A BRIEF HISTORY OF SPEECH PERCEPTION RESEARCH IN THE UNITED STATES

Michael Studdert-Kennedy and D. H. Whalen
Haskins Laboratories

INTRODUCTION

A hundred and fifty years ago, Alexander Melville Bell (1849) prefigured an insight that has come to shape research on speech perception only in recent decades: There is a powerful link between the way we perceive speech and the way we produce it. Bell's system of transcription, his "visible speech" (Bell 1867), reportedly allowed speakers who knew the system to reproduce exactly any utterance not only in languages they knew, but in languages they did not. Thus, by the intermediary of a phonetic script, Bell unfolded the imitative capacity implicit in every untutored child who automatically recovers from speech the articulatory gestures that shape it, and so learns to speak a native language.

Yet, curiously, modern studies of speech perception and speech production have generally followed separate paths at laboratories where only one or the other topic was of interest. Only quite recently have researchers begun to argue that a viable theory of speech perception must be grounded in a viable theory of speech production, and vice versa. The reaction to this stance, either for or against, defines much of the field of speech perception today.

EARLY WORK

Telephonic Communication

Early work, in the years after World War II, was largely guided by the demands of telephonic communication. Its aim was to estimate how much distortion (by frequency-bandwidth compression, amplitude peak-clipping, filtering, noise, and so on) could be imposed on the speech signal without seriously reducing its intelligibility (for a review, see Miller, 1951). Three general conclusions were surprising and important. First, speech is so resistant to distortion that we can throw away large parts of the signal without seriously reducing its intelligibility. Second, intelligibility does not depend on naturalness. These first two facts have enabled us to learn a great deal about the important information-bearing elements of speech by stripping it to its minimal acoustic skeleton.

A third conclusion, confirmed in many later studies, was that when speech perception breaks down in noise, it tends to do so along the dimension, or features, of traditional articulatory phonetics. English consonants, for example, are more likely to be confused within than across manner (stop, fricative, nasal) and voicing classes (Miller and Nicely, 1955). By corollary, place of articulation is the feature most susceptible to damage by noise; fortunately for the hearing-impaired, it is also the feature most easily seen on a talker's lips.

The Sound Spectrograph

Study of the auditory bases for articulatory perception became possible with the development of the sound spectrograph at Bell Laboratories during World War II (Koenig, Dunn, and Lacy 1946). The spectrograph provided a visual record not of the physical signal as it impinges on the ear, but of its time-varying Fourier transform as it is assumed to be represented at the output of the cochlea. Strictly, then, the representation is auditory (psychological), not acoustic (physical), and it was originally hoped that the spectrograph would enable deaf persons to use the telephone (Potter, Kopp, and Green 1947); but this proved impracticable because spectrograms are formidably difficult to read.

The difficulty arises from the astonishing variability of the speech signal, both within and among speakers. JOOS (1948), in a monograph still well worth reading, first described the variability. But experimental investigation awaited development of the Pattern Playback at Haskins Laboratories in New York.

The Pattern Playback

The Playback reconverted the visual pattern of a spectrogram into a speech sound sequence with a constant fundamental frequency (COOPER 1950; Cooper, LIBERMAN, and Borst 1951). Experimenters laid a transparent acetate loop over a spectrogram and traced the formant pattern with white paint. The pattern was then rolled at a constant speed, matched to the time scale of the spectrogram, beneath a strip of frequency-modulated light. The light was reflected from the painted portions of the pattern to a photocell that drove a loudspeaker, thus reproducing an approximation to the original sound. The Playback (and its more flexible computer successors at Haskins and elsewhere) permitted experimenters to manipulate the speech signal systematically, by pruning, deleting or exaggerating portions of a spectrographic pattern until they had isolated those pieces that determine the perception of a particular utterance.

One broad conclusion from the first perceptual studies has stood, and has guided research, for over 40 years: Information in the speech signal is not conveyed by an acoustic alphabet. The invariant phonetic segments of the perceived message do not correspond one-for-one to segments in the acoustic signal (LIBERMAN, COOPER, Shankweiler, and STUDDERT-KENNEDY 1967). Due to coarticulation, that is, due to the overlapping actions of articulators engaged by successive segments, segment boundaries become interleaved, and the acoustic pattern specifying a given segment varies with its context. Thus, in a typical consonant-vowel-consonant syllable, acoustic information for all three segments may be distributed, both temporally and spectrally, over the entire syllable. This lack of isomorphism

between signal and message has been, and continues to be, the central puzzle of speech perception.

ACOUSTIC FEATURES

Categorical Perception

Early work with synthetic speech revealed that tokens of syllables contrasting on a single phonetic feature could be constructed by manipulating a single acoustic variable. For example, by varying the direction of the second formant (F2) transition at the onset of a CV syllable, an experimenter could construct a continuum of a dozen or so items, separated by acoustically equal steps, ranging from /bæ/ to /dæ/ to /gæ/. If listeners were then asked to identify tokens from the continuum, they typically divided them into clear-cut categories, despite the absence of obvious acoustic boundary markers. Moreover, asked to discriminate between tokens two steps apart, say, on the continuum, listeners did little better than chance if they had assigned them to the same category, but performed very well if they had assigned them to different categories. The phenomenon was dubbed "categorical perception" (LIBERMAN, Harris, Hoffman, and Griffith 1957) to distinguish it from the "continuous perception" typical of non-speech continua, such as tones varying in pitch or loudness, for which discrimination is equally good across the entire continuum (see Harnad 1987, for a collection of articles).

Many experiments eventually established that the level of discrimination within categories varies with experimental method (e.g. Pisoni 1973; Carney, Widin and Viemeister 1977; Miller, Connine, Schermer and Kluender 1983), and that categorical perception is not confined to speech (e.g. Pastore, Ahroon, Baffuto, Friedman, Puleo and Fink 1977), or even perhaps to humans (e.g. Kuhl and Miller 1978). Nonetheless, the phenomenon does characterize speech, and widespread use of the identification/discrimination paradigm has proved fruitful in establishing phonological differences among languages (e.g. Miyawaki, STRANGE, Verbrugge, LIBERMAN, JENKINS and FUJIMURA 1975), infant capacity for speech perception (e.g. EIMAS, Siqueland, Jusczyk and Vigorito 1971) and the distinction between auditory and phonetic perception (e.g. Mann and Liberman 1983).

Quantal Theory

Among the offshoots of work on categorical perception was the quantal theory of speech (STEVENS 1972; 1989). Stevens attributed the lack of acoustic category boundary markers in synthetic speech studies to the fact that categories were there defined by articulatorily impossible variations in a single acoustic variable (e.g. F2 formant transitions) rather than by the whole-spectrum properties (e.g. grave-acute, compact-diffuse) of distinctive feature theory (JAKOBSON, Fant and HALLE 1951/1963; Chomsky and Halle 1968). Stevens' goal has been to derive the articulatory and acoustic properties of the postulated features by applying the acoustic theory of speech production to an idealized model of the vocal tract. The acoustic properties selected are those few that are both easy to articulate (because they are centered in regions of acoustic stability where large changes in some articulatory parameter have little acoustic effect) and easy to discriminate (because they are bounded by regions of acoustic discontinuity where small articulatory changes have a large effect).

Quantal theory thus rejects the claim that speech is not an acoustic alphabet. The theory proposes, rather, that the speech signal is a sequence of discrete spectral patterns, invariant across context, each integrated perceptually over brief intervals by property detectors characteristic of the mammalian auditory system. Note that temporal properties are explicitly excluded from the description of a feature; this omission has proved to be the central weakness of the theory's account of perception. A series of experimental studies of the acoustic structures that support stop consonant perception both by STEVENS' colleagues (e.g. BLUMSTEIN, Isaacs and Mertus 1982; Lahiri and Blumstein 1984) and by others (e.g. Kewley-Port, Pisoni and STUDDERT-KENNEDY 1983; Walley and Carrell 1983) have come down clearly in favor of dynamic, context-dependent formant patterns rather than of the gross, static spectral invariants postulated by quantal theory (for critiques of the theory, see the special issue of *Journal of Phonetics*, Volume 17, July, 1989).

ACOUSTIC CUES

Unlike features, cues are empirically defined properties of spectrally and temporally limited portions of the signal that have been shown (usually by manipulation of a synthesized syllable) to contribute to perception of a standard articulatory dimension. The invention of the Pattern Playback opened the way to systematic description of the acoustic cues for phonological categories. Within less than a decade of the initial work, a preliminary set of "minimal rules for synthesizing speech" was proposed (LIBERMAN, Ingemann, LISKER, DELATTRE and COOPER 1959).

Perhaps the most surprising discovery of this and later work was that virtually every phonetic contrast is carried by several spectrally and temporally distributed cues. The critical importance of time was first recognized by LISKER and ABRAMSON (1964) who showed by analysis of natural utterances, that the several spectro-temporal properties specific to perception of the voicing, aspiration or "tensity" of homorganic stops in many languages reflect the timing of laryngeal action relative to consonant release (voice onset time, or "VOT").

Other work showed that place of articulation is signaled in syllable-initial English stops by spectral properties of the release burst and of formant transitions at vowel onset (LIBERMAN, DELATTRE and COOPER 1952; Dorman, STUDDERT-KENNEDY and Raphael 1977); in syllable-initial fricatives by spectral properties of the frication noise and of its formant transitions into the vowel (e.g. Harris 1958; Whalen 1981); in the unaspirated stops of English [s]-stop clusters by duration of the stop closure, by spectral properties at the offset of the [s], and by the relation between those properties and those of the following vowel (Bailey and Summerfield 1980). Even for vowels, sometimes taken to be relatively static formant patterns (PETERSON and Barney 1952), critical information in a CVC syllable is carried not only by the nucleus, but by onset and offset transitions (e.g. Lindblom and Studdert-Kennedy 1967; STRANGE, Verbrugge, Shankweiler and Edman 1976).

In all these examples, cues do not occur in "simultaneous bundles", as posited for distinctive features, but in temporal sequences that reflect the course of articulatory action. Many studies of reciprocal relations among cues, as in so-called "trading relations" (e.g. REPP 1983; Kluender 1991), and of multiple cue function (e.g. Bailey and Summerfield 1980) have

indeed demonstrated that cues are *additive* components of a coherent pattern of sound, and that their coherence is intrinsic to the speech signal itself, imposed not by perception, but by the speaker's articulations. Further support for this conclusion comes from studies of sinewave speech and of lipreading.

SINE WAVE SPEECH

Sine wave speech is generated from a radically reduced copy of a spectrogram in which only the center frequencies of the formants are preserved. Intelligible speech can be constructed for semantically implausible, and therefore unpredictable, utterances from which all information about source (voicing, friction, plosive release), nasality, harmonic spectrum, and fundamental frequency has been removed, so that the listener hears no more than a crude approximation to the peak resonances of the changing cavity shapes and volumes of the vocal tract (REMEZ, Rubin, Pisoni and Carrell 1981; Remez, Rubin, Berns, Pardo and Lang 1994). Most listeners come to hear such bizarre combinations of whistles as speech after brief instruction and little or no practice. We do not infer from this work that the diverse acoustic properties of natural speech, eliminated from sine wave speech, have no function. We infer, rather, that these properties are integral components of the dynamic patterns of spectral change to which listeners are demonstrably sensitive.

LIPREADING

Studies of lipreading in recent years have taken on a new theoretical importance, largely precipitated by the well-known McGurk effect (McGurk and MacDonald 1976), in which mismatches between what is seen and what is heard can lead to speech perception that is based on portions of each modality. At issue is the question of whether the listener/viewer combines phonetic features extracted independently from the two channels (MASSARO 1987), or integrates optic and acoustic information into a continuous, time-varying, precategorical event structure (Summerfield 1987). Studies in which one or other signal is ambiguous if presented alone, but the combination is not (e.g. Green and Miller 1985; FOWLER and Dekle 1991) support the latter interpretation, as do studies in which prelinguistic infants prefer an acoustic-optic match to an acoustic-optic mismatch (MacKain, STUDDERT-KENNEDY, Spieker and Stern 1983; Kuhl and Meltzoff 1984). Such studies corroborate the conclusion, independently drawn from work on cue function and sine wave speech, that the information-bearing elements of speech are articulator movements, or gestures.

PROSODY

Prosody refers to the suprasegmental melody, amplitude and timing of speech (LEHISTE 1970; Martin 1972). A central concern has been its perceived isochrony, seemingly absent from the signal (Morton, Marcus and Frankish 1976). FOWLER (1979; 1980) has argued, however, that the perceived regularity is based on acoustic information about articulatory timing, concealed in the signal by gestural overlap. The onsets of gestures overlap, and so the acoustic output can be confusing. Others (Howell 1987; Pompino-Marschall 1989) have argued for an articulation-free acoustic basis, but their work seems to ignore the effects of later occurring information (Cooper, Whalen, and Fowler 1988). The competition between articulatory and acoustic explanations continues to inform this research.

SPECIALIZATION FOR SPEECH PERCEPTION?

The question of whether speech perception engages general auditory or specialized phonetic mechanisms first arose from attempts to devise an acoustic alphabet to substitute for the optic alphabet in a reading machine for the blind (LIBERMAN, et al. 1967). Despite innumerable attempts, no one was able to devise an acoustic alphabet that listeners could follow faster than Morse code, that is, a rate of some 10-15 words per minute, roughly a tenth of a typical English speaking rate. What accounts for our ease in following speech?

The answer hangs on the nature of the speech percept. On one view, perhaps the most widely held, the percept is auditory, an amalgam of cues that we have learned to associate with linguistic dimensions, or features (e.g. Diehl and Kluender 1989). Perceptual coherence then emerges from spectrotemporal diversity according to the Gestalt "laws" of visual perception, adapted to audition by Bregman (1990). (But see also the arguments in REMEZ, et al. 1994.) On this account, we follow speech with peculiar ease because of its Gestalt structure and because we have been hearing it continually since infancy.

On a second view, the direct realist view (FOWLER 1986; Best 1995), the percept is articulatory. Whether by ear, by eye, or by hand, we perceive the gestures that structure the energy in the signal. We follow speech with ease because speech has evolved to match our perceptual systems, and our perceptual systems have evolved to pick up information about objects and events in the world (Gibson 1979).

On a third view, the motor theory of speech perception (LIBERMAN and MATTINGLY 1985), the percept is again articulatory, but is achieved by a specialized computational device that has evolved to recover discrete phonetic gestures from the intricately shingled articulatory and acoustic structures that make rapid speech possible. Evidence consistent with a specialized mode of phonetic perception has come from studies of dichotic listening (Kimura 1967; STUDDERT-KENNEDY and Shankweiler 1970; Zatorre, Evans, Meyer and Gjedde 1992) and of so-called "duplex perception". In the latter, listeners are led to hear a synthetic sine wave transition as simultaneously a non-speech glissando and an integrated phonetic component of a stop-vowel syllable (e.g. Xu, Liberman and Whalen 1997).

DEVELOPMENT OF SPEECH PERCEPTION

A large and still growing body of work on infant speech perception began with a demonstration of categorical perception in one- and four-month-old infants (EIMAS, et al. 1971). Within a few years, research had shown that infants could discriminate virtually any speech contrast from any language during the first six months of life (e.g. Kuhl 1976), but that over the second half of the first year, they gradually lose the capacity to discriminate non-native contrasts (Werker, Gilbert, Humphrey and Tees 1981), especially those that are close to, but not the same as, native contrasts (Best 1995). Over this period, infants also become sensitive to recurrent word patterns, to phonotactic constraints in the surrounding language, and even to prosodic markers of clausal units. (For a comprehensive review, see Jusczyk 1997).

CURRENT TRENDS AND FUTURE DIRECTIONS

The past 10-15 years have seen a shift away from the segment and the invariance issue toward the word, and even longer stretches of the signal, where the goal is less to discover invariants than to understand how listeners master and exploit variability (e.g. Perkell and KLATT 1986). Among the growing points in the area are studies of word recognition, both in isolation (Elman and McClelland 1984; Pisoni and Luce 1987; Luce and Pisoni 1998) and in running speech (e.g. Marslen-Wilson 1973). Such work and continued research along older lines, revitalized perhaps by the new techniques of brain imaging now emerging, should make for an interesting history at the 25th International Congress of Phonetic Sciences in 2043.

REFERENCES

BAILEY, P. J. and Q. SUMMERFIELD. 1980. Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 536-563.

BELL, A. M. 1849. A new elucidation of the principles of speech and elocution.

BELL, A. M. 1867. *Visible Speech*. London: Simpkin & Co.

BEST, C. T. 1995. A direct realist perspective on cross-language speech perception. In W. Strange and J.J. Jenkins (Eds.), *Cross-language speech perception*. Timonium, MD: York Press.

BLUMSTEIN, S. E., E. ISAACS, and J. MERTUS. 1982. The role of the gross spectral shape as a perceptual cue to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America*, 72, 43-50.

BREGMAN, A. S. 1990. *Auditory scene analysis*. Cambridge, MA: M.I.T. Press.

CARNEY, A. E., G. P. WIDIN, and N. F. VIEMEISTER. 1977. Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 62, 961-970.

CHOMSKY, N. and M. HALLE, 1968. *The Sound Pattern of English*. New York: Harper and Row.

COOPER, A. M., D. H. WHALEN, and C. A. FOWLER. 1988. The syllable's rhyme affects its P-center as a unit. *Journal of Phonetics*, 16, 231-241.

COOPER, F. S. 1950. Spectrum analysis. *Journal of the Acoustical Society of America*, 22, 761-762.

COOPER, F. S., A. M. LIBERMAN, and J. M. BORST. 1951. The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proceedings of the National Academy of Science*, 37, 318-325.

DIEHL, R. L. and K. R. KLUENDER. 1989. On the objects of speech perception. *Ecological Psychology*, 1, 121-144.

DORMAN, M. F., M. STUDDERT-KENNEDY, and L. J. RAPHAEL. 1977. Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception and Psychophysics*, 22, 109-122.

*EIMAS, P. D., E. R. SIQUELAND, P. W. JUSCZYK, and J. VIGORITO. 1971. Speech perception in infants. *Science*, 171, 303-306.

ELMAN, J. L. and J. L. MCCLELLAND. 1984. Speech perception as a cognitive process: The interactive activation model. In N. J. Lass (Eds.), *Speech and language: Advances in basic research and practice*. New York: Academic Press. 337-374.

FOWLER, C. A. 1979. "Perceptual centers" in speech production and perception. *Perception and Psychophysics*, 25, 375-388.

FOWLER, C. A. 1980. Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 8, 113-133.

*FOWLER, C. A. 1986. An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14, 3-28.

FOWLER, C. A. and D. J. DEKLE. 1991. Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 816-828.

GIBSON, J. J. 1979. *The ecological approach to visual perception*. Boston: Houghton Mifflin.

GREEN, K. P. and J. L. MILLER. 1985. On the role of visual rate information in phonetic perception. *Perception and Psychophysics*, 38, 269-276.

HARNAD, S. (Ed.). 1987. *Categorical perception: The groundwork of cognition*. Cambridge: Cambridge University Press.

HARRIS, K. S. 1958. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, 1, 1-7.

HOWELL, P. 1987. Prediction of P-center location from the distribution of energy in the amplitude envelope: I. *Perception and Psychophysics*, 43, 90-93.

JAKOBSON, R., G. FANT, and M. HALLE. 1951/1963. *Preliminaries to Speech Analysis: The Distinctive Features and their Correlates*. Cambridge, MA: M.I.T..

JOOS, M. 1948. Acoustic phonetics. *Language*, 24, 2.

JUSCZYK, P. 1997. *The discovery of spoken language*. Cambridge, MA: M.I.T. Press.

KEWLEY-PORT, D., D. B. PISONI, and M. STUDDERT-KENNEDY. 1983. Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America*, 73, 1779-1793.

KIMURA, D. 1967. Functional asymmetry of the brain in dichotic listening. *Cortex*, 3, 163-178.

KLUENDER, K. R. 1991. Effects of first formant onset properties on voicing judgments result from processes not specific to humans. *Journal of the Acoustical Society of America*, 90, 83-96.

†KOENIG, W., H. K. DUNN, and L. Y. LACY. 1946. The sound spectrograph. *Journal of the Acoustical Society of America*, 18, 19-49.

KUHL, P. K. 1976. Speech perception in early infancy: The acquisition of speech-sound categories. In S.K. Hirsh, D.H. Eldredge, I.J. Hirsh and S.R. Silverman (Eds.), *Hearing and Davis: Essays honoring Hallowell Davis*. St. Louis: Washington University Press. 265-280.

KUHL, P. K. and A. N. MELTZOFF. 1984. The intermodal representation of speech in infants. 7, 361-381.

*KUHL, P. K. and J. D. MILLER. 1978. Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, 63, 905-917.

LAHIRI, A. and S. E. BLUMSTEIN. 1984. A re-evaluation of the feature coronal. *Journal of Phonetics*, 12, 133-146.

LEHISTE, I. 1970. *Suprasegmentals*. Cambridge, MA: M.I.T. Press.

*LIBERMAN, A. M., F. S. COOPER, D. P. SHANKWEILER, and M. STUDDERT-KENNEDY. 1967. Perception of the speech code. *Psychological Review*, 74, 431-461.

LIBERMAN, A. M., P. DELATTRE, and F. S. COOPER. 1952. The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *American Journal of Psychology*, 65, 497-516.

LIBERMAN, A. M., K. S. HARRIS, H. S. HOFFMAN, and B. C. GRIFFITH. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358-368.

LIBERMAN, A. M., F. INGEMANN, L. LISKER, P. C. DELATTRE, and F. S. COOPER. 1959. Minimal rules for synthesizing speech. *Journal of the Acoustical Society of America*, 31, 1490-1499.

*LIBERMAN, A. M. and I. G. MATTINGLY. 1985. The motor theory of speech perception revised. *Cognition*, 21, 1-36.

*LINDBLOM, B. E. and M. STUDDERT-KENNEDY. 1967. On the rôle of formant-transitions in vowel recognition. *Journal of the Acoustical Society of America*, 42, 830-843.

LISKER, L. and A. S. ABRAMSON. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384-422.

LUCE, P. A. and D. B. PISONI. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1-36.

MACKAIN, K. S., M. STUDDERT-KENNEDY, S. SPIEKER, and D. STERN. 1983. Infant intermodal speech perception is a left-hemisphere function. *Science*, 219, 1347-1349.

MANN, V. A. and A. M. LIBERMAN, 1983. Some differences between phonetic and auditory modes of perception. *Cognition*, 14, 211-235.

MARSLEN-WILSON, W. D. 1973. Linguistic structure and speech

shadowing at very short latencies. *Nature*, 244, 522-523.
MARTIN, J. G. 1972. Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, 79, 487-509.
MASSARO, D. W. 1987. *Speech perception by ear and eye: A paradigm for psychological enquiry*. Hillsdale, NJ: Lawrence Erlbaum Associates.
*MCGURK, H. and J. MACDONALD. 1976. Hearing lips and seeing voices. 264, 746-748.
MILLER, G. A. 1951. *Language and communication*. New York: McGraw-Hill.
*†MILLER, G. A. and P. E. NICELY. 1955. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 27, 338-352.
MILLER, J. L., C. M. CONNINE, T. M. SCHERMER, and K. R. KLUENDER. 1983. A possible auditory basis for internal structure of phonetic categories. *Journal of the Acoustical Society of America*, 73, 2124-2133.
*MIYAWAKI, K., W. STRANGE, R. VERBRUGGE, A. M. LIBERMAN, J. J. JENKINS, and O. FUJIMURA. 1975. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, 18, 331-340.
MORTON, J., S. MARCUS, and C. FRANKISH. 1976. Perceptual centers (P-centers). *Psychological Review*, 83, 405-408.
PASTORE, R. E., W. A. AHROON, K. J. BAFFUTO, C. J. FRIEDMAN, J. S. PULEO, and E. A. FINK. 1977. Common-factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 686-696.
PERKELL, J. S. and D. H. KLATT, (Ed.). 1986. *Invariance and variability in speech processes*. Hillsdale, NJ: Lawrence Erlbaum Associates.
†PETERSON, G. E. and H. L. BARNEY. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24, 175-184.
*PISONI, D. B. 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, 13, 253-260.
PISONI, D. B. and P. A. LUCE. 1987. Acoustic-phonetic representations in word recognition. *Cognition*, 25, 21-52.
POMPINO-MARSCHALL, B. 1989. On the psychoacoustic nature of the P-center phenomenon. *Journal of Phonetics*, 17, 175-192.
POTTER, R. K., G. A. KOPP, and H. C. GREEN. 1947. *Visible speech*. New York: Van Nostrand.
REMEZ, R. E., P. E. RUBIN, S. M. BERNS, J. S. PARDO, and J. M. LANG. 1994. On the perceptual organization of speech. *Psychological Review*, 101, 129-156.
REMEZ, R. E., P. E. RUBIN, D. B. PISONI, and T. D. CARRELL. 1981. Speech perception without traditional speech cues. *Science*, 212, 947-950.
REPP, B. H. 1983. Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. *Speech Communication*, 2, 341-362.
STEVENS, K. N. 1972. The quantal nature of speech: Evidence from articulatory-acoustic data. In E.E. David, Jr. and P.B. Denes (Eds.), *Human communication: A unified view*. New York: McGraw-Hill. 51-66.
STEVENS, K. N. 1989. On the quantal nature of speech. *Journal of Phonetics*, 17, 3-45.
STRANGE, W., R. R. VERBRUGGE, D. P. SHANKWEILER, and T. R. EDMAN. 1976. Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America*, 60, 213-224.
*STUDDERT-KENNEDY, M. and D. P. SHANKWEILER. 1970. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, 48, 579-594.
SUMMERFIELD, Q. 1987. Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd and R. Campbell (Eds), *Hearing by eye: The psychology of lip-reading*. London: Lawrence Erlbaum Associates. 3-51.
WALLEY, A. C. and T. D. CARRELL. 1983. Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 73, 1011-1022.
WERKER, J. F., J. H. V. GILBERT, K. HUMPHREY, and R. C. TEES. 1981. Developmental aspects of cross-language speech perception. *Child Development*, 52, 349-355.
WHALEN, D. H. 1981. Effects of vocalic formant transitions and vowel quality on the English [s]-[S] boundary. *Journal of the Acoustical Society of America*, 69, 275-282.
XU, Y., A. M. LIBERMAN, and D. H. WHALEN. 1997. On the immediacy of phonetic perception. *Psychological Science*, 8, 358-362.
ZATORRE, R. J., A. C. EVANS, E. MEYER, and A. GJEDDE. 1992. Lateralization of phonetic and pitch discrimination in speech processing. *Science*, 256, 846-849.

Note: References marked with an asterisk (*) also appear in the collection *Papers in speech communication: speech perception*, edited by Joanne L. Miller, RAYMOND D. KENT, and Bishnu S. Atal. Acoustical Society of America: Woodbury, NY, 1991.

References marked with a dagger (†) also appear in the collection *Readings in acoustic phonetics*, edited by ILSE LEHISTE M.I.T. Press: Cambridge MA, 1967.