# Acoustic Phonetics

- Display formats for acoustic phonetic information
- Vowel classification and features
- Consonant classification and features
- Coarticulation
- Suprasegmentals

NOTES

1) We'll begin by briefly describing spectrograms and other display formats that are useful for describing the acoustic structure phonemes.

2) Next we'll look at how vowels are classified and their most prominent acoustic features.

3) Ditto for consonants.

4) Coarticulation is an extremely important property of speech from both a production and perception standpoint. Unfortunately, it is less important for speech analysis and recognition. We will describe it, but not go into it in depth.

5) Suprasegmentals are also extremely important, but little used in recognition. This topic will also receive very brief treatment.

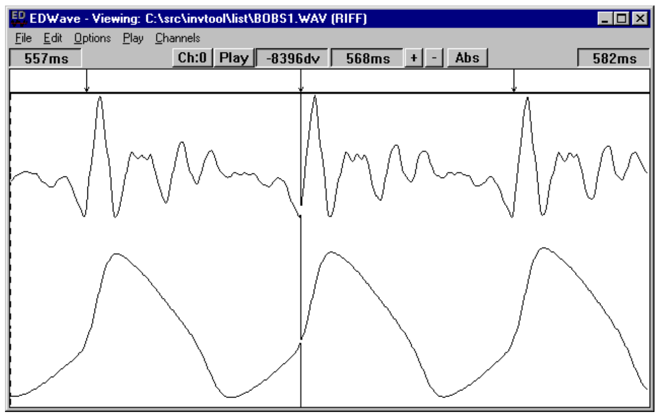
# Representation of Acoustic Phonetic Information

- Waveform and Spectrogram displays
- Auxiliary signals and information

NOTES

The next several slides are intended to familiarize everyone with the graphics we'll use in this class and throughout the remainder of the semester.

# Waveform of /ɔ/ with EGG signal




NOTES

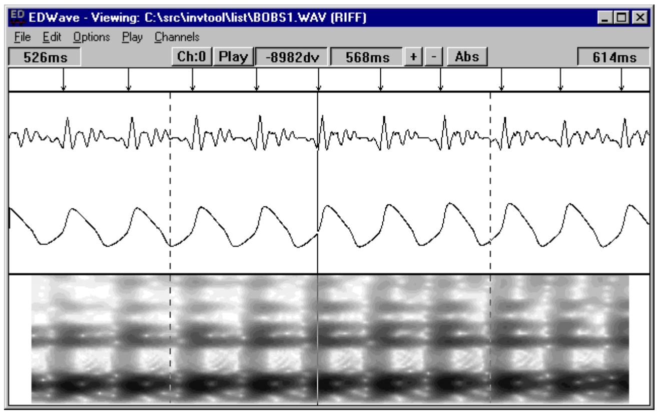
1) The display program is WEDW, a Windows based waveform editor. We'll look at it in depth a little later.

2) There are three types of graphic information shown here, two time signals (the speech waveform for nearly three pitch periods of a vowel (upper waveform), the Electroglotograph (EGG) signal (lower wavefrom), and onset markers calculated for each pitch period (the arrows in the band above the speech waveform).

3) The EGG signal is recorded from a device which measures impedance across the vocal folds during speech. It uses a surface electrode on either side of the larynx. When the vocal folds are in contact with one another, the impedance is relatively low and when the folds are abducted, impedance is high. For this reason, the EGG signal is essentially measuring vocal fold contact area. The EGG signal is also inversely related to air flow: when contact area is low, air flows; when contact area is high, air cannot flow.

4) With each vocal cord closure, the vocal tract is excited by an impulsive pressure change which initiates a pitch period. The vocal tract is essentially ringing from the impulse. The ringing dies down, especially during the open phase of the glottal cycle when the vocal tract is not closed on one end, but coupled with the air volume in the lungs.

# /ɔ/ Waveform, EGG, and Spectrogram

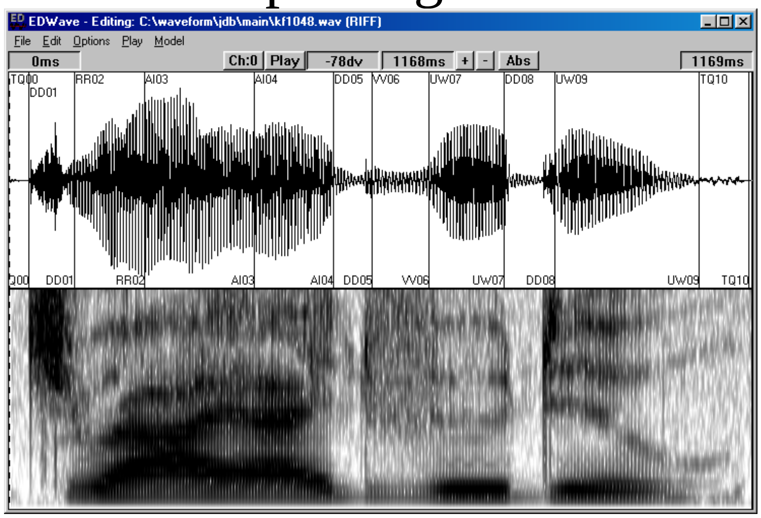



NOTES

This is a slightly broader view of the same vowel, EGG, and pitch marker display with the addition of a spectrogram. The spectrogram has axes of time (X axis), Frequency (Y axis), and intensity (darkness).

This spectrogram shows both horizontal bands of strong energy at the formant frequencies and vertical bands of energy at the onset of each pitch period. The pressure transient that excites the vocal tract is sufficiently like an impulse that it contains moderately strong energy throughout the spectrum, but the energy between formant frequencies dies down quickly after the impulse, leaving strong energy only at the formant frequencies through most of each pitch period.

# Waveform & Wideband Spectrogram



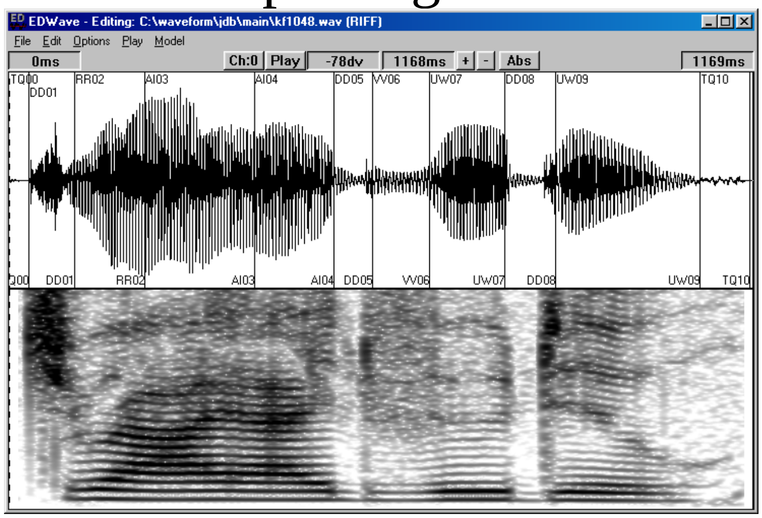

NOTES

This is a still broader view of an utterance. At this level, the fine structure of individual pitch periods is no longer visible. This figure also differs in that the pitch marker and EGG displays have been turned off (neither is of much use at this resolution).

The spectrogram more clearly shows the trajectories of formants throughout the voiced portions of the utterance, but the vertical striations due to pitch periods are hard to see.

This is a wideband spectrogram. The frequency resolution of the bandpass filters used to compute the spectrum was about 300 Hz. As a consequence, only gross spectral features are visible.

# Waveform and Narrowband Spectrogram

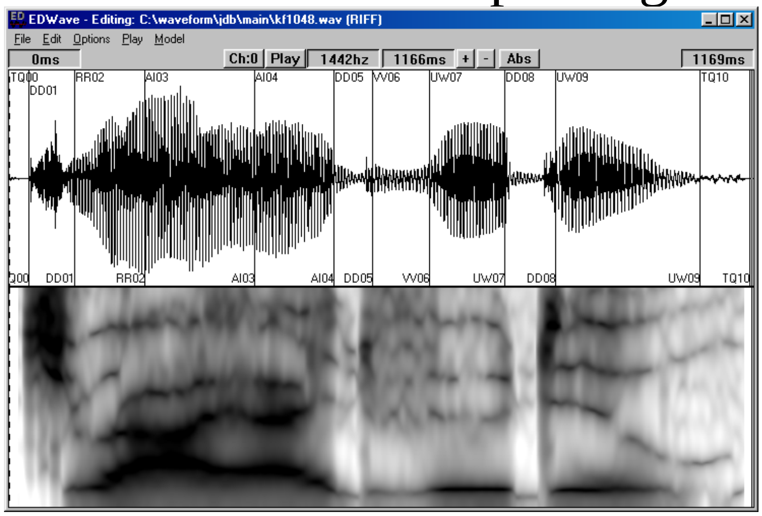



NOTES

This is a narrow band spectrogram of the same utterance. In the figure, the formant frequencies can still be seen fairly clearly, but there are many additional horizontal bands in the spectrogram. Moreover, the vertical striations are not visible at all.

The many vertical bands in this figure are harmonics. The 45 Hz bandwidth of the analyzing filters is sufficiently narrow to resolve individual harmonics. However, because it requires a large window of speech to produce such good frequency resolution, the temporal resolution of the narrow band spectrogram is less than that of the previous spectrogram.

# Waveform & LPC Spectrogram




NOTES

This final view of the same utterance uses a linear predictive coding (LPC) analysis to estimate the speech spectrum rather than a fast fourier transform (FFT) as in the previous examples.

The LPC analysis intentionally discards pitch information from the analysis.

## Vowel Classification

- High - Low dimension
  - Approximately how constricted the VT is.
  - Corresponds (inversely) to frequency of F1
- Front - Back dimension
  - Whether the constriction is toward the front or back of the oral cavity
  - Corresponds to Frequency of F2
- Rhotic
  - Indicates /r/ coloring of vowel
  - Corresponds to frequency of F3

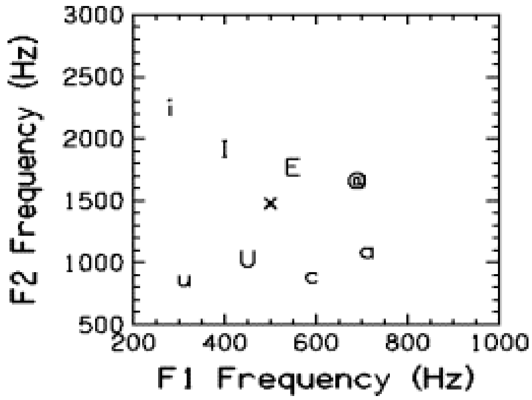
NOTES

1) For the most part, two formants are all that are needed to describe the perceptually crucial acoustic structure of vowels.

2) Fant has shown that for vowels with F2 and F3 high in frequency and close together, the two formants are perceptually well matched by a single formant that lies between F2 & F3. He called this F2'

3) For vowels that have an /r/ coloring such as the "er" sound in words like fur, one must consider the frequency of F3 (which is significantly lower than its neutral frequency). This is because the F1 & F2 frequencies of "er" overlap those of schwa.

# The Vowel Space




NOTES

This figure is pretty self-explanatory. The common English vowels are distributed fairly well in a two dimensional space defined by the F1 and F2 frequency for the vowel.

If we plotted vowels in a space that went from HIGH to LOW on the X axis and BACK to FRONT on the Y axis we would see a more coarsely quantized version of the same plot.

## Consonant Classification

- Voicing distinction
  - Voiced versus voiceless
  - Adducted versus Abducted vocal folds
- Manner distinction
  - Degree and type of constriction
- Place distinction
  - Location of constriction

NOTES

These are the main linguistic distinctions used to classify consonants.

# Consonant Manner by Place

| | Bilabial | Labio-dental | Dental | Alveolar | Paleto-alveolar | Palatal | Velar |
|---|---|---|---|---|---|---|---|
| Approximant | w | | | ɹ / l | | j | w |
| Fricative | | f v | θ ð | s z | ʃ ʒ | | |
| Stop | p b | | | t d | | | k g |
| Nasal | m | | | n | | | ŋ |
| Flap | | | | ɾ | | | |
| Affricate | | | | | tʃ dʒ | | |

NOTES

1) This table covers most of the English consonants, but not all. Neither /h/ nor glottal stop are shown, for instance.

2) These are also not all the possible places or manners for consonants throughout the world.

## Additional manners & places

- Glottal - stop /ʔ/ or approximant /h/
- In all, Ladefoged (1993) identifies 10 manners and 11 places of articulation needed to account for all languages.

NOTE

Didn’t I just say this?

# Phonemes & Allophones

- A Phoneme may vary in structure from one context to another.
  - /t/ in syllable initial position is aspirated $t^h$
  - /t/ in some syllable final contexts is voiced t̬
  - /t/ preceding some segments is dental t̪
  - /t/ before front vowels has different burst spectrum than before back rounded vowels

NOTES

1) Allophones are acoustic variations of phonemes that are not contrastive in a given language. That is, all variants are “heard” as the same phoneme in the language.

2) Things that are allophones in one language may not be in another language where they may be used contrastively.

# Acoustic or Phonetic features

- A phonetic property used to classify speech sounds.
- Features can be present or not, and when present can have two or more values.
- Features typically given with capitalized names like Voice or Stricture.
- Feature values given in square brackets like [+voice] (a value of Voice) or [stop] (a value of Stricture.

NOTES

1) Recent accounts of the structure of phonemes rely more heavily on descriptions in terms of a uniform set of features. Each phoneme is defined by its feature values.

# Features

- Voice
  - [+voice] b, d, g, m, n, …
  - [-voice] p, t, k, f, s, h, …
- Labial **-** uses one or more lips
- Coronal
  - [+anterior] on or in front of alveolar ridge (incl. r, l).
  - [-anterior] behind alveolar ridge (incl. front vowels).
- Dorsal **-** uses back of tongue.
- Stricture
  - [stop]
  - [fricative]
  - [approximant]
- Nasal
  - [+nasal]
  - [-nasal]
- Lateral
  - [+lateral] /l/
  - [-lateral] all other sounds
- Sibilant
  - [+sibilant] s, z, sh, zh, affric.
  - [-sibilant] all others
- Height
  - [maximum] all C except w, j
  - [{4,3,2,1} height]
- Back
  - [+back] w, k, g, ng
  - [-back] all other Cs
- Syllabic
  - [+syllabic] vowels
  - [-syllabic] consonants

NOTES

This is (part of) one featural system. Most English phonemes can be classified uniquely with just these features.

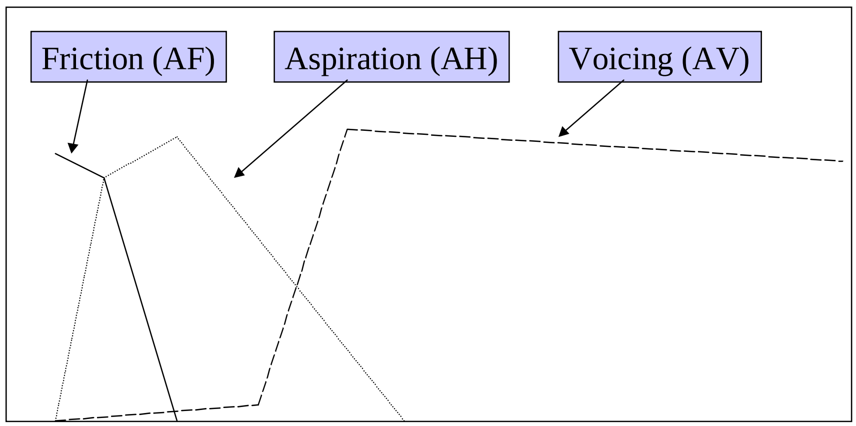
# Stop consonants

- Vocal tract fully closed resulting in low amplitude murmur or total silence
- Air pressure builds up behind constriction
- Release of constriction produces burst of air followed by aspiration followed by vowel transitions
- All these are cues to stop identity

NOTES

The next collection of slides look at place and/or voicing distinctions among segments within a manner class.

# Schema for stops




There are three phases in the onset of a stop consonant:

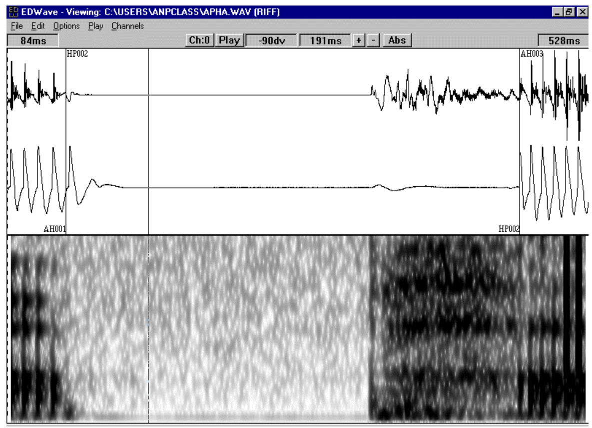
1. At the moment of release, there is a plosive burst of air which causes frication sound to be generated in the vocal tract from the location of the constriction. The friction dies down as the air pressure behind the point of constriction equalizes with the atmosphere causing the velocity of the air rushing out to slow down.

2. Shortly after the release, air begins to flow by the vocal cords which are beginning to close for the upcoming vowel. At some point, before the cords are close enough together to actually start voicing, the rushing air produces aspiration. The aspiration peaks at a moment when the vocal cords are close, but not close enough for voicing and then continues, diminishing, until voicing starts. For voiced stops, voicing may be continuous throughout the closure interval preceding the burst. In that case, there may be no aspiration and VOT is not a very useful construct.

3. Voicing onset occurs when the cords are sufficiently close to begin phonation. Shortly after this, the aspiration will die out (if modal voicing).

Note that the exact timing and extent of the three phases depends on (a) whether the stop is voiced or voiceless, (b) how rapidly the constriction opens, and (c) the air pressure at the time of release.

# Aspirated [p]



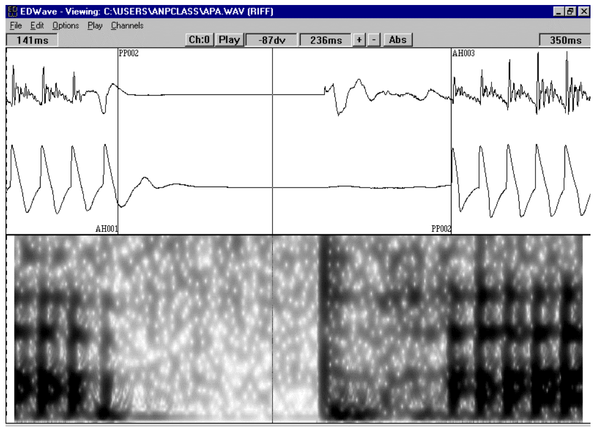

NOTES

1) This is the aspirated bilabial voiceless stop [$p^h$]. The EGG signal allows us to see what is happening at the level of the glottis, however, the adduction motion of the folds that produced the aspiration may be too slow to be very visible.

2) Note that the figure covers 444 msec. The VOT is around 100 msec.

3) Vocalic transitions of all formants into the consonant constriction interval are downward and out of the constriction interval, transitions are upward. This is the expectation for a bilabial place of articulation.

# Unaspirated [p]

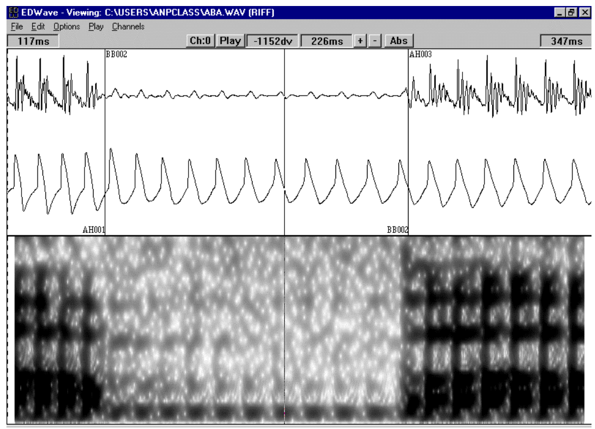


NOTES

1) This is the unaspirated bilabial voiceless stop [p].

2) Note this figure covers a time period of only 109 msec. Thus VOT is a little more than 20 msec.

# [b]




NOTES

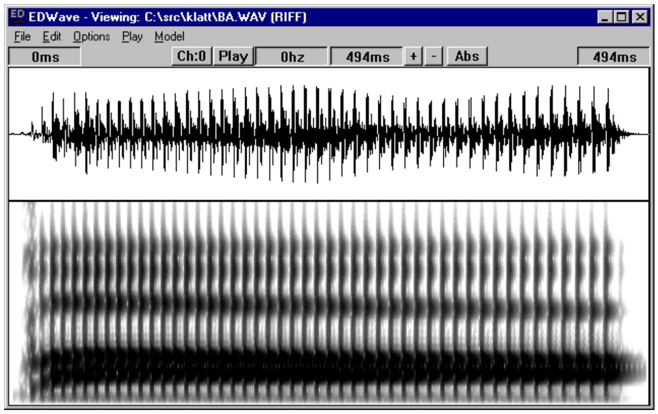
1) This is the voiced unaspirated bilabial stop [b].

2) Note that the EGG signal never ceases during the closure interval.

3) The burst is superimposed on the continued voicing. There is no VOT in this case.

4) Transitions, reflecting place of articulation, are similar to the previous voiceless examples, but may appear more prominent because continuous voicing allows them to be more visible.

# [ba]



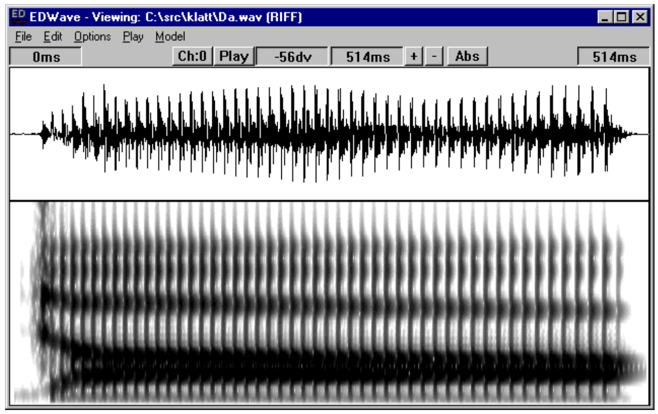

NOTES

1) This is a synthetic [ba]. The actual synthesis parameters are shown below.

2) Note the upward transitions at the vowel onset and that VOT is 0.

CONFIG
AV=VAR; AH=VAR; AF=VAR; F0=VAR; F1=VAR; F2=VAR; F3=VAR
F0=156; F1=649; F2=2508; F3=3050; F4=3300; FNZ=250
A1=60; A2=56; A3=52; A4=48; A5=44; A6=40; AB=36
B1=50; B2=70; B3=110; SW=0
FGP=0; BGP=100; FGZ=1500; BGZ=6000; B4=250; F5=3850
B5=200; F6=4900; B6=1000; FNP=250; BNP=100; BNZ=100
FRA=200; SR=11000; NWS=55; GAI=78; NFC=5
ENDCONFIG
*
* /ba/ template (all cues)
*
TIME = 0; F1=120; F2=900; F3=2200; F0=120; AV=48
TIME = 10; F1=200; F2=900; F3=2200; F0=120; AV=60; AF=72; AH=0
TIME + 10; AF=0; AH=00; AV=72
TIME + 10; F1=650; F2=1100; F3=2500; AV=72; AH=0
TIME + 30; F1=750; F2=1150; F3=2500; AV=72
TIME + 400; F1=750; F2=1000; F3=2300; F0=90; AV=72
TIME + 30; AV=0
END

# [da]

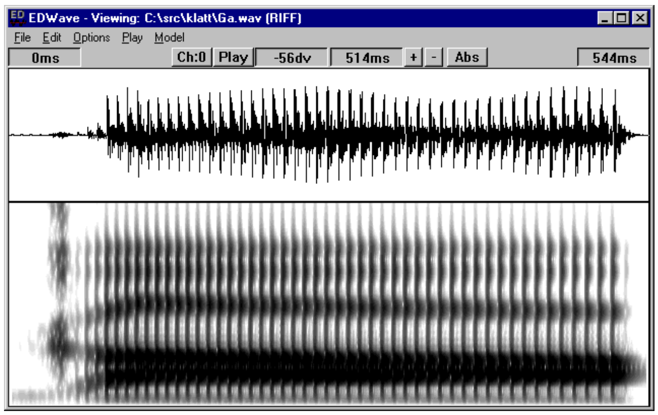



NOTES

Again, note transition cues at the vowel onset and the short VOT. F1 transition from consonants to vowels is always upward, but F2 transition reflects place of articulation of the consonant. Here F2 decreases, and F3 does slightly as well.

```
CONFIG
AV=VAR; AH=VAR; AF=VAR; F0=VAR; F1=VAR; F2=VAR; F3=VAR
F0=156; F1=649; F2=2508; F3=3050; F4=3300; FNZ=250
A1=60; A2=56; A3=52; A4=48; A5=44; A6=40; AB=36
B1=50; B2=70; B3=110; SW=0
FGP=0; BGP=100; FGZ=1500; BGZ=6000; B4=250; F5=3850
B5=200; F6=4900; B6=1000; FNP=250; BNP=100; BNZ=100
FRA=200; SR=11000; NWS=55; GAI=78; NFC=5
ENDCONFIG
*
* /da/ template (all cues)
*
TIME =   0; F1=120; F2=1800; F3=2700; F0=120; AV=48
TIME =  20; F1=200; F2=1800; F3=2700; F0=120; AV=48; AF=72; AH=67
TIME +  10;   AF=0;   AH=50;   AV=72
TIME +  20; F1=650; F2=1300; F3=2500;  AV=72;  AH=0
TIME +  30; F1=750; F2=1150; F3=2500;  AV=72
TIME + 400; F1=750; F2=1000; F3=2300;  F0=90; AV=72
TIME +  30; AV=0
END
```

# [ga]

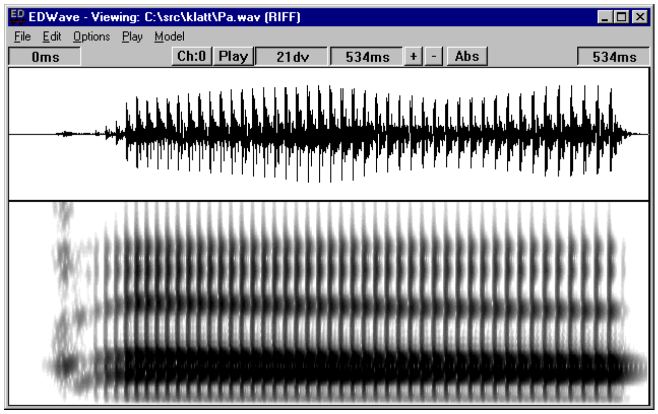



NOTE

1) Same comments: transition cues and VOT. Strong voicing starts latest in this case because the tongue dorsum covers a broader area so the constriction takes longer to open. For velar stops /g/ and /k/, F2 & F3 are at mid frequencies and often close together.

```
CONFIG
AV=VAR; AH=VAR; AF=VAR; F0=VAR; F1=VAR; F2=VAR; F3=VAR
F0=156; F1=649; F2=2508; F3=3050; F4=3300; FNZ=250
A1=60; A2=56; A3=52; A4=48; A5=44; A6=40; AB=36
B1=50; B2=70; B3=110; SW=0
FGP=0; BGP=100; FGZ=1500; BGZ=6000; B4=250; F5=3850
B5=200; F6=4900; B6=1000; FNP=250; BNP=100; BNZ=100
FRA=200; SR=11000; NWS=55; GAI=78; NFC=5
ENDCONFIG
*
* /ga/ template (all cues)
*
TIME =   0; F1=120; F2=1400; F3=1800; F0=120; AV=48
TIME =  40; F1=200; F2=1400; F3=1900; F0=120; AV=48; AF=72; AH=67
TIME +  10;   AF=0;   AH=64;   AV=52
TIME +  30; F1=650; F2=1200; F3=2300;  AV=72;  AH=0
TIME +  30; F1=750; F2=1150; F3=2500;  AV=72
TIME + 400; F1=750; F2=1000; F3=2300;  F0=90; AV=72
TIME +  30; AV=0
END
```

# [pa]

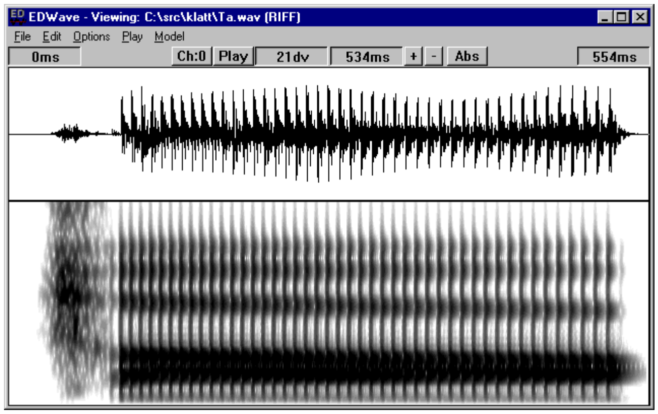



NOTES

1) voiceless bilabial compare transitions and VOT to [b] example.

```
CONFIG
AV=VAR; AH=VAR; AF=VAR; F0=VAR; F1=VAR; F2=VAR; F3=VAR
F0=156; F1=649; F2=2508; F3=3050; F4=3300; FNZ=250
A1=60; A2=56; A3=52; A4=48; A5=44; A6=40; AB=36
B1=50; B2=70; B3=110; SW=0
FGP=0; BGP=100; FGZ=1500; BGZ=6000; B4=250; F5=3850
B5=200; F6=4900; B6=1000; FNP=250; BNP=100; BNZ=100
FRA=200; SR=11000; NWS=55; GAI=78; NFC=5
ENDCONFIG
*
* /pa/ template (all cues)
*
TIME =   0; F1=120; F2=900; F3=2000; F0=120; AV=0
TIME =  40; F1=200; F2=900; F3=2000; F0=120; AV=0; AF=72; AH=67
TIME =  50; AH=72;   AV=00
TIME =  60; AF=0;    AV=40
TIME =  70; F1=650; F2=1100; F3=2300; AV=52;
TIME +  30; F1=750; F2=1150; F3=2500; AV=72;  AH=0
TIME + 400; F1=750; F2=1000; F3=2300; F0=90;  AV=72
TIME +  30; AV=0
END
```

# [ta]

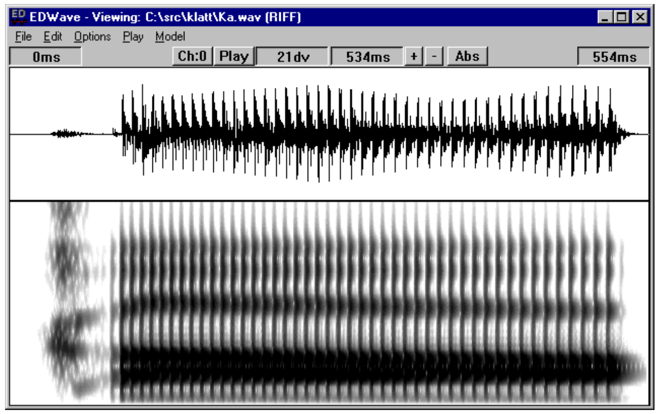



NOTES

/t/ is the voiceless alveolar stop. Note the burst and aspiration interval both are much longer than for /d/, the voiced cognate of /t/. Also, note that transitions can be seen in the aspiration interval.

```
CONFIG
AV=VAR; AH=VAR; AF=VAR; F0=VAR; F1=VAR; F2=VAR; F3=VAR
B1=VAR; B2=VAR; B3=VAR
F0=156; F1=649; F2=2508; F3=3050; F4=3300; FNZ=250
A1=60; A2=56; A3=52; A4=48; A5=44; A6=40; AB=36
B1=50; B2=70; B3=110; SW=0
FGP=0; BGP=100; FGZ=1500; BGZ=6000; B4=170; F5=3850
B5=200; F6=4900; B6=1000; FNP=250; BNP=100; BNZ=100
FRA=200; SR=11000; NWS=55; GAI=78; NFC=5
ENDCONFIG
*
* /ta/ template (all cues)
TIME =   0; F1=120; F2=1800; F3=2700; F0=120; AV=0
TIME =  40; F1=200; F2=1800; F3=2700; F0=120
          AV=0; AF=78; AH=67; B2=300; B3=200
TIME =  50; AH=78;   AV=00
TIME =  80; AF=40;   AV=40
TIME =  90; F1=650; F2=1200; F3=2550; AF=0; AV=72; B2=70; B3=110
TIME +  30; F1=750; F2=1150; F3=2500; AV=72;  AH=0
TIME + 400; F1=750; F2=1000; F3=2300; F0=90;  AV=72
TIME +  30; AV=0
END
```

[ka]

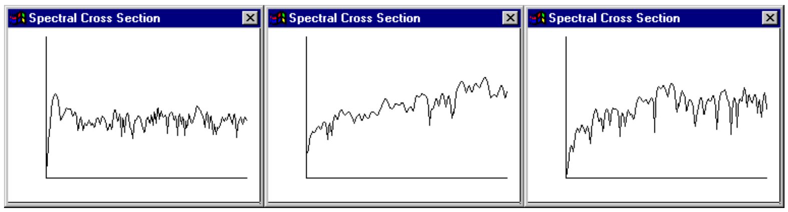



NOTES

Same comments as for /t/ regarding burst, aspiration, and transitions. As with /g/, /k/ tends to have the longest VOT and most gradual transitions of the stops because of the mass of the tongue dorsum.

```
CONFIG
AV=VAR; AH=VAR; AF=VAR; F0=VAR; F1=VAR; F2=VAR; F3=VAR
F0=156; F1=649; F2=2508; F3=3050; F4=3300; FNZ=250
A1=60; A2=56; A3=52; A4=48; A5=44; A6=40; AB=36
B1=50; B2=70; B3=110; SW=0
FGP=0; BGP=100; FGZ=1500; BGZ=6000; B4=250; F5=3850
B5=200; F6=4900; B6=1000; FNP=250; BNP=100; BNZ=100
FRA=200; SR=11000; NWS=55; GAI=78; NFC=5
ENDCONFIG
*
* /ka/ template (all cues)
*
TIME =   0; F1=120; F2=1400; F3=1800; F0=120; AV=0
TIME =  40; F1=200; F2=1400; F3=1900; F0=120; AV=0; AF=78; AH=67
TIME =  50; AH=78;   AV=00
TIME =  80; AF=0;    AV=40
TIME =  90; F1=650; F2=1200; F3=2300; AV=72;
TIME +  30; F1=750; F2=1150; F3=2500; AV=72;  AH=0
TIME + 400; F1=750; F2=1000; F3=2300; F0=90;  AV=72
TIME +  30; AV=0
END
```

# Burst Spectra




[p] [t] [k]

NOTES

In addition to transition cues for stops, the shape of the burst spectrum is a useful acoustic cue to consonant place, especially for voiceless stops.

Generally:

/p/ - Fairly flat burst spectrum

/t/ - Spectrum tilts upward reaching a peak at high frequencies.

/k/ - Spectrum tends to have a compact peak in mid frequency region with spectrum above the peak generally decreasing or level.

# Fricatives

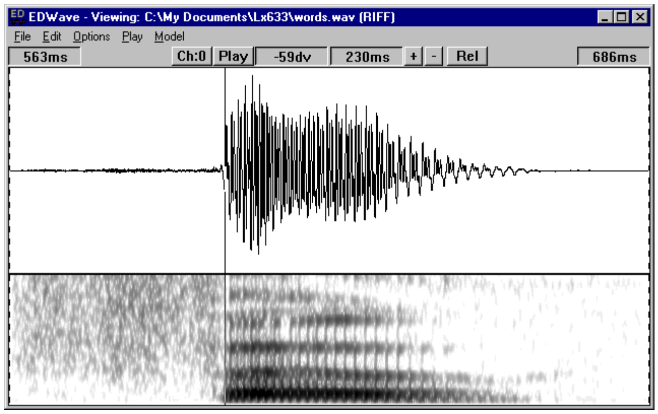
- Constriction slightly less than stop
- Continuous air flow generating friction
- Spectral characteristics due to location of constriction

NOTES

Usually, during most of a fricative, the constriction in the vocal tract is so narrow that the acoustic properties of the vocal tract behind the location of the constriction can be disregarded. That is, the vocal tract in front of the constriction and the vocal tract behind the constriction are decoupled.

Sometimes, especially in the transition interval between vowels and fricatives or vice versa, full tract formants can be seen as the vocal tract again begins to behave like a single tube.

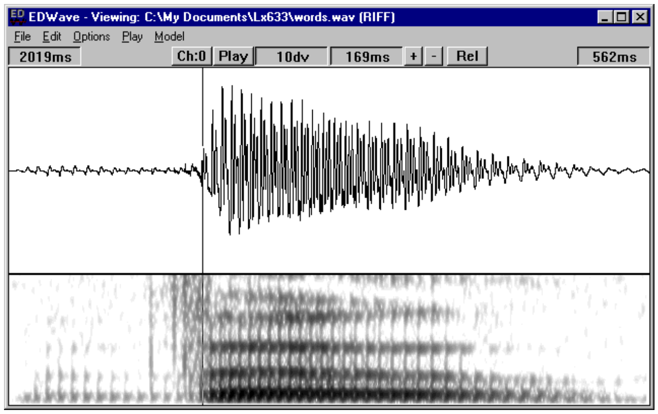
# [fu]




NOTES

/f/ has low energy overall and most of that at higher frequencies. Vocalic transitions should be similar to voiceless bilabial stop, but there may not be enough aspiration to clearly see them.

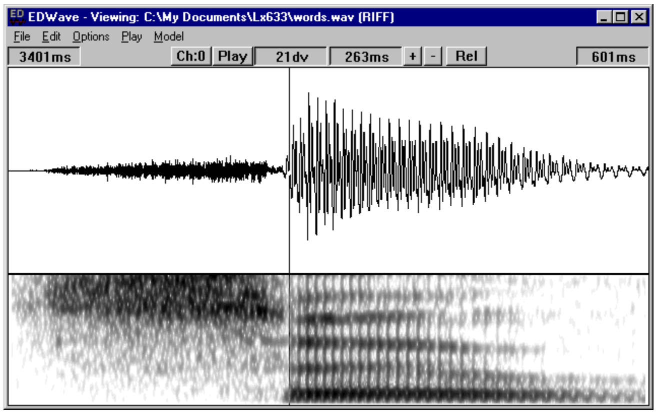
# [vu]




NOTES

/v/ shows the influence of voicing throughout constriction interval. Its transition pattern should parallel that of /f/, and transitions are often more visible because of continuous voicing.

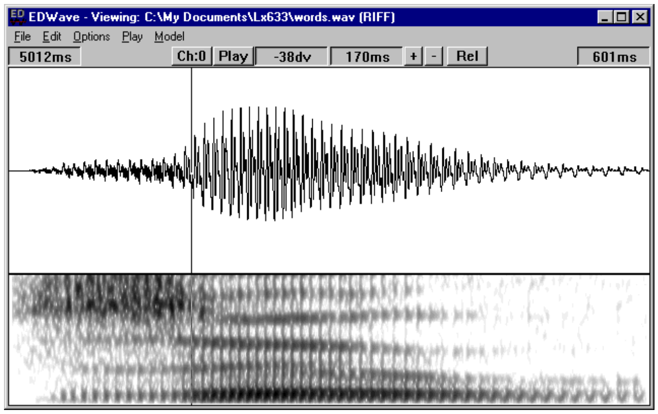
# [su]




NOTES

/s/ shows strong friction energy especially above 4-5 kHz. Transitions may be /t/-like but more gradual. Note the way F3 is excited by aspiration just before the onset of voicing. This is also happening with F2, but it is more difficult to see.

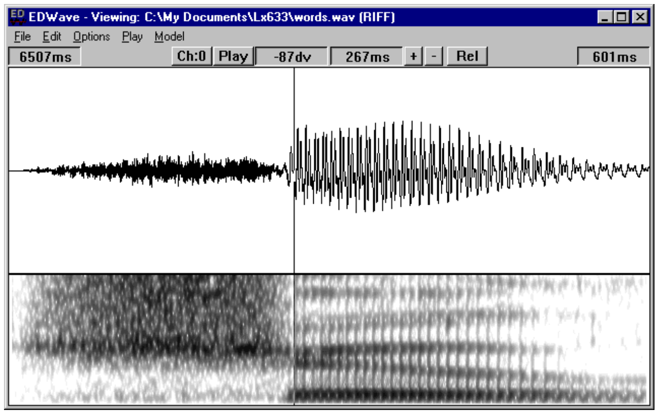
# [zu]




NOTES

For /z/, voicing can make lower formants visible in the closure interval. High frequency friction energy similar to /s/ may also be present. The constriction for /z/ in this case is not enough to decouple the front and back parts of the vocal tract (otherwise we would not be able to see F1-F3).

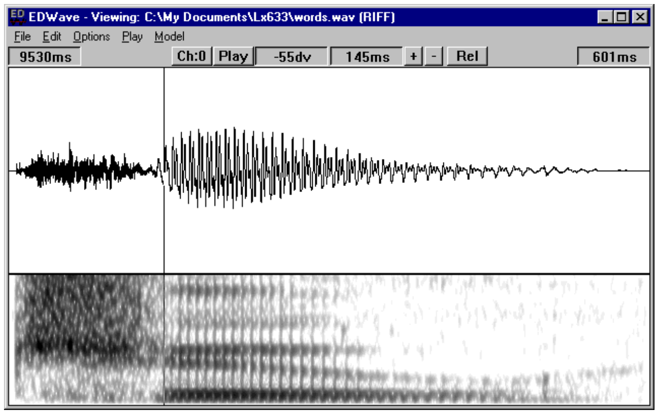
# [ʃu]




NOTES

/ʃ/, like /s/, shows strong friction energy but beginning at lower frequencies and often has decided peak in the region of F3. In fact, F2 & F3 may be merged in constriction interval. Transitions are like those for the release of a palatal stop or glide.

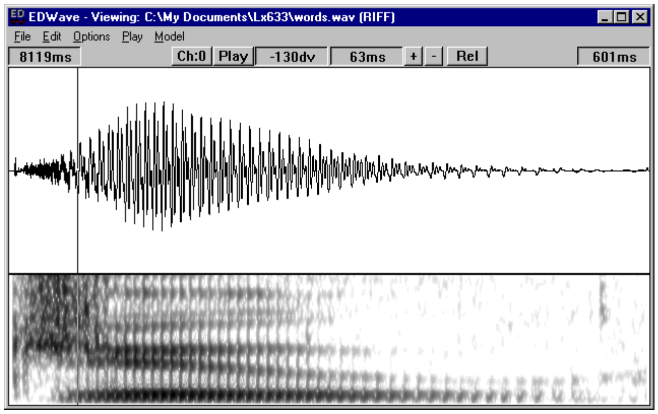
# [tʃ]




NOTES

The voiceless affricate has a sharp burst-like onset followed by frication. The frication interval is shorter than for the singleton fricative /ʃ/. There is a silent interval preceding the burst, but it was cut out of this figure.

# [dʒ]




NOTE

The voiced affricate /dʒ/ also has a sharp burst like onset and shorter fricative interval than for the singleton fricative /ʒ/. Voicing may or may not be visible throughout the closure and friction regions.

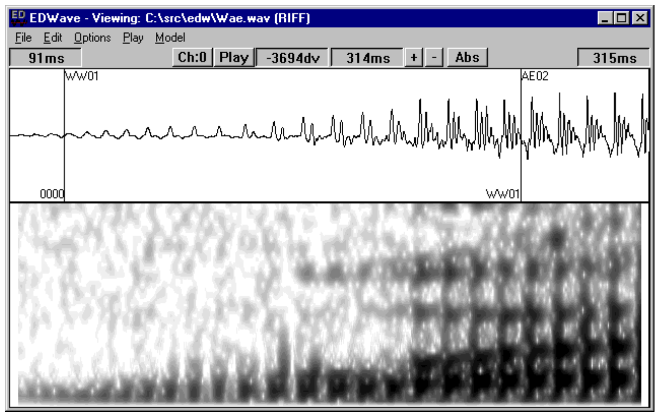
## Liquids and Glides

- Constriction less than for fricatives greater than for vowels
- Distinguished from stops by
  - Greater amplitude during constriction
  - Slower vocalic transitions than stops
- /l/ has additional features due to side passages

NOTES

This summarizes what we'll see in the next several slides.

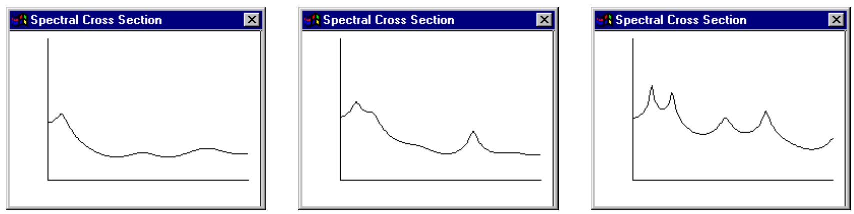
# [wæ]




NOTES

Constriction interval has very low F1 & F2. F3, when visible, is near that of following vowel, or maybe slightly higher. Transitions are much more gradual than for the voiced bilabial stop.

# [wæ] spectra




Murmur during constriction

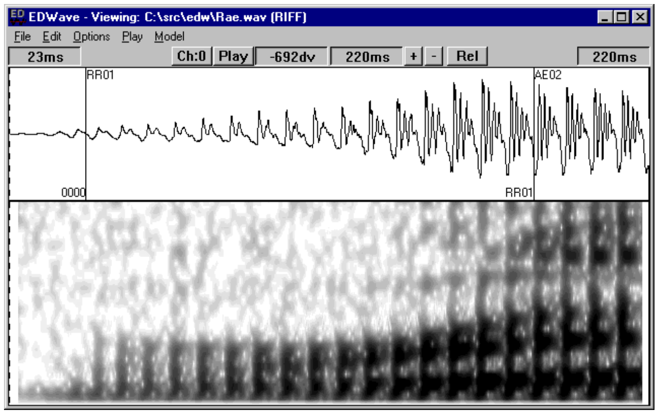
Release begun

Vocalic transition

NOTES

These figures show a sequence of spectra in stages from the murmur to the vocalic transition region. Note F1&F2 separating as release occurs. F3 fairly stable when it is finally visible.

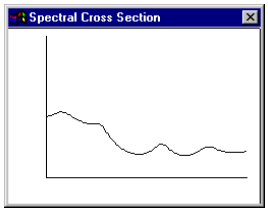
# [ɹæ]



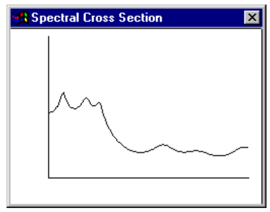

NOTES

F1, F2, and F3 are all low for /ɹ/. This low F3 is characteristic of /ɹ/ and of /ɹ/-like segments such as the vowel /ɚ/.

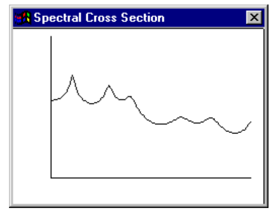
# [ɹæ] spectra




Murmur
during
constriction




Release begun

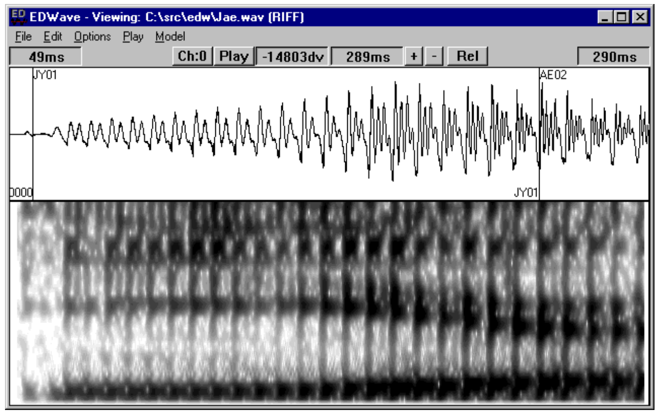



Vocalic transition

NOTES

Watch F3 emerge from F2 region and climb in this sequence of frames.

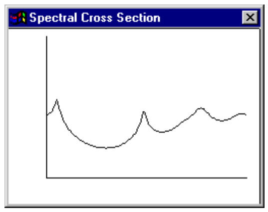
# [jæ]



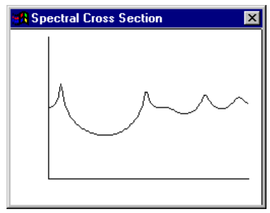

NOTES

Here F1 is low, but F2 and F3 are high and probably merged. They separate during the transition into the vowel. The frequency of the F2+F3 complex is higher for this palatal approximant than for a palatalized /g/ indicating that the constriction is further forward in the vocal tract.

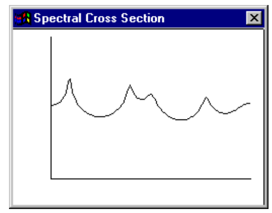
# [jæ] spectra




Murmur during constriction




Release begun

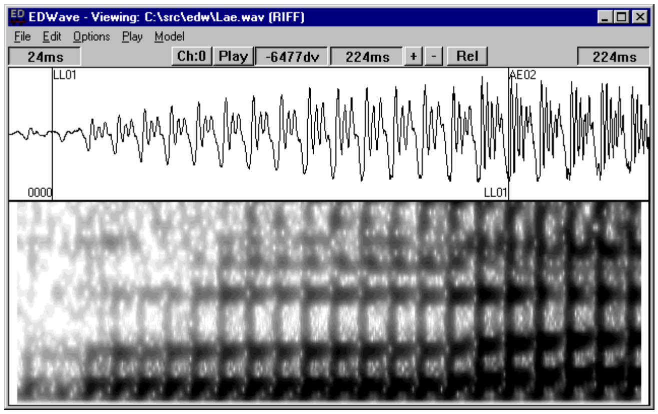



Vocalic transition

NOTES

Check for separation of F2 & F3.

# [læ]

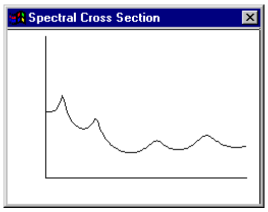



NOTES

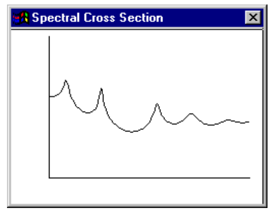
The lateral is a bit different. Transitions from constriction interval to vowel are often more rapid and may be almost step-like (compare to nasals below). Zeros and other resonances can appear during the constriction interval because there are typically two side channels for sound to propagate around the closed midline of the oral cavity.

F3 may be somewhat higher than for /w/ and stronger as well.

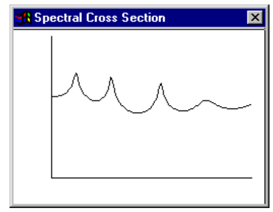
# [læ] spectra




Murmur
during
constriction




Release begun




Vocalic transition

NOTE

These LPC spectra may not show zeros in the constriction interval directly, but notice the broader bandwidth of the formants. F3 may be higher than neutral location which is contrastive with /w/.

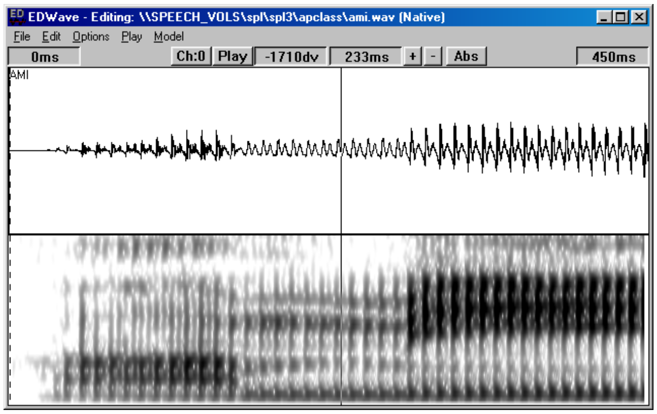
# Nasals

- Oral closure equivalent to stops at same place
- Velar port open
- Consequence:
  - Low frequency resonance ~250 Hz
  - Higher frequency resonances ~1000, 3000, 5000, etc.
  - Zeros depend on place of oral closure.

NOTES

Summarizes next series of slides.

# [m]




NOTES

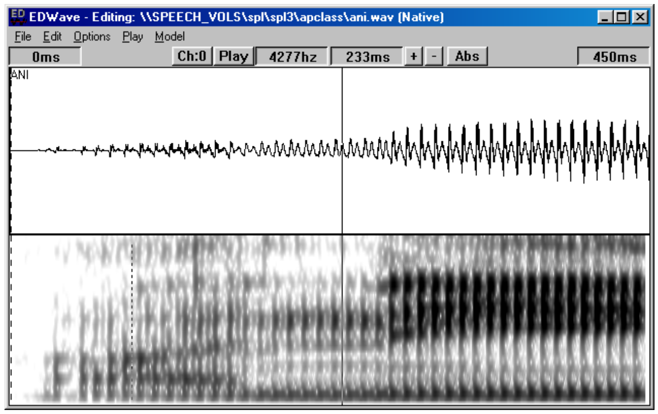
1) onset and offset (closure and release) of nasal very step-like due to abrupt coupling of nasal passage and rapid stop-like oral closure.

2) Resonances during constriction interval are due to nasal structure, not oral place of closure.

3) There are (potentially hard to find) zeros due to the location of the oral constriction.

4) There are oral constriction place cues in the vocalic transitions at closure and release. In this case the parallel falling and rising transitions suggest bilabial closure.

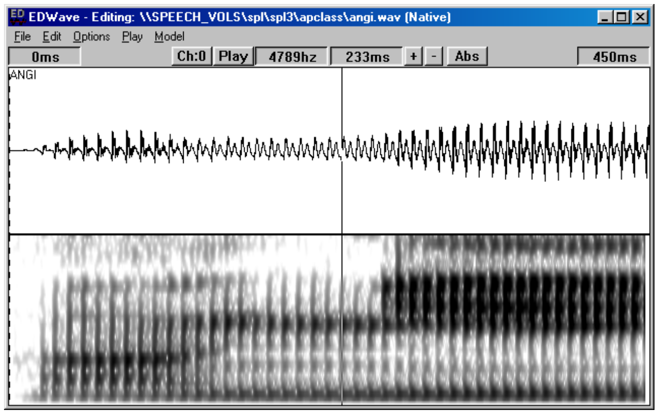
# [n]




NOTES

The differences between /m/ and /n/ are most prominent in the transitions into and out of the constriction interval. Note that F2 rises from the initial vowel into the constriction interval. F2 also rises during the release of the constriction because the F2 for /i/ is higher than the nominal F2 for the /n/ place of closure.

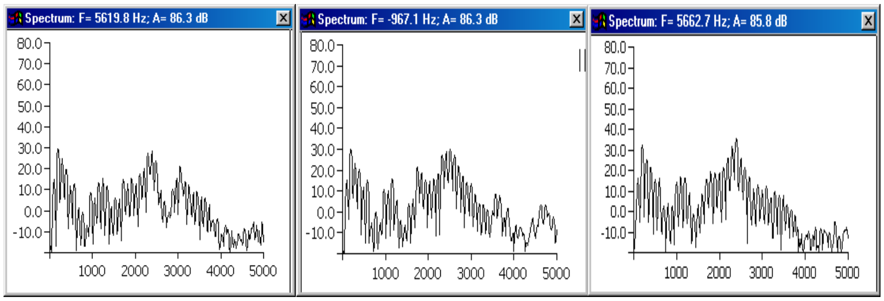
[ŋ]




NOTES

For the velar nasal, vocalic transitions resemble those of /g/. Note particularly the steep and long F2 transition which seems to merge with F3 and/or the nasal pole.

# Murmur Spectra




[m] [n] [ŋ]

NOTES

These three spectra were taken during the nasal constriction intervals of the nasals shown in the previous three slides.

See if you can find evidence of a zero that is lowest for /m/, higher for /n/, and highest for /ŋ/

# Coarticulation

- Change in the articulation of a segment attributable to surrounding segments.
- Types
  - Anticipatory
  - Carryover
- May change acoustic structure of a segment, but not its perceptual character.

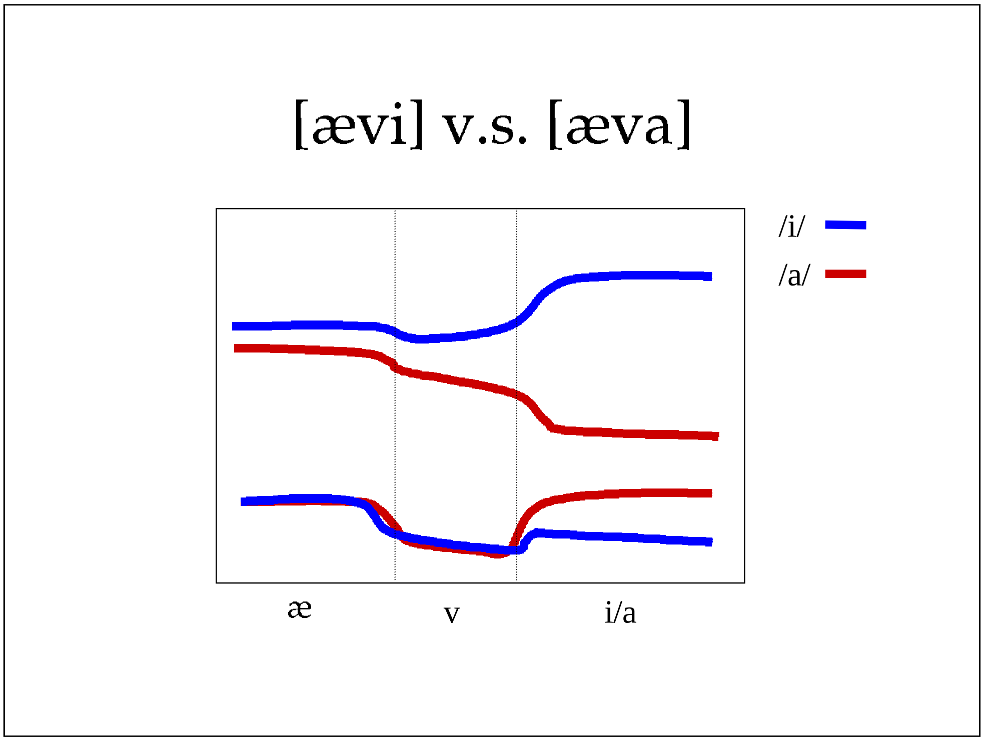
NOTES

Coarticulation results in shifting the articulation of segments toward the articulation of one or more surrounding segments. It can be thought of as representing an economy of movement principle wherein one tries to reduce the amount of articulator motion needed to realize a given sequence of phonemes. This economy principle is in opposition to the need to make the articulation of each segment sufficiently distinct and distinguishable.

In acoustic terms, coarticulation results in smoothing acoustic variation over time.

The effects of coarticulation are essentially imperceptible; we are not aware of the changes coarticulation introduces into the speech signal. Further, we may even depend upon the effects of coarticulation to make the speech signal easier to understand since it adds redundancy to the signal and slows the rate at which phonetic information must be extracted by spreading it over time.

A classic study of coarticulation was done by Ohman (1966) who showed that the articulation of a vowel in one syllable could be altered in anticipation of the articulation of an upcoming vowel in the next syllable. Martin and Bunnell (1980, 1981) showed that listeners actually make use of this anticipatory coarticulation information in decoding the acoustic speech signal.




NOTES

This figure illustrates a common type of anticipatory coarticulation. It shows schematic F1 & F2 patterns for the sequence [ævi] (blue lines) versus [æva] (red lines). In the [æ] of both cases, F1 is approximately the same, but F2 shows significant acoustic differences which anticipate the degree of fronting needed for the following vowel. This suggests that the talker articulated the [æ] further back in the oral cavity when in the context of a following [a] than in the context of a following [i]. The differences observed in F2 of the first vowel grow over the course of the intervocalic consonant [v] indicating still greater amounts of coarticulation present for the articulation of the consonant.

## Suprasegmental Structure

- Applies to:
  - Smallest unit syllables
  - Other units include {feet, phonological words, phonological phrases, intonation phrases}
- Involves:
  - Metrical structure (rhythm and timing)
  - Intonational structure (intonation contour and pitch accents)

NOTES

1) A foot is a unit of rhythm beginning with a stressed syllable and including all subsequent unstressed syllables up to the next stressed syllable.

2) Phonological words are in many cases simple lexical items, but can include more than one word when adjacent lexical items merge into a single phonological unit. The clitic group is an example of this where function words may be thought to have no boundary between them and an adjacent content word. For example, “the dog” may form a single articulatory entity.

3) In metrical phonology, syllables represent rhythmic “beats”. English as a strong tendency to alternation strong and weak beats, i.e. stressed syllables separated by unstressed syllables.

# Types of information in Prosody

- Lexical - word stress patterns
- Syntactic - Boundary effects
  - Boundary Tones
  - Phrase-final lengthening
- Semantic/pragmatic effects
  - Pitch accents
  - Focus

NOTES

1) Lexical stress is the pattern of syllabic prominence that is specifically identified with a word. It is primarily a lexical stress difference that distinguishes the verb conVICT from the noun CONvict (upper case representing the stressed syllable).

2) Boundary tones are things like the rising intonation at the end of a question or the falling intonation at the end of a sentence.

3) Phrase final lengthening is the tendency for the final syllable of phrases to be substantially longer in duration that the same syllable would be in a phrase medial position. We use the drawing out of syllables at the end of phrases (and to some extent smaller units as well) to signal the presence of a boundary.

4) Pitch accents are the high or low peaks/troughs in intonation which mark words for special emphasis or to guide listeners in interpreting utterances.

5) Focus is the placement of special prominence (including pitch accents, changes in duration, and changes in amplitude) on certain words in a sentence for pragmatic reasons such as to emphasize a particular contrast with a prior utterance.

# Rhythm

- The (generally) alternating pattern of prominence in speech.
- Speech can be viewed as a succession of metrical *feet*, each beginning with a stressed syllable and containing all unstressed syllables up to the next stressed syllable.
- For English, the “ideal” pattern is alternating strong-weak syllables.

NOTES

A slide to reiterate points made in the last notes section.

# Types of Timing

- Stressed timed
  - intervals between stressed syllables are *isochronous*
  - Example - English
- Syllable timed
  - Syllable durations are isochronous
  - Example - Japanese

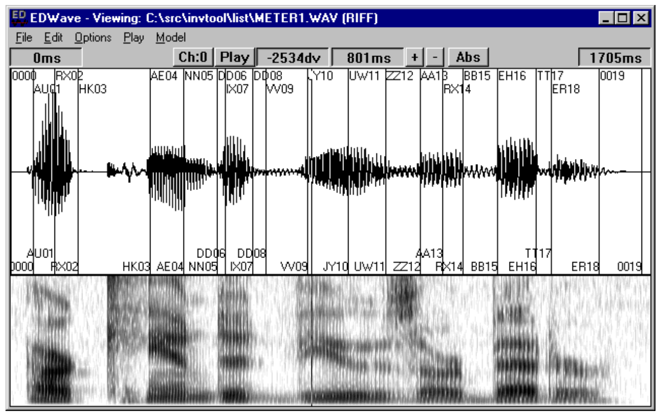
NOTES

We should examine the concept of timing in speech a bit more. Our linguistic intuition about the timing of English is that it has regularly timed stressed syllables (sort of the down beats) and a variable number from 0 to 5 or more weak beats between the major down beats. This suggests that stressed syllables should fall at regular intervals in English; a property called *isochrony*.

Many languages do not have this (intuitive) rhythmic structure. For example, in languages such as Japanese the tendency is for each syllable to have about the same duration. Such languages are called syllable-timed rather than stress-timed.

A large number of investigators have looked for the isochrony of stressed syllables in English, but despite our intuition, isochrony turns out to be the exception rather than the rule. It is no longer clear whether there is any broad metrical scheme to which we fit segment durations. Instead, segments may have inherent durations which are modified by a number of contextual and global constraints to result in the actual timing of English and other languages.

# Our [k$^h$ændɪd] views...




| | | | |
|---|---|---|---|
| 178 | 501 | 524 | 501 |

| | | | |
|---|---|---|---|
| 365 | 425 | 505 | 409 |

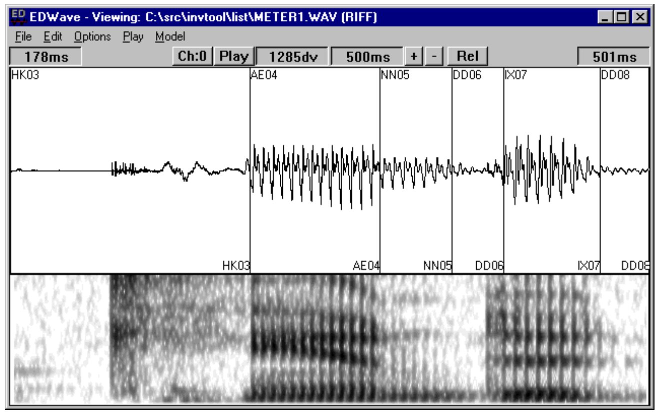
NOTES

This and the next several figures illustrate how timing can change to adjust for the rhythmic pattern of utterances, that is, for the pattern of stressed and unstressed syllables in an utterance.

The sentence is *our candid views are better* and the bars underneath the spectrogram show two ways of dividing the utterance into feet. The top bar of numbers divides the utterance at syllable onsets for the stressed syllable heading each foot. The first word *our* is unstressed and is considered unfooted. The remaining feet each contain two syllables (`*candid*|*views are*|`*better*) and vary in duration from 501 to 524 msec.

The bottom bar divides the utterance into feet, but measures from the vowel onset of each stressed syllable to that of the next stressed syllable. This tends to load part of the consonant onset of one foot into the end of the previous foot but makes sense in a perceptual sense because the perceived location of beats in syllables is closer to the vowel onset than to the phonological start of the syllable. In this example, however, the syllables are of more nearly equal duration when measured in the phonologically motivated manner.

# [k$^h$ændɪd]

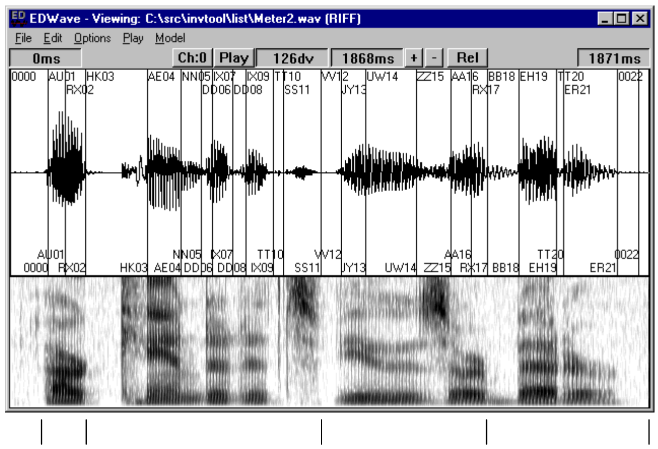



188 101 57 41 75 39

NOTES

This figure shows just the durations of the individual segments in the word *candid*. It illustrates how the stressed first syllable is greater in duration than the unstressed second syllable.

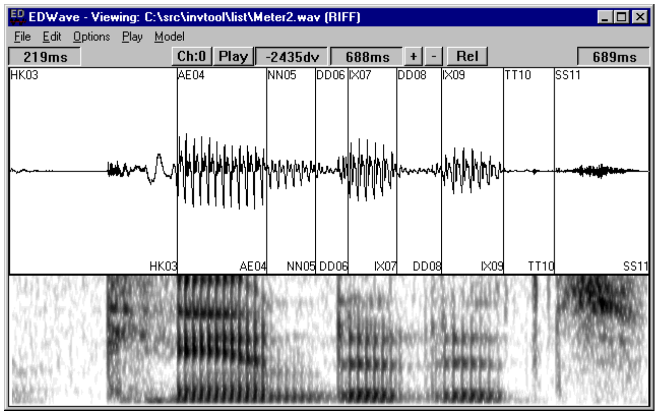
# Our [kʰændɪdɪts] views...




NOTES

This is the sentence *Our candidate's views are better*. It differs from the previous sentence in the addition of one unstressed syllable to the first foot. On the view that English has isochronous feet, the duration of the first foot should remain unchanged (with all its syllables compressed). Although times are not indicated, the foot boundary marks clearly are not equally spaced and the three-syllable foot is considerably longer than the two syllable feet.

# [k$^h$ændɪdɪts]




180 97 52 35 53 48 67 54 103

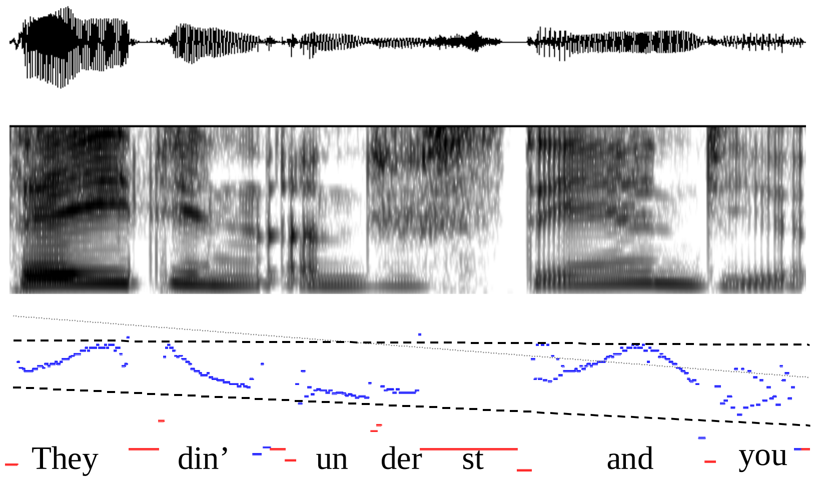
*188 101 57 41 75 39*

224

NOTES

This figure focuses on just the word *candidates* to compare it to the word *candid* in the previous utterance. The top row of numbers below the spectrogram are the durations in msec of each segment. The second row of numbers in italics are the measurements of the corresponding syllables in the word *candid*. Although the addition of the third syllable has added 224 msec, all the other syllables (and every segment except the second [d]) is shorter in this utterance than it was in the previous one. Adding another syllable to the foot has had the effect of compressing the durations of the first two syllables--or their segments--compared to those of an utterance containing only those two syllables.

# Intonation Contour




NOTES

This is a waveform, spectrogram, and F0 contour for the utterance *They didn't understand you.* In the F0 contour, voiced regions are blue while voiceless regions are red. It illustrates a number of principles of intonation. The darker dashed lines above and below the F0 contour trace the upper and lower bounds of the pitch contour. The bottom line shows the expected (Pierrehumbert, 1980) declination of F0 over the utterance, however, the top line does not. This is in disagreement with Pierrehumbert's model in which the upper lines is expected to have greater negative slope than the lower line (illustrated by the lighter dotted line). Exceptions to nearly any generalization in speech are not unusual.

1) There are two pitch accents present: the first on *they*; and the second on the stressed syllable of *understand*. When a polysyllabic word receives a pitch accent, it aligns to the primary stressed syllable of the word (only one syllable of a word receives primary lexical stress).

2) As is common for pitch accents, F0 rises over the nucleus of the accented syllable and peaks toward the end of the nucleus.

3) The fall in F0 over the word *didn't* is (arguably) not information bearing; it merely connects F0 at the end of the pitch accent to F0 in subsequent syllables.

4) The two low F0 syllables in *under* may simply be due to the fact that F0 tends to droop between pitch accents. It is also possible that there is a **low** tone here which is coupled with the subsequent **high** tone to add further strength to the second pitch accent.

5) F0 in the second accent is about equal to F0 in the first one. Since pitch starts out high and declines over an utterance, a pitch accent late in an utterance does not need to reach the same F0 value as one early in the sentence to be heard as equally prominent. Because these F0 values are equal, it is likely that the second accent will be heard as more prominent than the first one.

6) F0 declines very rapidly after the second accent due to the presence of a low boundary tone at the end of the utterance.