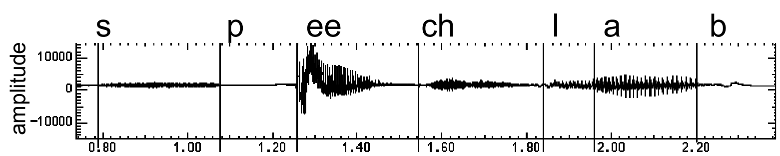
Statistical NLP
Spring 2009

Lecture 9: Speech Signal

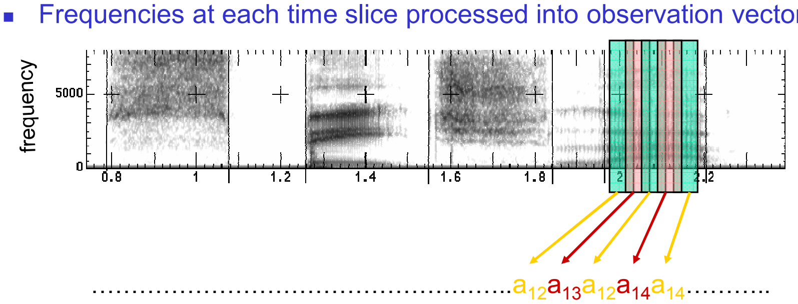
Dan Klein – UC Berkeley

Speech in a Slide

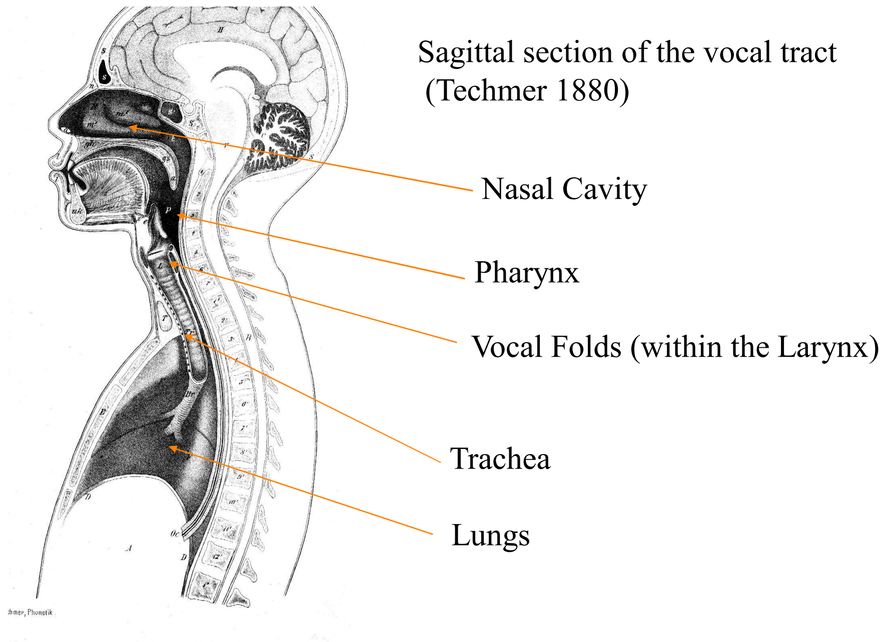
- Frequency gives pitch; amplitude gives volume



- Frequencies at each time slice processed into observation vectors



Sagittal section of the vocal tract (Techmer 1880)



Places of articulation

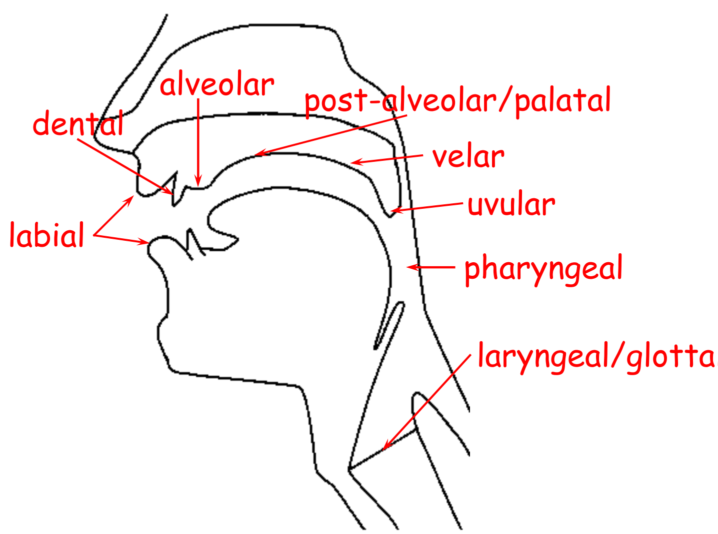
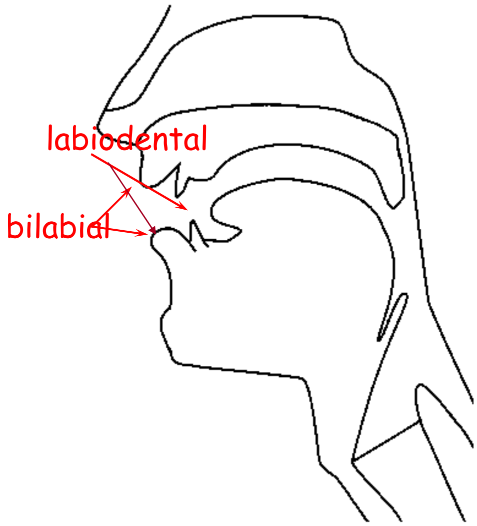


Figure thanks to Jennifer Venditti

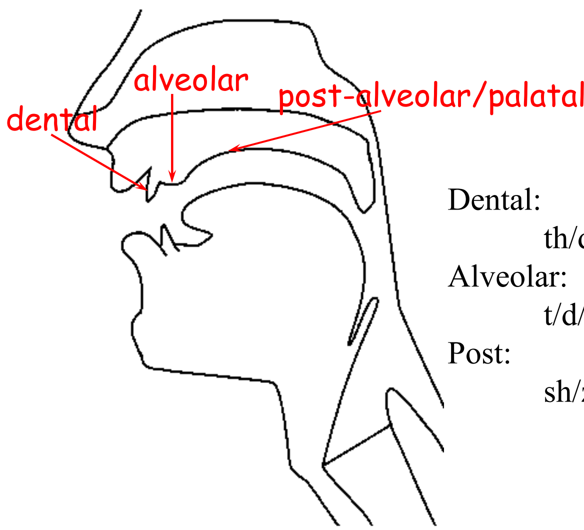
Labial place



Bilabial:
p, b, m
Labiodental:
f, v

Figure thanks to Jennifer Venditti

Coronal place



Dental:
th/dh
Alveolar:
t/d/s/z/l
Post:
sh/zh/y

Figure thanks to Jennifer Venditti

Dorsal Place

Velar:
k/g/ng

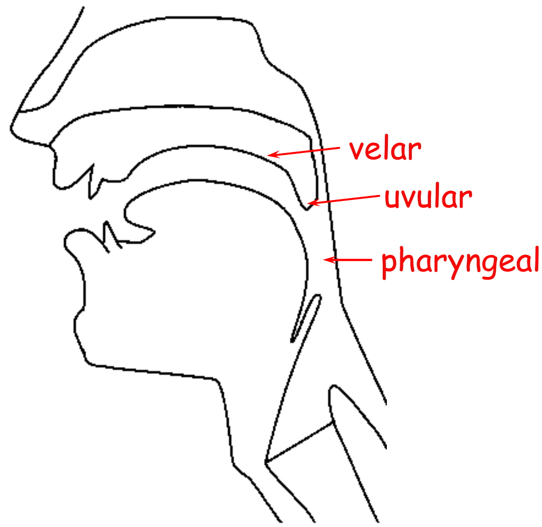
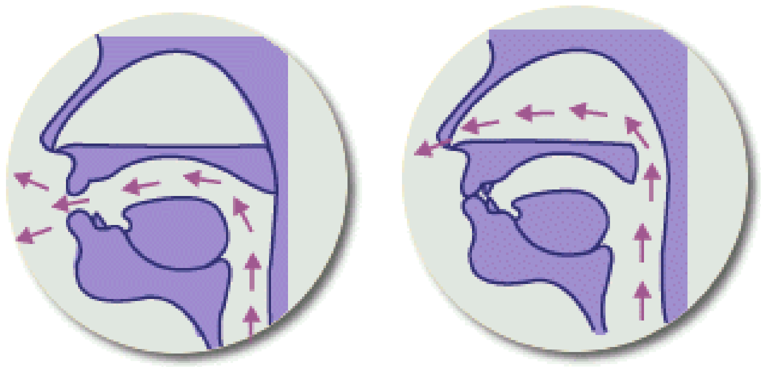


Figure thanks to Jennifer Venditti

Manner of Articulation

- Stop: complete closure of articulators, so no air escapes through mouth
- Oral stop: palate is raised, no air escapes through nose. Air pressure builds up behind closure, explodes when released
  - p, t, k, b, d, g
- Nasal stop: oral closure, but palate is lowered, air escapes through nose.
  - m, n, ng

Oral vs. Nasal Sounds



Thanks to Jong-bok Kim for this figure!

Vowels

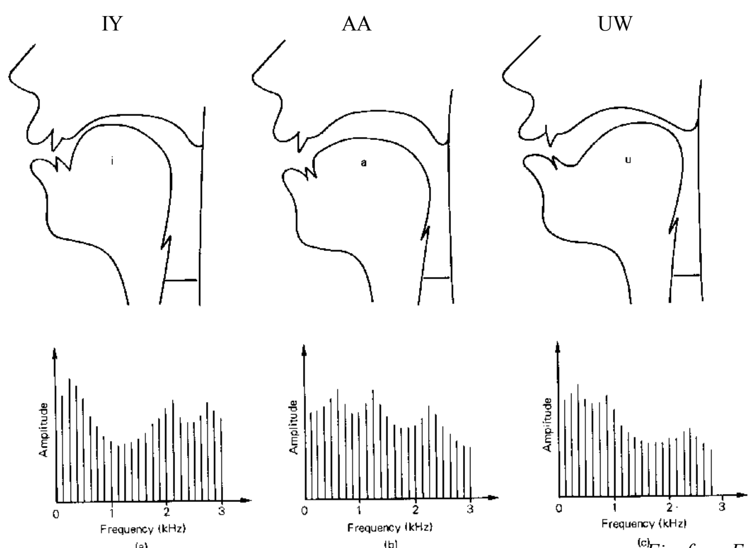
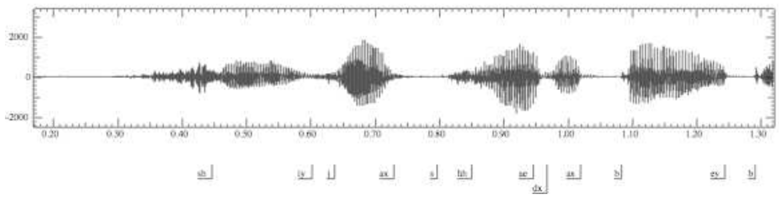


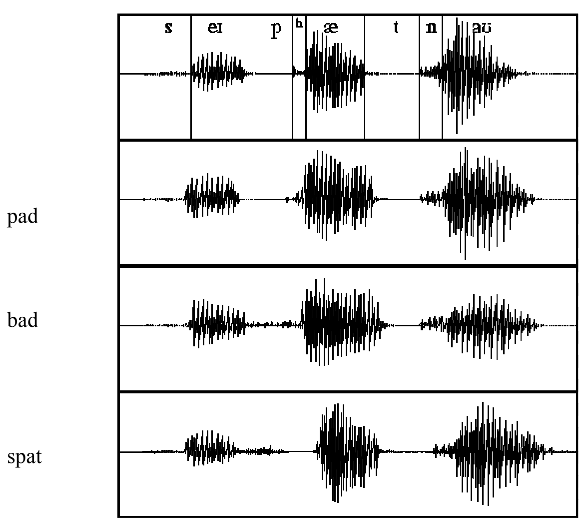
Fig. from Eric Keller

She just had a baby

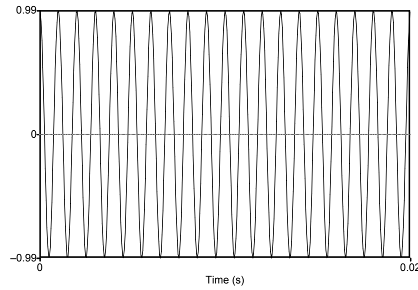


- What can we learn from a wavefile?
  - Vowels are voiced, long, loud
  - Length in time = length in space in waveform picture
  - Voicing: regular peaks in amplitude
  - When stops closed: no peaks: silence.
  - Peaks = voicing: .46 to .58 (vowel [iy], from second .65 to .74 (vowel [ax]) and so on
  - Silence of stop closure (1.06 to 1.08 for first [b], or 1.26 to 1.28 for second [b])
  - Fricatives like [sh] intense irregular pattern; see .33 to .46

Examples from Ladefoged

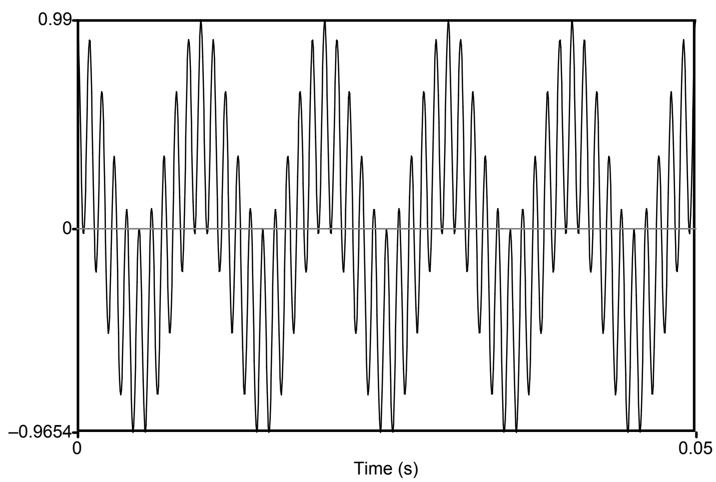


Simple periodic waves of sound



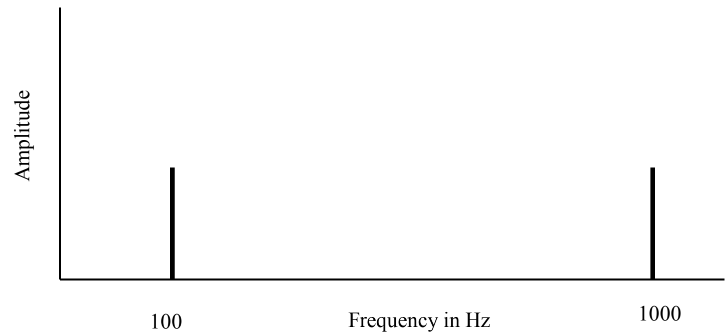
- **Y axis: Amplitude = amount of air pressure at that point in time**
  - **Zero is normal air pressure, negative is rarefaction**
- **X axis: time. Frequency = number of cycles per second.**
- **Frequency = 1/Period**
- **20 cycles in .02 seconds = 1000 cycles/second = 1000 Hz**

Complex waves: 100Hz+1000Hz

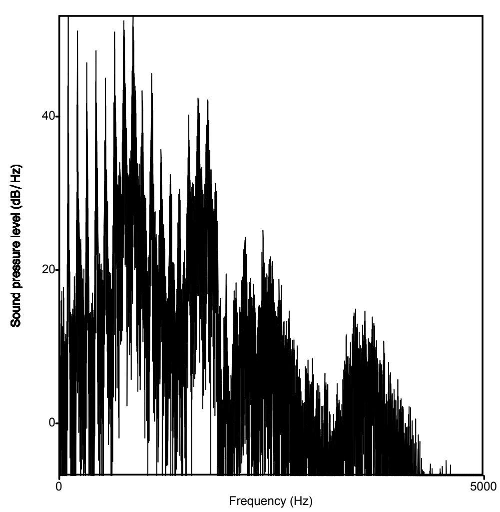


Spectrum

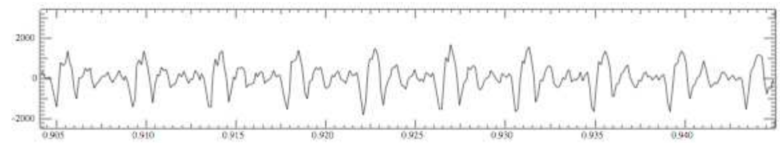
Frequency components (100 and 1000 Hz) on x-axis



Spectrum of an actual soundwave



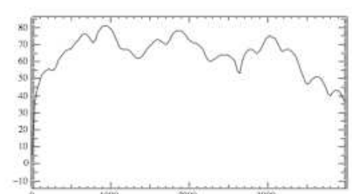
Part of [ae] waveform from “had”



- Note complex wave repeating nine times in figure
- Plus smaller waves which repeats 4 times for every large pattern
- Large wave has frequency of 250 Hz (9 times in .036 seconds)
- Small wave roughly 4 times this, or roughly 1000 Hz
- Two little tiny waves on top of peak of 1000 Hz waves

Back to Spectra

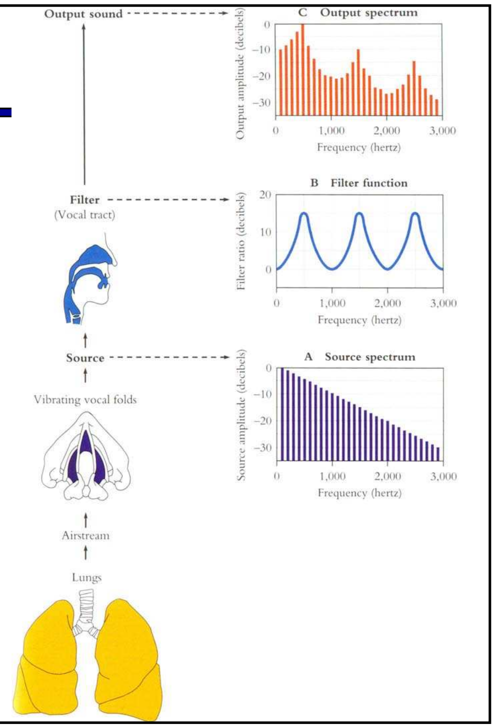
- Spectrum represents these freq components
- Computed by Fourier transform, algorithm which separates out each frequency component of wave.



- x-axis shows frequency, y-axis shows magnitude (in decibels, a log measure of amplitude)
- Peaks at 930 Hz, 1860 Hz, and 3020 Hz.

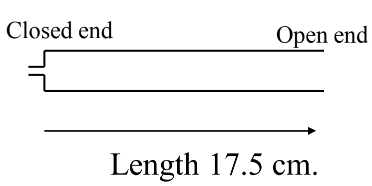
Why these Peaks?

- Articulator process:
  - The vocal cord vibrations create harmonics
  - The mouth is an amplifier
  - Depending on shape of mouth, some harmonics are amplified more than others



Resonances of the Vocal Tract

- The human vocal tract as an open tube



- Air in a tube of a given length will tend to vibrate at resonance frequency of tube.
- Constraint: Pressure differential should be maximal at (closed) glottal end and minimal at (open) lip end.

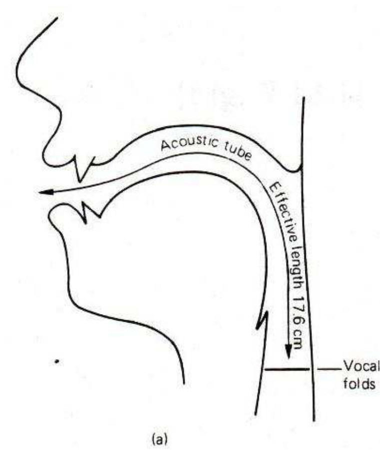
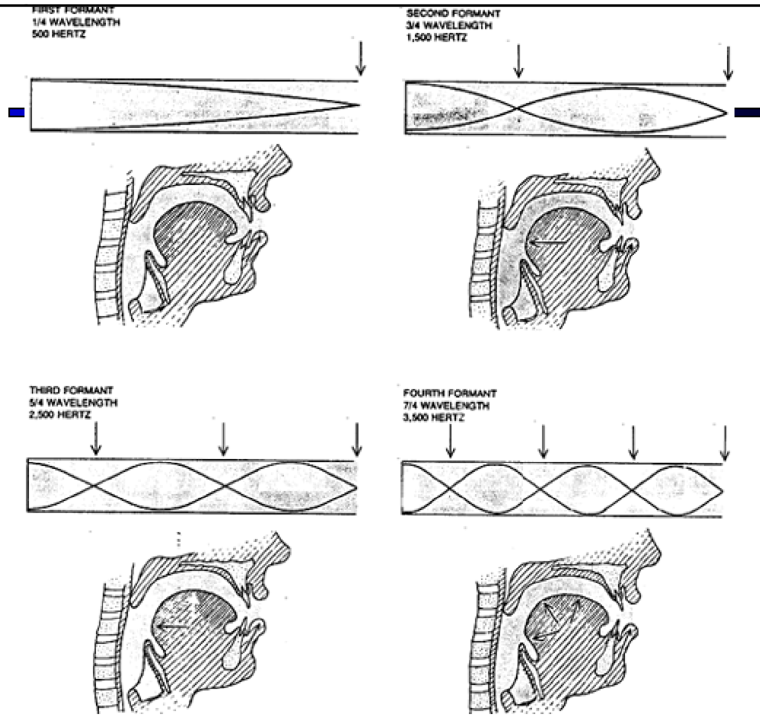


Figure from W. Barry Speech Science slides



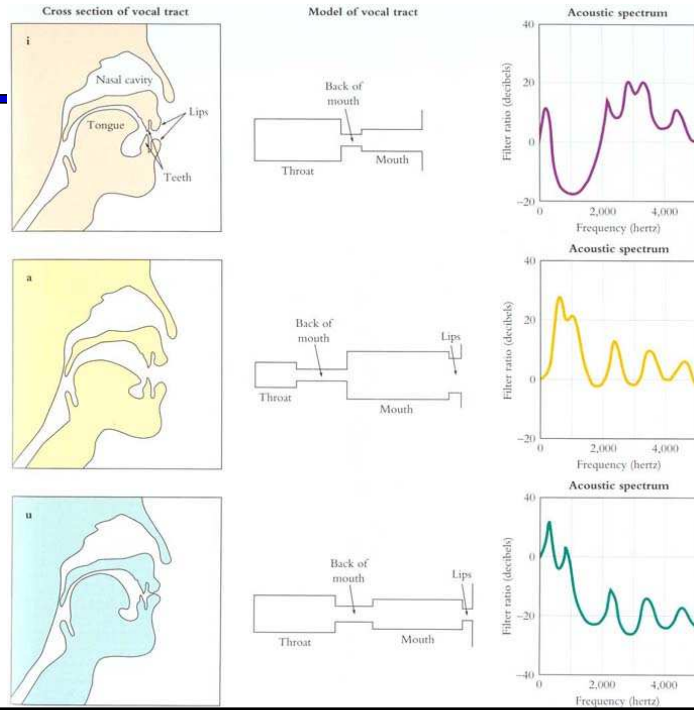
From Sundberg

Computing the 3 Formants of Schwa

- Let the length of the tube be L
  - $F_1 = c/\lambda_1 = c/(4L) = 35,000/4*17.5 =$ 500Hz
  - $F_2 = c/\lambda_2 = c/(4/3L) = 3c/4L = 3*35,000/4*17.5 =$ 1500Hz
  - $F_3 = c/\lambda_3 = c/(4/5L) = 5c/4L = 5*35,000/4*17.5 =$ 2500Hz

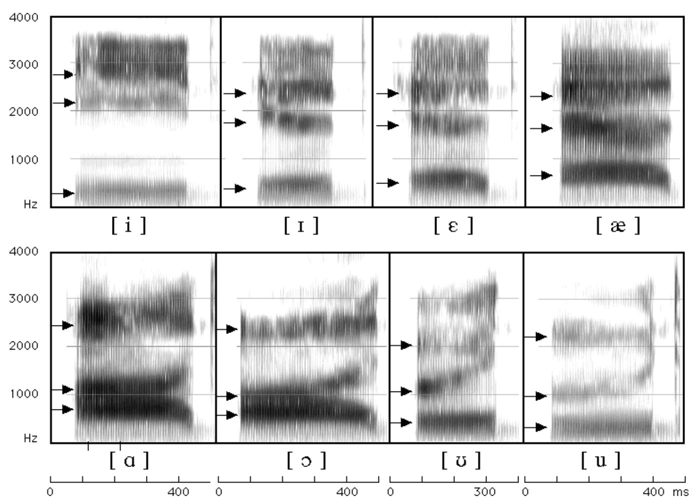
- So we expect a neutral vowel to have 3 resonances at 500, 1500, and 2500 Hz

- These vowel resonances are called formants



From Mark Liberman's Web site

Seeing formants: the spectrogram



American English Vowel Space

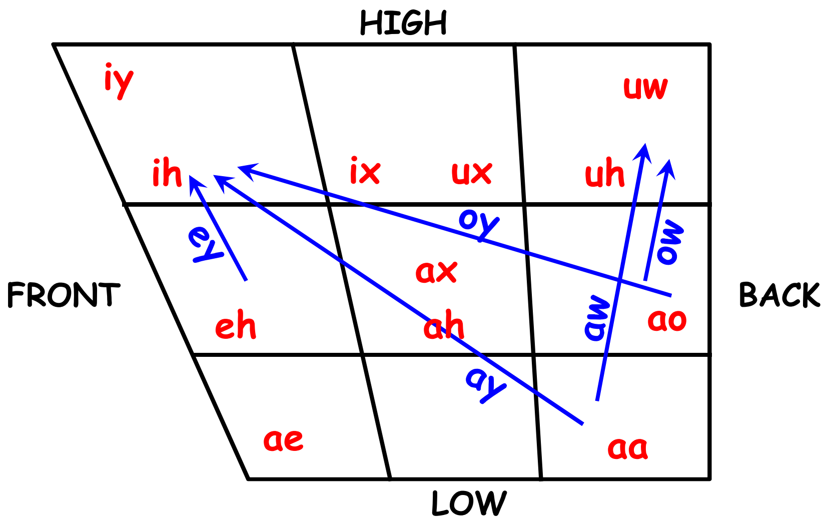
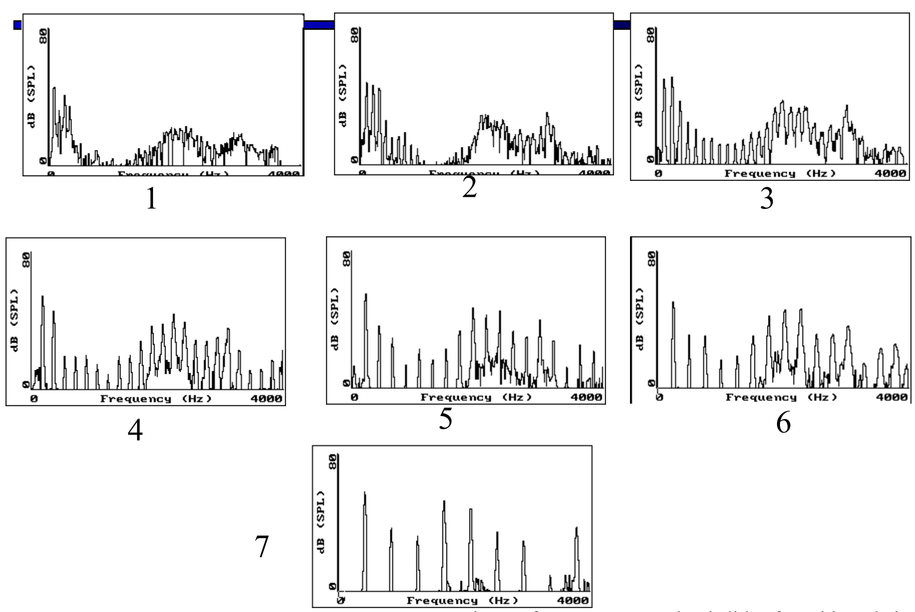


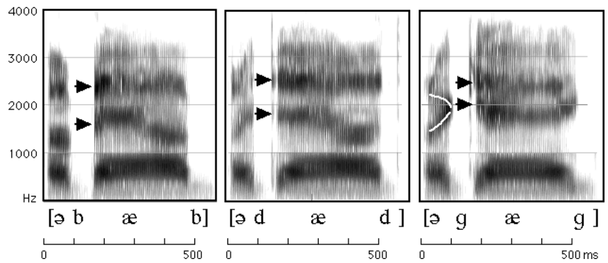
Figure from Jennifer Venditti

Vowel [i] sung at successively higher pitch.



Figures from Ratree Wayland slides from his website

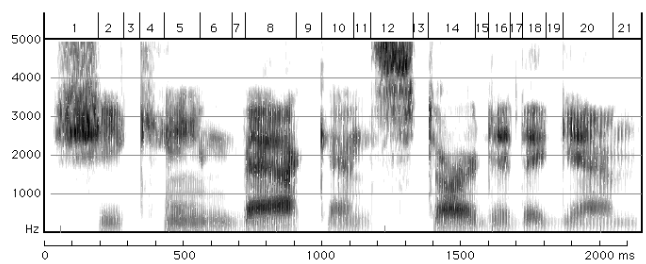
How to read spectrograms



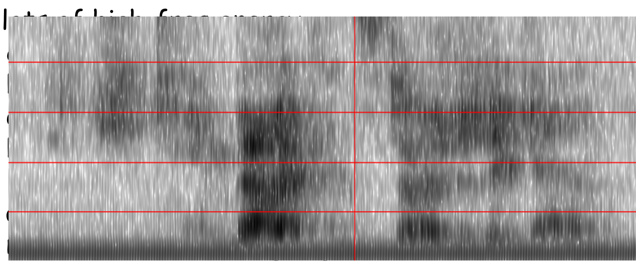
- bab: closure of lips lowers all formants: so rapid increase in all formants at beginning of "bab"
- dad: first formant increases, but F2 and F3 slight fall
- gag: F2 and F3 come together: this is a characteristic of velars. Formant transitions take longer in velars than in alveolars or labials

From Ladefoged "A Course in Phonetics"

She came back and started again



1.
3.
4.
5.
6.
7.
8. eginning
9.



From Ladefoged "A Course in Phonetics"