LING 576 Acoustic Phonetics
Spring 2009
Rod Casali
Topic number 11: Acoustic analysis and speech perception
3-26-09

Reading:
Beckman, Mary E. 1986. Stress and non-stress accent. Dordrecht: Foris Publications. Read pp. 141-144.
Bladon, Anthony. 1986. Phonetics for hearers. In Language for Hearers, ed. by Graham McGregor, 1-24. Oxford: Pergamon Press.
Casali, Roderic F. 1990. Contextual labialization in Nawuri. Studies in African Linguistics 21:319-346.
Wright, Richard. 2001. Perceptual cues in contrast maintenance. The Role of Speech Perception in Phonology, ed. by Elizabeth Hume and Keith Johnson, 251-274. San Diego: Academic Press.
The Bladon article will likely be difficult reading in places. Don't worry about understanding all the details. Focus mainly on what it says about ways in which acoustic signals are transformed during the hearing process.

1. Speech transmission (review)

Speech communication involves the encoding and decoding of information at several different levels (the "speech chain").

level	locus	function
mental sound representation	speaker	speech planning
↓		
articulatory events	speaker	speech production
↓		
acoustic signal	sound wave	speech transmission
↓		
sound percepts	hearer	speech perception
↓		
processed mental representation	hearer	speech processing

Some questions to consider:

- Which level in the speech chain would we ideally like a phonetic transcription to represent?
- Which level does a phonetic transcription typically represent in practice?
- Which level in the speech chain does a microphone and acoustic speech analysis software record?
- What implications do these facts hold for phonetic analysis?

A simplistic view:

- Phonetic transcriptions seek to represent facts at the *articulatory* level.
- Articulatory events correspond directly to abstract mental representations of sounds.

That is, speech sounds are represented mentally as bundles of articulatory features which are translated more or less directly into articulatory gestures.

- Phonetic instruments record the *acoustic* signal.
- The *acoustic* signal provides a faithful record of *articulatory* events, such that from a representation of the *acoustic* signal one can deduce what the *articulatory* facts are and provide an appropriate transcription of these facts.

Under this view, acoustic phonetic instruments and software are a way to "take the hearer out of the loop." They allow us to reliably recover, without use of our ears, the articulatory events that produced an utterance (and thus, ultimately, the abstract mental phonetic representation of the utterance).

What is wrong with this view?

In order to evaluate this view, we will consider in turn the following mappings among adjacent levels in the speech transmission process:

- Mental sound representation → articulatory events.
- Articulatory events → acoustic signal
- Acoustic signal → sound percepts

2. Abstract mental representations & articulatory events

Some standard assumptions of modern phonological theories:

- Native speakers of a language "store" abstract mental underlying representations of utterances.
- These underlying representations consist of segments composed of phonological features.
- Phonological features have substantive phonetic content.
- Phonological representations are related to surface phonetic representations by a system of rules and/or constraints that the native speaker has internalized.
- Surface phonetic representations also consist of features.
- The set of universal features that makes up phonetic representations is the same as the set of features that makes up the phonological representations.

A qualification: Whereas phonological features are generally assumed to be binary, these same features are sometimes assumed to take on scalar values in phonetic representations.

A further assumption shared by most (though not all) recent theories: Phonological and phonetic features are defined in terms of *articulatory* properties.

This means that, roughly speaking, phonetic features can be construed as *instructions to the vocal apparatus to execute particular articulatory gestures*.

Examples?

The relationship between abstract features and articulatory gestures is not always simple and one-to-one:

- A phonological / phonetic feature can sometimes be realized by more than one articulatory gesture.

Examples?

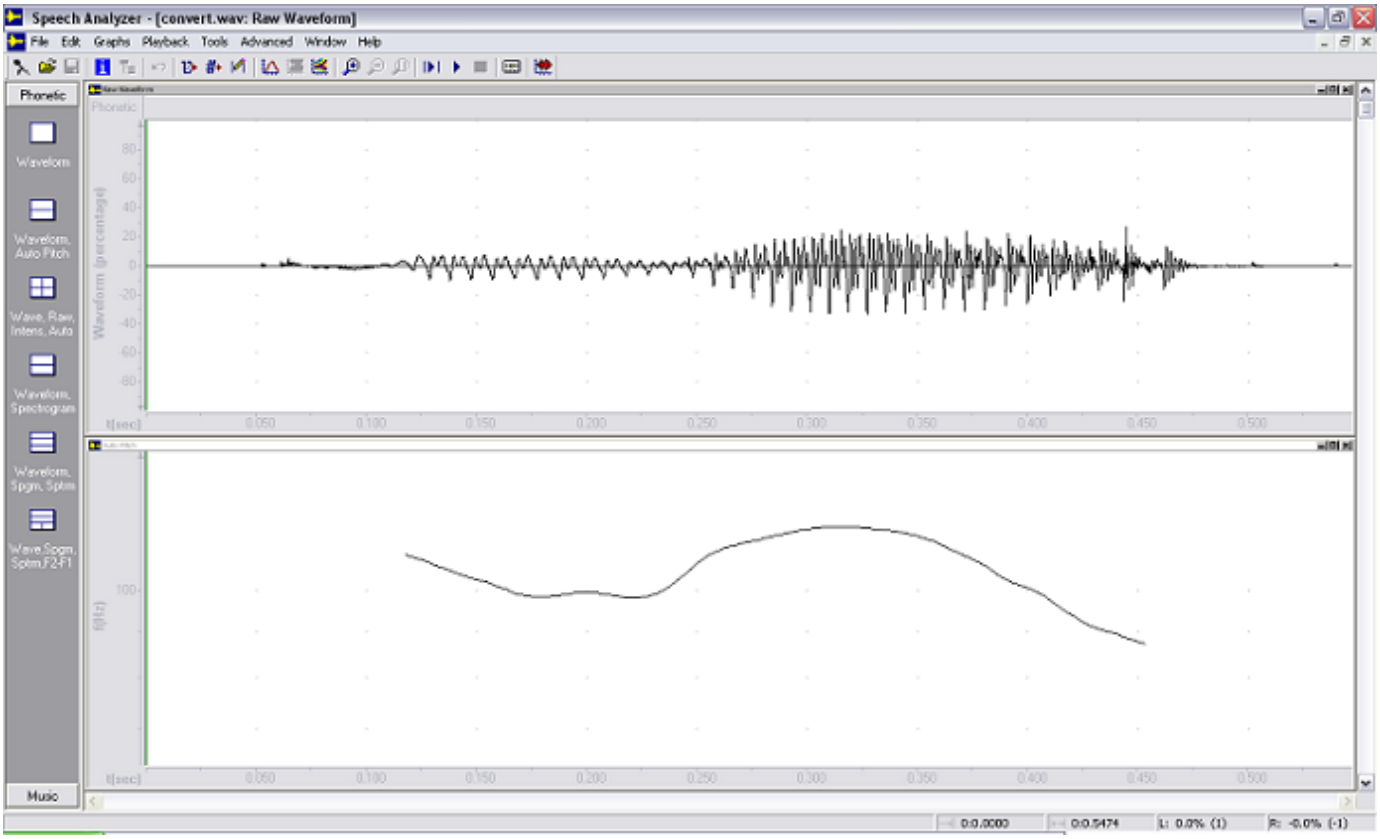
- A given articulatory gesture does not always correspond to the same phonological / phonetic feature.

Examples?

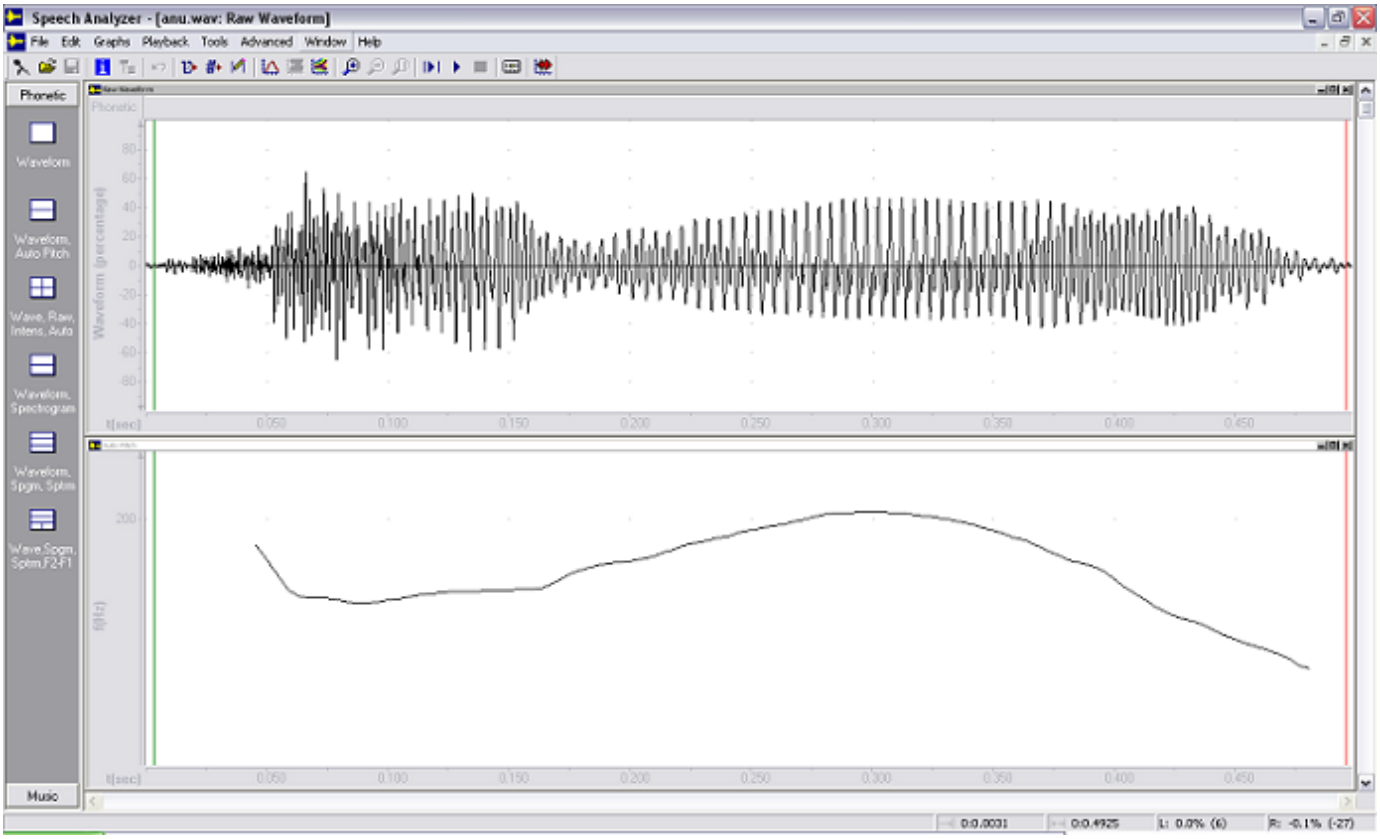
A particularly striking example involves the linguistic use of vocal fold vibration frequency (an articulatory feature).

Consider the two utterances whose waveforms and F_0 are displayed below.

Utterance 1: English verb convert pronounced in isolation:



Utterance 2: Nawuri word ànû 'five' pronounced in isolation:



Note that both utterances show a similar overall F_0 pattern, in which:

- F_0 is relatively low on the first syllable.
- F_0 on the second syllable starts considerably higher and falls.

Despite the similarity of the overall F_0 contours, *the phonetic features underlying these contours in the two languages are entirely different*.

- In the English example, the contour is the realization of *stress* on the final syllable.
- In the Nawuri example, the same contour is the realization of a surface L H-L tone melody.

Note that Nawuri is a tone language, and does not have stress at all. English, on the other hand, is a stress language and does not have lexical tone.

Assuming that the same articulatory gestures are used to produce the similar F_0 contours in the two languages, this example illustrates that a given articulatory gesture need not always correspond to the same phonetic feature.

3. Articulatory events & the acoustic signal

Key questions to consider:

- Does the same articulatory event always yield the same acoustic consequences?
- Is it always possible to reliably infer the articulatory event(s) that caused a given characteristic of the acoustic signal?

That is, does a particular kind of acoustic signal always have the same articulatory cause?

Discussion and examples:

4. The acoustic signal & sound percepts

There are general correlations between acoustic and perceptual features.

Some examples:

acoustic feature	perceptual correlate
amplitude	loudness
duration	perceived segment length
F_0	pitch
lowered F_2	perceived backness in vowels
lowered center of spectral energy	perceived labialization in fricative consonants

However, the relationship of acoustic to perceptual features is not simple and one-to-one.

Complications:

- Some perceptual features can be triggered by more than one acoustic feature.

Example: Although the primary acoustic correlate of rising or falling pitch is rising or falling F_0, *changes in amplitude* can also give rise to an impression of rising or falling pitch (Beckman 1986).

- The perceptual response to some acoustic features is highly non-linear.

This is the case for example with F_0 and pitch and with amplitude and loudness.

- Certain factors can causes major changes to the acoustic spectrum of sound without affecting how the sound is categorized perceptually.

Example: Shouted speech shows major shifts in vowel spectra. This does not however cause listeners significant problems in identifying vowel quality.

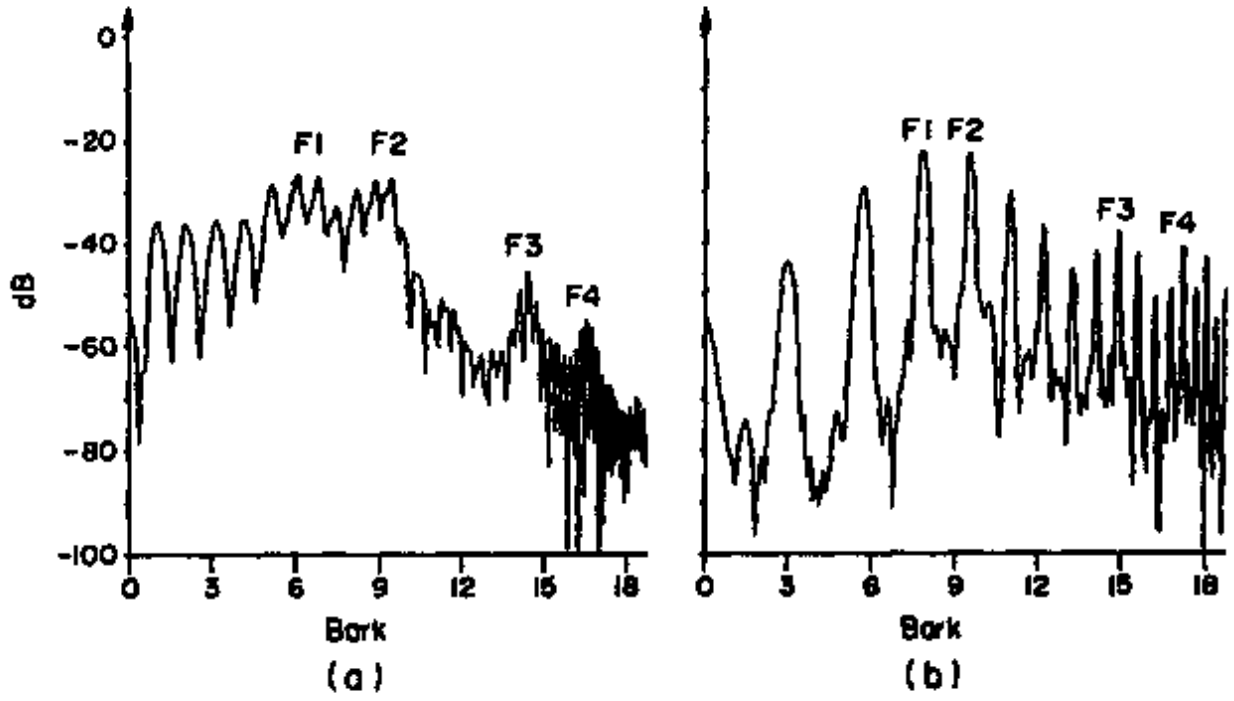


FIG. 3. Narrow-band spectra of a vowel [a] in "fire", (a) at normal loudness and (b) shouted; microphone distances 15 cm and 100 cm respectively. See text for comment on the substantial spectral differences.

-- From Bladon (1986:20)

A further example: Vowel formants are heavily dependent on the size of a speaker's vocal tract. The spectrum of a vowel as pronounced by a child can differ greatly from that of the

same vowel as pronounced by an adult. (There are also significant differences between male and female adults.)

Once again, these significant differences in spectral shape do not hinder listeners in correctly identifying vowel quality.

- Certain acoustic effects appear to be automatically filtered out during the course of speech perception.

There are *many* examples of this latter effect.

Exercise: Transcribe the words in the following sound files phonetically:

- lect5examplea.wav
- lect5exampleb.wav
- lect5examplec.wav
- lect5exampled.wav

Discussion:

Some other examples:

- Different vowel qualities have inherent effects on F_0. All else equal, higher vowels (e.g. [i], [u]) tend to have higher F_0 than non-high vowels. However, this effect is apparently compensated for in the perception of pitch (Beckman 1986:128-131).

This means for example that while the F_0 of a high vowel with a particular toneme in a tone language will be higher than the F_0 of a low vowel with the same toneme in the same context, a human listener will likely not hear the two vowels as having different pitches.

- Although lower vowels are generally (all else equal) longer in duration than higher vowels there is evidence that the speech perception mechanism corrects for this difference so that it is not generally perceived (Beckman 1986).
- In normal speech, high tone peaks (both in tone languages and intonational languages like English) tend to be progressively lower in F_0 toward the end of an utterance, a process known as *declination*. However, this effect is apparently not perceived (Yip 2002).

This means for example that in a word consisting of three high toned syllables, the second syllable will (all else being equal) have a lower F_0 than the first and the third syllable will

have a lower F_0 than the second. Listeners will not perceive the lowering effect however but will hear all three syllables as having the same pitch.

- A vowel that is actually nasal may be perceived as oral when it occurs next to a nasal consonant. (More generally, a vowel with a particular degree of nasalization will be perceived as less nasal when it occurs next to a nasal consonant than when it occurs in other contexts (Kawasaki 1986)).
- Labialization of consonants tends not to be perceived before round vowels (Casali 1990).

There is much evidence that human speech perception tends to filter out many features of the acoustic signal that could be attributed to automatic effects of the surrounding context.

"listeners' expectations in perceiving speech play a crucial role in giving rise to sound patterns in language . . . whatever a listener expects to hear, that is, some kind of automatic or commonly encountered perturbation of one segment by another, may be taken for granted and factored out of the phonetic percept constructed for a word, as long as the segment responsible for the perturbation is detected . . . If the perturbing segment is not detected, for whatever reason, then the perturbation is not expected and is not factored out; it is then included as part of the phonetic percept of the word."

--Kawasaki 1986:86,87

- What implications do these perceptual filtering effects have for normal speech communication?
- What implications do they have for instrumental acoustic speech analysis?

References

Beckman, Mary E. 1986. Stress and non-stress accent. Dordrecht: Foris Publications.

Bladon, Anthony. 1986. Phonetics for hearers. In Language for Hearers, ed. by Graham McGregor, 1-24. Oxford: Pergamon Press.

Casali, Roderic F. 1990. Contextual labialization in Nawuri. Studies in African Linguistics 21: 319-346.

Kawasaki, Haruko. 1986. Phonetic explanation for phonological universals: The case of distinctive vowel nasalization. In Experimental Phonology, ed. by John J. Ohala & Jeri J. Jaeger, 81-103. Orlando: Academic Press.

Yip, Moira. 2002. Tone. Cambridge: Cambridge University Press.