HCS 7367
Speech Perception Lab

Dr. Peter Assmann
Fall 2011

Animation example

```
x=fgen(200,60,0,1000,10000);
fp(x,10000);
axis tight
set(gca,'nextplot','replacechildren');
for j = 1:20
   f=(j-1)*200+200;
   x=fgen(f,60,0,1000,10000);
   fp(x,10000);
   F(j) = getframe;
end
movie(F,20)
```

Consonant production

- American English has 24 consonant phonemes
- The standard phonetic classification system uses three main features to classify the consonants:
  - **Place of articulation**
  - **Manner of articulation**
  - **Voicing**

English Consonants

Manner of articulation	Place of articulation →	Bilabial		Labio-dental		Dental		Alveolar		Palatal		Velar		Glottal
	Voicing →	–	+	–	+	–	+	–	+	–	+	–	+	–
	Stop	p	b					T	d			k	g	
	Nasal		m						n				ŋ	
	Fricative			f	v	θ	ð	s	z	ʃ	ʒ			h
	Affricate							tʃ	ʤ					
	Approximant		(w)						ɹ		j		w	
	Lateral								l					

Ladefoged, P. (1993). *A Course in Phonetics*. Harcourt Brace: Fort Worth. 3rd Ed.

Consonant classification

- **Place of articulation**
  - bilabial, labiodental, dental, alveolar, palatal, velar, glottal
- **Manner of articulation**
  - stops, nasal, fricatives, affricates, approximant, lateral
- **Voicing**
  - voiced, voiceless

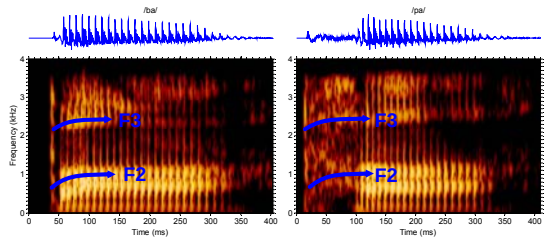
Stop consonants in English

Manner of articulation = stop consonants

Place of articulation

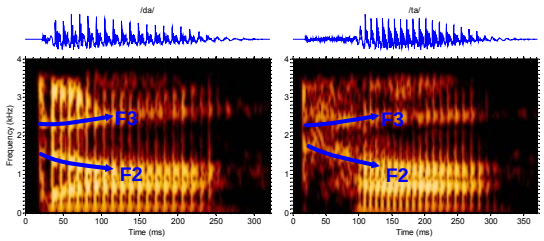
Voicing	bilabial	alveolar	velar
voiced (d)	/ba/	/da/	/ga/
voiceless (s)	/pa/	/ta/	/ka/

Formant transitions: /ba/ and /pa/



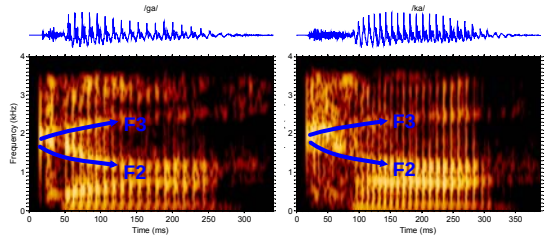
- **Bilabials: low F2 and F3 onset frequencies**

Formant transitions: /da/ and /ta/



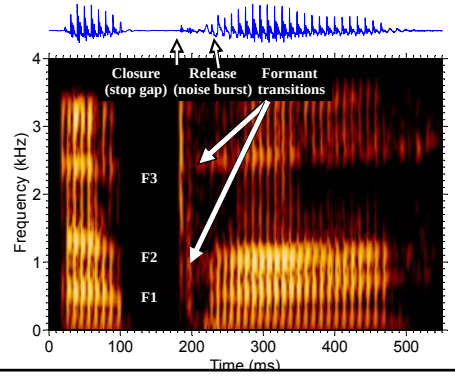
- **Alveolars: high F3 and relatively high F2 onset frequencies**

Formant transitions: /ga/ and /ka/



- **Velars: intermediate F2 and F3 frequencies; F2 and F3 appear to converge near the point of consonant release.**

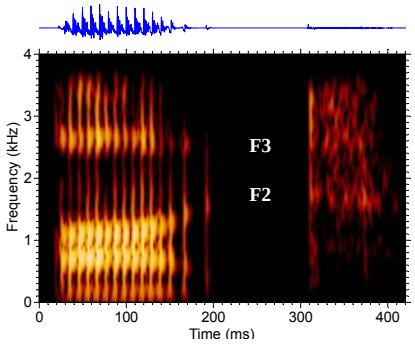
Bilabial stop



/apa/

- **low F2**
- **low F3**

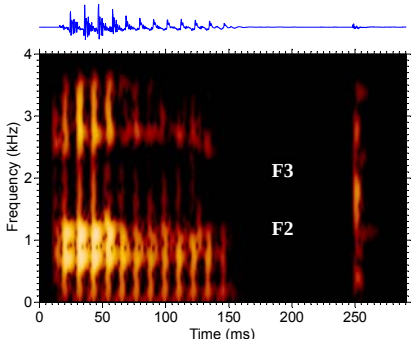
Alveolar stop



/at/

- **high F2**
- **high F3**

Bilabial stop



/ap/

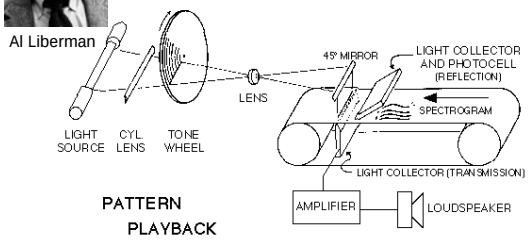
- **low F2**
- **low F3**

Continuum experiments

- Generate a continuum of sounds that span two phonetic categories. For example:
  - Voicing continuum: vary the voice onset time (VOT) in small steps (e.g. 10 ms) between the category endpoints, /pa/ and /ba/
  - Place continuum: vary formant transitions (F2 alone, or F2 and F3 combined) to see if this acoustic cue controls listeners' judgments of place of articulation, /ba/ vs. /da/.

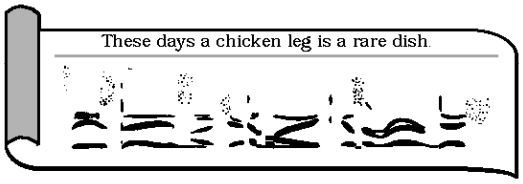
Pattern Playback
Haskins Labs

Al Liberman



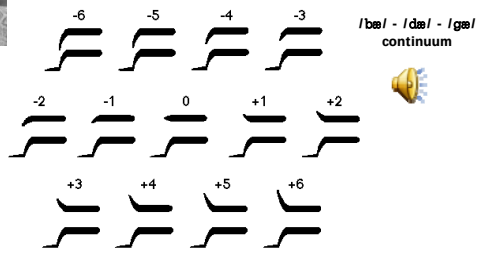
Source: http://www.haskins.yale.edu/haskins/MISC/PP/diagram.html

Pattern Playback
Haskins Labs



Source: http://www.haskins.yale.edu/haskins/MISC/PP/diagram.html

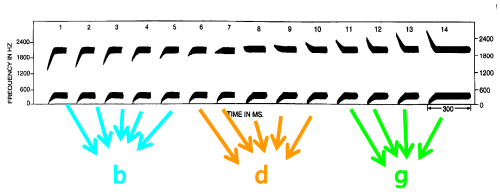
Pattern Playback
Haskins Labs



Source: http://www.haskins.yale.edu/haskins/MISC/PP/diagram.html

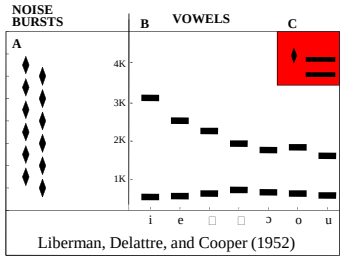
Categorical Perception of stop consonants

Equal acoustic changes → unequal auditory percepts



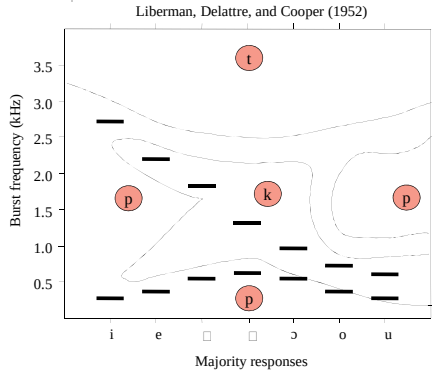
Liberman, Harris, Hoffman, and Griffith (1957)
Journal of Experimental Psychology 54, 358-368

Invariance problem for stop consonants



Liberman, Delattre, and Cooper (1952)

Invariance problem for stop consonants

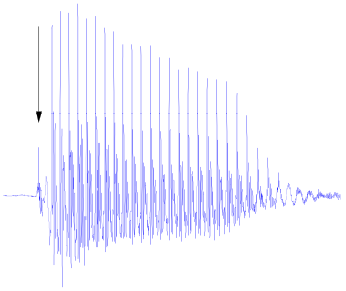


Burst spectrum

- Blumstein and Stevens (1979, 1980) **invariance hypothesis** (static spectral shape model)
  - The shape of the spectrum – sampled at the time of burst release – provides invariant cues specifying the place of articulation for the English stop consonants.

Static spectral shape model

Step 1: Find consonant burst onset



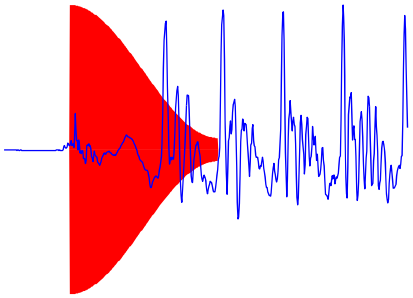
Acoustic landmarks

- Use the built-in Matlab command *ginput.m* to find the *x* and *y* coordinates of events in a waveform plot (such as the burst onset).
  - **Example**: *plot a dashed line to mark burst onset*

```
>> [x,y]=ginput(1);
>> plot([x x],[min(y) max(y)],':');
```

Temporal window

Step 2: Position a 25.6 ms half-Hamming window



Temporal window

```
>> load ba.mat
>> plot(ba(1:850)); % plot first 850 points
```

Temporal window

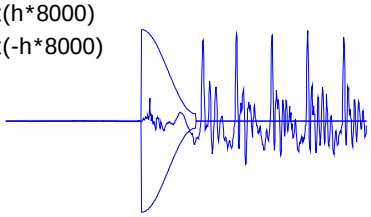
- Create time window

```
% second half of a 256-point Hamming window
>> h=hamming(256);
>> h(1:128)=[];
>> h=[ zeros(319,1); h; zeros(403,1) ];
```

Temporal window

- Superimpose temporal window on syllable:

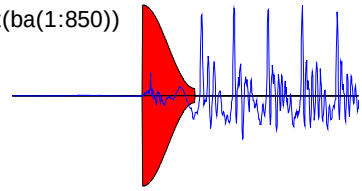
```
>> plot(ba(1:850))
>> hold on;
>> plot(h*8000)
>> plot(-h*8000)
```



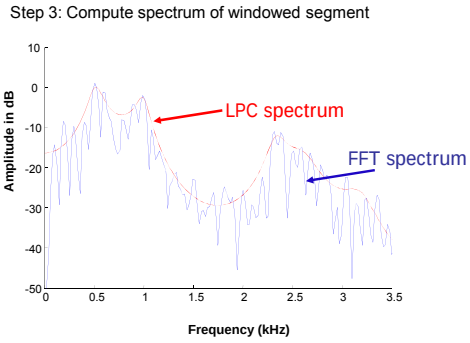
Temporal window

- Same, but with a red color patch

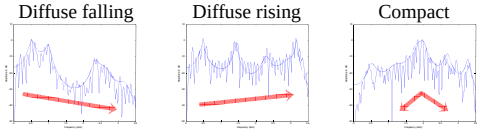
```
>> patch(1:length(h),h*max(abs(ba)),'r')
>> hold on;
>> patch(1:length(h),-h*max(abs(ba)),'r')
>> plot(ba(1:850))
```



Acoustic landmarks



Static spectral shape model



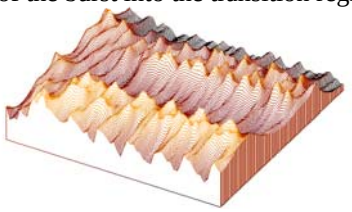
Burst spectra of syllable-initial stop consonants

Static spectral shape model

- Predictions:
  1) spectral shape for a given place of articulation is the same for different talkers and in different phonetic contexts.
  2) acoustic modifications that distort other aspects of the syllable but preserve the spectral shape near the burst do not affect consonant identity.

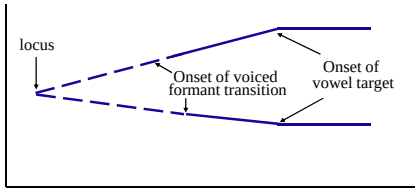
Dynamic properties

- Kewley-Port (1983) hypothesized that the identification of stop consonants is based on *time-varying changes* in the spectrum from the onset of the burst into the transition region.



Locus Equations

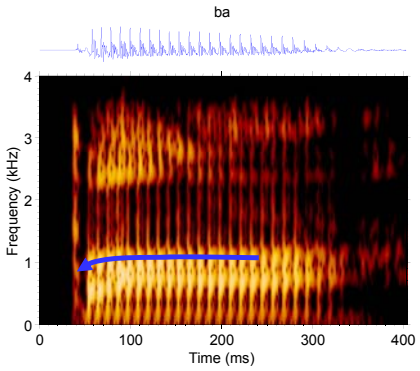
- Delattre et al. (1954) described the *locus* as the frequency location of F2 extrapolated back in time to the consonant release.



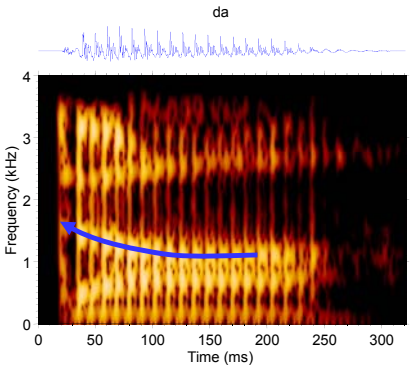
Locus equations

- Sussman et al. (1991, 1993) proposed that **locus equations** provide *invariant relational cues* for the perception of place of articulation in stop consonants.
- Locus equations describe the relationship between the frequency of F2 at burst onset and in the vowel.

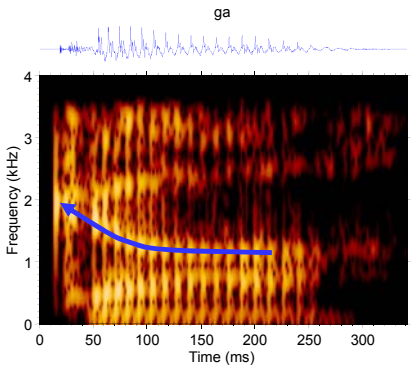
Locus equations



Locus equations

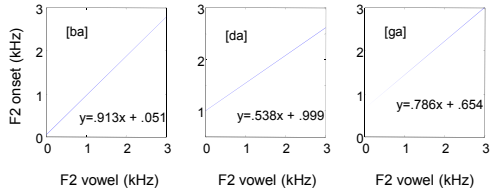


Locus equations



Locus equations

Sussman et al. (1991)



Lab assignment 2

Acoustic cues for place of articulation in stop consonants

Read the background papers:

- Blumstein SE, Stevens KN. (1979). Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants. J Acoust Soc Am. 66(4):1001-17.
  http://www.utdallas.edu/~assmann/hcs7367/blumstein_stevens79.pdf
- Sussman HM, McCaffrey, HA, Matthews SA (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. J Acoust Soc Am. 1991 90(3):1309-1325.
  http://www.utdallas.edu/~assmann/hcs7367/sussman91.pdf

Lab assignment 2

Download the recordings:

- Download the recordings of the CV syllables, /ba/, /da/, /ga/, /pa/, /ta/, /ka/ from the class web page:
  http://www.utdallas.edu/~assmann/hcs7367/cvc.zip
- Use these recordings to get started with the measurements.

Download stop consonants

ba	da	ga
pa	ta	ka

http://www.utdallas.edu/~assmann/hcs7367/cvc.zip

Make spectrogram plots

load ba.mat	load da.mat	load ga.mat
subplot(2,3,1)	subplot(2,3,1)	subplot(2,3,1)
sp(ba,rate)	sp(da,rate)	sp(ga,rate)
load pa.mat	load ta.mat	load ka.mat
subplot(2,3,2)	subplot(2,3,2)	subplot(2,3,2)
sp(pa,rate)	sp(ta,rate)	sp(ka,rate)

Identify formant transitions and bursts

Lab assignment 2

Supplement these with your own recordings:

Syllable-initial stops			Intervocalic stops			Syllable-final stops		
/bi/	/bɑ/	/bu/	/ ibi/	/aba/	/ubu/	/ ib/	/ab/	/ub/
/di/	/dɑ/	/du/	/ idi/	/ada/	/udu/	/ id/	/ad/	/ud/
/gi/	/gɑ/	/gu/	/ igi/	/aga/	/ugu/	/ ig/	/ag/	/ug/
/pi/	/pɑ/	/pu/	/ ipi/	/apa/	/upu/	/ ip/	/ap/	/up/
/ti/	/tɑ/	/tu/	/ iti/	/ata/	/utu/	/ it/	/at/	/ut/
/ki/	/kɑ/	/ku/	/ iki/	/aka/	/uku/	/ ik/	/ak/	/uk/

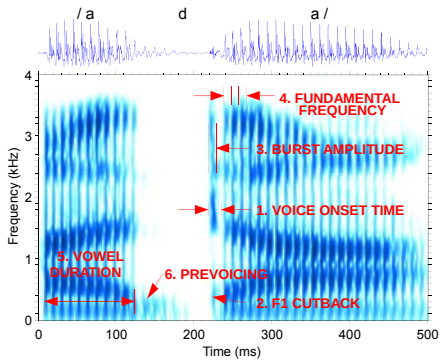
Static spectral shape model

- Use (a subset) of these recordings to investigate the hypothesis that the shape of the spectrum – sampled at the time of burst release – provides invariant cues specifying the place of articulation for the English stop consonants. Use FFT and LPC analyses to generate spectral plots for each of the recordings. Classify these as *diffuse falling*, *diffuse rising* or *compact*. Can you think of potential problems for this theory?

Locus equation model

- Locus equations describe the relationship between the frequency of F2 at burst onset and in the vowel.
- Use **Matlab** or **wavesurfer** to estimate the F2 frequency at burst onset and F2 in the vowel for each of the recorded syllables.
- Plot the locus equations published by Sussman et al. 1991 and superimpose your measured values. How well do the observed data match the predictions of the model?

Voicing cues in stop consonants



Voicing cues in stop consonants