

MIT OpenCourseWare
<http://ocw.mit.edu>

24.963 Linguistic Phonetics
Fall 2005

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

24.963
Linguistic Phonetics
Quantal Theory

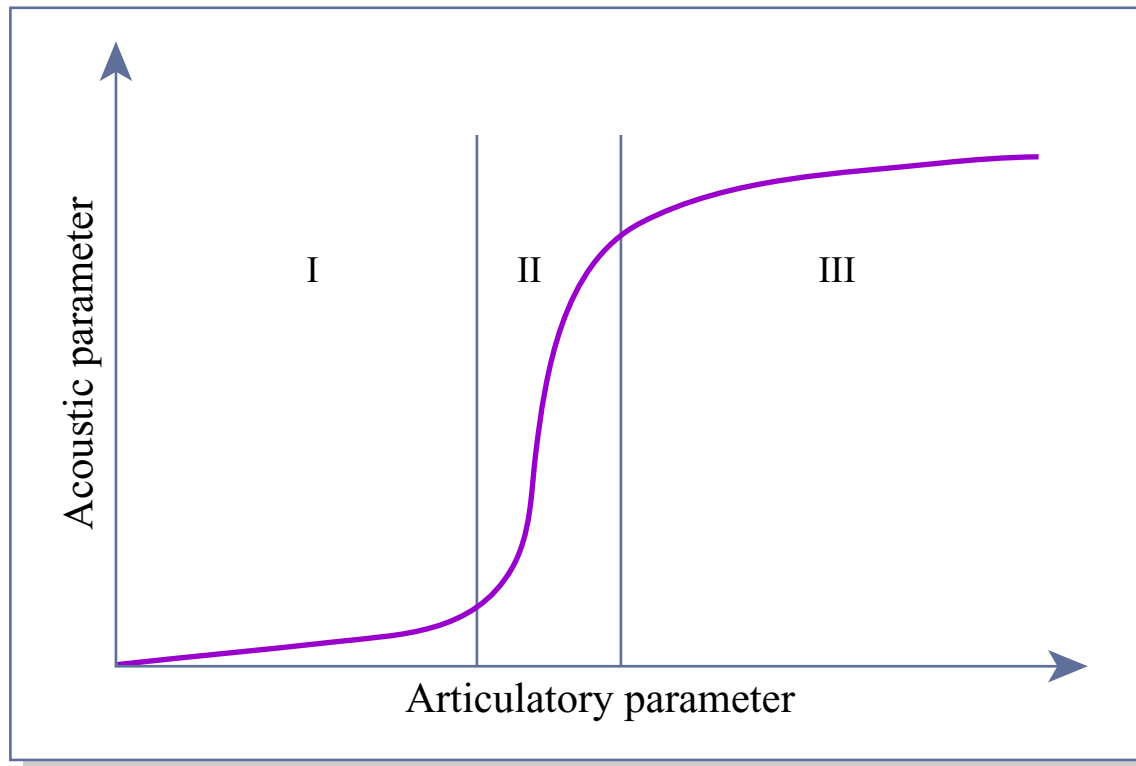


Image by MIT OpenCourseWare. Adapted from Stevens, K. N. "On the Quantal Nature of Speech." *Journal of Phonetics* 17 (1989): 3-46.

Reading for week 7:

- Johnson chapters 7 and 8.

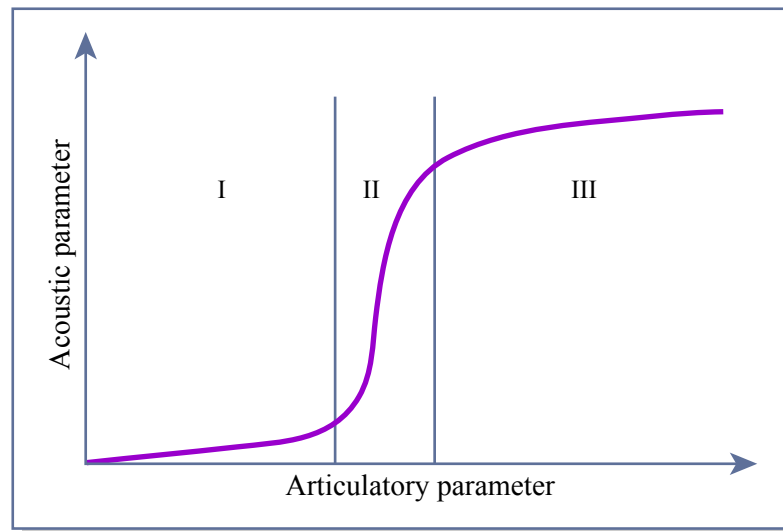
Assignments:

- 3rd acoustics assignment

Quantal Theory

Quantal relationship between articulatory and acoustic parameters
(Stevens 1972, 1989, etc)

- The acoustic difference between I and III is large - qualitatively different (Johnson's example: glottal aperture and voicing).
- The acoustic parameter is relatively insensitive to change in the articulatory parameter within regions I and II, hence:
 - articulation need not be precise.
 - continuous movement through the region will yield acoustic steady states.



Quantal Theory

Claim:

- Linguistic contrasts involve differences between regions I and III.
- More specifically, quantal relations provide the basis for distinctive features:
 - ‘The articulatory and acoustic attributes that occur within the plateau-like regions of the relations are, in effect, the correlates of the distinctive features’
(p.5)

Voicing and glottal aperture

- This example is not Stevens's, but it's a nice illustration of the insight behind the quantal theory and the potential complications it faces:
- Gradual change in articulatory parameters can result in abrupt, qualitative change in acoustic output - voicing is qualitatively different from voicelessness.
- Glottal aperture is only one of many parameters that affects voicing - glottal tension and pressure drop across the glottis are relevant also. How does this affect the identification of quantal regions (particularly as a basis for features)?
- Languages also contrast breathy vs. modal voice vs. creaky voice. Are these quantal distinctions?

Voicing and glottal aperture

- Glottal aperture is only one of many parameters that affects voicing - glottal tension and pressure drop across the glottis are relevant also. How does this affect the identification of quantal regions (particularly as a basis for features)?

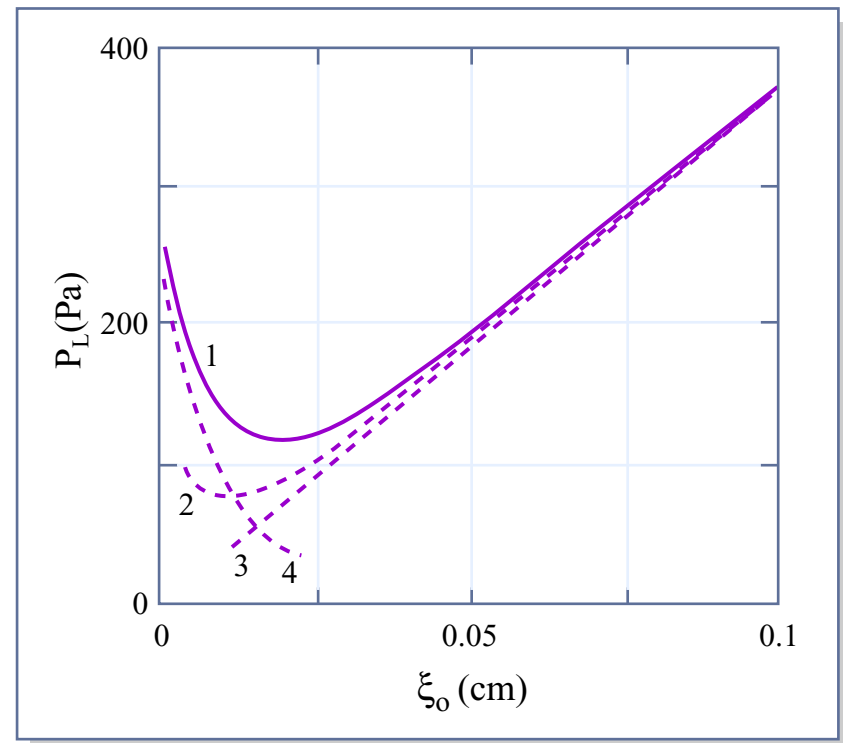
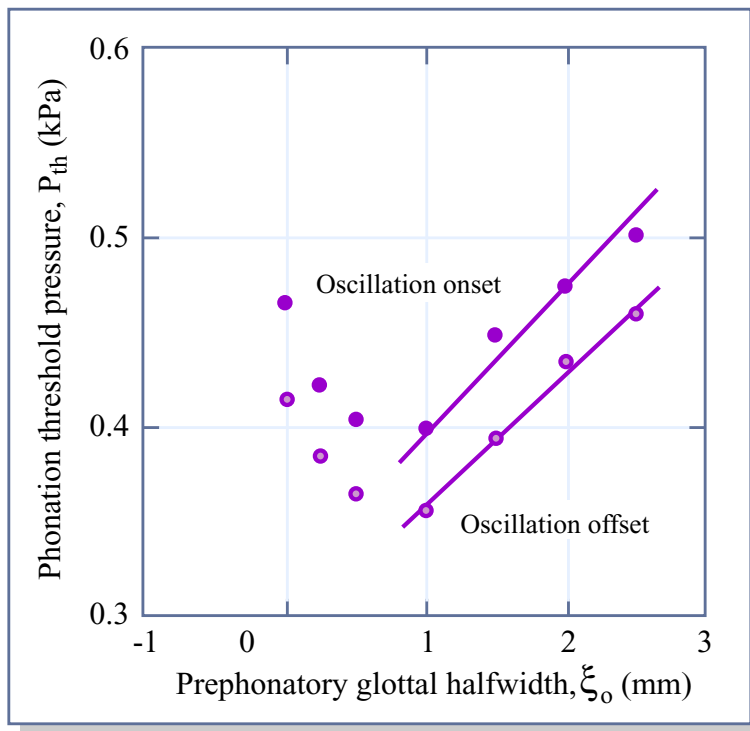
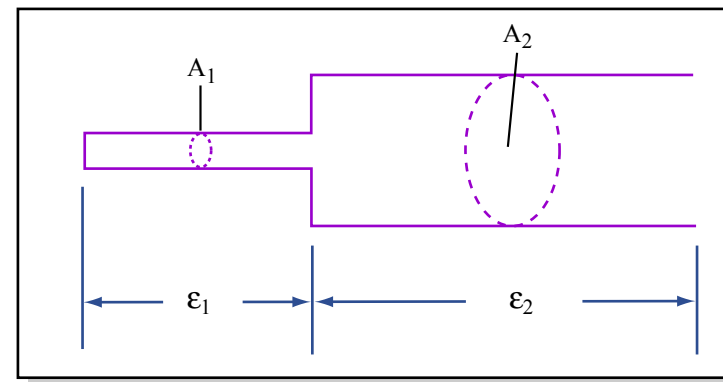
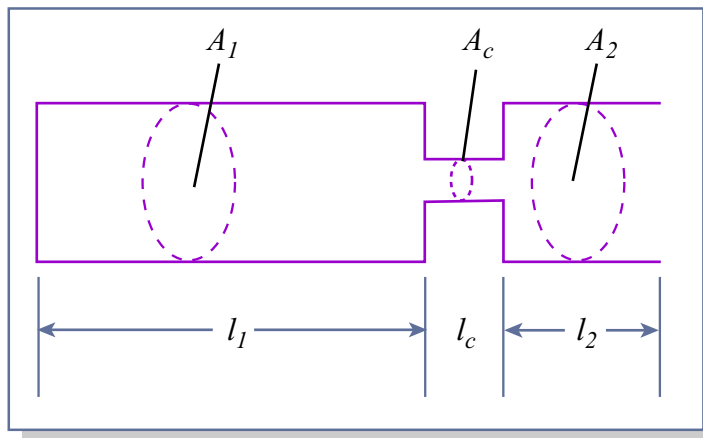
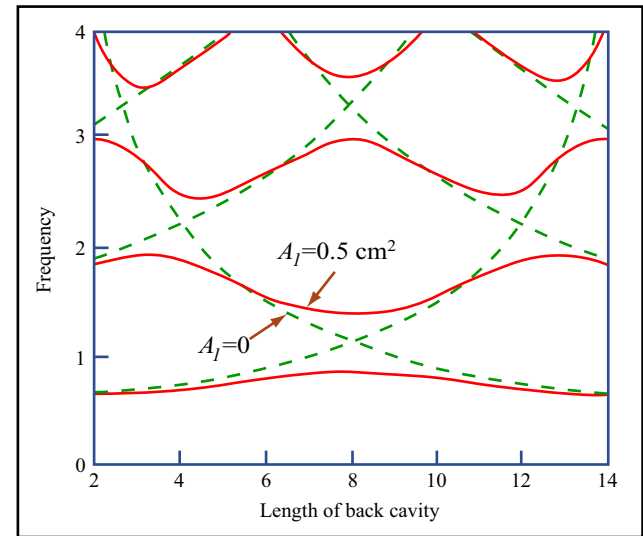
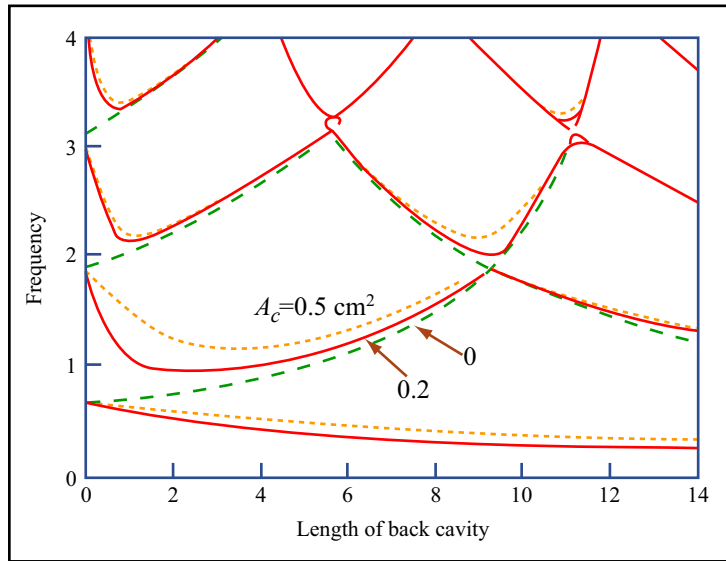


Image by MIT OpenCourseWare. Adapted from Titze, Ingo R., Sheila S. Schmidt, and Michael R. Titze. "Phonation Threshold Pressure in a Physical Model of the Vocal Fold Mucosa." *Journal of the Acoustical Society of America* 97 (1995).

Image by MIT OpenCourseWare. Adapted from Lucero, J. C. "The Minimum Lung Pressure to Sustain Vocal Fold Oscillation." *Journal of the Acoustical Society of America* 98 (1995): 779-784.

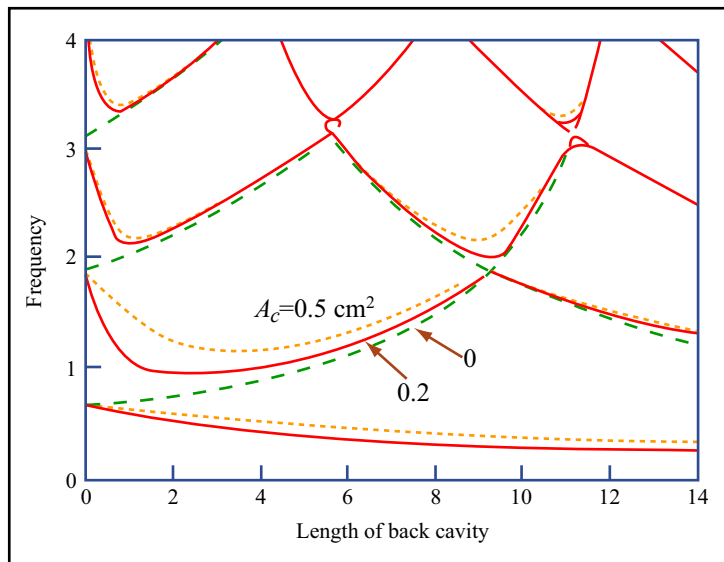
Quantal theory applied to vowels

- Regions of stability (quantal regions) for vowel formant frequencies occur where two formants converge.

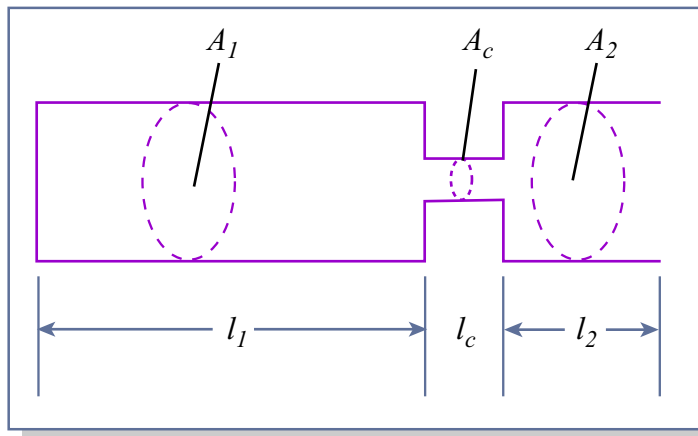


Quantal theory applied to vowels

- The three most common vowels cross-linguistically are [i, a, u]. Stevens argues that these are quantal vowels.



- High front [i] is produced at the convergence of F2 and F3 created by a narrow constriction in the palatal region.



Quantal theory applied to vowels

- Regions of stability (quantal regions) for vowel formant frequencies occur where two formants converge.
- Low [a] is produced at the convergence of F1 and F2 created by a narrow back cavity and a wide front cavity of equal length.

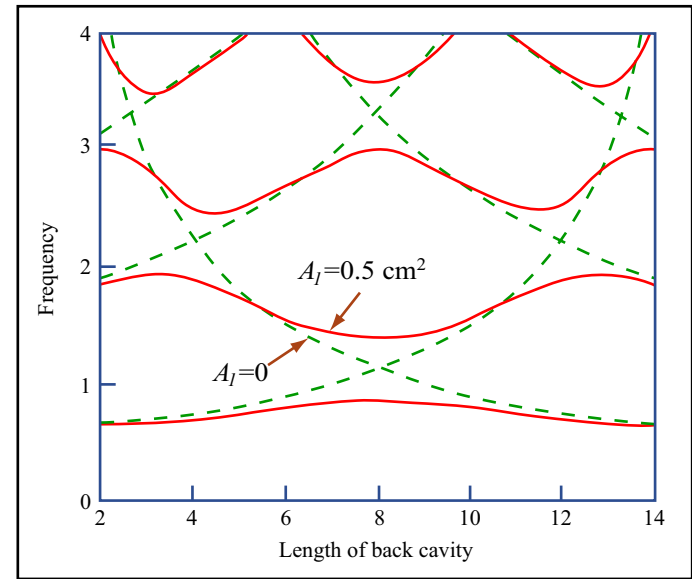


Image by MIT OpenCourseWare. Adapted from Stevens, K. N. "On the Quantal Nature of Speech." *Journal of Phonetics* 17 (1989): 3-46.

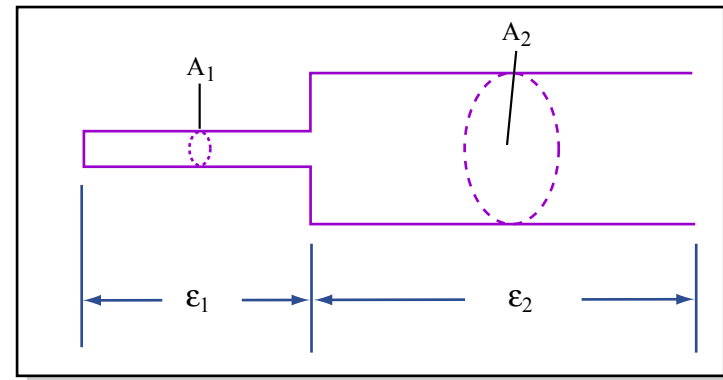
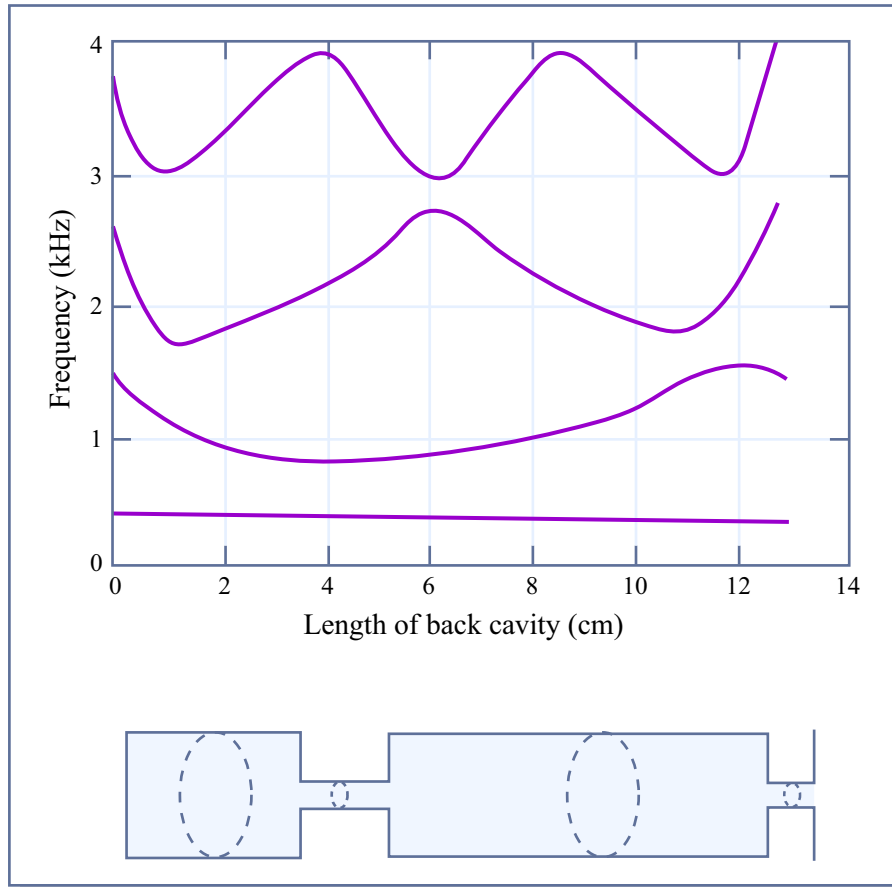


Image by MIT OpenCourseWare.

Quantal theory applied to vowels

- Regions of stability (quantal regions) for vowel formant frequencies occur where two formants converge.

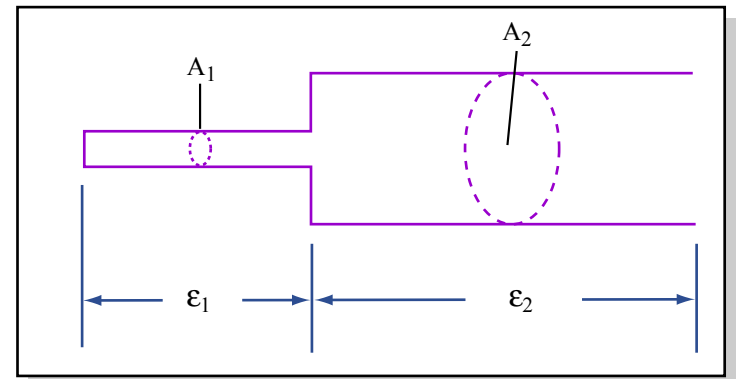
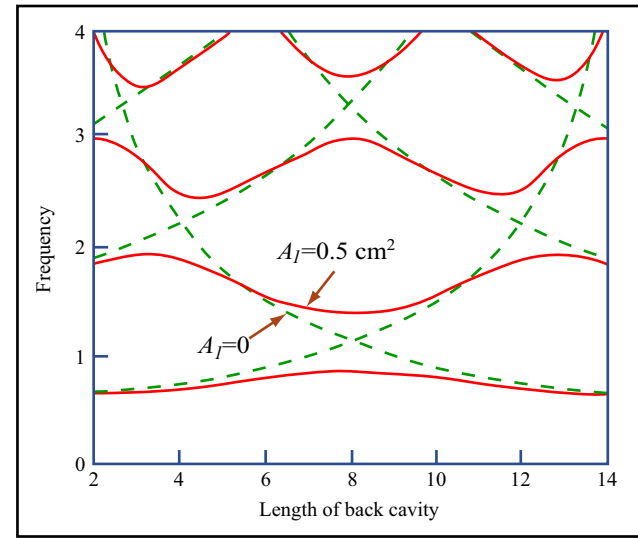


- High back rounded [u] is produced near a minimum in F2, in a region where F1 is relatively stable.

Image by MIT OpenCourseWare. Adapted from Stevens, K. N. "On the Quantal Nature of Speech." *Journal of Phonetics* 17 (1989): 3-46.

Quantal theory applied to vowels

- Are all convergences of F1 & F2 or F2 & F3 quantal regions?
- Some are not anatomically feasible - e.g. convergence of F2 and F3 at 12cm.
- Convergence of F2 and F3 at 4cm is said to be quantal vowel [æ].
- Note that this vowel is cross-linguistically relatively unusual.



Quantal theory applied to vowels

- Are mid vowels [e, o] quantal?
- We have suggested that the difference between high and mid vowels can be modeled as an increase in the area of the constriction.
- What is the effect on the formants?

$$\Delta F = \frac{c^2}{2\pi^2 l_c l_2 F_n} \times \frac{A_c}{A}, \quad F_1 = \frac{c}{2\pi} \sqrt{\frac{A_c}{V l_c}}$$

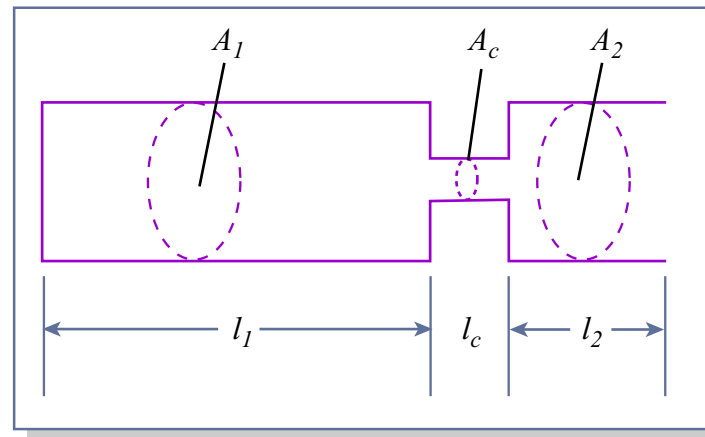


Image by MIT OpenCourseWare. Adapted from Stevens, K. N. "On the Quantal Nature of Speech." *Journal of Phonetics* 17 (1989): 3-46.

Stability with respect to multiple parameters

- There is no quantal relationship between constriction area and formant frequencies
- In fact formants are maximally sensitive to constriction area at the points of formant stability.
- Is a quantal relationship between one articulatory parameter and one acoustic parameter sufficient?
- Stevens seems concerned about this case - argues that:
 - although there is no minimum, the relationship between formants and constriction area is a shallow slope.
 - there may be non-monotonicity in the relationship between muscle activity and constriction area (p.15).

Stability with respect to multiple parameters

- What is the effect on F2 and F3 of varying constriction length? Consider the configuration where F2 and F3 converge.

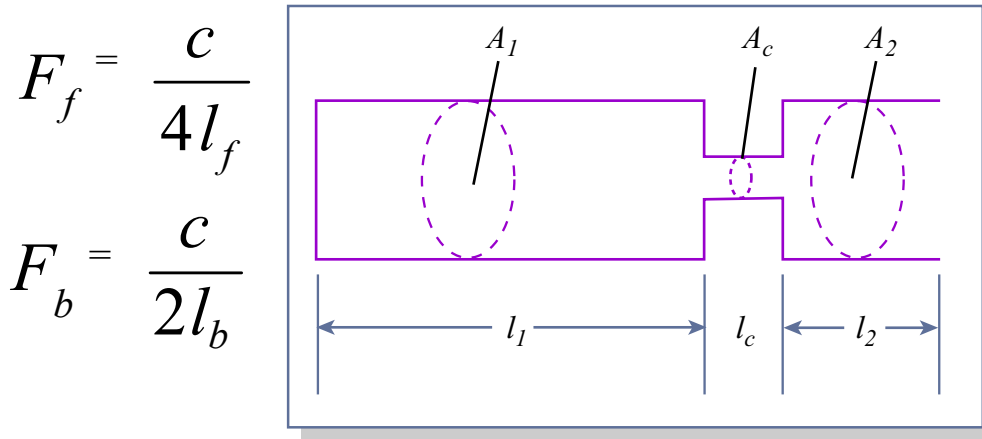


Image by MIT OpenCourseWare. Adapted from Stevens, K. N. "On the Quantal Nature of Speech." *Journal of Phonetics* 17 (1989): 3-46.

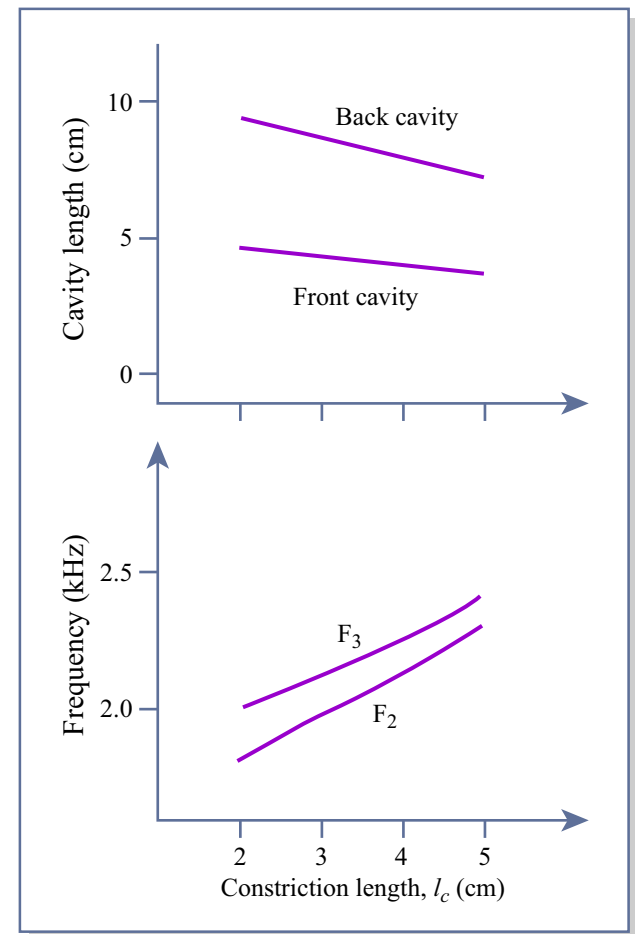


Image by MIT OpenCourseWare. Adapted from Lindblom, B., and O. Engstrand. "In what Sense is Speech Quantal?" *Journal of Phonetics* 17 (1989): 107-121.

Stability with respect to multiple parameters

- Vowel formants vary monotonically with degree of lip constriction.
- How undesirable is articulatory precision?
- Languages do not appear to take full advantage of the imprecision that quantal regions allow - e.g. differences between Danish and English [i].
- Note that Stevens has recently suggested that the quantal distinction between front and back vowels is based on the frequency of F2 relative to the first sub-glottal zero, not on convergence of F2 with F1/F3.

Lindblom's Theory of Adaptive Dispersion

- Liljencrants and Lindblom (1972), Lindblom 1986, 1990a,b
- An alternative explanation for the cross-linguistic preference for the vowels [i, a, u]: these vowels are at the extremes of the formant space of physiologically possible vowels.
- These vowels are maximally distinct from each other and therefore less likely to be confused by a listener.

Lindblom's Theory of Adaptive Dispersion

- A shift in perspective: preferred systems of contrasts vs. preferred sounds.
- There are many generalizations about possible inventories of contrasting sounds.

Lindblom's Theory of Adaptive Dispersion

- Common vowel inventories:

i	u	i	u	i	u
		e	o	e	o
				ɛ	ɔ
	a		a		a

Arabic,
Nyangumata,
Aleut, etc.

Spanish,
Swahili,
Cherokee, etc.

Italian,
Yoruba,
Tunica, etc.

- Unattested vowel inventories:

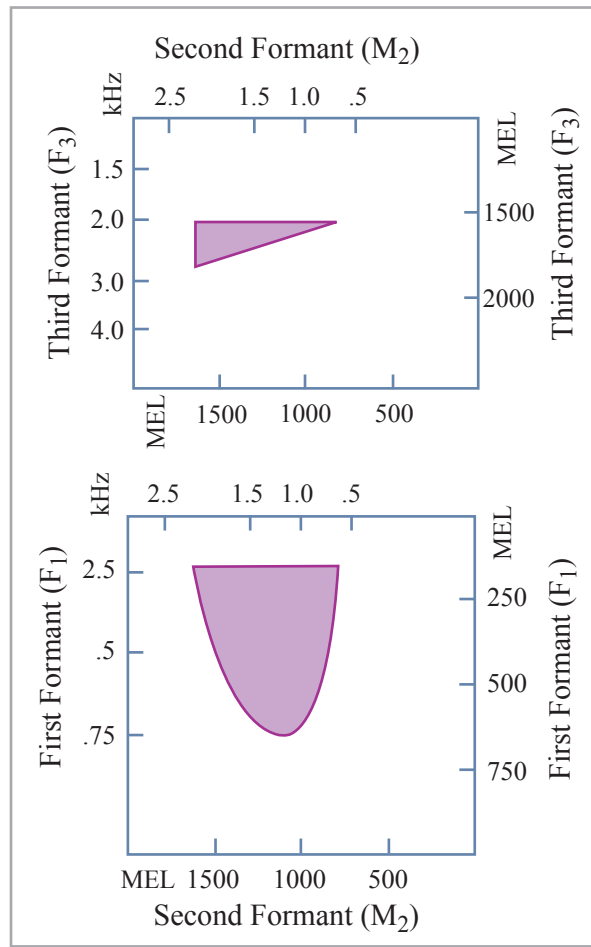
i	i	ɯ	i	u
e	e	ɤ	ɪ	ʊ
	a			

Lindblom's Theory of Adaptive Dispersion

- Lindblom's approach takes these generalizations as prior to generalizations about sounds: a preferred speech sound is one that appears in many preferred inventories.
- Specifically, sounds in a language are selected so as to best satisfy requirements that derive from the communicative function of language:
 - Maximize perceptual distinctiveness
 - Minimize effort

Liljencrants and Lindblom (1972)

- The role of perceptual contrast in predicting vowel inventories.
- The space of articulatorily possible vowels:



Images by MIT OpenCourseWare. Adapted from Liljencrants, Johan, and Bjorn Lindblom. "Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast." *Language* 48, no. 4 (December 1972): 839-862.

Liljencrants and Lindblom (1972)

- Perceptual distinctiveness of contrast between V_i and V_j : distance between vowels in perceptual vowel space

$$r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

where x_n is F2 of V_n in mel

y_n is F1 of V_n in mel

- Maximize distinctiveness: select N vowels so as to minimize E

$$E = \sum_{i=1}^{n-1} \sum_{j=0}^{i-1} \frac{1}{r_{ij}^2}$$

- Predicted optimal inventories
- Reasonable approximations to typical 3 and 5 vowel inventories are derived.
- Preference for [i, a, u] is derived.
- Problem: Too many high, non-peripheral vowels.
- Not enough mid non-peripheral vowels.

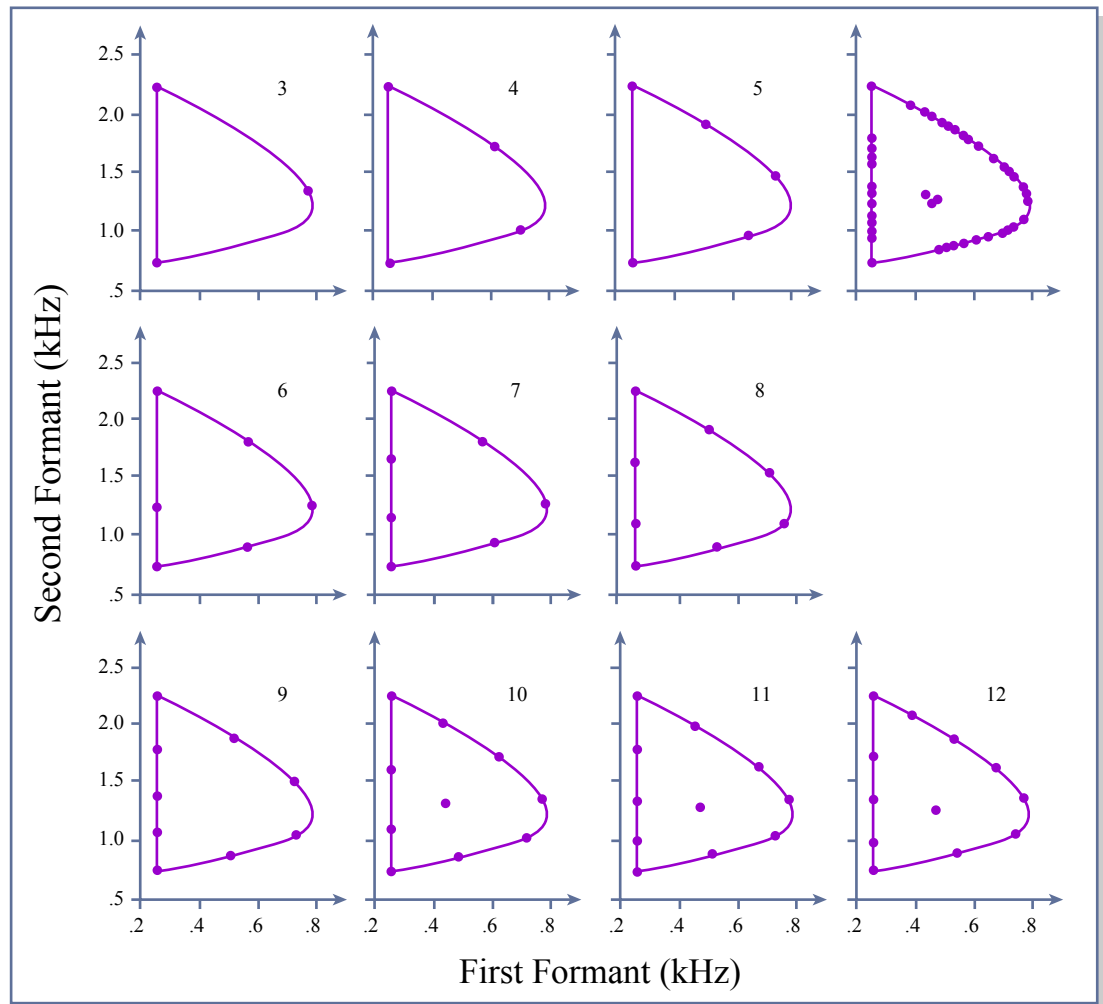


Image by MIT OpenCourseWare. Adapted from Liljencrants, Johan, and Bjorn Lindblom. "Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast." *Language* 48, no. 4 (December 1972): 839-862.

Liljencrants and Lindblom (1972)

- The excess of central vowels arise because measuring distinctiveness in terms of distance in formant space gives too much weight to differences in F2 (even after mel scaling).
- Recent work by Diehl, Lindblom and Creeger (2003) suggests that the greater perceptual significance of F1 probably follows from the higher intensity of F1 relative to F2.

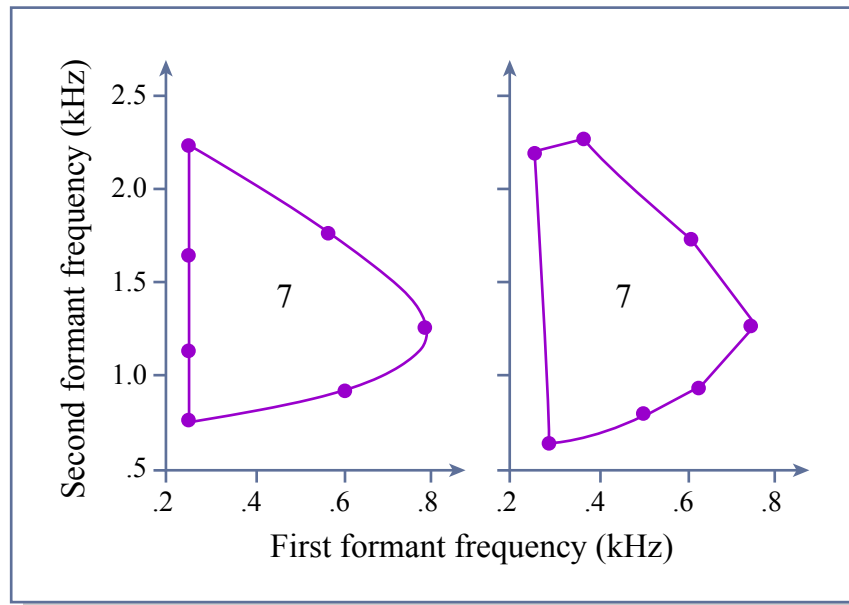


Image by MIT OpenCourseWare. Adapted from Diehl, R. L., B. Lindblom, and C. P. Creeger. "Increasing Realism of Auditory Representations Yields Further Insights into Vowel Phonetics." *Proceedings of the 15th International Congress of Phonetic Sciences*. Vol. 2. Adelaide, Australia: Causal Publications, 2003, pp.1381-1384.

Liljencrants and Lindblom (1972)

- The absence of interior vowels [ə, ø] is a result of the way in which overall distinctiveness is calculated.
- Each vowel contributes to E based on its distance from every other vowel.
- Interior vowels have a high cost because they are relatively close to all the peripheral vowels.
- One possible alternative is to maximize the minimum distance (Flemming 2005).

Problems with Adaptive Dispersion

- Specific instantiations of the model have made specific incorrect predictions (but some of the broad predictions are correct and models are improving).
- The model answers an inobvious question: ‘Given N vowels, what should they be?’ - what determines the size of inventories?
- TAD predicts a single best inventory for each inventory size. Why would languages have sub-optimal inventories?

24.963

Linguistic Phonetics

Source-filter analysis of fricatives

Noise source

- Turbulence noise - random pressure fluctuations.
- Turbulence can result when a jet of air flows out of a constriction into a wider channel (or open space).

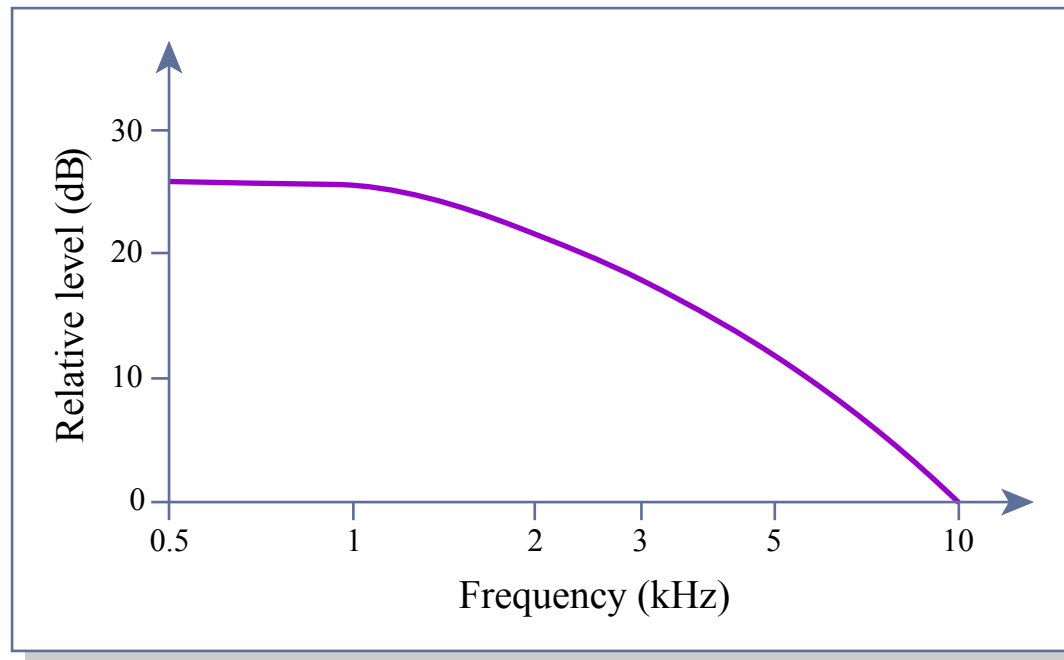


Image by MIT OpenCourseWare. Adapted from Stevens, K. N. "On the Quantal Nature of Speech." *Journal of Phonetics* 17 (1989): 3-46.

Noise source

- Turbulence can result when a jet of air flows out of a constriction into a wider channel (or open space).
- The intensity of turbulence noise depends on particle velocity.
- For a given volume velocity, particle velocity will be greater if the channel is narrower, so for a given volume velocity, narrower constrictions yield louder frication noise.

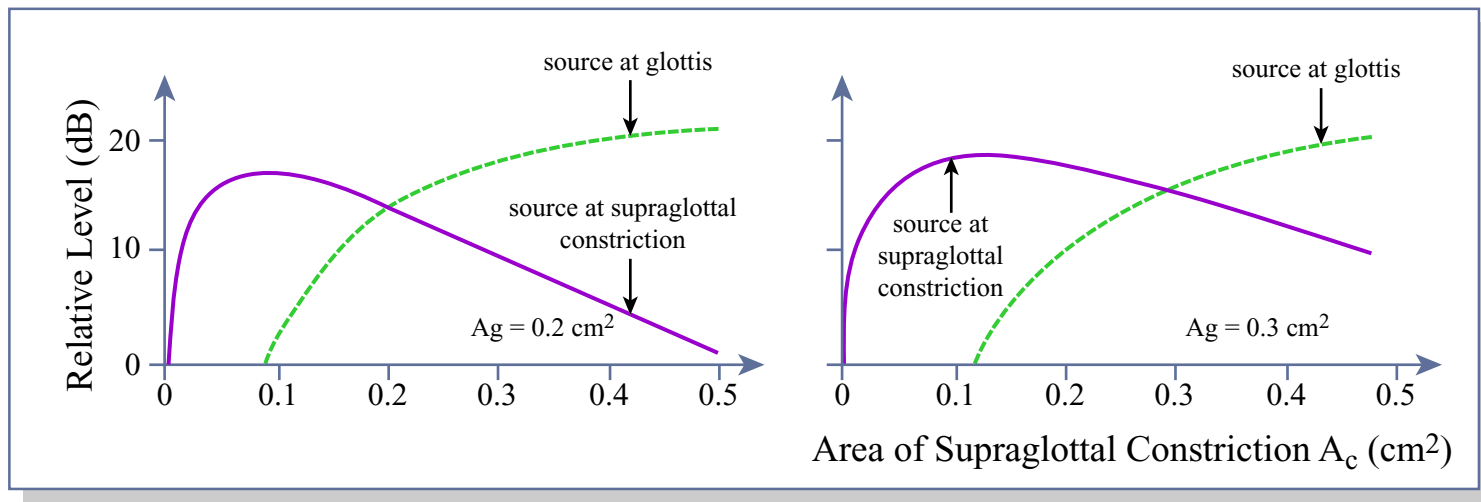
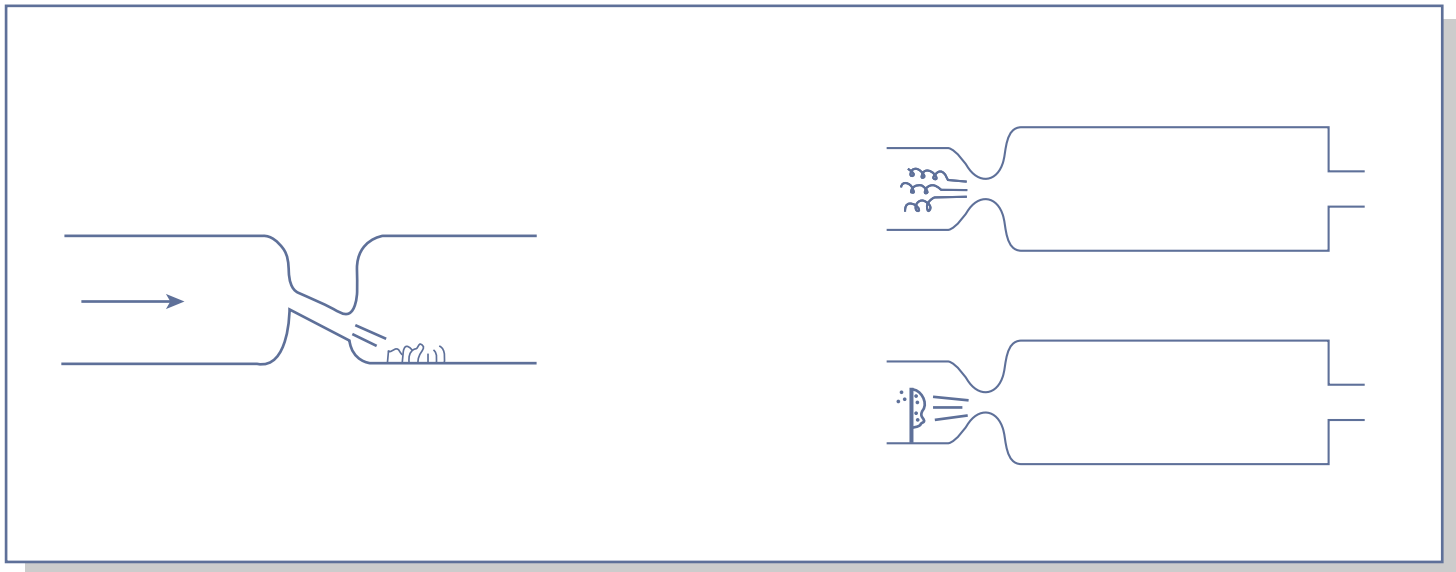


Image by MIT OpenCourseWare. Stevens, K. N. *Acoustics Phonetics*. Cambridge, MA: MIT Press, 1999.

Noise source

- Turbulence is also produced when an airstream strikes an obstacle (e.g. the teeth in [s]).
- The orientation of the obstacle to the direction of flow affects the amount of turbulence produced - the teeth are more or less perpendicular to the airflow in [s] and thus produce significant turbulence.
- The louder noise of strident fricatives is a result of downstream obstacles.



Filter characteristics

- The noise sources are filtered by the cavity in front of the constriction.
- In [h] the noise source is at the glottis, so the entire supralaryngeal vocal tract filters the source, just as in a vowel.
- So [h] has formants at the same frequency as a vowel with the same vocal tract shape, but the formants are excited by a noise source instead of voicing.
- The noise source generated at the glottis has lower intensity at low frequencies, so F1 generally has low intensity in [h].

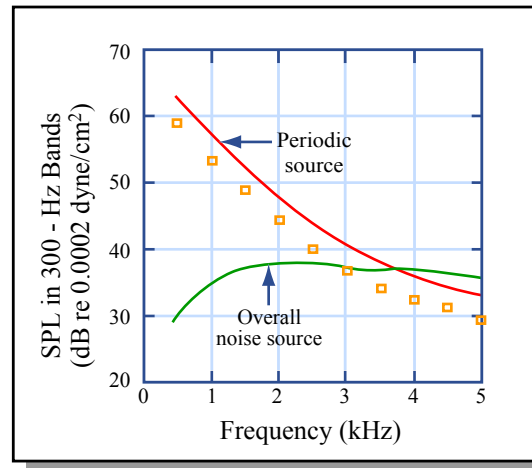


Image by MIT OpenCourseWare.

[h]

- The noise source generated at the glottis has lower intensity at low frequencies, so F1 generally has low intensity in [h].

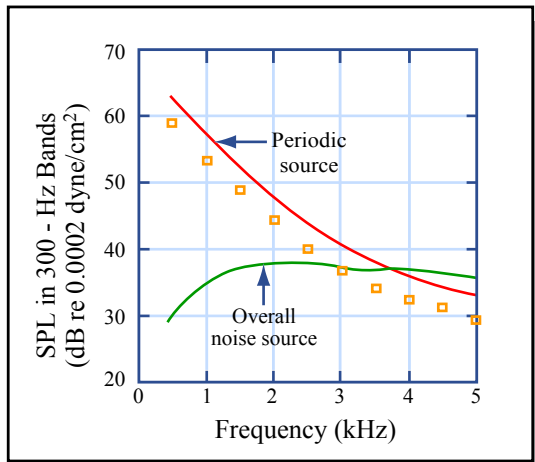


Image by MIT OpenCourseWare.

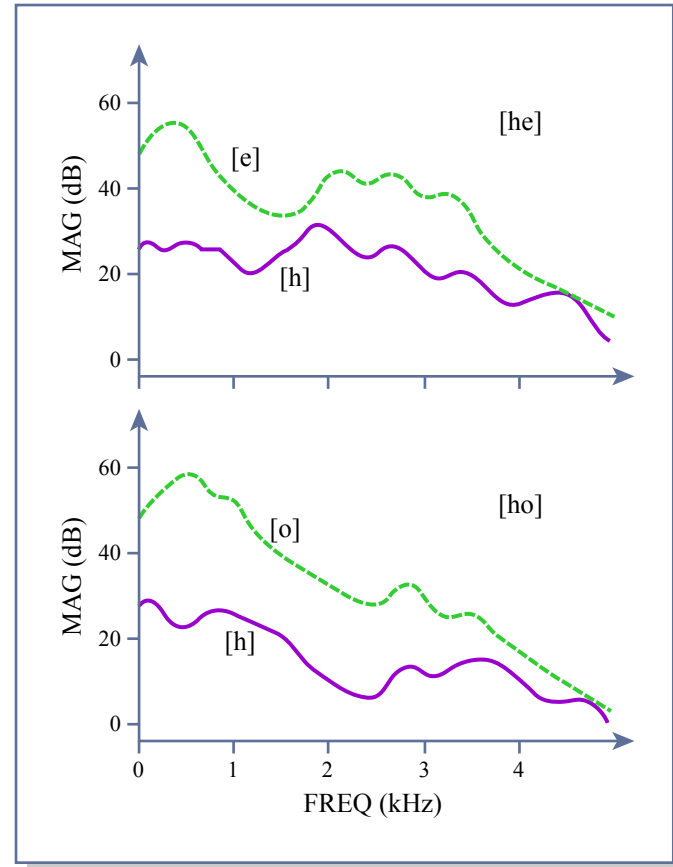
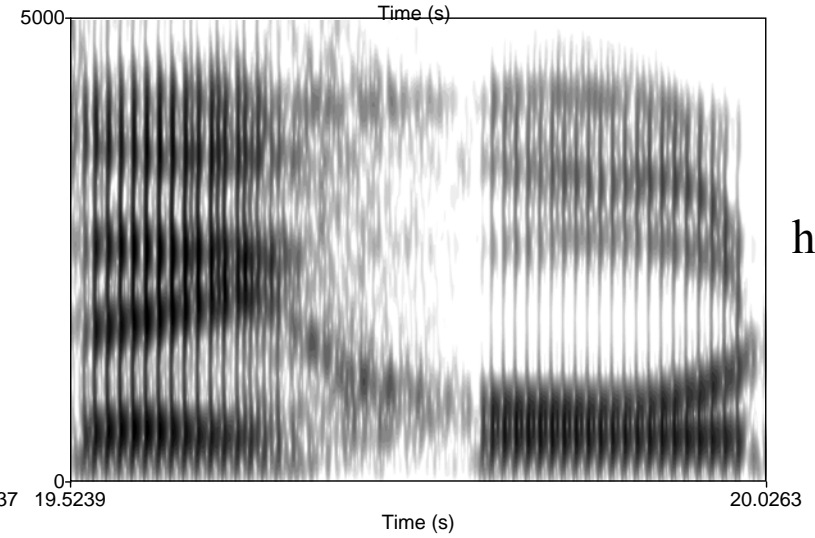
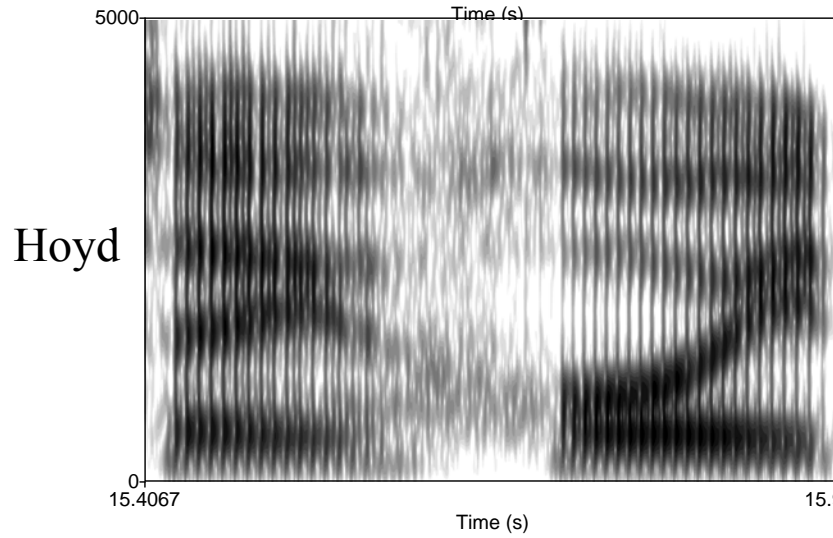
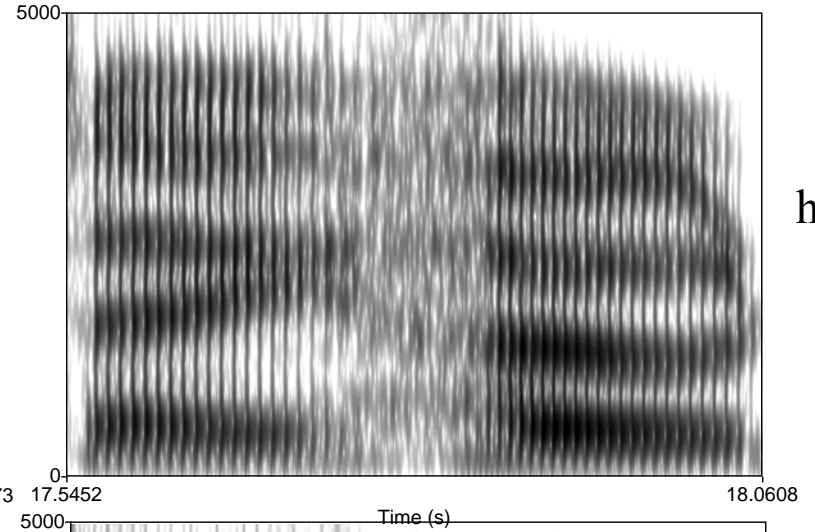
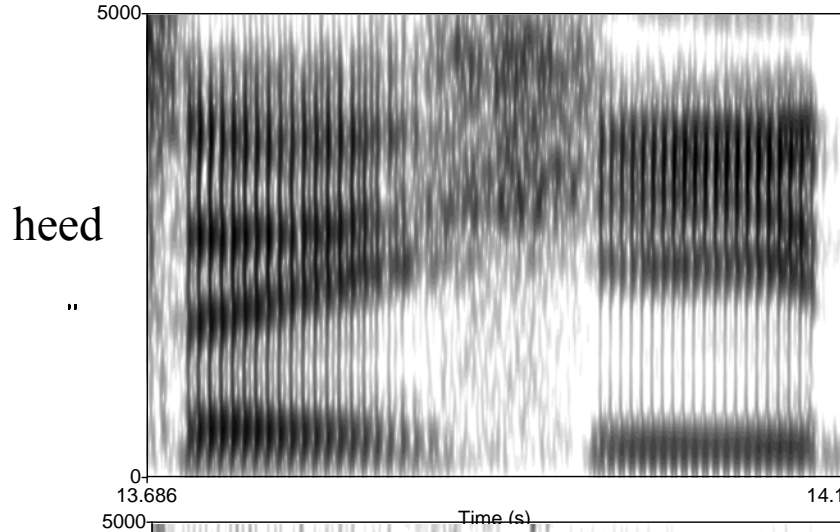


Image by MIT OpenCourseWare. Stevens, K. N. *Acoustic Phonetics*. Cambridge, MA: MIT Press, 1999.

[h]



Filter characteristics

- As the place of articulation shifts forward, the cavity in front of the noise source is progressively smaller.
- A smaller cavity has higher resonances, so other things being equal, the concentration of energy in the fricative spectrum is higher the closer the place of articulation is to the lips.

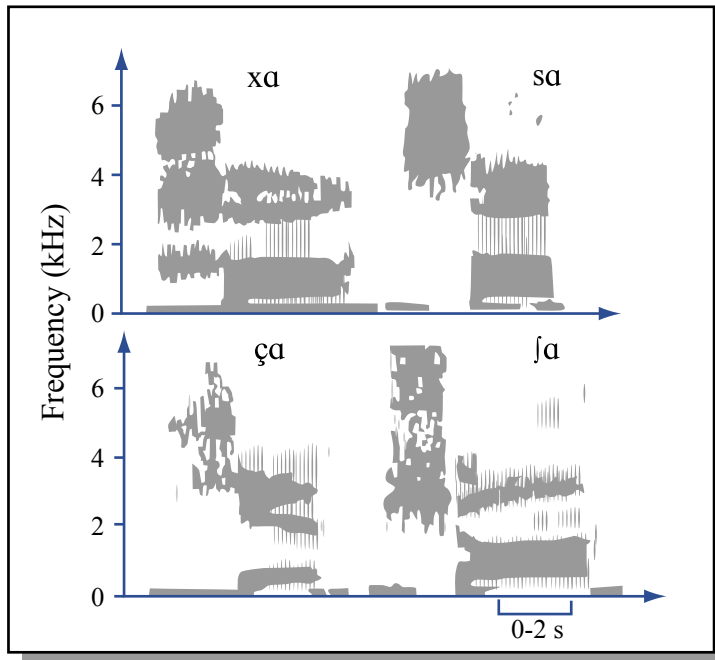


Image by MIT OpenCourseWare.

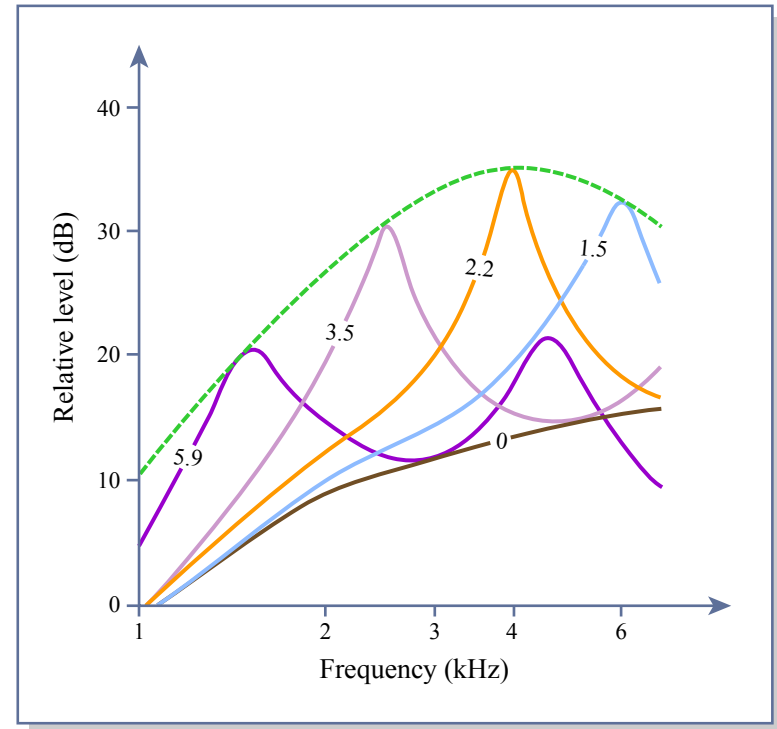
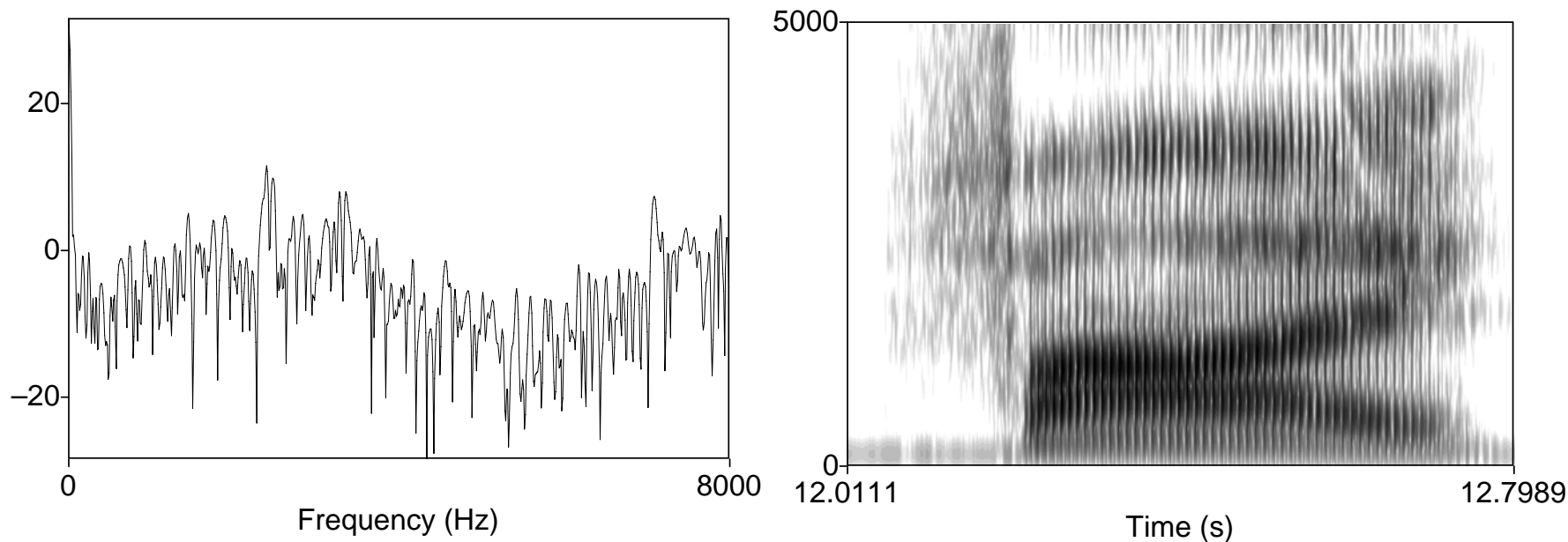


Image by MIT OpenCourseWare. Adapted from Stevens, K. N. "On the Quantal Nature of Speech." *Journal of Phonetics* 17 (1989): 3-46.

Filter characteristics

- The front cavity of a labial is so short (first resonance ~ 10 kHz) that it has little effect on the fricative spectrum, resulting in fricative noise spread over a wide range of frequencies with a broad low-frequency peak.
- This picture can be complicated by acoustic coupling with back cavity.



Filter characteristics

- Lip rounding lowers the resonant frequencies of the front cavity, just as in vowels.
- In coronals, the presence or absence of a sublingual cavity has a significant effect on the size of the front cavity.

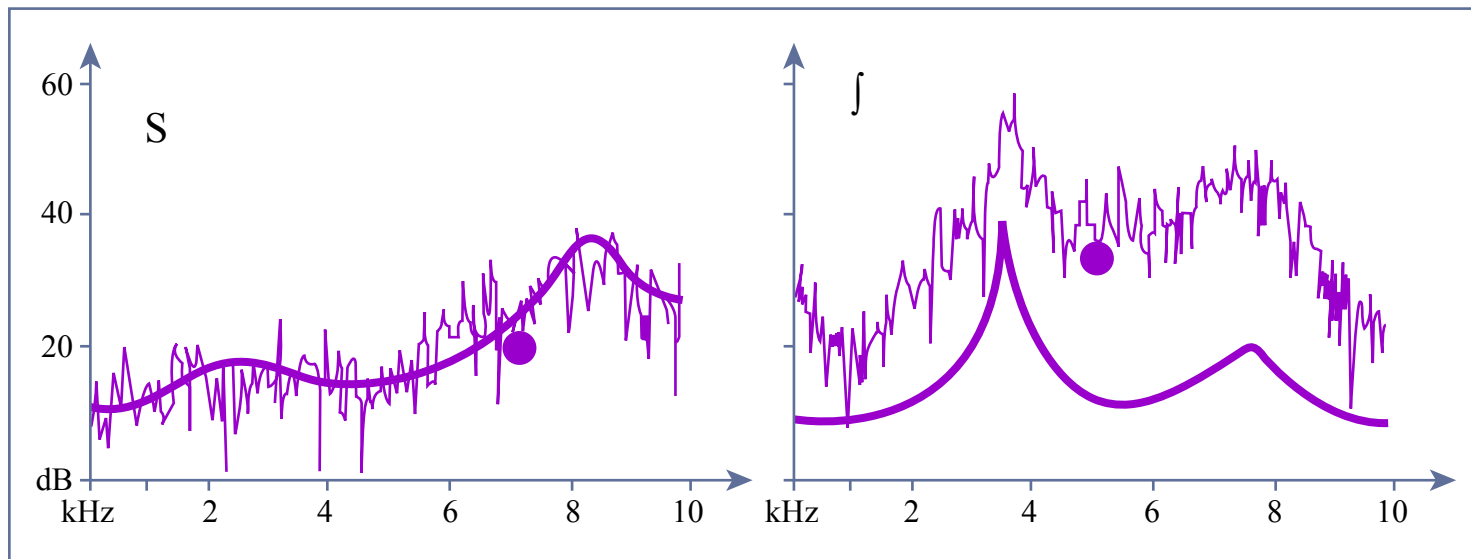


Image by MIT OpenCourseWare.

Source-filter analysis of stops

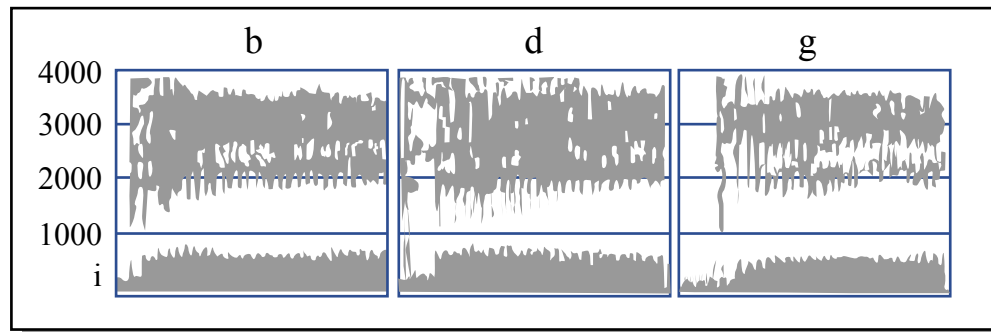


Image by MIT OpenCourseWare.

Stops

- Stops are complicated in that they involve a series of rapid changes in acoustic properties, but each component can be analyzed in similar terms to vowels and fricatives.
- A stop can consist of four phases: implosion (closure) transitions - closure - burst - release transitions

Closure

- Only source of sound is voicing, propagated through the walls of the vocal tract.
- The walls of the vocal tract resonate at low frequencies, so only low-frequency sound is transmitted ('voice bar').

Burst

- Consists of a transient, due to abrupt increase in pressure at release, followed by a short period of frication as air flows at high velocity through the narrow (but widening) constriction.
- Transient source is an impulse (flat spectrum) filtered by the front cavity.
- The frication is essentially the same as a fricative made at the same place of articulation.
- Alveolars have high frequency, high intensity bursts.
- Velar bursts are concentrated at the frequency of F2 and/or F3 at release.
- Labial bursts are of low intensity, with energy over a wide range of frequencies, with a broad, low-frequency peak.

Release transitions

- As the constriction becomes more open, frication ceases.
- The source at this time is at the glottis - either voicing or aspiration noise. This source excites the entire vocal tract as in a vowel (or [h]).
- The shape of the vocal tract, and thus the formants, during this phase are basically determined by the location of the stop constriction and the quality of adjacent vowels.
- The formants move rapidly as the articulators move from the position of the stop to the position for the vowel. The formant movements are usually called ‘formant transitions’.

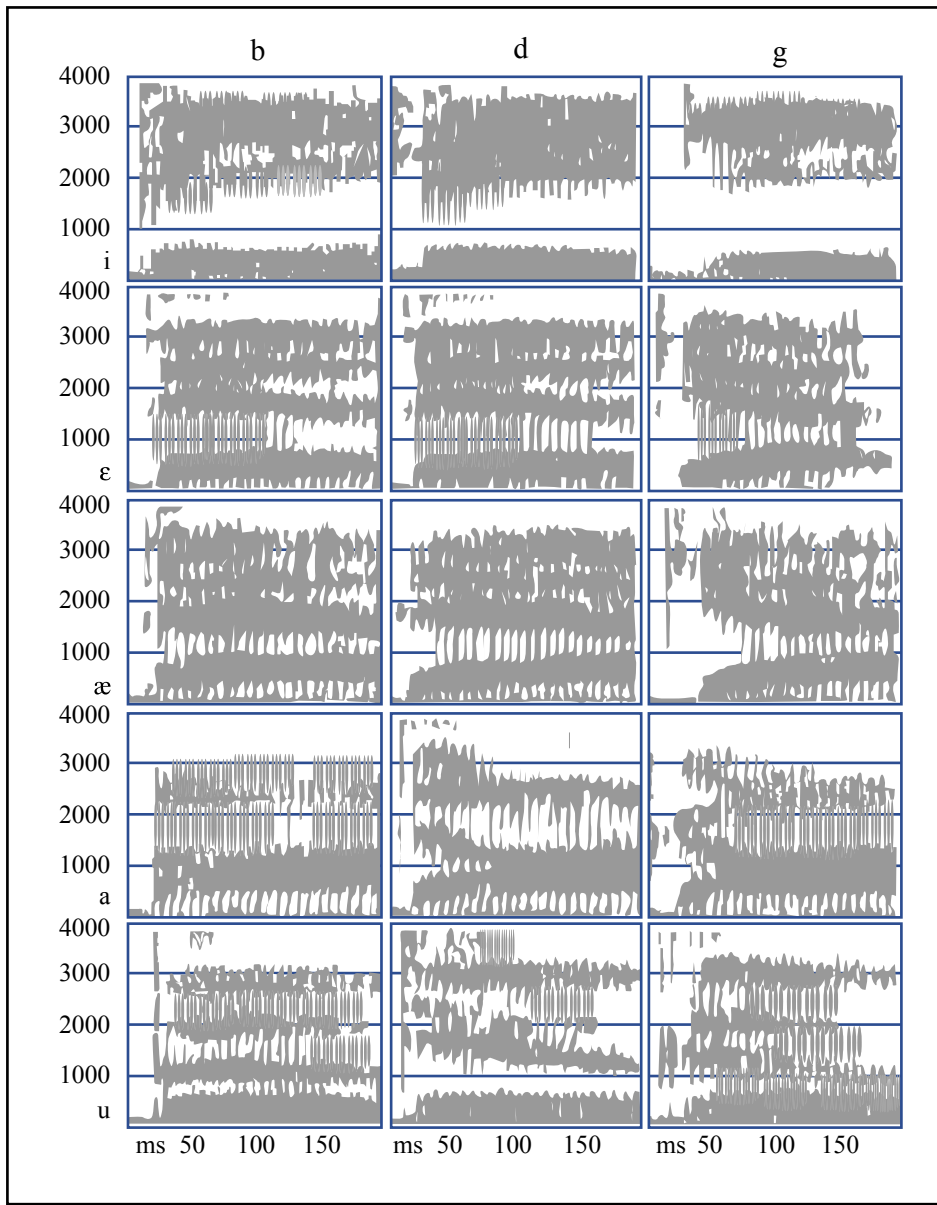


Image by MIT OpenCourseWare. Adapted from Ladefoged, Peter. *Phonetic Data Analysis*. Malden, MA: Blackwell, 2003.

Release transitions

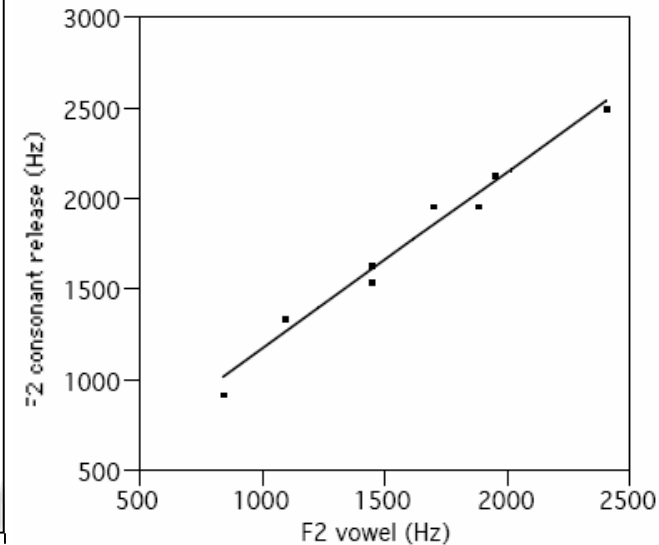
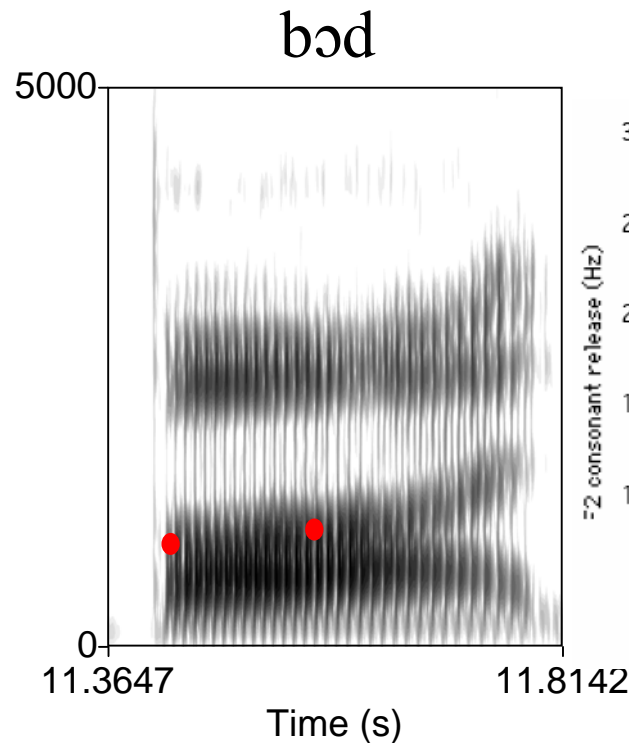
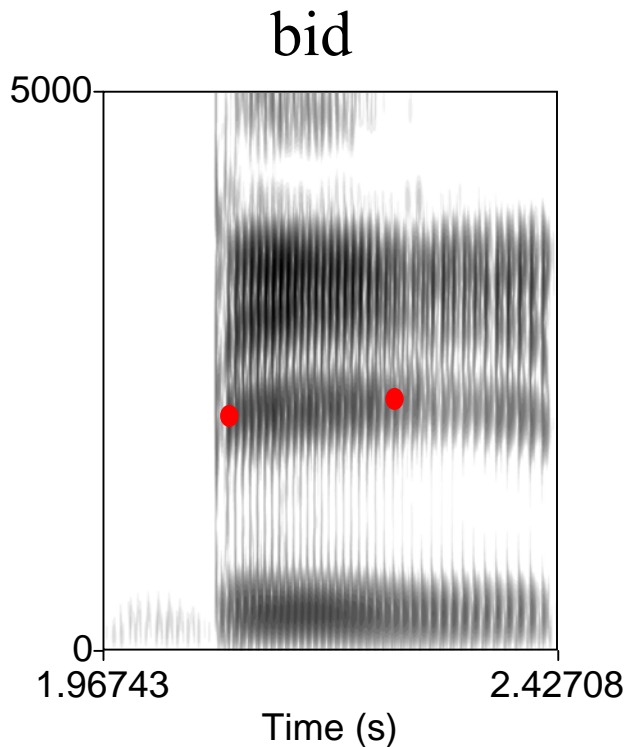
- In alveolar stops the formant transitions due to the tongue tip constriction are probably very rapid (Manuel and Stevens 1995), so the observed formant transitions appear to be due to tongue body movements.
- The tongue body is generally relatively front to facilitate placement of the tongue tip//blade, thus there is a relatively high F2 at release (~ 1800 - 2000 Hz) and high F3.
- Labial stops involve a constriction at the lips. The tongue position is determined by adjacent vowels, so the exact formant frequencies at release depend on these vowel qualities.
- The labial constriction always lowers formants, so F2 and F3 are generally lower at release of a labial than in the following vowel.

Release transitions

- Velar stops involve a dorsal constriction, but the exact location of this constriction depends on the neighbouring vowels.
- So the formant transitions of velars vary substantially, approximately tracking F2 of the adjacent vowel.
- F2 and F3 are often said to converge at velar closure. Under what conditions should this occur?
- Similar transitions are observed during the formation of a stop closure.
- Similar transitions are observed into and out of any consonant with a narrow constriction, e.g. fricatives, nasal stops.

Locus equations

- Typically F2 at the release of a consonant is a linear function of F2 at the midpoint of the adjacent vowel (Lindblom 1963, Klatt 1987, etc).
- The slope and intercept of this function depend on the consonant.



Locus equations

- The slope and intercept of this function depend on the consonant.

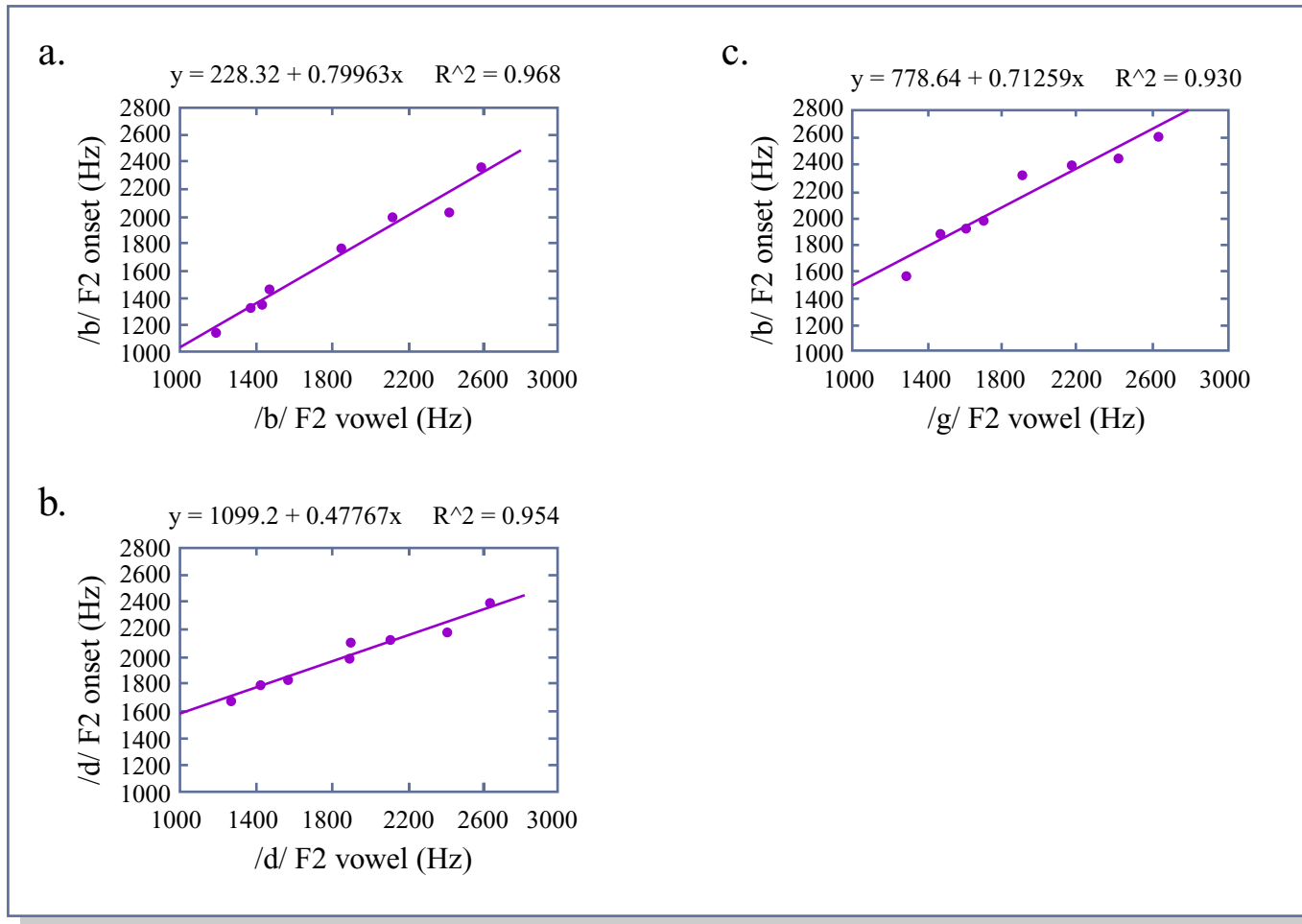


Image by MIT OpenCourseWare. Adapted from Fowler, C. A. "Invariants, Specifiers, Cues: An Investigation of Locus Equations as Information for Place of Articulation." *Perception and Psychophysics* 55, no. 6 (1994): 597-610.

Affricates

- The frication portion of the release of the stop is prolonged to form a full-fledged fricative.
- The fricative portion of an affricate is distinguished from a regular fricative by its shorter duration, and perhaps by the rapid increase in intensity at its onset (short rise time).