24.963 Linguistic Phonetics
Fall 2005

24.963
Linguistic Phonetics

Speech Perception

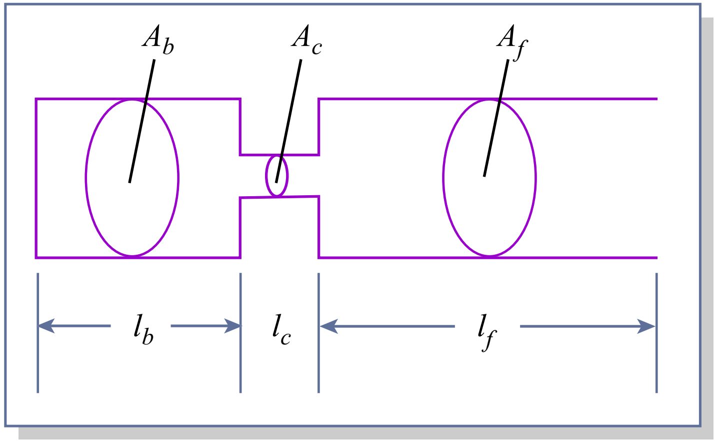
Reading for week 9:

- Keating (1990)
- Browman and Goldstein (1990)
- Work on final paper.

Acoustics assignment - question 2

velars stops are often described as being characterized by a convergence of F2 and F3 in the formant transitions at closure and release.

- F2 and F3 will only converge on exactly the same frequency under very specific conditions – what are these conditions?
- The back cavity must be twice the length of the front cavity.



$$F_b = \frac{c}{2L}$$

$$F_f = \frac{c}{4L}$$

Image by MIT OpenCourseWare. Adapted from Johnson, Keith. *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell Publishers, 1997.

Acoustics assignment - question 2

- Under what broader set of circumstances would we expect F2 to fall and F3 to rise in the release transitions of a velar? (Assume that the following vowel is produced with a vocal tract with uniform cross-sectional area).

Answer 1: Perturbation theory

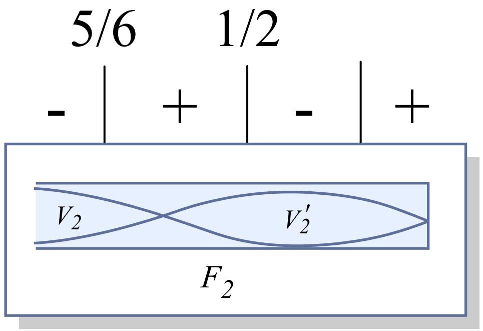


Image by MIT OpenCourseWare. Adapted from Johnson, Keith. *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell Publishers, 1997.

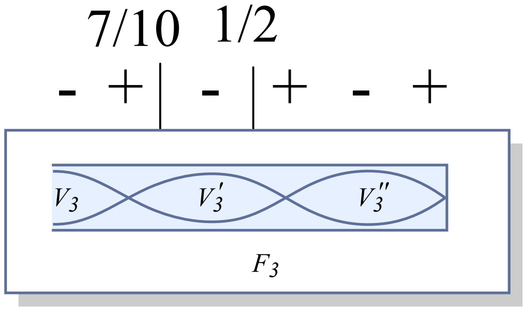
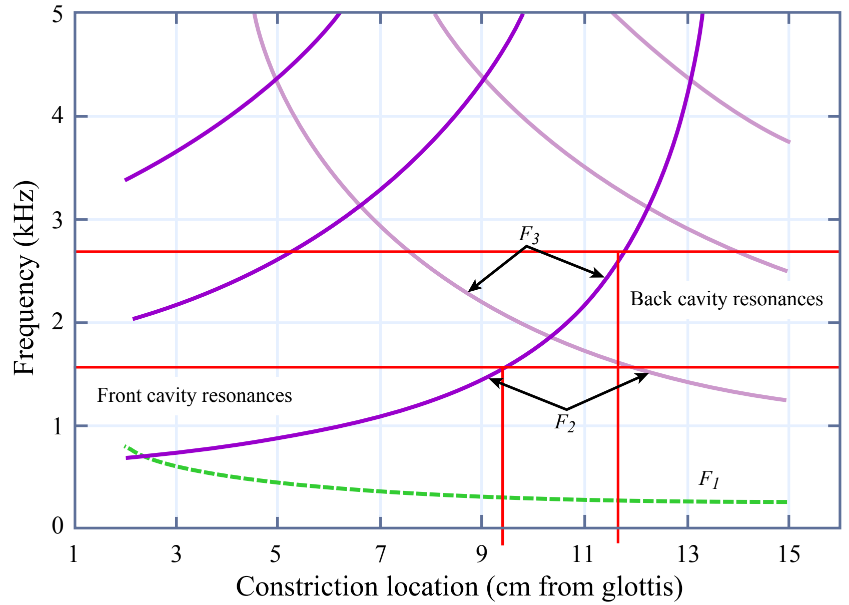


Image by MIT OpenCourseWare. Adapted from Johnson, Keith. *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell Publishers, 1997.

- A constriction between 0.5 and 0.7 of the length of the tube from the glottis should raise F2 and lower F3.

Acoustics assignment - question 2

Answer 2 - tube model



Resonant frequencies of the back tube (light lines), front tube (heavy lines) and Helmholtz resonance (dashed line) in the tube model. Frequencies are plotted as a function of different back tube lengths (l_b), with the length of the constriction fixed at 2 cm and the total length of the model fixed at 16 cm.

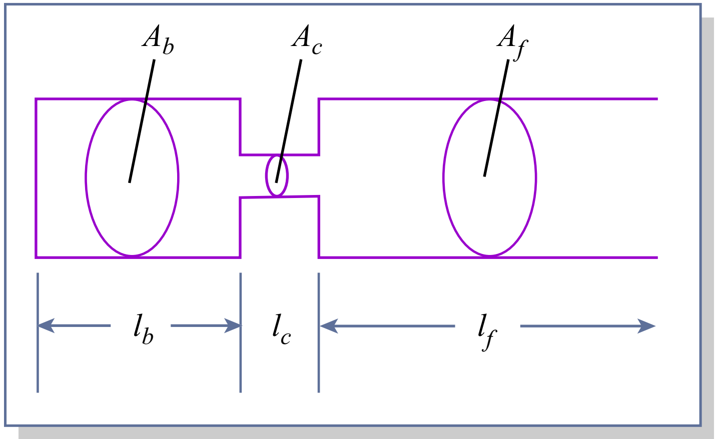


Image by MIT OpenCourseWare.

Images by MIT OpenCourseWare. Adapted from Johnson, Keith. *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell Publishers, 1997.

Modeling locus equations

Bilabial stop

A good fit to the observed locus equation can be obtained with the following model:

- The vocal tract at the onset of F2 is the same as for the following vowel except for the presence of a narrow lip constriction.
- Lip constriction modeled by adding a fixed length *d* to the front cavity.
- F2 is always a resonance of the front cavity.

a.

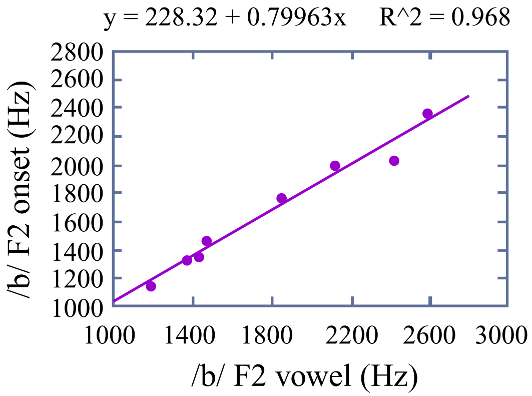


Image by MIT OpenCourseWare. Adapted from Fowler, C. A. "Invariants, Specifiers, Cues: An Investigation of Locus Equations as Information for Place of Articulation." *Perception and Psychophysics* 55 (1994): 597-610.

Modeling locus equations

Bilabial stop

- Calculate effective front cavity length of vowel from vowel F2.

  $$F_2 = \frac{c}{4L} \quad \text{so} \quad L = \frac{35000}{4F_2}$$

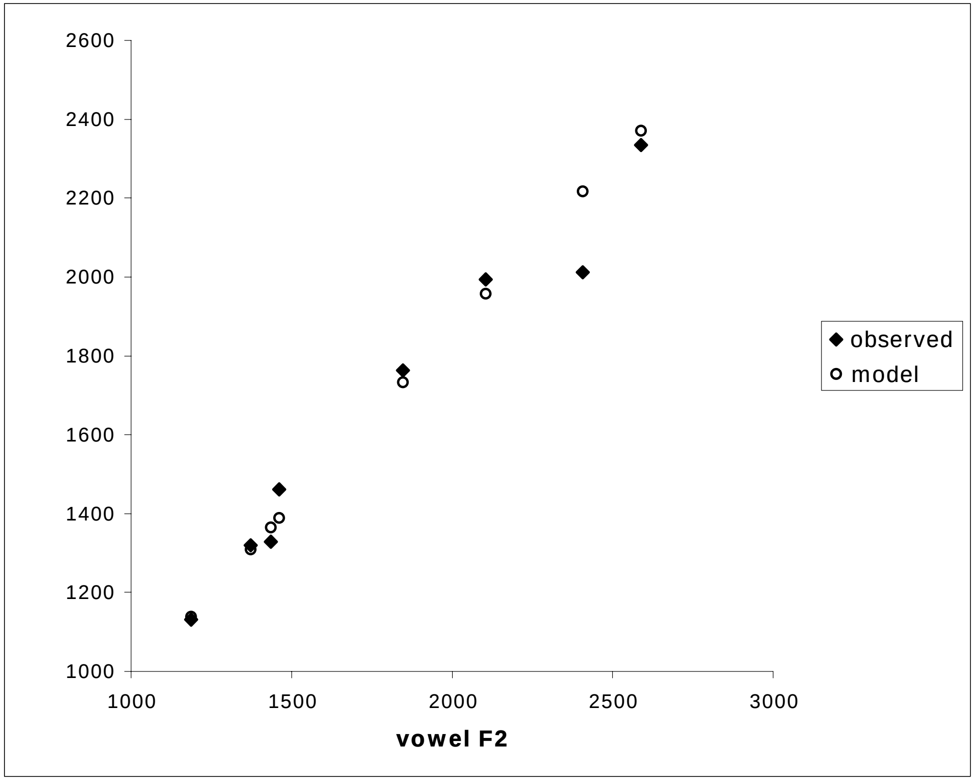
- Calculate predicted F2 at consonant release: $F_2 = \dfrac{c}{4(L+d)}$
- Estimate *d* to obtain best fit to Fowler's data.

Modeling locus equations

Bilabial stop

- Best overall fit: $d \approx 0.4$.
- Worst error is for [eɪ], perhaps because this is a diphthong.
- If we ignore [eɪ], best fit is obtained with $d = 0.31$

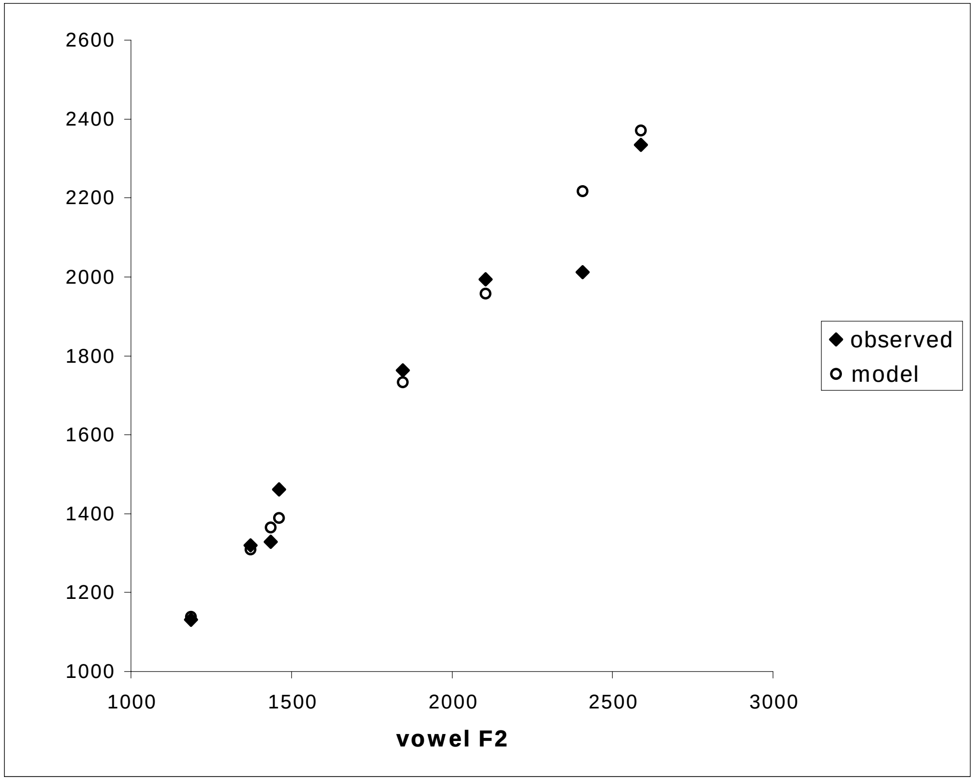
- Slope is less than 1 because adding a fixed length to the front cavity has less effect if the front cavity is longer.
- Locus equation is actually curved, but only slightly.
- Model fits round vowels well although we might expect less effect of a labial constriction.



Modeling locus equations

Bilabial stop

- The assumption that F2 is always a front cavity resonance is inconsistent with the common assumption that F2 is a back cavity resonance for [i] (at least).
- If this were the case, the model would predict that F2 in [i] should be unaffected by lip constriction, but only given that we ignore coupling.
- Lengthening front cavity may shift F2 to the front cavity.



Velar stop

A first model: Velars assimilate fully to the place of a following vowel.

- Would predict slope close to 1, in fact the slope in 0.7.
- Velars could not assimilate to low vowels since their constriction is in the pharynx.

Second model: Velar closure is formed at the point on the roof of the mouth that minimizes tongue body movement to the vowel.

- predicts that F2 at release should be close to vowel F2 for high vowels.
- This is correct for [i], but not for [u] - F2 at release is substantially higher, suggesting that the velar constriction is further forward.

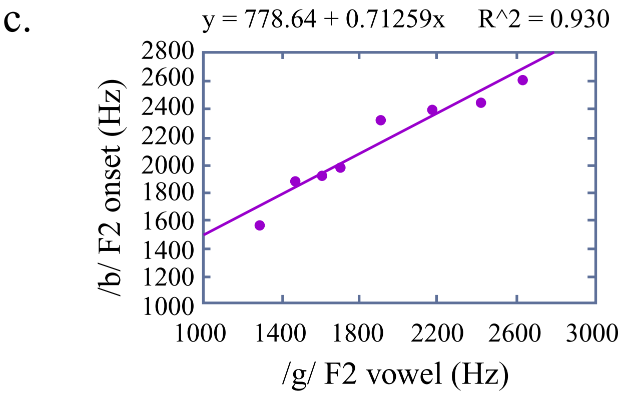


Image by MIT OpenCourseWare. Adapted from Fowler, C. A. "Invariants, Specifiers, Cues: An Investigation of Locus Equations as Information for Place of Articulation." *Perception and Psychophysics* 55 (1994): 597-610.

Velar stop

A final model: target and undershoot.

- F2 at consonant release is approximately equal to vowel F2 for high F2 values, while consonant F2 is progressively higher than vowel F2 for lower values of vowel F2.
- Suggests that velar closure is closest to the vowel constriction in front vowels and is in front of the vowel constriction for all other vowels.
- This could be interpreted as showing that the target for a velar closure is quite far forward (a little behind [i]).
- The actual constriction location is shifted towards the following vowel constriction, but only partially, hence slope < 1.

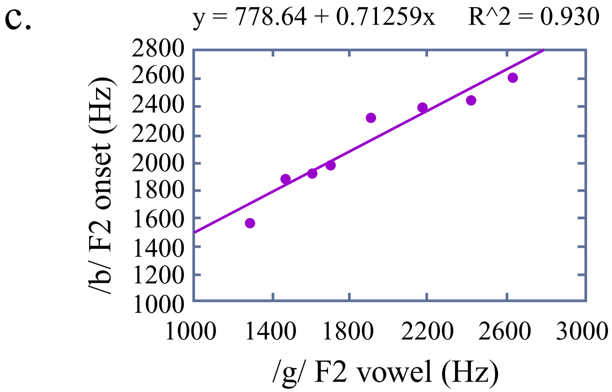


Image by MIT OpenCourseWare. Adapted from Fowler, C. A. "Invariants, Specifiers, Cues: An Investigation of Locus Equations as Information for Place of Articulation." *Perception and Psychophysics* 55 (1994): 597-610.

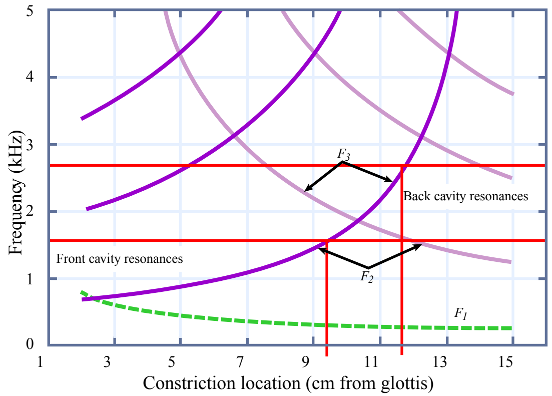
Velar stop

Shifting cavity affiliations of F2 and F3:

- F3 is supposed to be a front cavity resonance in [i], but a back cavity resonance in most other vowels.
- As a velar constriction shifts forward, frequencies of the front cavity resonances increase.
- For most constriction locations, this will be F2, but as F2 comes close to F3, F2 should level out, then fall if the constriction moves far enough forward.
- Can this explain the flattening of the velar locus slope?

Velar stop

- Can this explain the flattening of the velar locus slope? Not clear
- The line starts to level out after [æ] (F2 = 1912 Hz), but the nomograms suggest that F2 should not level out at such a low frequency.
- Alternative: /æ/ is a diphthong [eæ] [ɪæ], so F2 at stop release looks high given vowel F2 measured in the middle of the diphthong.
- My velar locus equation looks linear because my /æ/ is a monophthong.



Resonant frequencies of the back tube (light lines), front tube (heavy lines) and Helmholtz resonance (dashed line) in the tube model. Frequencies are plotted as a function of different back tube lengths (l_b, with the length of the constriction fixed at 2 cm and the total length of the model fixed at 16 cm.

Image by MIT OpenCourseWare. Adapted from Johnson, Keith. *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell Publishers, 1997.

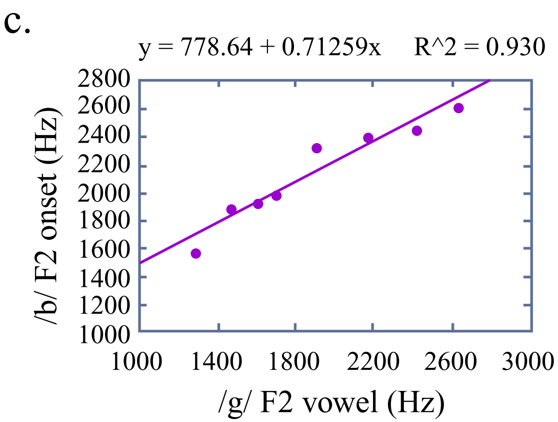


Image by MIT OpenCourseWare. Adapted from Fowler, C. A. "Invariants, Specifiers, Cues: An Investigation of Locus Equations as Information for Place of Articulation." *Perception and Psychophysics* 55 (1994): 597-610.

Alveolar stop

Ttarget and undershoot model.

- The preferred tongue body position in an alveolar is fairly far forward, facilitating contact between tongue tip and alveolar ridge.
- There is a tendency for the stop to assimilate to the tongue body position of the following vowel, but the assimilation is quite limited.
- So F2 at consonant release varies fairly narrowly around 2100 Hz.
- According to this analysis the shallower slope for alveolars (0.48) reflects greater resistance to tongue body assimilation compared to labials (0.8) and velars (0.71).

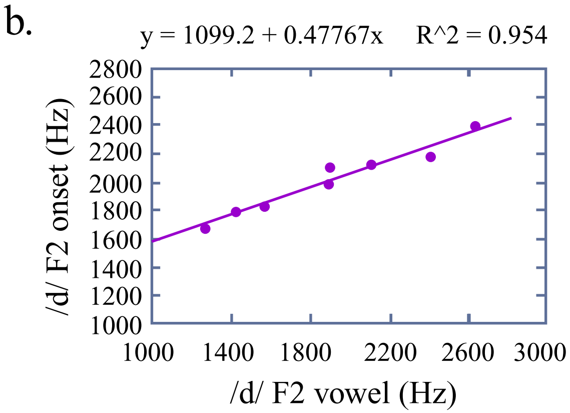


Image by MIT OpenCourseWare. Adapted from Fowler, C. A. "Invariants, Specifiers, Cues: An Investigation of Locus Equations as Information for Place of Articulation." *Perception and Psychophysics* 55 (1994): 597-610.

Speech perception

- The problem faced by the listener: To extract meaning from the acoustic signal.
- This involves the recognition of words, which in turn involves discriminating the segmental contrasts of a language.
- Much phonetic research in speech perception has been directed toward identifying the perceptual cues that listeners use.
- Production studies can reveal many differences between minimal contrasting words, e.g. contrasting vowels of English differ in formant frequencies and duration.
  - Are listeners sensitive to these differences in speech perception?
  - What is the nature of the perceptual representations of speech?

Speech perception

- The problem faced by the listener: To extract meaning from the acoustic signal.
- This involves the recognition of words, which in turn involves discriminating the segmental contrasts of a language.
- Much phonetic research in speech perception has been directed toward identifying the perceptual cues that listeners use.

Speech perception

- Production studies can reveal many differences between minimal contrasting words, e.g. contrasting vowels of English differ in formant frequencies and duration.
  - Are listeners sensitive to these differences in speech perception?
  - What is the nature of the perceptual representations of speech?
- These questions are adressed through perceptual experiments (cf. Johnson p.70).
- Most direct test of perceptual significance of an acoustic property: manipulate the acoustic property synthetically and see if perceptual response is affected. E.g. vary formant frequencies in synthetic vowels, and have subjects categorize the vowels.
- This type of experiment does not directly address the issue of perceptual representation.
- One technique for probing perceptual representation is Multi-Dimensional Scaling (technique for constructing perceptual spaces from confusion or similarity data). However, it is often difficult to relate MDS spaces back to ‘acoustic space’.

Cues to vowel quality

- The main cues to vowel quality are related to the frequencies of the first two or three formants.

Evidence:

- Synthesis experiments: Varying formant frequencies in synthetic vowels can change the vowel percept (Delattre, Liberman, Cooper, and Gerstman 1952, and many others).
- MDS studies: Shepard (1972) presents a three-dimensional MDS analysis of confusion data on American English vowels from Peterson and Barney (1952). Two nearly orthogonal axes correlate well (but non-linearly) with F1 and F2. It is possible to identify an axis which correlates well with F3, but it is also correlated with the F2 axis (not orthogonal).
  - Many other MDS studies of vowel confusion/similarity have found that the first two dimensions identified correspond well to F1, F2.
- It seems that a third dimension is required to accommodate rhotic vowels, but MDS analyses do not provide clear evidence that this dimension corresponds to F3.

Cues to vowel quality

- It is not clear whether listeners extract formant peaks per se or use parameters related to the shape of the whole spectrum. Either way, the perceptual dimensions seem to correlate well with formant frequencies (Plomp 1975, Shepard 1972).

Cues to consonant contrasts

- Place cues (Wright, Frisch and Pisoni 1999)

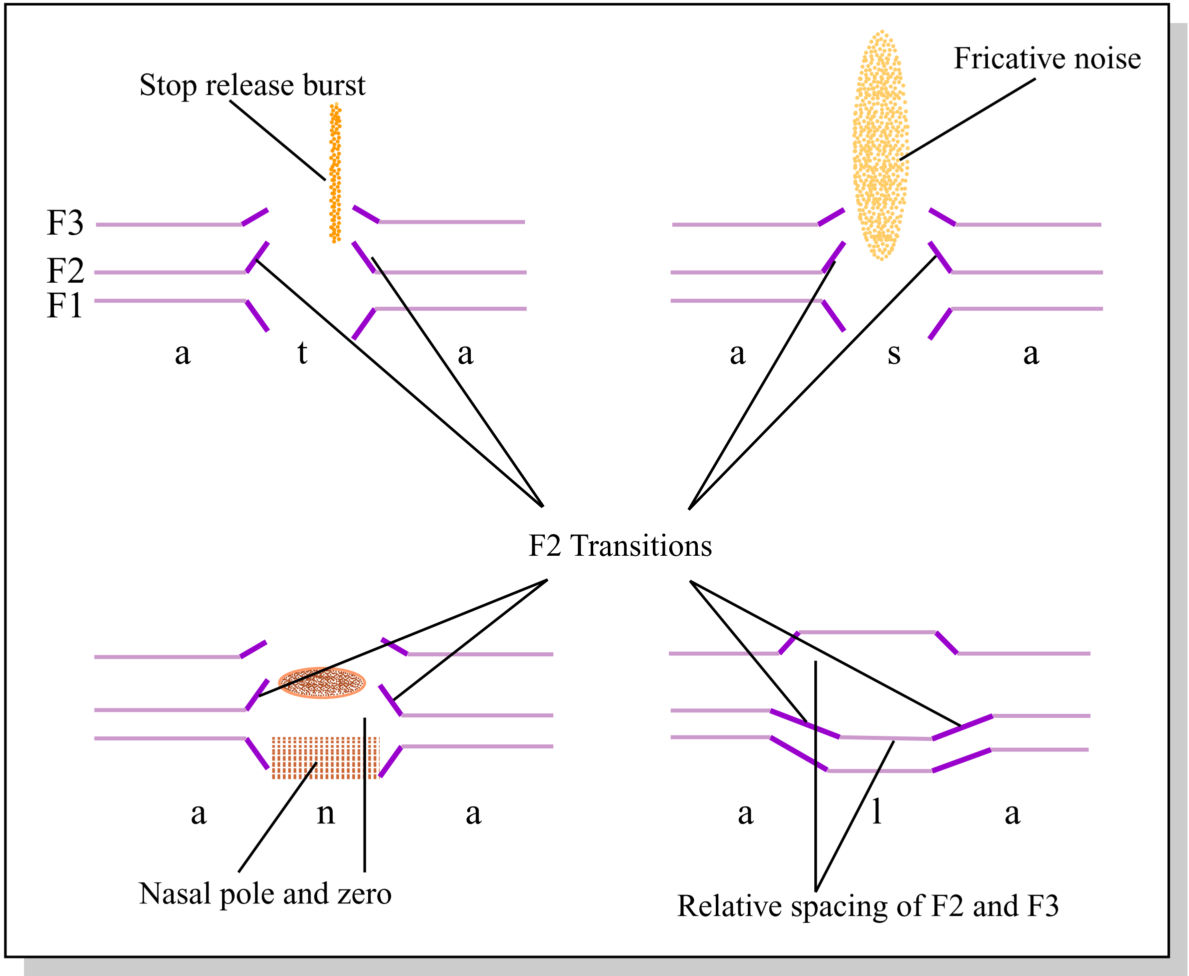


Image by MIT OpenCourseWare. Adapted from Wright, R., S. Frisch, and D. B. Pisoni. "Speech Perception." In *Wiley Encyclopedia of Electrical and Electronics Engineering*. Vol. 20. New York, NY: John Wiley and Sons, 1999, pp. 175-195.

Cues to consonant contrasts

- Manner cues (Wright, Frisch and Pisoni 1999)

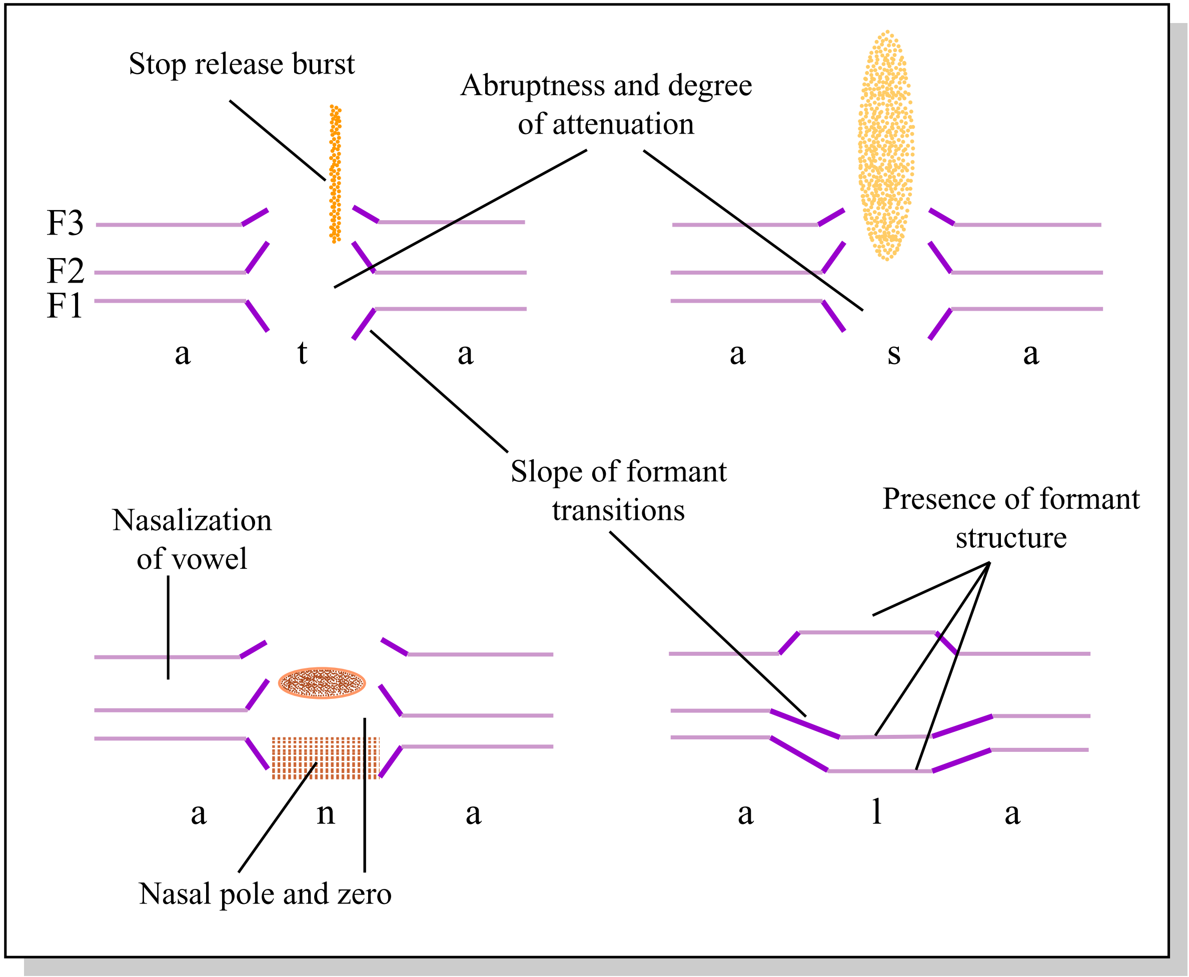

Stop release burst
Abruptness and degree of attenuation
F3
F2
F1
a
t
a
a
s
a
Slope of formant transitions
Presence of formant structure
Nasalization of vowel
a
n
a
a
l
a
Nasal pole and zero

Image by MIT OpenCourseWare. Adapted from Wright, R., S. Frisch, and D. B. Pisoni. "Speech Perception." In *Wiley Encyclopedia of Electrical and Electronics Engineering.* Vol. 20. New York, NY: John Wiley and Sons, 1999, pp. 175-195.

Cues to consonant contrasts

- Obstruent voicing cues (Wright, Frisch and Pisoni 1999)

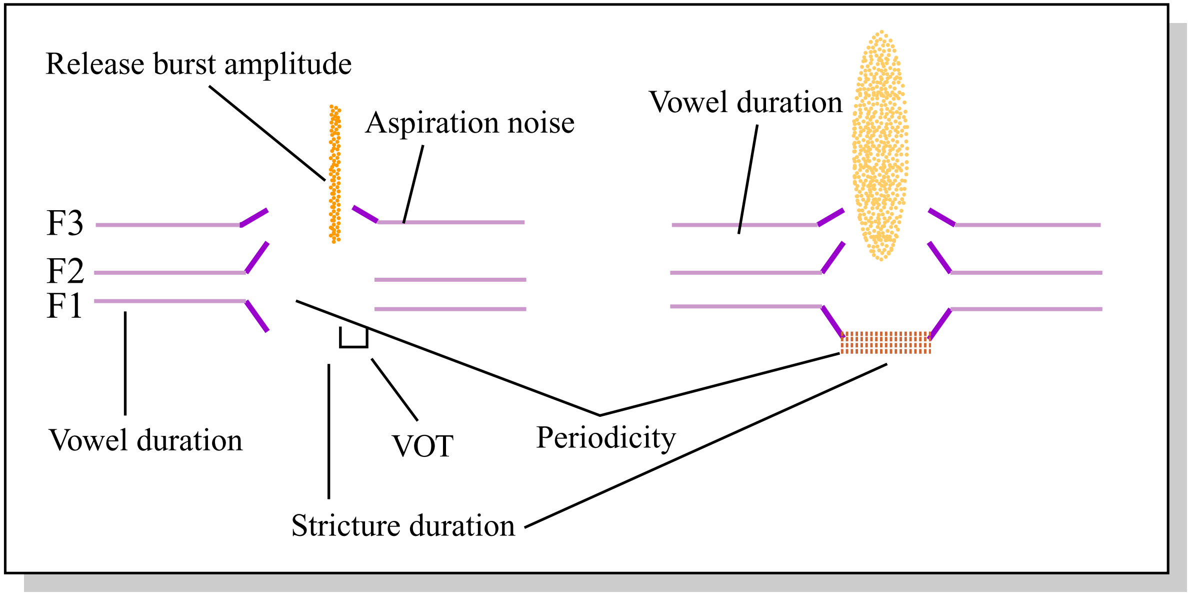


Image by MIT OpenCourseWare. Adapted from Wright, R., S. Frisch, and D. B. Pisoni. "Speech Perception." In *Wiley Encyclopedia of Electrical and Electronics Engineering*. Vol. 20. New York, NY: John Wiley and Sons, 1999, pp. 175-195.

The nature of acoustic cues

- There are multiple cues to every contrast – the speech signal is highly redundant.
- E.g. stop voicing in English

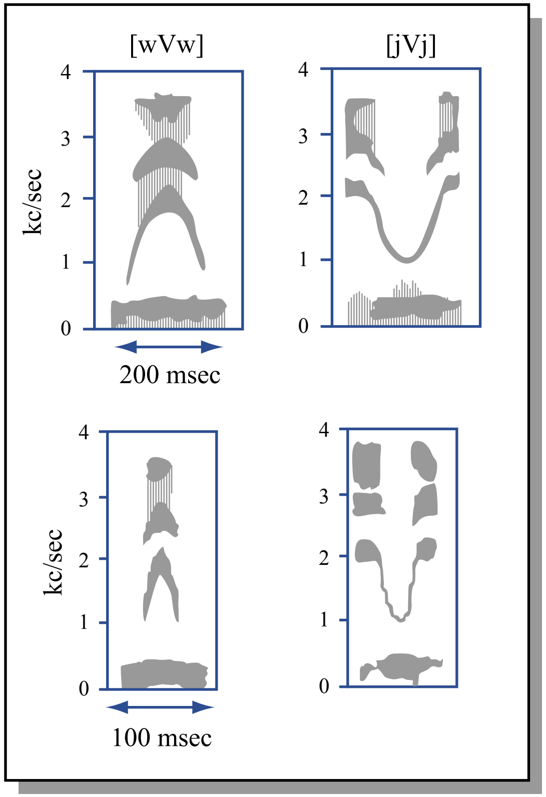
1. Low-frequency spectral energy, periodicity (Stevens and Blumstein 1981:29)
2. Voice onset time (Lisker 1975)
3. Amplitude of aspiration (Repp 1979)
4. Amplitude of release burst (Repp 1979)
5. Closure duration (Lisker 1957)
6. Duration of the preceding vowel (Massaro and Cohen 1983)
7. F1 adjacent to closure (Lisker 1975, Kingston and Diehl 1995)
8. f0 adjacent to the closure (Haggard, Ambler and Callow 1970)
9. Amplitude of F1 at release (Lisker 1986).

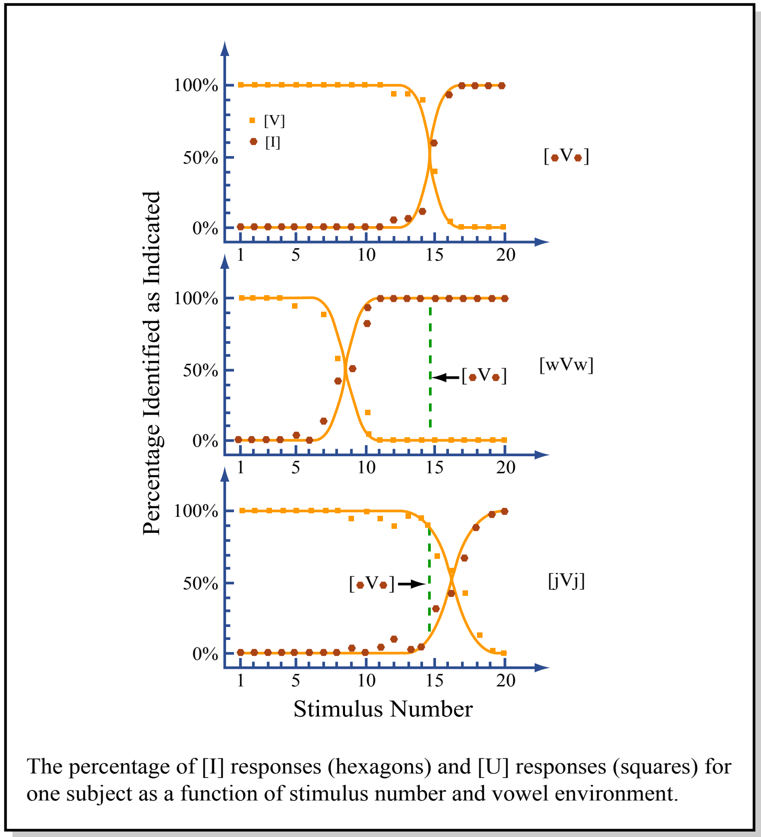
The nature of acoustic cues

- Cues to a contrast are temporally distributed and cues to more than one contrast may be present in the signal simultaneously (i.e. no strict segmentation).
- The availability and nature of the cues to a given contrast type vary systematically with context.
- Listeners take context into account in categorizing signals. E.g. Lindblom and Studdert-Kennedy 1967, Mann and Repp 1980.

Segmental context-dependence

- Lindblom and Studdert-Kennedy (1967)





The percentage of [I] responses (hexagons) and [U] responses (squares) for one subject as a function of stimulus number and vowel environment.

Image by MIT OpenCourseWare. Adapted from Lindblom, Bjšrn, and Michael Studdert-Kennedy. "On the Role of Formant Transitions in Vowel Recognition." *Journal of the Acoustical Society of America* 42 (1967): 830-843.

Image by MIT OpenCourseWare. Adapted from Lindblom, Bjšrn, and Michael Studdert-Kennedy. "On the Role of Formant Transitions in Vowel Recognition." *Journal of the Acoustical Society of America* 42 (1967): 830-843.

Segmental context-dependence

- Mann and Repp (1980)

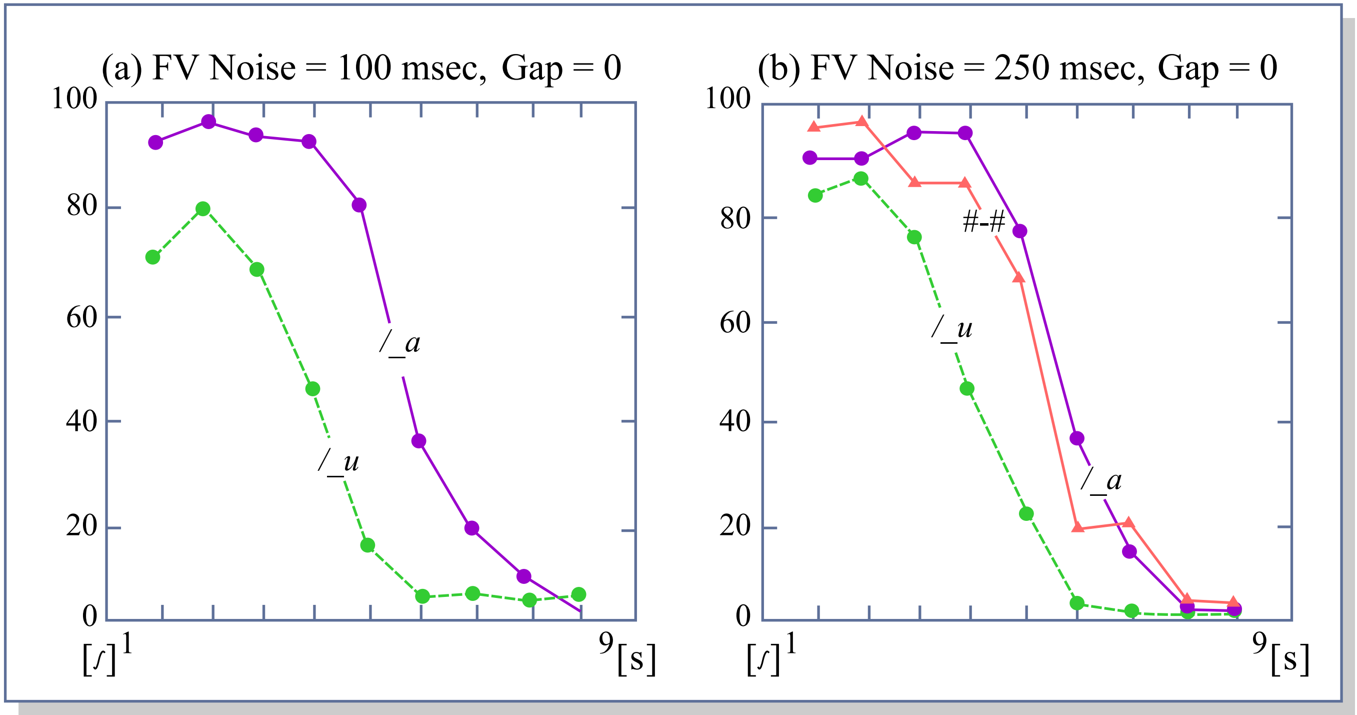


Image by MIT OpenCourseWare. Adapted from Mann, V. A., and B. H. Repp. "Influence of Vocalic Context on the Perception of [ ]-s] distinction: I. Temporal Factors." *Perception and Psychophysics* 28 (1980): 213-228.

Categorical perception

- Strict categorical perception is said to occur where discrimination performance is limited by identification performance, i.e. listeners only have access to category labels, so stimuli can only be distinguished if they are identified as belonging to different categories.
- Tested in two stages:
  - Identification of a synthetic continuum
  - Discrimination of stimuli from the continuum

Categorical perception

- E.g. Liberman (1970) place of articulation F2 transition continuum, b-d-g.

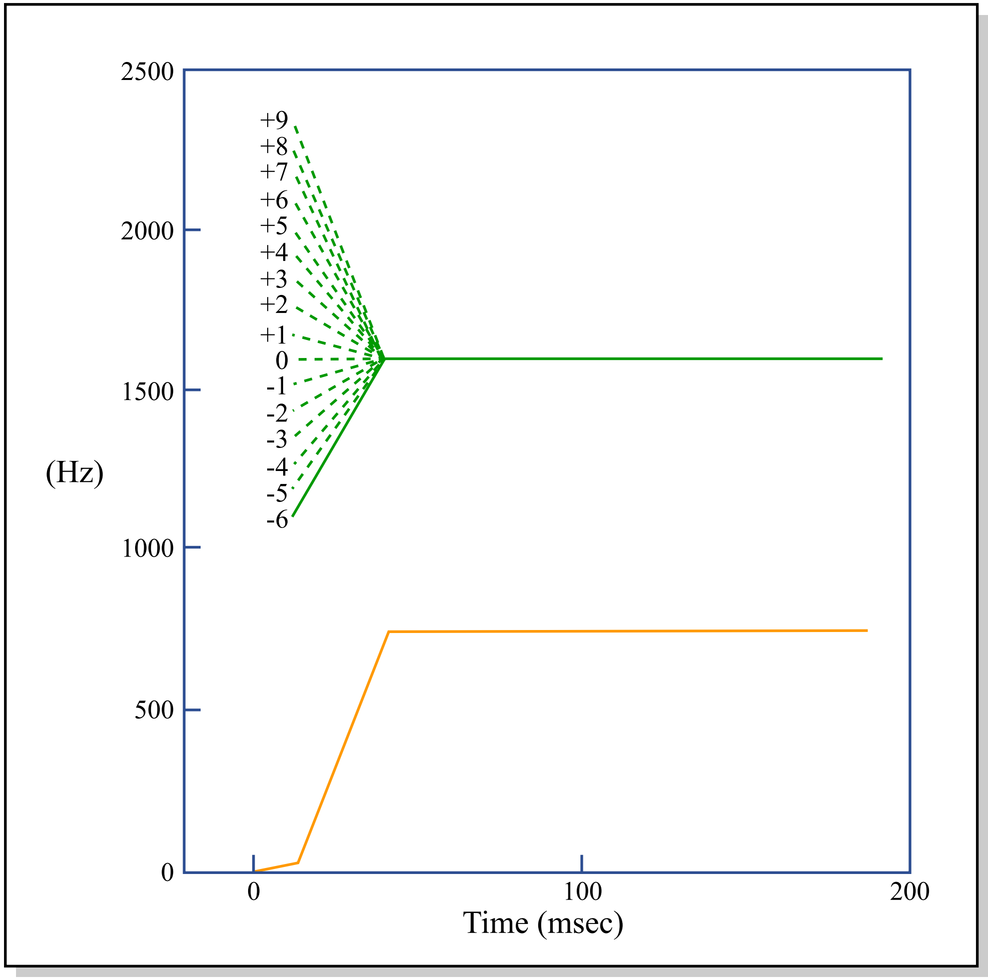


Image by MIT OpenCourseWare. Adapted from Liberman, A. M. "Some Characteristics of Perception in the Speech Mode." *Perception and its Disorders* 48 (1970): 238-254. And Liberman, A. M. "Discrimination in Speech and Nonspeech Modes." *Cognitive Psychology* 2 (1970): 131-157.

Categorical perception

- Identification: Subjects identify stimuli as b, d, g
- Discrimination: Subjects are presented with pairs of stimuli and asked to judge whether they are the same or different.

- Relatively abrupt transitions in identification functions.
- Peaks in discrimination function at the category boundary

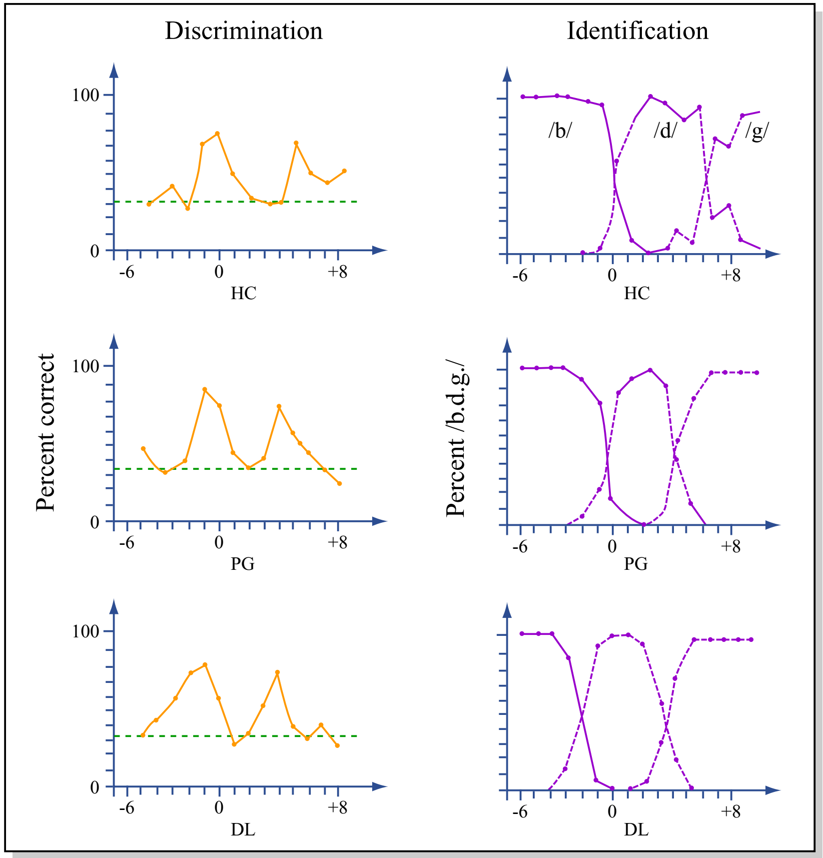


Image by MIT OpenCourseWare. Adapted from Liberman, A. M. "Some Characteristics of Perception in the Speech Mode." *Perception and its Disorders* 48 (1970): 238-254. And Liberman, A. M. "Discrimination in Speech and Nonspeech Modes." *Cognitive Psychology* 2 (1970): 131-157.

Categorical perception

- Discrimination has never been found to be precisely predictable from identification - Discrimination is always better than predicted.
- More loosely, categorical perception is sometimes said to be exhibited where there is a discrimination peak at the category boundary determined by identification, even if the relationship is not precisely as predicted.
- A sharp transition in the 'identification function' for a stimulus continuum is *not* categorical perception in any technical sense.

Why is categorical perception significant?

- The (loose) categorical perception pattern contrasts with the pattern observed in psychophysical experiments using non-speech stimuli:

  > "Typically, nonspeech stimuli that vary acoustically along a single continuum are perceived continuously, resulting in discrimination functions that are monotonic with the physical scale" (Luce and Pisoni, p.31).

- This contrast was used by Liberman and others to argue that speech perception is 'special' – i.e. it uses special mechanisms, not the general mechanisms of non-speech auditory perception.

Why is categorical perception significant?

- Vowels are not usually perceived categorically, even in the loose sense (Luce and Pisoni and refs there).
- The argument for specialness from categorical perception has been weakened by:
  - –Evidence for categorical perception of non-speech sounds (noise-buzz, Miller et al 1976).
  - –Evidence that Chinchillas perceive a VOT continuum categorically (Kuhl and Miller 1975).
  - –But the debate sputters on…

References

- Delattre, Pierre C., Alvin M. Liberman, Franklin S. Cooper, and Louis J. Gerstman (1952). An experimental study of the acoustic determinants of vowel color: observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word* 8, 195-210.
- Haggard, Mark P., Stephen Ambler, and Mo Callow (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America* 47, 613-17.
- Jun, Jongho (1995). *Perceptual and Articulatory Factors in Place Assimilation: An Optimality-Theoretic Approach*. PhD dissertation, UCLA.
- Kawasaki, Haruko (1982). An acoustical basis for universal constraints on sound sequences. Ph.D. dissertation, University of California, Berkeley.
- Kingston, John, and Randy L. Diehl (1995). Intermediate properties in the perception of distinctive feature values. Bruce Connell and Amalia Arvaniti (eds) *Papers in Laboratory Phonology IV* , Cambridge University Press, Cambridge.
- Lindblom, Björn, and Michael Studdert-Kennedy (1967). On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America* 42, 830-843.
- Lisker, Leigh (1957). Closure duration and the intervocalic voiced-voiceless distinctions in English. *Language* 33, 42-49.
- Lisker, Leigh (1975). Is it VOT or a first formant transition detector?. *Journal of the Acoustical Society of America* 57, 1547-51.

References

- Lisker, Leigh (1986). “Voicing” in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29.3-11.
- Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on the perception of [ ]-s] distinction: I. Temporal factors. *Perception & Psychophysics*, 28, 213-228.
- Massaro, Dominic W., and Michael M. Cohen (1983). Consonant/vowel ratios: An improbable cue in speech. *Perception and Psychophysics* 33, 501-5.
- Ohala, J.J. (1990) The phonetics and phonology of aspects of assimilation. M. Beckman and J. Kingston (eds) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. CUP, Cambridge.
- Plomp, Reinier (1975). Auditory analysis and timbre perception. Gunnar Fant and Michel Tatham (eds.) *Auditory Analysis and Perception of Speech*. Academic Press, New York, 7-22.
- Repp, Bruno (1979). Relative amplitude of aspiration noise as a cue for syllable-initial stop consonants. *Language and Speech* 22, 947-950.
- Shepard, Roger N. (1972). Psychological representation of speech sounds. Edward David and Peter Denes (eds.) *Human Communication: A Unified View*. McGraw-Hill, New York, 67-113.
- Steriade, Donca (1997). Phonetics in phonology: the case of laryngeal neutralization. Ms, UCLA.
- Stevens, Kenneth N., and Sheila E. Blumstein (1981). The search for invariant acoustic correlates of phonetic features. Peter D. Eimas and Joanne L. Miller (eds.) *Perspectives on the study of speech*. Lawrence Erlbaum, Hillsdale.
- Wright, R., Frisch, S., & Pisoni, D. B. (1999). Speech Perception. In J. G. Webster (Ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*, Vol. 20 (pp. 175-195). New York: John Wiley and Sons.