Acoustic phonetics & Speech perception

http://www-rohan.sdsu.edu/~gawron/functions_of_language

Jean Mark Gawron

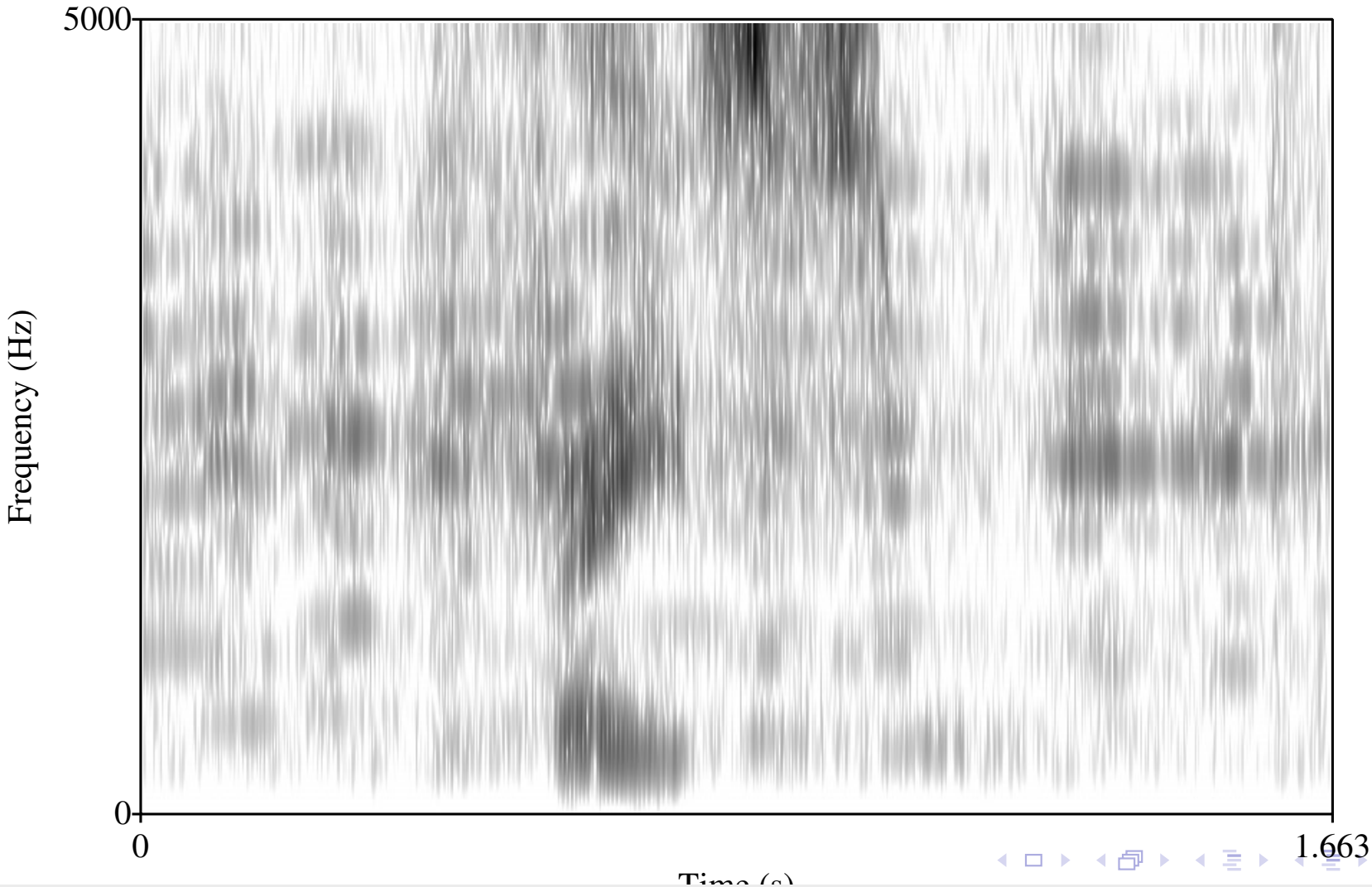
San Diego State University, Department of Linguistics

2012-01-25 Ling 525

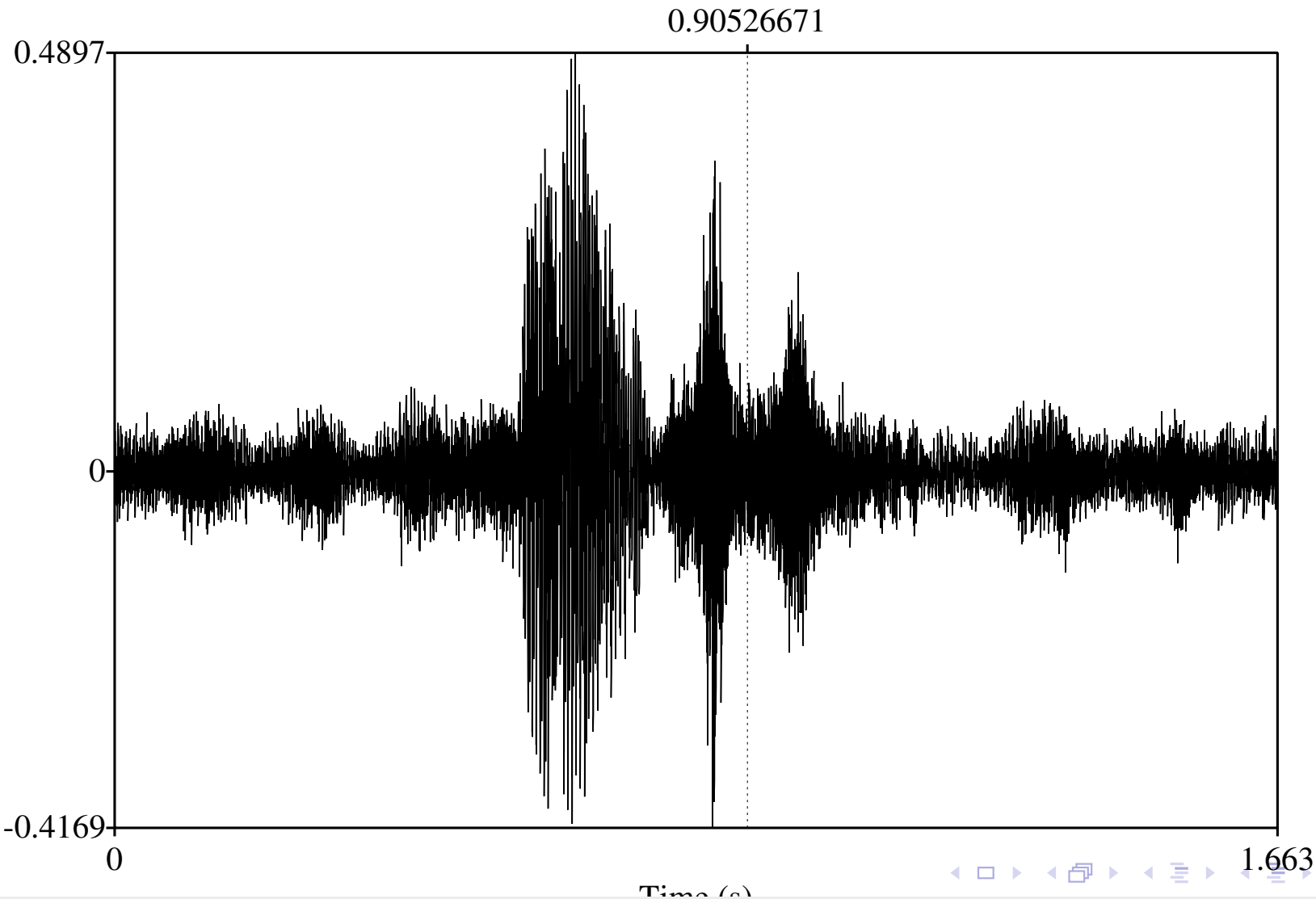
Outline

Outline

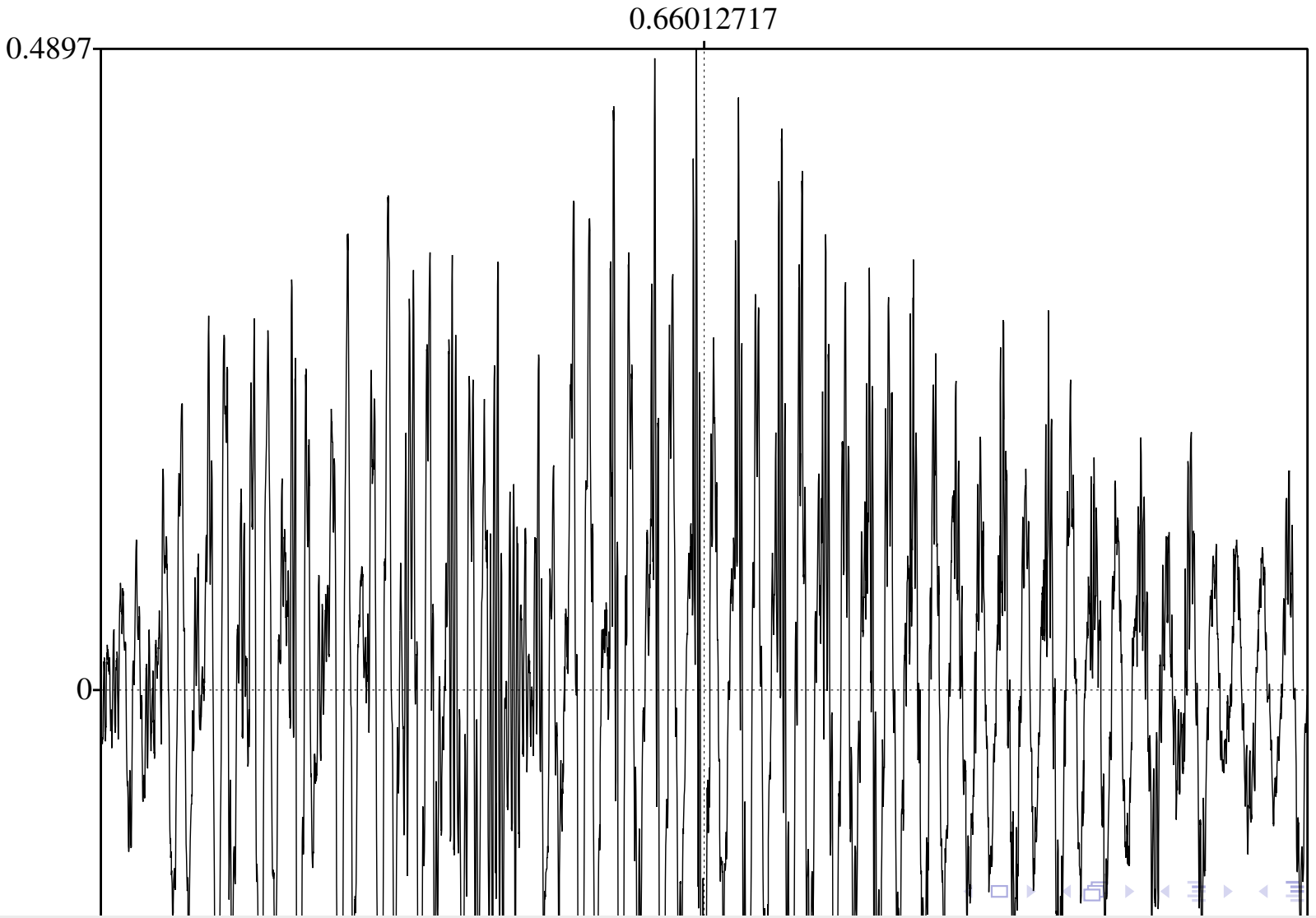
Spectrogram for *face*



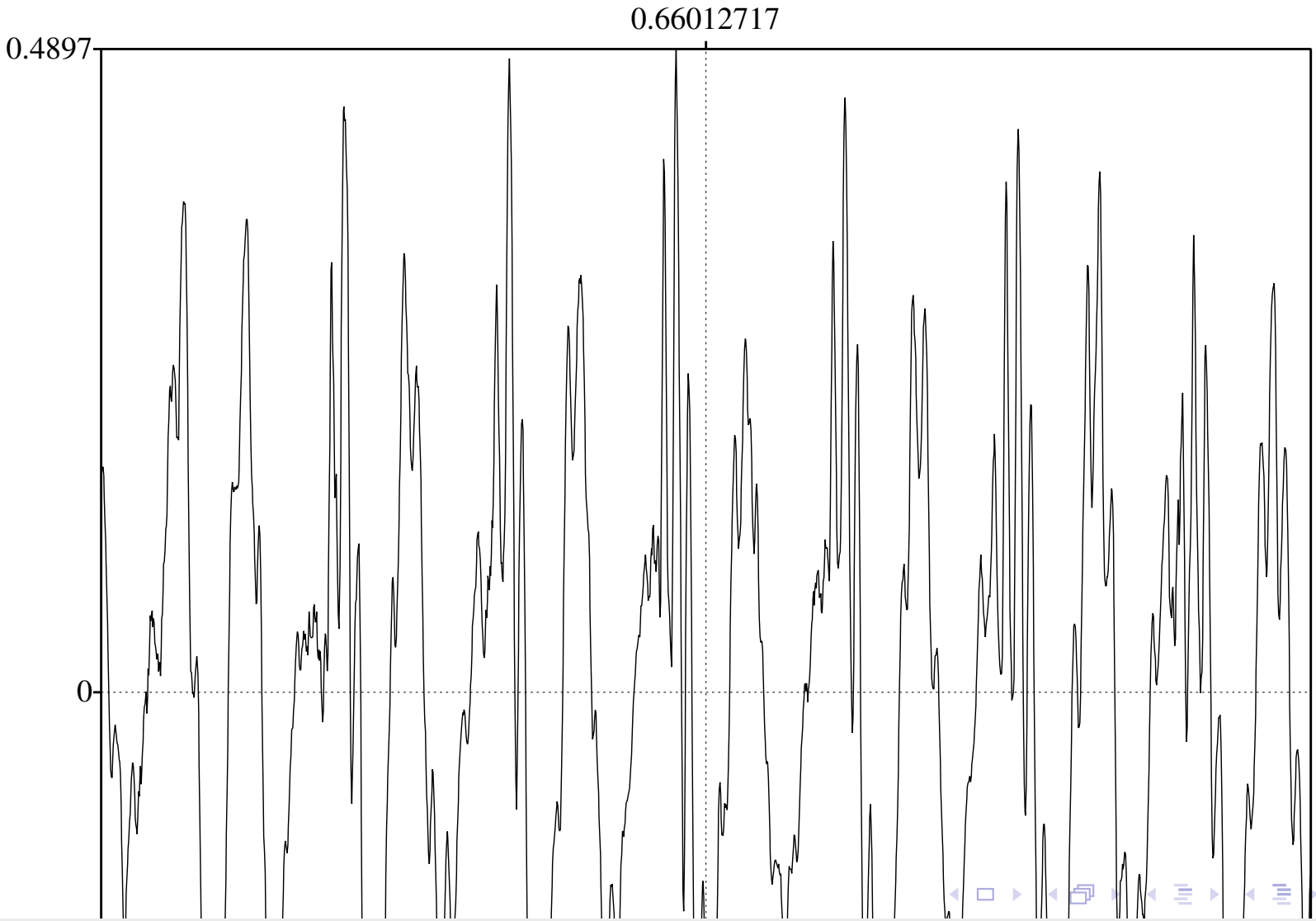
Waveform for *face*



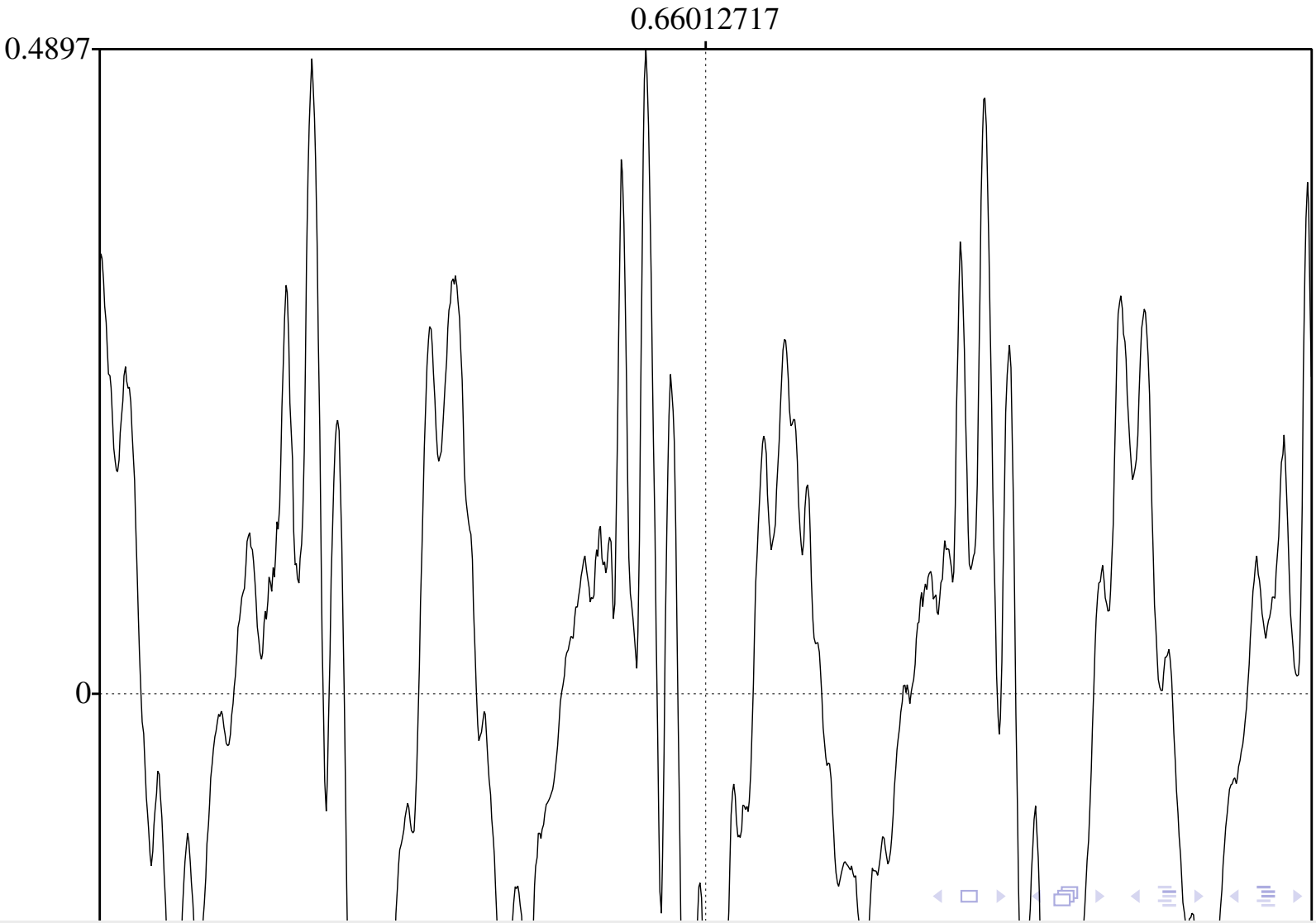
Waveforms for ej



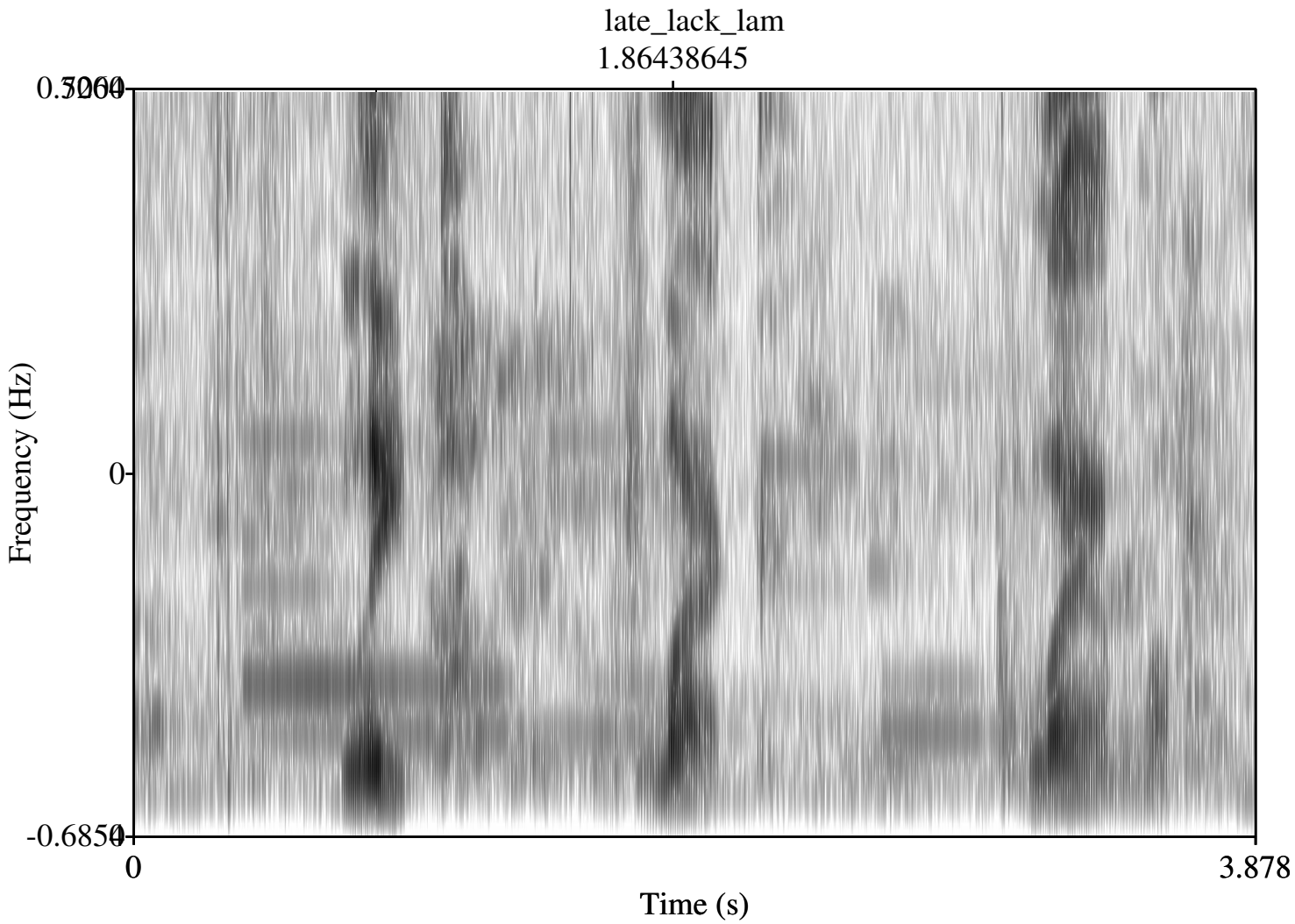
Waveforms for ej



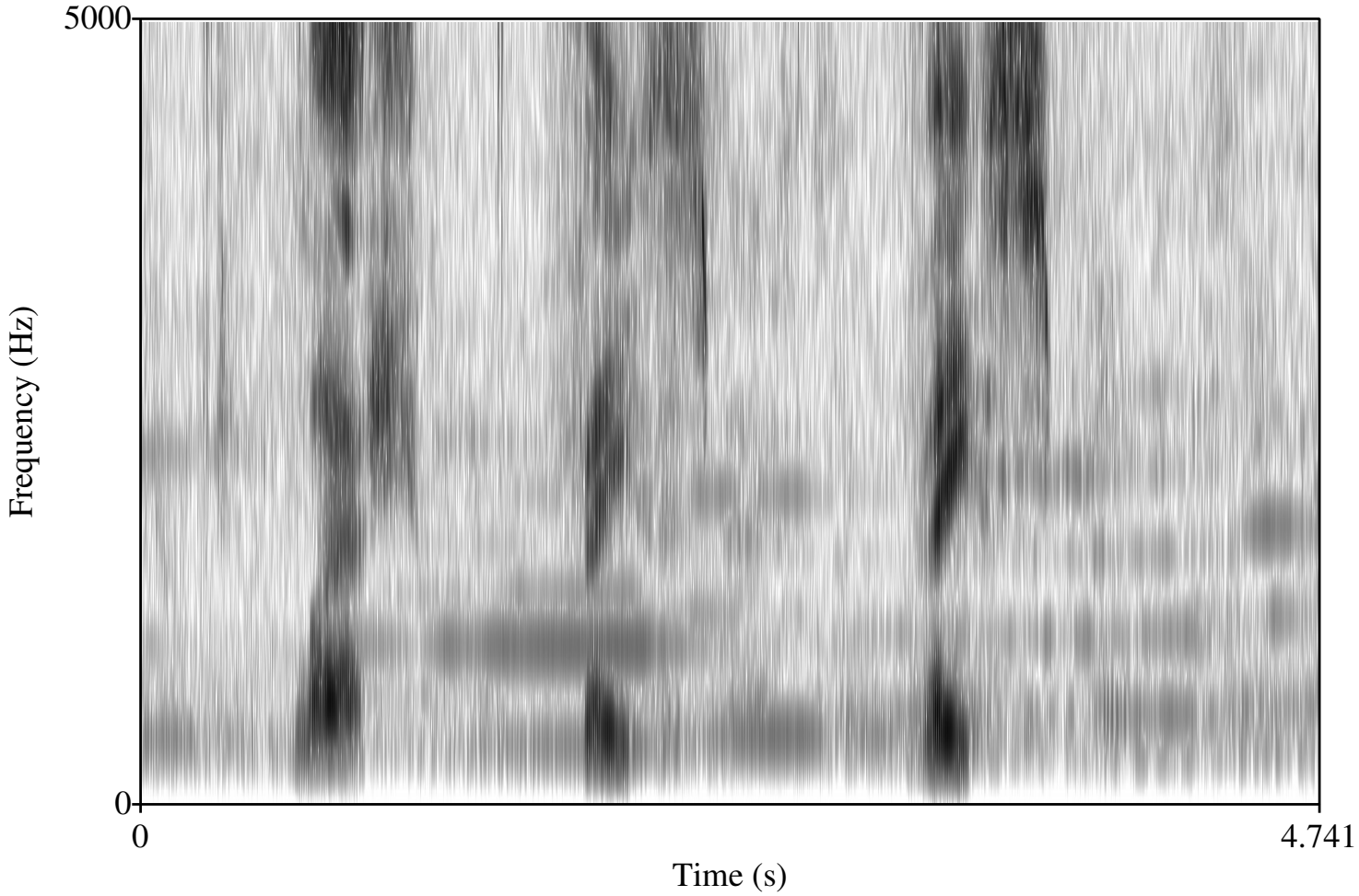
Waveforms for ej



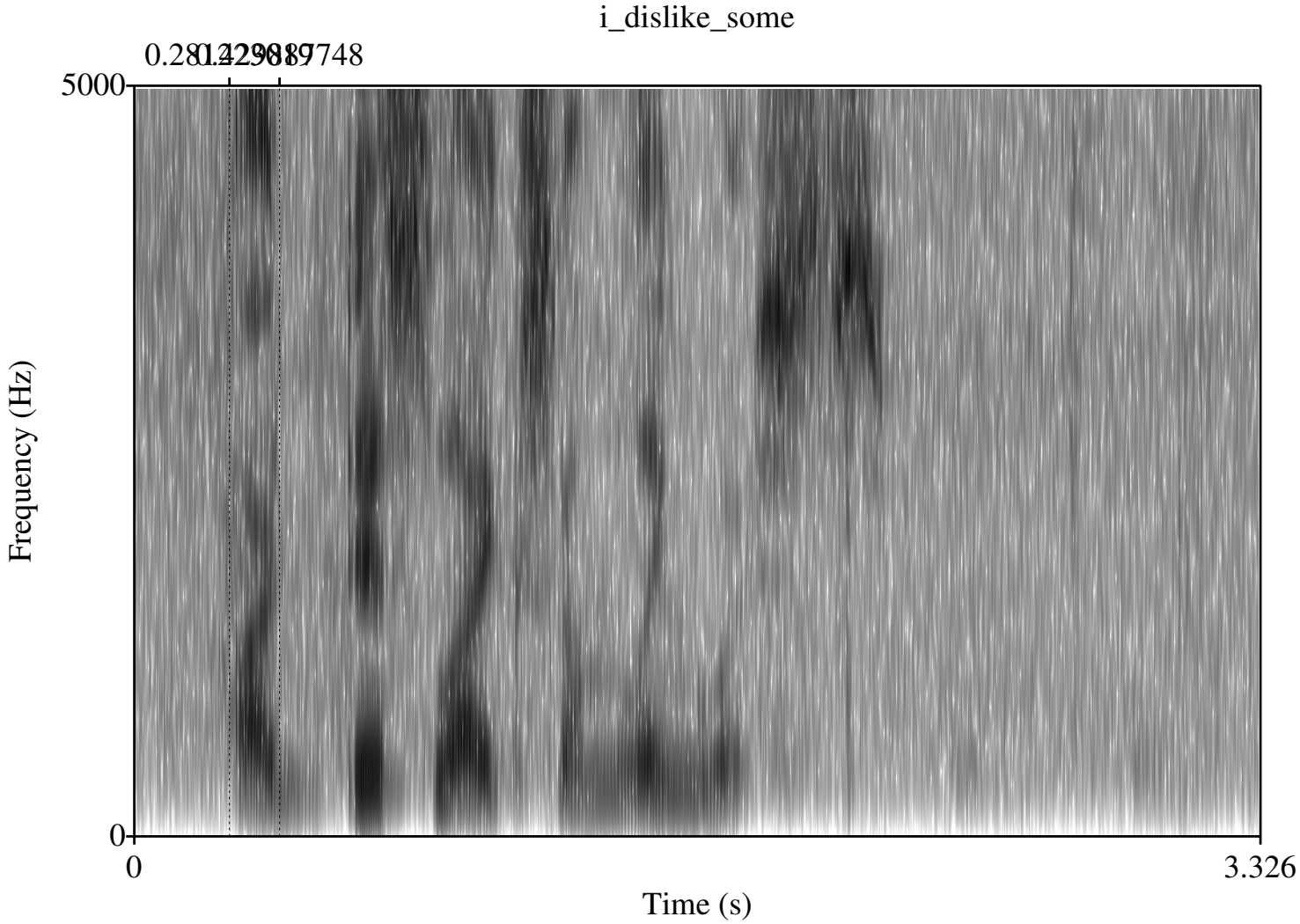
Spectrogram for *late, lack, lam*



Spectrogram for *lash, face, vase*

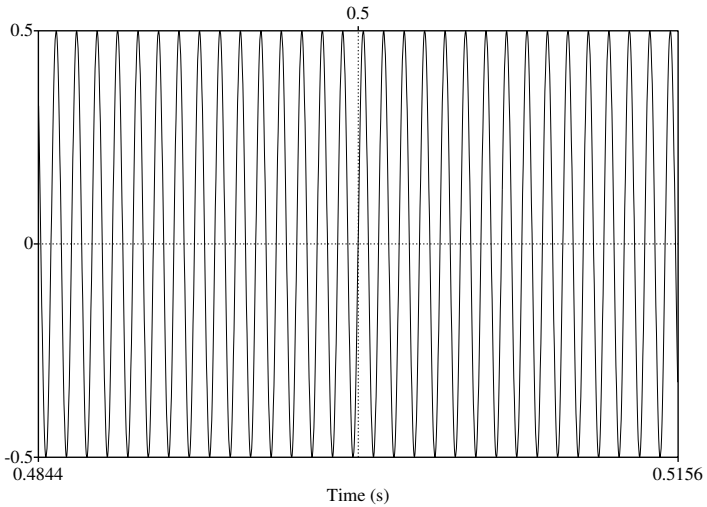


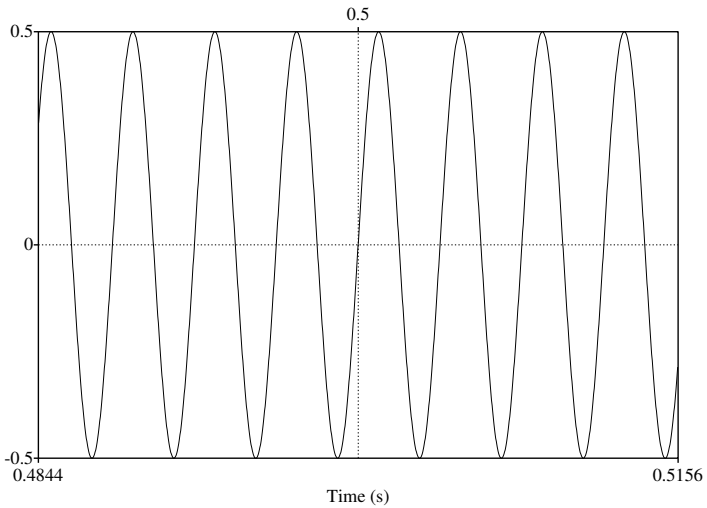
An utterance



Outline

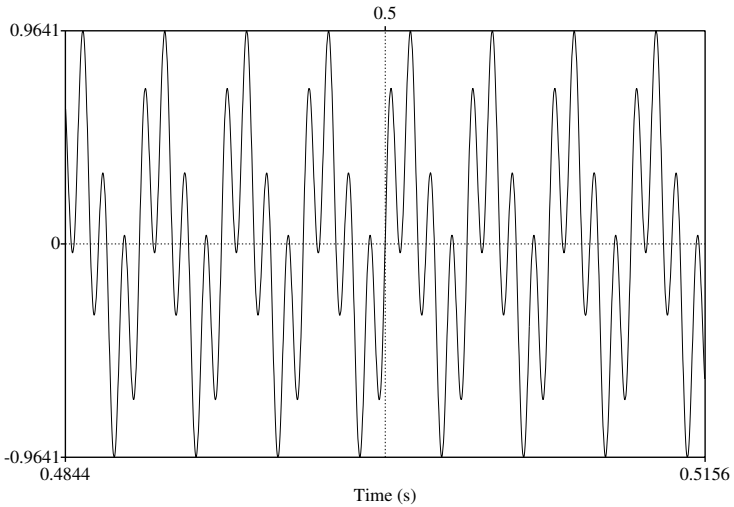
1000 Hz vs 250 Hz

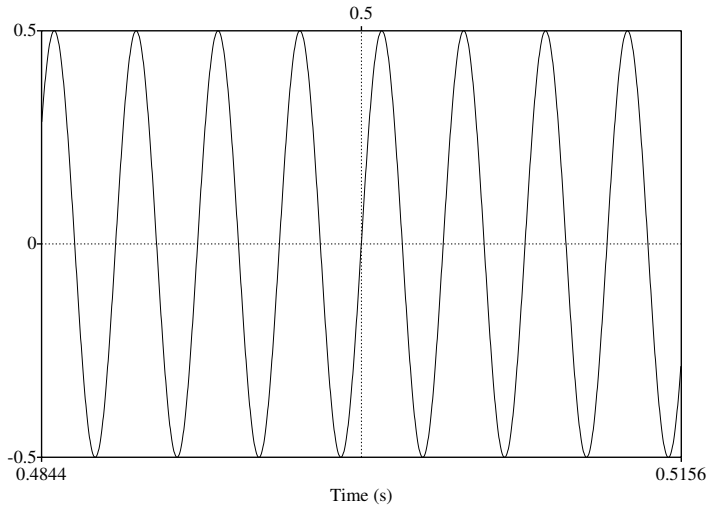


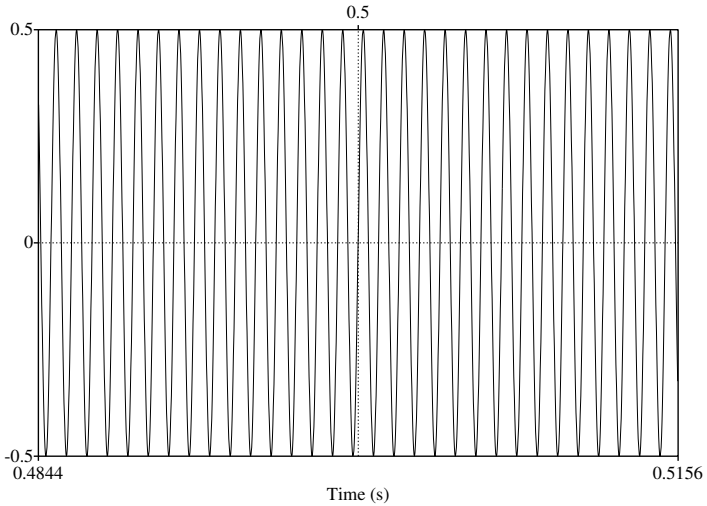


Sine Waves of 2 frequencies: 1000Hz and 250Hz

1000 Hz + 250 Hz



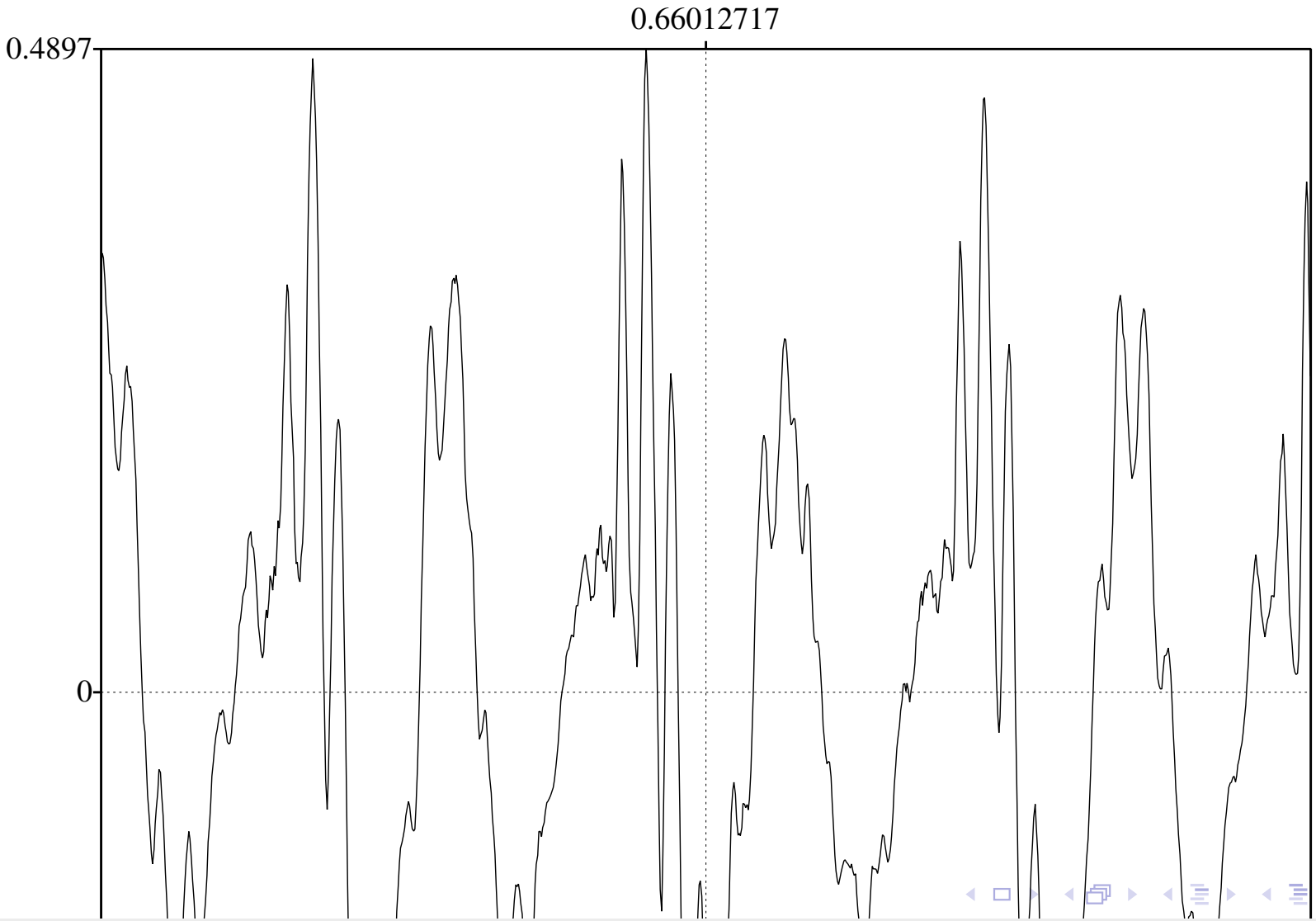




A sound's spectrum

Waveform for 1000Hz + 250Hz [top]; 1000Hz [bottom]

Waveform for ej



Sound components

- Even the most complex sound can be viewed as the sum of sine waves, although it may be a very complex sum of waves at many different frequencies

Source filter model

Sound components

- Even the most complex sound can be viewed as the sum of sine waves, although it may be a very complex sum of waves at many different frequencies
- These sine waves are the **components** of the sound.

Source filter model

Sound components

- Even the most complex sound can be viewed as the sum of sine waves, although it may be a very complex sum of waves at many different frequencies
- These sine waves are the **components** of the sound.
- The components may have different **intensities** (perceived as **volume**)

Source filter model

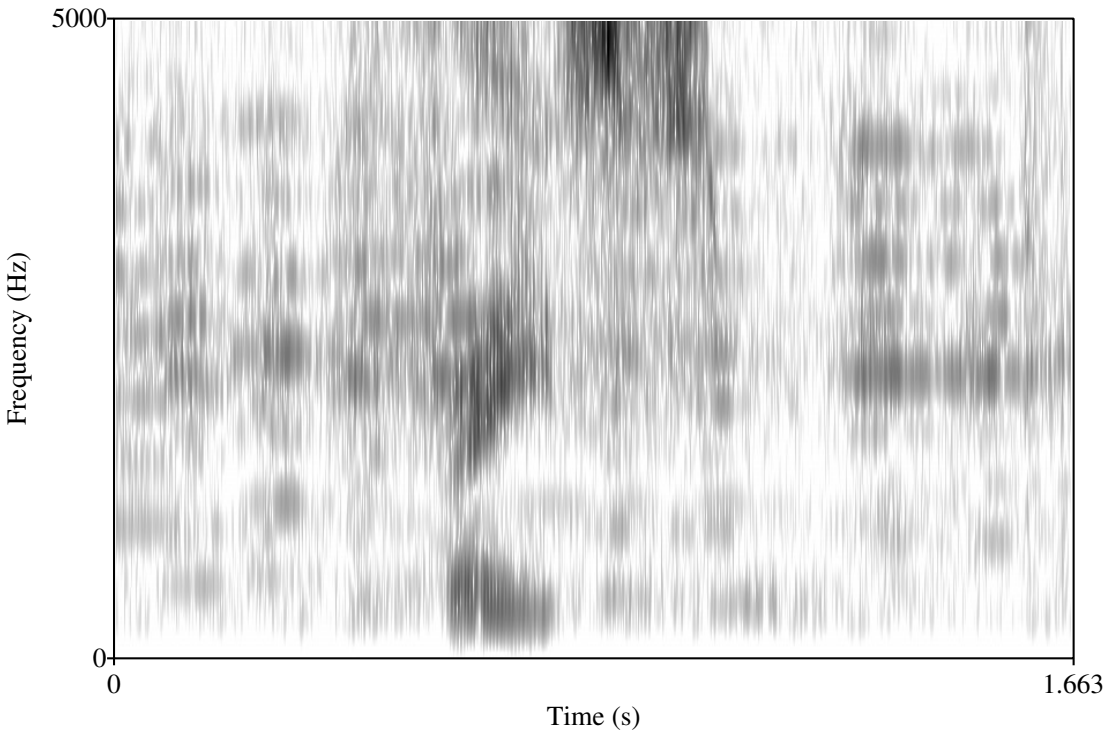
Sound components

- Even the most complex sound can be viewed as the sum of sine waves, although it may be a very complex sum of waves at many different frequencies
- These sine waves are the **components** of the sound.
- The components may have different **intensities** (perceived as **volume**)
- In our example: the complex sound has two components, 1 at 1000 Hz, another at 250 Hz, both at a volume of 1 dB.

Source filter model

Outline

Spectrogram for *face*



Formants

Formants

- A spectrogram has time as it horizontal axis, and frequency as its vertical axis.

Formants

- A spectrogram has time as it horizontal axis, and frequency as its vertical axis.
- So where is volume (intensity)?

Formants

- A spectrogram has time as it horizontal axis, and frequency as its vertical axis.
- So where is volume (intensity)?
- A dark band at frequency f represent times during which the sound components at f is especially intense.

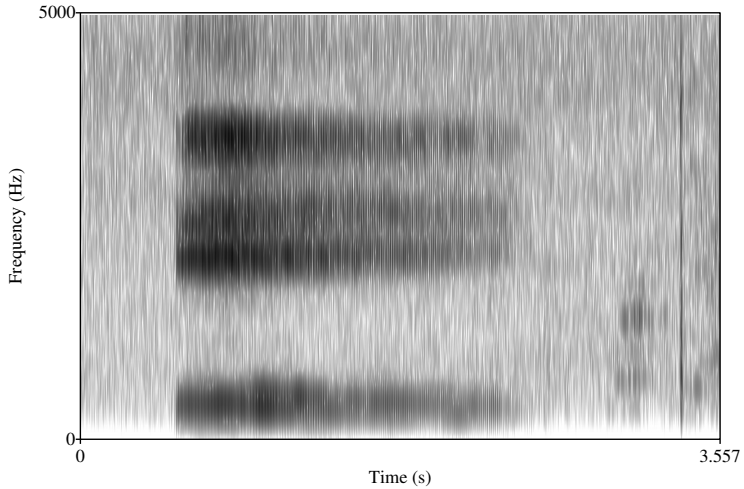
Formants

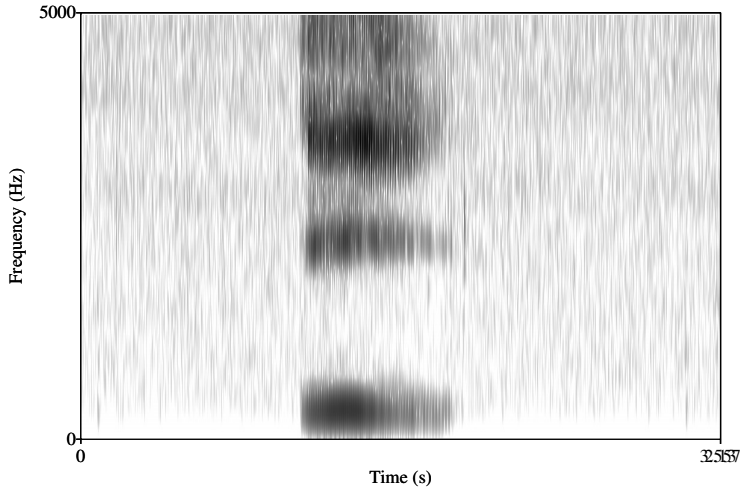
- A spectrogram has time as it horizontal axis, and frequency as its vertical axis.
- So where is volume (intensity)?
- A dark band at frequency f represent times during which the sound components at f is especially intense.
- The formants of a vowel are frequencies at which there is especially high intensity, visible as dark bands in the spectrogram.

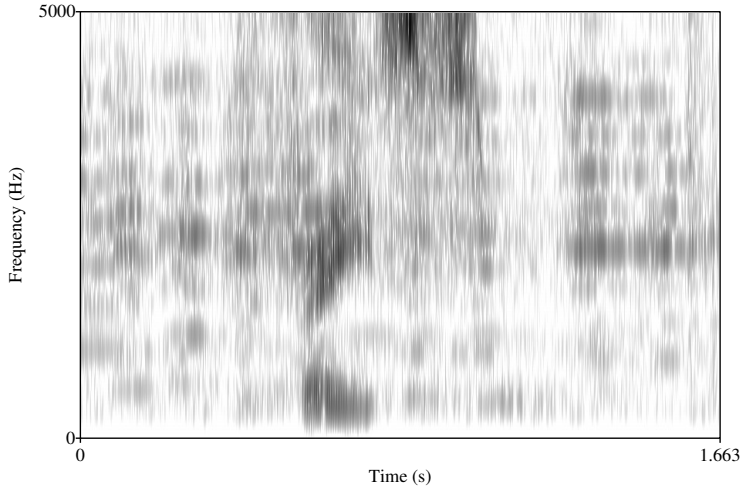
Formants

- A spectrogram has time as it horizontal axis, and frequency as its vertical axis.
- So where is volume (intensity)?
- A dark band at frequency f represent times during which the sound components at f is especially intense.
- The formants of a vowel are frequencies at which there is especially high intensity, visible as dark bands in the spectrogram.
- Can you find the formants for ej in the spectrogram on the next slide?

Formants for ej

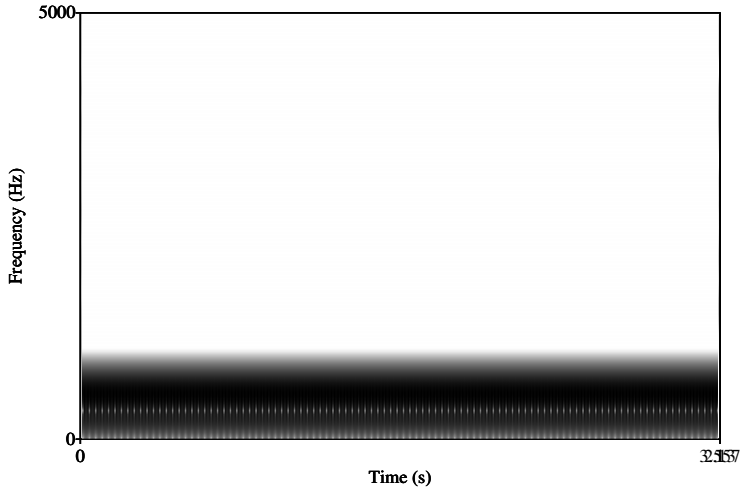


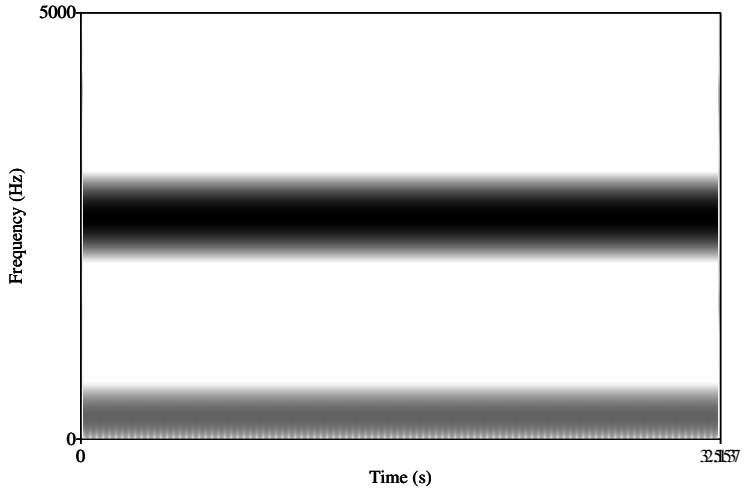




e and i[top]; feis [bottom]

Synthesizing i and u





Synthesized i and u

Outline

Inconsistencies of speech cues

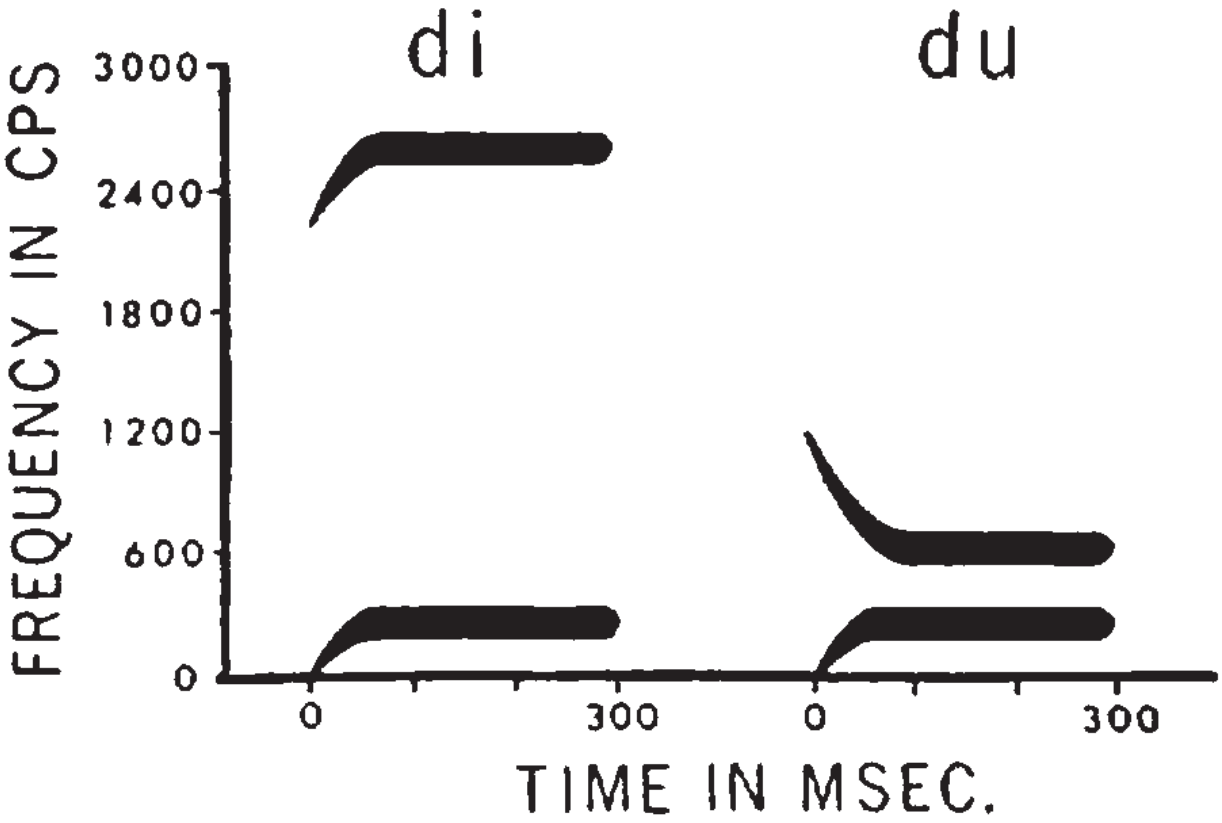


FIG. 1. Spectrographic patterns sufficient for the synthesis of /d/ before /i/ and /u/.

The decoding problem for /d/

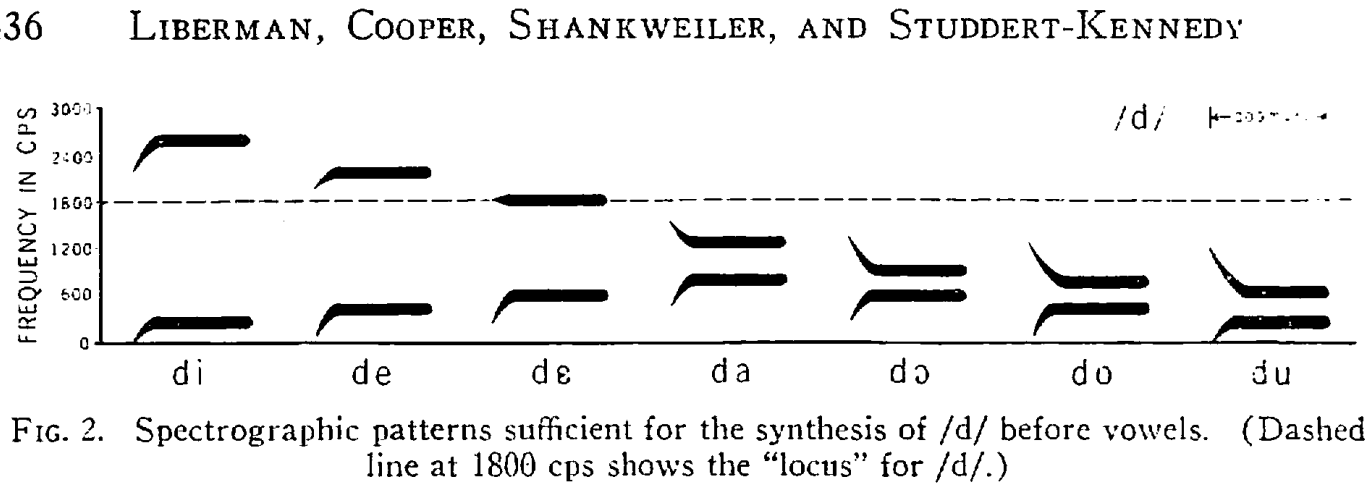


FIG. 2. Spectrographic patterns sufficient for the synthesis of /d/ before vowels. (Dashed line at 1800 cps shows the "locus" for /d/.)

From Liberman et al. (1967)

No consistent cue for /d/

- No /d/ segments of speech signal that can be consistently cut out and substituted for one another: /d/ before /u/ cannot be substituted for /d/ before /i/ .
- Conclusion: What unifies the different cases of d is the way they are produced (articulatory facts, place, manner, voicing).
- Facts like thse lead to complex models of segments in speech recognizers/synthesizers; in effect both kinds of systems need different models for d for each possible preceding and following segment.
- Is speech perception/categorization a uniquely human ability?

A special code

... speech is, for the most part, a special code that must often make use of special perceptual mechanism to serve as its decoder. On the evidence presented here we should conclude that most phonemes cannot be perceived by a straightforward comparison of the incoming signal with a set of stored phonemic patterns or templates... to find acoustic segments that are in any reasonable simple sennse invariant with linguistic (and perceptual) segemtns – that is to perceive without decoding – one must go to the syllable level or higher... We would suggest an additional possibility: the operations that occur in the speech decoder – including in particular the interdependence of perceptual and productive processes – may be in some sense similar to those that take place at other levels of grammar. If so, there would be a special compatiblity between the perception of speech sounds and the comprehension of language at higher stages.
(Liberman et al. 1967)

Animal categorization of speech

- Chinchillas can learn to distinguish voiced from unvoiced stops. Kuhl and Miller (1975) [see below]
- Japanese quail Kluender et al. (1987):

> *Japanese quail (*Coturnix coturnix*) learned a category for syllable-initial [d] followed by a dozen different vowels. After learning to categorize syllables consisting of [d],[b], or [g] followed by four different vowels, quail correctly categorized syllables in which the same consonants preceded eight novel vowels. Acoustic analysis of the categorized syllables revealed no single feature or pattern of features that could support generalization, suggesting that the quail adopted a more complex mapping of stimuli into categories. These results challenge theories of speech sound classification that posit uniquely human capacities.*

Outline

Summary/conclusion

- Recognizing /d/ acoustically is hard. There's no single thing unique to /d/ shared by all the different contexts that /d/ can occur in.
- But Japanese quail can learn a category for /d/ in all its different contexts.
- Conclusion: From the fact that acquiring a certain linguistic feature is difficult and that success at it is amazing, we shouldn't conclude that ability to acquire that feature evolved specifically for language.

Outline

Are there segments?

Working with recordings of real speech, Harris (1953) tried to arrive at "building blocks" by cutting tape recordings into segements of phoneme length and then recombing the segments to form new words. "Experiments indicated that speech based upon one buidliung block for each vowel not only sounds unnatural but is mostly unintelligible. Peterson, Wang and Sivertsen (1958) concluded that the smallest segments one can use are of half-syllable length.
(Liberman et al. 1967:441)

Context dependence of all segment realizations

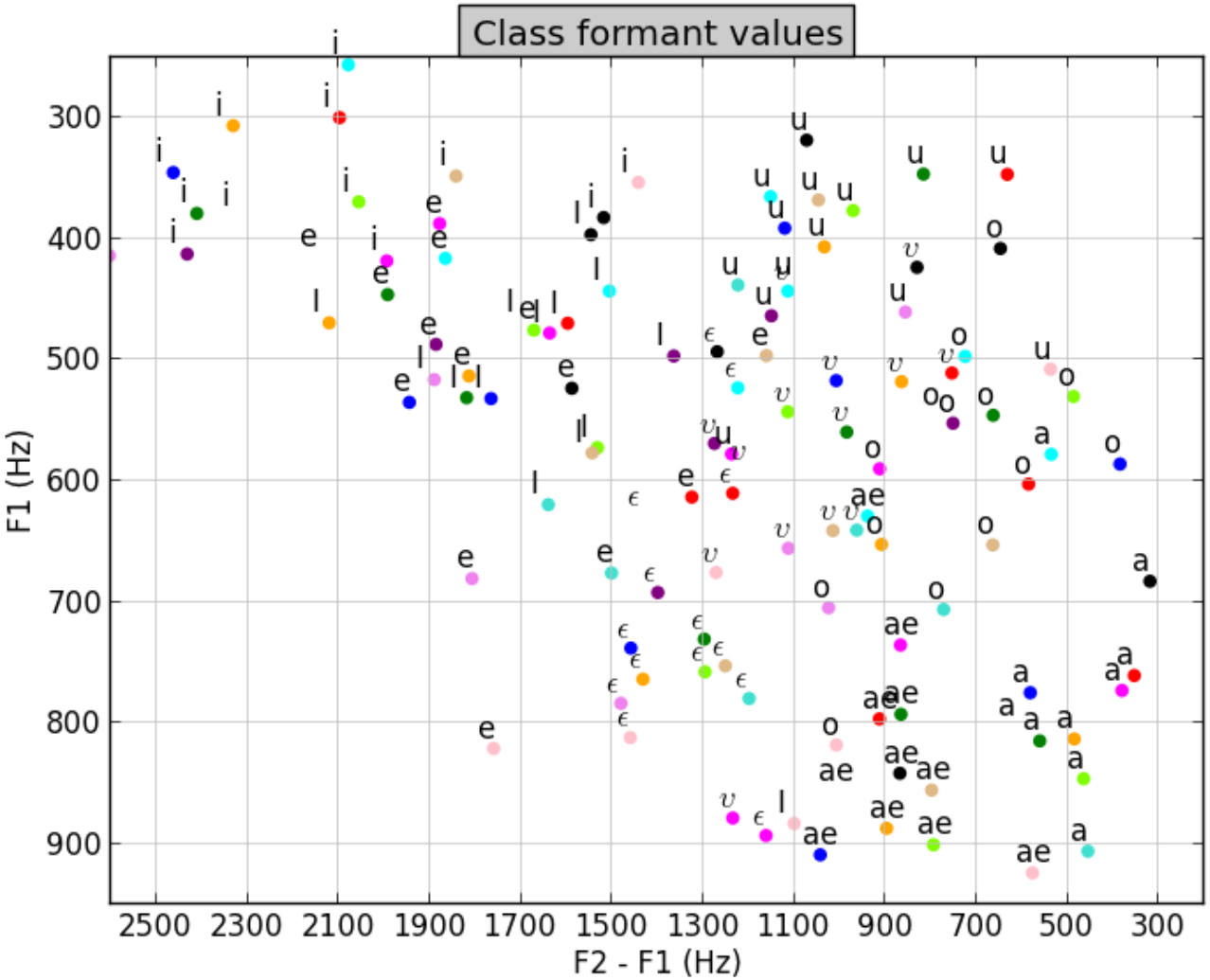
... vowels are rarely steady state in normal speech; most commonly these phonemes are articulated between consonants and at rather rapid rates. Under these conditions vowels also show substantial restructuring — that is, the acoustic signal at no point corresponds to the vowel alone, but rather shows, at any instant instant, the merged influences of the preceding pr following consonant...
(Liberman et al. 1967:441)

Summary

1. The realization of all segments in the acoustic signal is to some extent highly context dependent (moreso for consonants than vowels)
2. There actually aren't any **pure** invariants in the speech signal actually corresponding to normal linguistic units.
3. But talking about syllables or moras rather than segments helps: Although the realization of a consonant is somewhat affected by the preceding vowel, the effect of the preceding vowel is much less than the effect of the following vowel: That is, the vowel in the same syllable with the consonant is the main determinant of its acoustic realization.
4. Vowel realizations are affected by immediately preceding and following consonants, again, by members of teh same syllable.

Outline

Variability



Categorical perception

Categorical perception

1. Understanding speech requires **categorical perception**
2. To distinguish /dɪp/ from /t^hɪp/ requires classifying **voice onset times** (VOT), the time between the stop release and the onset of voicing that occurs with the following vowel.
3. Armenian VOT: Chapter 3

Categorical perception questions

1. Clear category instances at either end of a continuum
2. VOT
3. What happens in the middle?
4. Is there a category boundary?
5. In-category discrimination vs between-category discrimination

Categorical perception experiments

Adult subjects (p. 130)

Infants (p. 134)

Categorical perception in crickets

Sound has meaning

Frequency	Function	Meaning
4-5 kHz	Identification	attracts potential mates and con
25 kHz	Echolocation sounds of bats	predator, repels crickets

Is there a category boundary?

Experiment

Labeling	Vary sounds from 5 to 40 kHz.
	Categorical reponses, flight or attraction, never ignored
Discrimination	Habituate to sound at x kHz, play sound at y kHz, y > x
	Increased flight response only observed when y crosses category boundary

Kluender, K.R., R.L. Diehl, and P.R. Killeen. 1987.
Japanese quail can learn phonetic categories.
Science 237:1195–97.

Kuhl, P.K., and J.D. Miller. 1975.
Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants.
Science 190:69–72.

Liberman, A.M., F.S. Cooper, D.P. Shankweiler, and M. Studdert-Kennedy. 1967.
Perception of the speech code.
Psychological review 74(6):431.