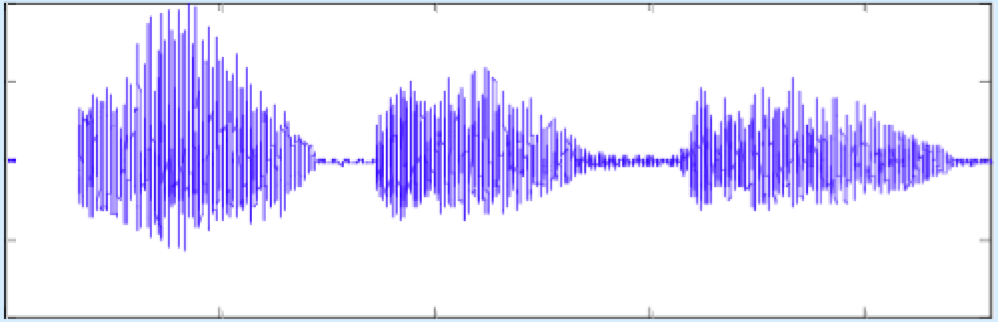
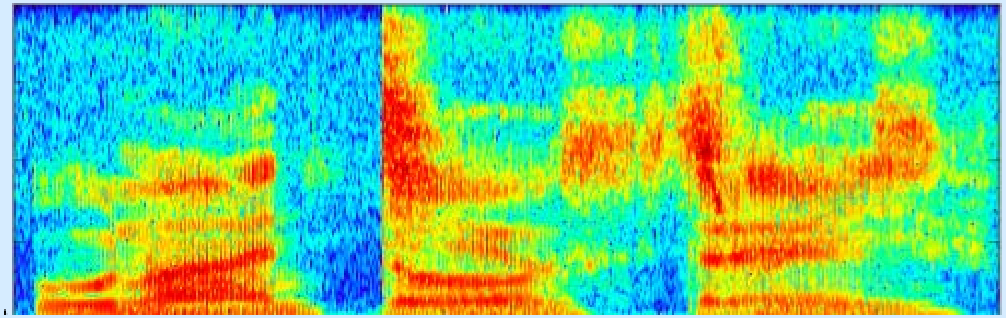
Acoustic Phonetics



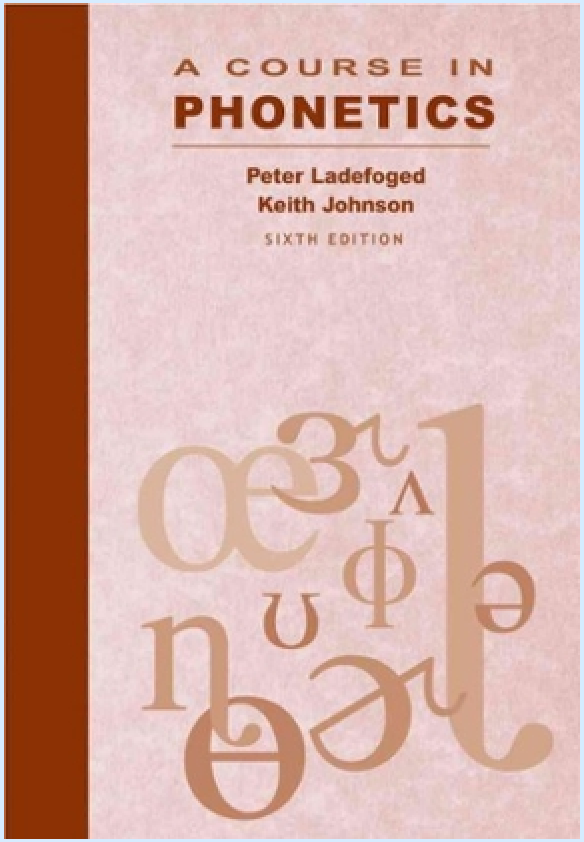


Anna Sfakianaki
Phonetician/Linguist (PhD)
Laboratory Teaching Staff, CSD UoC

CS-578 Digital Speech Signal Processing
Spring Term 2020-21
University of Crete

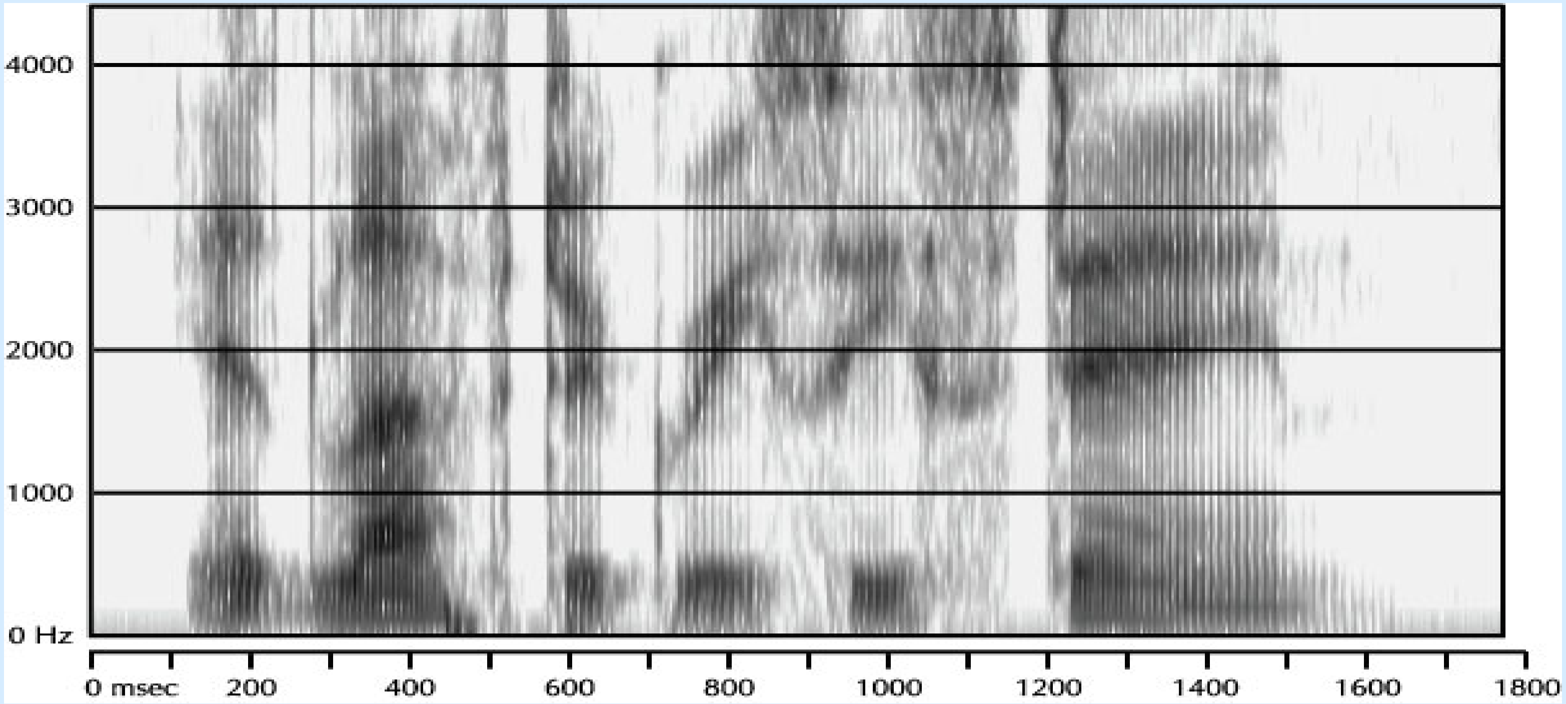
A COURSE IN PHONETICS

- Ladefoged & Johnson (2011)
- Material from this book was used in the slides.



QUESTIONS

- Which are the acoustic properties of speech?
- How do we “read” spectrograms?



https://home.cc.umanitoba.ca/

1. Formants

- Sounds differ from each other in three ways
  - pitch/frequency
  - loudness/intensity
  - 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
- **Overtones = Formants**
- The lowest 3 formants distinguish vowels from each other
  - F1 F2 F3

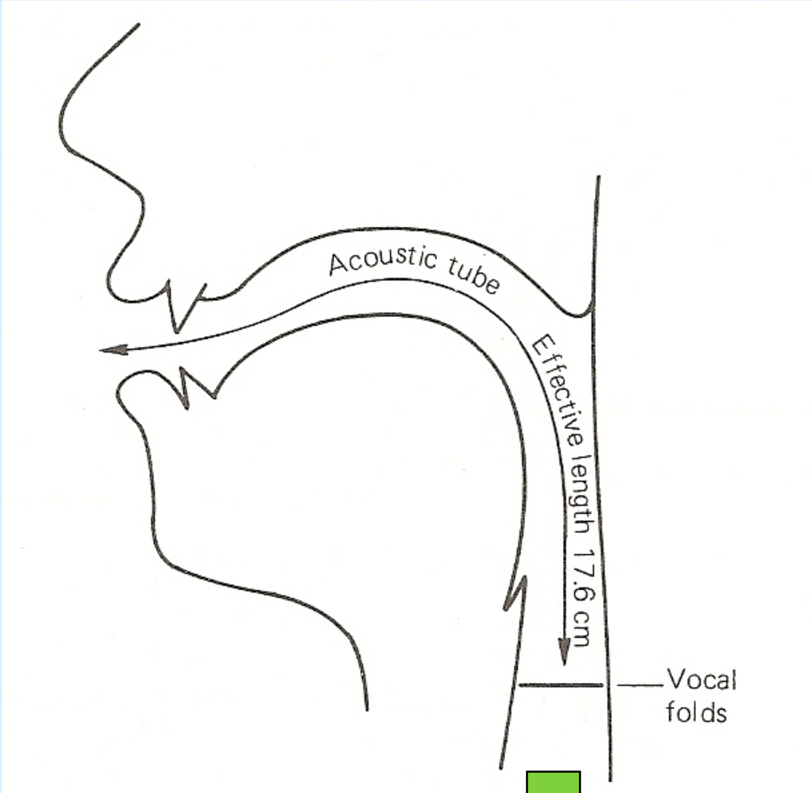
1.1 How do formants arise?

- 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.

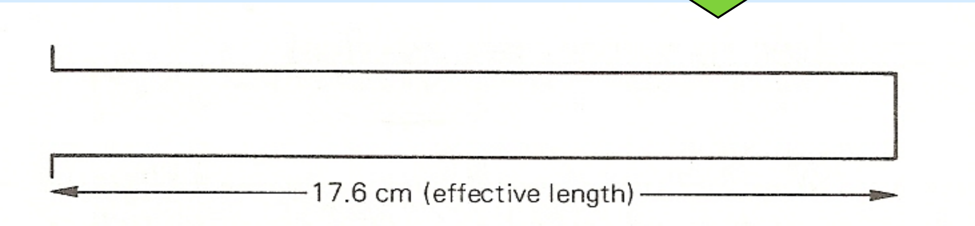
  What do you observe?

  - 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]

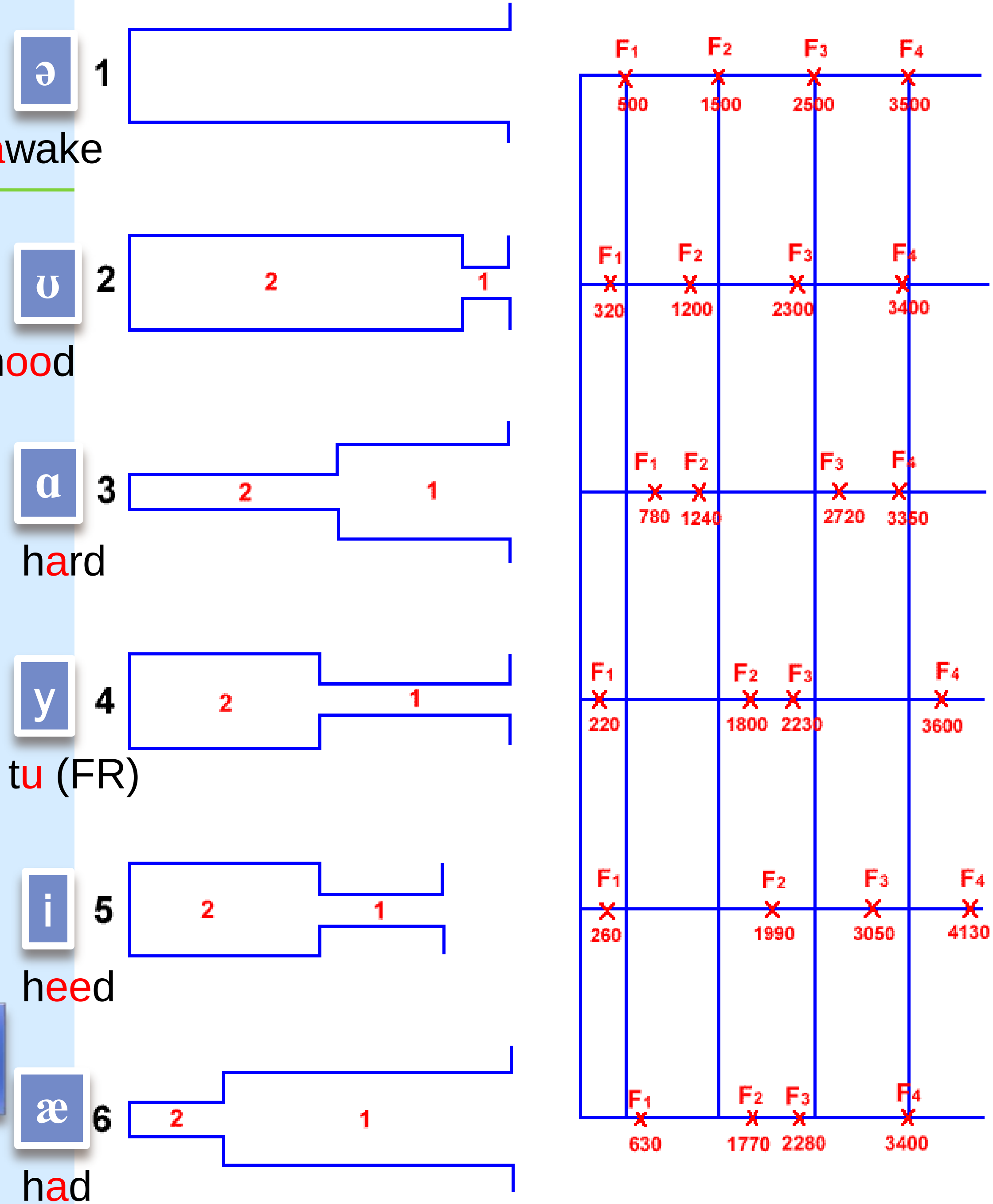


TUBE MODELS

- 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)

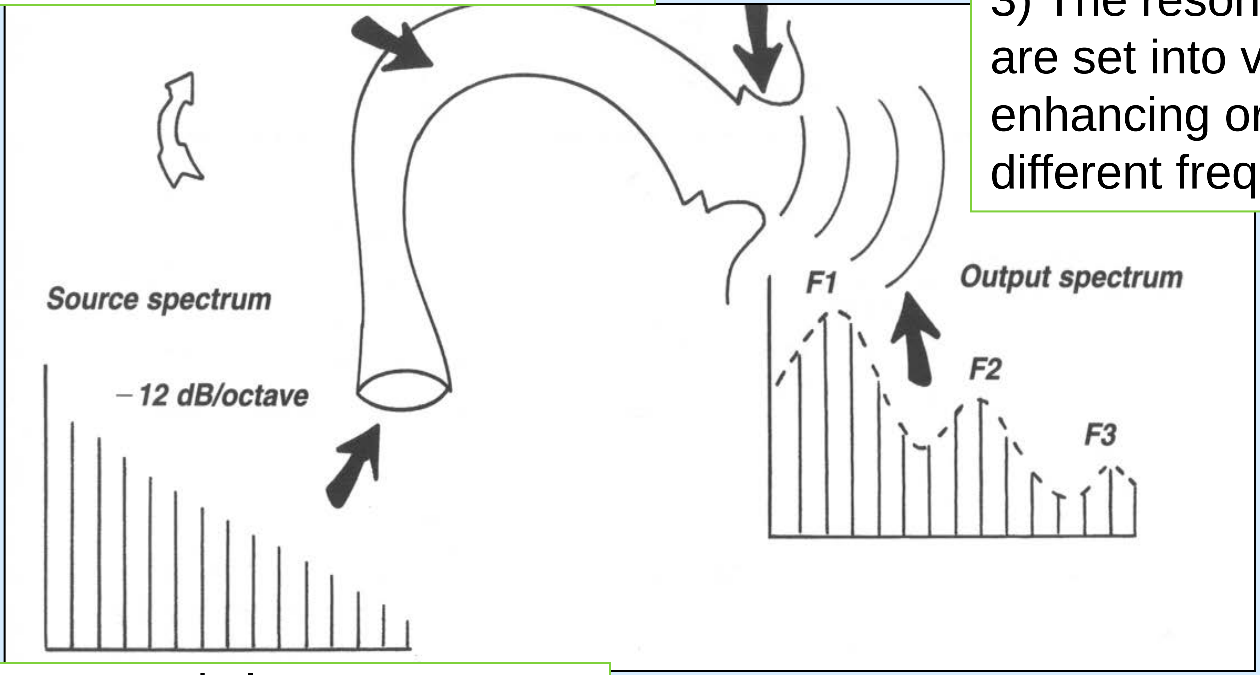
BOTTLE FLUTES – HELMHOLTZ RESONANCE

https://www.youtube.com/watch?v=UeJLoFh47SE

- The vocal tract has a complex shape → contains several bodies of air with different volumes → different overtones

2) These pulses act like sharp taps on the air in the vocal tract.

3) The resonating cavities are set into vibration, enhancing or damping different frequencies.



1) Vocal folds open and close sending out pulses of acoustic energy at different pitches and amplitudes.

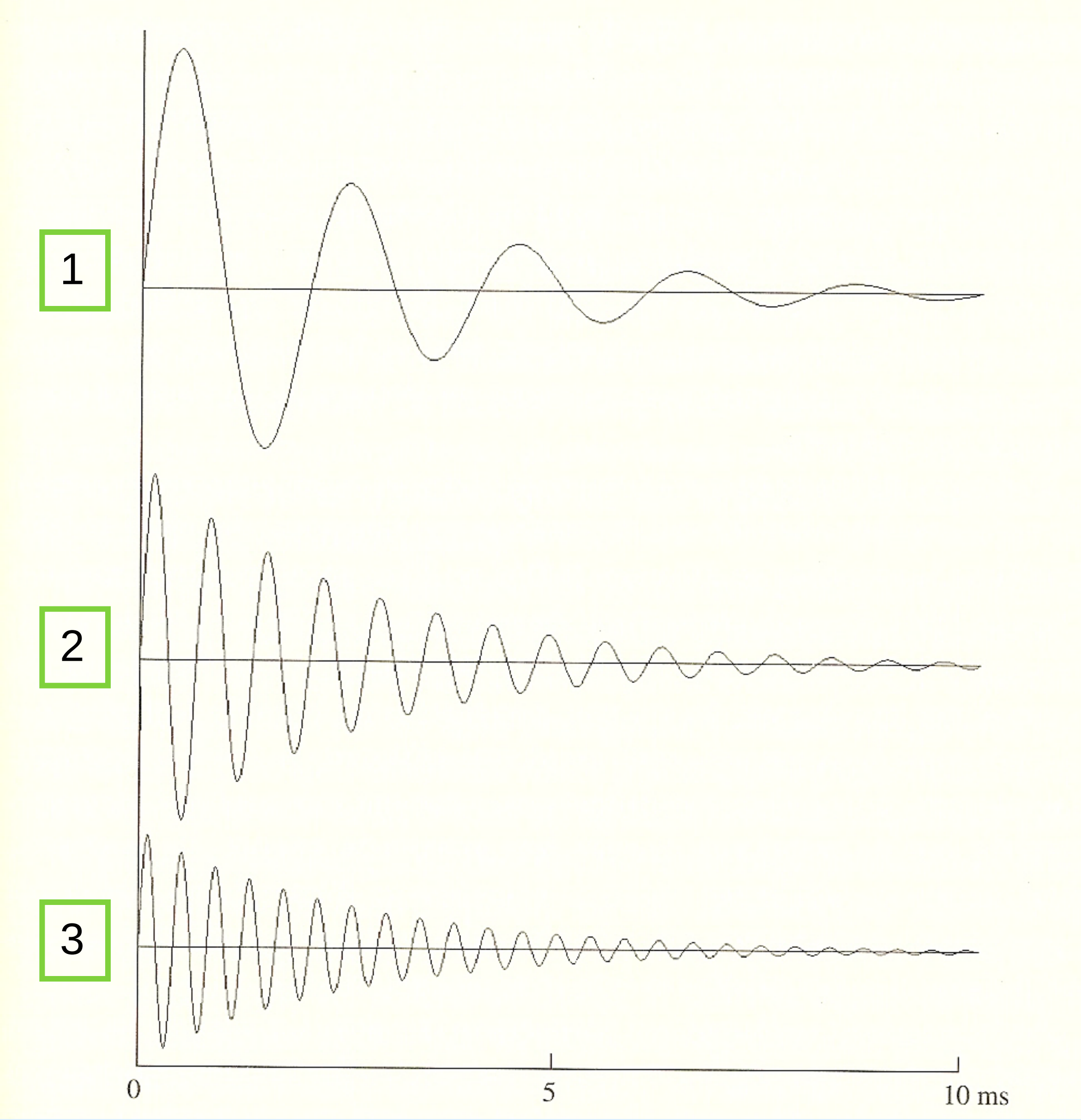
- **One** pulse in the vocal tract may produce **three** different waveforms.

In vowels, we actually hear the sum of these waveforms added together.

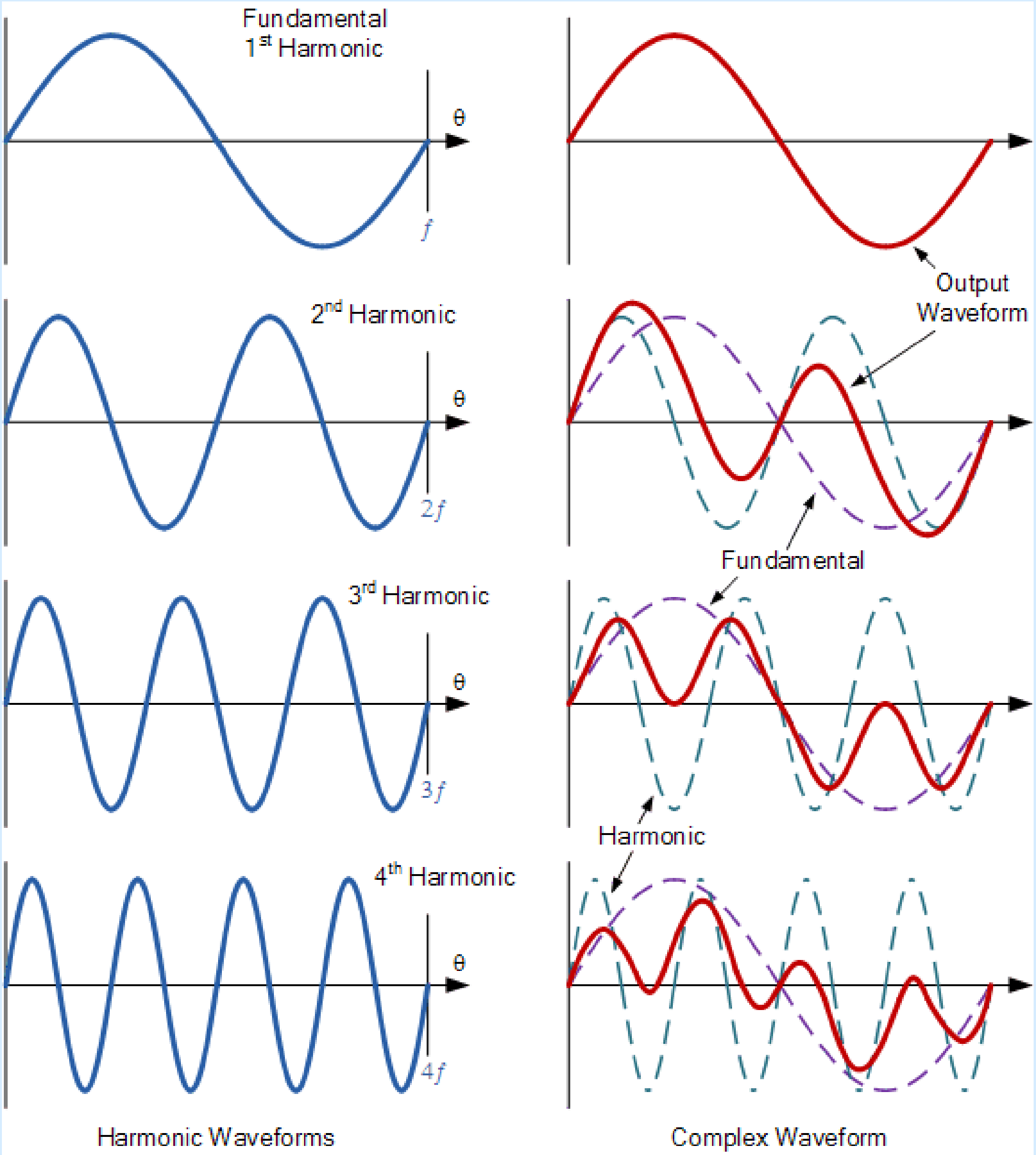
The air in the back of the vocal tract produces a low-frequency waveform.

The air in front of the tongue, a smaller cavity, produces a higher-frequency waveform.

A third mode of vibration of the air in the vocal tract may produce a sound of even higher frequency.



Sum of waveforms



1.2 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.

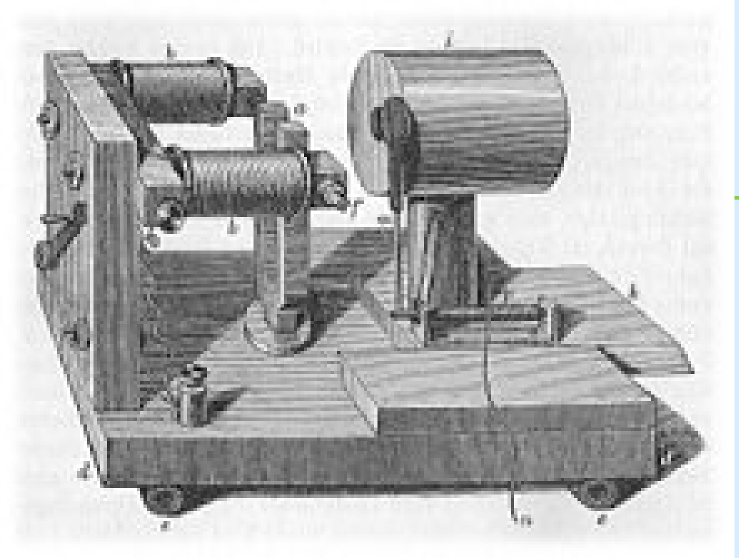
- Male voice: 120 Hz
- Female voice: 220 Hz
- Child voice: 260-280 Hz

- All voiced sounds are distinguishable due to their formants.

SOME HISTORY…

- The general theory of formants was stated by the great German scientist Hermann von Helmholtz (1821-1894) about 150 years ago.
- His scientific work covers the disciplines of Physiology, Psychology (theories of vision), Physics (energy, electrodynamics, thermodynamics), philosophy and aesthetics.
- He was **Hertz**'s supervisor and during his studies in 1879 he suggested that Hertz's doctoral dissertation be on testing Maxwell's theory of electromagnetism, resulting in Hertz's discovery of **electromagnetic waves**.

SOME HISTORY…


Helmholtz Resonator

- Helmholtz invented the **Helmholtz resonator** to identify the various frequencies or pitches of the pure sine wave components of complex sounds containing multiple tones.
- The Helmholtz resonator inspired **Alexander Graham Bell** to invent the **telephone** based on the harmonic telegraph principle.

- *"A given vowel is merely the rapid repetition of its peculiar note"* Robert Willis (English physicist)

↓

- A vowel is the rapid repetition (corresponding to the vibrations of the vocal folds) of its peculiar two or three notes (corresponding to its formants).
- All **voiced sounds** are distinguishable from one another by their formant frequencies.

SPEECH SYNTHESIS DEMO

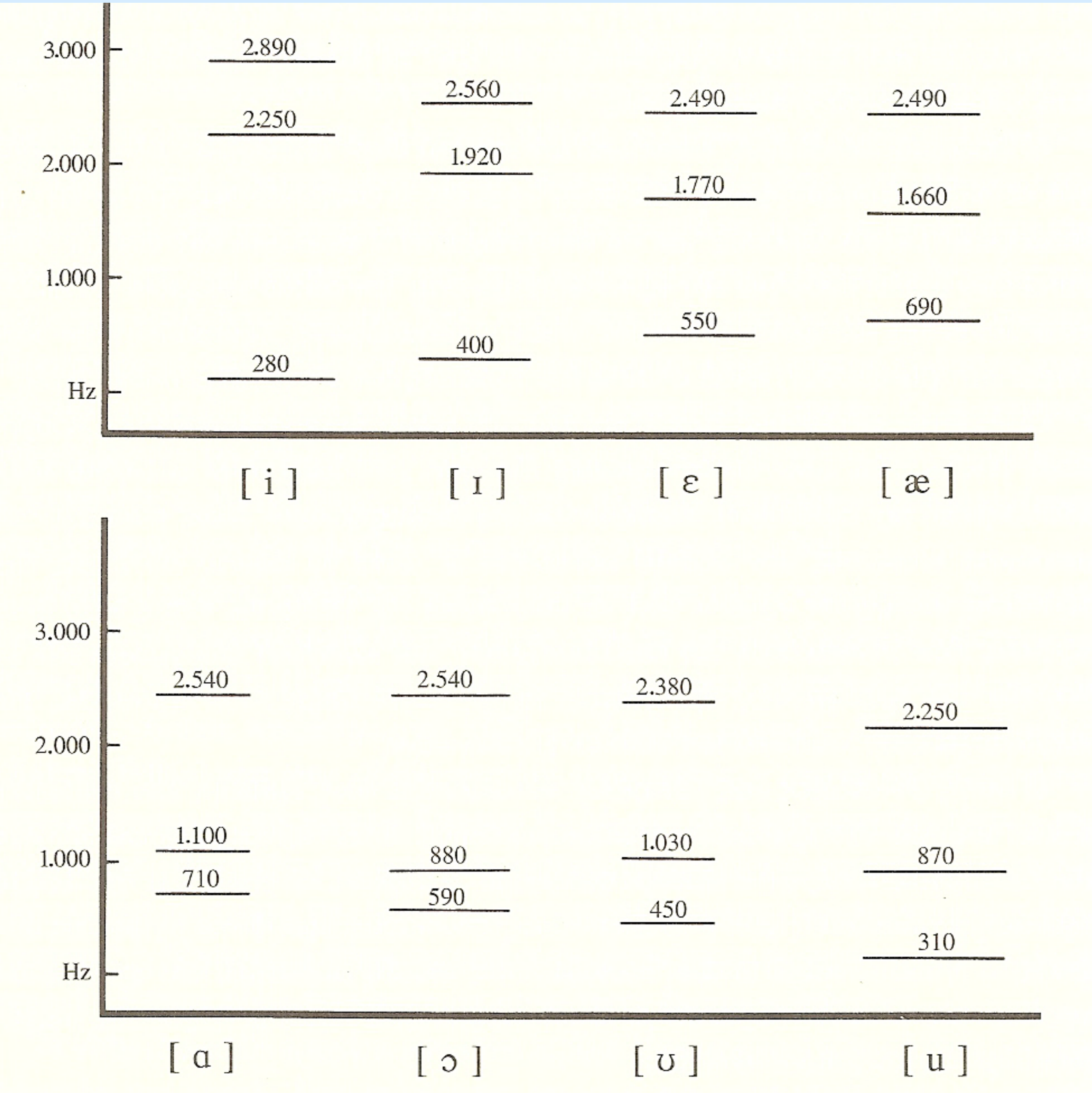
- The notion that vowels contain several different pitches at the same time is difficult to appreciate.
- The demo shows how a sentence is built from its component waves.
- This speech was synthesized in 1971 by Peter Ladefoged on a synthesizer at UCLA.
- **“A bird in the hand is worth two in the bush”**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

2. Acoustic Analysis

- It is possible to analyze sounds so that we can measure the actual frequencies of the formants and represent them graphically.
- Average of **F1**, **F2** and **F3** frequencies in eight American English vowels.

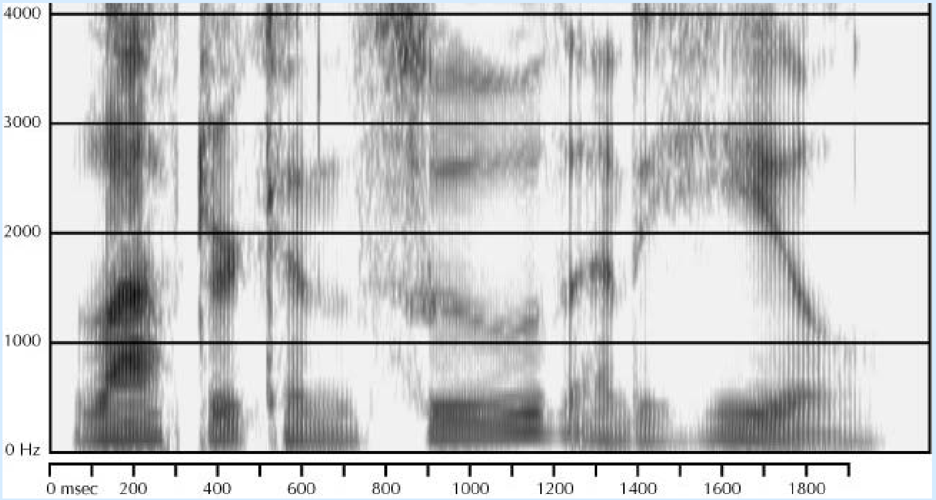
heed, hid, head, had, hod, hawed, hood, who'd

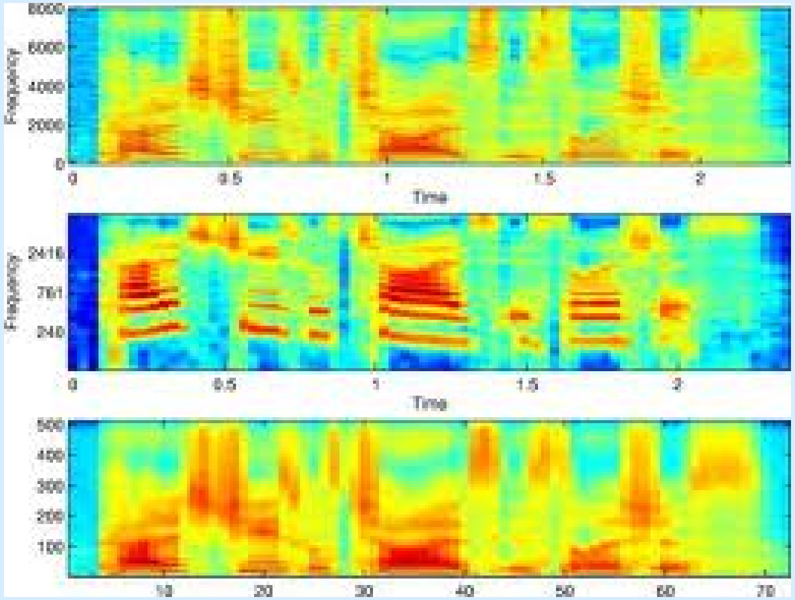
British vowels in hVd words produced by Peter Ladefoged: http://www.phonetics.ucla.edu/course/chapter1.1/chapter1.1.htm#four



2.1 Spectrogram

- Computer programs can analyze sounds and show their components. The display produced is called a spectrogram.
- In spectrograms
  - horizontal axis: time
  - vertical axis: frequency
  - degree of darkness or colour: formants



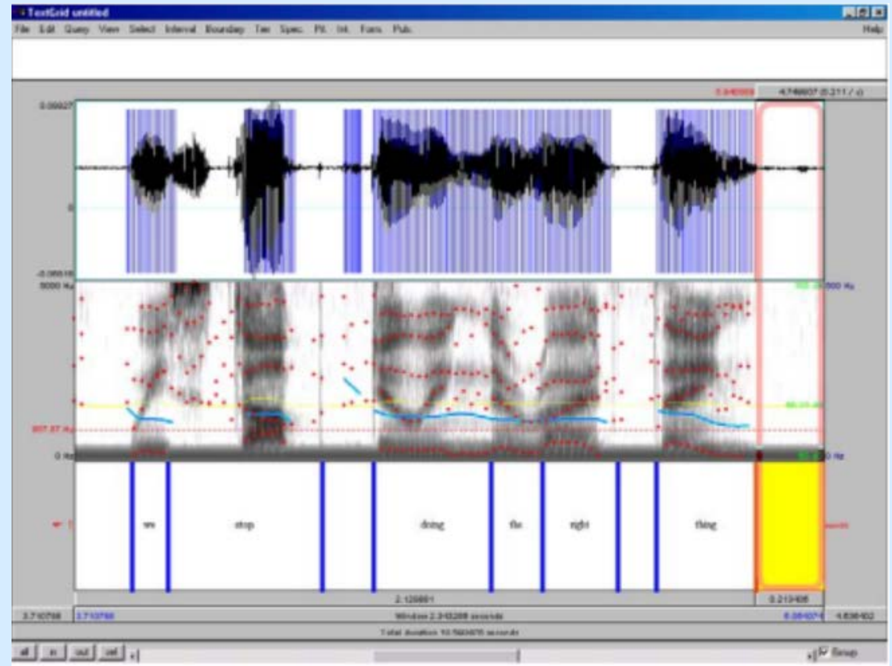


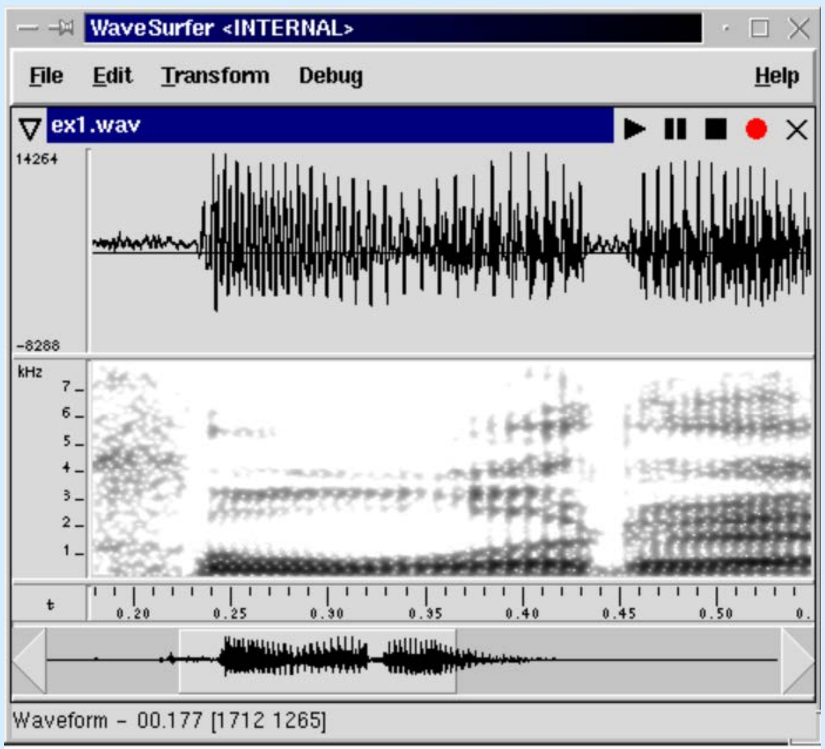
Spectrograms
Dark bands for concentrations of energy at particular frequencies showing the source and filter characteristics of speech

2.2 Computer Programs for acoustic analysis

(free access)

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



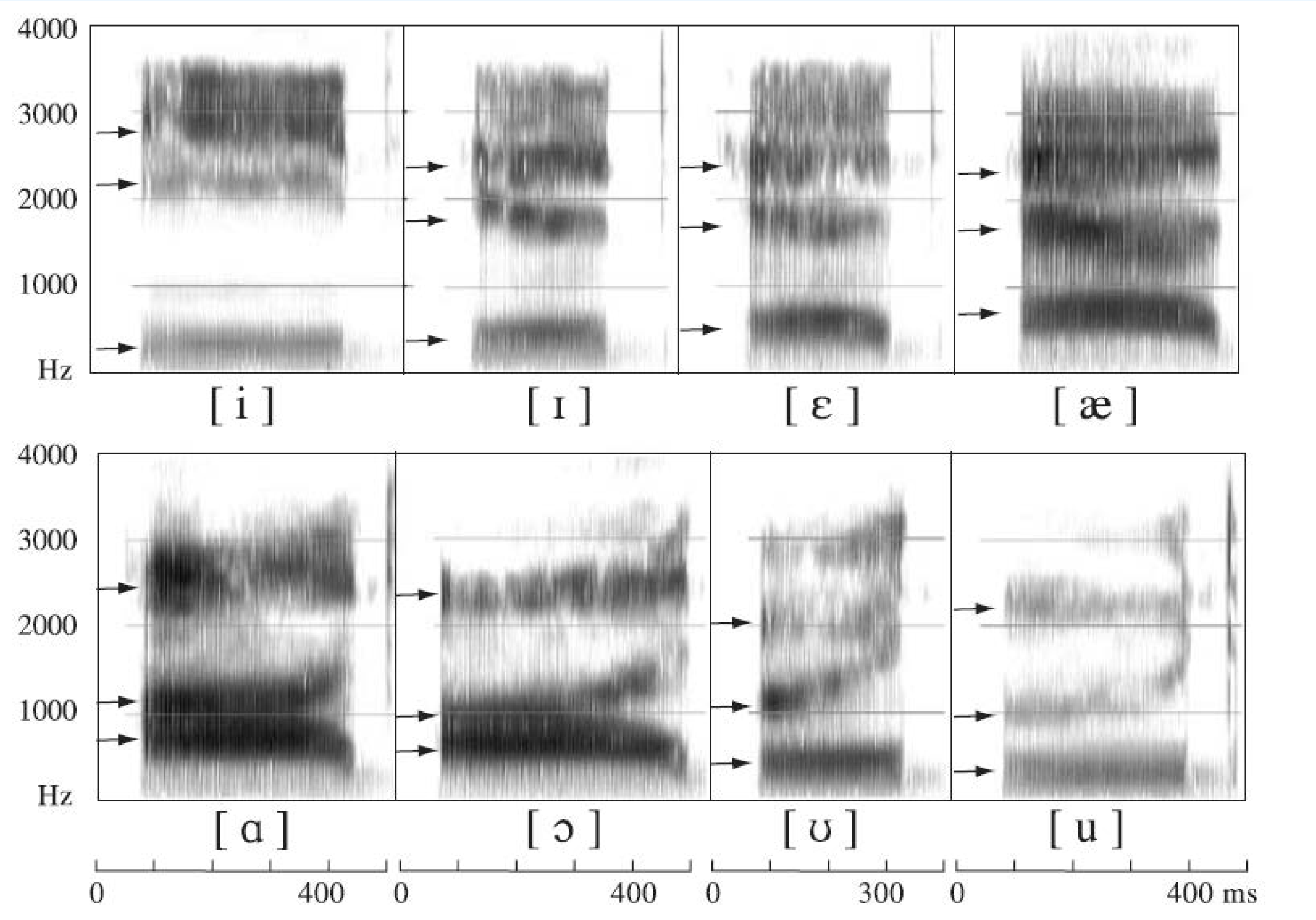


- **Wavesurfer**
  http://www.speech.kth.se/wavesurfer/
  KTH (Royal Institute of Technology, Stockholm)

2.3 Spectrograms of words (American English)

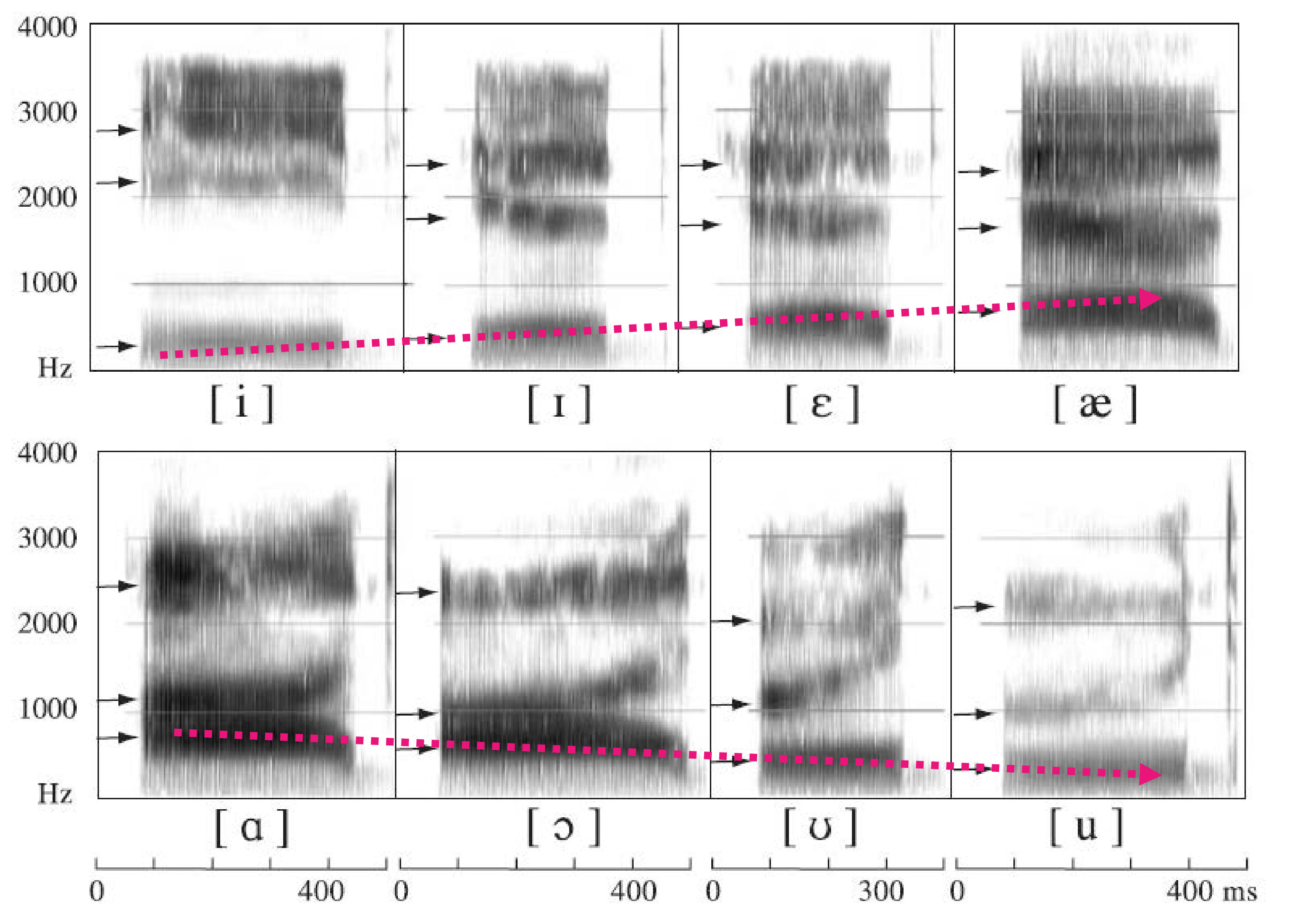
- The vertical scale goes up to 4000 Hz which is sufficient to show the component frequencies of vowels.
- The exact position of the higher formants varies a great deal from speaker to speaker. They are indicative of a person's voice quality.
- Observe the effect of the consonant at the end of the vowel.

heed, hid, head, had, hod, hawed, hood, who'd



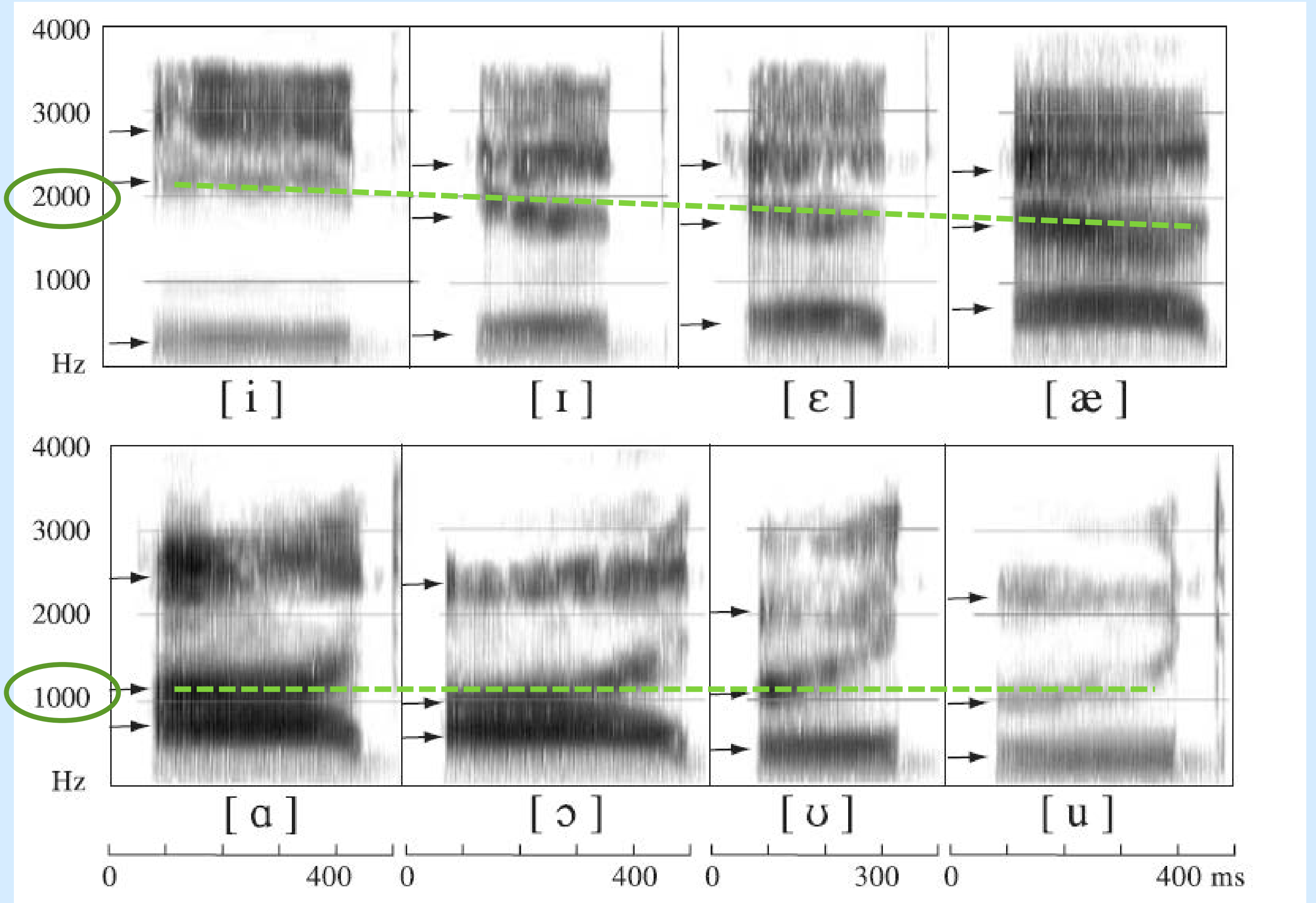
2.4 Formants in relation to traditional articulatory descriptions

- **F1**
  - increases from [i] to [æ] –as vowel height decreases.
  - decreases from [ɑ] to [u] –as vowel height increases.
- Hence F1 is **inversely related** to vowel height.



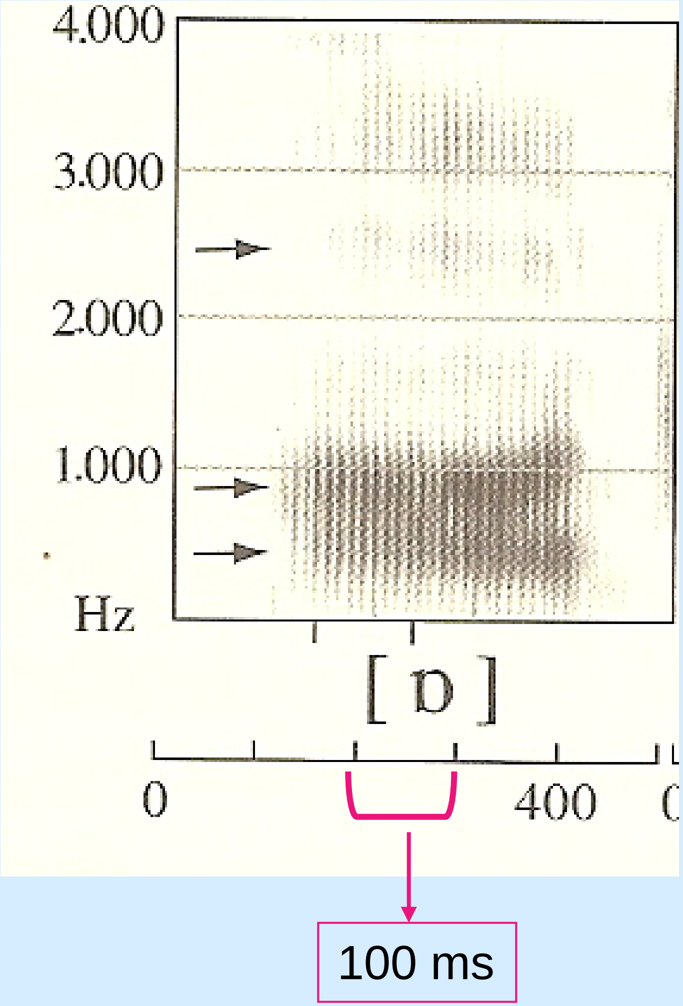
F2

- higher for front vowels
- lower for back vowels
- affected by lip rounding→ decrease of F2 & F3



ESTIMATING F0 FROM SPECTROGRAM

- Regularly spaced vertical lines → vocal fold vibration
- Momentary increase of acoustic energy due to vocal fold single movement → vertical lines
- Calculating pitch from lines in spectrogram
  - sparse → low pitch
  - dense → high pitch
  - here: 8-9 lines per 100 ms = 85 Hz



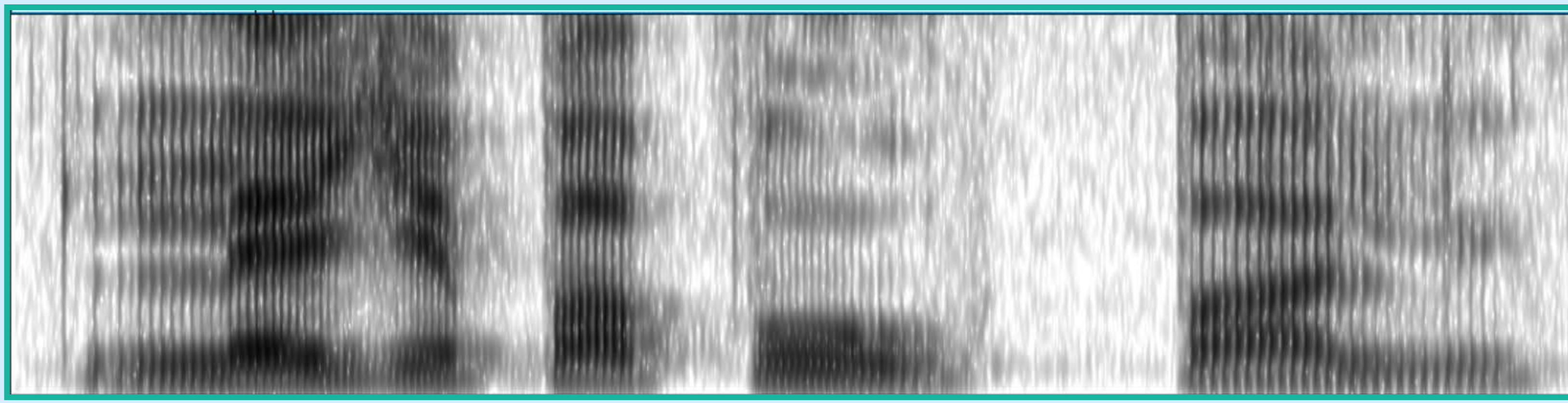
FEMALE VS MALE PITCH

- Women's voices usually have a higher pitch.
- The higher the F0 the more difficult it is to locate formants, because the harmonics interfere with the display of formants.

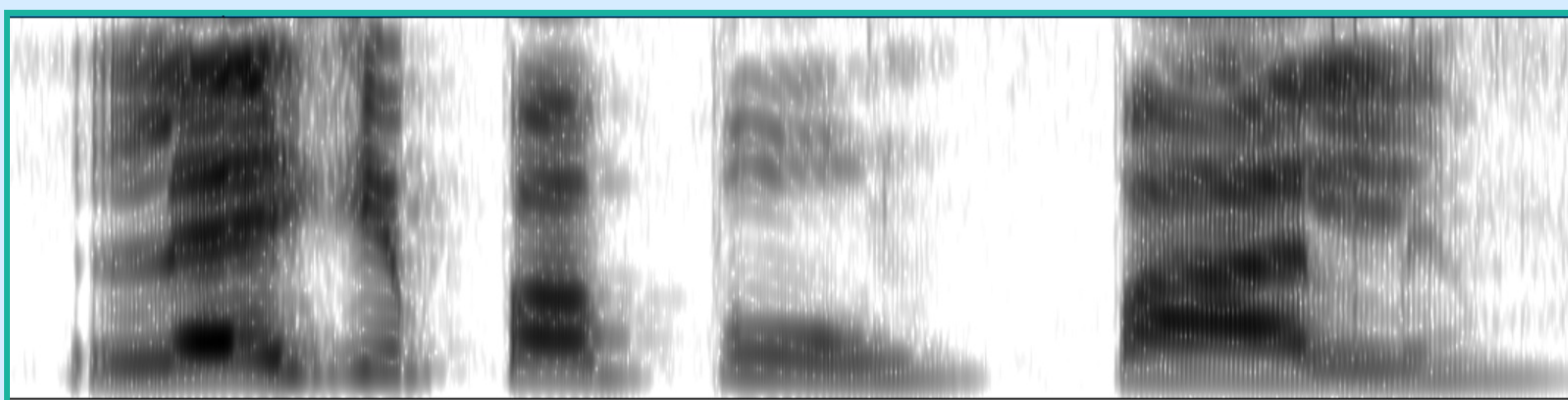
Greek phrase uttered by a male and a female Greek adult.

Λέγε «παππού» πάλι. (*Say "grandfather" again)*

male

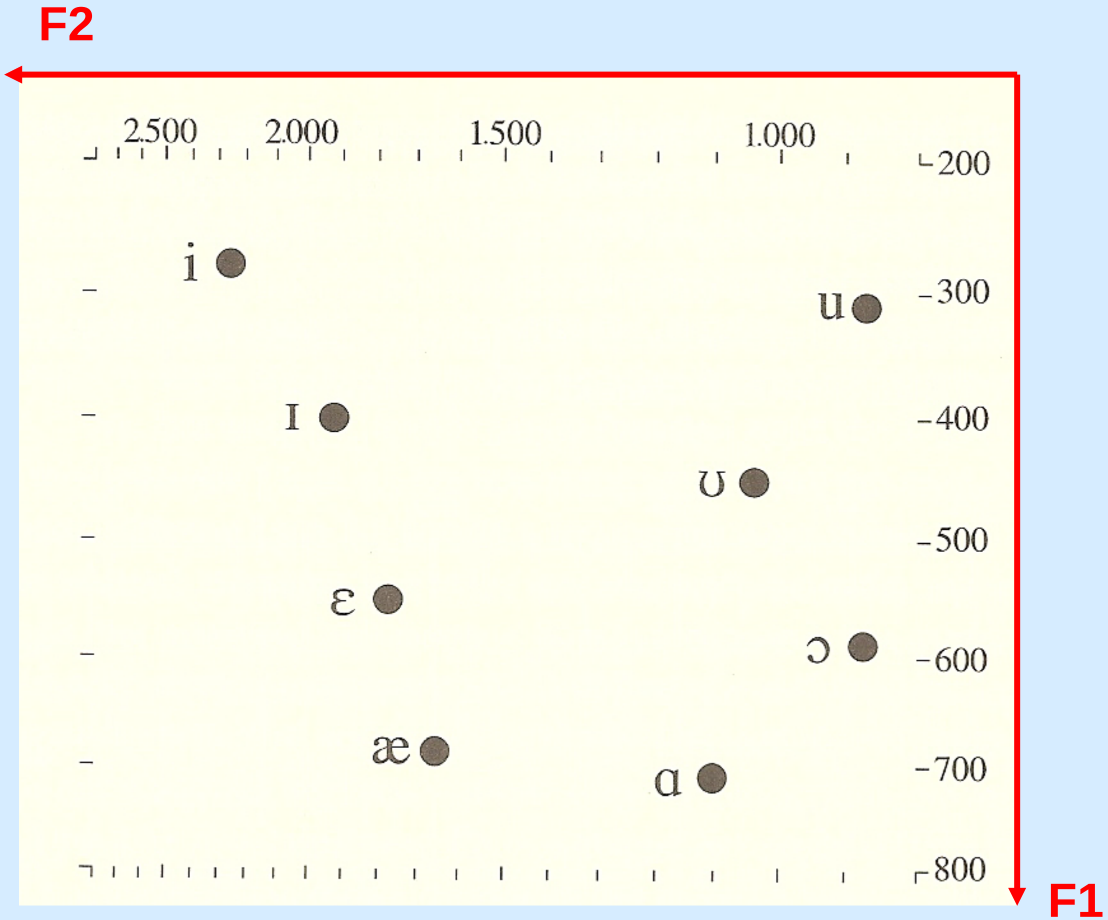


female



2.5 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).



TRADITIONAL VOWEL CHART

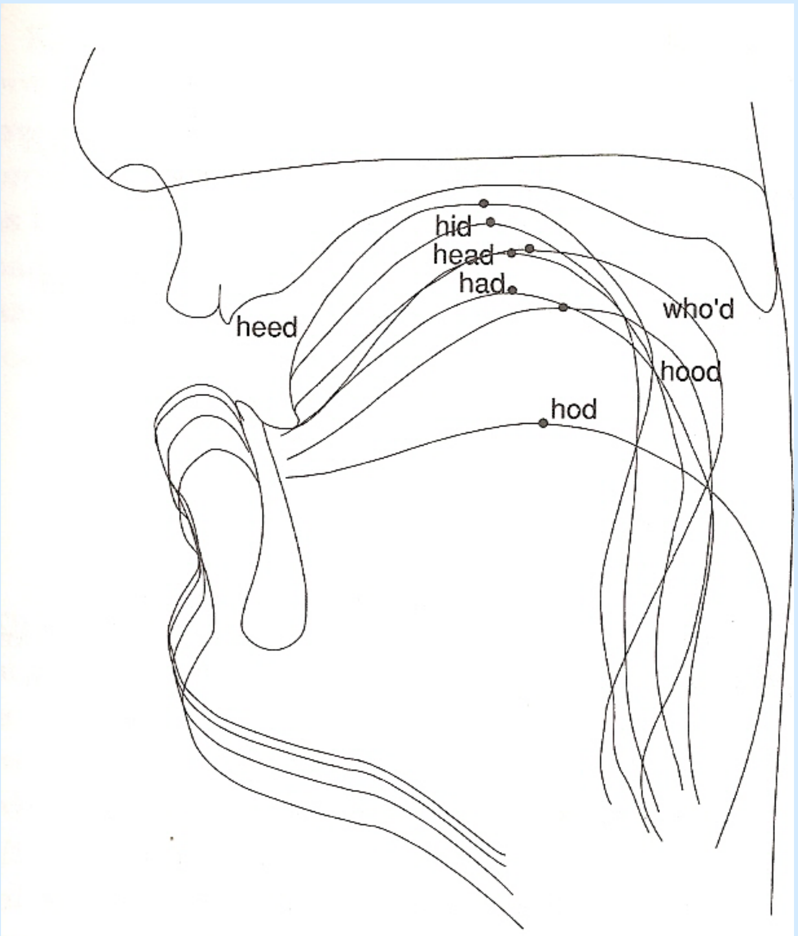
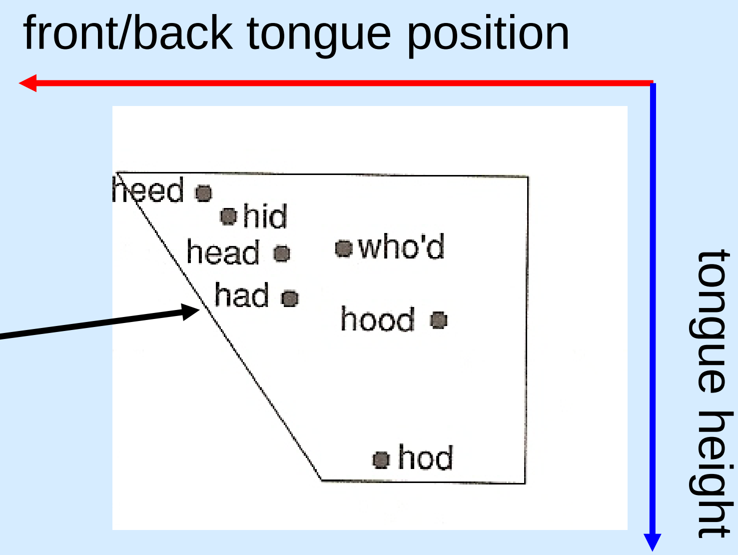
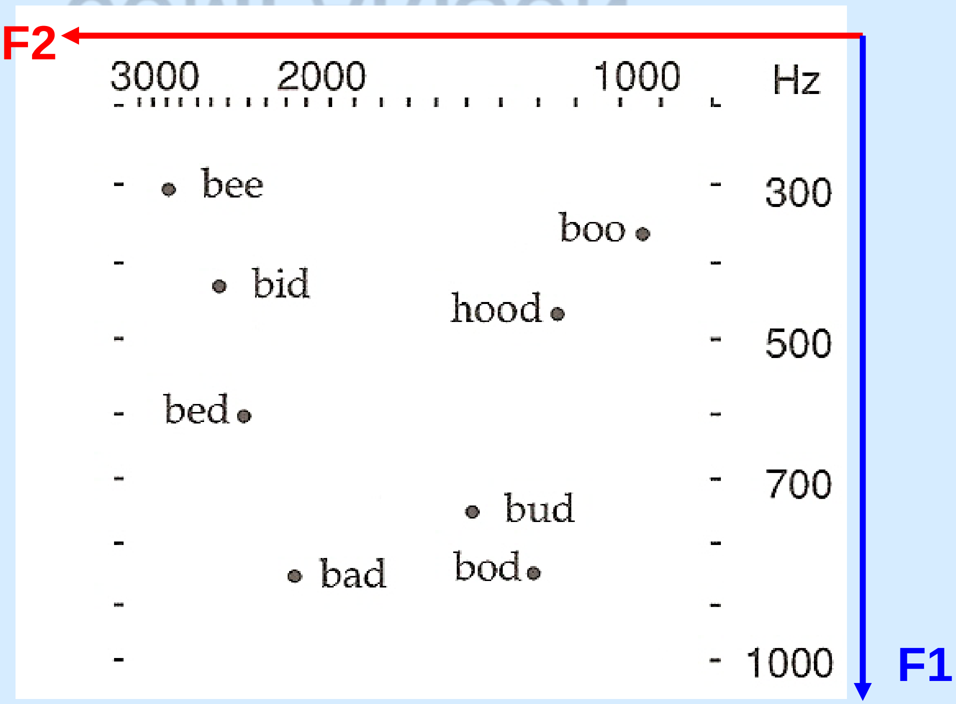


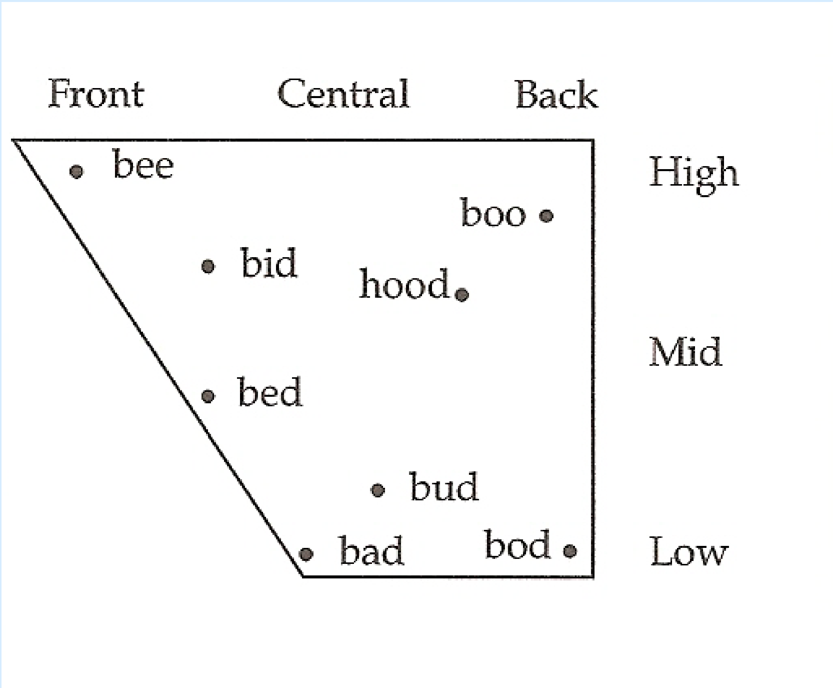
Chart based on X-ray data



Ladefoged, *Vowels & Consonants*, (2001:115)

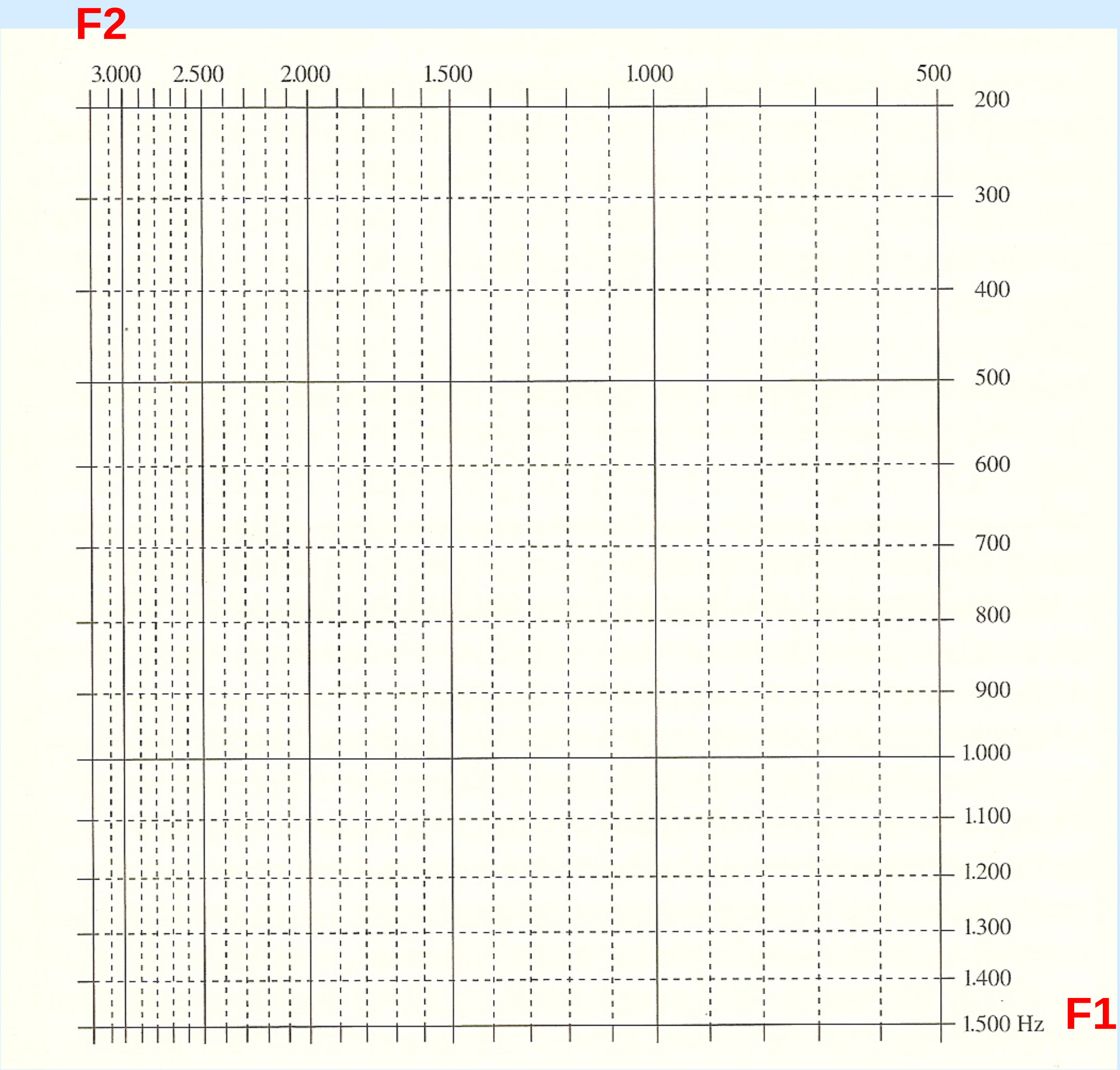
COMPARISON





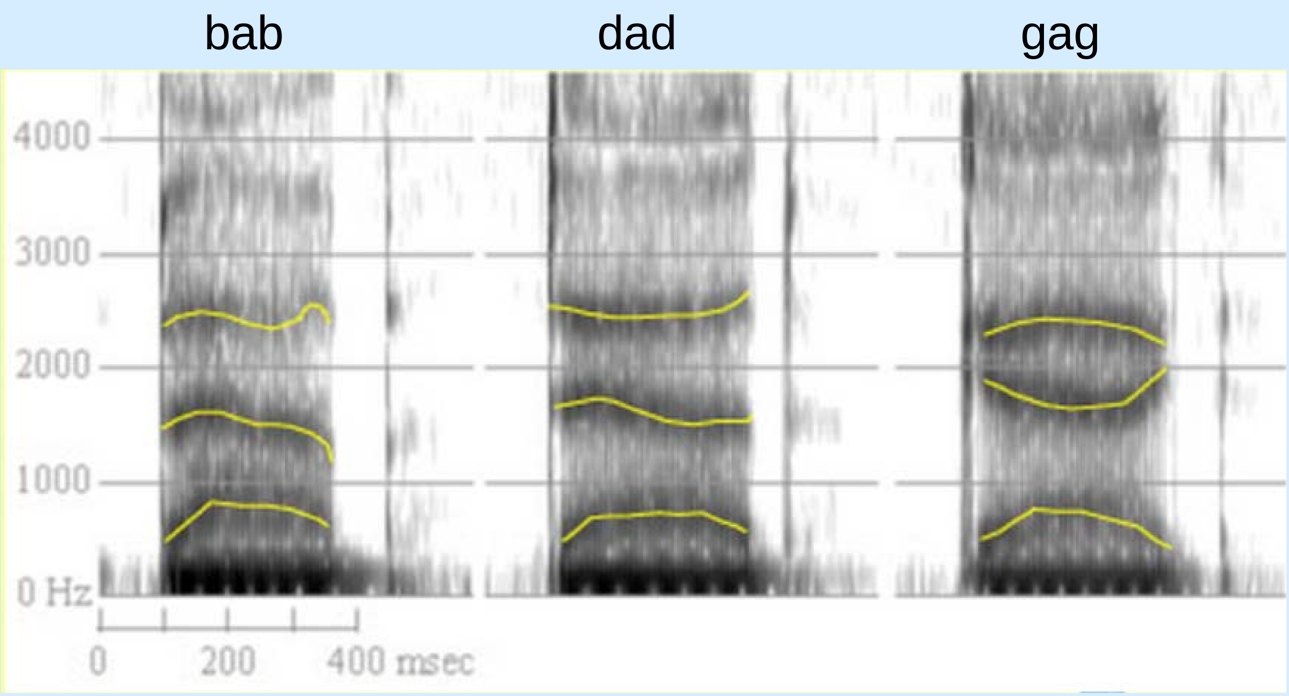
- *"Traditional vowel diagrams express acoustic facts in terms of physiological fantasies."* Oscar Russell (1930s)
- **Vowel height ⇔ F1, not actually tongue height**
- **Front – back dimension** { backness + lip rounding
- Degree of backness ⇔ F1-F2 difference
- The closer together F1 and F2, the more "back" a vowel sounds.

Exercise: Make your own F1 by F2 plot



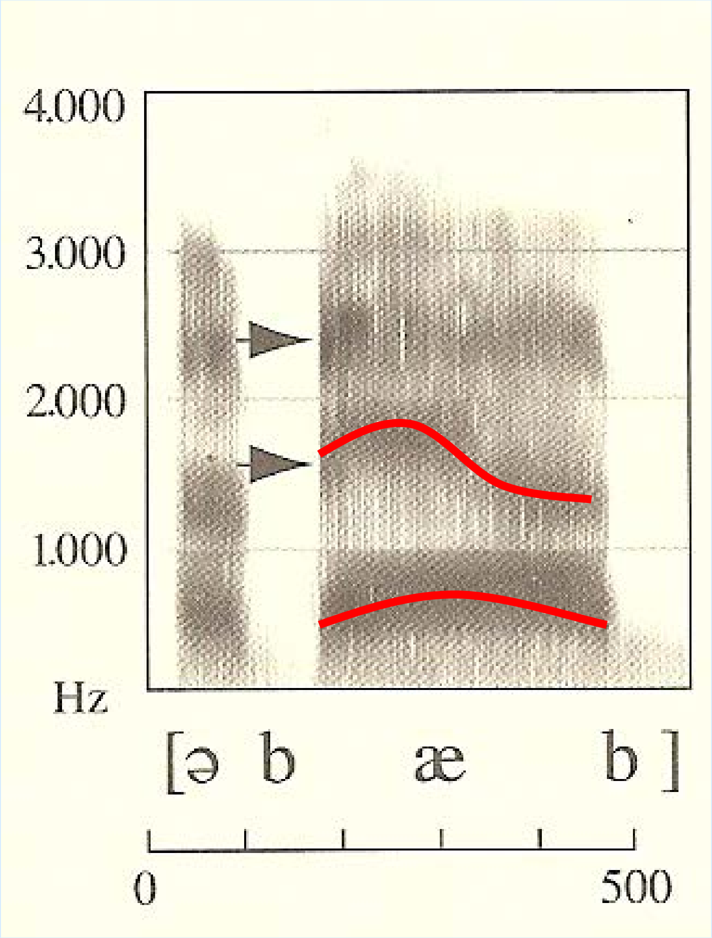
3. Acoustics of Consonants

- 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.



3.1 Stops

- Each of the stop sounds conveys its quality by its effect on the adjacent vowel.
- The formants of [æ] correspond to the particular shape of the vocal tract.
- During the production of [bæ] the formants correspond to the particular shape that occurs the moment the lips come apart.
- Closure of the lips causes a lowering of all formants.
- The syllable [bæb] will begin with formants in a lower position, then they will rapidly rise to the positions of [æ], and finally descend again as the lip closure is formed.

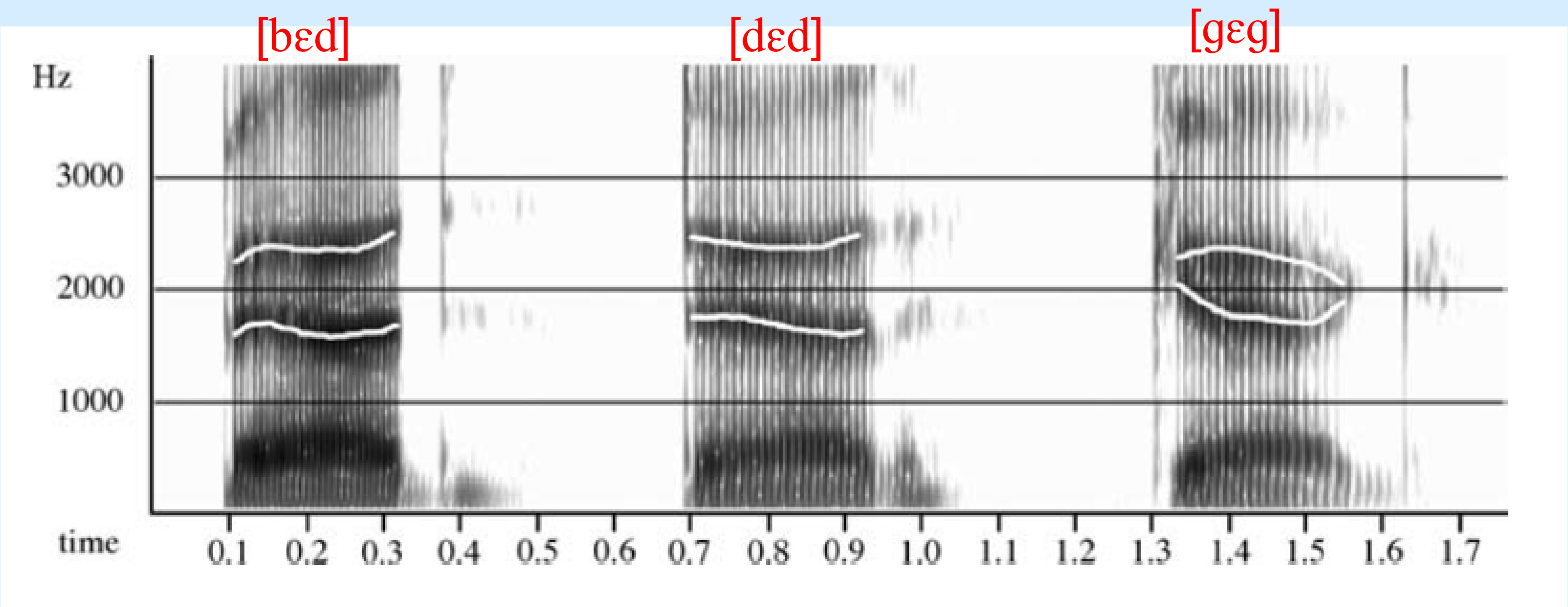


Anticipatory Coarticulation

- For the production of e.g. [bib] or [bab], the tongue will be in position for the vowel even when the lips are closed at the beginning of the word.
- This happens because the part of the tongue not involved in the formation of the consonant closure is already in position for the following vowel.
- 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.

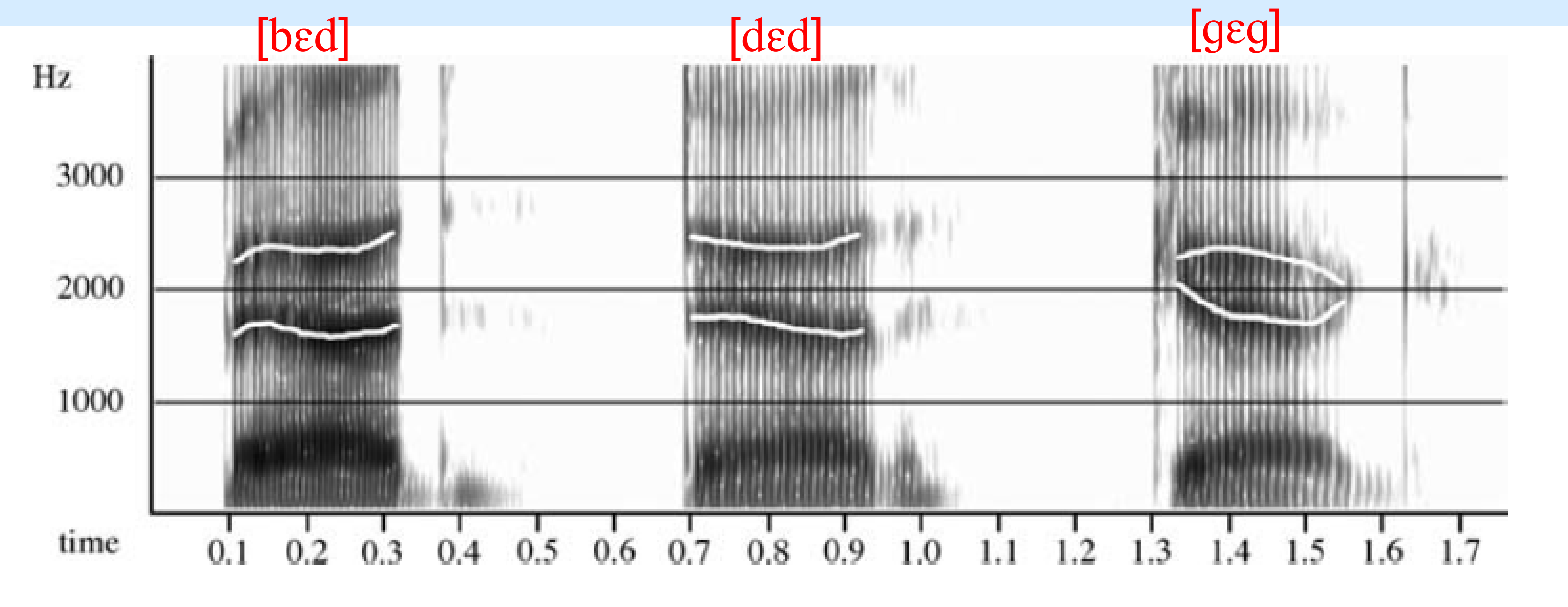
3.2 Formant transitions

- Faint voicing striations near the baseline for each of the stops [b, d, g] (voice bar).
- In all three words, **F1 rises from a low position** due to consonant closure, hence it does not distinguish one place of articulation from another.
- What distinguishes the three stops are the **onsets and offsets of F2 and F3**.



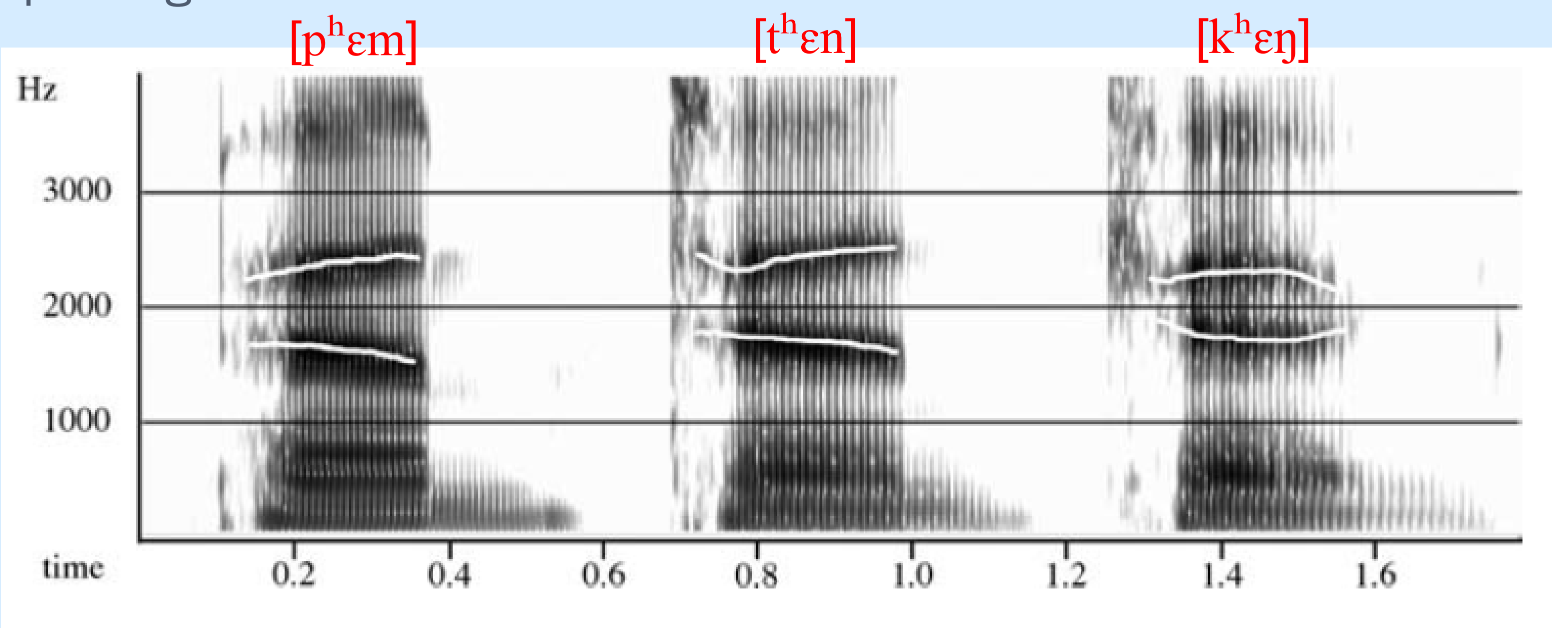
3.2 Formant transitions

- [bɛd]
  - F2 & F3 start at a lower frequency than in [dɛd].
  - F2 & F3 are noticeably rising from a low locus.
- [dɛd]
  - F2 is fairly steady at the beginning.
  - F3 drops a little.
- [gɛg]
  - Characteristic coming together of F2 & F3 → velar pinch



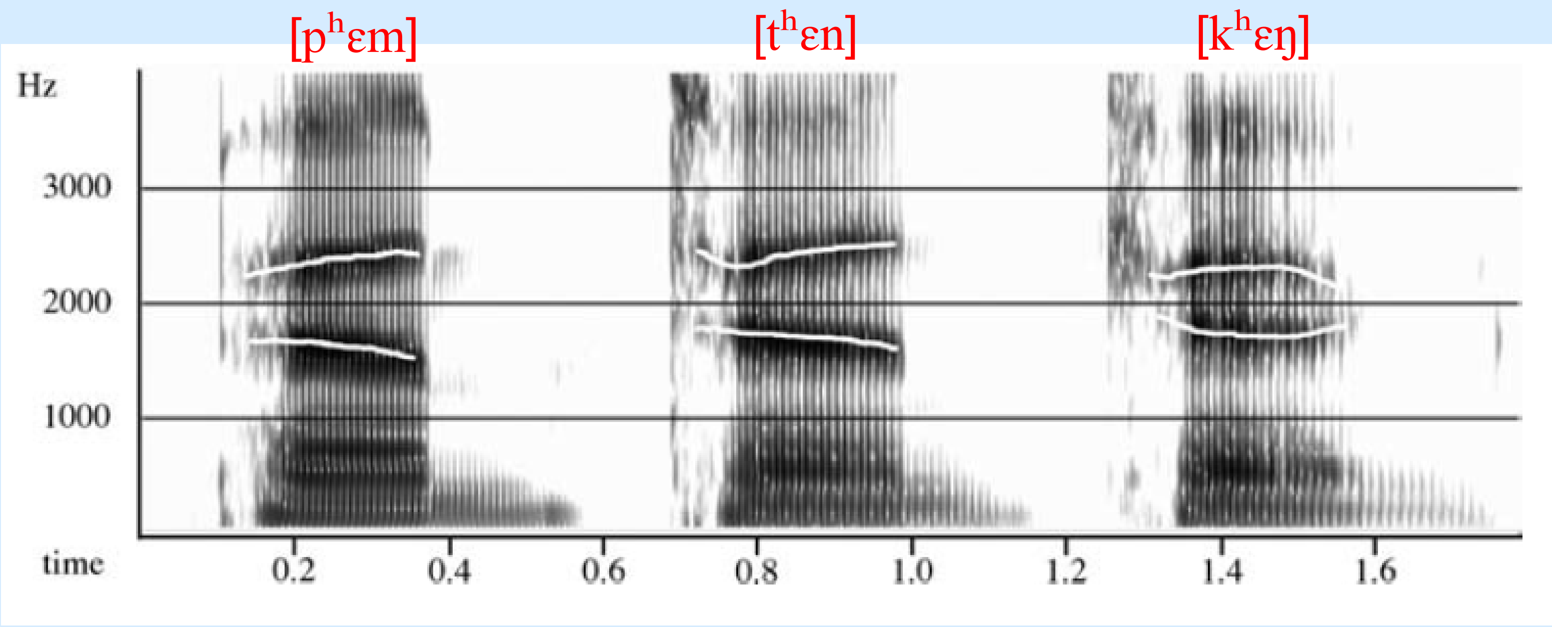
3.3 Voiceless stops

- The release of aspirated stops is marked by a sudden sharp spike→ lean vertical line.
- Period of aspiration noise→ absence of energy in F1 & no vertical striations
- Frequency & intensity
  - Whisper [t, t, t, k, k, k, p, p, p]. What do you observe?
  - [t] > [k] > [p]
- Intensity of [p] burst is sometimes so low that there is no evidence of it on a spectrogram.



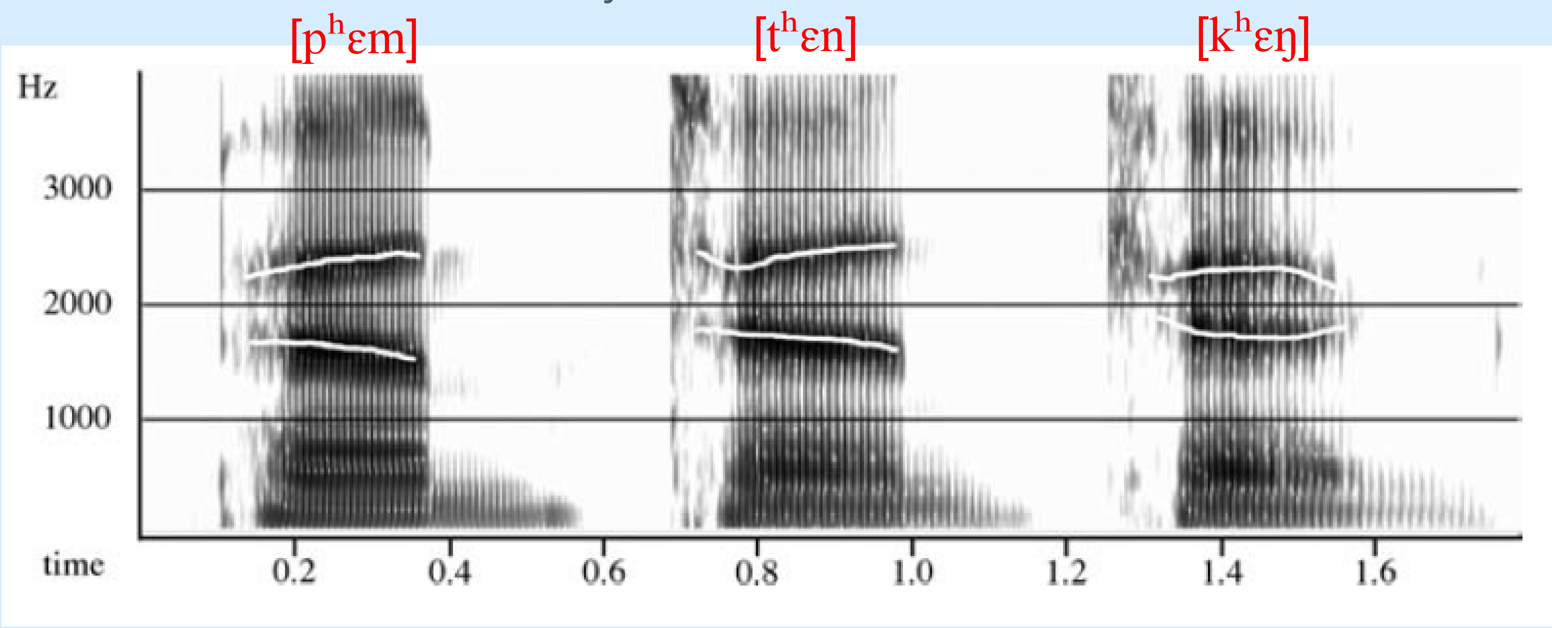
3.3 Voiceless stops

- Formant transitions also present in aspiration noise.
- [pʰɛm] : F2 & F3 rising into the vowel.
- [tʰɛn] : F2 steady, F3 dropping and then rising.
- [kʰɛŋ] : characteristic velar pinch



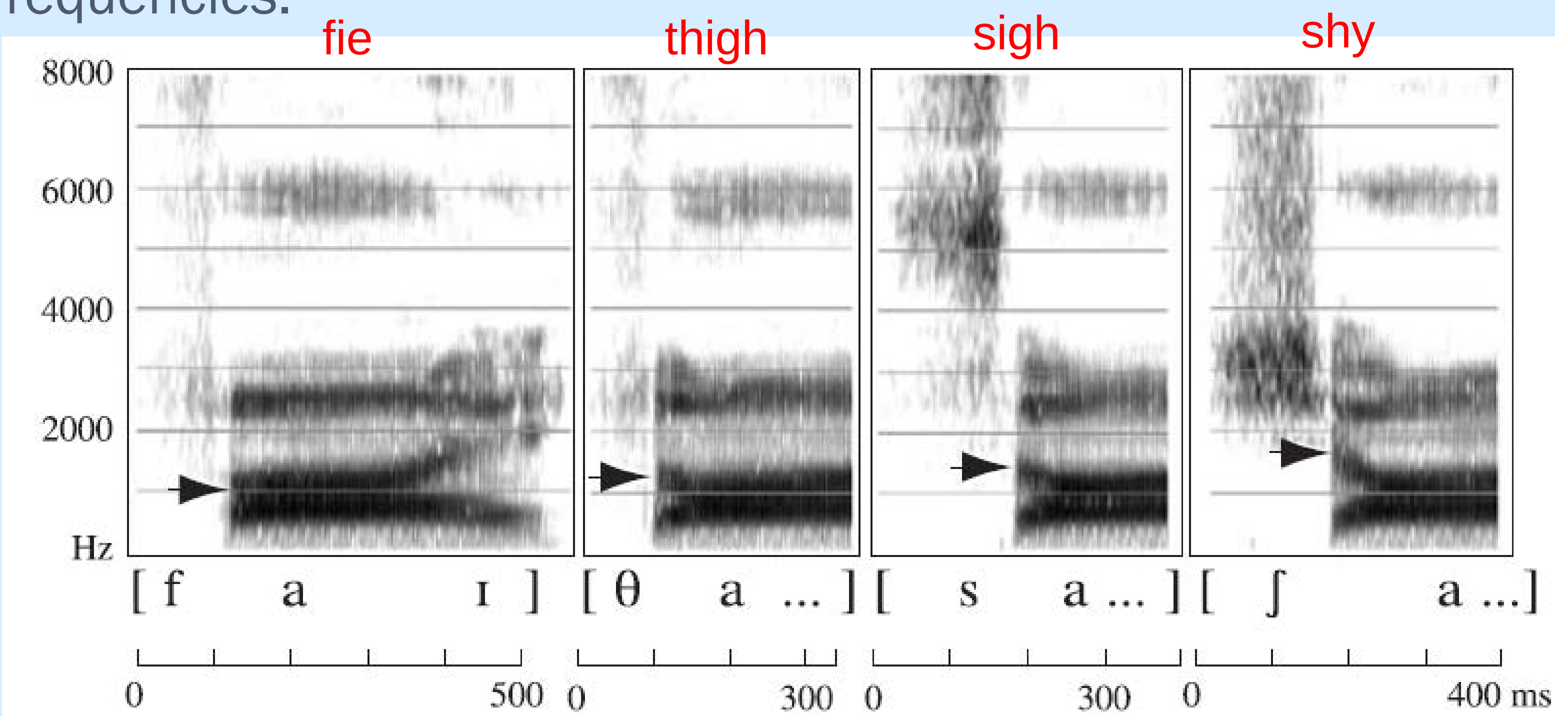
3.4 Nasals

- A clear mark of a nasal (and a lateral) is an **abrupt change** in the spectrogram at the time of the formation of the articulatory closure.
- A nasal has a formant structure similar to that of a vowel. **Differences:**
  - Bands are fainter.
  - Bands located in particular frequency locations depending on characteristic resonances of the nasal cavities.
- **F1: around 250 Hz**
- **Large region above F1 with no energy.**
- **F2 etc: varying according to speaker** (here below 2000 Hz).
- Place cues sometimes not very clear.



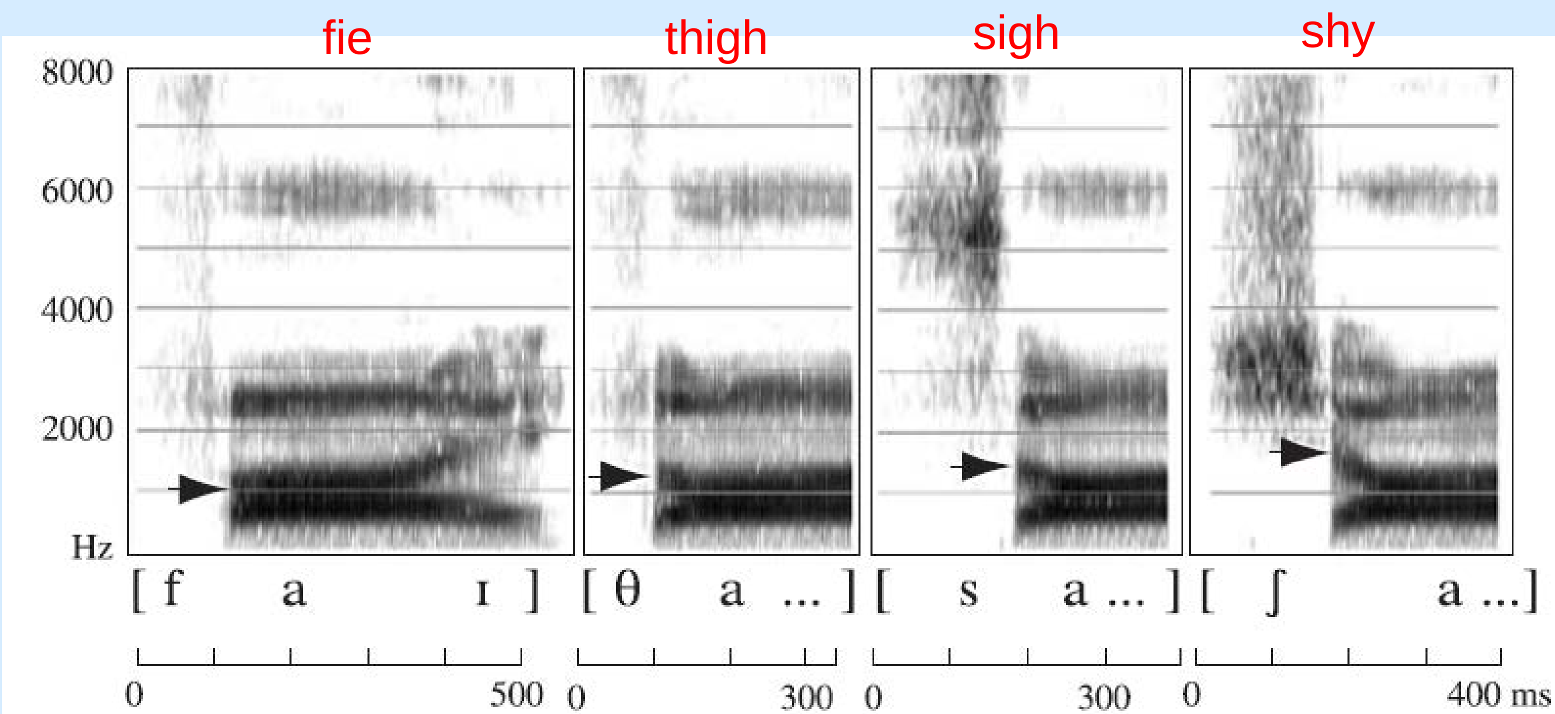
3.5 Voiceless fricatives

- Highest frequencies in speech occur over fricatives.
- Frequency scale increased to 8000 Hz.
- Diphthong [aɪ] : F1 & F2 start close together for low central [a] and move apart for high front [ɪ].
- Fricatives: Random energy distributed over a wide range of frequencies.



Voiceless fricatives [f, θ]

- Same pattern in [f] and [θ].
- Difference: Movement of F2 into following vowel.
  - Very little movement in [f].
  - In [θ], F2 starts around 1200 Hz and moves down.
- Often confused in noisy settings.
- Fallen together in some accents of English, such as London Cockney
  - *fin* and *thin* both pronounced with a [f].

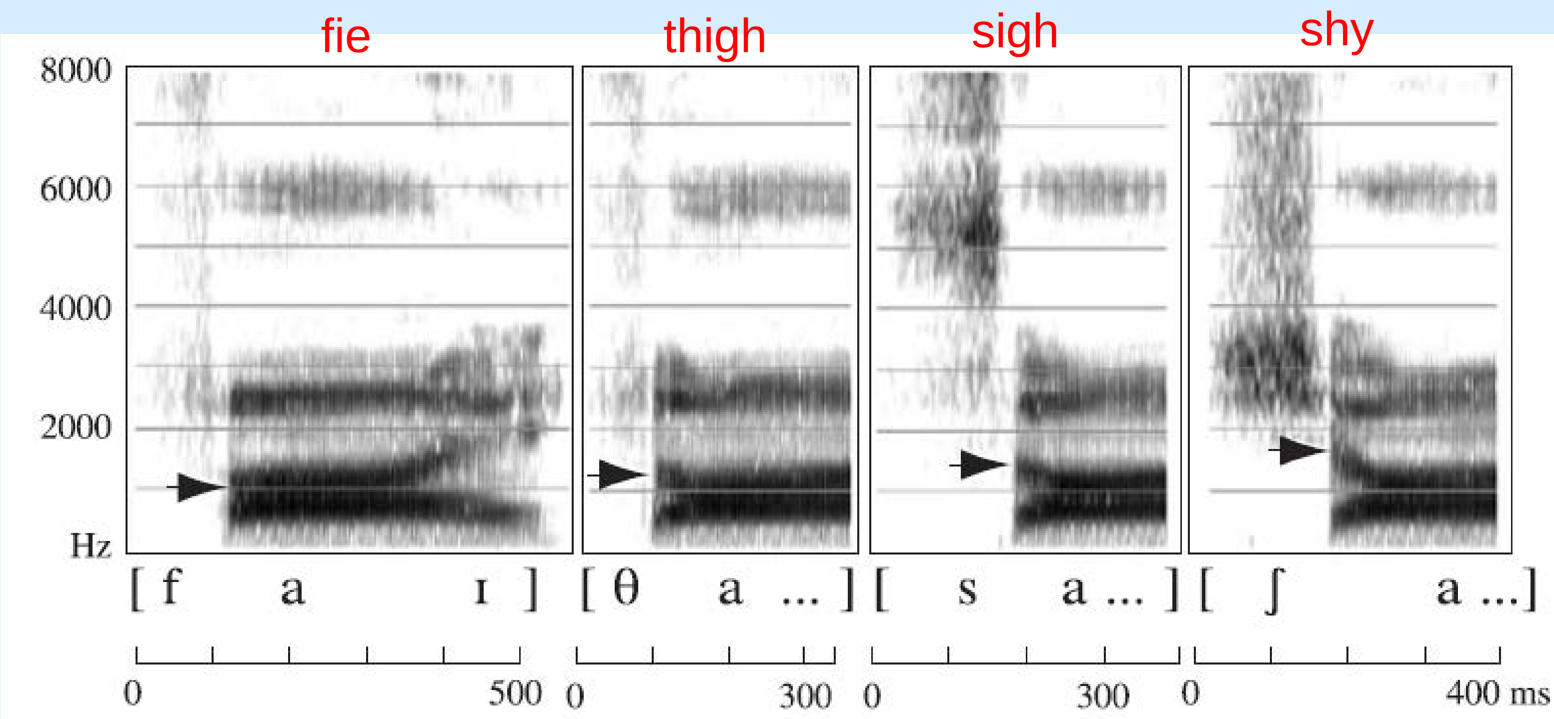


COCKNEY ACCENT

Learn the Cockney accent with Jason Statham (Go to 2:50 for ‘th’ sounds)
https://www.youtube.com/watch?v=1WvIwkL8oLc

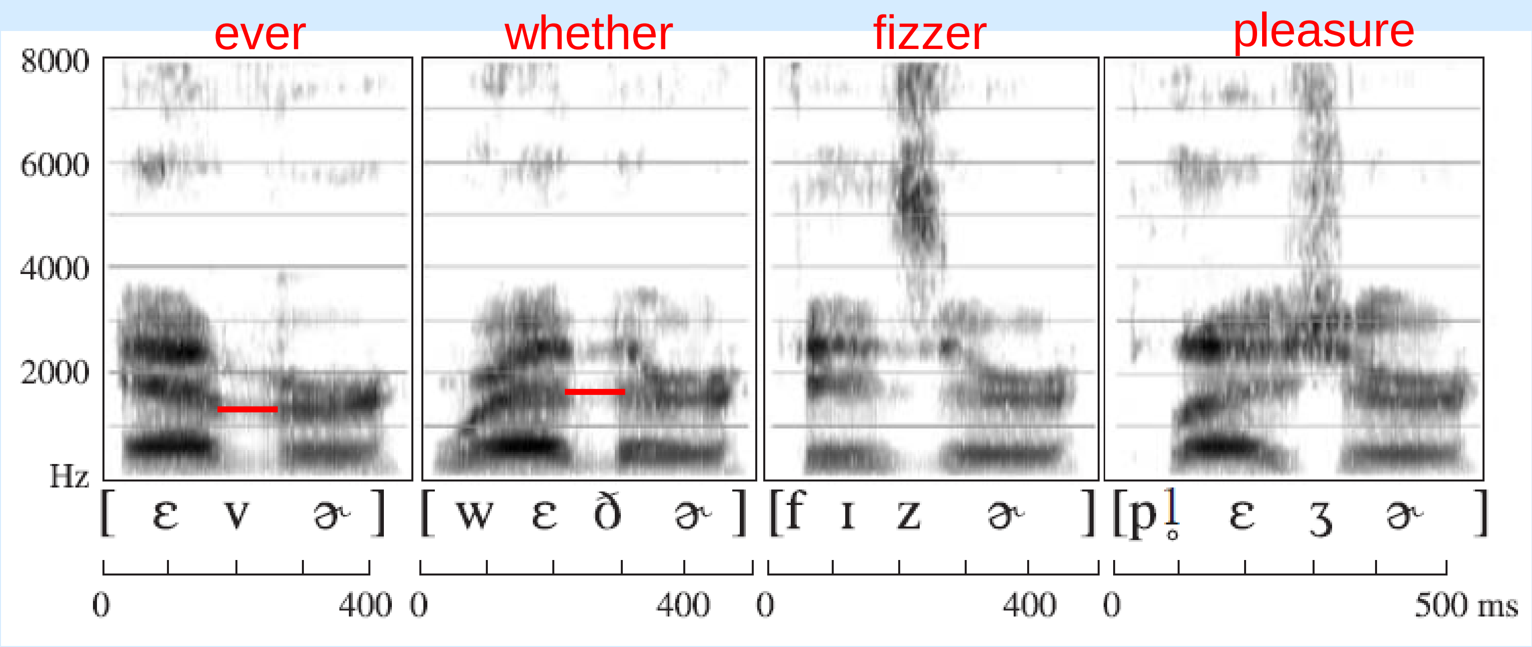
Voiceless fricatives [s, ʃ]

- The noise in [s] is centered at a high frequency, 5000 – 6000 Hz.
- In [ʃ] it is lower, extending down to about 2500 Hz.
- Both [s, ʃ] have **larger acoustic energy** and produce **darker patterns** than [f, θ]
- Both [s, ʃ] are marked with distinctive formant transitions.
- The locus of F2 transition increases throughout the words
  - [f] < [θ] < [s] < [ʃ] (see arrows in fig.)
- Before [ʃ] F2 of [a] is in a position comparable to its location in [i].



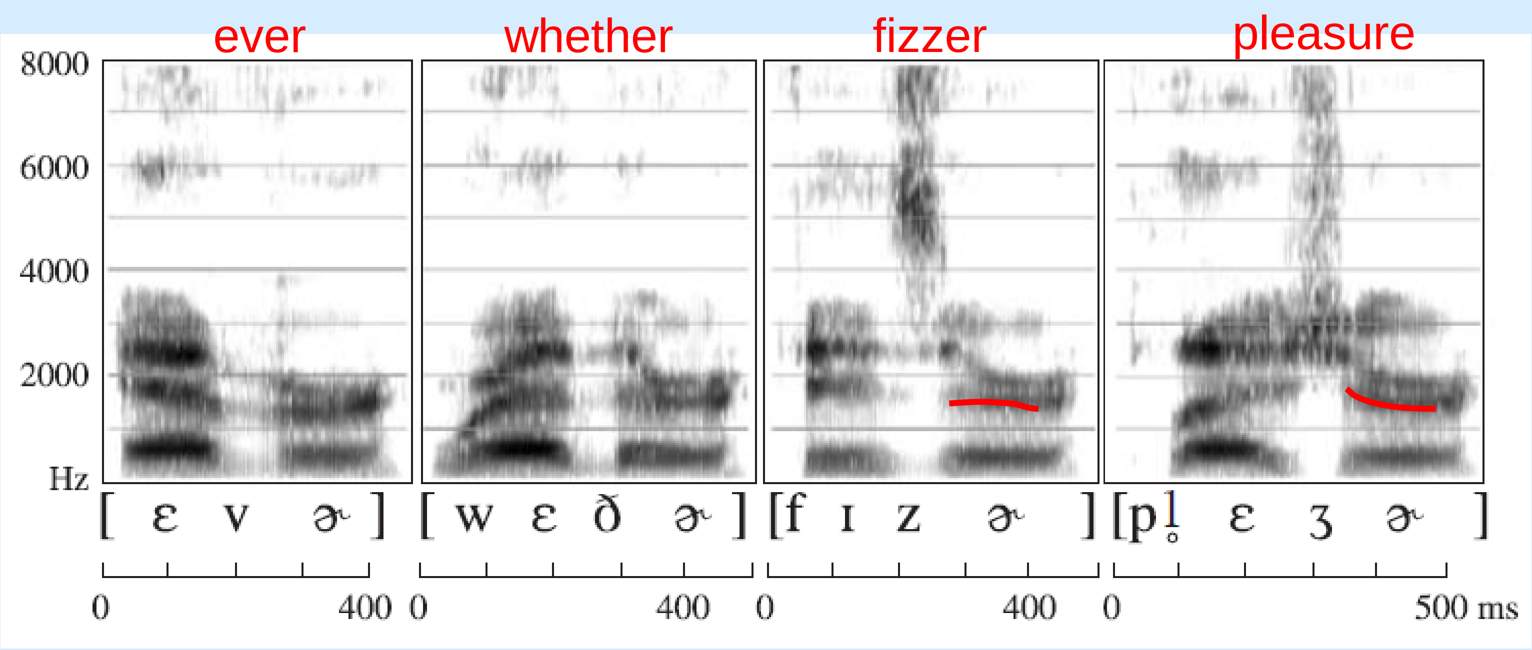
3.6 Voiced fricatives [v, ð]

- Voiced fricatives [v, ð, z, ʒ] have patterns similar to their voiceless counterparts [f, θ, s, ʃ].
- Voiced fricatives also have vertical striations indicative of voicing.
- Vertical striations due to voicing are apparent throughout [v] and [ð].
- The fricative component of [v] is very faint.
- F2 higher around [ð] than [v].



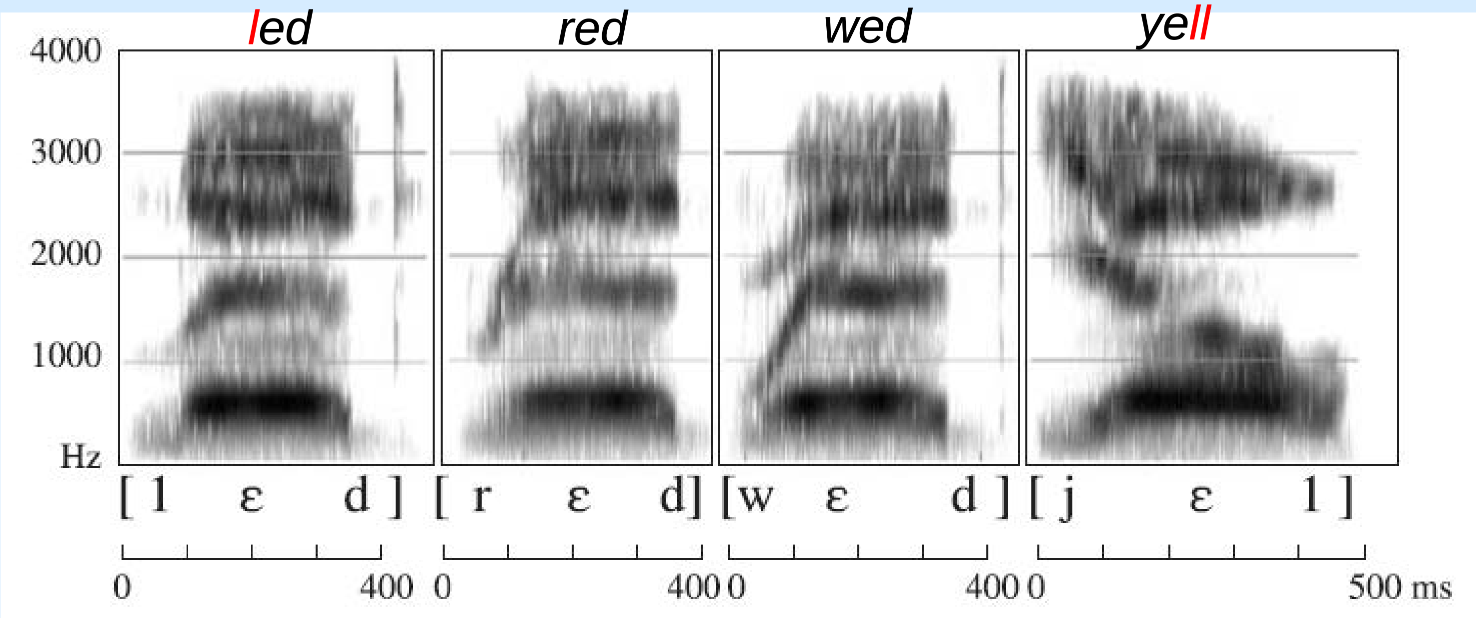
Voiced fricatives [z, ʒ]

- Fricative energy in higher frequencies very apparent in [z, ʒ].
- Voice bar
  - faint in [z]
  - hard to see in [ʒ] –vertical striations due to voicing in 6-8 kHz.
- F2 transition into [ɚ] is
  - level from[z]
  - descending from [ʒ]



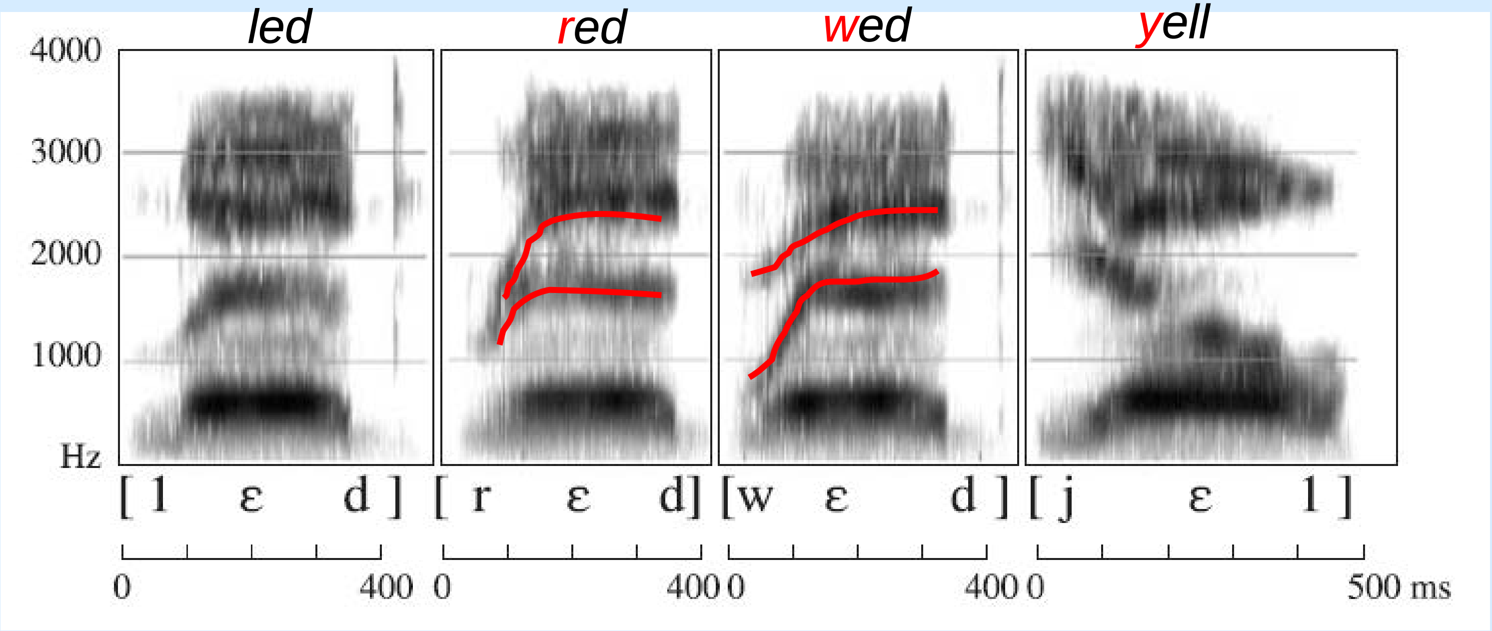
3.7 Lateral and central approximants

- Voiced approximants have formants not unlike those of vowels.
- The initial [l] has formants with center frequencies of approx. 250, 1100 & 2400 Hz, which change abruptly in intensity at the beginning of the vowel.
- A marked change in formant pattern is characteristic of voiced nasals and laterals.
- A final lateral may have little of no central contact, making it not really a lateral but a back unrounded vowel.
- A formant around 1100 or 1200 Hz is typical of most initial laterals for most speakers.



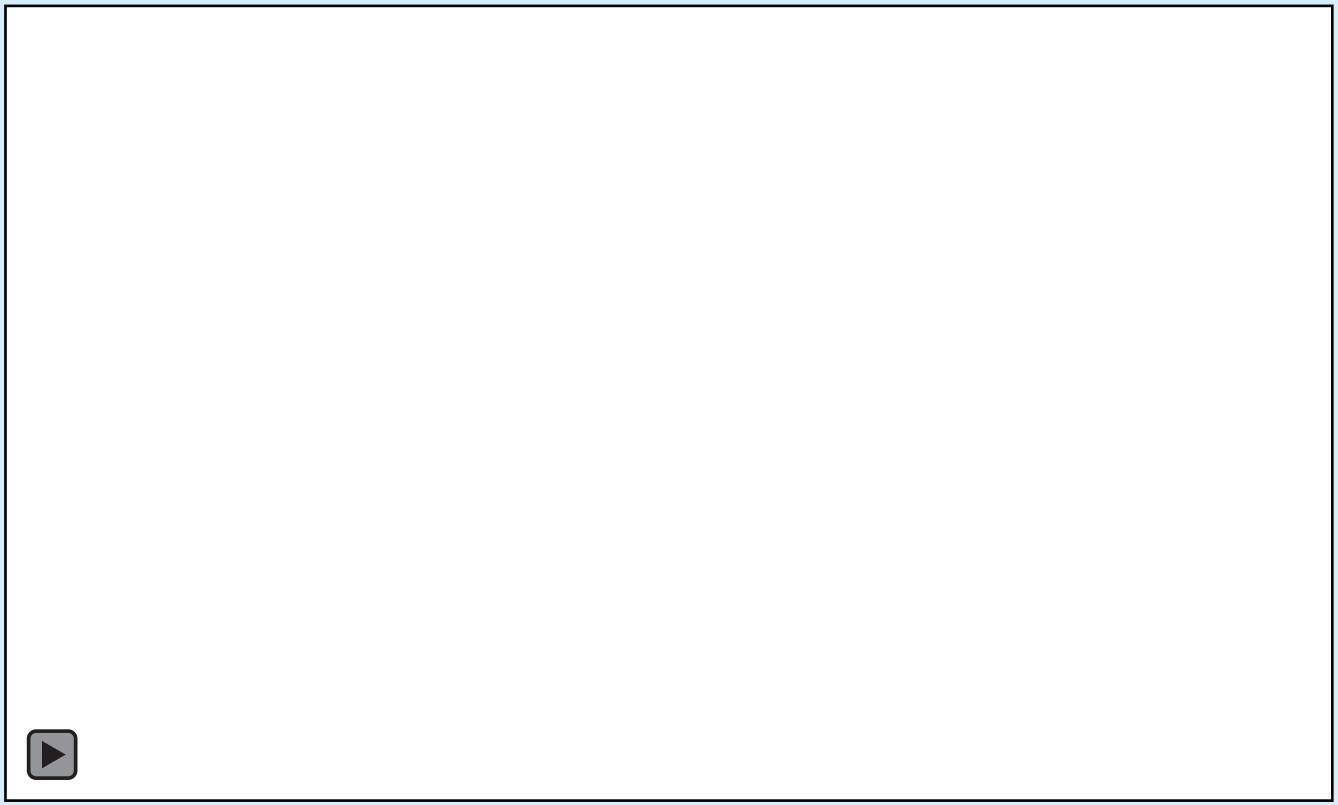
3.7 Lateral and central approximants

- The most obvious feature of approximant [ɹ] is the low frequency of F2 and F3.
- F3 begins at 1600 Hz!
- 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 <u>**semivowels**</u>, that is, semi versions of vowels [u] and [i] respectively.



3.7 Lateral and central approximants

[ɹ]



[w]

https://www.youtube.com/watch?v=e2vj6ma-854

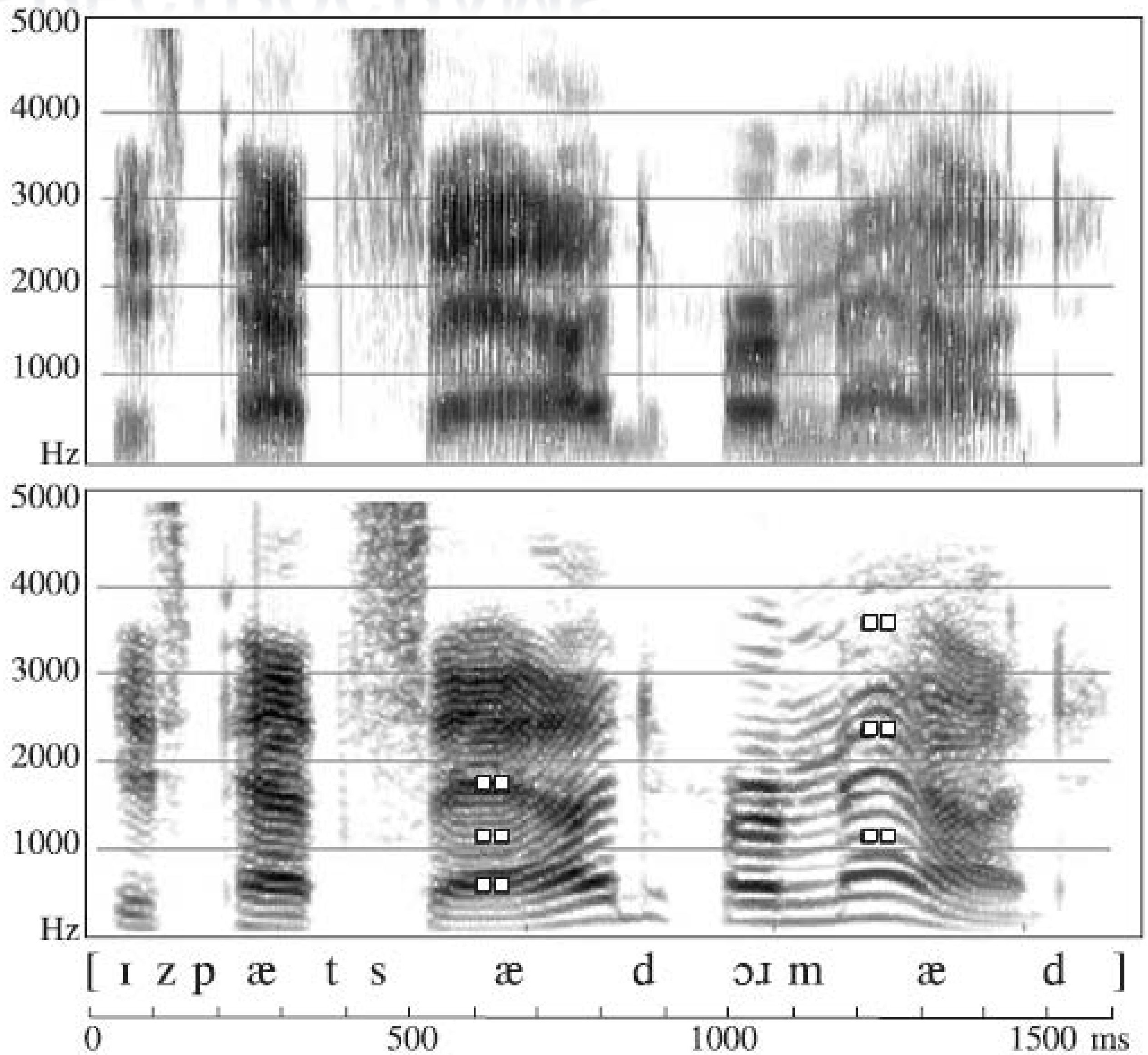
TABLE 8.1 Acoustic correlates of consonantal features. Note: These descriptions should be regarded only as rough guides. The actual acoustic correlates depend to a great extent on the particular combination of articulatory features in a sound and on the neighboring vowels.

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.

TYPES OF SPECTROGRAMS

wide-band spectrograms

narrow-band spectrograms



“Is Pat sad or mad?”

TYPES OF SPECTROGRAMS

Wide-band spectrograms

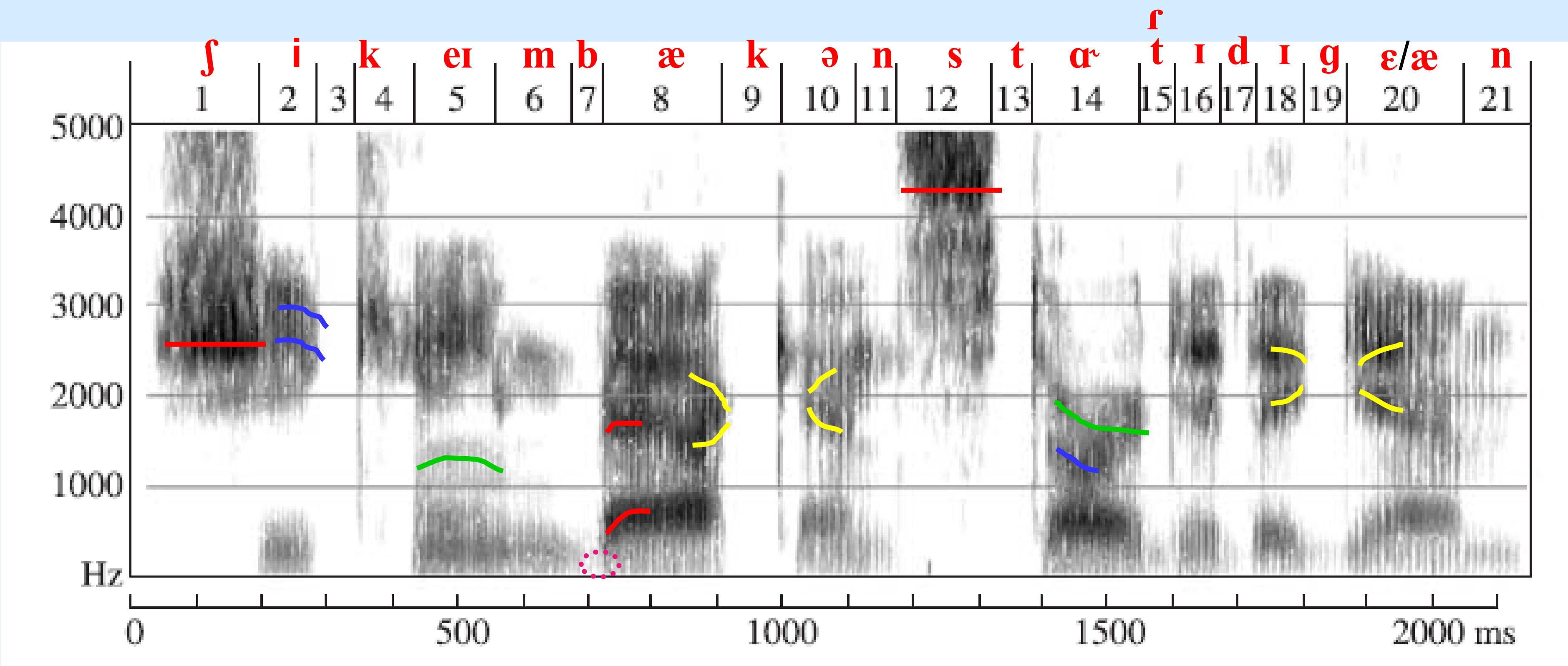
- Very accurate in the time dimension
  - They show each vibration of the vocal folds as a separate vertical line.
  - They indicate the precise moment of a stop burst with a vertical spike.
- Less accurate in the frequency dimension
  - There are usually several component frequencies present in a single formant, all of them lumped together in one wide band on the spectrogram.

Narrow-band spectrograms

- More accurate in the frequency dimension (at the expense of accuracy in the time dimension).
  - The spikes of stop releases are smeared in the time dimension in the narrow-band spectrogram.
  - The frequencies that compose each formant are visible.

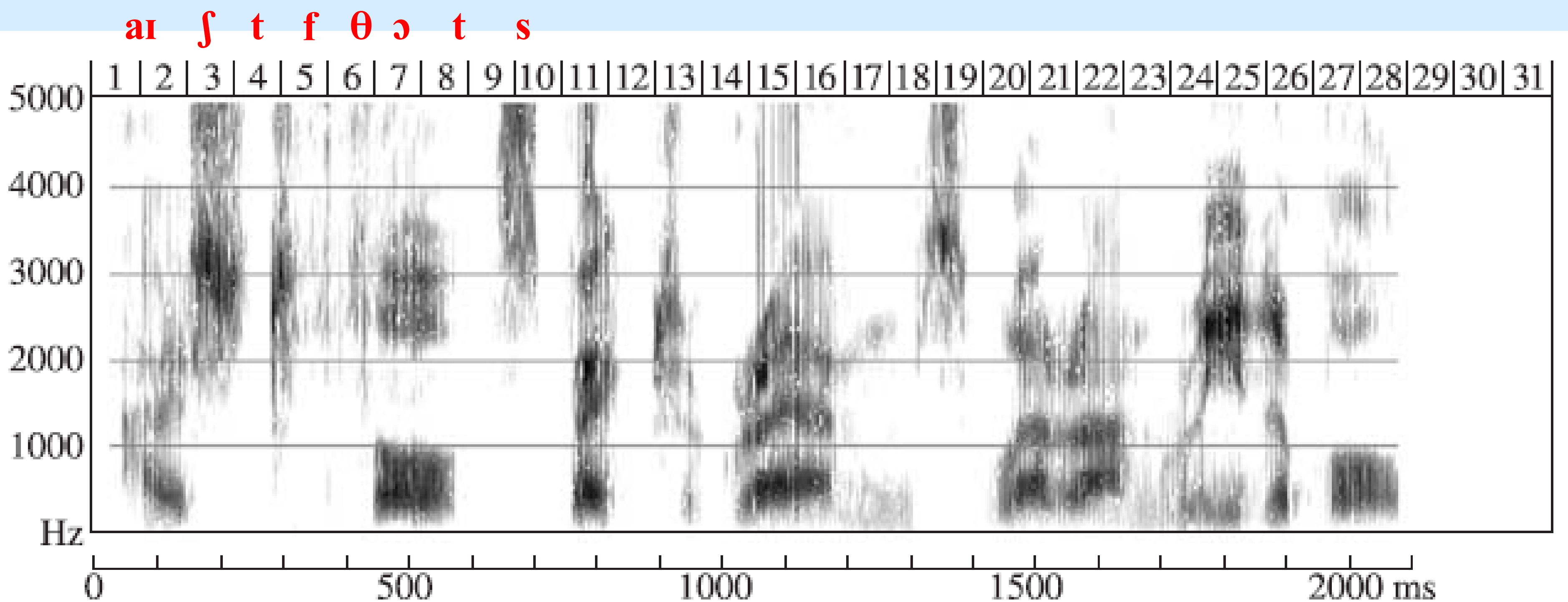
INTERPRETING SPECTROGRAMS (I)

- In connected speech, many of the sounds are more difficult to distinguish.
- Transcribe the segments in the following phrase
  “She came back and started again.” (American English)
- [ʃi keɪm bæk ən ˈstɑ˞ɾɪd ˈɪgæn]



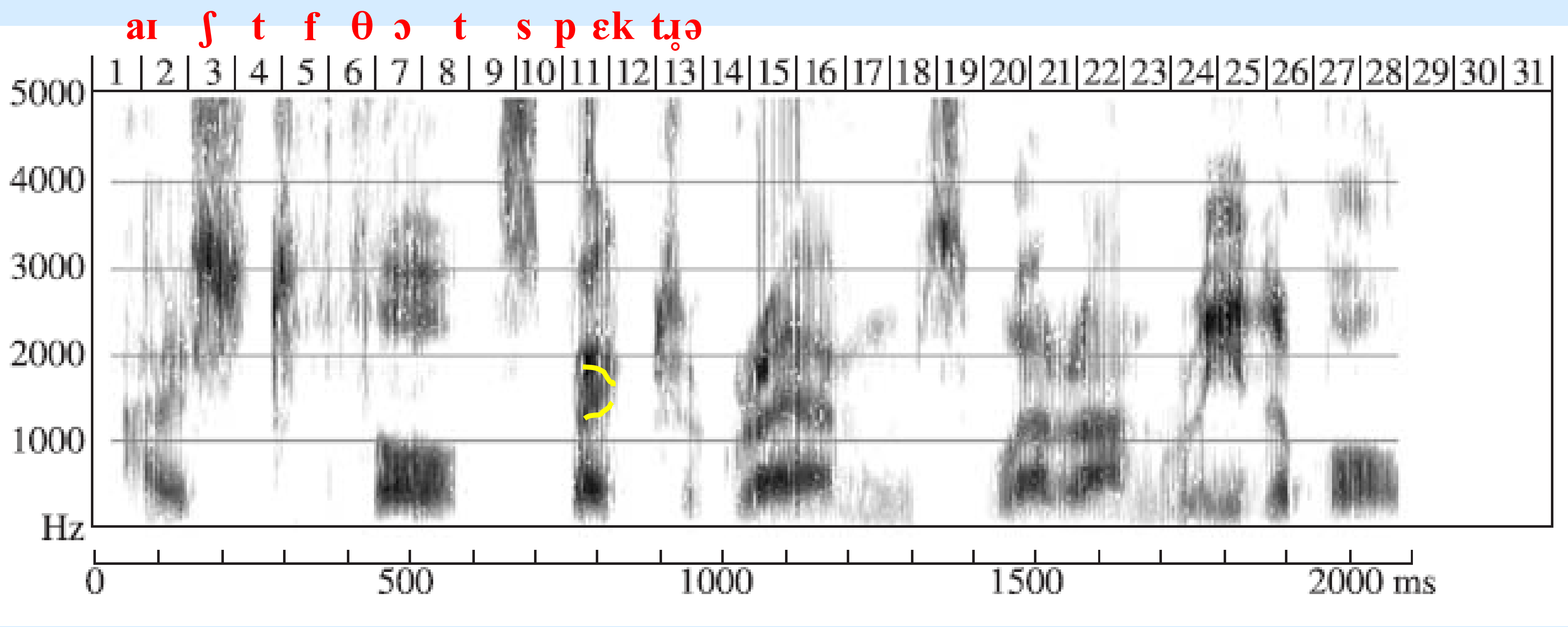
INTERPRETING SPECTROGRAMS (II)

- *I should have thought spectrograms were unreadable. (British English)*
- We first find obvious things first, i.e. [s, ʃ] which stand out.
- Start at the beginning, and find the vowel [aɪ] in the first word.
- The vowel in *thought* before [s]. And then the [t] in *thought*.
- It seems as if the whole of the phrase “*should have*” was pronounced without any voicing: [aɪʃtf ˈθɔt]



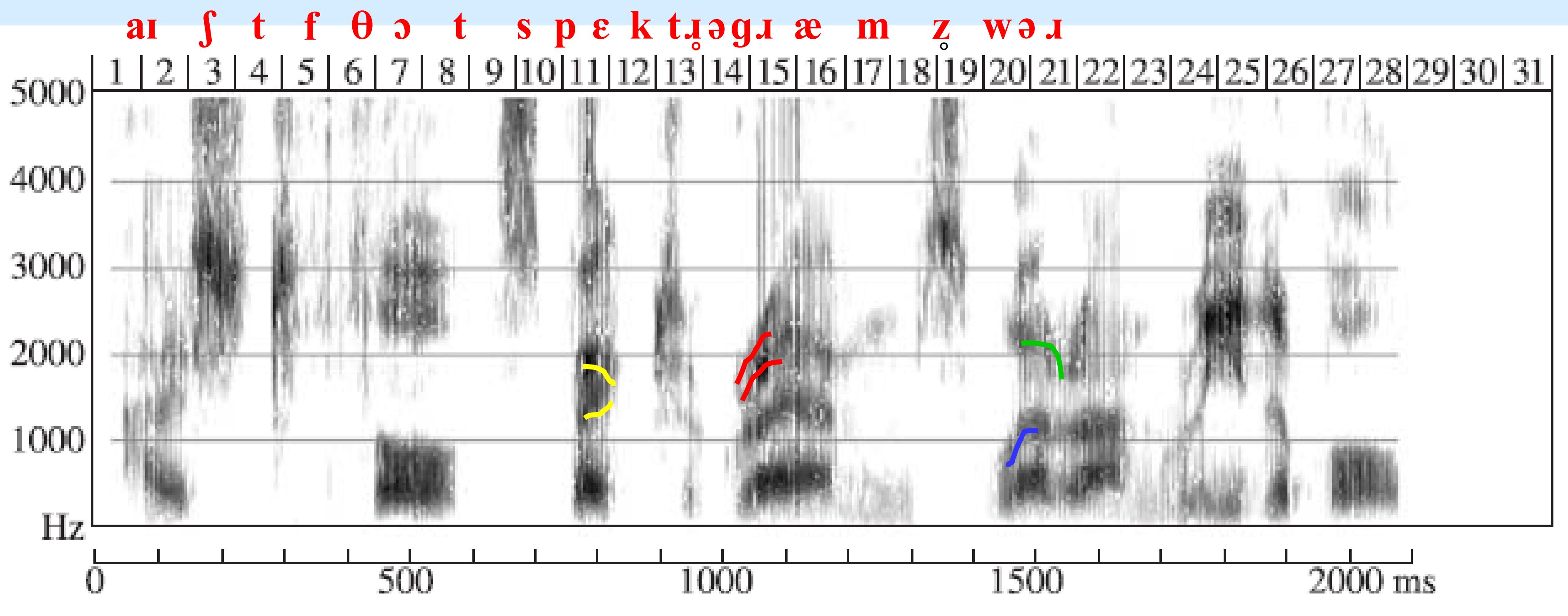
INTERPRETING SPECTROGRAMS (II)

- *I should have thought spectrograms were unreadable.*
- Try to transcribe *spectrograms were unreadable,* remembering that some of the sounds you might have expected to be voiced might be voiceless.
- No aspiration after [p].
- [ɛ] is very short but you can see the coming together of F2 and F3 for the [k].
- [t] is highly aspirated, so the following [r] becomes voiceless. Same with [ə].



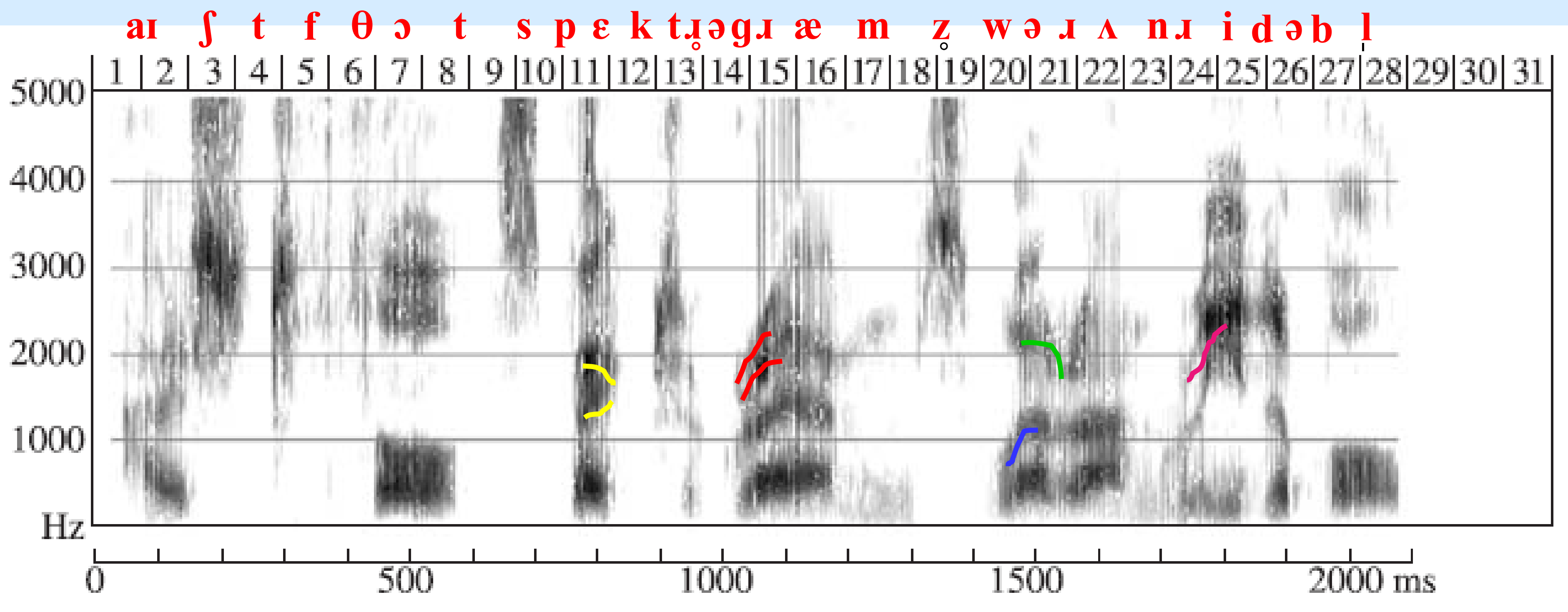
INTERPRETING SPECTROGRAMS (II)

- *I should have thought spectrograms were unreadable.*
- The velar stop [g] is released into an [ɹ] located by the lowering of F3 and F4.
- The fricative after the [m] appears to be voiceless and of less intensity than a [s].
- The [w] is distinguishable by the low F2 of the following vowel.
- The lowering of F3 marks the [r] in *were*.



INTERPRETING SPECTROGRAMS (II)

- *I should have thought spectrograms were unreadable.*
- The lowering of F3 marks the [r] in *unreadable*.
- [d] and [ə] are very short.
- The final syllabic /l/ looks like a back vowel.



INDIVIDUAL DIFFERENCES

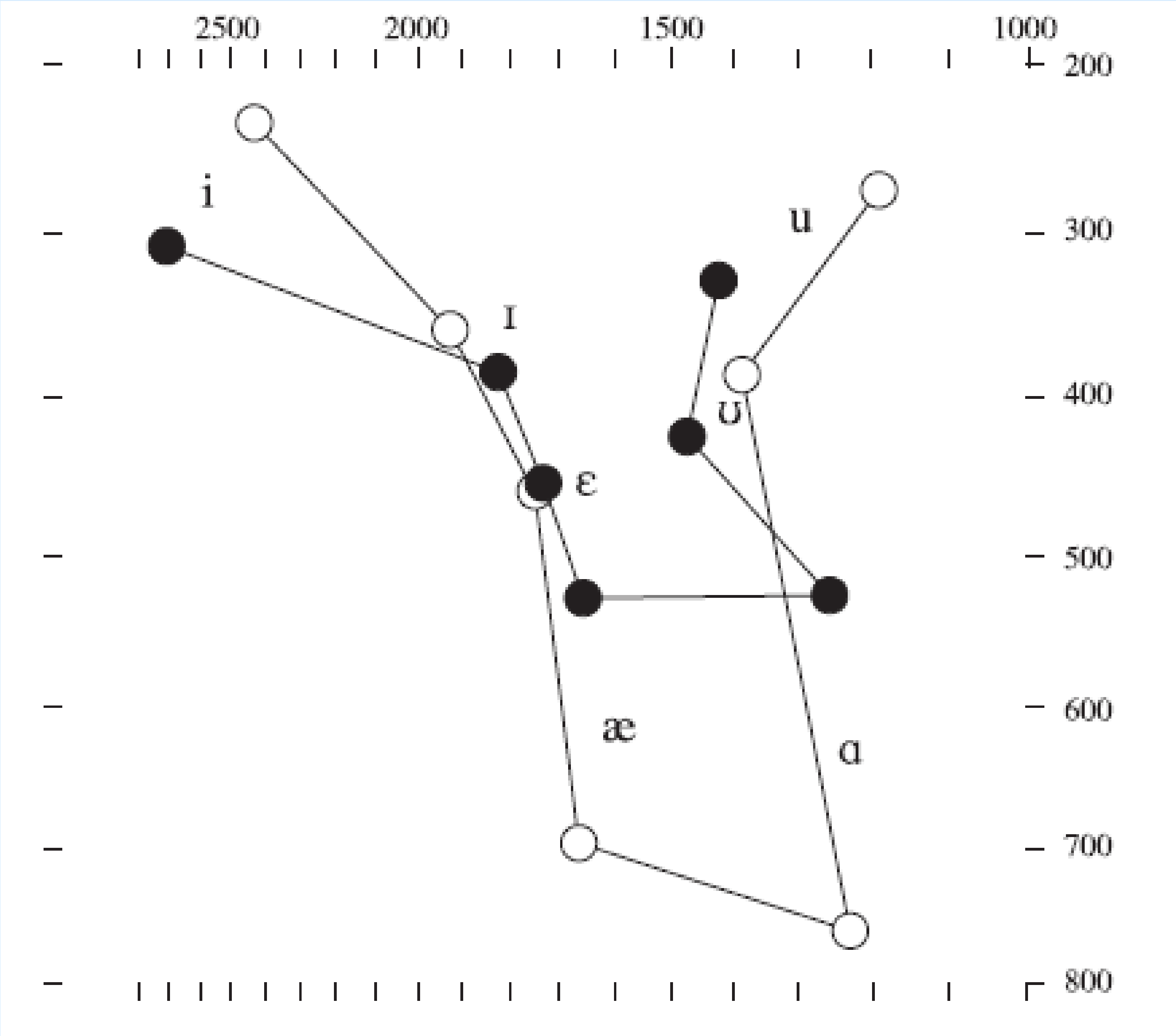
- It is important to know what sort of differences exist between different speakers.
  1. When trying to measure features that are linguistically significant, one must know how to discount purely individual features.
  2. When trying to find out whether a speaker has speech problems.
  3. For valid speaker identification in forensic situations.

- Individual variation is readily apparent when studying spectrograms → **relative quality**

INDIVIDUAL DIFFERENCES

- Same phonetic quality
  - Similar relative positions
  - Different absolute values

Vowels pronounced by 2 speakers of Californian English.



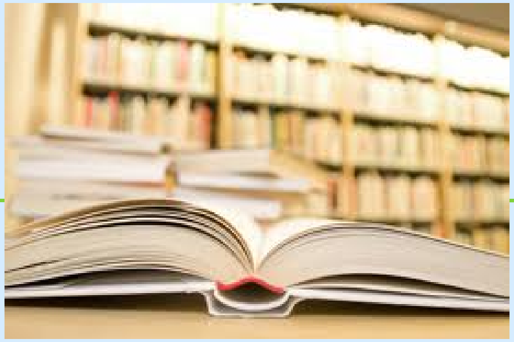
INDIVIDUAL DIFFERENCES

- No simple technique to average out individual characteristics so that a formant plot shows only the phonetic qualities of vowels.
- **F4** indicator of individual's head size
  - Express values of other formants as percentages of the mean F4.
  - F4 values are not usually reported.
- Phoneticians do not really know how to compare acoustic data on the sounds of one individual with those of another.
- We cannot write a computer program that will accept any individual's vowels as input and then output a narrow phonetic transcription.

PHONETICS ASSIGNMENT

- Do the waveform and spectrogram reading exercises on Phonetics Assignment_CS-578_2020-21.pdf
- Submit to asfakianaki@csd.uoc.gr by March 22^{nd}
- 10% of grade (only timely submissions)

Read…



- Ladefoged & Johnson “A course in phonetics”, chapter 8
- Ladefoged “Vowels & Consonants”, chapter 7
- Lieberman & Blumstein “Speech physiology, speech perception & acoustic phonetics”, chapter 5, pp. 51-73
- Clark & Yallop “An Introduction to Phonetics & Phonology”, chapter 7
- Ladefoged, “Elements of Acoustic Phonetics”, chapter: “The Production of Speech”

…& visit

- https://corpus.linguistics.berkeley.edu/acip/course/chapter8/
  Material for chapter 8 from UC Berkley Linguistics, “A course in phonetics” including online exercises
- 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.linguistics.ucla.edu/people/hayes/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.youtube.com/watch?v=Gg4IHbiITd0
  Introduction to spectrogram analysis (FloridaLinguistics.com)