

Different Speech-Processing Mechanisms Can Be Reflected in the Results of Discrimination and Dichotic Listening Tasks¹

JAMES E. CUTTING²

Haskins Laboratories and Yale University

The relative peakedness of diotic ABX discrimination functions for certain speech stimuli and the relative magnitude of the right ear advantage that they yield in dichotic listening tasks have been thought to be functionally parallel measures of speech processing. The results of the present study suggest that this is not always the case.

The results of two very different experimental paradigms have been taken as primary evidence distinguishing the perception of speech from the perception of other auditory events. These results are the discriminability functions associated with categorical perception, and the right ear advantage in dichotic listening (see Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967). Other auditory events, which for the most part can be handily classified as nonspeech, typically yield neither of these results (for an overview, see Studdert-Kennedy & Shankweiler, 1970; Mattingly, Liberman, Syrdal & Halwes, 1971). The most parsimonious explanation for the apparently unique processing of speech in both experimental paradigms is that a similar mechanism or set of mechanisms underlies both effects.

Categorical Perception and ABX Discrimination

Usually one can perceive and discriminate many more stimuli along a physical continuum than he can identify (Pollack, 1952, 1953). Often equal increments of physical change yield equal increments of perceptual change. However, for certain kinds of stimuli, especially certain kinds of speech stimuli, this is untrue. The best-studied continua in speech perception have been the dimensions of place of articulation as manifested by second- and third-formant transitions (Liberman, Harris, Hoffman & Griffith, 1957; Pisoni, 1971, 1973a), and voice-onset time

¹ Supported by NICHD Grant No. HD-01994 to the Haskins Laboratories. Requests for reprints should be sent to the author at the Haskins Laboratories, 270 Crown Street, New Haven, CT 06510.

² Now at Wesleyan University, Middletown, CT, and at Haskins Laboratories, New Haven, CT.

(Abramson & Lisker, 1965, 1970). Both dimensions are relevant to the perception of different stop consonants in most languages. Particular adjacent stimuli along these acoustic continua are difficult if not impossible to discriminate. Typically, subjects are able to discriminate items only as well as they can label them differently, and this labeling process is very nearly categorical. Thus, categorical perception implies a finely tuned ability to discriminate items which are acoustically similar but phonetically different, coupled with an inability to discriminate items which are acoustically similar but phonetically identical.

Lieberman, Harris, Kinney, and Lane (1961) have shown that the discriminability peaks associated with categorical perception occur for certain speech sounds but not for acoustically similar nonspeech sounds. Mattingly *et al.* (1971) have shown that these peaks occur for entire speech patterns, but not for the phonetically relevant acoustic cues excised and presented in isolation as "chirps." Furthermore, Eimas and his colleagues (Eimas, Siqueland, Jusczyk & Vigorito, 1971; Eimas, in press; Cutting & Eimas, in press) have shown that young infants perceive speech stimuli in a categorical manner and "chirp" stimuli in a more continuous manner, a result functionally identical to that of adults.

For the most part, however, the discrimination results have been found with adult subjects using ABX stimulus triads. On such trials the subject judges which of the first two stimuli in the triad is identical to the third stimulus. Pisoni (1971, 1973a, 1973b) has rigorously examined the processes underlying ABX discriminations drawing on previous work by Fujisaki and Kawashima (1968, 1970). A perceptual model which stems from these papers includes both auditory and phonetic short-term memory components in order to explain the relative peaks and troughs in the discrimination function. The peaks appear to result from the engagement of the phonetic component of memory. The troughs, on the other hand, appear to reflect engagement of auditory memory. Performance here typically remains substantially above chance. The difference in performance between within-phoneme-boundary and across-boundary pairs may be interpreted as the difference in relative efficiency of the two types of memory. Thus, the shape of the discrimination function stems from both auditory and phonetic processes.

Right Ear Advantages

In a dichotic listening situation, where one stimulus is presented to the right ear and another to the left ear, the subject is often unable to report all items that were presented on a particular trial. Regardless of whether both items are sentences (Bever, 1971), meaningful words (Bartz, Satz, Fennel & Lally, 1967), or nonsense syllables (Shankweiler & Studdert-

Kennedy, 1967), the listener is generally able to report more accurately the information presented to the right ear than to the left. This highly replicable result is called the right ear advantage and has been explained, in part, in terms of the general processing capabilities of the cerebral hemispheres. Clinical evidence indicates that language processing occurs primarily in the left hemisphere for right handers (Milner, 1967; Geschwind, 1970), and that the crossed pathways from ear-to-cortex are more prominent during transmission than the uncrossed pathways (Pulletti & Celesia, 1970). Thus, in the dichotic situation, stimuli presented to the right ear have privileged access to the speech processor in the left hemisphere.

Certain aspects of the sound pattern of a speech utterance appear to require relatively more left hemisphere processing than others. For example, Shankweiler and Studdert-Kennedy (1967) found that stop consonants in consonant-vowel (CV) nonsense syllables yielded a large right ear advantage, while steady-state vowels yielded a considerably smaller, nonsignificant right ear advantage. Other classes of phonemes such as liquids (Day & Vigorito, 1973; Crystal & House, 1974; Cutting, in press) often yield results intermediate between stops and vowels. To account for these results, Cutting (in press) has suggested that, along with a generalized speech processor in the left hemisphere, there may be an auditory analyzer whose task it is to compute frequency transitions and other purely auditory aspects of the signal. Indeed, Darwin (1971) found a right ear advantage for fricatives with transitions in CV syllables and no ear advantage for fricatives without transitions. The data of Cutting (in press) suggest that this auditory device is engaged equally for speech and nonspeech signals. Thus, nonspeech signals conveying transient information may occasionally yield right ear advantages (Halperin, Nachshon & Carmon, 1973).

The Two Phenomena Compared

Although ABX discrimination tasks and dichotic listening tasks are quite different and appear to overload the perceptual system in a very different manner, there are impressive parallels between them. On empirical grounds, stop consonants yield sharply defined discrimination peaks and large right ear advantages, whereas vowels typically yield less extreme results. (Nonspeech sounds, of course, often yield relatively flat discrimination functions and *left* ear advantages.) On theoretical grounds, recent accounts of both phenomena have included both auditory and phonetic components. The present experiment was designed to compare the two phenomena in as direct a manner as possible. A specific attempt was made to devise a situation in which two tasks, ABX discrimination and dichotic recognition, would yield divergent results.

METHOD

Stimuli

Four arrays of seven stimuli each were synthesized on the Haskins Laboratories parallel resonance synthesizer. All stimuli consisted of two formants, were 300 msec in duration, and had the same fundamental frequency (112 Hz). Two arrays consisted of CV syllables which began with either [b] or [d]. One array was synthesized with the vowel [a] and the other with the vowel [æ]. The steady-state formant frequencies for [a] were 743 Hz for the first formant (F1) and 1232 Hz for F2, while the corresponding values for [æ] were 743 and 1620 Hz. Formant transitions were 50 msec in duration. The F1 transition was identical for both vowels, increasing linearly in frequency from a value of 437 Hz. Seven different F2 transitions were synthesized for each vowel: Their initial values were 616, 769, 921, 1075, 1232, 1386, and 1541 Hz for the [ba]-[da] array; and 769, 921, 1075, 1232, 1386, 1541, and 1695 Hz for the [bæ]-[dæ] array. All F2 transitions were linear.

Two other arrays of stimuli were synthesized. They were identical to the CV stimuli except that the F1 transition was inappropriate for any particular phoneme segment. Instead of increasing in frequency for 50 msec, it decreased from a value of 894 to 743 Hz for both vowels. The F2 transitions were identical to the seven used in the [ba]-[da] and [bæ]-[dæ] arrays. Since these stimuli resembled the CV stimuli along many dimensions, but did not have F1 transitions corresponding to specific consonants, they were designated C'V stimuli. Schematic spectrograms of a CV and a C'V stimuli are shown in Fig. 1, along with a display of the various possible F2 transitions.

In addition an eighth stimulus was added to each of the four arrays: [ga] and [gæ] with very rapidly decreasing F2 transitions, and two C'V stimuli corresponding to them.

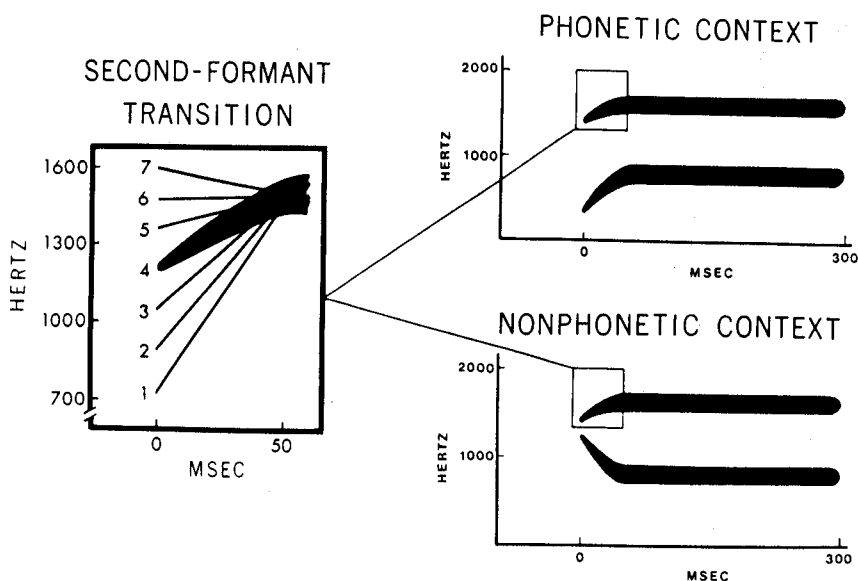


FIG. 1. Schematic spectrograms of stimuli with speech-relevant and speech-irrelevant first-formant transitions, and a display of their second-formant transition variations.

ABX Discrimination Tapes

Members of each AB comparison were selected by pairing each stimulus with the item two steps removed along the F2 continuum; that is, stimulus 1 was paired with stimulus 3, 2 with 4, 3 with 5, 4 with 6, and 5 with 7, yielding five possible pairs. For each pair there were four possible ABX comparisons: ABA, ABB, BAB, and BAA. Two different random sequences of 80 items were prepared: (5 ABX pairs) \times (4 ABX comparisons per pair) \times (4 arrays of stimuli). The members of each ABX triad were separated by 1 sec, with 4 sec between triads.

Dichotic Recognition Tapes

Stimuli 2 and 6 were selected from each of the four seven-item arrays used in the ABX task. Added to them was each of the eighth stimuli, yielding the CVs [ba, da, ga, bæ, dæ, gæ] and the C'Vs corresponding to them. Trials consisted of a dichotic pair, followed by one second of silence, followed by a diotic probe. Dichotic pairs of CV stimuli were matched so that items shared neither consonant nor vowel: thus, [ba] was paired with [dæ] or [gæ] and [bæ] with [da] or [ga]. A similar rule was applied to the C'V pairs in that items shared neither vowel nor F2 transition normally associated with a particular stop in that vowel context. On one quarter of the trials the diotic probe was the stimulus that had been presented to the left ear, on a second quarter it was the stimulus that had been presented to the right ear, and on half the trials it was a third stimulus not previously presented. Two different random sequences of 96 items were prepared: (6 pairings per stimulus class) \times (2 classes of stimuli—CV and C'V) \times (2 channel arrangements) \times (2 possible legitimate probes) + 48 similar trials in which the probe was not a member of the dichotic pair.

Subjects and Apparatus

Sixteen Yale University undergraduates participated in two tasks, diotic ABX discrimination and dichotic recognition. All subjects were right-handed native American English speakers with no history of hearing difficulty, and no previous experience in dichotic listening or in listening to synthetic speech. Audio tapes were played on an Ampex AG500 dual-track tape recorder, and signals were sent through a listening station to Grason-Stadler earphones (Model TDH39-300Z).

Procedure

Before the ABX task began, subjects were instructed to write down which stimulus, *A* or *B*, was identical to the third member of the triad on each trial. They listened to four practice trials to familiarize themselves with the task and the stimuli. Before the dichotic recognition task began, they were instructed to attend to one ear for a block of trials, and to write down *Y* for yes if they thought the probe had been presented to the attended ear, and *N* for no if it was not. Counterbalancing of the monitored ear was done within subjects, and counterbalancing of ear-to-channel assignments was done across subjects. Subjects listened to six practice trials before monitoring a given ear. Half of the subjects participated in the discrimination task before the dichotic recognition task, while the others participated in reverse order.

RESULTS

Discrimination

Consonant-vowel stimuli yielded results typical of those associated with categorical perception. Figure 2 (top) shows that the discrimina-

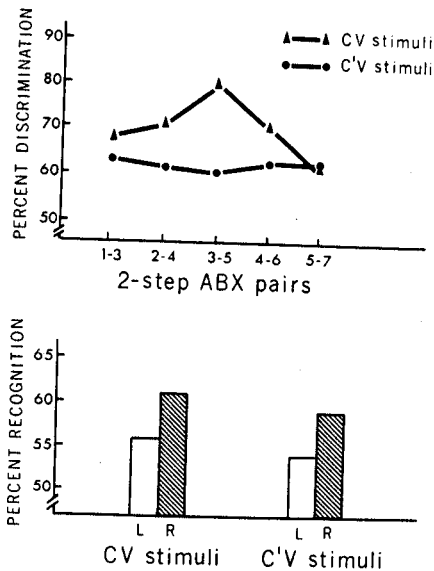


FIG. 2. The results of the ABX discrimination task (top) and of the dichotic recognition task (bottom).

bility of stimuli 3-5 pairs averaged 80% correct, much superior to all other CV pairs. Stimulus 3 was typically heard as [ba] or [bæ] and Stimulus 5 as [da] or [dæ] according to preliminary identification results. Thus these stimuli unambiguously belonged to different phoneme categories. The C'V stimuli did not yield any discrimination peak; all comparisons were within a few percentage points of 60% correct. In general subjects were better at discriminating CV stimuli than C'V stimuli ($F(1,15) = 27.9, p < .001$), but this superiority was largely a result of the interaction of the shapes of the two discrimination functions, as shown in Fig. 2. The interaction of the array of two-step comparisons with the stimulus classes, CV and C'V, was significant ($F(4,15) = 4.5, p < .025$). Thirteen of 16 subjects discriminated 3-5 CV pairs better than 3-5 C'V pairs, and no subject performed better on the C'V comparison. There was no significant difference between the CV and C'V pairs for any other two-step comparison. Furthermore, the discrimination functions shown in Fig. 2 were typical of all 16 subjects.

Dichotic Recognition

The CV and C'V stimuli yielded nearly identical right ear advantages, as shown at the bottom of Fig. 2. For the CV pairs, subjects were 61% correct in recognizing the attended stimulus in the right ear, while only

57% correct in recognizing that stimulus in the left ear. The corresponding scores for the C'V stimuli were 59 and 55%. Thus, both classes of stimuli yielded a 4% right ear advantage ($F(1,15) = 5.35, p < .05$). There was no significant difference in the two ear advantages, nor in their levels of overall performance. Neither ear advantage, however, was significant by itself.

DISCUSSION

The results of the present study suggest that the relative peakedness of discrimination functions normally associated with categorical perception and the relative magnitude of the right ear advantage in dichotic listening are *not* functionally parallel results. Furthermore, in certain situations the mechanisms which underlie them cannot be exactly the same. Before considering the ramifications of this finding, however, it is necessary to consider the stimuli, paradigms, and results in more depth.

The C'V stimuli are essentially unidentifiable in any way other than in a purely arbitrary fashion. This is partly due to the fact that they could never have been produced by a human vocal tract. On the other hand, CV stimuli, like all speech items, have the peculiar feature of naming themselves—that is, providing their own nonarbitrary label—precisely because they could be produced by any normal human vocal tract. This difference between the two classes of stimuli determined the nature of the dichotic listening task. Most dichotic tasks which employ speech stimuli are tasks in which subjects identify, in an oral or written form, the stimuli they heard. Since the C'V stimuli could not be reproduced by the subjects, in either oral or written forms, it was necessary to devise either a recognition task or an identification task which used arbitrary labels for the C'V items. The recognition task was selected so that ease of identification would not enter as a variable into the subjects response. Furthermore, the task was similar to that used by Kimura and Folb (1968) and Spellacy and Blumstein (1970). No special modification was necessary for the discrimination task.

The results of the dichotic recognition task showed that both CV and C'V stimuli yielded right ear advantages. Although significant when considered in conjunction, neither ear advantage was significant by itself. Ideally, one would wish both ear advantages significant to demonstrate that the left hemisphere was superior in processing each kind of stimulus. Nevertheless, the fact that the ear advantages were nearly identical in magnitude and that taken together they did yield a significant ear advantage is a compelling result. That the ear advantages are small is not unusual. Small ear differences appear to be a hallmark of recognition tasks (see again, the results of Kimura & Folb, 1968; Spellacy & Blumstein, 1970).

Perceptual Mechanisms

The mechanisms underlying the results of the discrimination task are quite transparent. The peak in the CV function appears to result from the engagement of the phonetic component of short-term memory; the absence of a peak in the C'V functions at any position appears to result from the absence of any opportunity for phonetic memory coding to occur. On the other hand, the above-chance performance for all five C'V comparisons and for the endpoint CV comparisons appears to reflect the relative strength of purely auditory memory.

The mechanisms underlying the results of the dichotic recognition task are less transparent. There are two candidates: a generalized speech processor and an auditory analyzer. Cutting (in press) has shown that the left hemisphere system appears to excel its counterpart not only in the phonetic processing of speech, but also in processing certain purely auditory aspects of the signal. The two mechanisms appear to be independent and additive. In the present study, however, it is not clear whether a phonetic mechanism, an auditory mechanism, or both account for the results. Consider these three possibilities in reverse order.

First, CV and C'V dichotic pairs may have yielded nearly identical results because they invoked both auditory and phonetic left hemisphere mechanisms to a similar extent. After all, both classes of stimuli contained formant transitions which would need to be analyzed, and both had vowel segments upon which basic phonetic decisions could be made. As a part of this explanation is the assumption that C'V stimuli are speech stimuli. This assumption is, I think, entirely valid. The C'V stimuli sound like speech syllables with a very garbled beginning. The "garbling" of the stimulus onsets results from inappropriate formant transitions. Nonetheless, backwards speech stimuli, such as those used by Kimura and Folb (1968), could be no less garbled and yet they too yield a right ear advantage. The major argument against this explanation is that the ear advantages in the present study are quite small, perhaps too small for there to be convincing evidence of two components underlying them. The data of Cutting (in press) show that CV and C'V stimuli yield much larger right ear advantages in a temporal-order judgment task.

The second possibility is that both classes of stimuli, CV and C'V, invoked only a phonetic decision-making mechanism in the left hemisphere. Since members of each dichotic pair differed in both transitions and vowels, subjects may have used only the vowel dimension of the stimuli as the basis of their judgments. This explanation, though tenable, seems unlikely, since Cutting (in press) found that, for both identification and temporal-order judgment tasks, dichotic speech stimuli differing in both transitions and vowels yielded larger right ear advantages than

pairs of different steady-state vowels. Such a result strongly implies that transitions are processed in this situation, and that they contribute to the ear advantage.

Third, and perhaps most likely, both classes of stimuli may have invoked only auditory processing mechanisms, and the results may reflect only the superiority of the left hemisphere system for this type of auditory perception and memory. The task, after all, did not require phonetic identifications. Speech processing per se may not have been involved at all. Day and her co-workers (Day & Cutting, 1970; Day, Cutting & Copeland, 1971; Wood, Goff & Day, 1971) have shown that the speech processor can be disengaged in a variety of dichotic and diotic tasks. Indeed, perhaps the reason that the ear advantages in the present study were comparatively small in relation to those of other studies (see Studdert-Kennedy & Shankweiler, 1970; Cutting, in press) is that only an auditory processor was engaged in the left hemisphere. Cutting (in press) has suggested that the increment in right ear advantage due to auditory processing is less than that due to phonetic coding, or speech processing in the traditional sense.

CONCLUSION

Evidence from many discrimination studies and dichotic listening studies have suggested that there is a functional parallel between the relative sharpness of discrimination curves and the relative magnitude of the right-ear advantage. The present study demonstrates that this is not always so. Speech discrimination functions appear to be a result of both auditory and phonetic processing, whereas a right ear advantage may result from the engagement of one or both types of processing mechanisms.

REFERENCES

- ABRAMSON, A. S., & LISKER, L. 1965. Voice onset time in stop consonants: Acoustic analysis and synthesis. In *Proceedings of the Fifth International Congress of Acoustics*, Liege.
- ABRAMSON, A. S., & LISKER, L. 1970. Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the Sixth International Congress of Phonetics Sciences*, Prague. Pp. 15-25.
- BARTZ, W. H., SATZ, P., FENNEL, E., & LALLY, J. R. 1967. Meaningfulness and laterality in dichotic listening. *Journal of Experimental Psychology* 73, 204-210.
- BEVER, T. G. 1971. The nature of cerebral dominance in speech behavior of the child and adult. In R. Huxley and E. Ingram (Eds.) *Language Acquisition: Models and Methods*. New York: Academic Press.
- CRYSTAL, T. H., & HOUSE, A. S. 1974. Effect of local signal level on differential performance in dichotic listening. *Journal of the Acoustical Society of America*, 55, 434(A).
- CUTTING, J. E. In press Two left-hemisphere mechanisms in speech perception. *Perception and Psychophysics*.
- CUTTING, J. E., & EIMAS, P. D. In press. Phonetic feature analyzers and the processing of

- speech in infants. In J. F. Kavanagh and J. E. Cutting, (Eds.) *The Role of Speech in Language*. Cambridge, MA: MIT Press.
- DARWIN, C. J. 1971. Ear differences in the recall of fricatives and vowels. *Quarterly Journal of Experimental Psychology*, **23**, 46-62.
- DAY, R. S., & CUTTING, J. E. 1970. Levels of processing in speech perception. Paper presented to the 10th Annual Meeting of the Psychonomic Society, San Antonio, November.
- DAY, R. S., CUTTING, J. E., & COPELAND, P. M. 1971. Perception of linguistic and nonlinguistic dimensions of dichotic stimuli. Paper presented at the 11th Annual Meeting of the Psychonomic Society, St. Louis, November 9th.
- DAY, R. S., & VIGORITO, J. M. 1973. A parallel between encodedness and the ear advantage: Evidence from a temporal-order judgment task. *Journal of the Acoustical Society of America*, **53**, 358(A).
- EIMAS, P. D. In press. Speech perception in early infancy. In L. B. Cohen, and P. Salapatek (Eds.) *Infant Perception*. New York: Academic Press.
- EIMAS, P. D., SIQUELAND, E. R., JUSCZYK, P., & VIGORITO, J. M. 1971. Speech perception in infants. *Science*, **171**, 303-306.
- FUJISAKI, H., & KAWASHIMA, T. 1968. The influence of various factors on the identification and discrimination of synthetic speech sounds. *Reports of the Sixth International Congress on Acoustics*, Tokyo, August 16th.
- FUJISAKI, H., & KAWASHIMA, T. 1970. Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute*, Vol. 29. Tokyo: Faculty of Engineering, University of Tokyo. Pp. 207-214.
- HALPERIN, Y., NACHSHON, I., & CARMON, A. 1973. Shift in ear superiority in dichotic listening to temporal pattern nonverbal stimuli. *Journal of the Acoustical Society of America*, **53**, 46-50.
- GESCHWIND, N. 1970. The organization of language and the brain. *Science*, **170**, 940-944.
- KIMURA, D., & FOLB, S. 1968. Neural processing of backwards-speech sounds. *Science*, **161**, 395-396.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. 1967. Perception of the Speech Code. *Psychological Review*, **74**, 431-461.
- LIBERMAN, A. M., HARRIS, K. S., HOFFMAN, H. S., & GRIFFITH, B. C. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, **54**, 358-368.
- LIBERMAN, A. M., HARRIS, K. S., KINNEY, J. A., & LANE, H. L. 1961. The discrimination of relative onset time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, **61**, 379-388.
- MATTINGLY, I. G., LIBERMAN, A. M., SYRDAL, A. K., & HALWES, T. 1971. Discrimination in speech and nonspeech modes. *Cognitive Psychology*, **2**, 131-157.
- MILNER, B. 1967. Brain mechanisms suggested by studies of temporal lobes. In C. H. Millikan and F. L. Darley (Eds.) *Brain Mechanisms Underlying Speech and Language*. New York: Grune and Stratton. Pp. 122-145.
- PISONI, D. B. 1971. On the nature of categorical perception of speech sounds. Ph. D. thesis, University of Michigan (Psycholinguistics); *Haskins Laboratories: Status Report on Speech Research* a Suppl.
- PISONI, D. B. 1973a. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, **13**, 253-260.
- PISONI, D. B. 1973b. The role of auditory short-term memory in vowel perception. *Haskins Laboratories: Status Report on Speech Research*, **SR-34**, 89-118.
- POLLACK, I. 1952. The information in auditory displays. *Journal of the Acoustical Society of America*, **24**, 745-749.

- POLLACK, I. 1953. The information in elementary auditory displays. II. *Journal of the Acoustical Society of America*, **25**, 765-769.
- PULLETTI, F., & CELESIA, G. G. 1970. Functional properties of the primary cortical auditory area in man. *Journal of Neurosurgery*, **32**, 244-247.
- SHANKWEILER, D., & STUDDERT-KENNEDY, M. 1967. Identification of consonants and vowels presented to left and right ears. *Quarterly Journal of Experimental Psychology*, **19**, 59-63.
- SPELLACY, F., & BLUMSTEIN, S. 1970. The influence of language set on ear preference in phoneme recognition. *Cortex*, **6**, 430-439.
- STUDDERT-KENNEDY, M., & SHANKWEILER, D. P. 1970. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, **48**, 579-594.
- WOOD, C. C., GOFF, W. R., & DAY, R. S. 1971. Auditory evoked potentials during speech perception, *Science*, **173**, 1248-1251.