

24.963

Linguistic Phonetics

Perception - cues

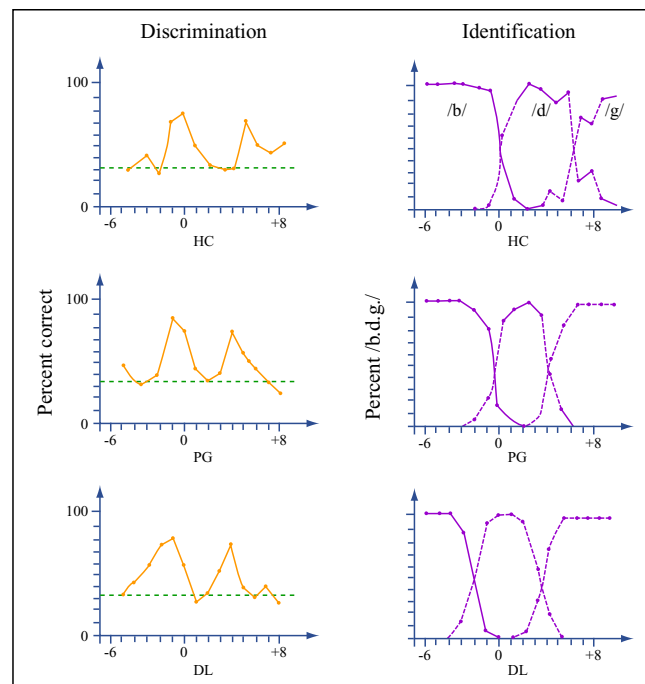


Image by MIT OCW.

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.

- Reading for next week: Johnson ch. 9
- Assignments:
 - Acoustics assignment 4, due 11/3
 - Talk to me about a paper topic.

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.

Perceptual Cues

- Production studies can reveal differences between minimally contrasting words, e.g. aspirated and unaspirated stops in Mandarin differ in VOT.
 - Are listeners sensitive to these differences in speech perception?
- Most direct test of perceptual significance of an acoustic property: manipulate the acoustic property synthetically (or by editing, resynthesis) and see if perceptual response is affected.
 - E.g. vary VOT in synthetic CV syllables (or by editing natural utterances), and have listeners classify the syllables as voiced or voiceless.

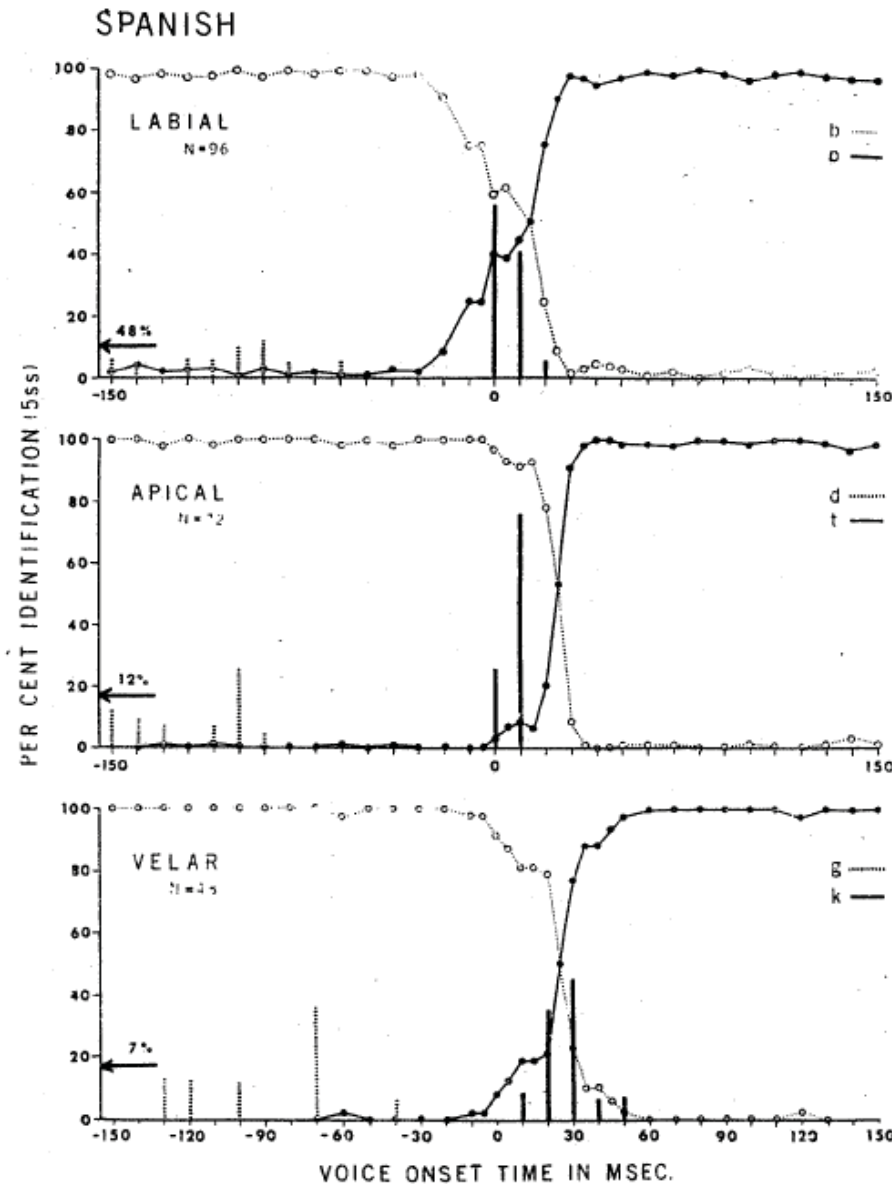
VOT as a cue to stop voicing and aspiration contrasts

- Lisker & Abramson (1970)

- Synthesized a VOT continuum for /Ta/ syllables.
 - Three steady state formants for [a]
 - Formant transitions for labial, coronal, dorsal stop places (3 continua)
 - 37 VOT variants -150 ms to +150 ms
 - Negative VOT = voicing bar during stop closure
 - Positive VOT filled with aspiration noise
 - 10 ms steps, except 5 ms steps from -10 to +50.
- Stimuli presented to
 - 5 speakers of Latin American Spanish
 - 12 speakers of American English
 - 8 speakers of Thai
- Subjects classified stimuli using orthographic labels (forced choice).
 - e.g. 'ba', 'pa'

VOT - Lisker & Abramson (1970)

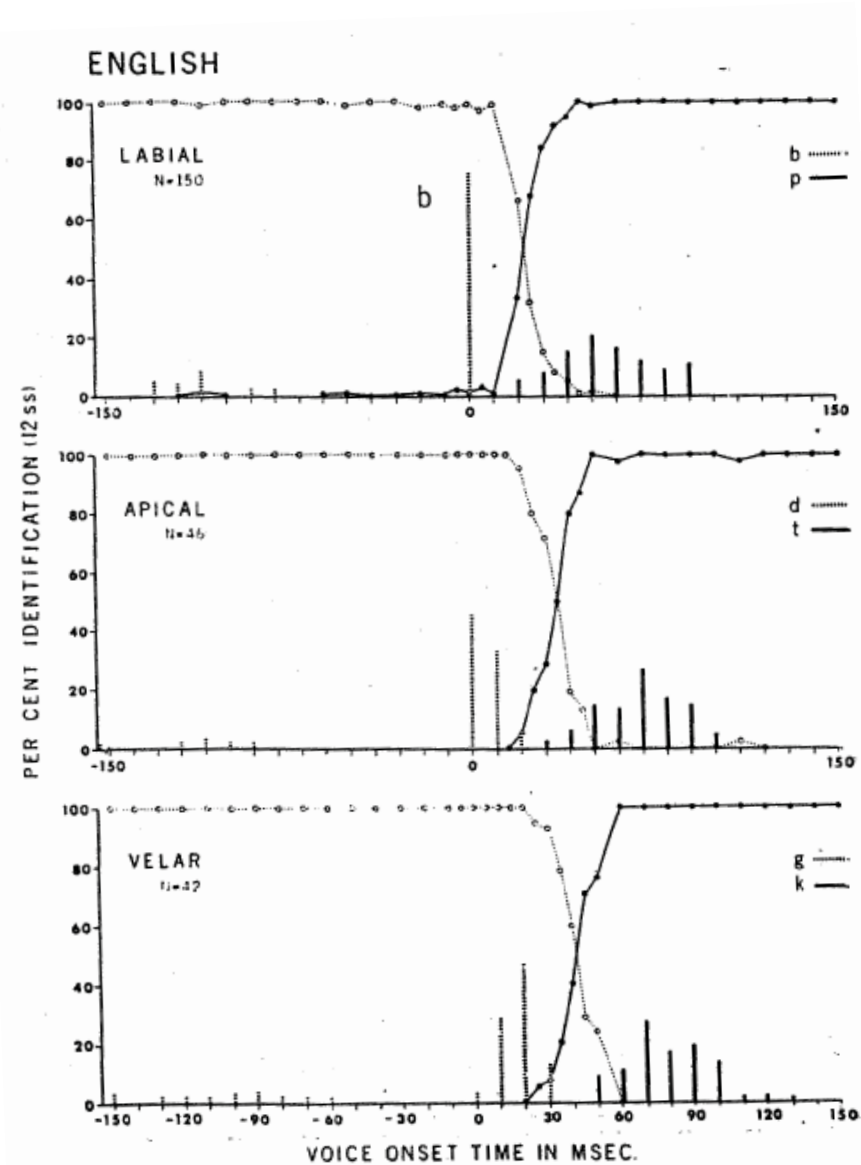
- Histograms show frequencies of different VOT ranges from a production study.
- Lines are identification functions for voiced (dashed) and voiceless (solid).
- Spanish contrasts voiced vs. voiceless unaspirated [b vs. p]



Courtesy of Arthur Abramson and Leigh Lisker. Used with permission.
Source: Lisker, Leigh, and Arthur S. Abramson. "The voicing dimension: Some experiments in comparative phonetics." In Proceedings of the 6th international congress of phonetic sciences, pp. 563-567. Academia Prague, 1970.

VOT - Lisker & Abramson (1970)

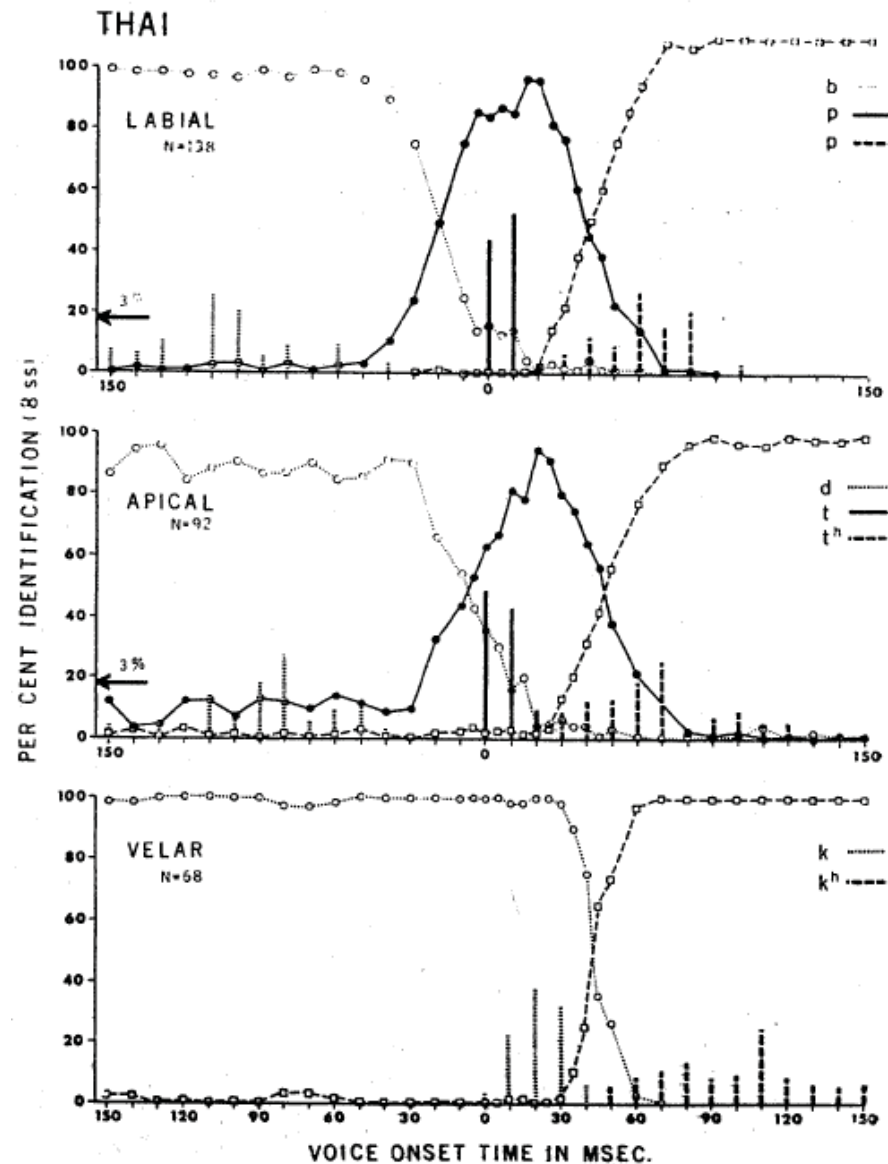
- In initial position, English usually contrasts voiceless unaspirated with aspirated [p, p^h], with the occasional voiced stop [b~p].
- Place affects VOT in production and interpretation of VOT in perception
 - Velar > alveolar > labial



Courtesy of Arthur Abramson and Leigh Lisker. Used with permission.
Source: Lisker, Leigh, and Arthur S. Abramson. "The voicing dimension: Some experiments in comparative phonetics." In Proceedings of the 6th international congress of phonetic sciences, pp. 563-567. Academia Prague, 1970.

VOT - Lisker & Abramson (1970)

- In initial position, Thai contrasts voiced, voiceless unaspirated and aspirated stops [b, p, p^h].



Courtesy of Arthur Abramson and Leigh Lisker. Used with permission.
 Source: Lisker, Leigh, and Arthur S. Abramson. "The voicing dimension: Some experiments in comparative phonetics." In Proceedings of the 6th international congress of phonetic sciences, pp. 563-567. Academia Prague, 1970.

Excursus: 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.

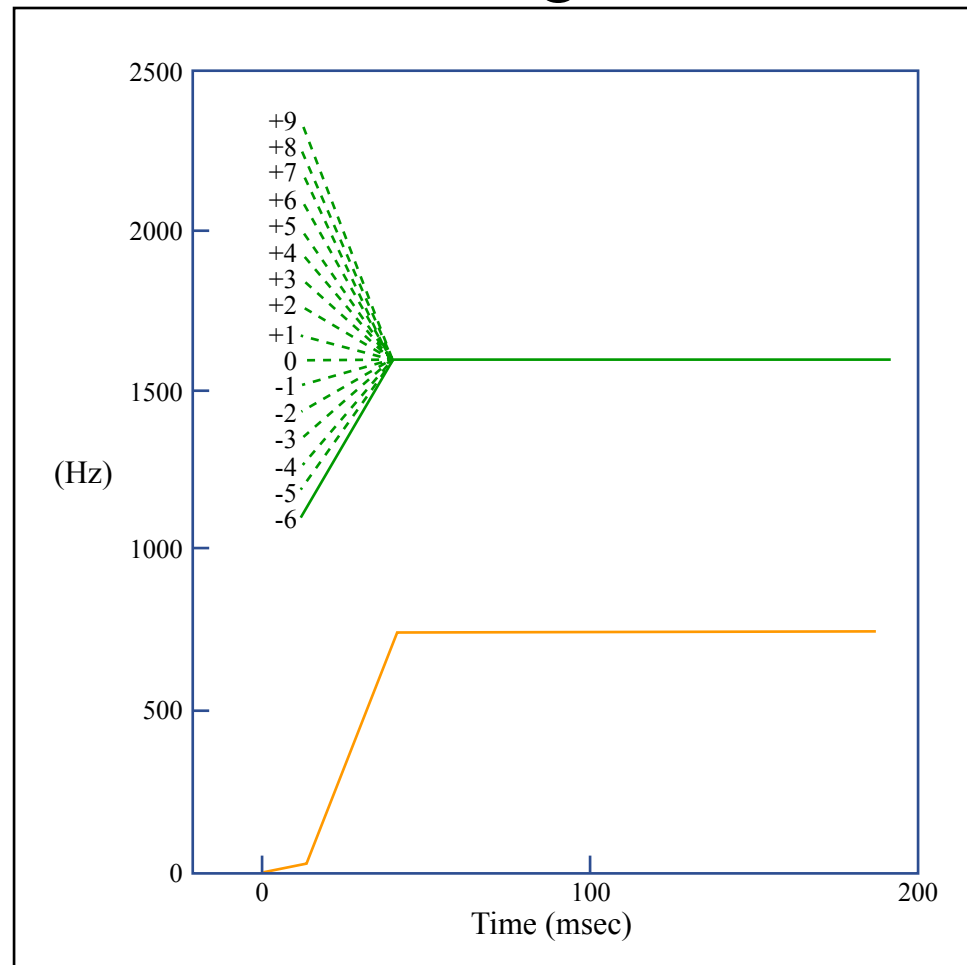


Image by MIT OCW.

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

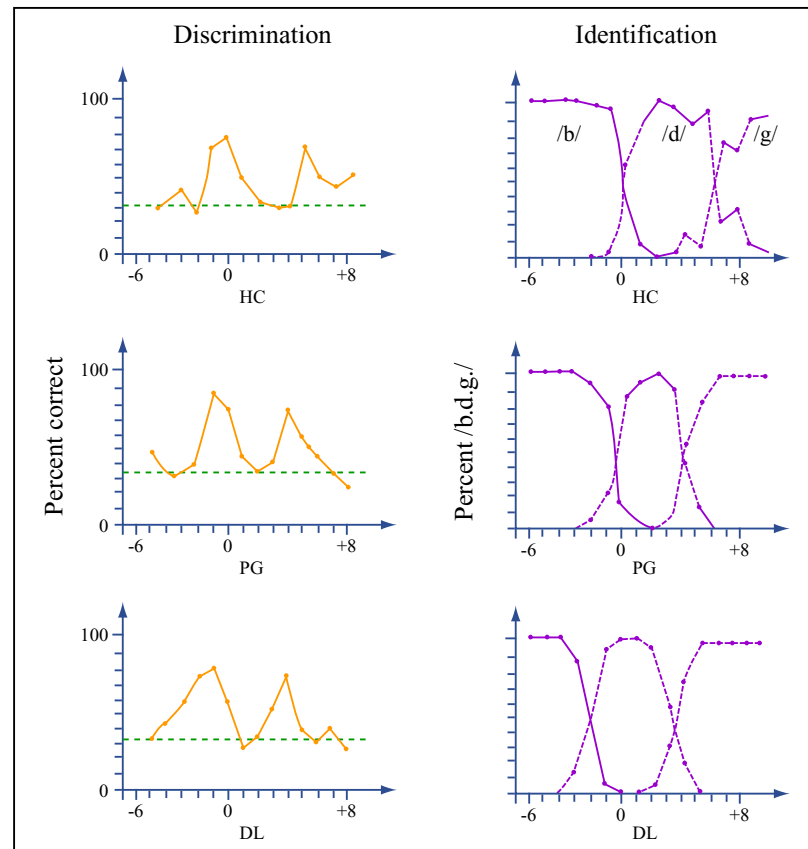


Image by MIT OCW.

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 by strict categorical perception.
- A sharp transition in the ‘identification function’ for a stimulus continuum is *not* categorical perception in any technical sense.

Why is categorical perception significant?

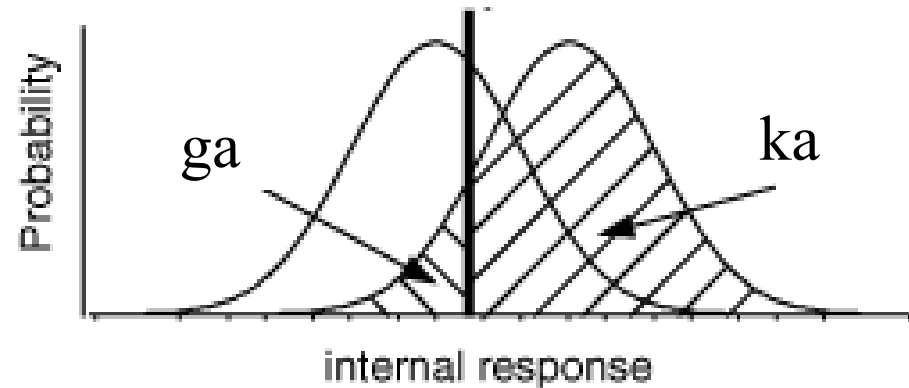
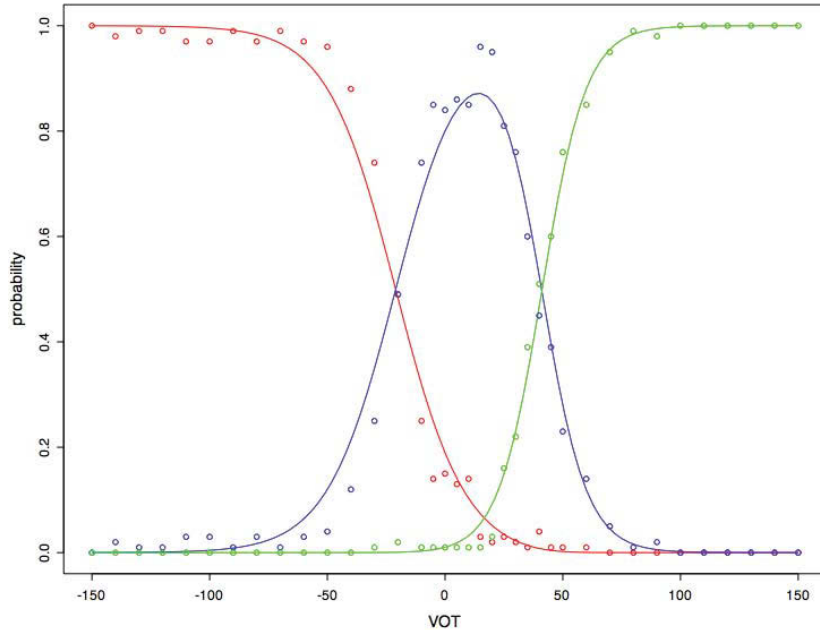
- The (loose) categorical perception pattern contrasts with the pattern observed in psychophysical experiments using non-speech stimuli:
 - “Typically, nonspeech stimuli that vary acoustically along a single continuum are perceived continuously, resulting in discrimination functions that are monotonic 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).
- It has been argued that perception is actually essentially continuous, with categorical effects arising from a categorical decision (natural or experimentally imposed) (e.g. Massaro).

Modeling categorical perception

- ‘Noisy’ continuous perception plus a decision criterion can account for the shape of identification functions



© Source Unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>.

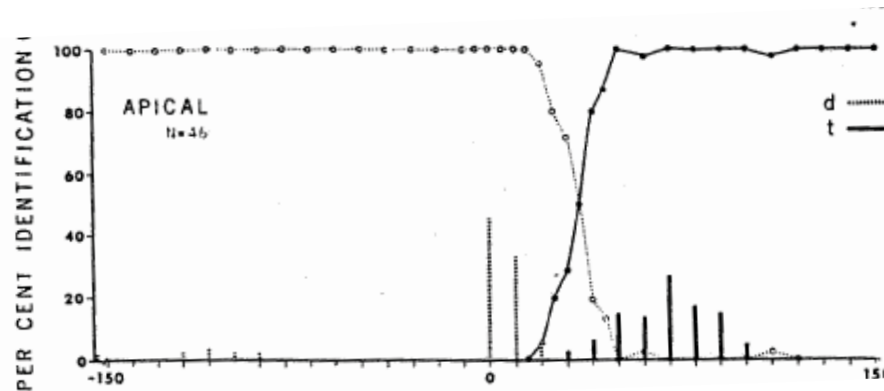
- Doesn't explain discrimination peak at boundary.

Modeling categorical perception

- The occurrence of discrimination peaks at category boundaries follows from a Bayesian model of discrimination (Feldman et al 2009).
- When presented with two stimuli in a discrimination task, the listener is trying to locate the stimuli on a perceptual dimension (e.g. VOT).
- Given the presence of noise in the signal (and perceptual process), the perceived VOT of a stimulus may differ from the actual VOT.
- Listener has to estimate the most likely VOT for that stimulus given the perceived value, and their knowledge of the prior probabilities of different values of VOT.
 - Best estimate of VOT is shifted towards values with higher probability.

Modeling categorical perception

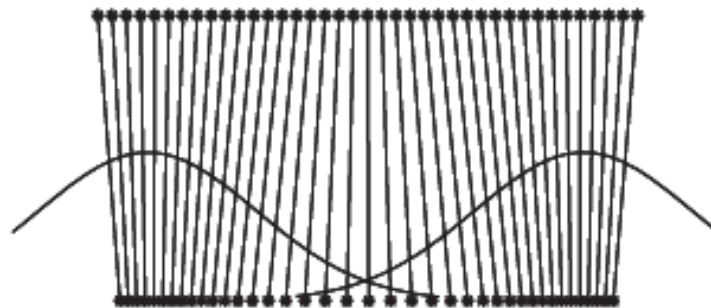
- VOT values close to the category boundary have a low probability, so perceived VOT values near the boundary are shifted towards the category centers, resulting in better discrimination



Courtesy of Arthur Abramson and Leigh Lisker. Used with permission.
Source: Lisker, Leigh, and Arthur S. Abramson. "The voicing dimension: Some experiments in comparative phonetics." In Proceedings of the 6th international congress of phonetic sciences, pp. 563-567. Academia Prague, 1970.

(b)

Actual Stimulus



Perceived Stimulus

Cues

- Perception experiments based on synthetic/edited speech have been used to establish the perceptual significance of a wide variety of acoustic correlates of linguistic contrasts.

Cues to consonant contrasts

- Place cues (Wright, Frisch and Pisoni 1999)

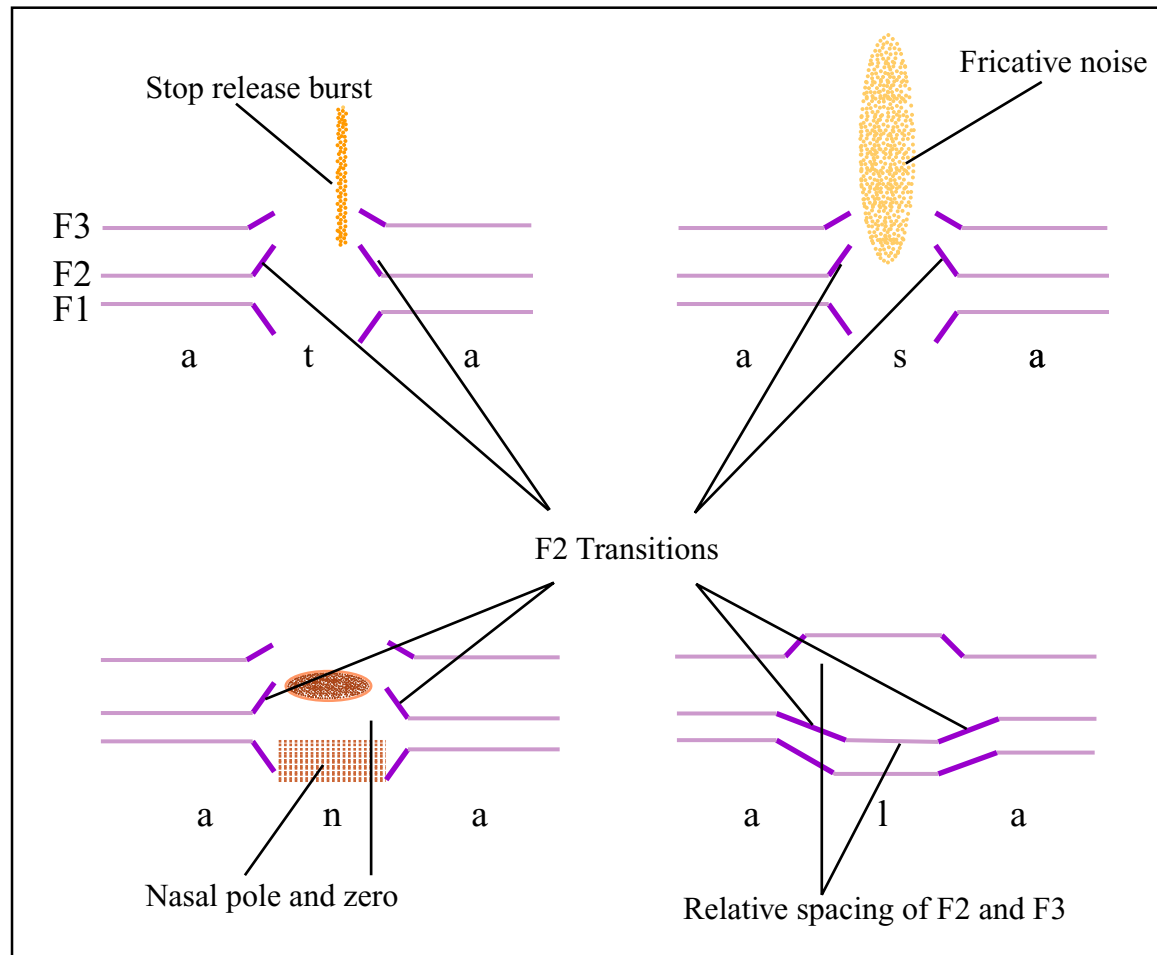


Image by MIT OCW.

Adapted from Wright, R., S. Frisch, 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)

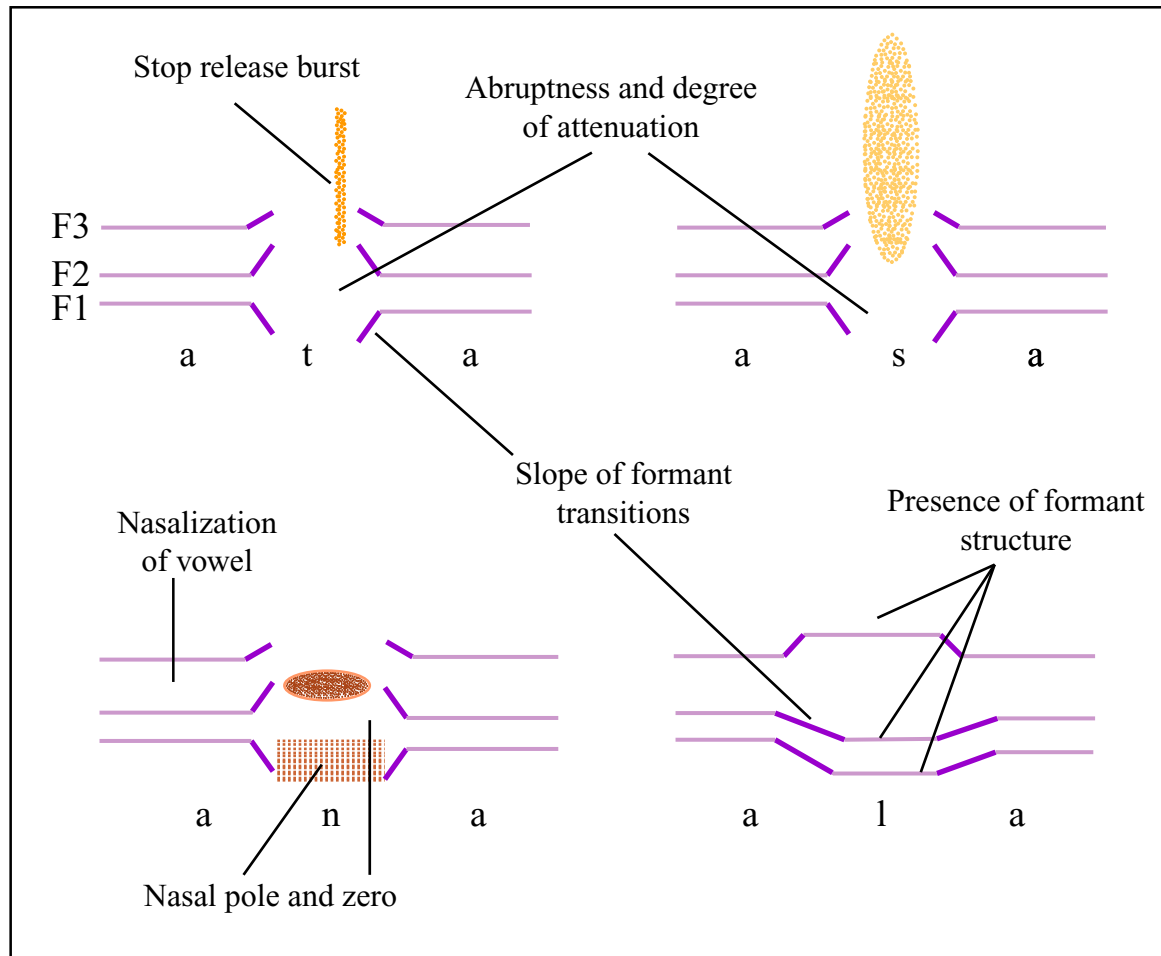


Image by MIT OCW.

Adapted from Wright, R., S. Frisch, 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)

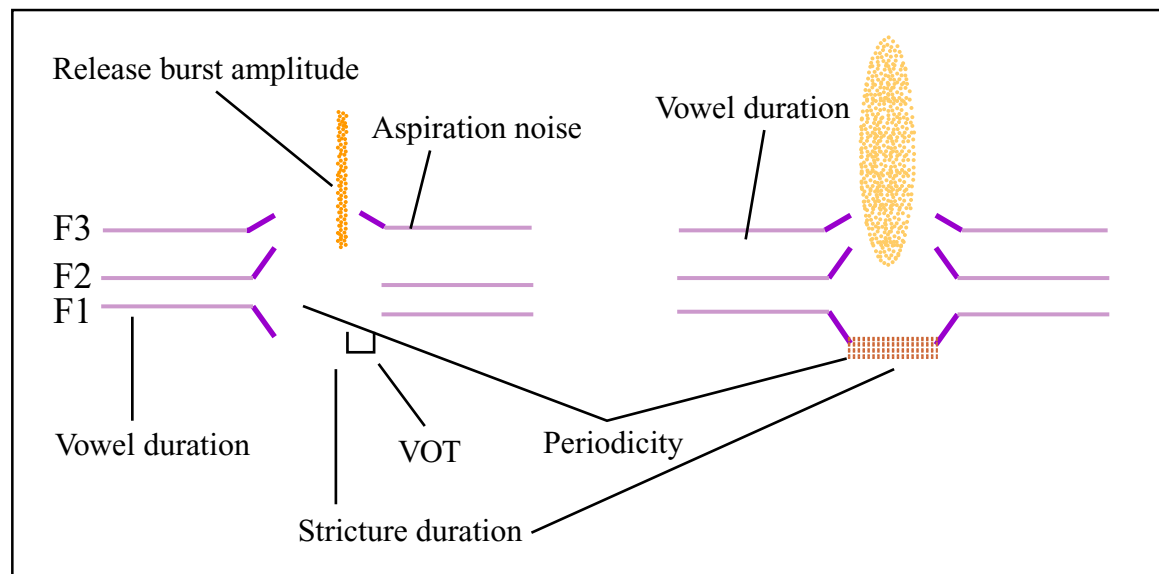


Image by MIT OCW.

Adapted from Wright, R., S. Frisch, 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

- Properties of cues:
 - There are multiple cues to every contrast
 - these cues combine to distinguish sounds
 - cues can vary in their relative strengths
 - Individual cues can vary in strength - e.g. longer VOT is a stronger cue to voicelessness.

The nature of acoustic cues

- There are multiple cues to every contrast
- E.g. stop voicing in English
 1. Low-frequency spectral energy, periodicity (Stevens and Blumstein 1981:29)
 2. Voice onset time (Lisker & Abramson 1970)
 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).

F_0 as a cue to stop voicing

Ohde (1984)

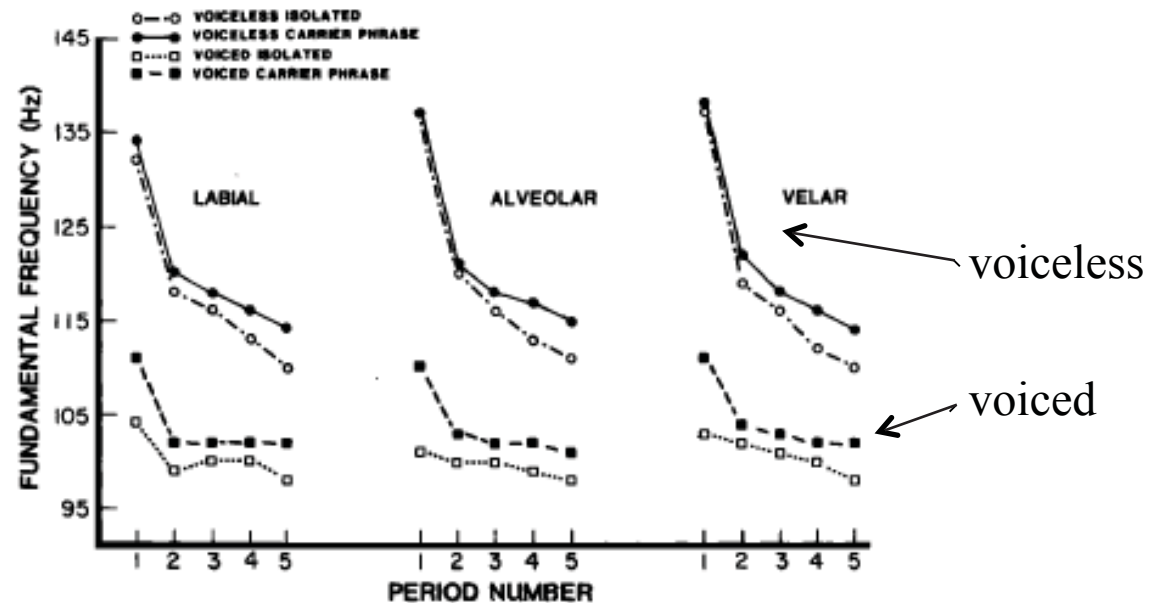
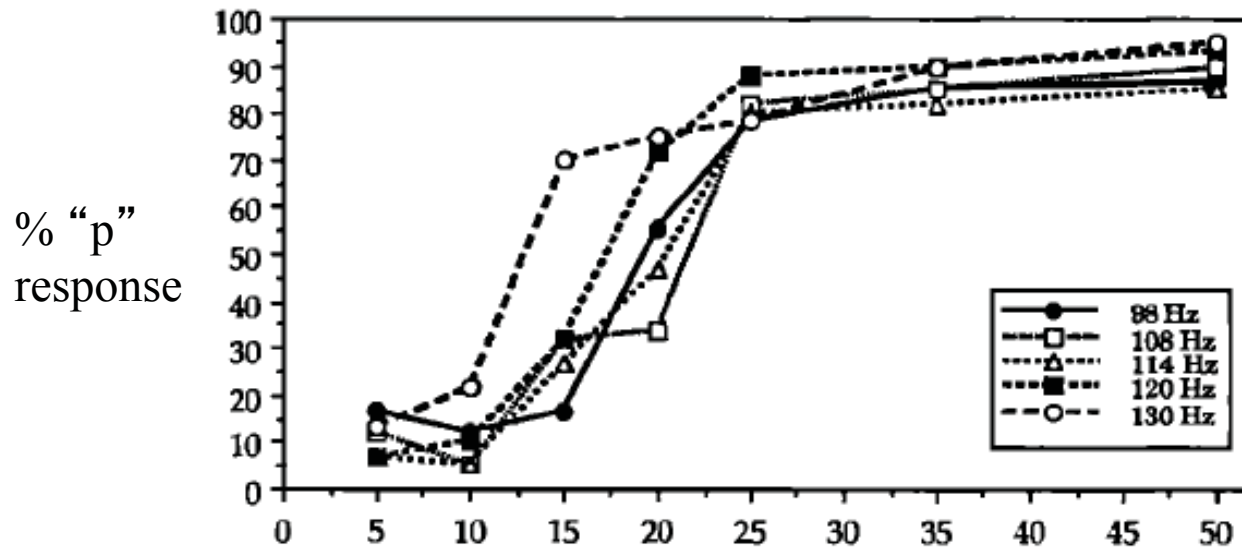


FIG. 2. Average F_0 from voicing onset to the fifth glottal period for voiceless aspirated and voiced stops as a function of linguistic context and place of articulation.

© The Journal of the Acoustical Society of America. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>.
Source: Ohde, Ralph N. "Fundamental frequency as an acoustic correlate of stop consonant voicing." The Journal of the Acoustical Society of America 75, no. 1 (1984): 224-230.

- F_0 is higher after voiceless obstruents than after voiced obstruents (other things being equal)

F₀ as a cue to stop voicing



© The Journal of the Acoustical Society of America. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>.
Source: Whalen, Douglas H., Arthur S. Abramson, Leigh Lisker, and Maria Mody. "F₀ gives voicing information even with unambiguous voice onset times." The Journal of the Acoustical Society of America 93, no. 4 (1993): 2152-2159.

VOT

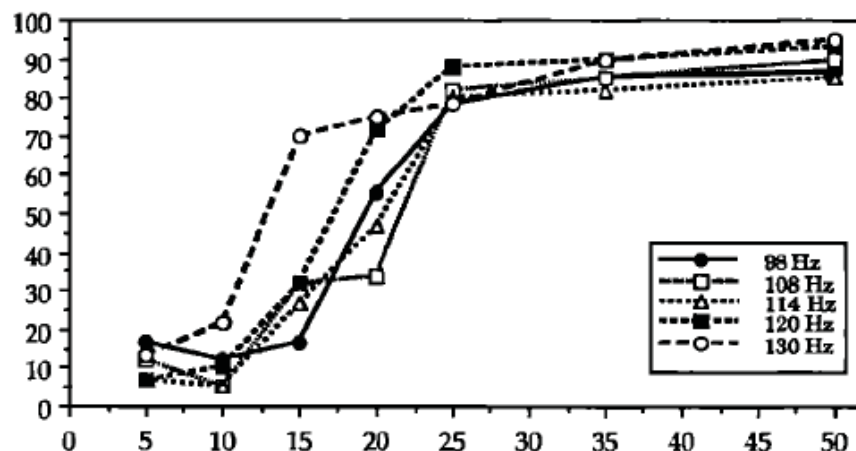
- This difference is exploited as a cue - higher F₀ onset results in more voiceless percepts, for a given VOT.
- Whalen et al (1993): Synthetic [ba-pa] continuum.
 - varied VOT
 - varied F₀ onset, with a linear trajectory to 114 Hz over the first 50 ms of the vowel

Figure removed due to copyright restrictions.

Source: Figures 1 & 2, Raphael, Lawrence J. "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English." *The Journal of the Acoustical Society of America* 51, no. 4B (1972): 1296-1303.

Cue strength

- The multiple cues to voicing all contribute to voicing judgments
 - 20 ms VOT, 98 Hz is ambiguous (~55% voiceless)
 - 20 ms VOT, 120 Hz is less ambiguous (>70% voiceless)
- Cues can differ in strength.



© The Journal of the Acoustical Society of America. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>.

Source: Whalen, Douglas H., Arthur S. Abramson, Leigh Lisker, and Maria Mody. "F₀ gives voicing information even with unambiguous voice onset times." The Journal of the Acoustical Society of America 93, no. 4 (1993): 2152-2159.

- Higher VOT (e.g. 40ms) is a stronger cue to voicelessness than lower values (e.g. 20ms)
- Typical values of VOT provide stronger cues to voicing than typical f₀ differences

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.
- 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.
- 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.
- Lisker, Leigh (1986). “Voicing” in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29.3-11.

References

- 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.
- Ohde, Ralph (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America* 75(1), 224-230.
- 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.

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

24.915 / 24.963 Linguistic Phonetics
Fall 2015

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