Vocal Tract Acoustics

R. D. Kent

Journal of Voice

1993

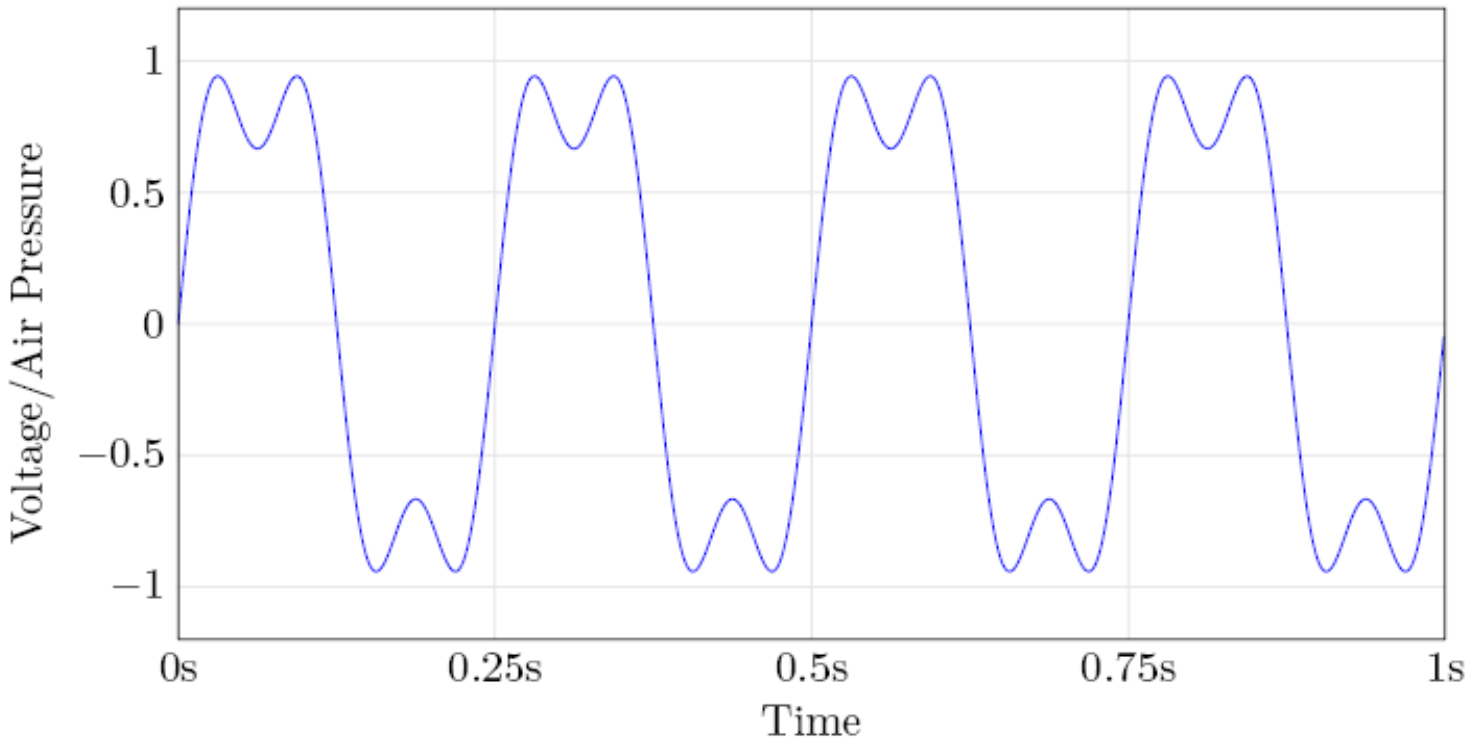
Presented by Daniel Felps

Motivation

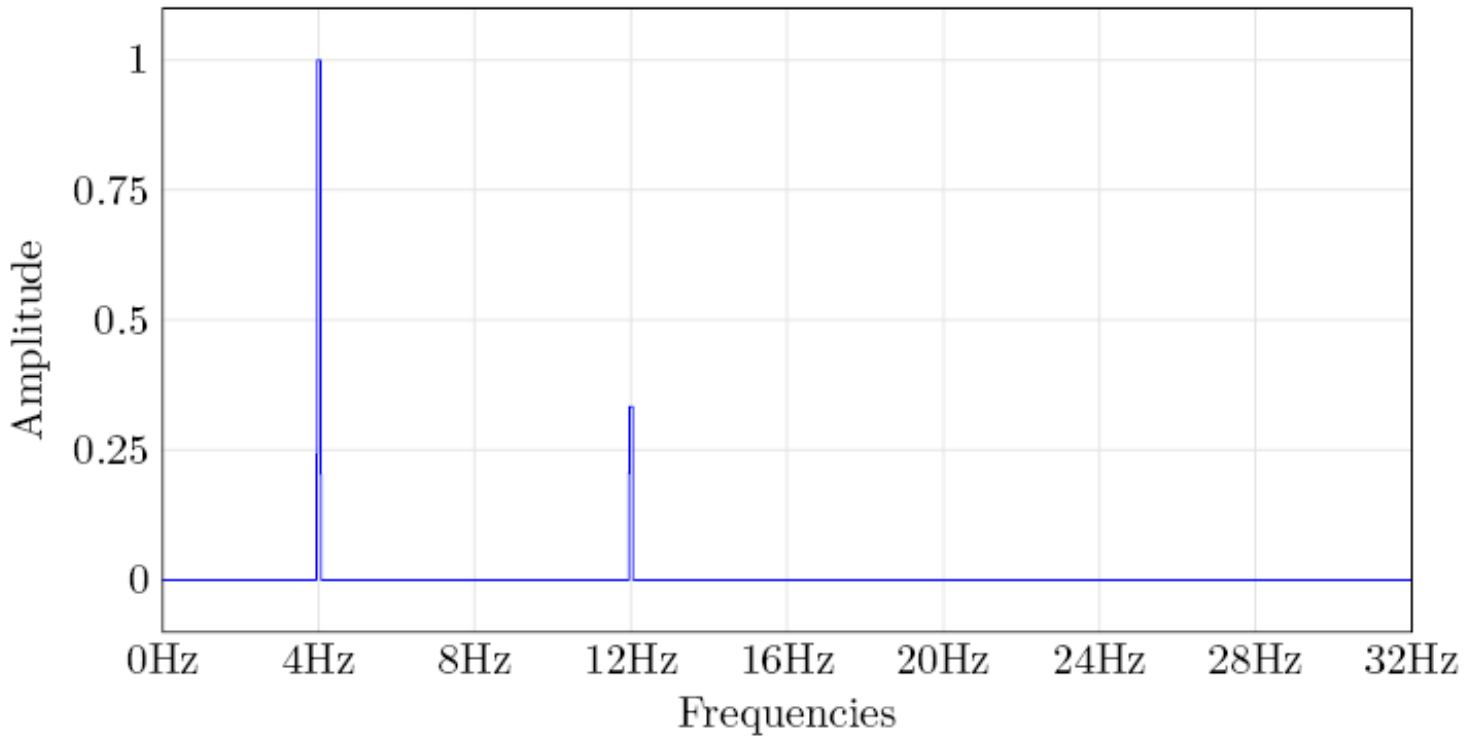
- This is an excellent paper to kick off speech recognition
  - High level
  - Overview of source-filter theory
  - It introduces many common terms in speech processing (pitch, formant, LPC, spectrograms)

Time domain

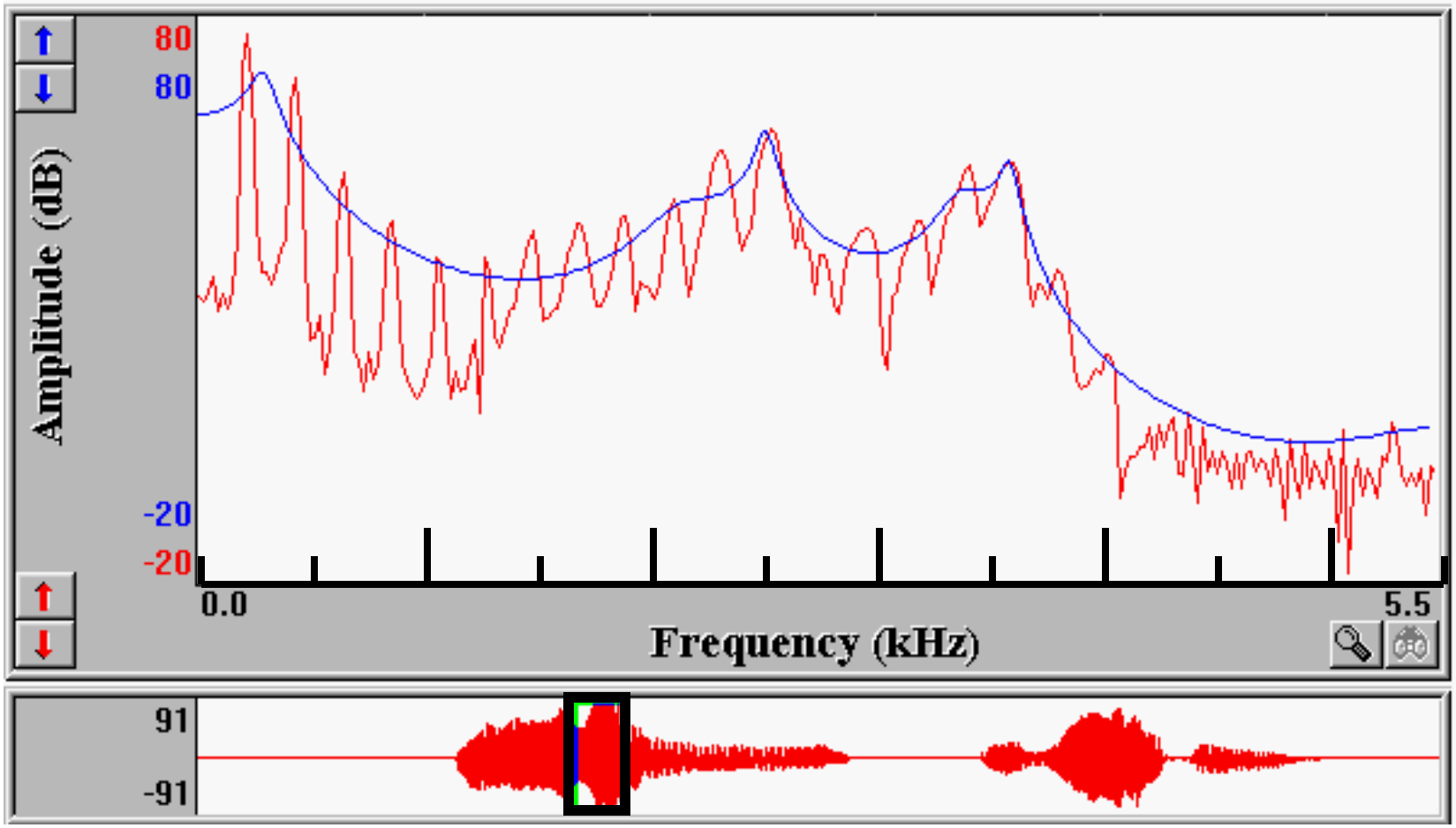
$$y(t) = \sin(4t) + \frac{\sin(12t)}{3}$$



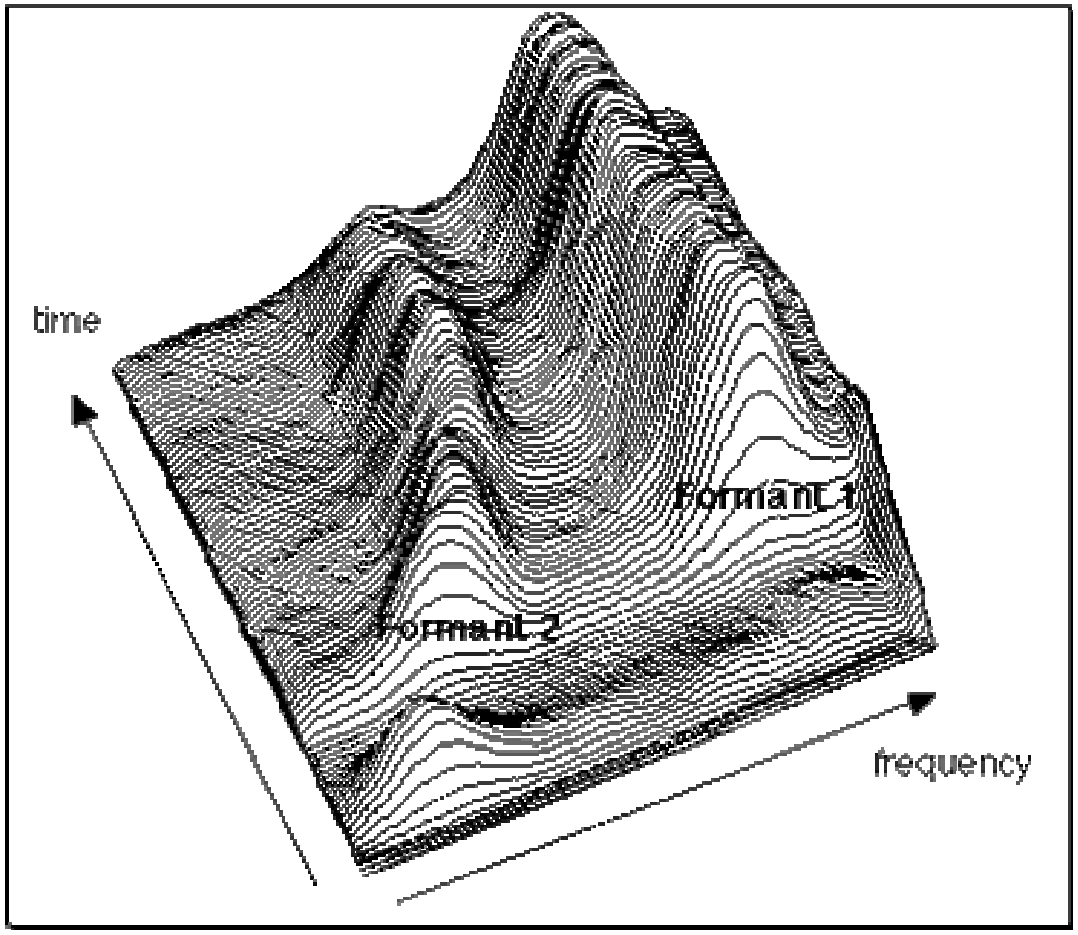
Frequency domain



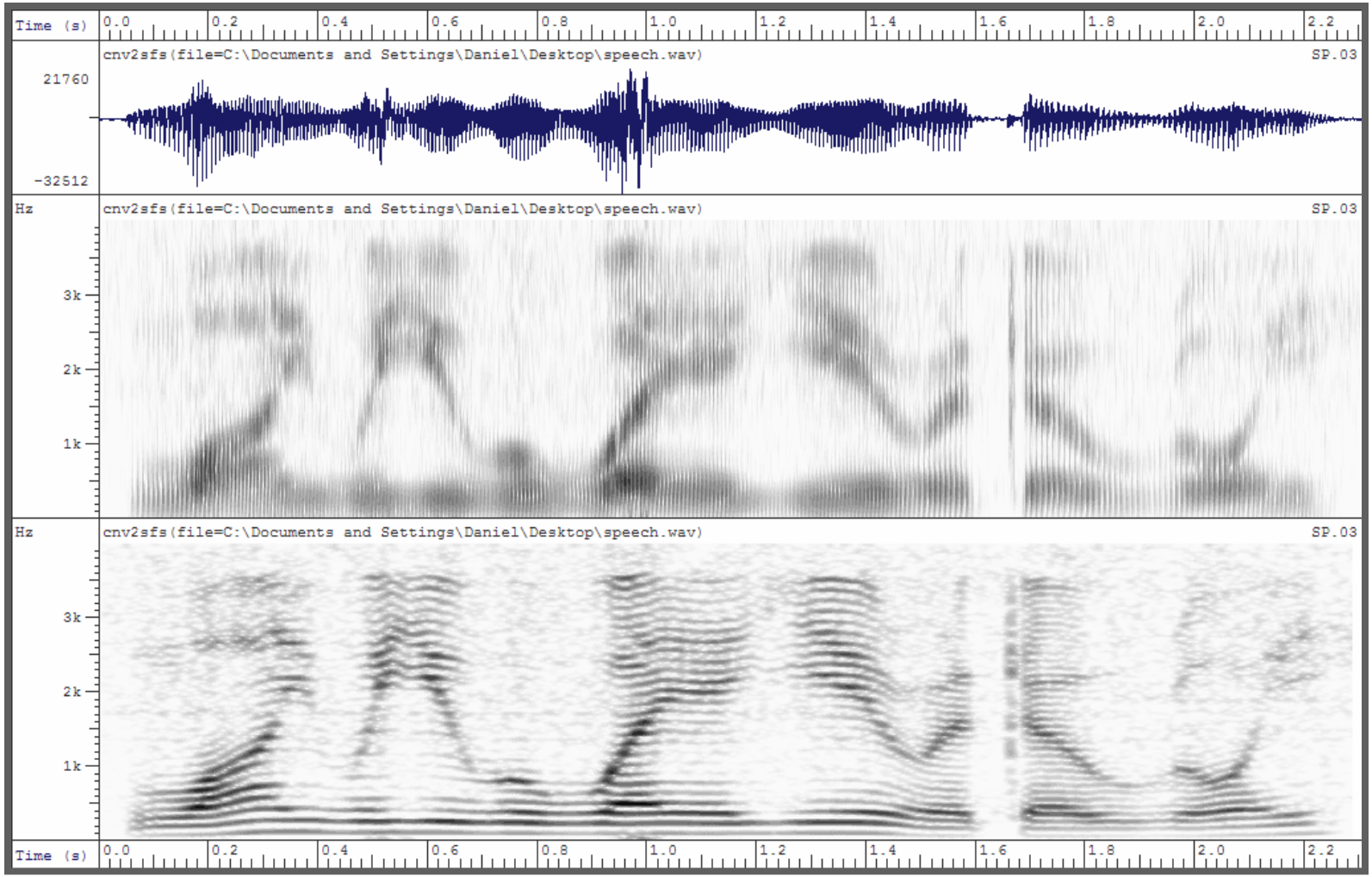
Laboratory instruments for speech analysis



Waterfall spectrogram



Wideband and Narrowband



Acoustic theory of speech production

- Source-filter theory proposed by Gunnar Fant in 1960
- Breaks speech into 2 parts
  1. Source
     - Laryngeal voicing
     - Turbulent noise
     - Transient
  2. Filter

Source-filter theory for vowels

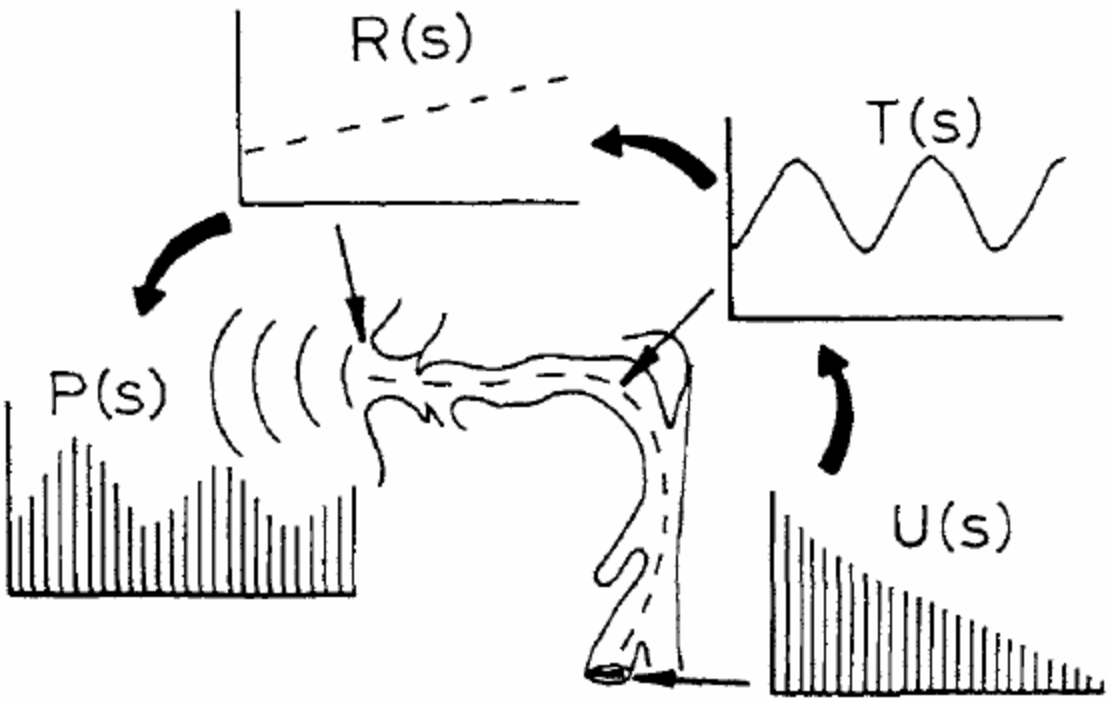


FIG. 2. Diagram of the vocal tract showing the affiliation of vocal tract regions with the major terms of the source-filter theory.

Source

- All vowels are voiced
- Periodic source

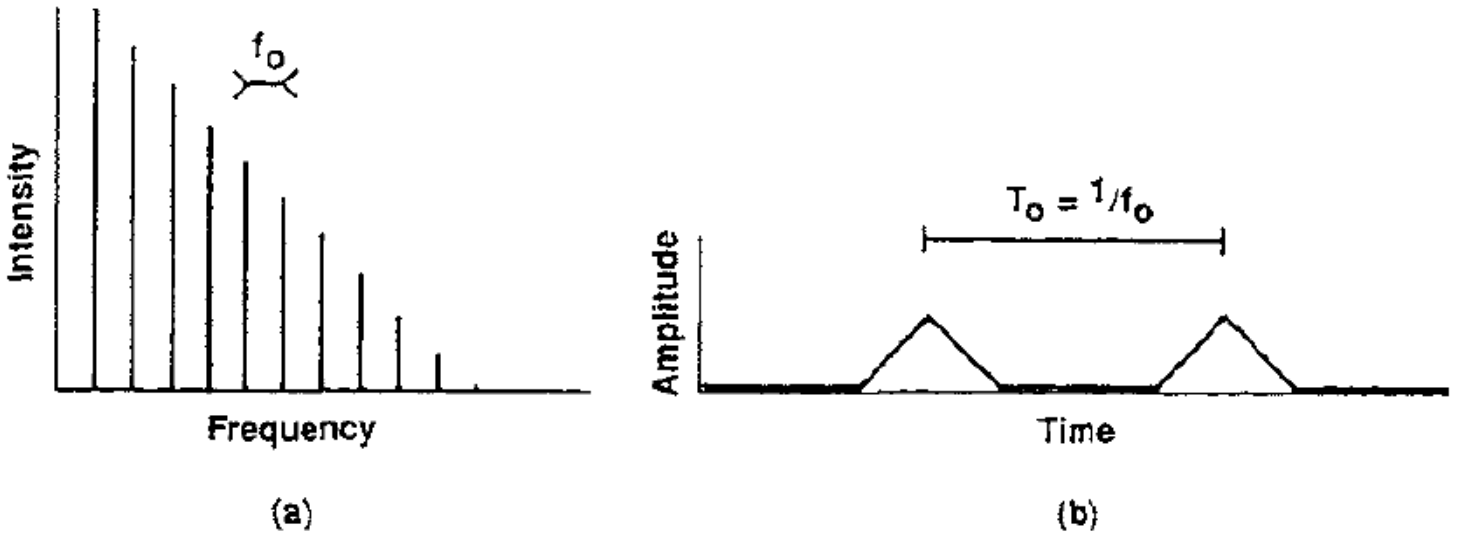


FIG. 3. **a** and **b**: Idealized form of the glottal spectrum, U(f), and the associated waveform, u(t).

Filter

- The filter is defined by the resonances of the vocal tract

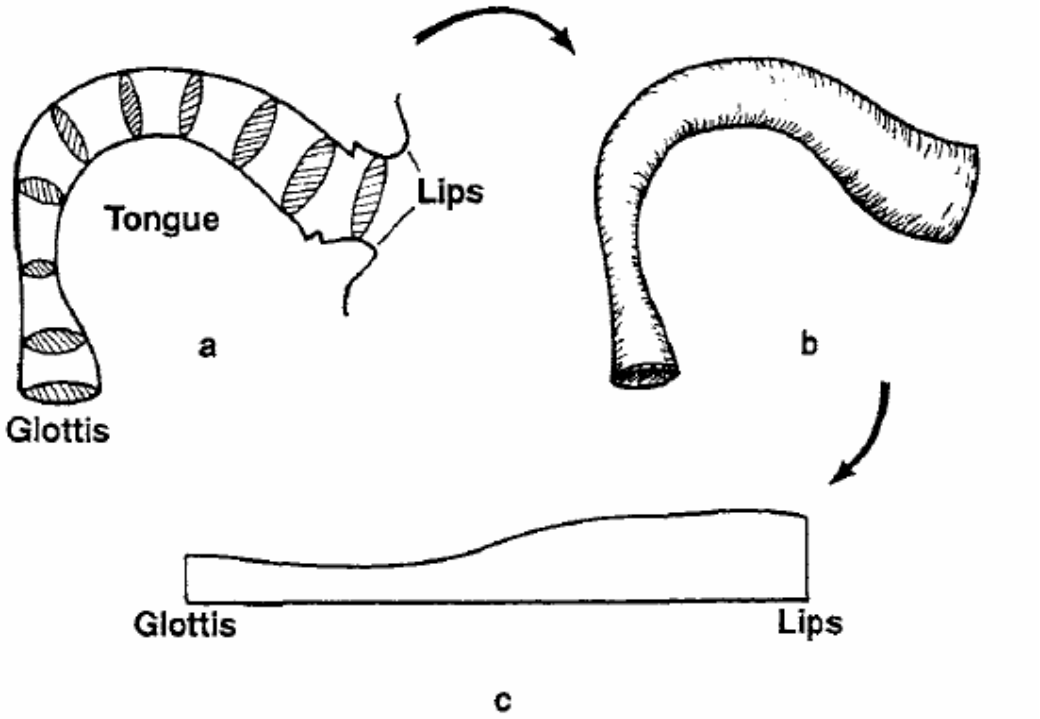


FIG. 4. Vocal tract area function shown as (a) curved vocal tract with selected points of cross-dimension measurement, (b) derived area function for a curved tube, and (c) area function for an equivalent straight tube.

Single tube resonances

$$F_n = \frac{(2n-1)c}{4l}$$

- Average male vocal tract is 17 cm long
- This makes speech recognition tough

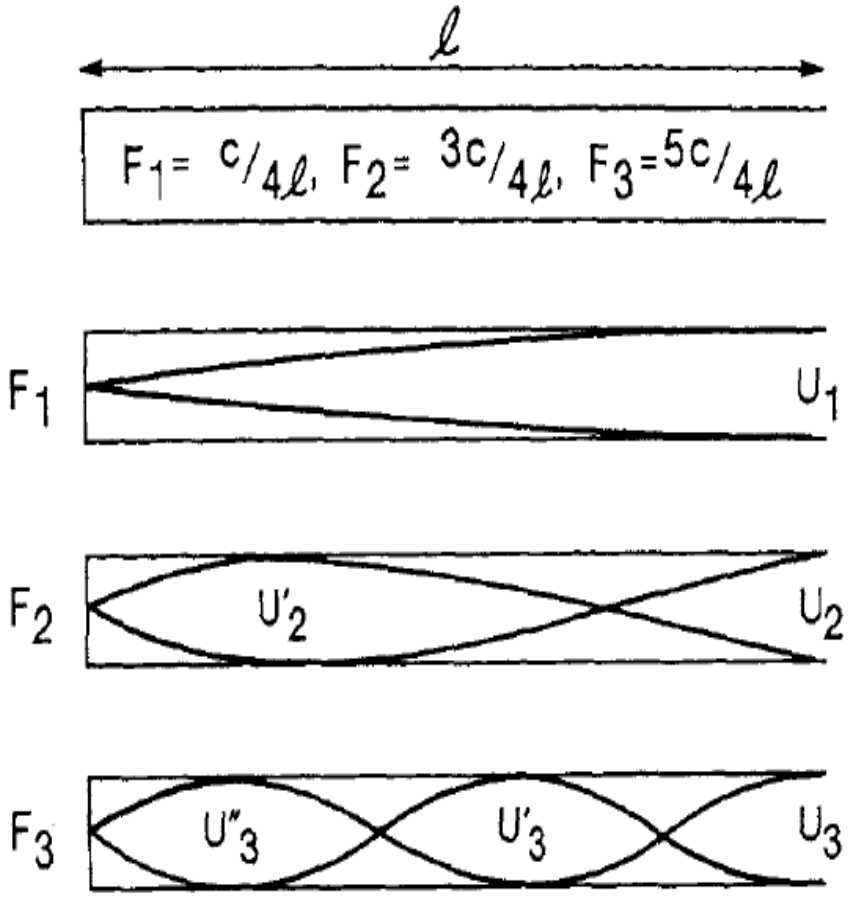
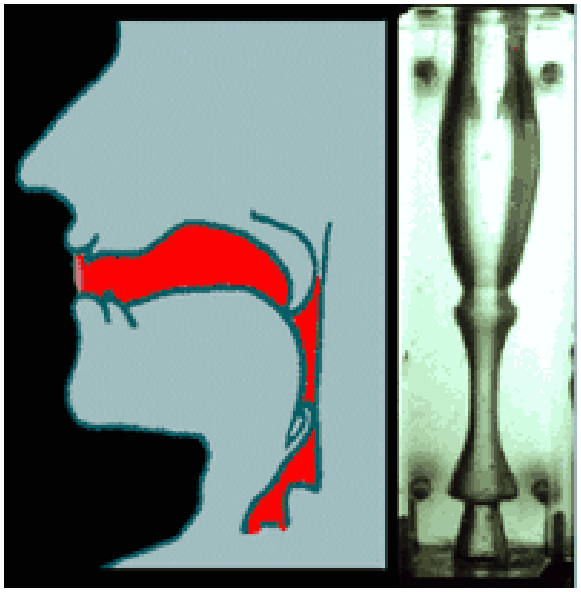


FIG. 5. Straight tube closed at one end (glottis) and open at the other (lips), showing stationary distribution of volume velocity for the first three formants, F1, F2, and F3. The resonances of the tube are given by the odd-quarter wavelength relationship (a tube of this configuration will resonate with maximal intensity to a sinusoid whose wavelength is four times the tube length).

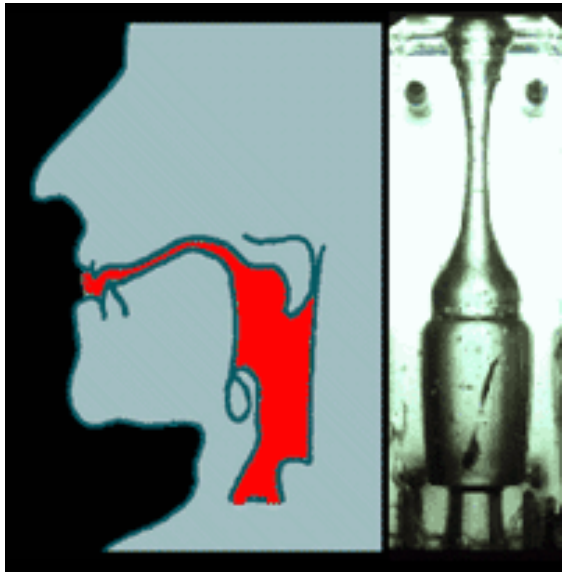
Duck Call

- How do they work?

AH



EE



Vowel formant patterns

- F1 frequency generally varies with the *up and down* tongue movement
- F2 frequency generally varies with the *front to back* tongue movement

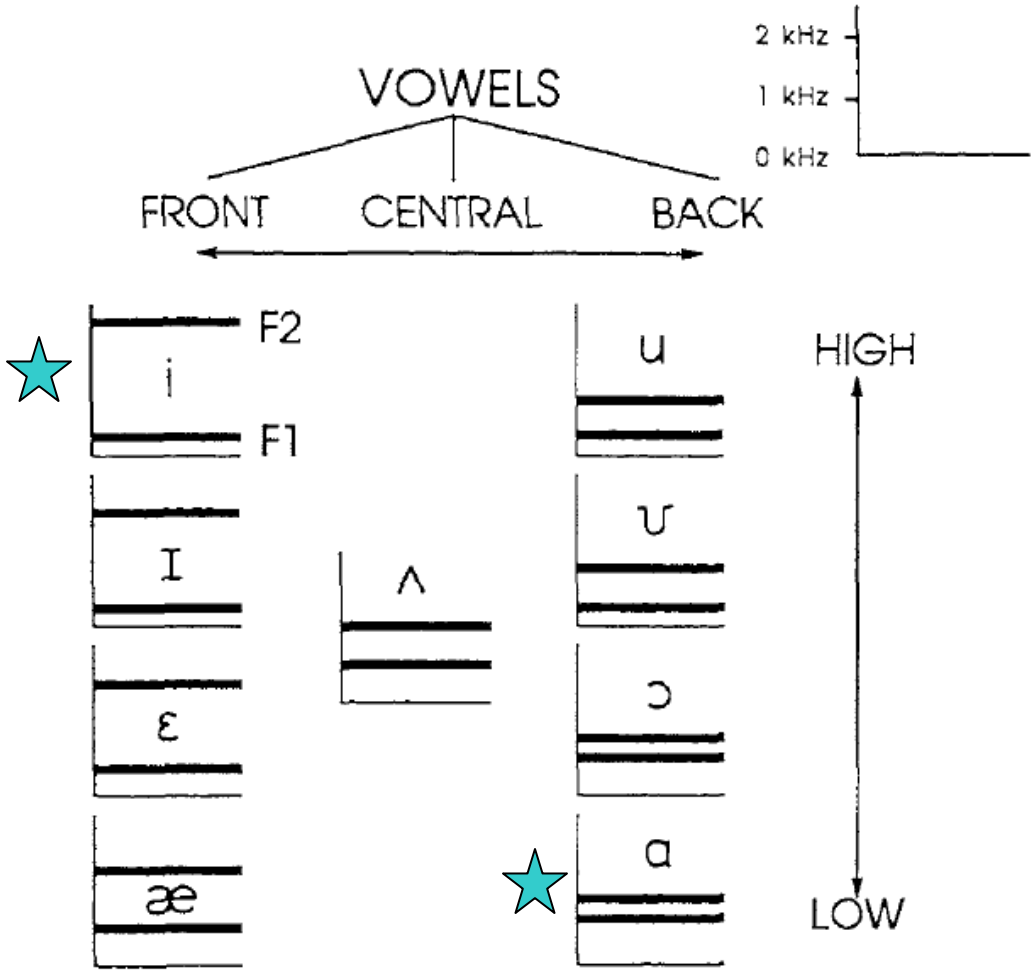


FIG. 6. Acoustic-articulatory relations for vowels. Front vowels are associated with a fairly wide F2–F1 separation, back vowels with a narrow F2–F1 separation. Therefore, F2–F2 separation correlates with advancement or retraction of the tongue. High vowels are associated with a low F1, low vowels with a high F1. Therefore, F1 frequency correlates with tongue height (or jaw opening). The effect of lip rounding, not shown, is to lower all formant frequencies. In English, only the back vowels and r-colored vowels are rounded.

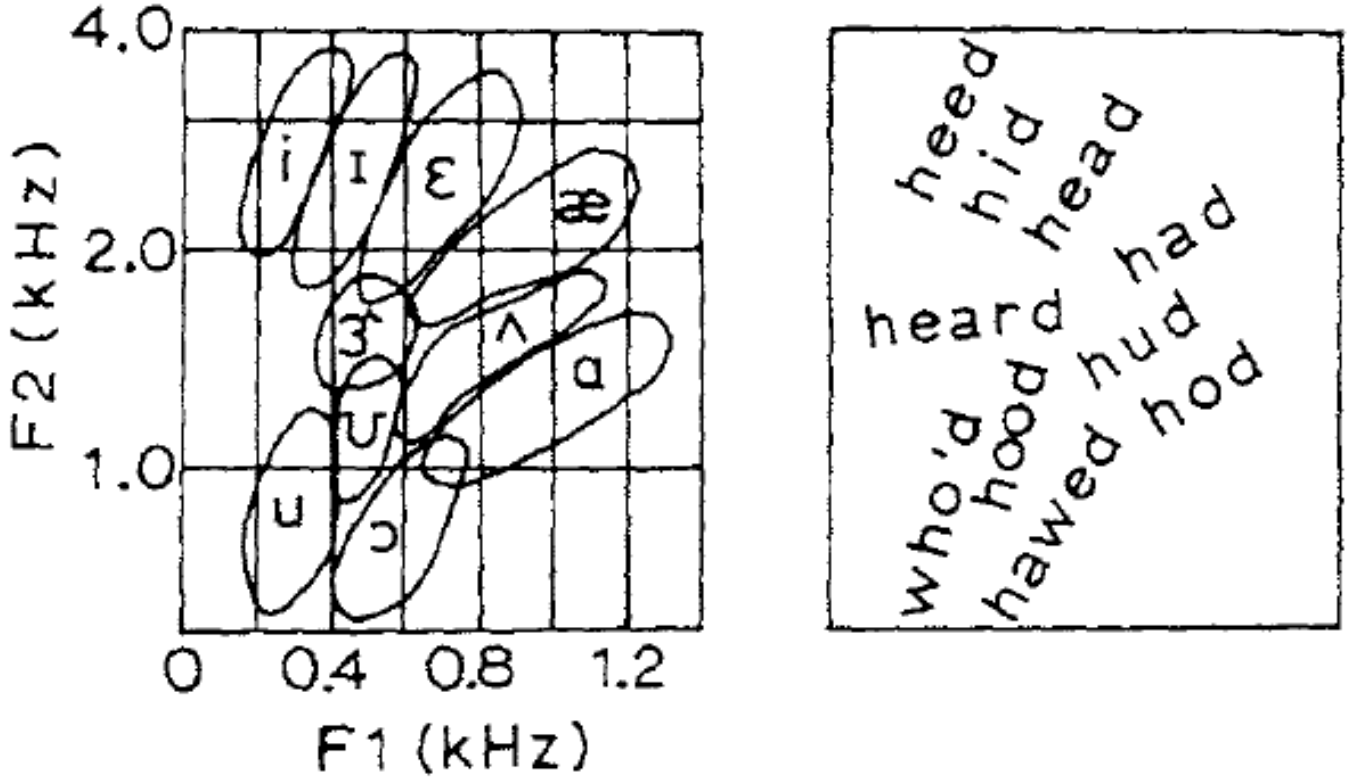


FIG. 7. Left: F1–F2 vowel chart with ellipses drawn to enclose the data for a large group of men, women and children. Values for men are at the end of the ellipses closest to the origin, values for women are close to the middle of the ellipses, and values for children are at the end of the ellipses farther from the origin. **Right:** The accompanying graph shows the approximate location of keywords for each vowel phonetic symbol shown in the ellipses in (a).

Relating vocal tract shape for vowels to acoustic output

- Constriction parameterization
  1. Size and location of constriction
  3. Ratio of mouth opening to length
  - A nomogram is graphical computation device (slide rule)

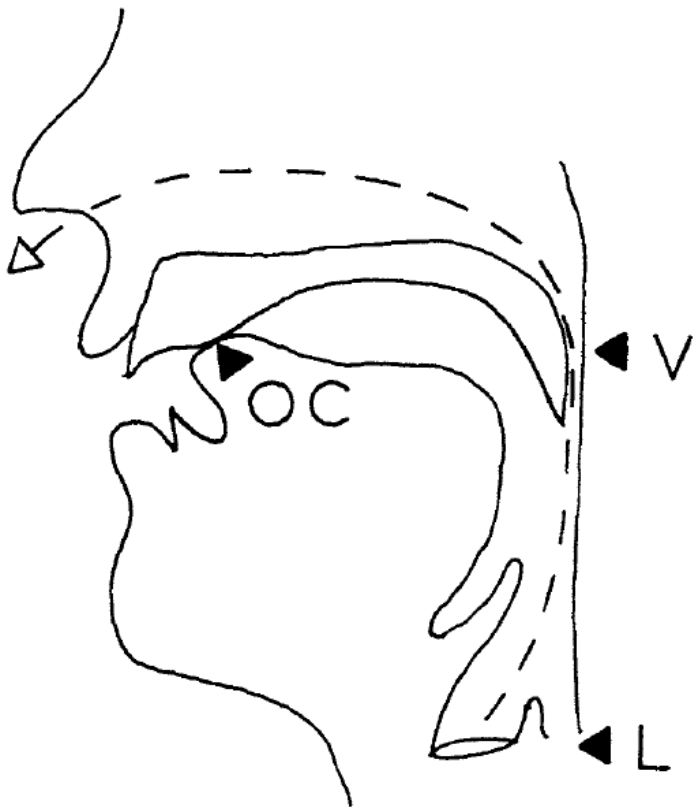
- Statistical relationship
  1. Tongue (2)
  3. Lip
  4. Jaw
  - I would guess these would be the first 4 principal components

- Articulatory relationship
  - Understand the way the tongue, lips, or jaw effect the acoustic signal
  - Quantal nature of articulation
    - Nonlinearities exist between vocal tract configuration and acoustic signal

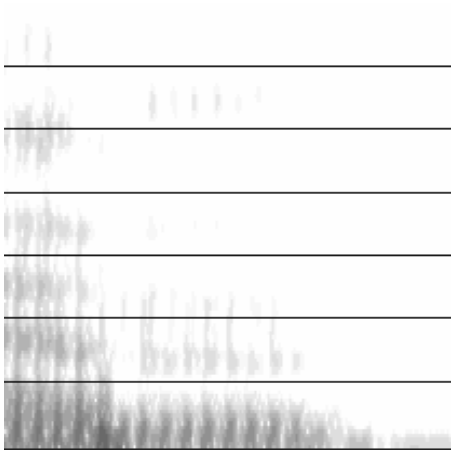
Source-filter theory for consonants

- Each category of consonants must be looked at individually
- Consonants have lower sound levels than vowels, but contribute significantly to intelligibility

Nasals /n/

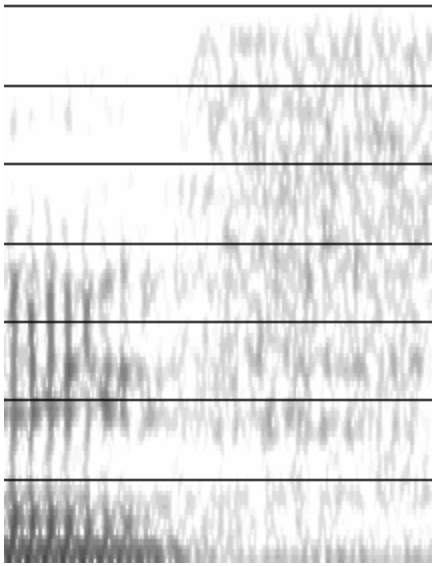


- Nasals involve blocking the mouth completely and letting the air come out of your nose
- Antiformants



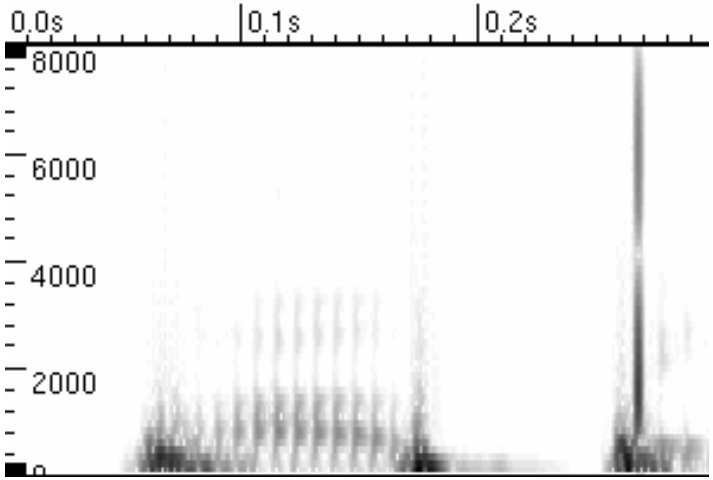
Fricatives /f/

- Fricatives involve letting the air slide through a narrow opening in the mouth
- Generate turbulence noise



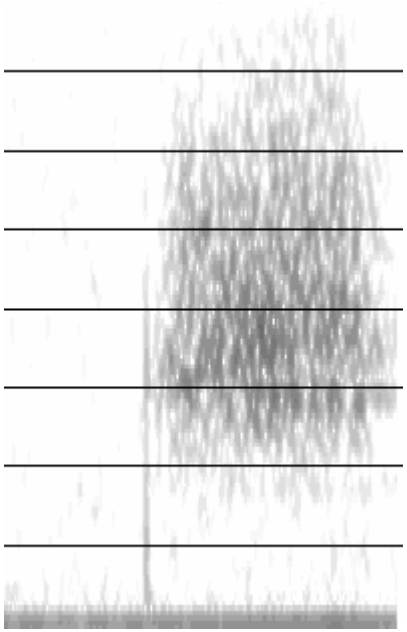
Stops /p/

- Stops must be described with cues
  1. Stop gap
  2. Release burst
  3. Formant transitions



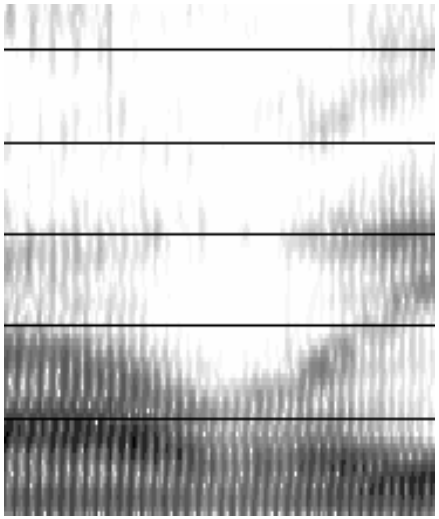
Affricates /tʃ/

- Affricates begin as stops and slide into fricatives, and hence are represented as a stop followed by a fricative



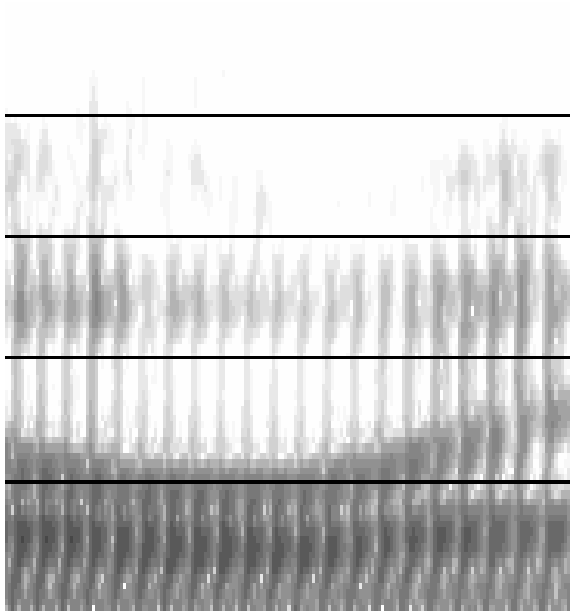
Liquids /l/

- Liquids are sometimes called "laterals" because of the sideways motion involved in producing them
- Resembles nasals and has antiformants



Glides /w/

- Also known as a semi-vowel
- Formant patterns change gradually



Acoustic measures of speech and voice

- Numerous features can be extracted from a speech signal
- Table 2 compares the abilities of techniques to extract certain measurements

Measurements

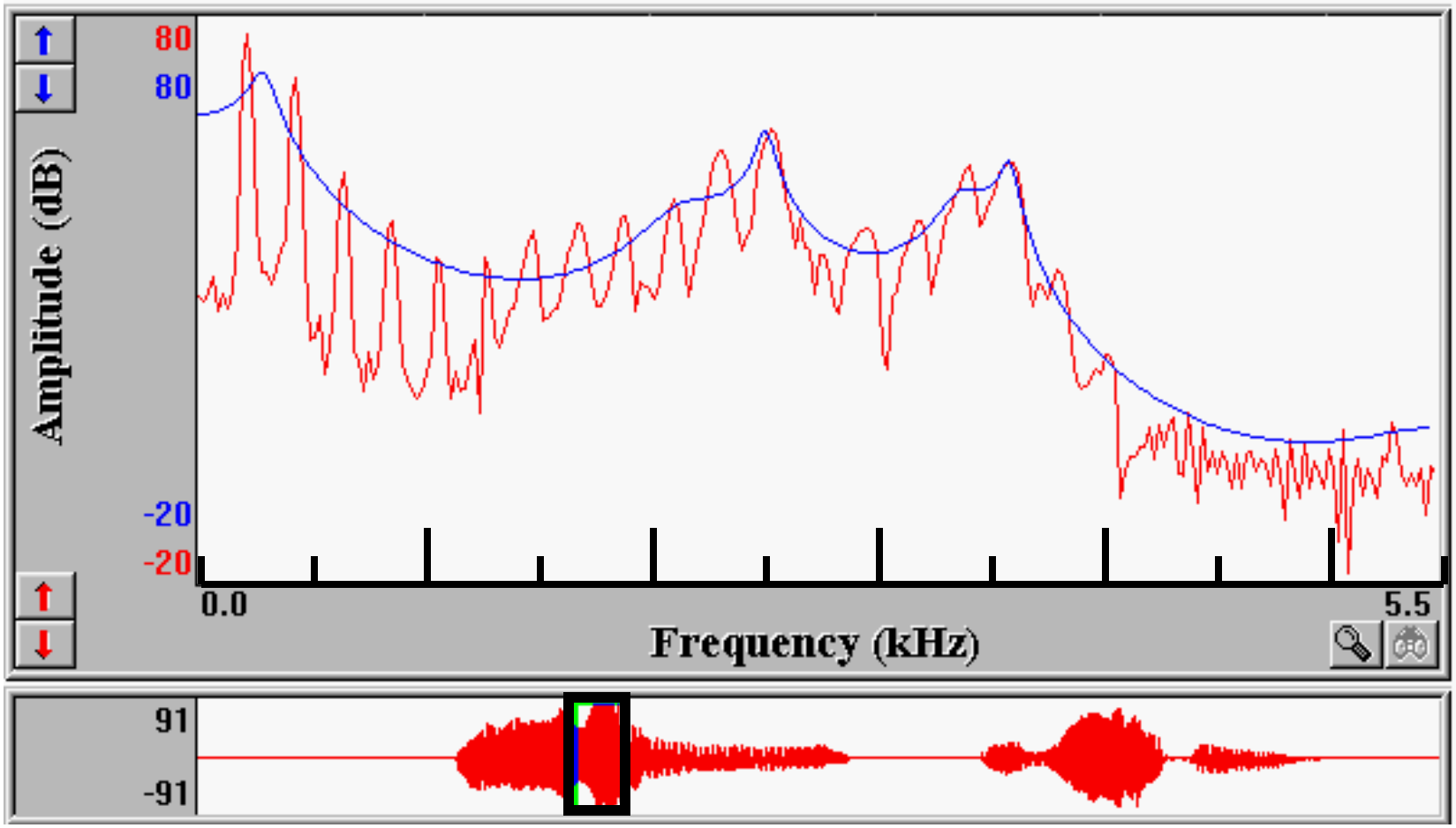
- *Voice onset time* is the length of time that passes between when a consonant is released and when voicing begins.
- *Voicing energy* is the ratio of the maximum amplitude value of a glottal cycle at the center of the fricative to the maximum amplitude value of a glottal cycle at the center of the following vowel.
- *Amplitude rise time* is the time between 10 and 90% of the peak amplitude.

- *Jitter* is the average absolute difference between consecutive periods, divided by the average period.
- *Shimmer* is the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude.

Prospects for automated, multidimensional analysis

- The paper gives the example of the difference in dysarthric speech
- We will see many more applications this semester

Still a mystery?



What can we tell?

- We know it is voiced since pitch harmonics are present
- The speaker is probably female, since the frequency of the pitch harmonics looks to be around 200
- Using Table 1, and the F1 and F2 values, we can guess the vowel and therefore the position of the tongue

Last slide

- Hopefully we better understand vocal tract acoustics from 3 perspectives
  1. Acoustic theory of speech production
     - Source-filter
  2. Methods for acoustic analysis
     - LPC, spectrogram
  3. Acoustic measures
     - Formants, pitch
- Any questions?