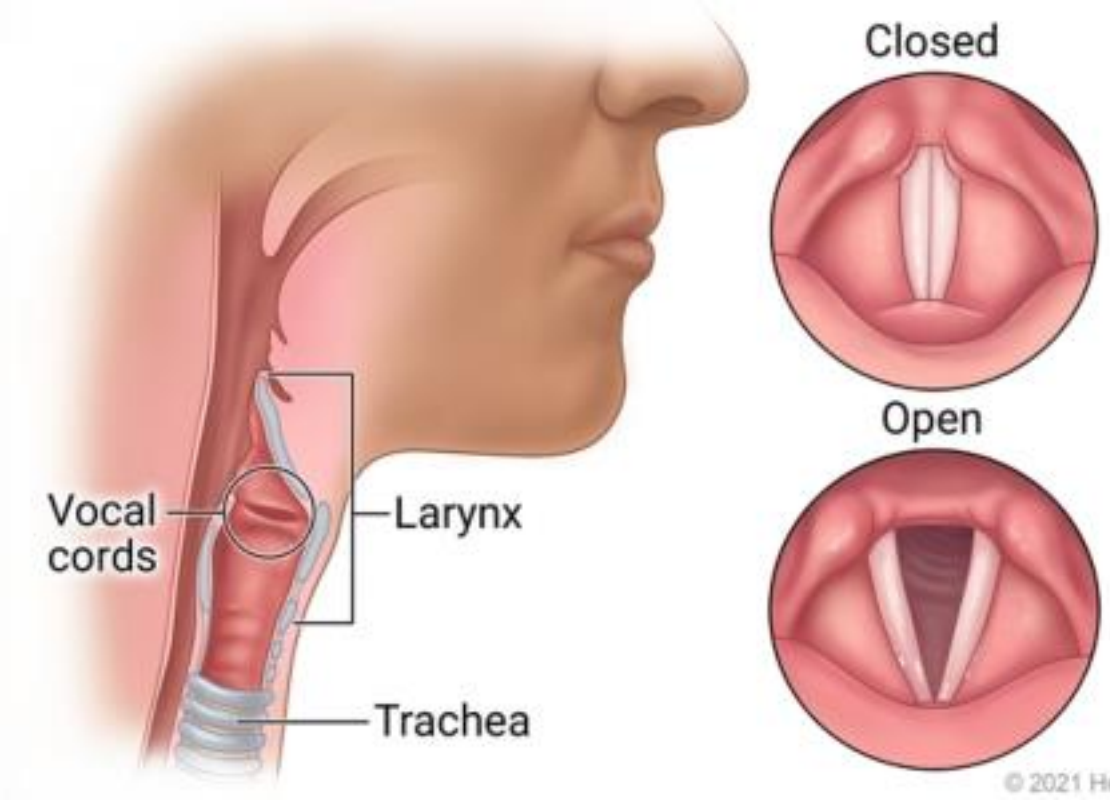
- Voicing
- Manner of articulation
- Place of articulation



Voicing

Vocal folds

- Open vocal folds:
  - breathing
  - production of voiceless sounds
- Closed vocal folds:
  - production of voiced sounds (phonation)



Exercise:
[fffffffffvvvvvvvvfffffffffffvvvv]

Voicing

Vocal folds

- Open vocal folds:
  - breathing
  - production of voiceless sounds

[p] voiceless

https://www.youtube.com/watch?v=LsAjJwC4JTQ

- Closed vocal folds:
  - production of voiced sounds (phonation)

[b] voiced

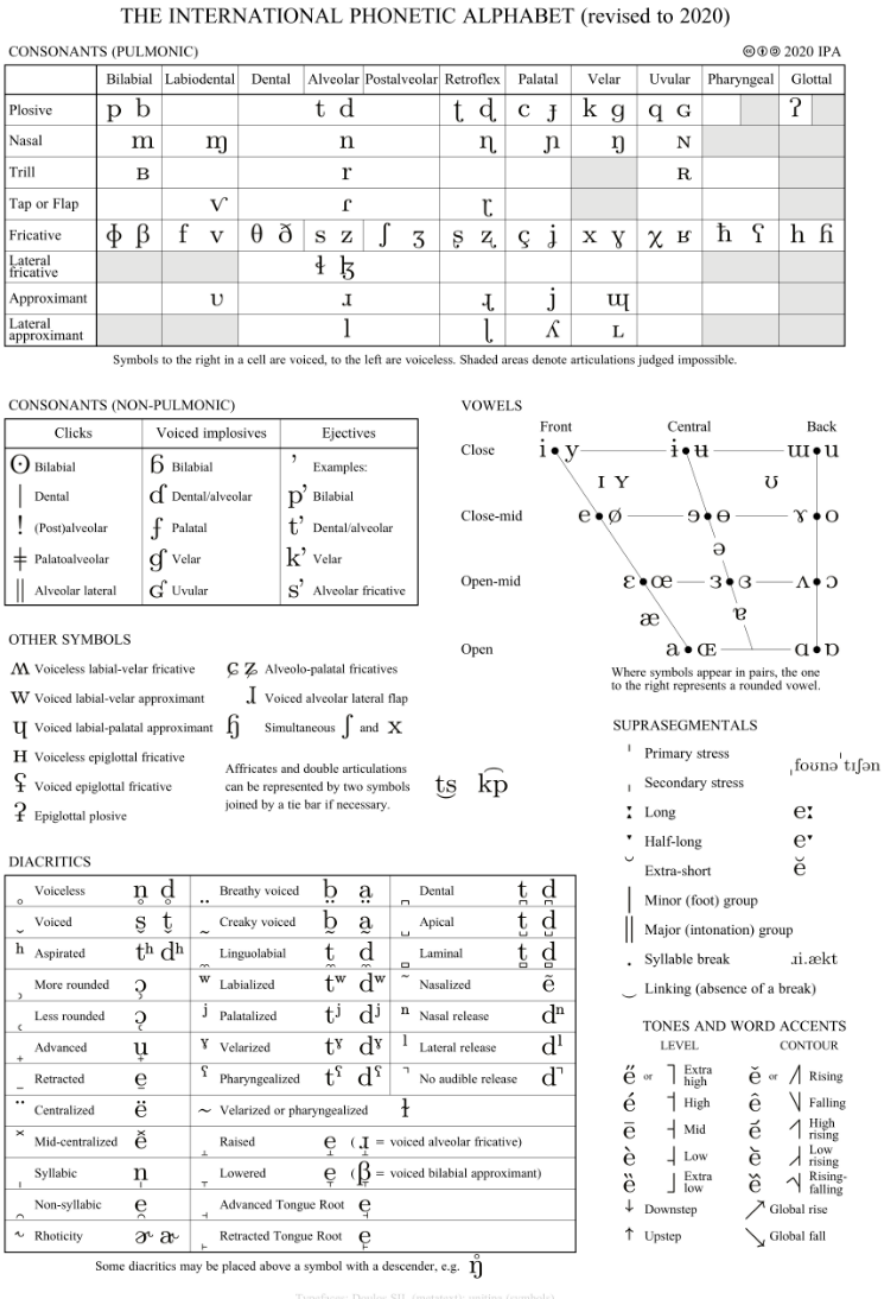
https://www.youtube.com/watch?v=eSaT1Cg1FbU

Consonants

- Voicing — Are vocal folds open or closed?
- Manner of articulation — How is the air constricted?
- Place of articulation — Where is the air constricted?

IPA Chart (2020)

Download from:
https://www.internationalphoneticassociation.org/IPAcharts/IPA_chart_orig/IPA_charts_E.html



THE INTERNATIONAL PHONETIC ALPHABET (revised to 2020)

CONSONANTS (PULMONIC) ©①⑤ 2020 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
ʘ Bilabial	ɓ Bilabial	ʼ Examples:
ǀ Dental	ɗ Dental/alveolar	pʼ Bilabial
ǃ (Post)alveolar	ʄ Palatal	tʼ Dental/alveolar
ǂ Palatoalveolar	ɠ Velar	kʼ Velar
ǁ Alveolar lateral	ʛ Uvular	sʼ Alveolar fricative

OTHER SYMBOLS

ʍ Voiceless labial-velar fricative
w Voiced labial-velar approximant
ɥ Voiced labial-palatal approximant
ʜ Voiceless epiglottal fricative
ʢ Voiced epiglottal fricative
ʡ Epiglottal plosive
ɕ ʑ Alveolo-palatal fricatives
ɺ Voiced alveolar lateral flap
ʃ͡ʒ Simultaneous ʃ and x

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary. t͡s k͡p

DIACRITICS

◌̥ Voiceless	n̥ d̥	◌̤ Breathy voiced	b̤ a̤	◌̪ Dental	t̪ d̪
◌̬ Voiced	s̬ t̬	◌̰ Creaky voiced	b̰ a̰	◌̺ Apical	t̺ d̺
◌ʰ Aspirated	tʰ dʰ	◌̼ Linguolabial	t̼ d̼	◌̻ Laminal	t̻ d̻
◌̹ More rounded	ɔ̹	◌ʷ Labialized	tʷ dʷ	◌̃ Nasalized	ẽ
◌̜ Less rounded	ɔ̜	◌ʲ Palatalized	tʲ dʲ	◌ⁿ Nasal release	dⁿ
◌̟ Advanced	u̟	◌ˠ Velarized	tˠ dˠ	◌ˡ Lateral release	dˡ
◌̠ Retracted	e̠	◌ˤ Pharyngealized	tˤ dˤ	◌̚ No audible release	d̚
◌̈ Centralized	ë	~ Velarized or pharyngealized	ɫ		
◌̽ Mid-centralized	e̽	◌̝ Raised	e̝ (ɹ̝ = voiced alveolar fricative)		
◌̩ Syllabic	n̩	◌̞ Lowered	e̞ (β̞ = voiced bilabial approximant)		
◌̯ Non-syllabic	e̯	◌̘ Advanced Tongue Root	e̘		
◌˞ Rhoticity	ə˞ a˞	◌̙ Retracted Tongue Root	e̙		

Some diacritics may be placed above a symbol with a descender, e.g. ŋ̊

VOWELS

	Front	Central	Back
Close	i • y	ɨ • ʉ	ɯ • u
	ɪ ʏ		ʊ
Close-mid	e • ø	ɘ • ɵ	ɤ • o
		ə	
Open-mid	ɛ • œ	ɜ • ɞ	ʌ • ɔ
	æ	ɐ	
Open	a • ɶ		ɑ • ɒ

Where symbols appear in pairs, the one to the right represents a rounded vowel.

SUPRASEGMENTALS

ˈ Primary stress ˌfoʊnəˈtɪʃən
ˌ Secondary stress
ː Long eː
ˑ Half-long eˑ
◌̆ Extra-short ĕ
| Minor (foot) group
‖ Major (intonation) group
. Syllable break ɹi.ækt
‿ Linking (absence of a break)

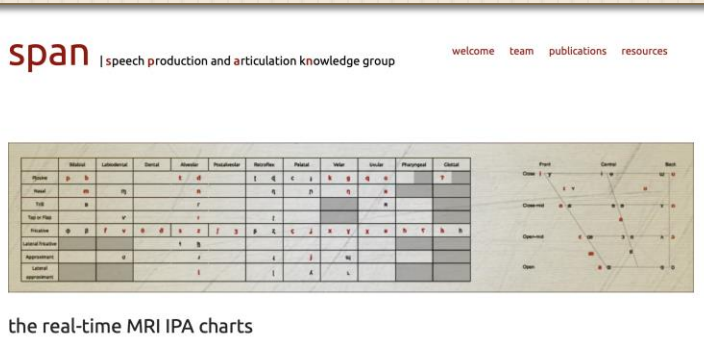
TONES AND WORD ACCENTS

LEVEL		CONTOUR	
e̋ or ˥	Extra high	ě or ˧˥	Rising
é ˦	High	ê ˥˩	Falling
ē ˧	Mid	e᷄ ˦˥	High rising
è ˨	Low	e᷅ ˩˧	Low rising
ȅ ˩	Extra low	e᷈ ˧˥˧	Rising-falling
↓ Downstep		↗ Global rise	
↑ Upstep		↘ Global fall	

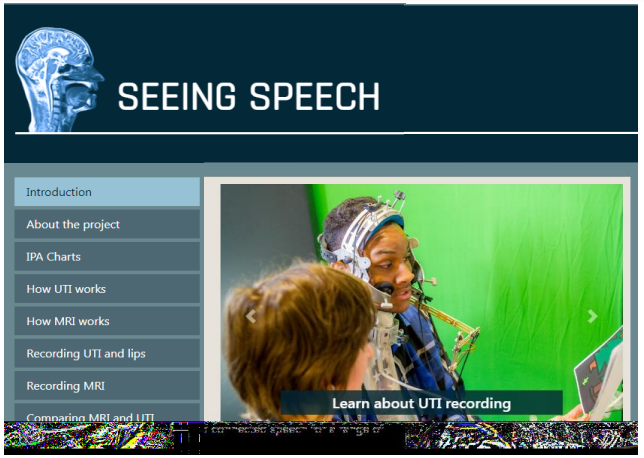
Interactive IPA Charts

IPA Online:
https://teaching.ncl.ac.uk/ipa/links.html

- **SPAN** (**S**peech **P**roduction and **A**rticulation K**n**owledge Group) University of Southern California https://sail.usc.edu/span/rtmri_ipa/



- **SEEING SPEECH**
6 Scottish Universities (Scottish Consortium)
https://www.seeingspeech.ac.uk/



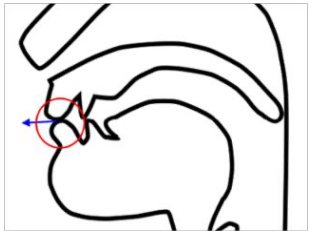
- **eNunciate** (**A Visual Language Learning Tool**) The University of British Columbia
https://enunciate.arts.ubc.ca/

Manner of Articulation

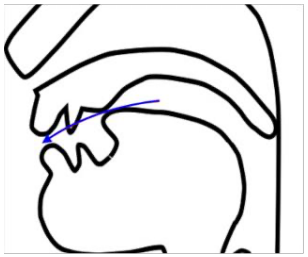
Consonants

Plosive
Nasal
Trill
Tap or flap
Fricative
Lateral fricative
Approximant
Lateral approximant

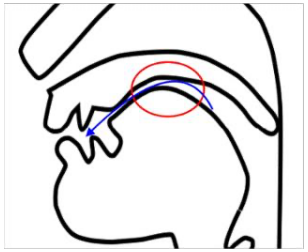
Complete blockage of air flow



Partial blockage of air flow (turbulence)

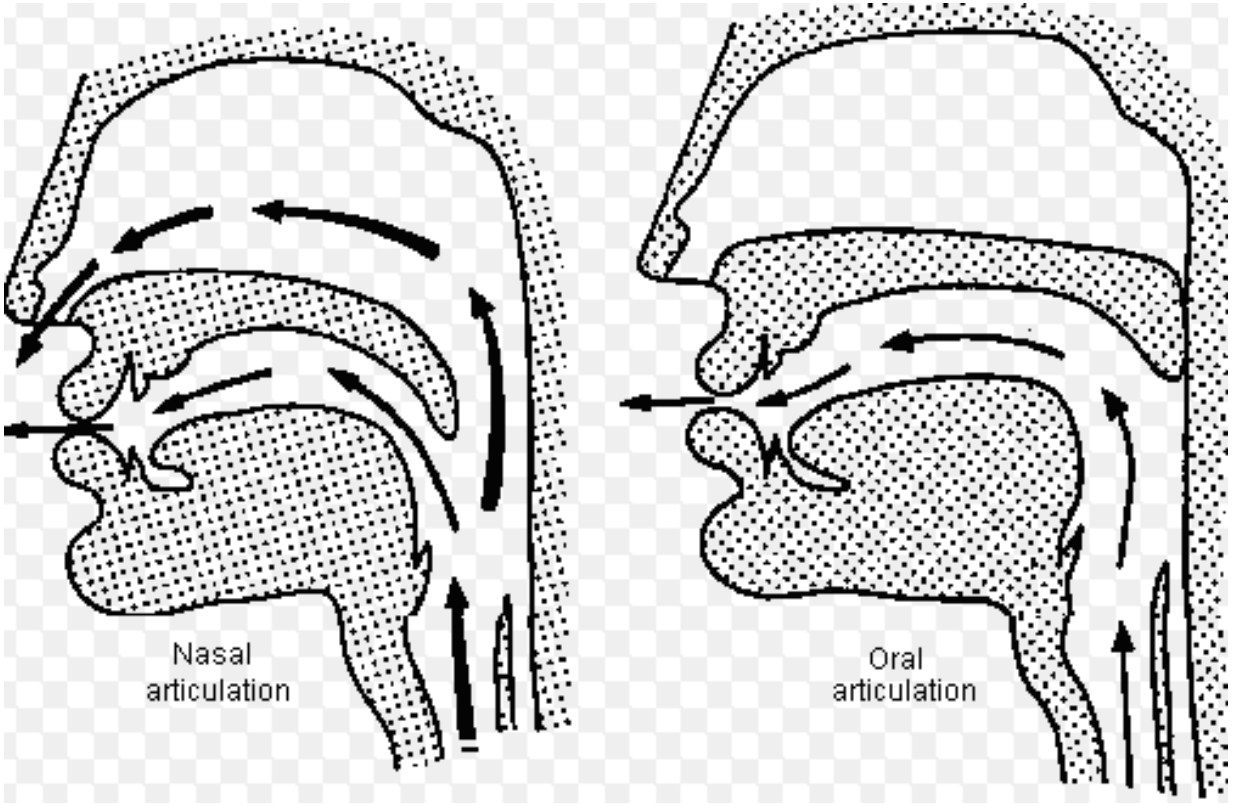


Partial blockage of air flow (no turbulence)



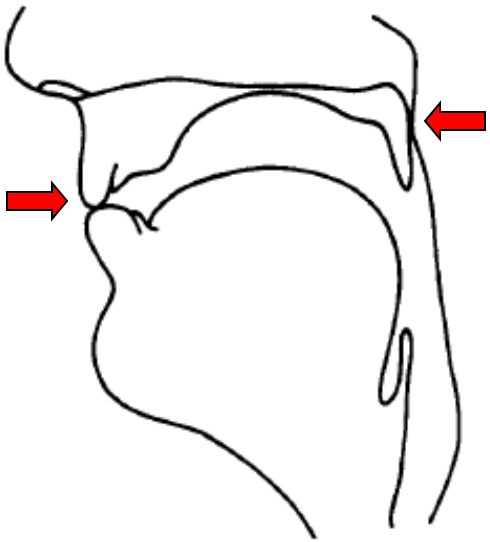
Nasal vs. oral articulation

- Nasal sounds
- Oral sounds

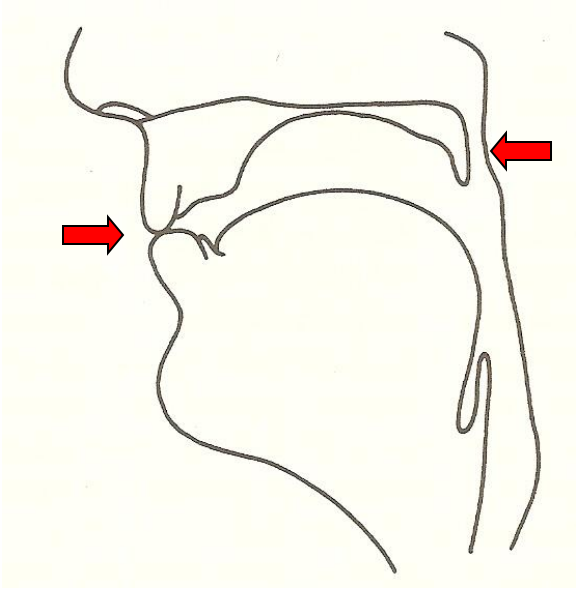


Manner of Articulation

Oral Stop



Nasal Stop



Place of Articulation

Basic places of articulation

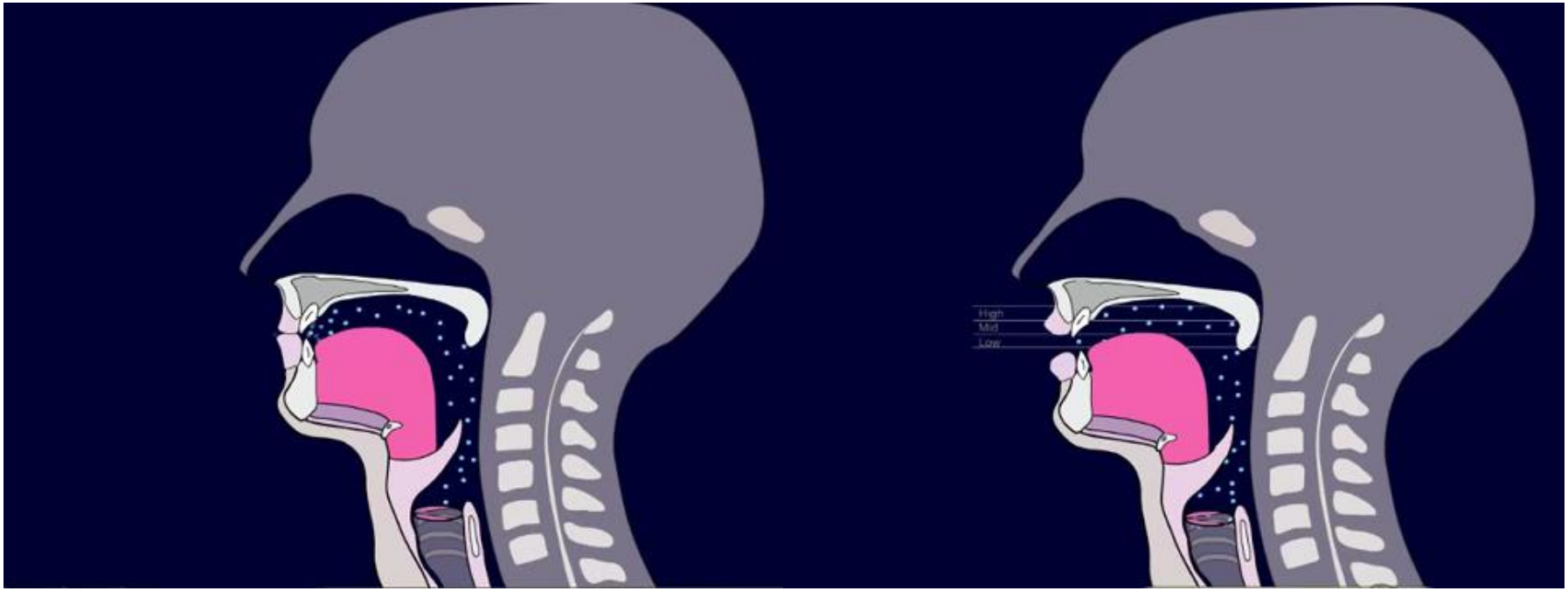
Articulator	Articulation
• lips	labial
• tongue tip/blade	coronal
• tongue dorsum	dorsal

Example: “topic”

Articulatory description of Consonants

Voicing	Place of Articulation	Manner of Articulation
[p]	voiceless	bilabial	stop
[z]	voiced	alveolar	fricative
[ʔ]	voiced	velar	nasal

Consonants - Vowels

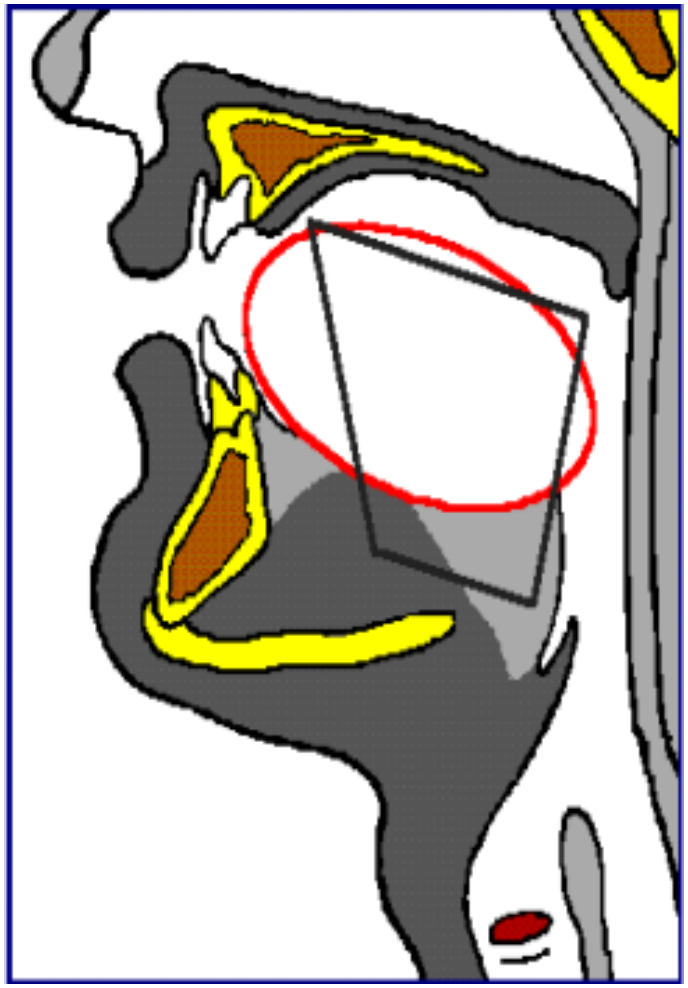


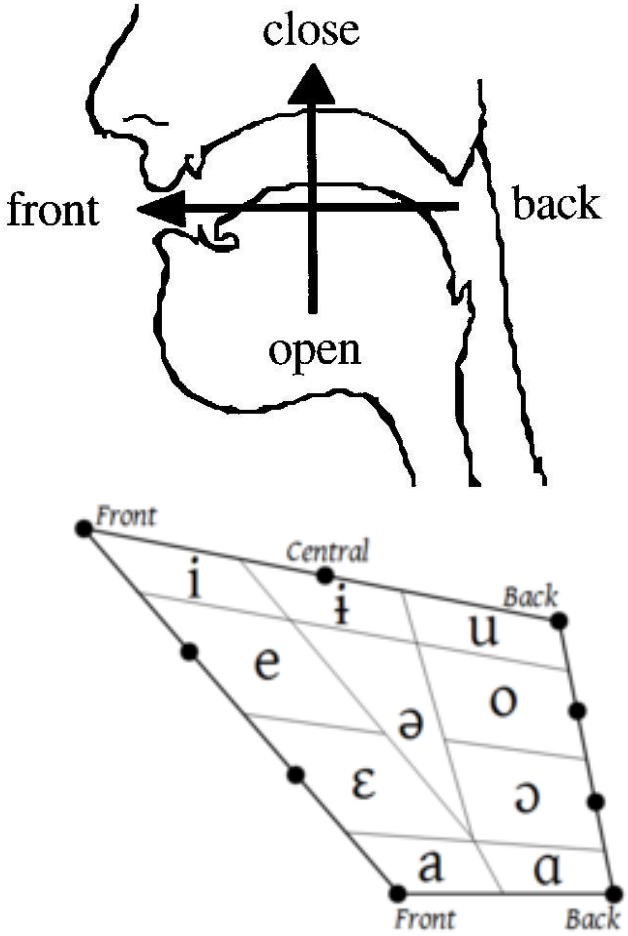
Articulation of vowels

- Articulators do not come very close together → the passage of the airstream is relatively unobstructed.

- Articulatory description focuses on
  - Position of highest point of the tongue
  - Position of the lips

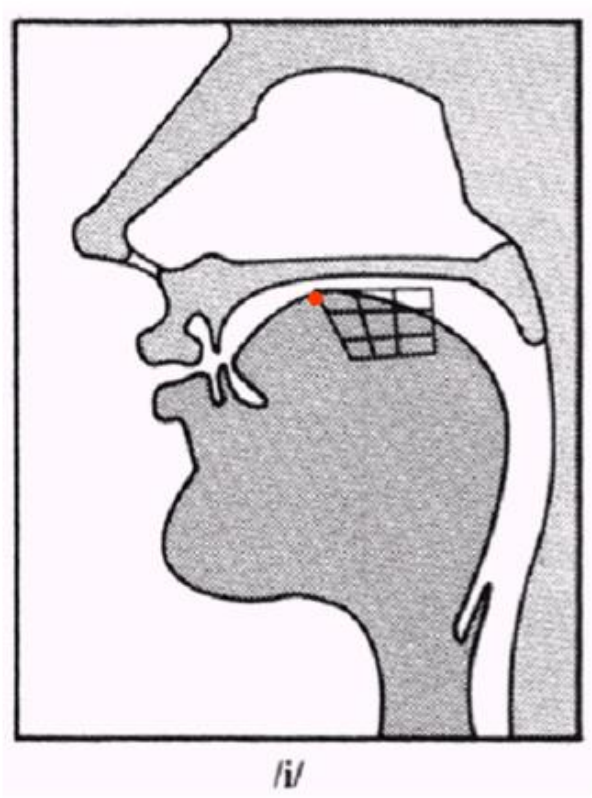
Tongue position





Principles of the IPA (1949)

Tongue position-vowel quadrilateral



UCLA tongue video

- X ray video of tongue and lip movement during production of vowels /i, e, a, o, u/.

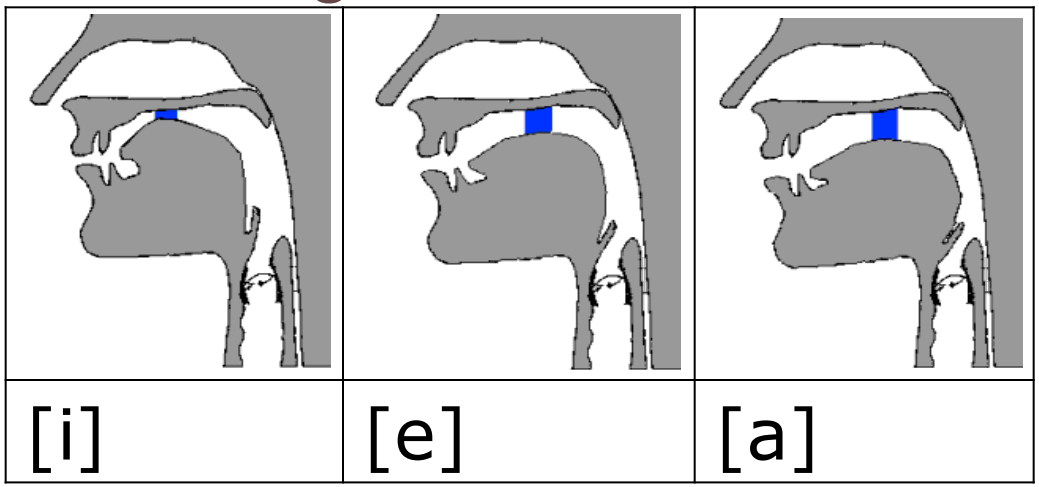
- Download from

http://www.phonetics.ucla.edu/vowels/chapter11/tongue.html

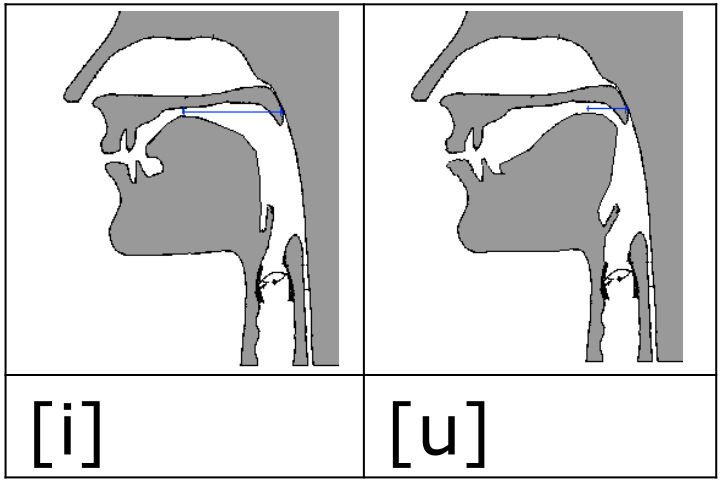
Articulatory description of vowels

1. **height** of tongue body
2. **front-back** position of the tongue
3. degree of lip rounding

high/low dimension

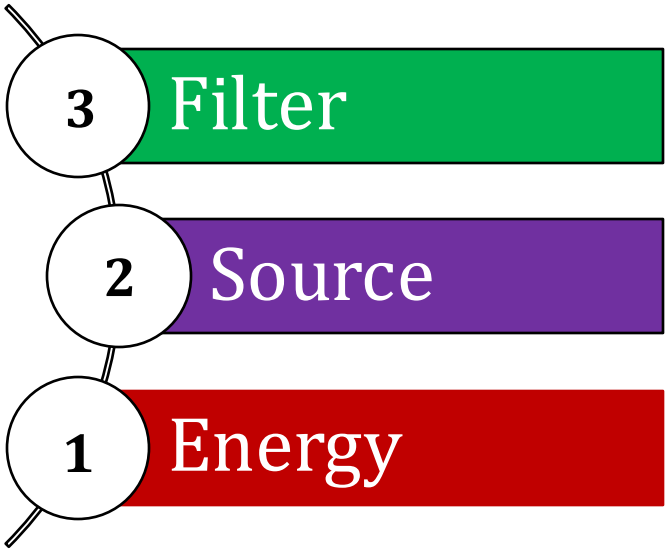


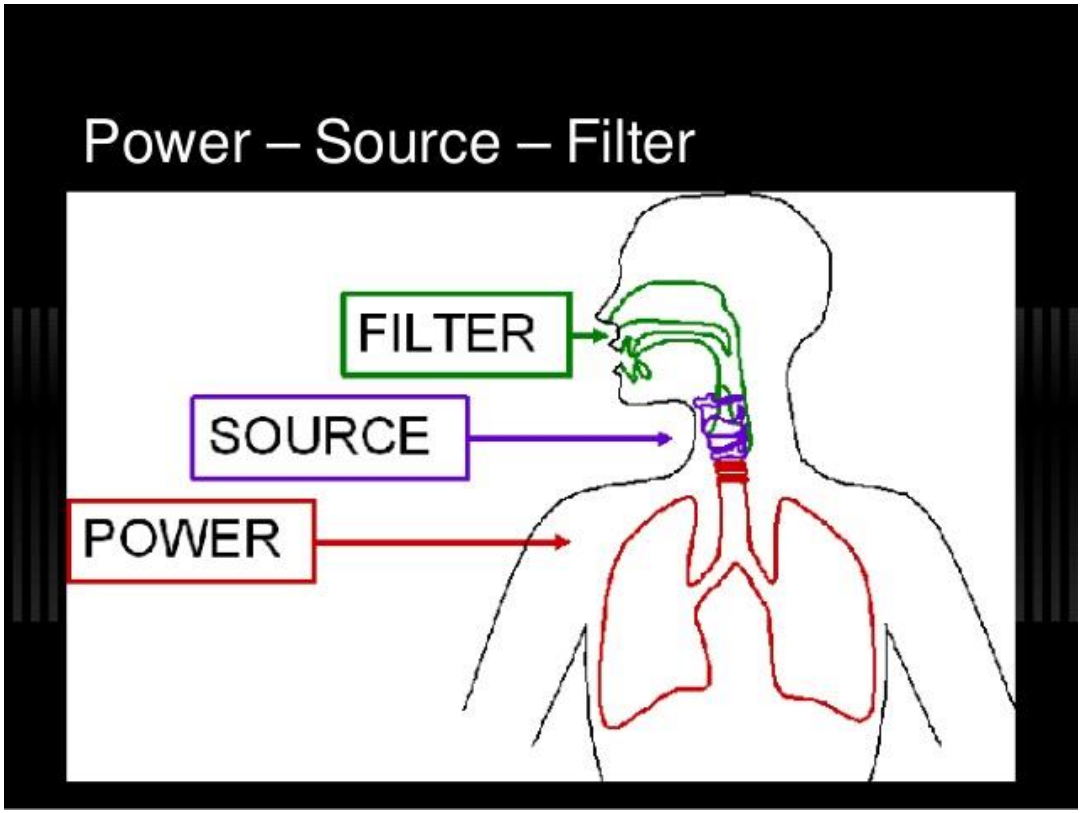
front/back dimension



Acoustics of Vowels & Consonants

Source-Filter Theory





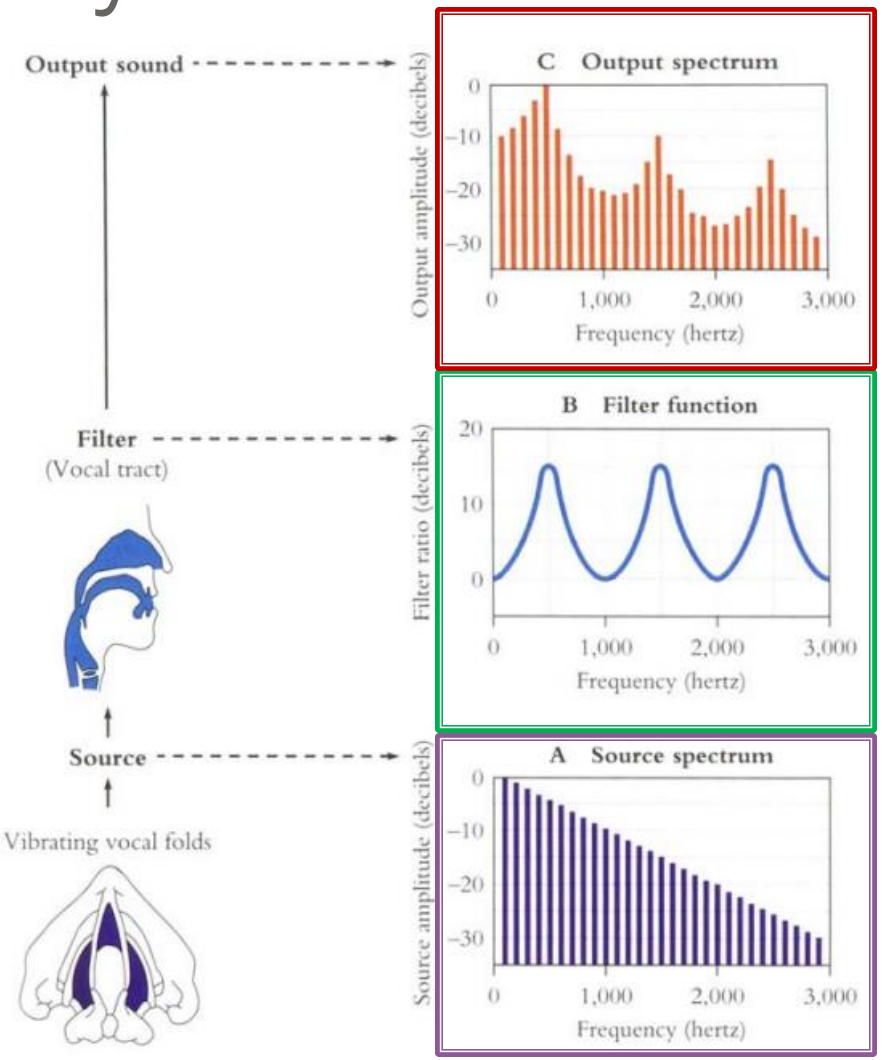
https://www.vocalsonstage.com/vocals-on-stage-blog/resonance-and-articulation

Source-Filter Theory

- The output spectrum is formed by the filter and is different for each sound.

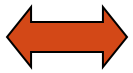
- The filter **amplifies** or **diminishes** frequency components and varies according to **vocal tract shape**.

- Vocal fold vibration (for voiced sounds) produces the source spectrum.
  - **Spectrum:** Energy of the signal distributed with frequency



Xavier Anguera Miró
http://www.xavieranguera.com/tdp_2011/4-Source-Filter-Models.pdf

Formants

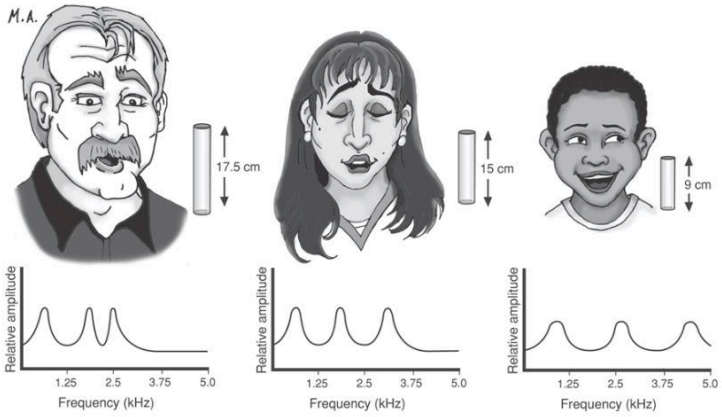
- Sounds differ from each other in three ways
  1. **pitch/frequency**
  2. **loudness**
  3. **quality**
- A vowel sound contains a number of different pitches simultaneously
  - pitch at which it was spoken
  - various overtone pitches that give it its distinctive quality
- Vowel Quality  Overtone Structure
- **<u>Overtones = Formants</u>**
- The lowest 3 formants distinguish vowels from each other
  - **F1** **F2** **F3**

Fundamental Frequency (F0)

- **Fundamental frequency**: number of vocal fold vibrations per second.
- Vocal folds must be vibrating in order to have F0.
- It corresponds to variations in pitch (speech melody or intonation).
- Vocal folds may vibrate faster or slower giving higher or lower pitch to the sound, BUT the formants of the sound remain the same as long as vocal tract shape remains unchanged.

- Male voice: 120 Hz
- Female voice: 220 Hz
- Child voice: 260-280 Hz

- All voiced sounds are distinguishable due to their formants.



Behrman (2021)

How do formants arise?

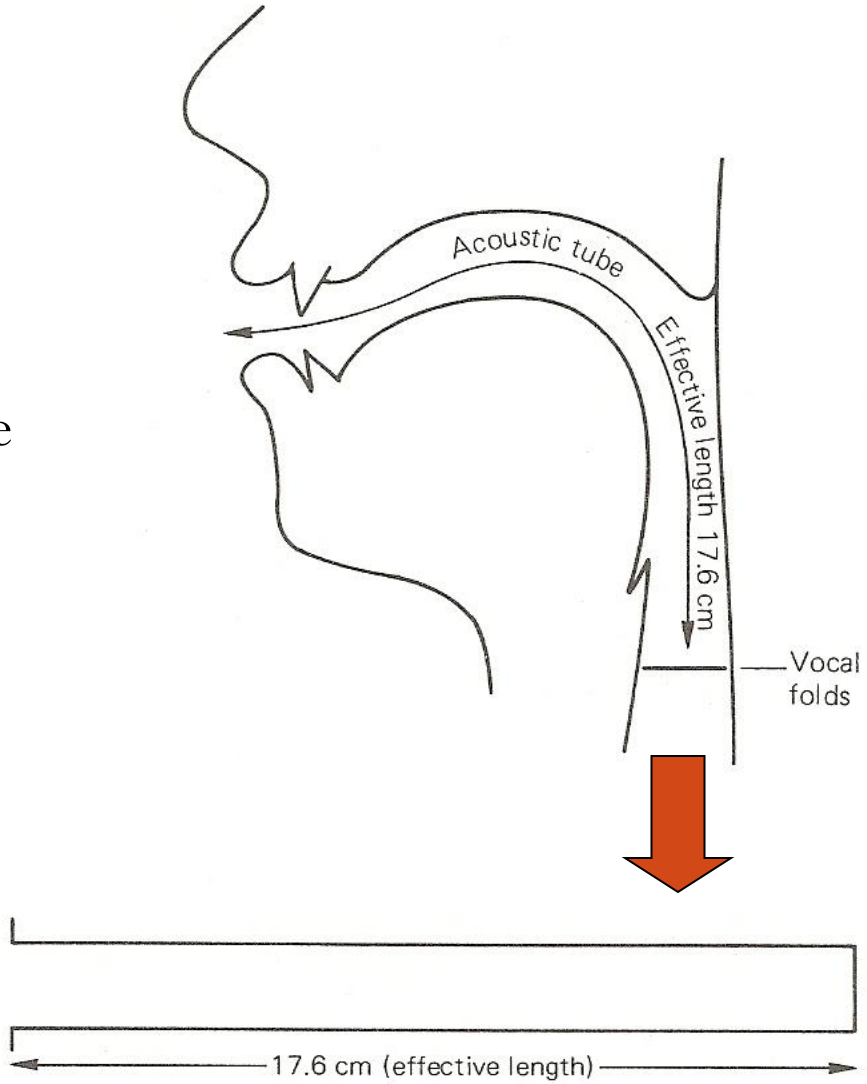
- The air in the vocal tract acts like the air in a bottle.
  - Tap on a bottle.
  - Open your mouth, make a glottal stop and flick a finger against your neck just to the side and below the jaw.

  What do you observe?

  - Articulate [i, e, a, o, u] without producing sound.

  What do you observe?

Pitch of F1 going up for [i, e] and down for [a, o, u]

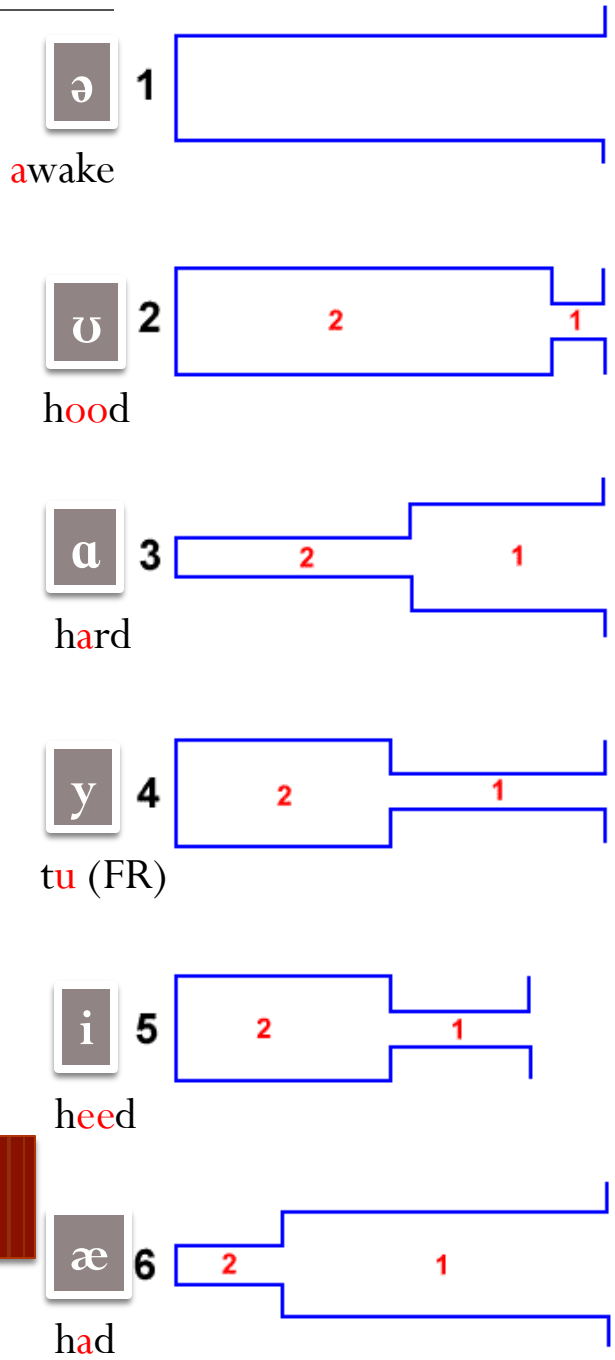


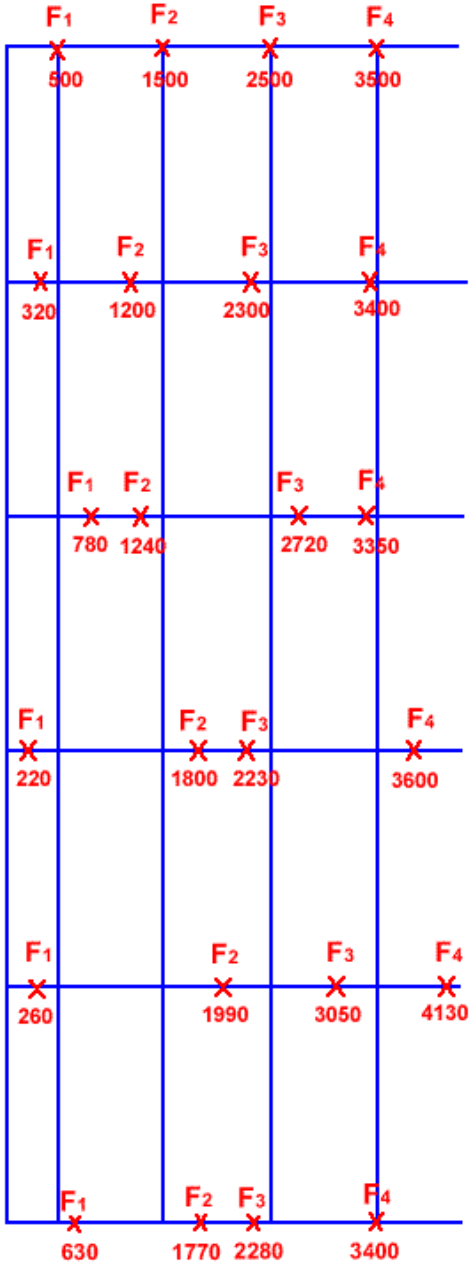
Tube models

- Formants that characterize different vowels are the result of the different shapes of the vocal tract.
- Any body of air will vibrate in a way that depends on its size and shape.
  - Blow across the top of
    - an empty bottle
    - partially filled bottle

  What do you observe?

Great volume of air → low-pitched note
Small volume of air → high-pitched note

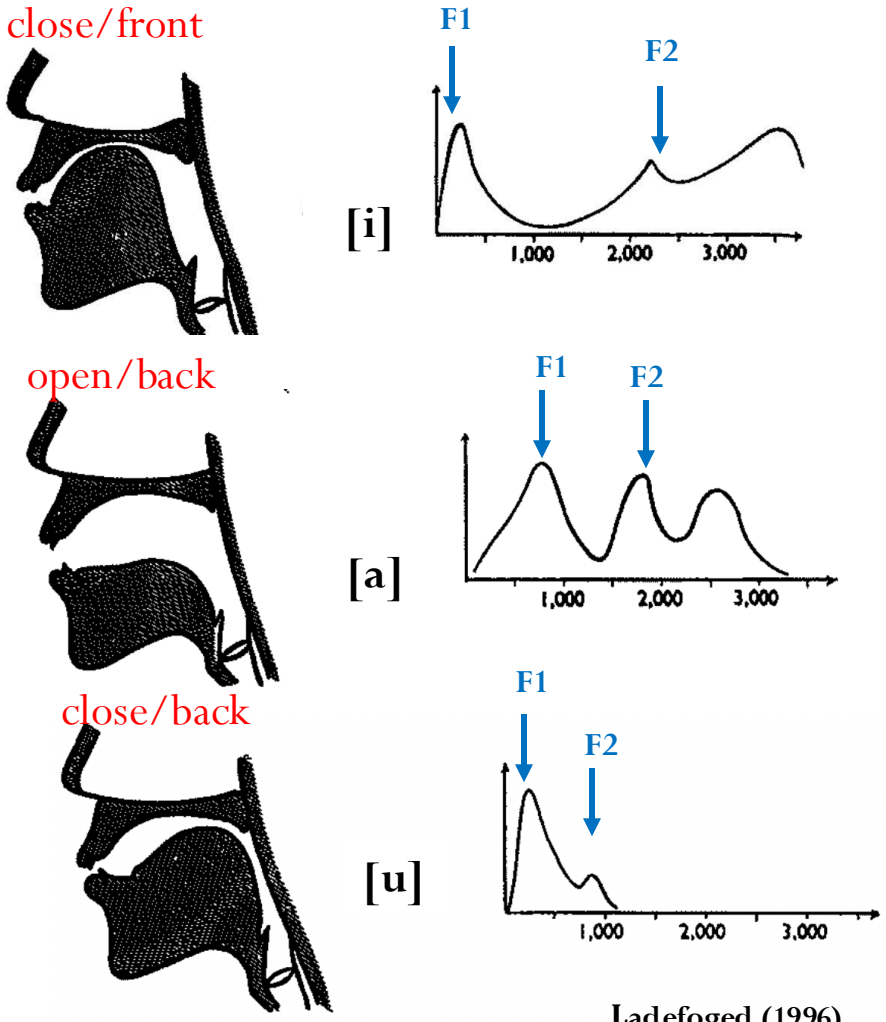




Adapted from Fant (1960)

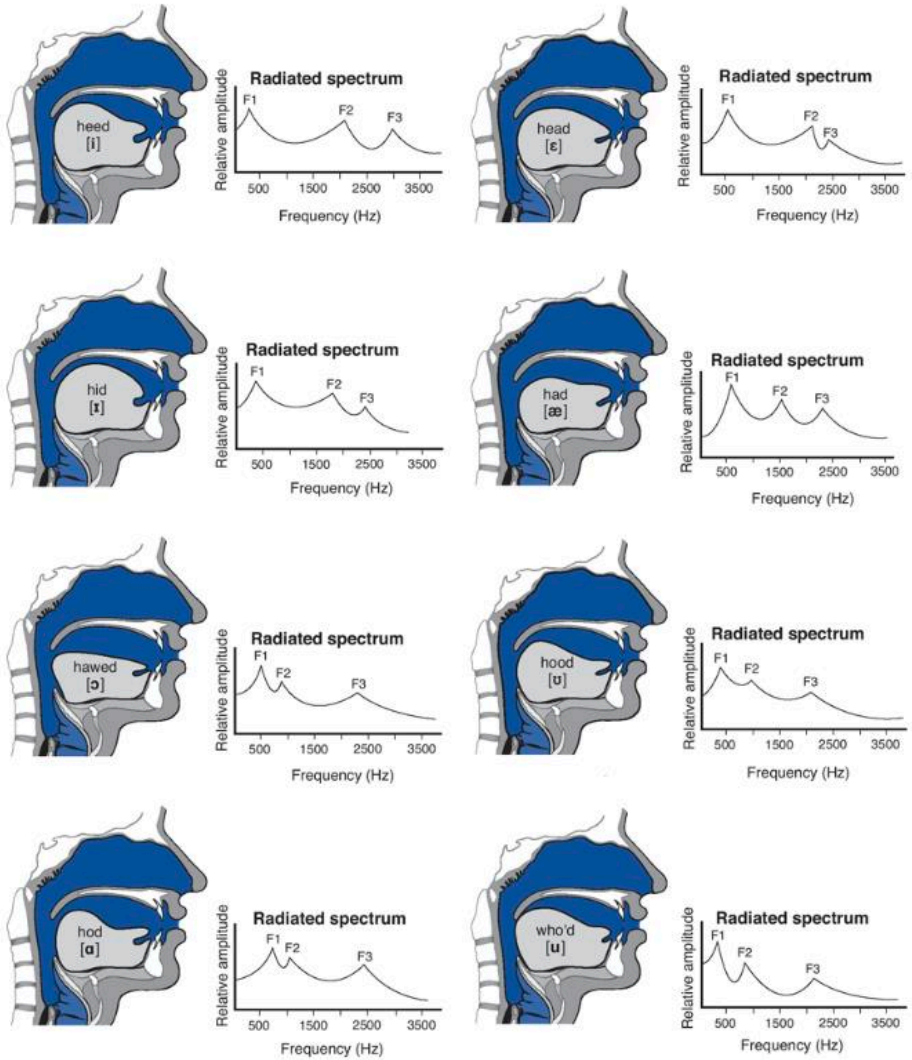
Spectra and Formants

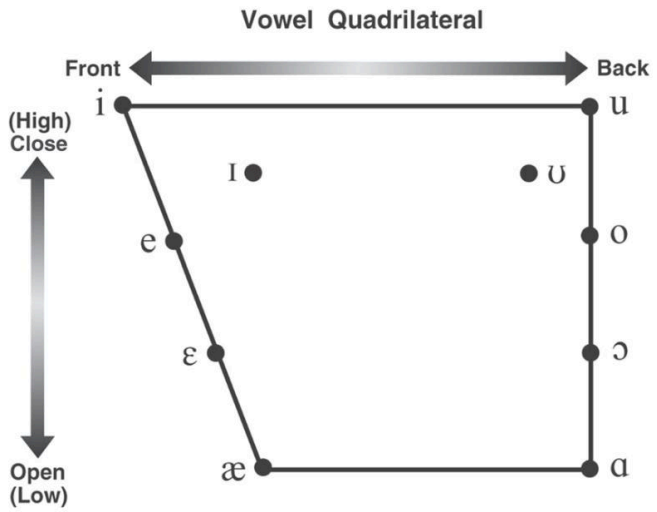
- Frequencies that are amplified, receive more energy and correspond to **formants**.
- Thus every speech sound corresponds to a different spectrum, and different formants.
  - [i]: F1 and F2 at a distance
  - [a]: F1 and F2 close
  - [u]: F1 and F2 close



Ladefoged (1996)
(altered)

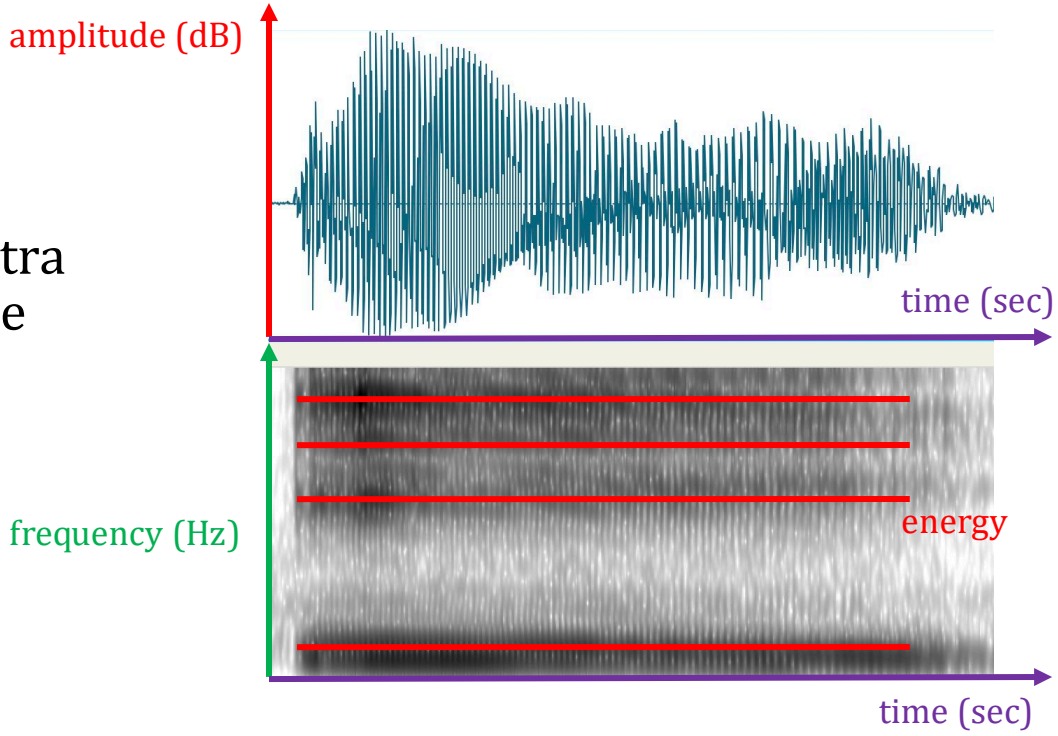
Spectra and Formants

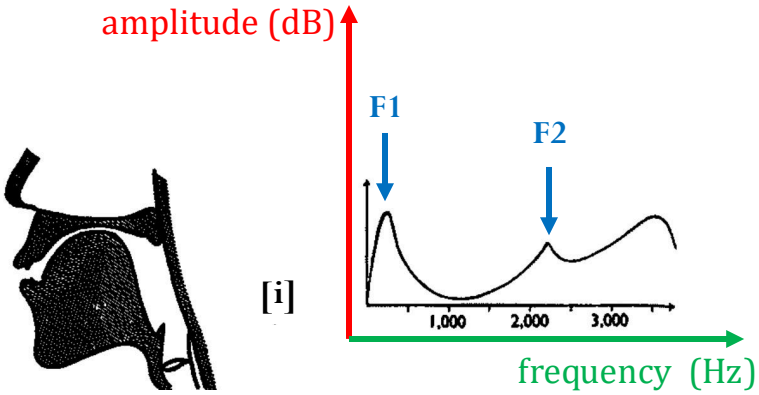




Spectrum vs Spectrogram

- Spectrum: distribution of energy with frequency
  - (amplitude vs frequency) **[2D] two-dimensional**
- Spectrogram: series of spectra at consecutive points in time
  - (frequency vs time vs amplitude/energy) **[3D] three-dimensional**

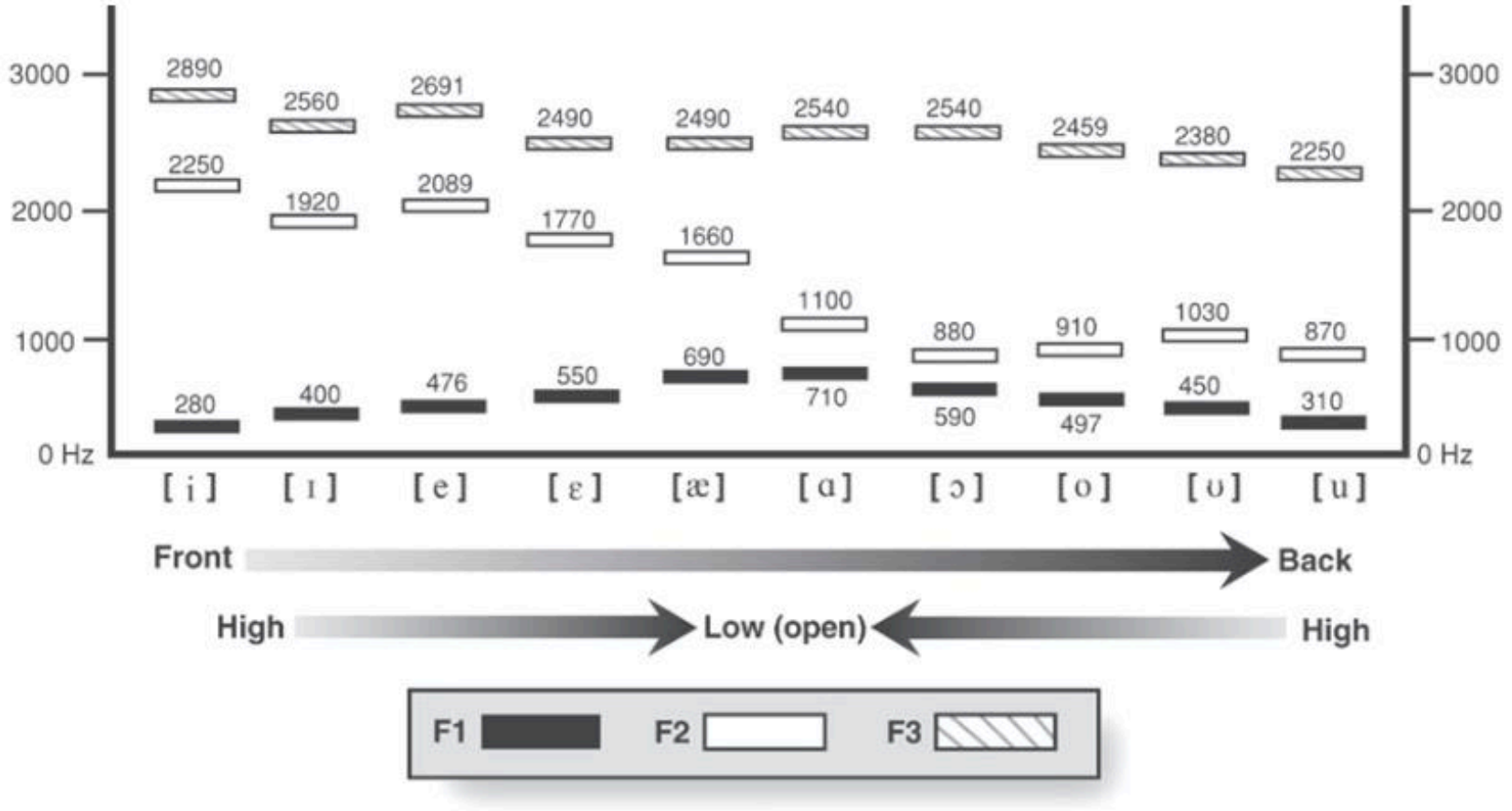




Spectrograms

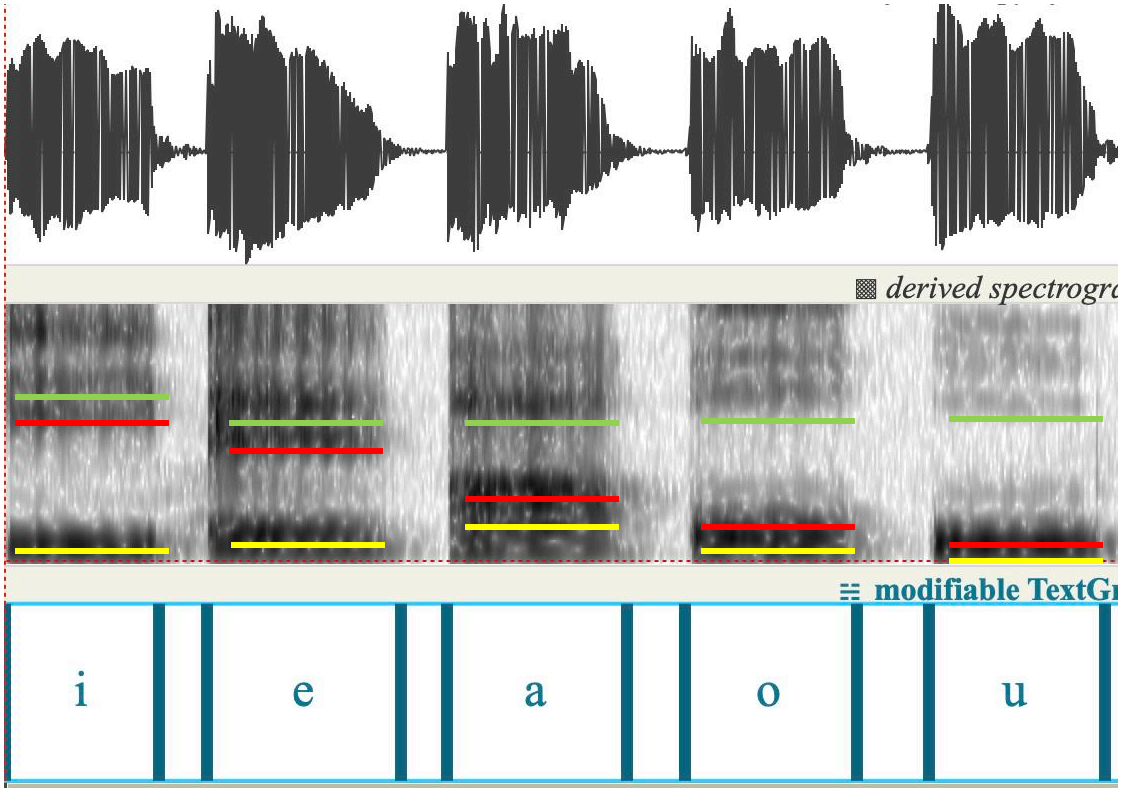
Dark bands for concentrations of energy at particular frequencies showing the source and filter characteristics of speech

Acoustics of vowels



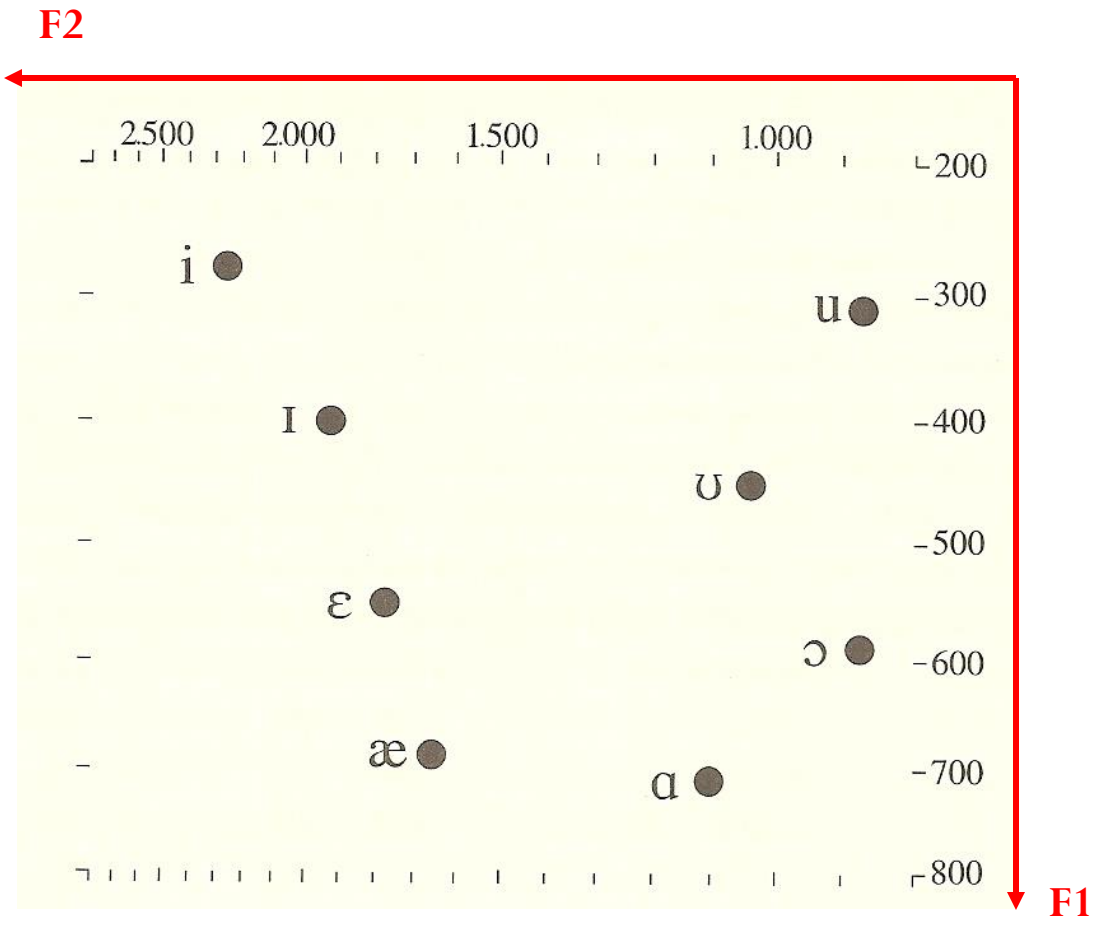
Acoustics of Greek vowels

- F1: Formant 1
  Inversely related to tongue **height**.
  High values = low (open) vowel
- F2: Formant 2
  Related to frontness (or rather F2-F1)
  High values = front vowel
- F3: Formant 3
  Related to roundedness and rhotacization
  Low values = rhotacization / roundedness



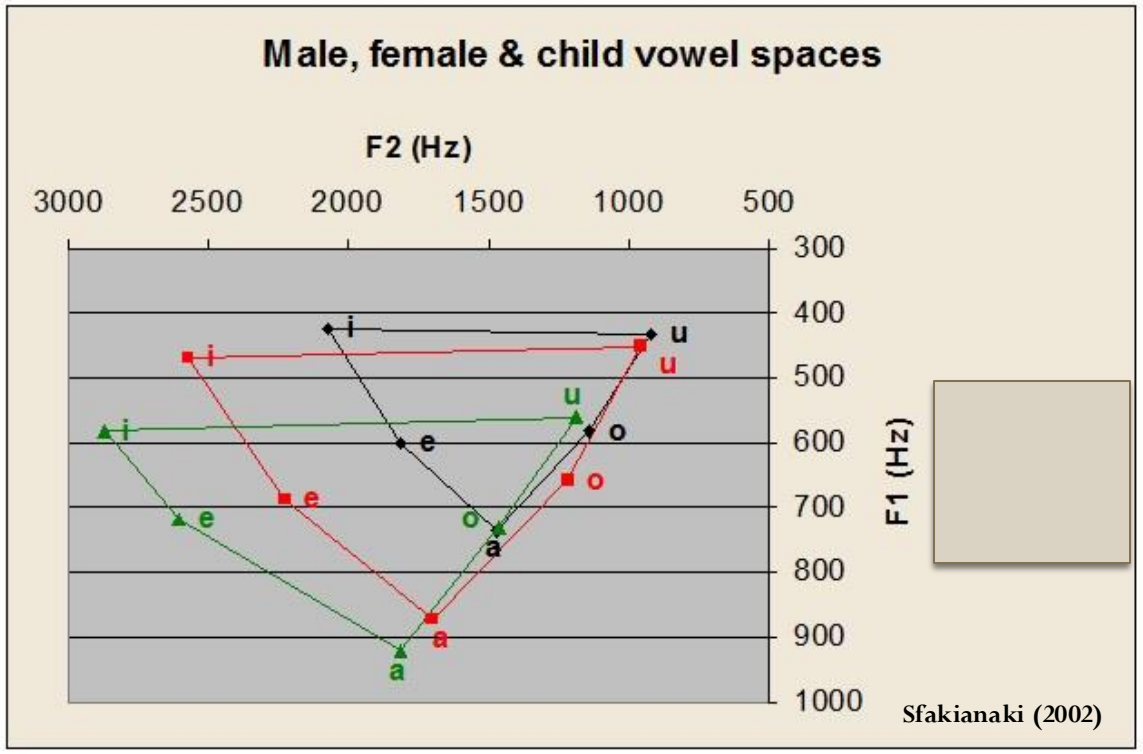
F1 by F2 plot

- Zero frequency is placed at the top right corner because formants are inversely related to traditional articulatory parameters.
- F2 scale not as expanded as F1, due to less prominent energy (F1: 80% of vowel energy).



Greek vowel space

- Formant values are influenced by anatomical characteristics (vocal tract and vocal fold size, etc.)
  - Lower in men, higher in women, even higher in children
- Formant values are also influenced by phonetic context.



Speech synthesis demo

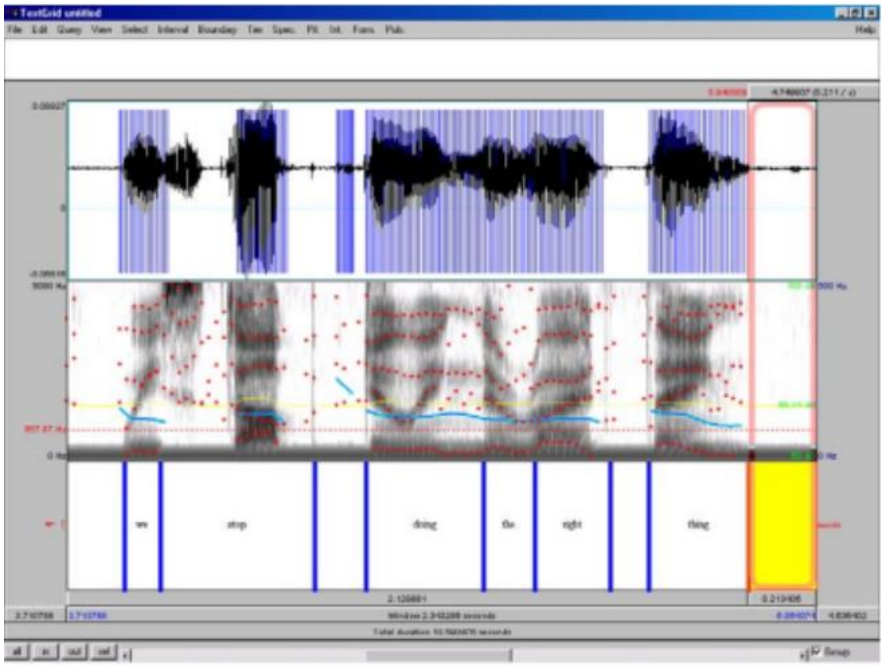
- The notion that vowels contain several different pitches at the same time is difficult to appreciate.
- The demo shows how a sentence is built from its component waves.
- This speech was synthesized in 1971 by Peter Ladefoged on a synthesizer at UCLA.
- **“A bird in the hand is worth two in the bush”**
  «Κάλιο πέντε και στο χέρι παρά δέκα και καρτέρει» (Greek translation)
- See the demo here:
  https://linguistics.berkeley.edu/acip/course/chapter8/speechbird/

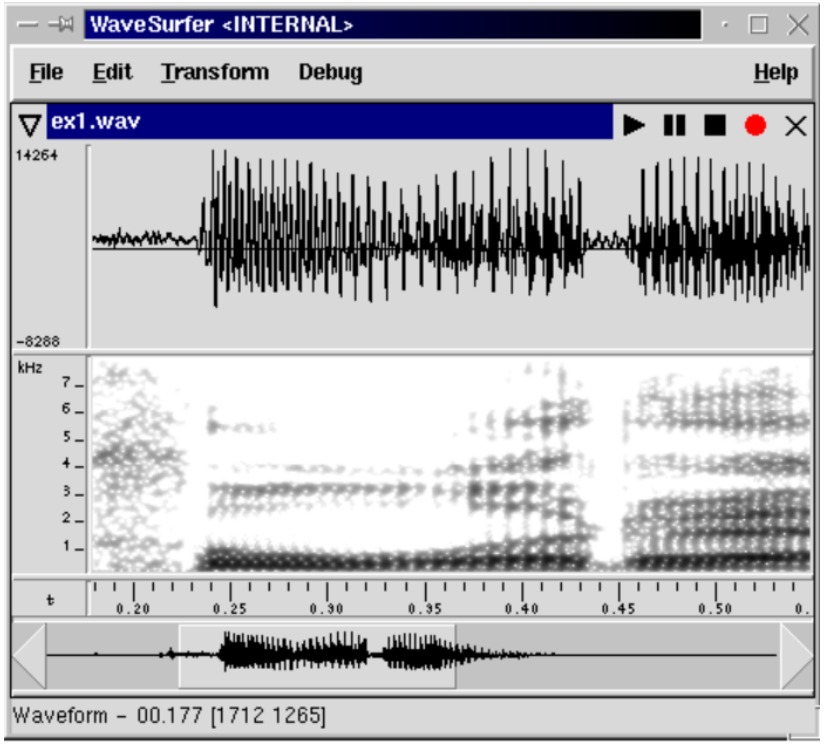
Computer Programs for acoustic analysis

(free access)

- **Praat**
  http://www.fon.hum.uva.nl/praat/

  University of Amsterdam

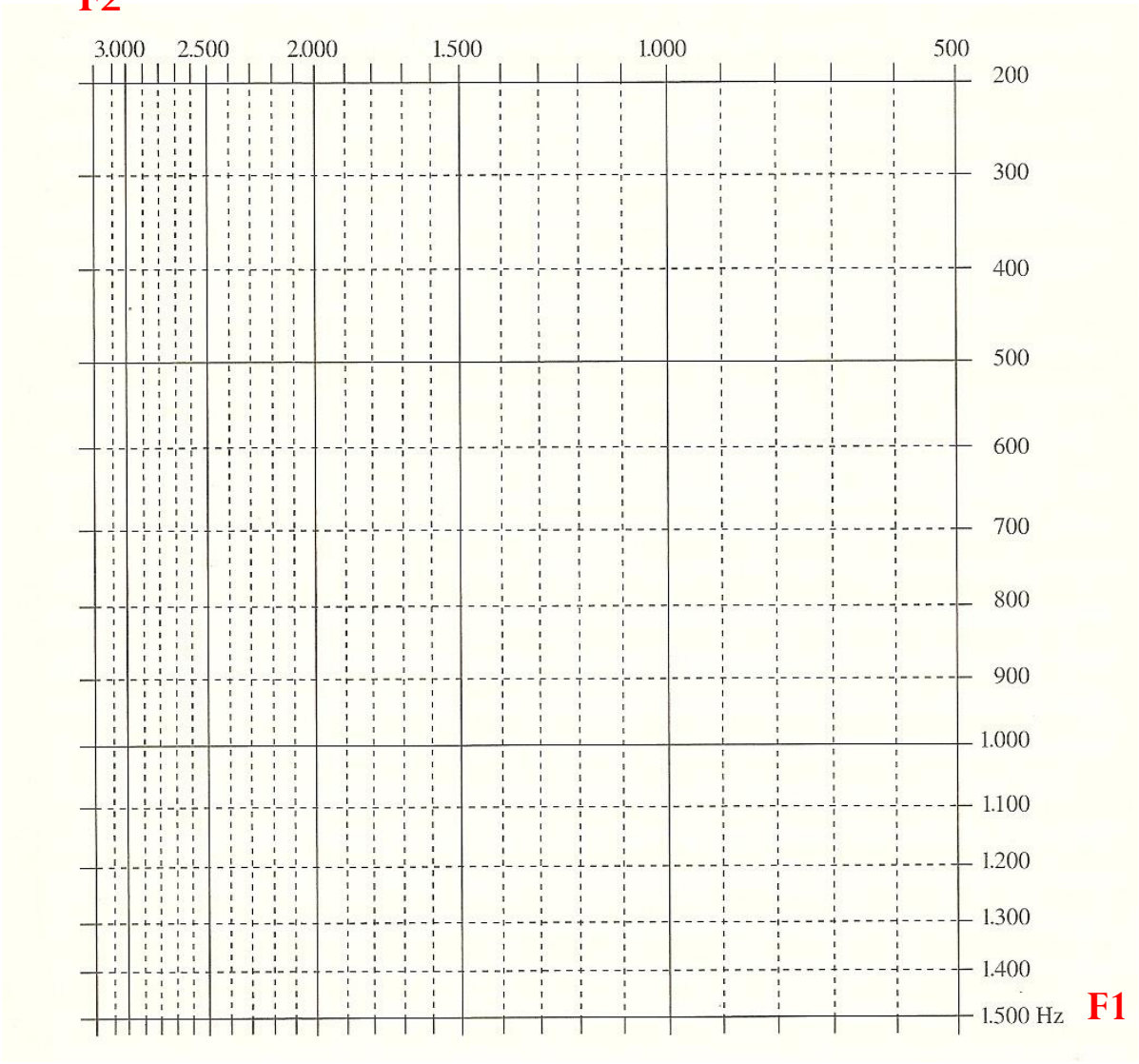




- **Wavesurfer**
  http://www.speech.kth.se/wavesurfer/

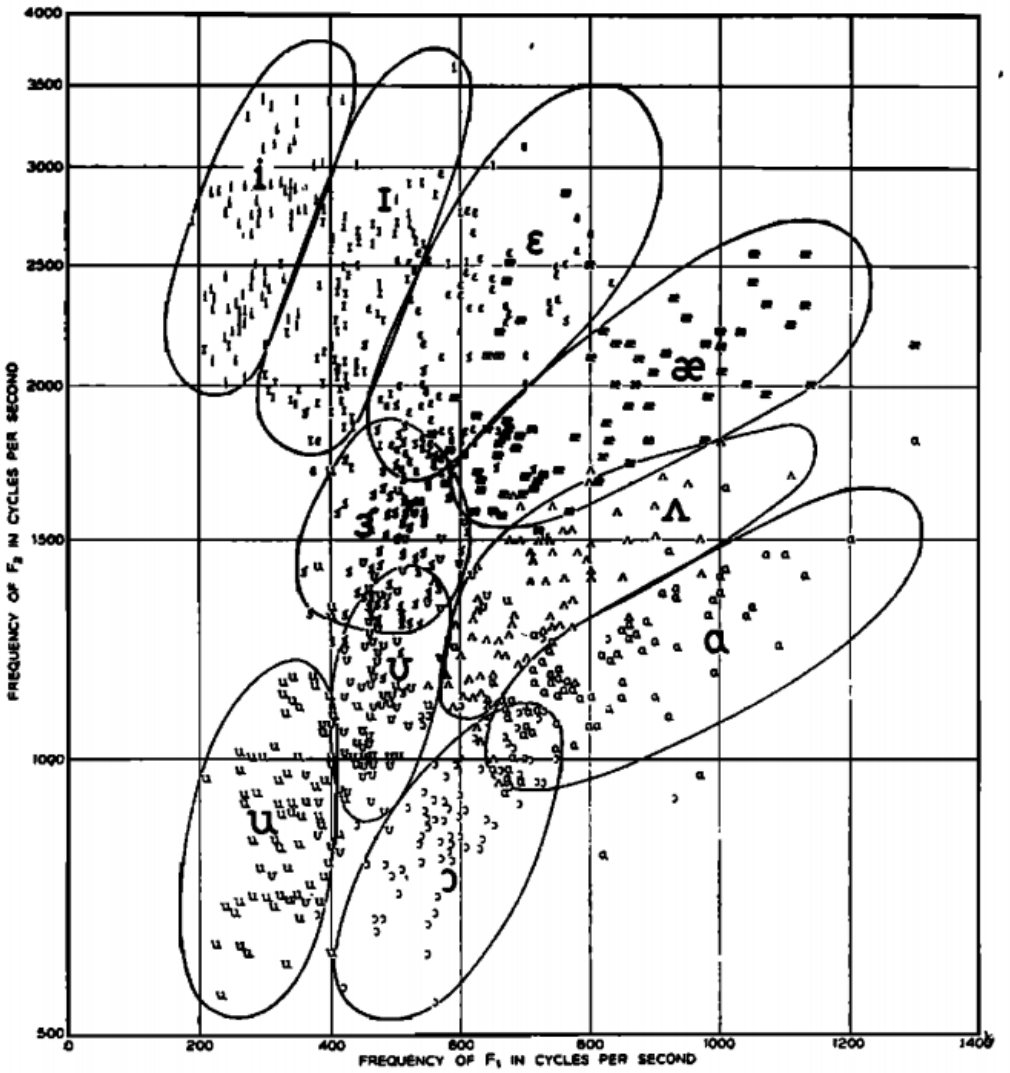
  KTH (Royal Institute of Technology, Stockholm)

Exercise: Make your own F1 by F2 plot



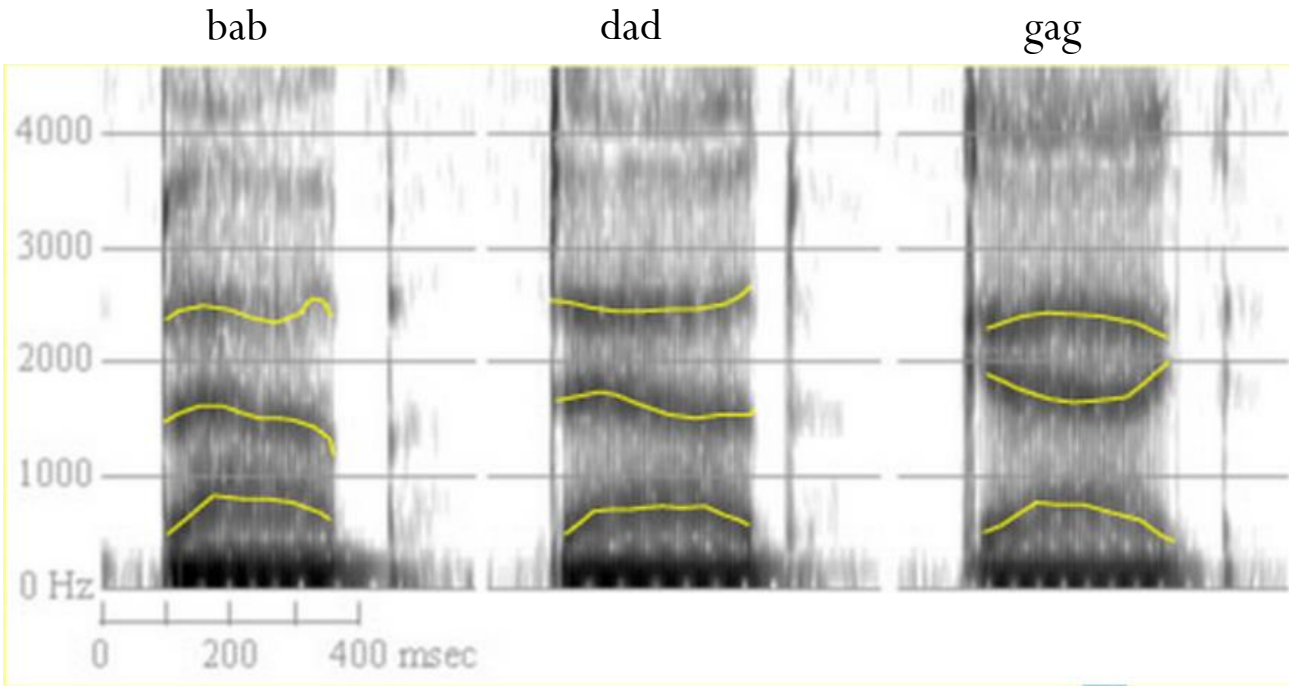
Speaker variation

- Peterson & Barney (1952)
  - 76 speakers
    - 33 men, 28 women, 15 children
  - Variability in vowel production
  - Overlap in formant frequencies



Acoustics of Consonants

- The acoustic structure of consonants is usually more complicated than that of vowels.
- In many cases, there is no distinguishable feature during the consonant articulation itself, e.g. silence part of [p, t, k].
- We have to look for the identity of the consonant at the beginning or the ending of the vowel beside it.



Anticipatory Coarticulation and Loci

- The formants at the moment of consonantal release will vary according to vowel.
- The apparent point of origin of the formant for each place of articulation is called the **locus** of that place of articulation.
- The locus depends on adjacent vowels.

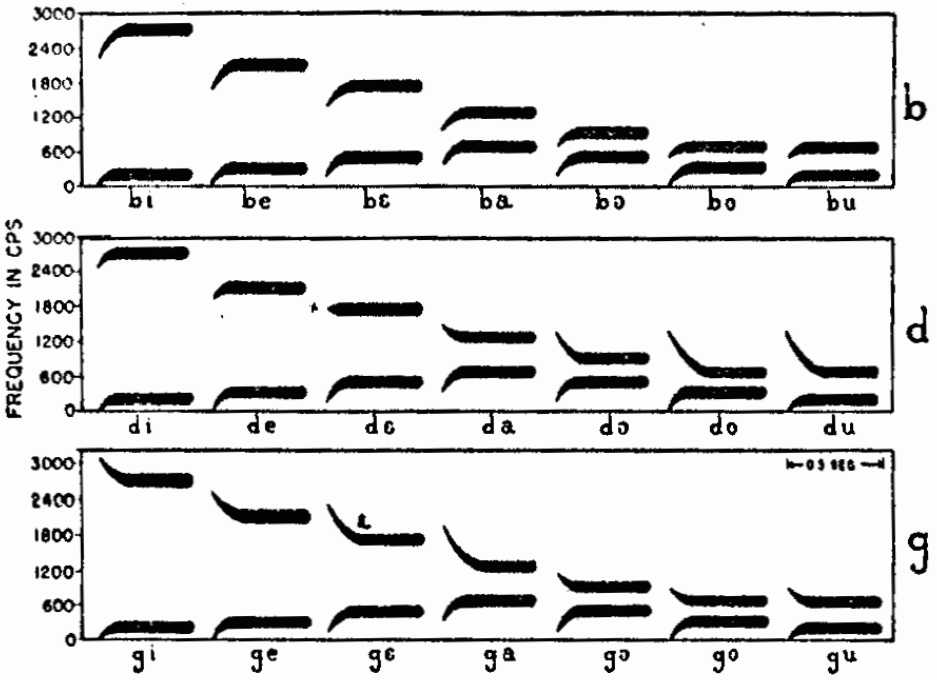
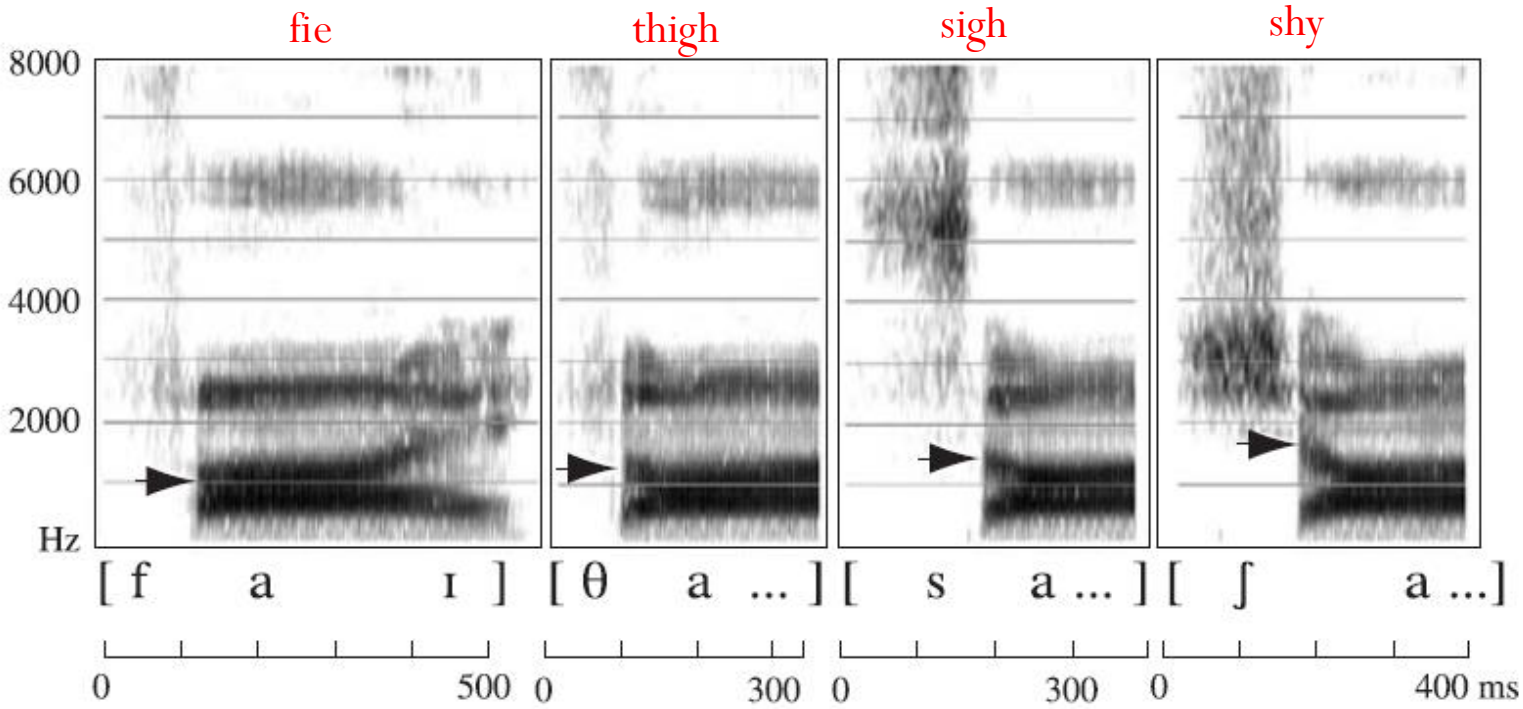


FIG. 1. Synthetic spectrograms showing second-formant transitions that produce the voiced stops before various vowels.

Delattre, 1969

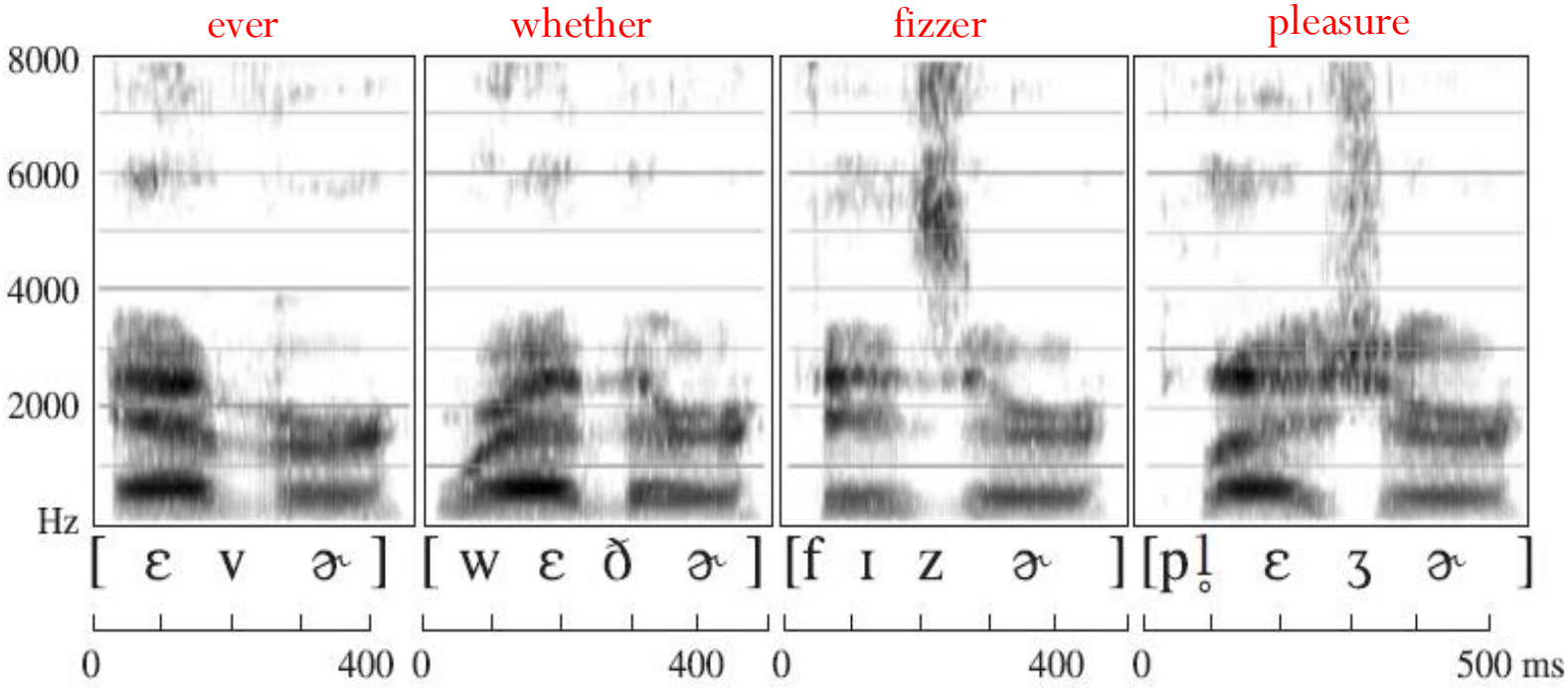
Voiceless fricatives

- Highest frequencies in speech occur over fricatives.
- Frequency scale increased to 8000 Hz.
- Both [s, ʃ] have **larger acoustic energy** and produce **darker patterns** than [f, θ]



Voiced fricatives [v, ?]

- Voiced fricatives [v, ?, z, ?] have patterns similar to their voiceless counterparts [f, ?, s, ?].
- Voiced fricatives also have vertical striations indicative of voicing.



Lateral and central approximants

- A final lateral may have little of no central contact, making it not really a lateral but a **back unrounded vowel**.
- The most obvious feature of approximant [ɹ] is the low frequency of F2 and F3.
- There is great similarity between *red* and *wed*. Young children have difficulty trying to distinguish them.
- The approximant [w] also starts with a low position for all three formants.
- F2 of [w] has the sharpest rise, as if it were a very short [u].
- The movements of formants for [j] are like those of a very short [i].
- This is why [w] and [j] are appropriately called **semivowels**, that is, semi versions of vowels [u] and [i] respectively.

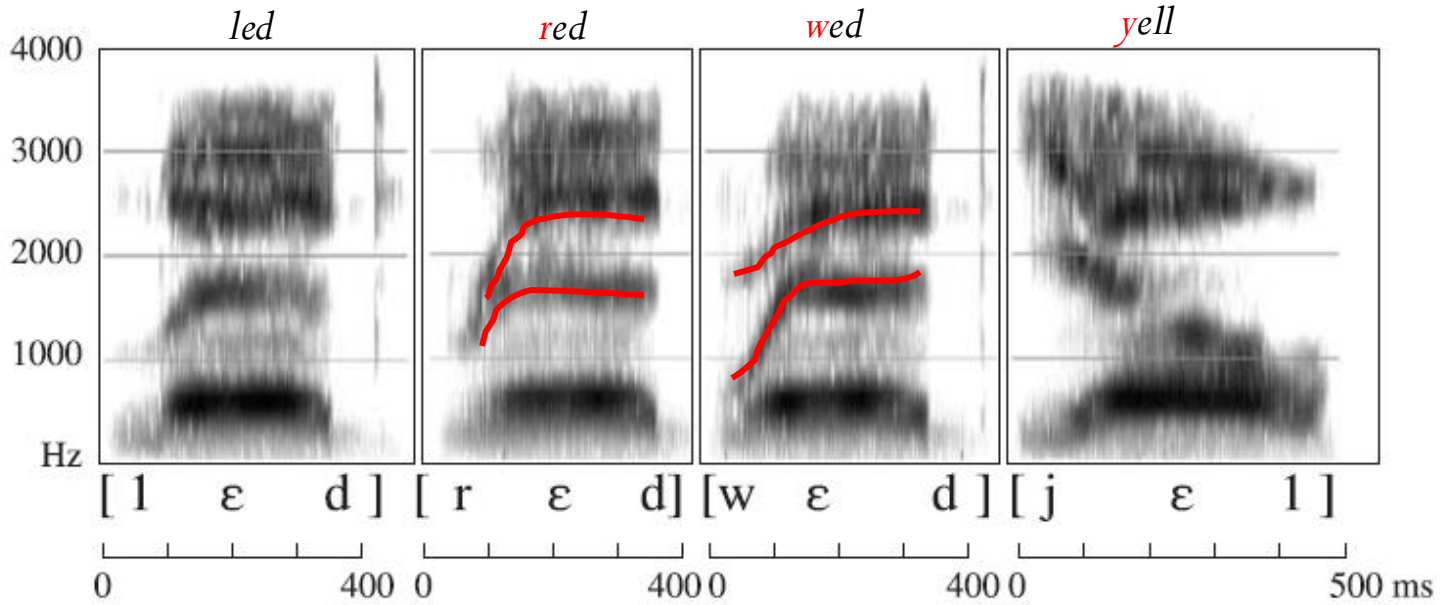


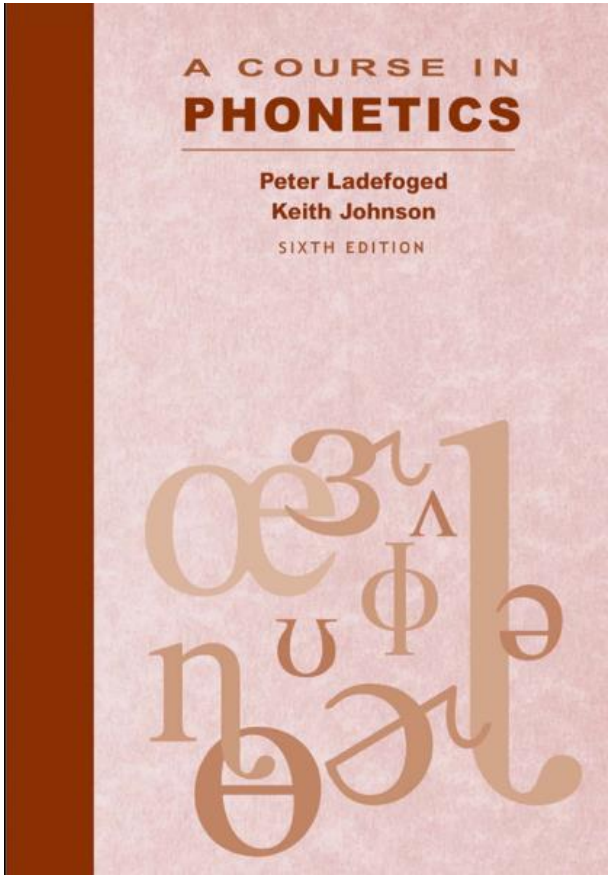
TABLE 8.1 Acoustic correlates of consonantal features. Note: These descriptions should be regarded only as rough guides. The actual acoustic correlates depend to a great extent on the particular combination of articulatory features in a sound and on the neighboring vowels.

Voiced	Vertical striations corresponding to the vibrations of the vocal folds.
Bilabial	Locus of both second and third formants comparatively low.
Alveolar	Locus of second formant about 1700–1800 Hz.
Velar	Usually high locus of the second formant. Common origin of second and third formant transitions.
Retroflex	General lowering of the third and fourth formants.
Stop	Gap in pattern, followed by burst of noise for voiceless stops or sharp beginning of formant structure for voiced stops.
Fricative	Random noise pattern, especially in higher frequency regions, but dependent on the place of articulation.
Nasal	Formant structure similar to that of vowels but with nasal formants at about 250, 2500, and 3250 Hz.
Lateral	Formant structure similar to that of vowels but with formants in the neighborhood of 250, 1200, and 2400 Hz. The higher formants are considerably reduced in intensity.
Approximant	Formant structure similar to that in vowels, usually changing.

Ladefoged & Johnson, 2011:204

Acknowledgements

- Material for this presentation has been adapted mainly from chapters 1 and 8 of
  - Ladefoged, P., & Johnson, K. (2011). *A course in phonetics.* (6th ed.). Canada: Wadsworth, Cengage Learning



Read & visit...

- Ladefoged & Johnson "Articulation & Acoustics", chapters 1 and 8 (A course in phonetics", 6th ed.)
- Visit the websites (for Articulation)
  - https://corpus.linguistics.berkeley.edu/acip/course/chapter1/ (Material from UC Berkeley Linguistics for the book "A course in phonetics")
  - http://soundsofspeech.uiowa.edu/index.html#english (Mobile App: Interactive Phonetic Library for American English)
  - https://www.enl.auth.gr/speakgreek/library.html (Interactive Phonetic Library for Greek)
  - http://smu-facweb.smu.ca/~s0949176/sammy/ (Interactive Sagittal Section)

- Visit the websites (for Acoustics)
  - https://www.compadre.org/books/?ID=46&About=1
    An Interactive eBook on the physics of sound (Indiana University Southeast)
  - http://zonalandeducation.com/mstm/physics/waves/waveAdder/WaveAdder1.html
    Wave Adder
  - http://www.youtube.com/watch?v=Gg4IHbiITd0
    Introduction to spectrogram analysis (FloridaLinguistics.com)
  - https://brucehayes.org/103/SpectrogramReading/index.htm
    Spectrogram reading practice (by Bruce Hayes, UCLA)
  - http://home.cc.umanitoba.ca/~robh/howto.html
    Monthly Mystery Spectrogram Webzone –Rob Hagiwara's professional web-space
  - http://www.acoustics.hut.fi/publications/files/theses/lemmetty_mst/chap4.html
    Problems in Speech Synthesis (Helsinki University of Technology)