

BCS 561: Speech Perception and Recognition

Spring 2006

Instructor: Richard Aslin

Wednesdays 1:15 - 3:15 PM

Meliora 418

Course Description

Provides an overview of the theories and empirical findings on human speech perception and recognition. Topics include an overview of phonetics, categorical perception, speech perception by nonhumans and by human infants, perception of nonnative speech sounds, intermodal perception of speech, and the interaction between speech perception and production. Additional topics include talker, rate, and dialect variability, perception of sine-wave and noise-vocoded speech, gradient effects in consonant perception, cochlear implants and plasticity, brain correlates of speech perception (fMRI, MEG, ERP), and automatic speech recognition.

Schedule of topics and list of readings

February 1: Historical perspective on speech perception

Readings:

Liberman, A. M. (1996). Introduction: Some assumptions about speech and how they changed. In A. M. Liberman (Ed.), *Speech: A special code*. Cambridge, MA: MIT Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431-461.

Liberman, A. M. and Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4, 187-196.

February 8: Categorical perception and gradiency

Readings:

Pisoni, D.B. and Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 15(2), 285-290.

Miller, J.L. (1997). Internal structure of phonetic categories. *Language and Cognitive Processes*, 12, 865-869.

McMurray, B., Tanenhaus, M., and Aslin, R. (2002). Gradient effects of within-category phonetic variation on lexical access, *Cognition*, 86(2), B33-B42.

Kuhl, P. K. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, but monkeys do not. *Perception & Psychophysics*, 50, 93-107 (NOTE: read only Experiments 1 and 2).

Lotto, A. J., Kluender, K. R., and Holt, L. L. (1998). Depolarizing the perceptual magnet effect. *Journal of the Acoustical Society of America*, 103, 3648-3655.

Optional readings:

Emmorey, K., McCullough, S. and Brentari, D. (2003). Categorical perception in American Sign Language. *Language and Cognitive Processes*, 18(1), 21-45.

Gerrits, E., and Schouten, M.E.H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, 66(3), 363-376.

Harnad, S. (1987). *Categorical Perception: The Groundwork of Cognition*. New York: Cambridge University Press.

Pisoni, D.B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13(2), 253-260.

February 15: Motor and gesture theories

Readings:

Fowler, C. A. and Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, 36, 171-195.

- Goldstein, L. and Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In N. O. Schiller and A. Meyer (Eds.) *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*. Berlin: Mouton de Gruyter (pp. 159-207).
- Lindblom, B. (1996). Role of articulation in speech perception: Clues from production. *Journal of the Acoustical Society of America*, 99, 1683-1692.
- Remez, R. E. (1996). Critique: Auditory form and gestural topology in the perception of speech. *Journal of the Acoustical Society of America*, 99, 1695-1698.
- Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. *Journal of the Acoustical Society of America*, 99, 1718-1725.
- Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, 99, 1730-1741.

Optional readings:

- Rizzolatti, G. and Arbib, M. A. (1998). Language within our grasp. *Trends in Neuroscience*, 21, 188-194.
- Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, 15, 399-402.

February 22: Speech perception by non-humans

Readings:

- Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences*, 4, 258-267.
- Kuhl, P. K. and Miller, J. D. (1975). Speech perception by the Chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190, 69-72.
- Kuhl, P. K. (1986). Theoretical contributions of tests on animals to the special-mechanisms debate in speech. *Experimental Biology*, 45, 233-265. [Skip pp. 253-265]

- Kluender, K. R., Diehl, R. L., and Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science*, 237, 1195-1197.
- Lotto, A. J., Kluender, K. R., and Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America*, 102, 1134-1140.
- Dent, M. L., Brittan-Powell, E. F., Dooling, R. J., and Pierce, A. (1997). Perception of synthetic /ba/-/wa/ speech continuum by budgerigars (*Melopsittacus undulatus*). *Journal of the Acoustical Society of America*, 102, 1891-1897.
- Sinnott, J. M. and Gilmore, C. S. (2004). Perception of place-of-articulation information in natural speech by monkeys versus humans. *Perception and Psychophysics*, 66, 1341-1350.

Optional Readings:

- Trout, J. D. (2001). The biological basis of speech: What to infer from talking to the animals. *Psychological Review*, 108, 523-549.
- Dooling, R. J., Okanoya, K., and Brown, S. D. (1989). Speech perception by budgerigars (*Melopsittacus undulatus*). *Perception and Psychophysics*, 46, 65-71.

March 1: Speech perception by infants

Readings:

- Eimas, P. D., Siqueland, E. R., Jusczyk, P. W. and Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303-306.
- Werker, J. F. and Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49-63.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by six months of age. *Science*, 205, 606-608.
- Rosenblum, L. D., Schmuckler, M. A., and Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, 59, 347-357.

- Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101-111.
- Kuhl, P. K., Tsao, F-M., and Liu, H-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, 100, 9096-9101.
- Vouloumanos, A. and Werker, J. F. (2004). Tuned to the signal: the privileged status of speech for young infants. *Developmental Science*, 7, 270-276.
- McMurray, B. and Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition*, 95, B15-26.
- Werker, J. F. and Yeung, H. H. (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences*, 9, 519-527.

Optional Readings:

- Jusczyk P. W. (1997). *The discovery of spoken language*. MIT Press.
- Werker, J. F. and Curtin, S. (2005). PRIMIR: a developmental framework of infant speech processing. *Language Learning and Development*, 1, 197-234.

March 8: Brain mechanisms of speech perception

Readings:

- Scott, S. K. and Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26, 100-107.
- Price, C., Thierry, G., and Griffiths, T. (2005). Speech-specific auditory processing: where is it? *Trends in Cognitive Sciences*, 9, 271-276.
- Benson, R. R., Whalen, D. H., Richardson, M., Swainson, B., Clark, V. P., Lai, S. and Liberman, A. M. (2001). Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain and Language*, 78, 364-396.
- Vouloumanos, A., Kiehl, K. A., Werker, J. F. and Liddle, P. F. (2001). Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience*, 13, 994-1005.

Liebenthal, E., Binder, J. R., Piorowski, R. L. and Remez, R. E. (2003). Short-term reorganization of auditory analysis induced by phonetic experience. *Journal of Cognitive Neuroscience*, 15, 549-558.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A. and Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *NeuroImage*, 24, 21-33.

Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T. and Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex*, 15, 1621-1631.

Blumstein, S. E., Myers, E. B. and Rissman, J. (2005). The perception of voice onset time: An fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, 17, 1353-1366.

Optional Readings:

Scott, S. K. and Wise, R. J. S. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition*, 92, 13-45.

Zatorre, R. J., Belin, P. and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, 6, 37-46.

March 22: Cue reliability and trading relations

Readings:

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustic Society of America*, 101, 2299-2310.

Ernst, M. O. and Bulthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8, 162-169.

Francis, A. L., Baldwin, K. and Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics*, 62, 1668-1680.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustic*

Society of America, 107, 2711-2724.

Mayo, C. and Turk, A. (2004). Adult-child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions. *Journal of the Acoustic Society of America, 115, 3184-3194.*

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., and McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience, 2, 89-108.*

Repp, B. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech-mode of perception. *Psychological Bulletin, 92, 81-110.*

Optional Readings:

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28, 349-366.*

Mitterer, H., Csepe, V., Honbolygo, F., and Blomert, L. (in press). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science.*

March 29: Effects of indexical variables (talker, rate, dialect) on speech perception

Readings:

Nygaard, L. C. (2005). Perceptual integration of linguistic and nonlinguistic properties of speech. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception*. Blackwell (pp. 390-413).

Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America, 115, 3171-3183.*

McLennan, C. T. & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory & Cognition, 31, 306-321.*

Clarke, C. M., & Garrett, M. (2004). Rapid adaptation to foreign accented speech. *Journal of the Acoustical Society of America*, 116, 3647-3658.

Kraljic, T. & Samuel, A. G. (in press). How general is perceptual learning for speech? *Psychonomic Bulletin & Review*.

Evans, B. B., & Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America*, 115, 352-361.

Optional Readings:

Goldinger, S. (1998). Echoes of echoes: An episodic theory of lexical access. *Psychological Review*, 105, 251-279.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, 60, 355-376.

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener- and item-related factors. *Journal of the Acoustical Society of America*, 106 (4), 2074-2085.

April 5: Perception of non-speech signals (Remez guest speaker)

Readings:

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America*, 61, 1352-1361.

Holt, L. L. (2005). Temporally nonadjacent sounds affect speech categorization. *Psychological Science*, 16, 305-312.

Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, 212, 947-950.

Remez, R. E., Pardo, J. S., Piorkowski, R. L., and Rubin, P. E. (2001). On the bistability of sinewave analogues of speech. *Psychological Science*, 12, 24-29.

Remez, R. E. (2005). Perceptual organization of speech. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception*. Blackwell (pp. 28-50).

Shannon, R. V., Zeng, F-G., Kamath, V., Wygonsky, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270, 303-304.

Wade, T. and Holt, L. L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *Journal of the Acoustical Society of America*, 118, 2618-2633.

Optional Readings:

Jusczyk, P. W., Pisoni, D. B., Walley, A., and Murray, J. (1980). Discrimination of relative onset time of two-component tones by infants. *Journal of the Acoustical Society of America*, 67, 262-270.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., and Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review*, 101, 129-156.

Holt, L. L., Lotto, A. J., and Diehl, R. L. (2004). Auditory discontinuities interact with categorization: Implications for speech perception. *Journal of the Acoustical Society of America*, 116, 1763-1773.

Mirman, D., Holt, L. L., and McClelland, J. L. (2004). Categorization and discrimination of nonspeech sounds: Differences between steady-state and rapidly-changing acoustic cues. *Journal of the Acoustical Society of America*, 116, 1198-1207.

Barker, J. and Cooke, M. (1999). Is the sine-wave speech cocktail party worth attending? *Speech Communication*, 27, 159-174.

Davis, M. H., Johnsruide, I. S., Hervais-Adelman, A., Taylor, K. & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, 134, 222-241.

Keidel, J. L. (unpublished ms). Perceptual change induced by learning of novel auditory categories. University of Wisconsin.

April 12: Levels of analysis and processing of speech signals

Readings:

Lahiri, A. & Marslen-Wilson, W. (1991). The mental representation of lexical form: A

phonological approach to the recognition lexicon. *Cognition*, 38, 245-294.

Pierrehumbert, J. (2003). Probabilistic Phonology: Discrimination and Robustness. In R. Bod, J. Hay and S. Jannedy (Eds.), *Probability Theory in Linguistics*. MIT Press.

Magnuson, J.S., Tanenhaus, M.K., Aslin, R.N. (under review). Which words compete? Dynamic similarity during spoken word recognition.

Salverda, A. Dahan, D, Tanenhaus, M., Crosswhite K., Masharov, M., & McDonough, J. (under review). Effects of prosodically-modulated sub-phonetic variation on lexical neighborhoods.

Vitevitch, M.S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28, 735-747.

Ferreira, V. S. & Griffin, Z. M. (2003). Phonological influences on lexical (MIS)selection. *Psychological Science*, 14, 86-90.

Optional Readings:

Wheeldon, L. & Waksler, R. (2004). Phonological underspecification and mapping mechanisms in the speech recognition lexicon. *Brain and Language*, 90, 401-412.

Stevens, K. N. (2005). Features in Speech Perception and Lexical Access. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception*. Blackwell (pp. 125-155).

April 19: Speech perception via cochlear implants and adaptive plasticity

Readings:

Pisoni, D. B. (2005). Speech perception in deaf children with cochlear implants. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception*. Blackwell (pp. 494-523).

Shannon, R. V., Zeng, F-G., and Wygonski, J. (1998). Speech recognition with altered spectral distribution of envelope cues. *Journal of the Acoustical Society of America*, 104, 2467-2476.

Rosen, S., Faulkner, A., and Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *Journal of the Acoustical Society of America*, 106, 3629-3636.

Eisenberg, L. S., Shannon, R. V., Martinez, A. S., Wygonski, J., and Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *Journal of the Acoustical Society of America*, 107, 2704-2710.

Burkholder, R. A., Pisoni, D. B., and Svirsky, M. A. (in press). Perceptual learning and nonword repetition using a cochlear implant simulation. *JEP:HPP*. [Progress Report No. 26, Indiana University, Speech Research Laboratory]

Optional Readings:

Loizou, P. C. (1998). Mimicking the human ear. *IEEE Signal Processing Magazine*, September, 101-130.

April 26: Automatic speech recognition devices

Readings:

Scharenborg, O., Norris, D., Bosch, L., and McQueen, J. M. (2005). How should a speech recognizer work? *Cognitive Science*, 29, 867-918.

Zue, V. (2004). Eighty challenges facing speech input/output technologies. Paper presented at a conference "From Sound to Sense", MIT.

Wet, F., Weber, K., Boves, L., Cranen, B., Bengio, S., and Boulard, H. (2004). Evaluation of formant-like features on an automatic vowel classification task. *Journal of the Acoustical Society of America*, 116, 1781-1792.

Optional Reading:

Knill, K. and Young, S. (1997). Hidden markov models in speech and language processing. In S. Young & G. Bloothoof (Eds.), *Corpus-based methods in language and speech processing*. Kluwer (pp. 27-68).