LING 576 Acoustic Phonetics
Spring 2009
Rod Casali
Topic number 8: Spectrogram Reading 1 - Identifying Sounds by Type
3-5-09

Reading:
Fujimura, Osamu & Donna Erickson. 1997. Linguistic phonetic descriptions. In The Handbook of Phonetic Sciences, ed. by William J. Hardcastle & John Laver, 65-115. Oxford, and Cambridge Massachusetts: Blackwell. Read pages 65-83.
Ladefoged, Peter. 2003. Phonetic data analysis. Malden, Massachusetts and Oxford: Blackwell. Read Chapter 6.

1. Overview

The ability to quickly identify various types of sounds in a spectrogram is a useful skill.

- Spectrograms play a central role in acoustic phonetics, and a high comfort level in reading them is an essential prerequisite for many types of acoustic investigation.

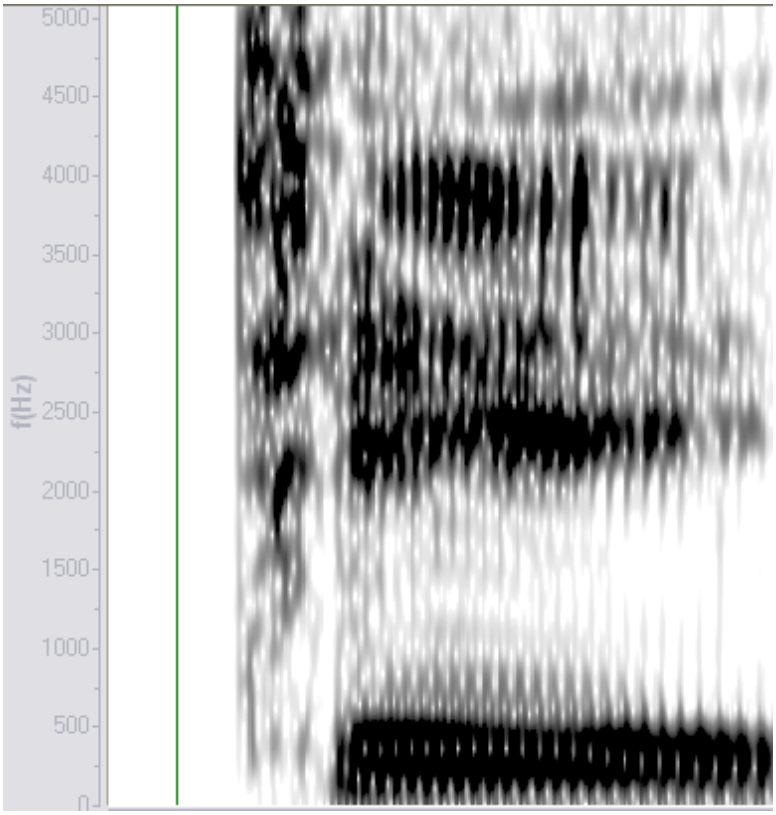
Spectrograms are an essential "tool of the trade."

- Correctly interpreted, spectrograms can provide valuable cues to the identity of speech sounds.

2. Oral vowels (review)

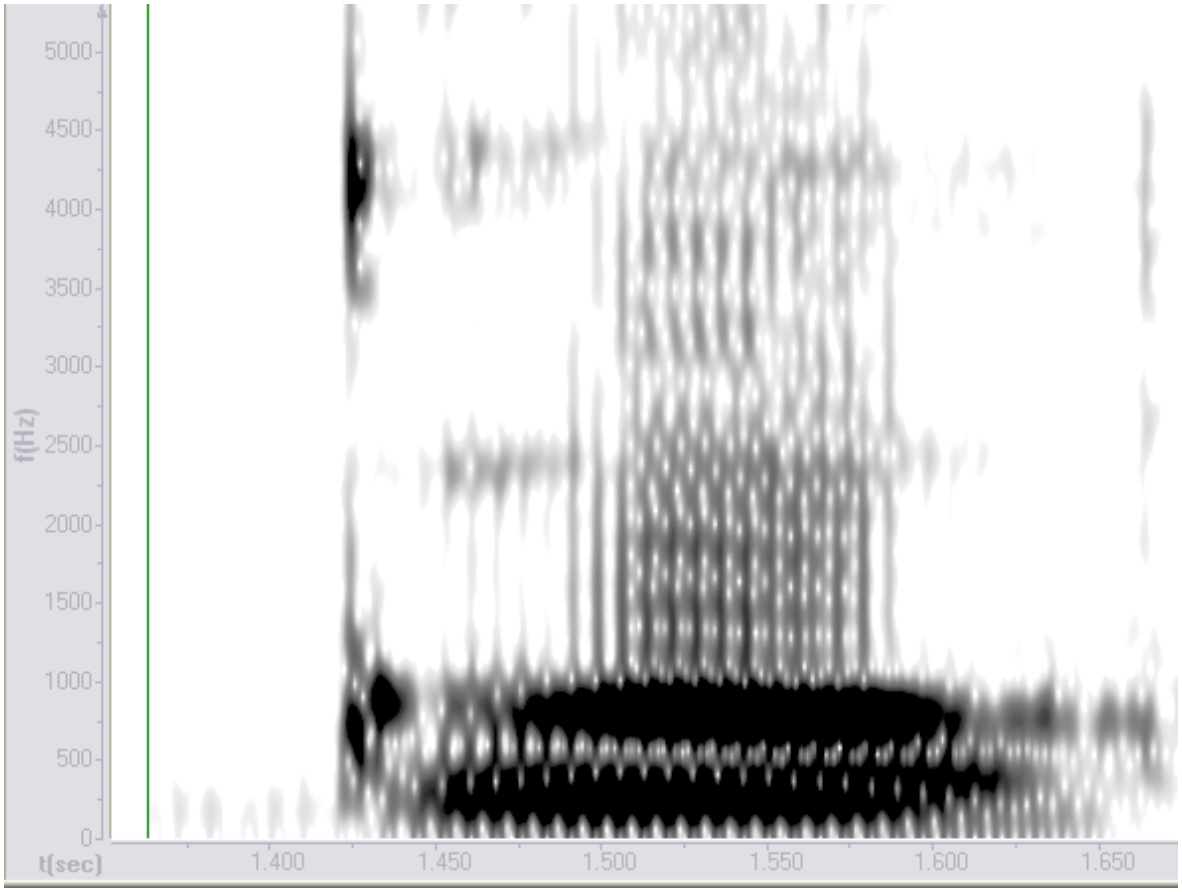
Non-diphthongal oral vowels show well-defined formants which persist in a reasonably steady state for some period of time. (Expect however to see changes--"transitions" in the vicinity of consonants.)

- Front vowels (especially high front vowels) show widely spaced F_1 and F_2.



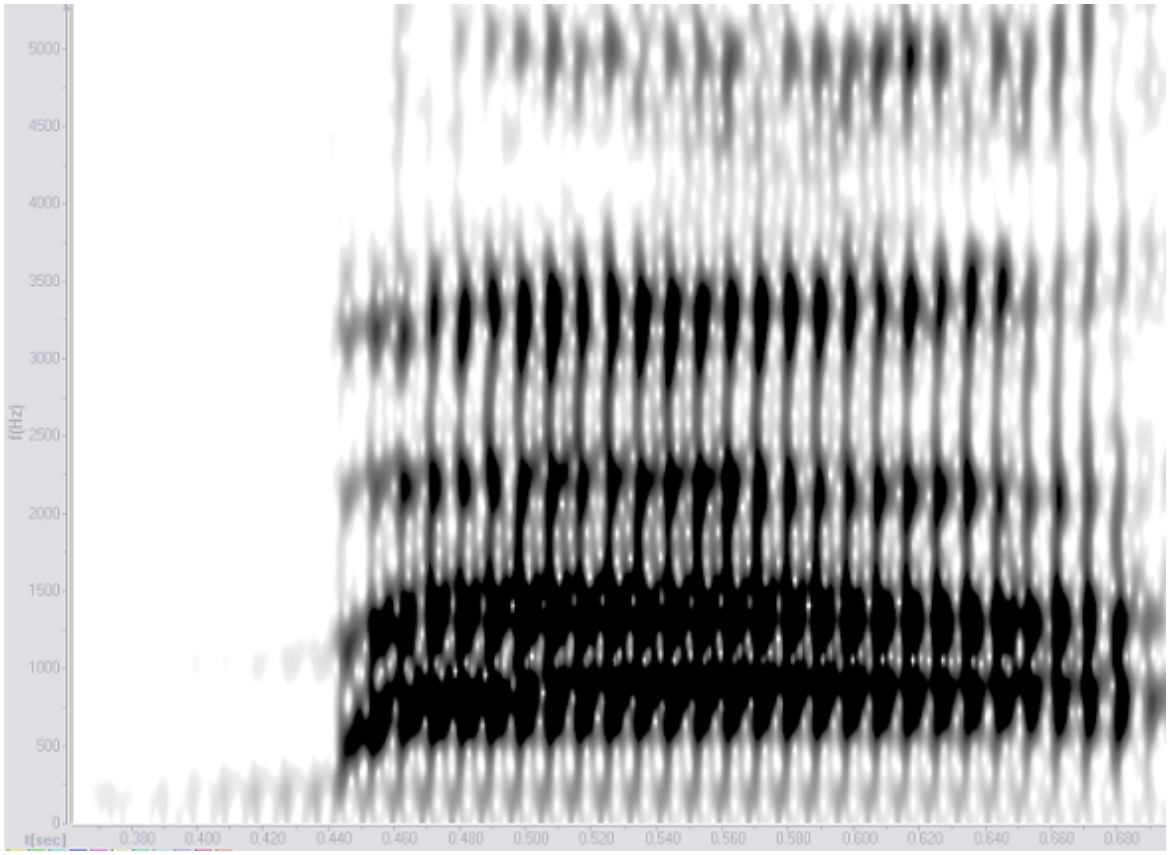
-- syllable [ti]

- Back round vowels (especially lower vowels) show F_1 and F_2 close together.



-- syllable [gu]

- The low central vowel /a/ shows very high F_1 (typically well above 500 Hz and sometimes above 900 Hz) and F_2 reasonably close to F_1 but clearly distinct from it.



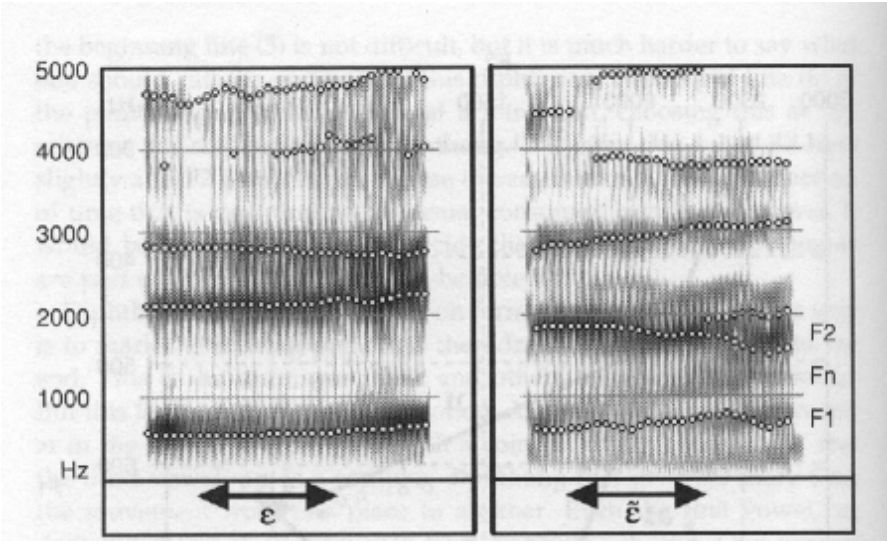
-- syllable [ba]

- Schwa-like mid central vowels have F_1, F_2, F_3 evenly spaced (about 1000 Hz) apart.

3. Nasal vowels

Nasal vowels are typically characterized by:

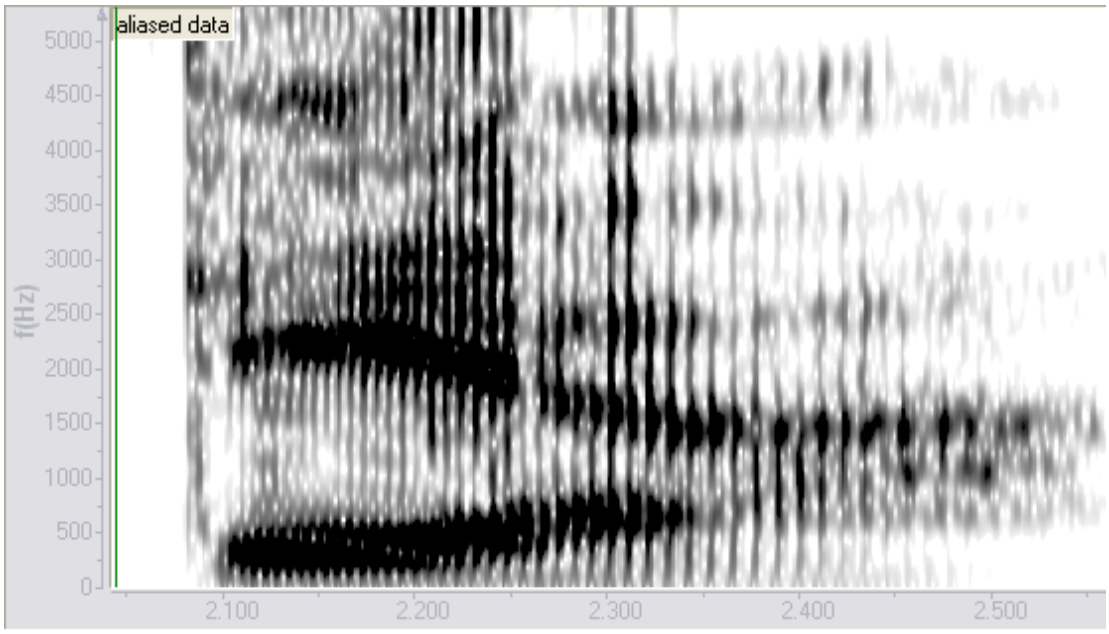
- Weakening of F_1.
- An additional "nasal" formant (sometimes around 1000 Hz).
- Lowering of some formant values.
- Increased formant bandwidth.



-- From Ladefoged reading, p. 136

4. Diphthongs

Diphthongs are characterized by significant changes in one or both of the first two formants during the vowel segment.



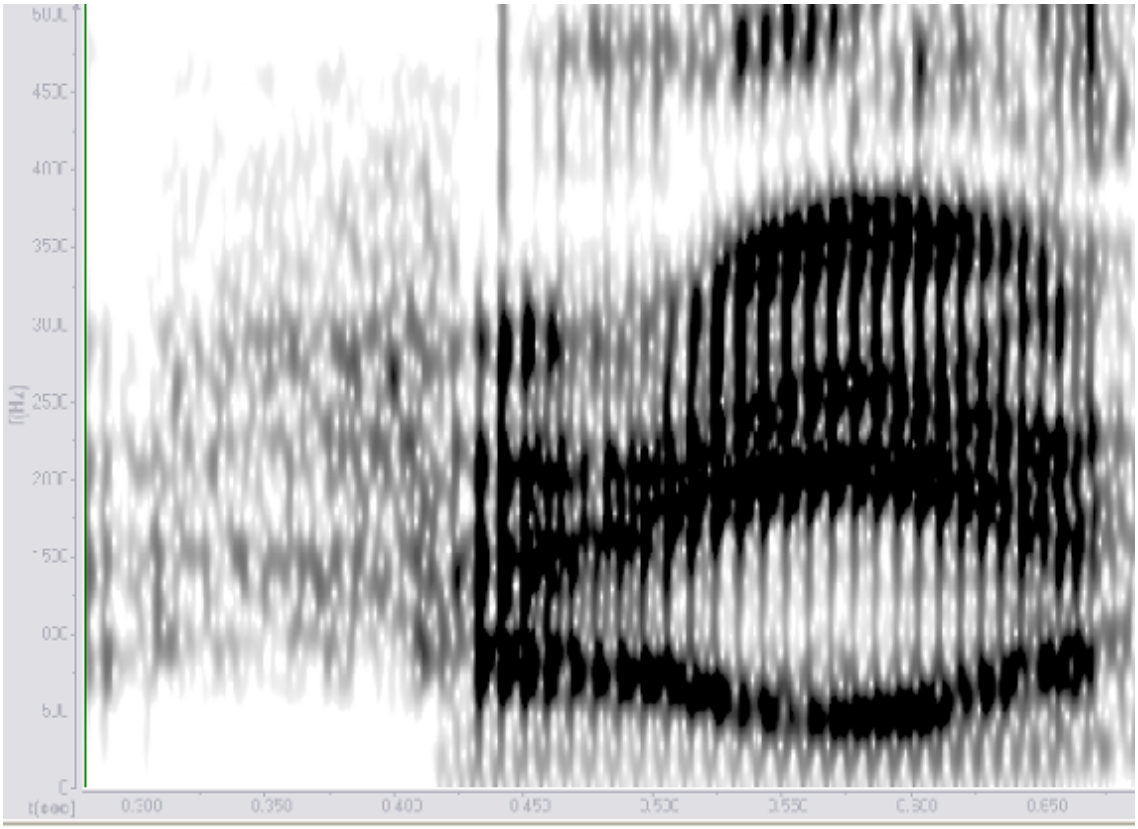
-- sequence [tia]

5. Approximants

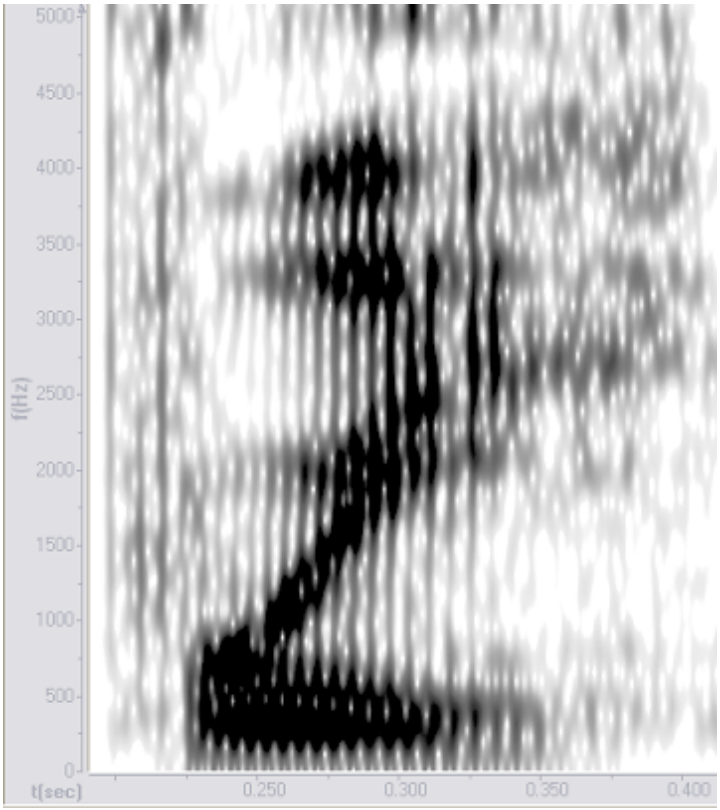
Approximants are essentially vocoids (vowel segments) of short duration, sometimes with a greater degree of stricture (narrower approximation).

Sometimes, a greater level of stricture leads to more extreme formant values (lower F_1 in [j], lower F_2 in [w]) than is typical for the corresponding vowels.

Spectrographically, they show formants similar to those of vowels but with significantly shorter steady state portions. (In some cases, the formant changes may be changing constantly throughout.)



-- word [haja]



-- syllable [pwɪ]

6. Nasals

Like vowels, nasals have visible formants.

Differences:

- Lower intensity.
- Different characteristic formant values.

These different values entail that there are typically sharp discontinuities in formants at the boundary between a nasal consonant and preceding or following vowel.

All nasals typically have a formant in the vicinity of 200 Hz and another above 2000 Hz.

There will typically be an additional formant somewhere between 1,000 to 2,000 Hz as well. (This will be lowest for [m], slightly higher for [n], highest for [ŋ].)

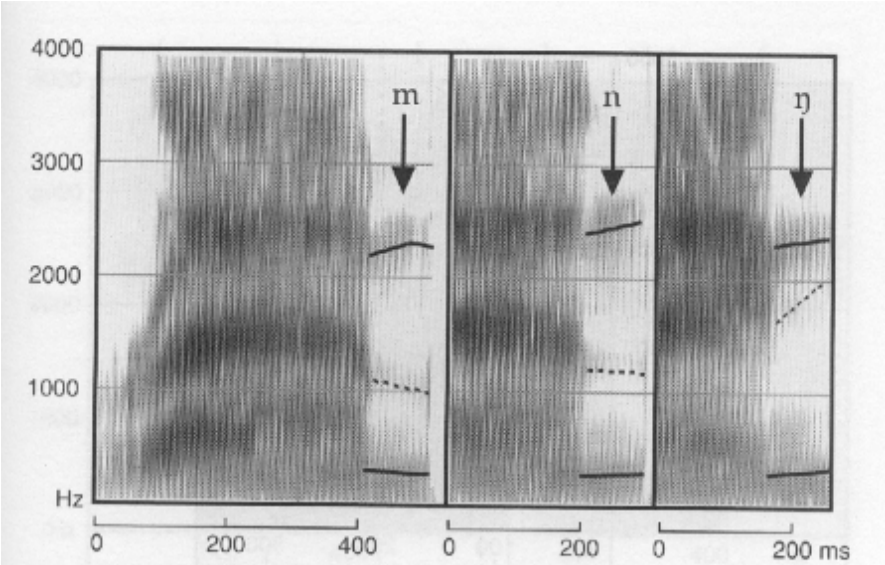


Figure 6.6 The nasals at the ends of the words *ram, ran, rang*. Only the ends of the last two words are shown.

-- from Ladefoged reading, p. 145

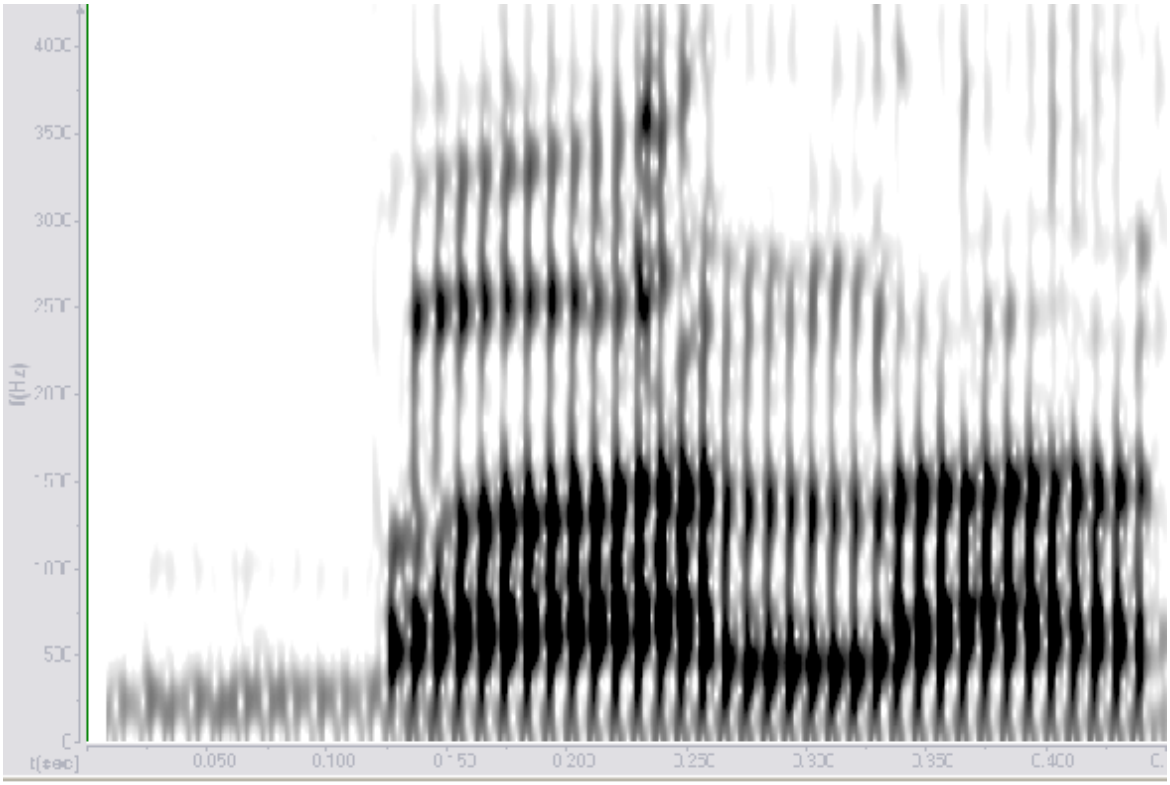
7. Lateral [l]
Laterals looks similar to nasals in spectrograms.

Characteristics:

- Vowel-like formants, but with weaker intensity.

Formant intensity may however tend to be greater than in nasals?

- Sharp discontinuity with respect to neighboring vowels in formant locations.
- F_2 around 1200 Hz (considerably lower if velarized--see Ladefoged reading p. 147).



-- sequence [bala]

8. Rhotics

Rhotics ("r-sounds") come in several varieties:

- Alveolar approximant (English r)

This is quite rare cross-linguistically.

- Alveolar flap / tap [ɾ]
- Retroflex tap / flap
- Alveolar trill
- Uvular trill

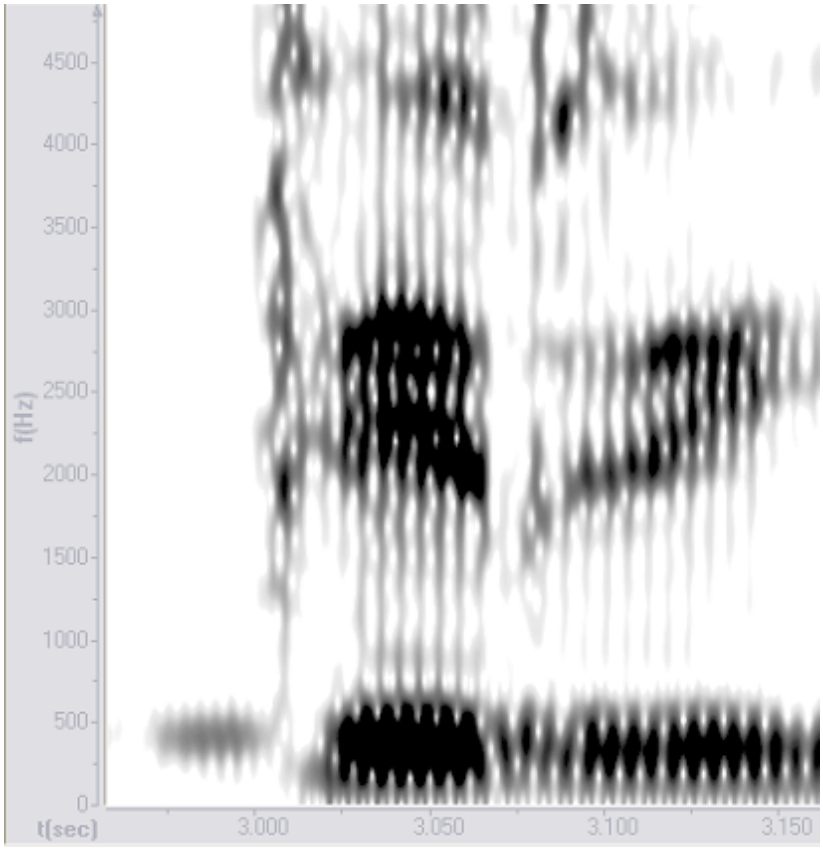
These sounds differ greatly in their articulatory properties.

What most / all of them have in common acoustically is a lowered F_3. (For at least some rhotics, F_2 may be lowered also.)

Perhaps the most common rhotic sound is an alveolar flap / tap.

Acoustically, this looks like a very short stop (silent or near-silent interval).

Sometimes weak formants are visible also.



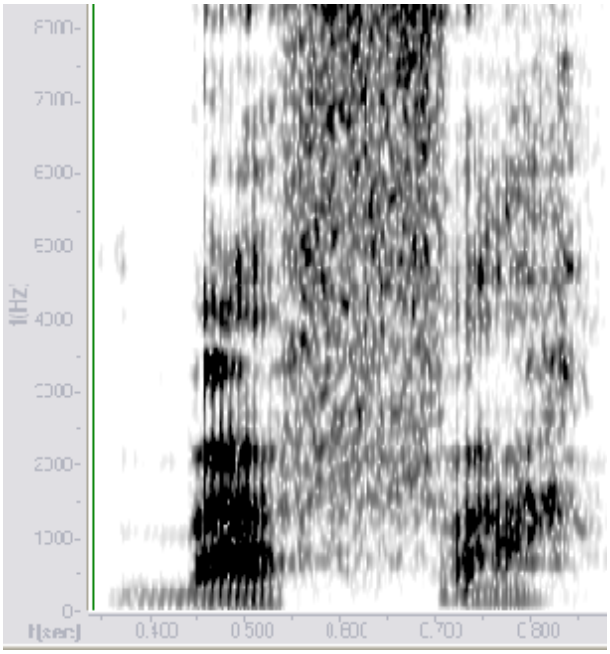
-- sequence [diɾi]

An alveolar trill will look spectrographically like a succession of taps / flaps interrupted by very short vowel-like intervals. (See Ladefoged reading p. 152.)

9. Fricatives

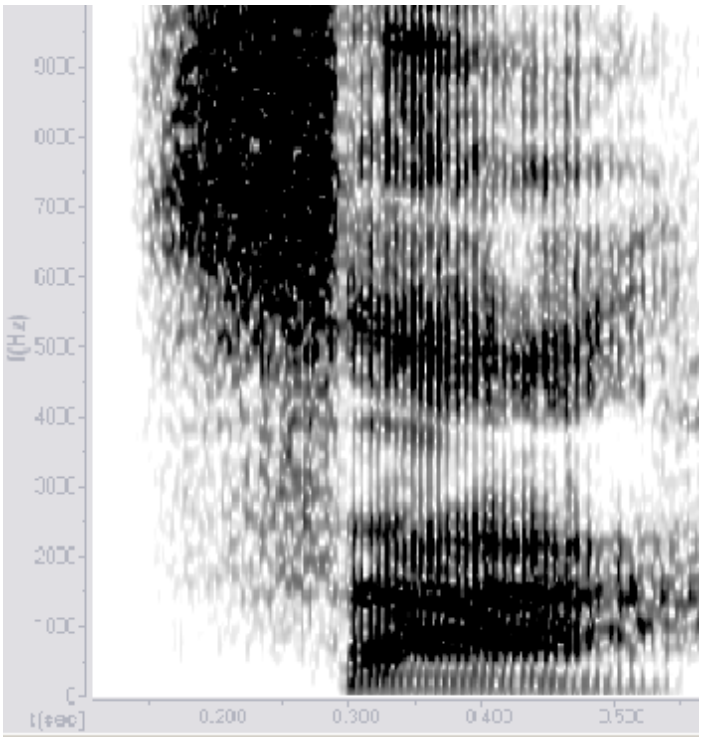
Fricatives show aperiodic energy over a range of frequencies.

Non-strident fricatives such as [f] and [θ] are spectrally diffuse; energy is spread out over a wide range of frequencies.



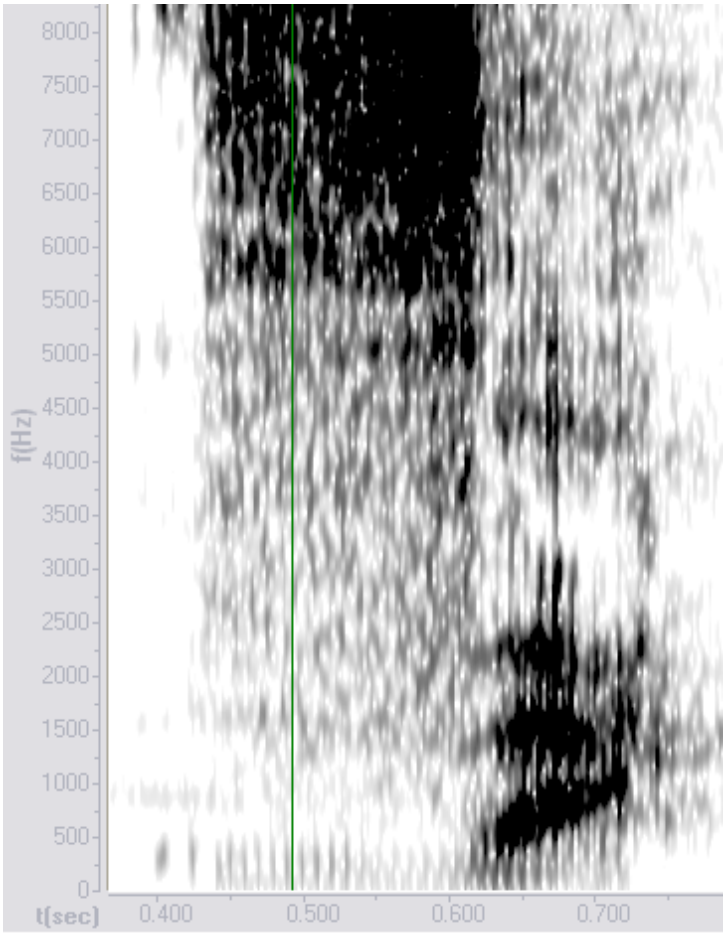
-- sequence [bafa]

Strident fricatives, e.g. [s], [ʃ]. have concentrations of energy in particular frequency ranges. It is possible to identify a center frequency of the spectrum, which will be higher for [s] than [ʃ]. For [s], most of the intense spectral energy is often above about 5000 Hz.



-- syllable [sa]

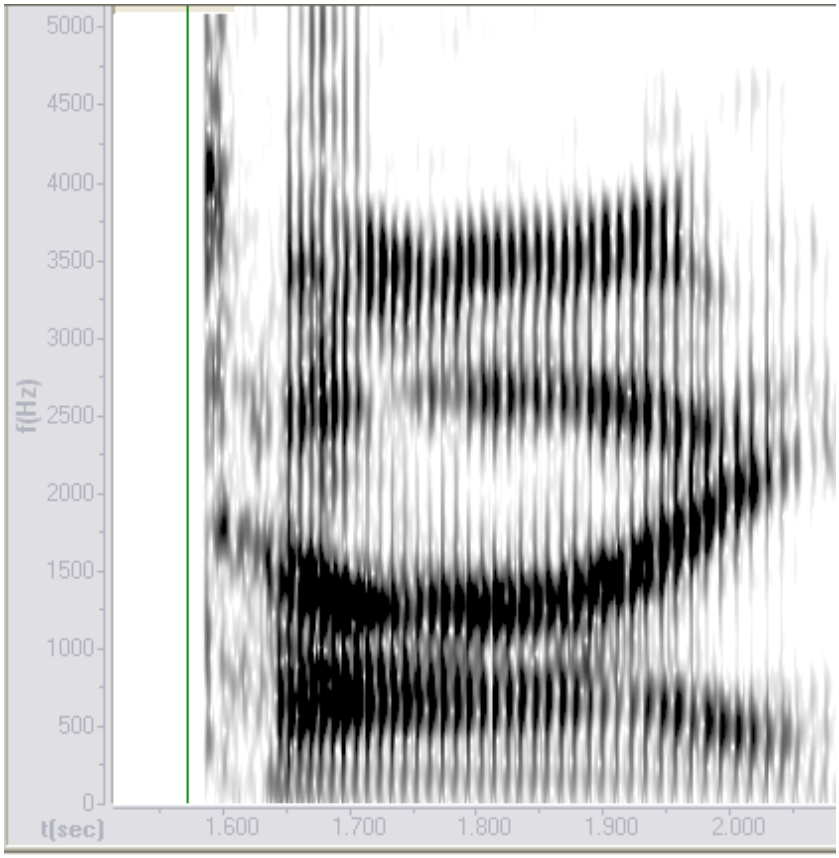
Voiced fricatives show a low frequency "voicing bar" in addition to the frication noise.



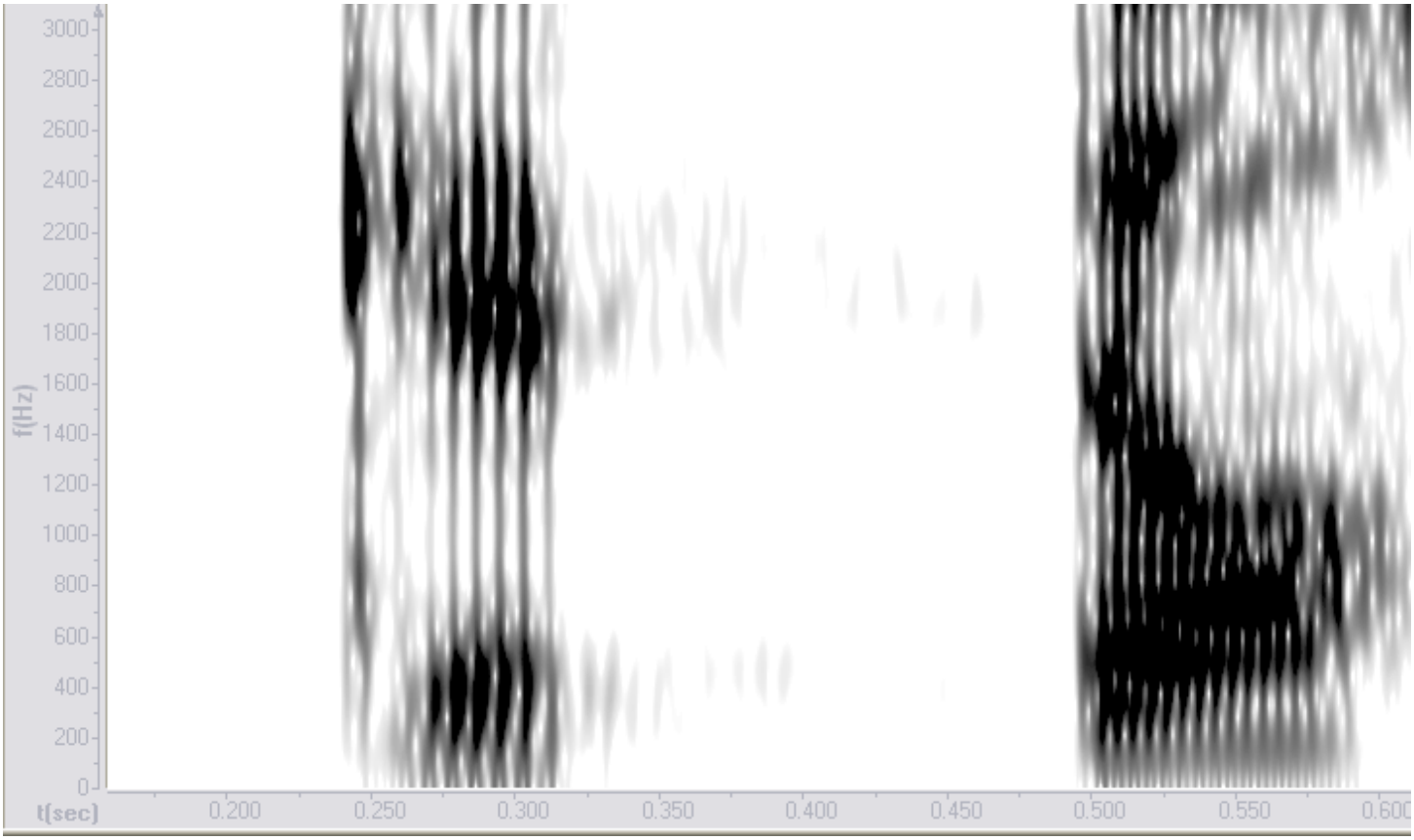
-- syllable [za]

10. Stops

Voiceless stops consist primarily of an interval of silence, in which no energy is visible at all. This is typically followed by a release burst or transient. There may also be a period of aspiration before voicing begins in a following vowel.

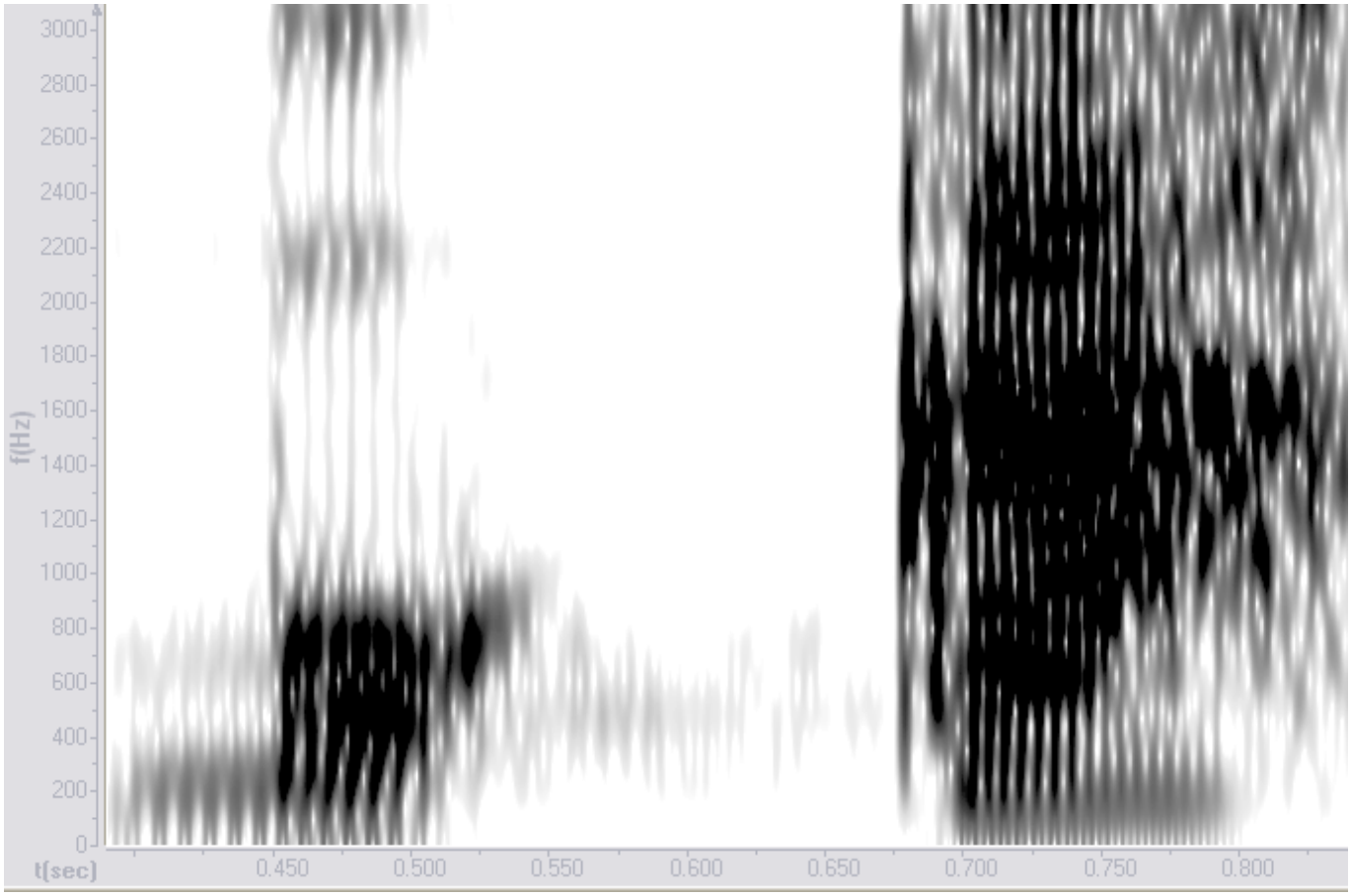


-- English word *tie*



-- Tafi word [kɪtɔ]

Voiceless velar stops often have a double release burst.



-- Tafi word [bʊka]

In pre-pausal position, voiceless stops in many languages are often unreleased.

Voiced stops show a release burst as well. (Aspiration on voiced stops is rare.)

Voiced stops also show a voicing bar. (See the word-initial [b] in the Tafi example [bʊka] above.)

11. Complex segments

In general, complex segments look spectrographically like a combination of the relevant simple segments, e.g., affricates look (as one would expect) like a stop followed by a fricative, prenasalized stops look like a nasal followed by a stop, etc.