PERCEPTION OF SPEECH VERSUS NON-SPEECH
Summary of Moderator's Introduction

David B. Pisoni, Speech Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA. 02139, U.S.A.

Historically, the study of speech perception may be said to differ in a number of ways from the study of other aspects of auditory perception. First, the signals typically used to study the functioning of the auditory system were simple, discrete and typically differed along only a single dimension. In contrast, speech signals involve very complex spectral and temporal relations. Secondly, most of the research dealing with auditory psychophysics that has accumulated over the last thirty years has been concerned with the discriminative capacities of the sensory transducer and the functioning of the peripheral auditory mechanism. In the case of speech perception, however, the relevant mechanisms are centrally located and intimately related to more general cognitive processes that involve the encoding, storage and retrieval of information in memory. Moreover, experiments in auditory psychophysics have typically focused on experimental tasks and paradigms that involve discrimination rather than identification or recognition, processes thought to be most relevant to speech perception. Thus, it is generally believed that a good deal of what has been learned from research in auditory psychophysics and general auditory perception is only marginally relevant to the study of speech perception and to an understanding of the underlying perceptual mechanisms.

Despite these obvious differences, investigators have, nevertheless, been quite interested in the differences in perception between speech and nonspeech signals. That such differences might exist was first suggested by the report on the earliest findings of categorical discrimination of speech by Liberman et al. (1957). And it was with this general goal in mind that the first so-called "nonspeech control" experiment was carried out by Liberman et al. (1961) in order to determine the basis for the apparent distinctiveness of speech sounds.

Numerous speech-nonspeech comparisons have been carried out over the years since these early studies, including several of the contributions to the present symposium. For the most part, these experiments have revealed quite similar results. Except until quite recently, performance with nonspeech control signals failed to show the same discrimination functions that were observed with the parallel set of speech signals (Cutting and Rosner, 1974; Miller et al., 1976; Pisoni, 1977). In addition, the nonspeech signals were typically responded to by subjects at levels approximating chance performance. Such differences in perception between speech and nonspeech signals have been assumed to reflect basically different modes of perception-- a "speech mode" and an "auditory mode". Despite some attempts to explain away this dichotomy, additional evidence continues to accumulate as suggested by several of the new findings summarized in the papers included in this section.

There have been, however, a number of problems involved in drawing comparisons between speech and nonspeech signals that have raised several questions about the interpretation of the results obtained in these earlier studies. First, there is the question of whether the same psychophysical properties found in the speech stimuli were indeed preserved in the nonspeech control condition. Such a criticism seems quite appropriate for the original /do/--/to/ nonspeech control stimuli which were simply inverted spectrograms as well as the well-known "chirp" and "bleat" control stimuli of Mattingly et al. (1971) that were created by removing the formant transitions and steady-states from speech context and then presenting them in isolation to subjects for discrimination. Such manipulations while nominally preserving the speech cue obviously result in a marked change in the spectral context of the signal which no doubt affects the detection and discrimination of the original formant transitions. Such criticisms have been taken into account in the more recent experiments comparing speech and nonspeech signals as summarized by Dr. Dorman and Dr. Liberman in which the stimulus conditions remain identical across different experimental manipulations. However, several additional problems still remain in making comparisons between speech and nonspeech signals. For example, subjects in these experiments rarely if ever receive any

experience or practice with the nonspeech control signals. With complex multidimensional signals it may be quite difficult for subjects to attend to the relevant attributes of the signal that distinguish it from other signals presented in the experiment. A subject's performance with these nonspeech signals may therefore be no better than chance if he/she is not attending selectively to the same specific criterial attributes that distinguish the speech stimuli. Indeed, not knowing what to listen for may force a subject to "listen" for an irrelevant or misleading property of the signal itself. Since almost all of the nonspeech experiments conducted in the past were carried out without the use of feedback to subjects, a subject may simply focus on one aspect of the stimulus on one trial and an entirely different aspect of the stimulus on the next trial.

Setting aside some of these criticisms, the question still remains whether drawing comparisons in perception between speech and nonspeech signals will yield some meaningful insights into the perceptual mechanisms deployed in processing speech. In recent years, the use of cross-language, developmental and comparative designs in speech perception research has proven to be quite useful in this regard as a way of separating out the various roles that genetic predispositions and experiential factors play in perception. For example, while it is cited with increasing frequency that chinchillas have been shown to categorize synthetic stimuli differing in VOT in a manner quite similar to human adults, little if anything is ever mentioned about the chinchilla's failure to carry out the same task with stimuli differing in the cues to place of articulation in stops, a discrimination that even young prelinguistic infants have been shown to be capable of making. Such comparative studies are therefore useful in speech perception research to the extent that they can specify the absolute lower-limits on the sensory or psychophysical processes inherent in discriminating properties of the stimuli themselves. However, they are incapable, in principle, of providing any further information about how these signals might be "interpreted" or coded within the context of the experience and history of the organism.

Cross-language and developmental designs have also been quite useful in providing new information about the role of

early experience in perceptual development and the manner in which selective modification or tuning of the perceptual system takes place. Although the linguistic experience and background of an observer was once thought to strongly control his/her discriminative capacities in a speech perception experiment, recent findings strongly suggest that the perceptual system has a good deal of plasticity for retuning and realignment even into adulthood. The extent to which control over the productive abilities remains plastic is still a topic to be explored in future research.

To what extent is it then useful to argue for the existence of different modes of perception for speech and nonspeech signals? Some investigators such as Dr. Ades and even Dr. Massaro would like to simply explain away the distinctions drawn from earlier work on the grounds of parsimony and generality. But this is a curious position to maintain as it is commonly recognized, not only in speech perception research but in other areas of perceptual psychology, that stimuli may receive differential amounts of processing or attention by the subject, that subjects may organize the interpretation of the sensory information differently under different conditions and that the sensory trace of the initial input signal may show only a faint resemblance to its final representation resulting from encoding and storage in memory. It is hard to deny that a speech signal elicits a characteristic mode of response in a human subject-- a response that is not simply the consequence of an acoustic waveform leaving a meaningless sensory trace in the auditory periphery. Such observations suggest to me that, just as in the case of "species-typical responding" observed in the behavior of numerous other organisms, the existence of a speech mode of perception is a way of capturing certain aspects of the way human observers typically respond to speech signals that are familiar to them. Such a conceptualization does not, at least in my view, commit one to the view that human listeners cannot respond to speech in other ways more closely correlated with the sensory or psychophysical attributes of the signals themselves. To explain away the speech mode, however, is to deny the fact that a certain subset of possible acoustic signals generated by the human vocal tract are used in a distinctive and quite systematic way by both

talkers and listeners to communicate by spoken language, a species-typical behavior that is restricted, as far as I know, to homo sapiens. Past experiments comparing the perception of speech and nonspeech signals have been quite useful in characterizing how the phonological systems of natural languages have, in some sense, made use of the general properties of sensory systems in selecting out the inventory of phonetic features and their acoustic correlates. The relatively small number of distinctive features and their acoustic attributes observed across a wide variety of diverse languages suggests that the distinctions between speech and nonspeech signals still remain fundamental ones setting apart research on speech perception from the study of auditory psychophysics and the field of auditory perception more generally.

References

Cutting, J.E. and Rosner, B.S. (1974) "Categories and boundaries in speech and music", Perc. Psych. 16, 564-570.

Liberman, A.M., Harris, K.S., Hoffman, H.S. and Griffith, B.C. (1957) "The discrimination of speech sounds within and across phoneme boundaries", J. Exp. Psych. 54, 358-368.

Liberman, A.M., Harris, K.S., Kinney, J.A. and Lane, H.L. (1961) "The discrimination of relative onset time of the components of certain speech and non-speech patterns", J. Exp. Psych. 61, 379-388.

Mattingly, I.G., Liberman, A.M., Syrdal, A.K. and Halwes,s, T.G. (1971) "Discrimination in speech and non-speech modes", Cogn. Psch. 2, 131-157.

Miller, J.D., J.D., Wier, C.C., Pastore, R., Kelly, W.J. and Dooling R.J. (1976) "Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception", JASA 60, 410-417.

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.