Pisoni, D.B. (1977): "Identification and discrimination of the relative onset of two component tones: Implications for voicing perception in stops", JASA 61, 1352-1361.

Stevens, K.N. (1972): "The quantal theory of speech: Evidence from articulatory-acoustic data", in Human communication: A unified view, E.E. David, Jr. and P.B. Denes (eds.), New York: McGraw-Hill.

COMMENTS FROM THE PANELISTS

The symposium on the perception of speech and nonspeech began with a brief summary statement by each of the contributors. This was followed by a panel discussion dealing with several issues that came up during the presentations. Finally, a number of questions and comments from the general audience were presented, followed by further discussion by the members of the panel. The highlights of these discussions and interactions are summarized below in an attempt to capture the flavor of the general issues and problems that surfaced as a result of this symposium.

Dr. Ades began his presentation by summarizing his paper contributed to the symposium and offering several comments on the introductory remarks given earlier by Professor Pisoni. Dr. Ades reiterated several times in this presentation that he personally believed that speech perception was, in some sense, "unique" or "special" despite the weak evidence usually cited from identification and discrimination experiments. He argued that the differences in perception between speech and nonspeech signals or consonants and vowels could be accounted for by differences in the range or spacing of signals. Dr. Ades criticized the recent data presented by Professor Pisoni showing equivalent ranges for consonants and vowels on the grounds that these data were collected in an identification rather than a discrimination paradigm. Most of Dr. Ades' specific remarks were directed, however, at narrow experimental questions, particularly the use of high uncertainty discrimination paradigms which provide relatively low estimates of discriminability.

Dr. Divenyi argued for the operation of two stages of processing in auditory perception regardless of whether the signals are complex auditory patterns or speech signals. According to Dr. Divenyi, speech is simply one class of complex signals with which the listener has had extensive experience and familiarity. Dr.

Divenyi described his two-stage model of auditory processing. The first stage, the auditory stage, involves the sensory analysis and coding of signals by the peripheral auditory system. The representation of signals at this stage is something like a neurogram reflecting the frequency selectivity of the auditory system. The second stage, the temporal stage in Dr. Divenyi's model, involves the analysis and coding of temporal information or patterns in both speech and nonspeech signals. Dr. Divenyi argued that the differences in perception between speech and nonspeech signals were due to differences in listening strategies brought about by learning and experience with speech and other sounds. Thus, in listening to speech several different strategies are available to the listener for centering or positioning the listening band differently. Dr. Divenyi concluded that there were no structural differences in perception between the so-called "speech mode" and "nonspeech modes" of processing. The distinctiveness of speech arises, according to Dr. Divenyi, from mere exposure and familiarity with speech and not because of any specialized processing by the auditory system.

Professor Dorman summarized his recent research which was carried out in collaboration with Drs. Bailey and Summerfield. This work was concerned with the perception of speech and nonspeech stimuli differing in the cues to place of articulation. Professor Dorman stated that his interest in these comparisons grew out of several questions surrounding whether infants can perceive speech signals as speech rather than simply complex nonspeech patterns. The methodology employed in these studies using adult subjects involved comparisons dissociating the location of the "phonetic" boundary from the location of the "acoustic" boundary. The results of these tests showed differences in the loci of the boundaries depending on whether the nonspeech stimuli were heard as speech or nonspeech. Accordingly, Professor Dorman argued for the operation of two modes of processing nonspeech signals having speech-like properties. Furthermore, Professor Dorman implied that the dissociation of these two modes could be assessed by looking at differences in the location of category boundaries when the same stimuli are perceived as speech or nonspeech.

Professor Massaro departed from his symposium contribution by focusing on his general model of auditory information processing which postulates both structures for storage of information in memory and processes for carrying out various operations on this information. According to Professor Massaro's model, the earliest stage of processing involves acoustic feature analysis and is similar for speech and nonspeech signals alike. Processing here is not influenced by higher-order knowledge or context from long-term memory. Professor Massaro claimed that his general model could account for the differences observed in perception between speech and nonspeech without assuming the existence or operation of a specialized "speech mode" of processing. According to Professor Massaro, a listener's higher-order knowledge and his experience with speech affects the way acoustic features are treated and integrated at what he calls the primary stage of recognition in his model. Thus, a two stage model is also assumed to be necessary for perception of speech stimuli although the same two processes may be employed with other nonspeech stimuli.

Professor Liberman's remarks on duplex perception were summarized very briefly by Professor Studdert-Kennedy.[1] Using a variation of the so-called "Rand Effect", Professor Liberman has shown that listeners can simultaneously perceive a phonetic event (i.e., a CV syllable) and an auditory event (i.e., a chirp). Professor Liberman has argued that these results imply that both auditory and phonetic processes are carried out together simultaneously in parallel and that a distinct phonetic subsystem exists for processing speech signals, a subsystem which is separate from processes used to perceive other auditory signals.

Dr. Summerfield summarized his symposium paper with Dr. Bailey by emphasizing that the information for phonetic perception must be found in the acoustic signal itself which reflects the consequences of articulation of speech. Dr. Summerfield suggested that the phonetic information in the signal could be properly characterized by detailed examination of the articulatory control that gives rise to acoustic patterning in speech production and by a detailed examination of how the distinctiveness of this articulatory patterning is enhanced by auditory processing of speech signals.

---

1) Professor Liberman was not present at the congress.

Dr. Summerfield emphasized that this research strategy would be possible without having to assume any need for articulatory mediation in speech perception.

DISCUSSION

Following the individual summary statements, there was a general discussion among the panel members which was then opened up to the audience for additional questions and comments. Several broad and narrow issues appeared to emerge from the symposium papers and summary presentations as well as from the preliminary discussions that the panel members held before the symposium began.

Professor Studdert-Kennedy summarized these issues briefly before beginning the panel discussion. The first, and perhaps most general issue, concerned comparisons made in perception between speech and nonspeech signals. Specifically, it appeared that everyone agreed more or less that speech perception is in some sense special although not everyone agreed on precisely in what way it is special. Thus, the question of whether speech is a special process is one that still remains and apparently is one that continues to occupy the attention of numerous investigators working in speech perception even today.

Closely associated with the speech-is-special issue is a set of somewhat more narrowly defined experimental issues related to how one would be able to demonstrate clearly what the presumed special properties of speech are. That is, some concern was expressed among several members of the panel with the currently available methods and research paradigms used in speech perception research, particularly the use of discrimination procedures to assess differences between speech and nonspeech signals. During Dr. Ades' summary statement and later during the panel discussion, he repeated his dissatisfaction and skepticism with the traditional methods of comparing identification and discrimination of speech and nonspeech and consonants and vowels.

Another, somewhat broader issue that emerged from these discussions concerned the question raised by Summerfield and Bailey in their paper of whether there are, in fact, "characteristic" acoustic properties of speech signals that result directly from articulation and whether these properties are distinct from the properties of nonspeech signals. This particular issue highlights

the clear separation of views that emerged at the symposium by Divenyi and Massaro, for example, who suppose instead that there really are no distinctively different or unique acoustic correlates of speech sounds that separate them from the class of nonspeech signals in the listener's environment. According to both of these investigators, differential processing by a human observer is not required or determined by properties of the signal itself but rather by experience, training, context and higher-order knowledge. The early stages of perceptual processing are therefore the same for speech and nonspeech signals alike.

Finally, the issues surrounding the development of speech perception, particularly the recent findings with young prelinguistic infants, were also cited as a potentially important topic for further discussion. Professor Studdert-Kennedy wondered to what extent it is reasonable to suppose that an organism such as a young infant who does not "know" a language can respond to an acoustic signal as though it were conveying language--that is as though the signal were speech.

The panel discussion began with several additional remarks about the use of discrimination paradigms in speech perception research. Dr. Ades suggested that he could see little use for additional discrimination experiments in the future. Dr. Divenyi repeated several of his earlier comments on the need for two stages of processing in auditory perception to deal with all the relevant empirical phenomena in the literature. Moreover, he restated his claims again about the role of perceptual strategies in determining what a listener focuses his attention on in speech perception.

In responding to Dr. Ades' remarks about discrimination testing, Dr. Massaro felt that discrimination experiments should proceed in parallel with categorization experiments to illuminate the nature of processing speech and nonspeech. Moreover, Dr. Massaro summarized the results of recent experiments that manipulated several acoustic cues at the same time in order to explore how listeners integrate or combine information in complex multidimensional signals.

Professor Studdert-Kennedy suggested that the discussion seemed to point toward general agreement about the need for levels and stages of processing in perception, particularly speech perception. Professor Studdert-Kennedy also noted at this time that

one of the major reasons for postulating two levels in speech perception was the earlier work of Fujisaki suggesting the possibility that two kinds of auditory memory or coding were operating in categorical perception experiments.

The discussion then turned to the issue of how speech is distinguished acoustically from nonspeech signals. Dr. Summerfield pointed out that the contrast between speech and nonspeech might be more profitably examined in terms of different styles of processing--one appropriate for real world "events" (i.e., speech signals generated by a human vocal tract) and the other being appropriate for a relatively unnatural mode of processing where the object of interest is a "nonevent". Dr. Summerfield also suggested that there are reliable acoustic markers in the speech signal that inform a listener that the signal is speech rather than nonspeech. For example, the posture of the vocal apparatus during speech production is unique to speaking. There are both short- and long-term changes in variations in intensity and rise-time which are indicators of speech that may act as "trigger-features" to engage a speech mode of processing.

Professor Dorman then suggested a possible experimental paradigm to compare speech and nonspeech more directly by examination of "trading relations" between different types of acoustic cues in both contexts. If the trading relations differ between the two contexts, speech and nonspeech, then one could argue for distinctly different modes of processing for speech vs. nonspeech signals.

After the members of the symposium panel completed their discussion of these issues, the moderator opened the discussion to members of the general audience in attendance. Professor Stevens raised the issue again of what markers or characteristics distinguish speech from nonspeech signals. Professor Stevens suggested that it is not necessary to make reference to articulation in speech perception because all speech signals have three or four criterial acoustic properties that set them apart from all nonspeech signals. The first property involves the rate of amplitude variations over time. A basic property of speech is that it has a syllabic structure creating amplitude fluctuations between consonants and vowels. A second property of speech is shown in the spectra of speech signals. If the spectra of speech are sampled

at any point in time, the resulting analysis will display characteristic peaks and valleys. A third property of speech is the fact that these spectra change with time. That is, there are well-defined acoustic correlates to the changing articulatory gestures in speech production. The spectra of speech can also change rapidly or slowly over time. Professor Stevens suggested that one might speculate that speech signals are acoustic signals that the auditory system "likes" because it is easy to extract properties from signals of this kind.

Dr. Waterson then raised the question of the usefulness of the present kinds of experiments carried out on speech vs. nonspeech. She argued that almost all of the research has used European-based languages with either European or American subjects and the tests employ language-specific features such as VOT. That is, the contrasts are presented in the language of the subjects. She wondered what sorts of results would be obtained if the subjects were presented with sounds from more exotic languages.

Professor Kuhl questioned the claim made earlier in the introduction by Professor Pisoni concerning the chinchilla's apparent inability to discriminate some of the cues to place of articulation in stop consonants. Professor Kuhl pointed out that the chinchilla's failure to discriminate /d/ from /g/ is due to a basic sensory limitation involving the length of their basilar membrane and not any inherent perceptual or cognitive limitation. Professor Kuhl also took issue with another remark of Professor Pisoni's in his introduction concerning the usefulness of certain kinds of comparative designs involving animal subjects and what these results could provide for understanding human language. Professor Kuhl stated that very pertinent information about "processing" species-specific acoustic signals may be provided by looking at animal models, particularly animals in which "vocal learning" is a salient characteristic such as the acquisition of bird song or coos by certain species of monkeys. Unfortunately, Professor Kuhl did not provide any further details about precisely what kinds of information would be obtained from these animal studies nor how the perceptual processing by these animals could be compared to the analyses carried out by humans.

Professor Kuhl also touched on the issue of a predisposition for processing certain salient acoustic attributes by human infants.

Such salient properties might serve to "focus" the infants' attention on certain aspects of the speech signal at a very early age. Moreover, Professor Kuhl repeated the suggestion, made by several others, that there is the strong possibility that the selection of speech sounds in language was guided, in some sense, by evolutionary constraints on the close match between both speech production and speech perception.

Dr. Klatt pointed out an important methodological difference in the results presented in the introduction by Professor Pisoni and the findings obtained by Professor Dorman on sine-wave analogs of CV syllables. Professor Pisoni showed well-defined labelling data for three categories of FMs corresponding to rising, level and falling, whereas Professor Dorman only reported two categories corresponding to rising and falling. Dr. Klatt suggested that this is a potentially important issue worthy of further study with fine-grained discrimination techniques which reduce the use of category labels. Dr. Klatt raised the question again of whether speech signals are somehow structured along "natural" auditory or psychophysical distinctions and/or constraints from the way speech is produced by the articulatory system.

Professor Fourcin offered an additional property, variations in fundamental frequency, that should be added to Professor Stevens' list for distinguishing between speech and nonspeech signals. Professor Fourcin also emphasized the need to look at pattern learning as the abstraction of invariants in complex stimuli, a topic that received little, if any, attention by members of the symposium.

Following the questions and comments from the audience, each of the panel members provided several additional final remarks elaborating on the statements they made earlier or commenting on some specific item raised in the general discussion. For the most part, however, the symposium on speech vs. nonspeech served to solidify a general sense of agreement among various investigators as to the value of comparisons in perception between speech and nonspeech signals. The issue of whether speech is special was discussed extensively throughout the symposium and led to a consensus that such a broad distinction is no longer meaningful, although nearly everyone believed that speech perception was somehow special or unique in its own way. A central issue that emerged

from this symposium was a concern with identifying the distinctive acoustic properties of speech signals that set them apart from other nonspeech signals in the listener's environment. There was also some attention devoted to questions of perceptual development in infants and issues surrounding perceptual predispositions for processing speech signals. Finally, there was a continued lively debate and interaction throughout the symposium on research methodology, particularly the use of discrimination paradigms in speech perception and the relevance of these sorts of data to categorization and recognition of phonemes in speech.