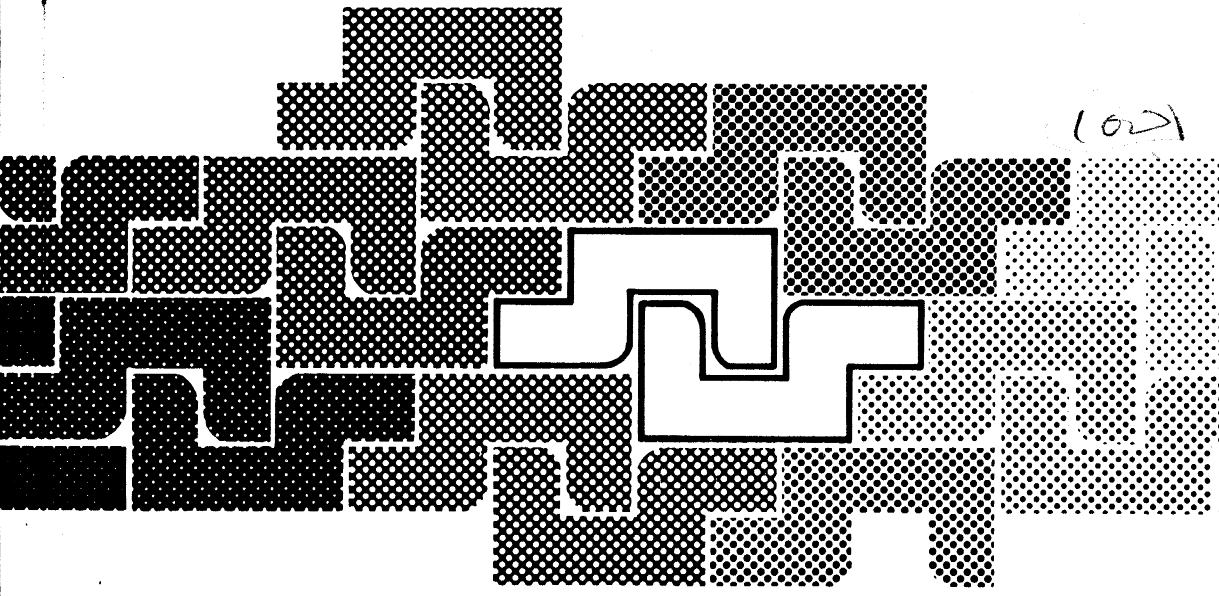

The Psychophysics of Speech Perception

Edited by

M.E.H. Schouten

NATO ASI Series

Series D: Behavioural and Social Sciences - No. 39

THE ROLE OF PSYCHOPHYSICS IN UNDERSTANDING SPEECH PERCEPTION*

Bruno H. Repp

Haskins Laboratories,
270 Crown Street, New Haven, CT 06511, USA

0. INTRODUCTION

The purpose of this workshop is to discuss the psychophysics of speech perception. The program includes a variety of topics that presumably fall under this heading and that demonstrate that the psychophysics of speech perception is alive and well. Yet it is not really obvious what the psychophysics of speech perception is, what its goals and limitations are, and whether it is indeed a circumscribed area of investigation. It seems useful, therefore, to pose these basic questions explicitly and to include them in our discussions along with the many specific issues addressed by our research. The purpose of my paper is to stimulate such discussion by presenting a particular, possibly controversial, view of speech perception, psychophysics, and the relation between the two.

My presentation has five parts. First, I will attempt to define the psychophysics of speech perception and to discuss some of its assumptions and limitations. Then, turning to the second half of my title, I will consider briefly what it might mean to "understand" speech perception. Next, I will sketch a general view of phonetic perception and follow this with a discussion of what I believe to be the major research questions from that perspective. Finally, I will suggest a relatively novel application of psychophysics in the research enterprise I have envisioned.

1. WHAT IS THE PSYCHOPHYSICS OF SPEECH PERCEPTION?

I am starting with the assumption that there is indeed a psychophysics of speech perception--a particular area of scientific inquiry that the title of this workshop is intended to refer to. If so, what distinguishes the psychophysics of speech perception from the investigation of speech perception in general?

Psychophysics, as traditionally defined, is the science of describing the relationship between objective (physical) and subjective (psychological) dimensions. In a typical experiment, physical characteristics of a series of stimuli are measured or manipulated, and the subjects' judgments are obtained on an explicit or derived

---

*Preparation of this paper was supported by NICHD Grant HD-01994 to Haskins Laboratories. I am grateful to Al Bregman, Bob Crowder, Jim Flege, Ignatius Mattingly, Robert Remez, and Michael Studdert-Kennedy for helpful comments on an earlier draft.

numerical scale. The resulting stimulus-response relationship is often described in the form of a function, such as Weber's law or Stevens' exponential curves. However, there are many other ways of describing stimulus-response relationships, and it would be unwise to exclude any particular descriptions from the domain of psychophysics. Since virtually all speech perception research involves eliciting subjects' responses to stimuli that have been manipulated in some way, it seems to me that, at first blush, the psychophysics of speech perception is the only kind of research on speech perception that exists, especially if we exclude psycholinguistic topics such as word recognition and sentence comprehension, which concern the perception of meaning.

Is the title of this workshop then a tautology? Perhaps not. In fact, the term "psychophysics" is not commonly applied to all of the research on speech perception. Therefore, it has certain connotations that derive from the kinds of experiments it is explicitly associated with. That is, even though the boundaries of psychophysics are not clearly defined and may include a large variety of topics and methods, those researchers who consider themselves psychophysicists represent certain typical theoretical attitudes and preferences. Thus, psychophysics may be considered a particular approach to the study of speech perception that, without necessarily being programmatic, characterizes a fair amount of work in the field. I presume that, in choosing the title for this workshop, the organizers wished to highlight this approach, which I will now attempt to characterize.

1.1. FOCUS ON THE AUDITORY MODALITY

One attitude I associate with a pschophysical approach to speech perception is a preoccupation with psychoacoustics. Indeed, all presentations at this workshop are concerned with aspects of auditory speech perception. This is not to say that research on speech perception via the visual and tactile senses is not often psychophysical in character; in fact, much of it is, and several participants in this workshop have made important contributions to it. Nevertheless, this research has often been the province of specialists outside the mainstream of speech perception research. One consequence of this is that many speech perception researchers place special emphasis on auditory processes and thereby miss the more general insights to be gained from a multimodal approach.

Tactile speech perception, to be sure, is uncommon and requires special transduction devices; moreover, it is not clear whether tactile information feeds directly into the speech perception system the way auditory and visual information does (except for the Tadoma method, where articulation is felt directly). Visual speech perception, by contrast, is extremely common, especially in conjunction with listening. The extent to which auditory and visual information is integrated was strikingly demonstrated by McGurk and MacDonald (1976), who presented conflicting information in the two modalities and found that visual information may override the auditory information without the perceiver's awareness. In such instances, subjects believed they heard what in fact they saw. More often, the conscious percept represents a compromise between the inputs from the two modalities (Massaro & Cohen, 1983a; Summerfield, in press). It appears, therefore, that speech information from the two sensory modalities converges upon a common mental representation. As Summerfield (1979) and others have argued, the information seems to be represented internally in a common metric that is amodal in nature.

If this kind of argument is accepted, it follows that not too much weight should be attached to descriptions of speech information that are tied to one modality. Rather, the basis for speech perception must be sought in information that is modality-independent and can be described in a common vocabulary. Such a vocabulary is provided by articulatory kinematics and/or by the dynamic parameters that underlie articulatory processes. To be sure, articulations taking place in the back of the vocal tract are transmitted exclusively by acoustic means, whereas movements of lips and jaw are prominent in the optic signal. This partial dissociation should not detract us from the fact, however, that in each case the information is about articulatory position and motion or, more abstractly, about the changing area function of the speaker's oral cavity.

Alternatively, it might be assumed that cues from different modalities are integrated in the process of categorical decision making, without recourse to a common metric (Massaro & Cohen, 1983a; Summerfield, in press). However, the question then arises: What motivates the integration in the first place? If the internal representations of stimuli are modality-specific, they can be related only through some form of association, either innate or acquired. In Massaro and Cohen's model, the associations reside in attribute lists which constitute phonetic category prototypes. Although this model seems to account well for audiovisual syllable perception, it seems less able to handle the intersensory integration of continuous dimensions such as speaking rate (Green & Miller, 1985) or prelinguistic infants' ability to recognize auditory-visual or visual-proprioceptive correspondences (Kuhl & Meltzoff, 1982; Meltzoff & Moore, 1985). A description of the stimulus information in articulatory terms eliminates the need to hypothesize independent mental representations of modality-specific correlates of articulation (see Yates, 1985), and it emphasizes the fact that the relation between visual and auditory manifestations of speech is nonarbitrary and possibly innately specified.

While it is generally taken for granted that we see the moving articulators when we look at them, not abstract optic patterns, there has been some reluctance in the field to accept the analogous proposition (Gibson, 1966; Neisser, 1976; Studdert-Kennedy, 1985) that, when we listen to speech, we hear the moving articulators and not the auditory patterns that constitute the proximal stimulus. Instead, researchers have been intensely preoccupied with acoustic variables such as formant transitions, delayed voicing onset, rise time, and so forth, as if the corresponding auditory percepts were the primary objects of speech perception. Whether they are is open to question, however (see, e.g., Liberman & Mattingly, 1985). Their prominent role in speech research may in large part be due to traditional techniques of acoustic analysis and synthesis, rather than to any compelling theoretical considerations. Many issues in the psychoacoustics of speech perception might never have been considered, had methods of articulatory analysis and synthesis preceded spectrographic and formant-based methods. As it is, we need to ponder whether these psychoacoustic issues are really pertinent to speech perception, or whether they merely have been forced upon us by the instruments we have had available. In other words, if we had only articulatory synthesizers as well as devices that extract area functions from the acoustic (and/or optic) signal, what would be the theoretical status of phenomena such as backward masking, adaptation, contrast, spectral integration, etc., in speech perception research? How much would we lose if we talked only about articulation and not about acoustics at all?

1.2 FOCUS ON METHODOLOGY

A second tendency that may reasonably be associated with a psychophysical approach is a focus on methodology. Certainly in classical psychophysics the methods by which stimulus-response mappings are obtained have been of overriding concern. There are many examples of a similar concern in speech perception research. Many experiments have compared performance in different discrimination paradigms, such as AX, 4IAX, ABX, fixed versus roving standard, etc. (e.g., Pisoni & Lazarus, 1974; MacKain et al., 1981; Rosner, 1984; Macmillan et al., this volume), and even in the many studies using only a single method its choice has usually been a matter of concern. Other studies have compared different identification tasks, such as binary classification, numerical rating scales, absolute identification, and perceptual distance scaling (e.g., Ganong & Zatorre, 1980; Massaro & Cohen, 1983b; Vinegrad, 1972). In fact, it may be argued that most of categorical perception research, as well as much research on selective adaptation, contrast, auditory memory, etc., has been exercises in methodology. To be sure, the variations in methods have usually served to test some reasonable models or hypotheses, and I do not mean to imply that this research has been worthless. Nevertheless, the questions asked in such experiments often are somewhat removed from the original phenomena that stimulated the research; in other words, they have become methodological variations on a common theme, and sometimes variations themselves have become the themes for further variations.

Take categorical perception. The category boundary effect (Wood, 1976)--the well-known finding that discrimination performance is higher across a phonetic category boundary than within categories--is important because it tells us that the acoustic structure of speech is not very transparent to the typical listener, who habitually focuses only on linguistically significant information. Numerous studies have shown that the strength of the effect varies with methodological factors such as discrimination paradigm, interstimulus interval, training, instructions, language experience, types of stimuli, etc. (see review by Repp, 1984). The large majority of these studies has been concerned with subjects' ability to discriminate small acoustic differences among speech stimuli. This ability, not surprisingly, can be enhanced by training, reduction of stimulus uncertainty, short interstimulus intervals, etc. The studies that have shown this are prime examples of the psychophysics of speech perception, and they include many an elegant piece of experimentation. However, the important aspect of categorical perception that seems directly relevant to speech communication is not subjects' apparent inability to dicriminate linguistically irrelevant differences along certain stimulus continua but rather their attention to linguistically distinctive information in the speech signal. To be sure, statements have been made in the literature (Studdert-Kennedy et al., 1970;

Liberman & Mattingly, 1985) to the effect that human listeners simply cannot perceive certain auditory properties of speech sounds, and this has, of course, been grist for the psychophysical mill. Apart from dismissing such extreme claims, however, little has been learned from all these studies about speech perception beyond the truism that perception within categories is not categorical. Rather, they have revealed some things about auditory discrimination and the methodological variables affecting it. Equivalent information could have been obtained by using nonspeech stimuli, and indeed one of the aims of psychophysical methodology (though this is rarely acknowledged) is to enable listeners to perceive speech as if it were a collection of arbitrary sounds. This leads me to another, related bias I associate with the psychophysics of speech perception.

1.3 FOCUS ON THE SOUNDS OF SPEECH

One possible definition of the psychophysics of speech perception is that it is the study of the perception of the sounds of speech. Unfortunately, the term "speech sounds" has often been used indiscriminately to denote both linguistically significant categories and acoustic components of the speech signal (and/or the auditory impressions associated with them). A clear distinction needs to be made between the auditory/acoustic and linguistic/articulatory domains, however (cf. Repp, 1981); the term "speech sounds" is appropriate for the former, whereas "phonetic categories" (or "phonemes") is appropriate for the latter. With this distinction in mind, my claim is that psychophysics is concerned, for the most part, with speech sound perception rather than with phoneme perception. It seems likely, however, that, except in very special circumstances, the sounds of speech as such do not play an important role in speech communication (see also Linell, 1982; Liberman & Mattingly, 1985; Traunmüller, this volume). Rather, I presume it is the more abstract, articulatory information that is used by listeners to decode the linguistic message. In fact, the only context in which the auditory qualities of speech segments may have a communicative function is in poetry, where an (unconscious) apprehension of the segmental sound pattern may enhance connotative and aesthetic qualities of the text (Fónagy, 1961; Hrushovski, 1980). Paradoxically, it seems that, so far, poetry has not attracted the attention of psychophysicists. (See, however, Marks, 1978.)

Why should one be interested in perceptual qualities that do not serve any important function in speech communication? There could be many valid reasons, such as questions about the auditory processing of complex sounds, the consequences of hearing impairment, skills of analytic perception, etc.--all topics worthy of scientific investigation. Nevertheless, these topics may be largely irrelevant to the perception of phonetic structure, and their study may therefore not contribute to our understanding of speech perception. To the hard-core psychophysicist, speech is primarily an acoustic signal of unusual complexity, which presents a challenge to the auditory system and to the experimenter's ingenuity. However, since this acoustic complexity is precisely what the speech perception system is equipped to handle, the speech signal actually has a very simple structure when viewed from the inside, as it were. For the speech perceiver, and for the speech researcher, perceptual complexity is defined by different criteria, such as the relative familiarity of a language, dialect, or foreign accent, the rate of speech, or the fidelity of the acoustic signal. In other words, perceptual complexity is defined not absolutely but in terms of deviations from expectancies. In the case of synthetic or degraded speech, an acoustically simpler signal may pose a perceptual problem.

1.4 FOCUS ON THE NAIVE LISTENER

The bias that I have just portrayed--that psychophysics tends to be concerned with linguistically irrelevant aspects of speech--may seem to apply only to a small portion of speech research. After all, most speech perception experiments do require subjects to respond with phonemic categories (strictly speaking, with alphabetic symbols) to the speech sounds they hear, and not with numerical ratings or other kinds of nonphonetic responses. However, it is often assumed, if only implicitly, that the phonemic or orthographic symbols employed by listeners are simply convenient labels for auditory experiences. Hand in hand with that assumption goes the much-discussed hypothesis that phonetic categories, and particularly the boundaries between them, reflect constraints imposed by the mammalian auditory system (see, e.g., Kuhl, 1981; Liberman & Mattingly, 1985). This hypothesis dovetails with another bias of psychophysical research.

Classical psychophysics is rarely concerned with subjects' experience prior to an experimental session, except for task-specific training received under controlled conditions. Essentially, psychophysics is about basic processes of perceptual translation, most often from a continuous physical dimension to a continuous psychological dimension. If categories are to be employed as responses in a psychophysical task, they are usually defined within the limited context of the experimental situation, often exemplified by the extremes of a stimulus dimension. The boundaries between such categories are either arbitrary--e.g., they may just bisect a stimulus continuum and hence depend on its range--or, if they are not (as is more often the case with speech) they are assumed to coincide with a psychoacoustic discontinuity that gave rise to the categories in the first place. Although subjects obviously have much experience with the categories of speech outside the laboratory, this experience is often considered irrelevant because the psychoacoustic basis for the category division is assumed to be present in the stimuli. (If the stimuli are synthetic and unfamiliar-sounding, so much the better.) At best, language experience may have taught subjects to attend to one particular discontinuity and to ignore another, hence certain cross-language differences in boundary location.

These assumptions are perfectly appropriate within the framework of psychophysics. Indeed, in the quest for an elegant description of the perceptual translation from the objective to the subjective realm, any intrusion of pre-experimental knowledge is undesirable. Imagine an experiment involving the perceived similarity of various round shapes, in which subjects judge two shapes as more similar than the others because both happen to look like the same familiar object (e.g., an apple). This would be an undesirable artefact (Titchener, 1909, called it the "object error") that might distort the true psychophysical function underlying the similarity judgments. This function is assumed to be universal and independent of prior experience.

There is considerable evidence, however, that many, perhaps all, phonetic distinctions rest on linguistic, not psychoacoustic criteria (see Rosen & Howell, in press; Repp & Liberman, in press). These criteria are acquired--or, if innate, are modified--through experience with spoken language. Rather than referring to particular auditory experiences, phonetic category labels--once certain orthographic and linguistic conventions are stripped off--denote specific articulatory maneuvers whose auditory correlates, though systematic, are largely irrelevant. This is most strikingly demonstrated by the finding that phonetic structure can be perceived in auditorily anomalous stimuli composed of time-varying sinusoids that imitate formant movements and thus retain information about the changing shape of the vocal tract (Remez et al., 1981; Remez, this volume). The articulatory patterns characteristic of a language presumably have evolved according to articulatory and linguistic constraints (Lindblom, 1983; Ohala, 1983), and it seems unlikely that auditory limitations have played a significant role, except in the very general sense that phonetic contrasts that are difficult to discriminate tend to be avoided or, if they occur, may lead to language change (Ohala, 1981; Bladon, this volume). I will argue below that listeners refer to their knowledge of language-specific articulatory norms when listening to speech. This reference is external to the experimental situation and inside the listener. Rather than emerging from acoustic properties of the stimulus or the stimulus ensemble, the phonetic structure imposed by the talker and recovered by the listener represents a learned conventional pattern constrained by universal articulatory possibilities.

Since it is the linguistic structure that is important in speech communication, and not the auditory properties of speech components, it is natural that human listeners focus their attention on the former and not on the latter. This attention to a discrete representation of speech influences subjects' judgments in a variety of psychophysical tasks designed to assess the psychological transformation of acoustic stimulus dimensions. For example, it is probably responsible for the category boundary effect in categorical perception experiments, as hypothesized long ago by proponents of the so-called dual-process model (e.g., Fujisaki and Kawashima, 1969, 1970; Pisoni, 1973; Samuel, 1977). However, some researchers committed to psychophysical approaches (e.g., Macmillan et al., 1977) have taken these perceptual nonlinearities to be inherent in the auditory stimulus representation. Although auditory nonlinearities do seem to occur along certain acoustic dimensions of speech, they may be unrelated to the discontinuities imposed by the mental organization of the listener (see, e.g., Watson et al., 1985; Howell & Rosen, 1983; Rosen & Howell, in press; Schouten, this volume). The same may be said about so-called phonetic trading relations and context effects (see review by Repp, 1982) which, for the most part, reflect not psychoacoustic interactions among signal components but the listener's imposition of multidimensional criteria in the process of phonetic categorization (Derr & Massaro, 1980; Massaro, this volume; Repp, 1983; however, see also Diehl, this volume).

By these arguments, speech is a particularly unwieldy object for psychophysical and psychoacoustic experimentation. If questions of auditory perception are to be addressed, why not use simpler stimuli?

If questions of speech communication are to be addressed, why use a psychophysical approach? As Massaro (this volume) aptly points out, a large part of modern speech perception research consists of either (a) applying reductionistic models to laboratory phenomena in a search for the auditory mechanisms that accomplish phonetic categorization, or (b) appealing to "special" mechanisms that do the job. Both enterprises have been sterile--the first in that it has not revealed any relevant mechanisms, and the second in that it has postponed or even relinquished the search for them. One problem with both approaches is that they represent models of speech perception according to which linguistically distinctive information somehow must emerge from the stimulus alone, without recourse to long-term mental representations of linguistic knowledge. One notable exception has been the work of Massaro and his collaborators who have consistently pursued the idea that speech perception proceeds by reference to internal category "prototypes" (see Massaro & Cohen, 1980a; Massaro, this volume). Their model, and similar ideas in the literature, lead the way toward a relational (or systemic) theory of speech perception, to be sketched further below.

2. UNDERSTANDING SPEECH PERCEPTION

The goal of speech perception researchers is to understand (or explain) speech perception--that much is obvious. However, what does this really mean? What is speech perception, and what does understanding (explaining) it entail? Probing these questions too deeply leads to profound epistemological issues. I offer only a few comments for discussion.

2.1 TWO DEFINITIONS OF PERCEPTION

The term "perception" is being used in different ways by different researchers, as has been pointed out by Chistovich (1971) and Shepard (1984), among others. An example of one usage is provided by Massaro's recent writings on categorical perception (Hary & Massaro, 1982; Massaro & Cohen, 1983b; Massaro, in press, and this volume). He argues that "categorical results do not imply categorical perception": The perception of speech continua is revealed to be continuous if only the right methods are employed. According to Hary & Massaro (1982), "a central issue in auditory information processing is whether certain auditory continua are perceived categorically rather than continuously" (p. 409). That is, it must be one or the other: Perception is entirely a function of the input. Perception is thus equated with sensory transduction--an immutable process that is insensitive to attention and experience. Of course, this is exactly what psychophysics is concerned with. The goal of speech perception research, in this view, is to find out what speech perception really is like, once all constraints imposed by attentional and experiential factors have been removed. The classification by reference to prototypes, which plays such a prominent part in Massaro's model, apparently is a post-perceptual process in his definition.

This view needs to be contrasted with a definition of perception that includes categorization and attentional filtering. According to this (my preferred) view, perception is what occurs when the transduced stimulus meets the mental structures (the "model of the world") laid down by past experience and possibly by genetic transmission (Hayek, 1952; Shepard, 1980, 1984; Yates, 1985). The result of perception is the outcome of that encounter, not the input to it. According to Fodor (1983, p. 40), "what perception must do is so to represent the world as to make it accessible to thought" through processes of transduction and inference. Categorical perception, and the apparent invariance of the categorical percept, represent the outcome of the inferential process. To find behavioral evidence of the (largely) continuous, transduced information that feeds into this process, a listener's perceptual strategy must be altered through instructions and training, or some measure of decision uncertainty (e.g., reaction time) must be obtained. Since there are a variety of mental structures a stimulus may relate to, there are often alternative ways of perceiving the same input, depending on the perceiver's experience (i.e., form of the mental representations) and attention (i.e., selection from among them). Thus, in this view, categorical results do imply categorical perception, and noncategorical results imply noncategorical perception.

Speech perception thus can mean different things depending on the situation and the subject's strategies. In addition, it has a double meaning from another perspective, depending on whether "speech" is taken to refer to the stimulus or the percept. Psychophysical research can be snugly accommodated under the stimulus-based definition that speech perception is whatever occurs when speech signals are presented to a listener. I favor a percept-based definition--that speech perception occurs when a stimulus is perceived as speech, i.e., when the listener interprets the stimulus in relation to the linguistic system. By that definition, many psychophysical experiments deal not with speech perception but with the perception of speechlike auditory stimuli. This distinction is not intended as a value judgment (indeed, psychophysical research generally surpasses speech perception research in rigor and methodological sophistication), but as a separation of largely independent domains of inquiry.

2.2 TWO DEFINITIONS OF UNDERSTANDING

What does it mean to understand (or explain) speech perception? According to one view, it involves building or programming a machine to recognize speech. For example, Chistovich (1980) presented this approach as the one taken by the Leningrad group. This pragmatic goal of "teachability" deserves our respect (for a critique, see Studdert-Kennedy, 1985). Even though the operations of the machine may not resemble those of the human brain, a speech recognition algorithm approximating human capabilities would represent a useful model of speech perception and thus increase our understanding of the process. Unfortunately, it seems that psychophysics has little to contribute to this enterprise. Psychoacoustic and physiological research has uncovered transformations in the auditory system that could be simulated by a speech processing system. However, incorporating auditory transforms into the machine representation of speech apparently does not improve speech recognition scores (Blomberg et al., 1986). This is perhaps not surprising. Machine representations need to capture the relationships between stimulus properties and precompiled knowledge structures (Shepard, 1980), and relational properties are likely to be largely invariant under transformations. Moreover, transformations of the input cannot result in an information gain, let alone in the magical emergence of properties that cannot also be computed by a central algorithm, so the most detailed coding of the speech signal is likely to be the most useful one for machines. Unless the goal is to build an analog of a complex biological system (and we are far from that stage), insights derived from psychophysical and psychophysiological research are likely to be of little use to computers. The essential problem to be solved in speech recognition research, I presume, is not that of stimulus coding but that of phonetic knowledge representation and utilization.

The alternative approach to scientific explanation is a purely theoretical one. Scientists and other human beings, of course, can perceive speech and need not (cannot) be taught explicitly, so the teachability criterion does not apply. This approach to explanation, therefore, is fundamentally different from that provided by the automatic speech recognition research. Theory construction, in psychology at least, is a cognitive act subject to individual preferences, sociological factors, and philosophical considerations (see Toulmin, 1972). One person's explanation may be another's tautology.

A variety of scientific philosophies are evident in the speech perception field, and their coexistence for a number of years suggests that they represent, in large part, individual preferences and not theories subject to empirical disconfirmation. What is worse, they do not agree on what really needs to be explained about speech perception. Rather than discussing the current theories or endorsing any of them, I am going to present a personal view below, at the danger of adding to the general confusion. My own ideas are neither fully worked out nor entirely original. (See, for example, Bregman, 1977; Elman & McClelland, 1984, 1986; Hayek, 1952; Liberman & Mattingly, 1985; Massaro & Oden, 1980a; Shepard, 1980, 1984; Yates, 1985.) Whatever their merit, however, they may serve as a useful basis for discussion at this workshop. After presenting my view, I will discuss what seem to be the major research questions from this perspective and what role psychophysics might play in this enterprise.

3. SPEECH PERCEPTION AS A RELATIONAL PROCESS

Phonetic perception--i.e., the perception of the phonological structure of speech without regard to its semantic content--has often been considered a purely input-driven process, to be contrasted with the largely knowledge-driven processes of language understanding (e.g., Marslen-Wilson & Welsh, 1978; Studdert-Kennedy, 1982). That is, it is often assumed that phonological structure <u>is</u> in the speech signal (e.g., Gibson, 1966; Fowler, 1984; Stevens & Blumstein, 1981) or <u>emerges</u> from it via specialized neural processes (Liberman & Mattingly, 1985). The present proposal contrasts with these views in that it assumes that speech perception requires two complementary ingredients: the input signal <u>and</u> the perceiver's internal representation of the speech domain. In other words, I am assuming that phonological structure emerges, especially in its language-specific details, from the relation between a stimulus and a "phonetic lexicon" in the perceiver's head which (in mature individuals) provides an exhaustive knowledge base representing all the characteristics associated with the structural units of a language.

In this view, it is not the stimulus as such (or its auditory transform) that is perceived, but rather its relationship to the phonetic knowledge base; perception thus is a relational process, a two-valued function. Its output is also two-valued: The relation of the input to the pre-existing internal structures yields (potential) awareness of the structure that provides the best fit, plus some measure of goodness of fit which may be experienced as degree of confidence or uncertainty.

How is the phonetic knowledge represented in the brain? One possible conceptualization is in terms of "prototypes" (schemata, norms, ideals, logogens, basic categories) abstracted from language experience (cf. Massaro & Oden, 1980; Flege, 1986; Yates, 1985). The mechanisms enabling this abstraction during language acquisition are unknown and may either reside in a specialized "module" (Fodor, 1983; Liberman & Mattingly, 1985) or represent general neural design principles (e.g., Grossberg, 1986). Language-specific phonetic categories are assumed to "crystallize" around central tendencies extracted from the variable input under the guidance of linguistic distinctiveness criteria. How this occurs is one of the great unsolved questions in speech research.

Just like the stimulus itself, the contents of the listener's knowledge base can be described in acoustic (optic), auditory (visual), or articulatory terms; that is, the lexicon is assumed to contain information about typical articulatory motions and their acoustic and optic concomitants, as well as possibly about their underlying dynamic parameters. The articulatory information is primary in so far as it also serves to control speech production and silent (imagined) speech, because it relates more directly to linguistic and orthographic symbols, and because it unites the different sensory modalities (as pointed out earlier). Whatever metaphor is used to describe the knowledge base--and we cannot expect to capture in words the state of a complex neural network--the important consequence of having it is that a perceiver is able, at each moment in time, to evaluate the information in the speech signal as to whether it fits the language norms. Deviations from these expectations may be perceived as unnaturalness, foreign accent, or indiviual speaker characteristics; or they may pass unnoticed.

Speech that is pronounced clearly, free of noise, and typical of the language is perceived "directly": The appropriate prototypes "resonate" to the input (Shepard, 1984). Ambiguous or degraded speech is represented in terms of its relative similarities to the most relevant prototypes. Whenever a decision is required, one prototype is selected that provides the best fit to the input (cf. Massaro & Oden, 1980). Explicit linguistic category decisions, however, are basically a response phenomenon governed by (laboratory) task requirements. Whether or not overt categorical decisions are made, the structural linguistic information is always present, being implicit in the prototypes and their relations to each other (cf. Lindblom et al., 1983). The size of the "perceptual units," and with it the size of the prototypes, is variable, being a joint function of cognitive accessibility and real-time requirements (cf. Warren's, 1981, LAME model). Thus, even though explicit recognition of individual phonemes is likely to be a function of literacy and linguistic awareness (cf. Mattingly, 1972; Morais et al., 1979), phonemic structure is nevertheless implicit in the prototype inventory: For example, /b/ is perceived when all prototypes transcribable as /b.../ are "active," i.e., resemble the input (cf. Elman & McClelland, 1984, 1986).

Properties of the speech signal become linguistic information only by virtue of their relation to the listener's knowledge base. One could imagine that the stimulus is represented in terms of a "similarity vector" (Chistovich, 1985) containing relative deviations from prototypes in some perceptual metric. This form of coding may be viewed as an effective way of information reduction, though it is by no means clear that the brain needs such a reduction the way we need it when thinking about the system's operation. That is, a similarity vector is better thought of as a set of potentials or relationships, not of physically instantiated quantities.

In my view, the "special" nature of speech, which has received so much emphasis in the past (e.g., Liberman, 1982), resides primarily in the fact that speech is a unique system of articulatory and acoustic events. In contrast to adherents of the modularity hypothesis (Fodor, 1983; Liberman & Mattingly, 1985) I suspect that the mechanisms of speech perception are general--i.e., that they can be conceptualized in terms of domain-independent models, such as adaptive systems theory (Grossberg, 1986), interactive activation theory (Elman & McClelland, 1984), or information integration theory (Massaro & Oden, 1980). In other words, I believe that the specialness of speech lies in those properties that define it as a unique phenomenon (i.e., its production mechanism, its peculiar acoustic properties, its linguistic structure and function) but not in the way the input makes contact with mental representations in the course of perception. That is, as long as we can only rely on models of the perceptual mechanism, it is likely that significant similarities will obtain across different domains, even though the physiological substrates may be quite different. This is a consequence of the relatively limited options we have for constructing models of perception and decision making.

To go one step further: If speech is special but speech perception is not, it follows that there is a lot to be learned about speech, but relatively little about speech perception. This conclusion, for what it is worth, suggests a "vertical" research strategy (giving a twist to Fodor's, 1983, arguments): The way to learn more about the speech system is to investigate its many special characteristics. This is a multidisciplinary venture, a task for the specialist called "speech researcher." By contrast, study of speech perception as such is open to a "horizontal" approach by psychologists interested in perception in general. However, there is comparitively little to be learned about that process. While there are lots of interesting facts to be uncovered about speech, the "mechanisms" of perception are a figment of the scientist's imagination (as is the mechanistic analogy itself). It is quite likely that, once we know enough about speech and have characterized the perceiver's knowledge in a suitably economic form, we also will have explained speech perception in its essential aspects.

4. A PROGRAM FOR SPEECH PERCEPTION RESEARCH

From the perspective I have adopted, there are four major questions for research on speech perception: What is the phonetic knowledge? How is it used? How is it acquired? How can it be modified?

4.1 DESCRIPTION OF THE KNOWLEDGE BASE

Before we can ask any questions about speech perception, we need to know what speech is, so we can account for the perceiver's expectations. This seemingly obvious requirement is often neglected by psychologists who plunge into speech perception experiments without considering the relevance of acoustic, articulatory, and linguistic phonetics. Even so useful a tool as Massaro's "fuzzy logical model" of information integration (Massaro & Oden, 1980a, 1980b) yields parameters characterizing phonetic prototypes whose relation to the normative properties of English utterances often remains unclear. It is often assumed that these properties will emerge from studies involving the classification of acoustically impoverished stimuli (see also Samuel, 1982). This is unlikely, however, because perceivers have detailed expectations about the full complement of acoustic properties, including those held constant in a given experiment, and they will often shift their criteria for stimulus classification along some critical dimension to compensate for the constancy or absence of others. While demonstration of this fact may be a worthwhile goal of some experiments, a more important point is that the perceivers' expectations can be assessed directly and independently (at least to a first approximation) by collecting facts about the acoustic and articulatory norms of their language, which constitute their knowledge base. Ever since Chomsky's (1965, 1968) seminal publications, the study of syntax, semantics, and phonology has been considered part of cognitive science, leading to a description of the language user's knowledge. I would like to add (normative) phonetics: The study of articulatory norms, too, yields a description of the average listener-speaker's "competence" (cf. Tatham, 1980).

I am thus proposing that the study of acoustic and articulatory phonetics be part and parcel of speech perception research. Incidentally, psychologists, with their thorough understanding of measurement and sampling problems, are especially well equipped to conduct phonetic and articulatory research, which too often has taken a case study approach in the past. Representative measurements are also important for automatic speech recognition research (Klatt, 1986). They would not make experimental determinations of prototypical perceptual parameters superfluous but rather provide a basis for their interpretation: The normative characteristics of a language are what a perceiver ought to have internalized. If deviations from the norm and/or individual differences emerge from such a comparison, the search for their causes should be an interesting and important undertaking.

In what form phonetic knowledge is represented in the brain is a question that cannot be answered conclusively by psychologists, who may choose from a number of alternative conceptualizations. As Shepard (1980, p. 181) has aptly stated, "there are many possible levels of description, and although they may appear very different in character, the various levels all pertain to the same underlying system. In this respect, the internal representation is no different from the external object." Choosing one particular level of description is basically a matter of preference and, perhaps, parsimony.

4.2 PERCEPTUAL WEIGHTS AND DISTANCES

One empirical question that psychologists may usefully address, however, is how phonetic knowledge is applied. Since a clear, unambiguous stimulus poses no challenge to the perceptual system and therefore cannot reveal its workings (cf. Shepard, 1984), the principal question is how phonetic ambiguities created by realistic signal degradation or by deliberate signal manipulation are resolved (explicitly) by the perceiver in the absence of lexical, syntactic, or other higher-order constraints. In such a situation, the perceiver must make a decision based on the perceptual distances of the input from the possible phonetic alternatives (prototypes) stored in his or her permanent knowledge base. The decision rule may be assumed to be straightforward: Select the prototype that matches the input most closely. However, what determines the degree of the match? What makes an ambiguous utterance more similar to one prototype than another? In other words, what is the phonetic distance metric, what are the dimensions of the perceptual space in which it operates, and what are the perceptual weights of these dimensions?

There are opportunities for the useful application of psychophysical methods here, since the distance metric may be, in part, a function of auditory parameters (see, e.g., Bladon & Lindblom, 1981). However, the relative importance of different acoustic dimensions for a given phonetic contrast cannot be predicted from psychophysical data alone, since it depends heavily on the nature and magnitude of the differences among the relevant prototypes, in combination with their auditory salience. Traditional psychophysics is concerned with perceptual similarities and differences between stimuli, whereas the present application requires a multidimensional psychophysics dealing with the similarity of stimuli to mental representations. The many confusion studies in the literature (beginning with Miller and Nicely, 1955) would seem to be about this issue, but the data have always been analyzed in terms of stimulus-stimulus, not stimulus-prototype similarities (which they indeed represent), and it is possible that important information has been missed. Research such as Massaro's modelling of information integration in phoneme identification (e.g., Derr & Massaro, 1980; Massaro & Oden, 1980a, 1980b) is an exemplary effort from the present viewpoint, despite certain limitations. Massaro has found again and again that stimulus attributes are evaluated in an independent and multiplicative (or log-additive) fashion in phonetic classification, and this has obvious implications for the nature of a phonetic distance metric. Many experiments on the perceptual integration and relative power of acoustic cues (e.g., Abramson & Lisker, 1985; Bailey & Summerfield, 1980; Lisker et al., 1977; Repp, 1982) also contribute relevant information. Experiments that avoid the fractionation of acoustic signals into "cues" and search for a phonetic distance metric based on more global spectral properties (Klatt, 1982, 1986) are promising but still at a very early stage.

Even though perceptual distances may reflect certain facts about auditory processing, these influences on phonetic perception are probably limited. The principal reason is that the mental structures that determine speech categorization have been built up from past experience with speech that underwent essentially the same auditory transformations as the current input is undergoing. That is, all transformations occurring during stimulus transduction are necessarily represented in the central knowledge base. Therefore, it makes relatively little difference whether we think of the input as sequences of raw spectra and of the mental categories as prototypical spectral sequences (e.g., Klatt, 1979), or whether we consider both in terms of some auditory transform or collection of discrete cues. It is the relation between the two that matters, and that relation is likely to remain topologically invariant under transformations. Only nonlinear transformations will have some influence on phonetic distances (Klatt, 1986).

4.3 PERCEPTUAL DEVELOPMENT

In addition to asking how phonetic knowledge is utilized, we must ask how and when it is acquired. Much developmental and comparative research in the past has focused on auditory discrimination abilities, and the approach has been quite psychophysical in character. The "categorical" effects that have been observed in infants and animals may not reflect phonetic perception but certain psychoacoustic discontinuities on speech continua (Jusczyk, 1985, 1986), although this suggestion becomes doubtful in view of findings (Sachs & Grant, 1976; Soli, 1983; Watson et al., 1985) that the category boundary effect can be trained away in adults. Alternatively, category boundary effects in infants may reflect an innate predisposition for perceiving a universal articulatory inventory (Werker et al., 1981). The interpretation of these data is uncertain at present. Speech perception research in older children (e.g., Elliot et al., 1981; Tallal & Stark, 1981) also has often focused on their auditory abilities, not specifically on their criteria for phonetic identification and on the nature of their phonetic knowledge. Only more recently, following the lead of researchers such as Kuhl (1979) and Werker et. al. (1981) has phonetic categorization in infancy been studied more carefully. A finding of special significance is the discovery (Werker & Tees, 1984) that infants' ability to perceive phonetic contrasts foreign to their parents' language declines precipitously before 1 year of age. This stage seems to mark the beginnings of a language-specific phonetic lexicon. It is an important research endeavor to trace the accumulation and refinement of phonetic knowledge through different stages of development, and much work remains to be done (see Jusczyk, this volume).

4.4 PERCEPTUAL LEARNING

Another question of great theoretical and practical importance is how the phonetic knowledge, once it is established in the mature adult, can be augmented and modified. This concerns the process of second language learning and also, to some extent, the skills acquired by professional phoneticians (and even by subjects in a laboratory task, although their skills may be rather temporary). Furthermore, there is the very interesting question of bilingualism--the separation and interaction of two different, fully established phonetic knowledge bases. Until recently, little rigorous research had been carried out in this predominantly education-oriented area. Research is burgeoning, however, and is yielding interesting results (see Flege, in press).

Another, related question is to what extent reduced or distorted auditory input over longer time periods affects the internal representation of phonetic knowledge. For example, it has been reported recently that otitis media in childhood (Welsh et al., 1983) or monaural hearing deprivation in adulthood (Silman et al., 1984) may result in reduced speech perception capabilities. Certainly, the congenitally hearing-impaired must have a very different representation of their limited phonetic experiences, and hearing impairments acquired later in life may distort the knowledge base as well. It has often been observed that the speech perception of the hearing-impaired is not completely predictable from assessments of auditory capacity (e.g., Tyler et al., 1982). One reason for this may be that there are distortions, not only in the auditory processing of speech (to which they are commonly attributed), but in the mental representations that hearing-impaired listeners refer to in phonetic classification. Such distortions are especially likely to result when hearing deteriorates progressively at a rate that exceeds the rate at which mental prototypes can be modified: A listener then expects to hear things that the auditory system cannot deliver. On the other hand, if the prototypes are degraded from many years of impoverished auditory experience, then there is little hope of improving speech perception by "improving" the acoustic signal, at least not without extensive training to rebuild the prototypes (cf. Sidwell & Summerfield, 1985).

5. MAKING PSYCHOPHYSICS MORE RELEVANT TO SPEECH RESEARCH

One characteristic of the psychophysical approach is that it is domain-independent. The psychophysical methods applied in the study of speech perception are essentially the same as those applied in research on auditory, visual, or tactile perception of nonspeech stimuli. Indeed, the generality across different stimulus domains and modalities of Weber's law or the law of temporal summation has been an important discovery. Such laws are in accord with behaviorist and information-processing orientations in psychology, which assume that perception and cognition are governed by general-purpose, domain-independent processes. The description of such processes is an important part of psychological research.

By focusing on domain-independent laws of perception, however, psychophysics essentially ignores those features that are specific to speech and whose investigation is critical to an understanding of speech perception as distinct from perception in general. Of course, there are many aspects that speech shares with nonspeech sounds and even with stimuli in other modalities. Research on the perception of those, however, leads only to an understanding of sound perception, temporal change perception, timbre perception, even categorization--in short, of all the things that speech perception has in common with nonspeech perception. What is missing is the main ingredient: the content. To understand speech perception fully, research needs to focus on the unique properties of speech, which include the facts that it is articulated (and hence peculiarly structured), capable of being imitated by a perceiver, and perceived as segmentally structured for purposes of linguistic communication. I see at least one way in which the sophisticated methods of psychophysics could be adapted to these special features and thus be made more relevant to speech research.

Psychoacoustic approaches to speech perception deal with both stimulus and response at some remove from the mechanism that is directly responsible for most (if not all) special properties of speech: the vocal tract. A more speech-relevant psychophysics might examine the articulatory source of the acoustic signal in relation to what is probably the most direct evidence that perception has occurred--the perceiver's vocal reproduction of what has been heard (or seen). I am thus proposing an <u>articulatory psychophysics</u> based on the realization that speech is constituted of motor events (cf. Liberman & Mattingly, 1985). Its goal would be to describe the lawful relationships between a talker's articulations and a listener's perception or imitation of them.

A first step in this enterprise would be to look at the speech signal not in terms of its acoustic properties, but in terms of the articulatory information that it conveys. This is done most easily by generating the stimuli using an articulatory synthesizer or an actual human talker, perhaps in conjunction with analytic methods for extracting the vocal tract area function from the acoustic signal (e.g., Atal et al., 1978; Ladefoged et al., 1978; Schroeder & Strube, 1979). Articulatory synthesis studies in the literature (e.g., Lindblom and Sundberg, 1971; Rubin et al., 1981; Abramson et al., 1981; Kasuya et al., 1982) illustrate this approach. A second step would be to examine subjects' articulatory (rather than just written) response to speech stimuli. Studies of vocal imitation (e.g., Chistovich et al., 1966; Kent, 1973; Repp & Williams, 1985) have commonly analyzed stimulus-response relationships in terms of acoustic parameters and thus fall somewhat short of the stated goal. In the wide field of speech production research, there are few studies that have required subjects to listen to speech stimuli and reproduce them; almost always the task has been to read words or nonsense materials, and measurements have focused on normative productions characteristic of a language, not on talkers' imitative or articulatory skills. The final step towards a true articulatory psychophysics would be to measure subjects' articulatory response to articulatorily defined stimuli, generated either by an articulatory synthesizer or by a human model whose articulators are likewise monitored. An important (though necessarily crude) example of this still rare approach is the work of Meltzoff and Moore (see 1985) on facial imitation in infancy. More detailed studies of adult subjects should benefit from the development of more economic descriptions of articulation and its underlying control parameters (Browman & Goldstein, 1985; Kelso et al., 1985).

Such studies would assess how articulatory dimensions such as jaw height, lip rounding, mouth opening, or velar elevation--or perhaps more global articulatory parameters such as the vocal tract area function--are apprehended by a listener/speaker, and how they are translated and rescaled to fit his or her own articulatory dimensions. Rather than relating physical stimulus parameters to some subjective auditory scale that is irrelevant to speech communication, the psychophysical function would relate equivalent articulatory measures in the model speaker and the imitator. Such functions would relate more directly to questions of speech acquisition and phonetic language learning than any measure of auditory perception. Even though articulatory psychophysics is likely to encounter various influences of linguistic categories on the subject's articulatory response, reflecting aspects of motor control that have become established through habit and practice, at least it would bypass the stage of overt categorical decisions (cf. Chistovich et al., 1966) which characterizes so many laboratory tasks. It may be possible to overcome these articulatory habits through training, and such training may not only yield better estimates of articulatory information transfer but also potential practical benefits for second-language learners and speech pathologists (more so than training in auditory discrimination). An ancillary, hitherto little-investigated topic is that of articulatory awareness--a talker's ability to consciously observe and manipulate his or her articulators.

6. SUMMARY

Before summing up, one qualification is in order concerning the role of psychophysics in understanding speech perception. I have argued that this role is limited, and undoubtedly many will disagree with this opinion. In addition, however, I have followed the custom of the mainstream speech perception literature (and my own proclivities) by considering speech perception to be synonymous with the perception of phonetic structure. There are many other aspects of speech, however, such as intonation, stress, speaking rate, effort, rhythm, emotion, voice quality, speaker characteristics, room reverberation, and separation from other environmental sounds. All these aspects are worthy of detailed investigation, and although speech-specific knowledge also plays a role in their perception (e.g., Tuller & Fowler, 1980; Darwin, 1984; Ainsworth & Lindsay, 1985), auditory psychophysics probably has a more important contribution to make to research on these topics. The perception of subtle gradations becomes especially important in the registration of paralinguistic information. Thus, in yet another sense, the relevance of psychophysics to speech perception depends on how broadly or narrowly the field of speech perception research is defined.

In this paper I have tried to do five things. First, I have attempted to characterize the psychophysics of speech perception in terms of certain biases: heavy emphasis on the auditory modality; preoccupation with methodology; treatment of speech as a collection of sounds; neglect of the perceiver's knowledge and expectations. This characterization may well seem a caricature to those who espouse a broad definition of psychophysics. However, even though only a small part of speech perception research may fit my description, it represents an extreme (a prototype of psychophysical orthodoxy, as it were) that, though only rarely instantiated in its pure form, nevertheless exerts a certain "pull" on research in the field.

Second, I have tried to ask what it means to understand speech perception. Far from giving a satisfactory answer to this difficult question, I have made two points: Perception can be defined narrowly as a rigid process of transduction, or more broadly as a flexible process of relating the input to a knowledge base; I favor the second definition. As to understanding, it can mean producing some tangible evidence, such as a good recognition algorithm, or it can remain largely a matter of personal indulgence. My sympathies are with the former approach, but my own research has been very much within the latter.

Third, I have characterized speech perception as the application of detailed phonetic knowledge. I have argued that the mechanisms of speech perception may be quite general, but that the system as a whole is unique, thus stating a modified (possibly trivial) version of the modularity hypothesis (Fodor, 1983; Liberman & Mattingly, 1985). This has led me further to suggest that speech perception, when considered divorced from the whole system, is a relatively shallow topic for investigation, and that a better understanding of speech perception will result indirectly from studying the whole "speech chain" (Denes & Pinson, 1963).

Fourth, I have discussed four major research questions that follow from the view taken here: Description of the phonetic knowledge; rules of its application; time course of its acquisition; and its modifiability in adulthood. The first and third of these topics are considered central to speech research. Many traditional core questions of speech perception, together with opportunities for the application for psychophysical methods, are contained in the second topic and thus are assigned a secondary role. Special emphasis is placed on articulatory and acoustic phonetics as a means for gaining insight into the language user's perceptual knowledge.

Finally, I have proposed the possibility of an articulatory psychophysics as a way of increasing the relevance of psychophysical methods to speech research.

In sum, I have painted a somewhat pessimistic picture of speech perception research, and in particular of the contribution of psychophysical approaches. This should not be taken as an assault on auditory psychophysics as such; on the contrary, the investigation of auditory function is an important area in which much excellent work is being done, as illustrated by many contributions to this workshop. What is at issue is the relevance of this general approach to the study of speech perception. If my paper stimulates discussion of this fundamental question, it will have served its purpose.

REFERENCES

1. Abramson, A., and Lisker, L. (1985). Relative power of cues: F0 versus voice timing. In V.A. Fromkin (Ed.), Phonetic linguistics. Essays in honor of Peter Ladefoged. New York: Academic. Pp. 25-33.
2. Abramson, A.S., Nye, P.W., Henderson, J.B., and Marshall, C.W. (1981). Vowel height and the perception of consonantal nasality. Journal of the Acoustical Society of America, 70, 329-339.
3. Ainsworth, W.A., and Lindsay, D. (1986). Perception of pitch movement on tonic syllables in British English. Journal of the Acoustical Society of America, 79, 472-480.
4. Atal, B.S., Chang, J.J., Mathews, M.V., and Tukey, J.W. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. Journal of the Acoustical Society of America, 63, 1535-1555.
5. Bailey, P.J., and Summerfield, Q. (1980). Information in speech: Observations on the perception of [s]-stop clusters. Journal of Experimental Psychology: Human Perception and Performance, 6, 536- 563.
6. Bladon, R.A.W., and Lindblom, B.E.F. (1981). Modeling the judgment of vowel quality. Journal of the Acoustical Society of America, 69, 1414-1422.
7. Blomberg, M., Carlson, R., Elenius, K. and Granström, B. (1986). Auditory models as front ends in speech recognition systems. In

J.S. Perkell and D.H. Klatt (Eds.), Invariance and variability in speech processes. Hillsdale, N.J.: Erlbaum. Pp. 108-114.

8. Blumstein, S.E., and Stevens, K.N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. Journal of the Acoustical Society of America, 66, 1001-1017.

9. Browman, C.P., and Goldstein, L.M. (1985). Dynamic modeling of phonetic structure. In V.A. Fromkin (Ed.), Phonetic linguistics. Essays in honor of Peter Ladefoged. New York: Academic. Pp. 35-53.

10. Bregman, A.S. (1977). Perception and behavior as compositions of ideals. Cognitive Psychology, 9, 250-292.

11. Chistovich, L.A. (1971). Problems of speech perception. In L.L. Hammerich, R. Jakobson, and E. Zwirner, (Eds.), Form & substance. Copenhagen: Akademisk Forlag. Pp. 83-93.

12. Chistovich, L.A. (1980). Auditory processing of speech. Language and Speech, 23, 67-75.

13. Chistovich, L.A. (1985). Central auditory processing of peripheral vowel spectra. Journal of the Acoustical Society of America, 77, 789- 805.

14. Chistovich, L.A., Fant, G., de Serpa-Leitao, A., and Tjernlund, P. (1966). Mimicking and perception of synthetic vowels. Quarterly Progress and Status Report (Royal Technical University, Speech Transmission Laboratory, Stockholm), 2, 1-18.

15. Chomsky, N. (1965). Aspects of the theory of syntax. Cambridge, MA: MIT Press.

16. Chomsky, N. (1968). Language and mind. New York: Harcourt, Brace & World.

17. Darwin, C.J. (1984). Perceiving vowels in the presence of another sound: Constraints on formant perception. Journal of the Acoustical Society of America, 76, 1636-1647.

18. Denes, P.B., and Pinson, E.N. (1963). The speech chain. Bell Telephone Laboratories.

19. Derr, M.A., and Massaro, D.W. (1980). The contribution of vowel duration, F0 contour, and frication duration as cues to the /juz/-/jus/ distinction. Perception & Psychophysics, 27, 51-59.

20. Elliot, L.L., Longinotti, C., Clifton, L.-A., and Meyer, D. (1981). Detection and identification thresholds for consonant-vowel syllables. Perception & Psychophysics, 30, 411-416.

21. Elman, J.L., and McClelland, J.L. (1984). Speech perception as a cognitive process: The interactive activation model. In N.J. Lass (Ed.), Speech and Language: Advances in research and practice. Vol. 10. New York: Academic. Pp. 337-374.

22. Elman, J.L., and McClelland, J.L. (1986). Exploiting lawful variability in the speech wave. In J.S. Perkell and D.H. Klatt (Eds.), Invariance and variability in speech processes. Hillsdale, NJ: Erlbaum. Pp. 360-380.

23. Flege, J.E. (in press). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), Human communication and its disorders, Vol. 1. Norwood, NJ: Ablex.

24. Fodor, J. (1983). The modularity of mind. Cambridge, MA: MIT Press.

25. Fõnagy, I. (1961). Communication in poetry. Word, 17, 194-218.

26. Fowler, C.A. (1984). Segmentation of coarticulated speech in perception. Perception & Psychophysics, 36, 359-368.

27. Fujisaki, H., and Kawashima, T. (1969). On the modes and mechanisms of speech perception. Annual Report of the

Engineering Research Institute (Faculty of Engineering, University of Tokyo), 28, 67-73.

28. Fujisaki, H., and Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. Annual Report of the Engineering Research Institute (Faculty of Engineering, University of Tokyo), 29, 207-214.
29. Ganong, W.F., III, and Zatorre, R.J. (1980). Measuring phoneme boundaries four ways. Journal of the Acoustical Society of America, 68, 431-439.
30. Gibson, J.J. (1966). The senses considered as perceptual systems. Boston: Houghton Mifflin.
31. Green, K.P., and Miller, J.L. (1985). On the role of visual rate information in phonetic perception. Perception & Psychophysics, 38, 269-276.
32. Grossberg, S. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E.C. Schwab and H.C. Nusbaum (Eds.), Pattern recognition by humans and machines, Vol. I. New York: Academic.
33. Hary, J.M., and Massaro, D.W. (1982). Categorical results do not imply categorical perception. Perception & Psychophysics, 32, 409-418.
34. Hayek, F.A. (1952). The sensory order. Chicago: University of Chicago Press.
35. Howell, P., and Rosen, S. (1983). Production and perception of rise time in the voiceless affricate/fricative distinction. Journal of the Acoustical Society of America, 73, 976-984.
36. Hrushovski, B. (1980). The meaning of sound patterns in poetry. Poetics Today, 2, 39-56.
37. Jusczyk, P.W. (1985). On characterizing the development of speech perception. In J. Mehler and R. Fox (Eds.), Neonate cognition: Beyond the blooming buzzing confusion. Hillsdale, NJ: Erlbaum. Pp. 199-229.
38. Jusczyk, P.W. (1986). Toward a model of the development of speech perception. In J.S. Perkell and D.H. Klatt (Eds.), Invariance and variability in speech processes. Hillsdale, NJ: Erlbaum. Pp. 1-18.
39. Kasuya, H., Takeuchi, S., Sato, S., and Kido, K. (1982). Articulatory parameters for the perception of bilabials. Phonetica, 39, 61-70.
40. Kelso, J.A.S., Vatikiotis-Bateson, E., Saltzman, E.L., and Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. Journal of the Acoustical Society of America, 77, 266-280.
41. Kent, R.D. (1973). The imitation of synthetic vowels and some implications for speech memory. Phonetica, 28, 1-25.
42. Klatt, D.H. (1979). Speech perception: a model of acoustic-phonetic analysis and lexical access. Journal of Phonetics, 7, 276-312.
43. Klatt, D.H. (1982). Prediction of perceived phonetic distance from critical-band spectra: A first step. Proceedings of the IEEE International Conference on Acoustic, Speech, and Signal Processing, Paris, France. Pp. 1278-1281.
44. Klatt, D.H. (1986). Problem of variability in speech recognition and in models of speech perception. In J.S. Perkell and D.H. Klatt (Eds.), Invariance and variability in speech processes. Hillsdale, NJ: Erlbaum. Pp. 300-319.

45. Kuhl, P.K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. Journal of the Acoustical Society of America, 66, 1668-1679.
46. Kuhl, P.K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. Journal of the Acoustical Society of America, 70, 340-349.
47. Kuhl, P.K. (1985). Categorization of speech by infants. In J. Mehler and R. Fox (Eds.), Neonate cognition: Beyond the blooming buzzing confusion. Hillsdale, NJ: Erlbaum. Pp. 231-262.
48. Kuhl, P.K., and Meltzoff, A.N. (1982). The bimodal perception of speech in infancy. Science, 218, 1138-1141.
49. Ladefoged, P., Harshman, R., Goldstein, L., and Rice, L. (1978). Generating vocal tract shapes from formant frequencies. Journal of the Acoustical Society of America, 64, 1027-1035.
50. Liberman, A.M. (1982). On finding that speech is special. American Psychologist, 37, 148-167.
51. Liberman, A.M., and Mattingly, I.G. (1985). The motor theory of speech perception revised. Cognition 21, 1-36.
52. Lindblom, B. (1983). Economy of speech gestures. In P.F. MacNeilage (Ed.), The production of speech. New York: Springer-Verlag. Pp. 207-246.
53. Lindblom, B., MacNeilage, P., and Studdert-Kennedy, M. (1983). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, and D. Dahl (Eds.), Explanations of linguistic universals. The Hague: Mouton. Pp. 181-203.
54. Lindblom, B.E.F., and Sundberg, J.E.F. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. Journal of the Acoustical Society of America, 50, 1166-1179.
55. Linell, P. (1982). The concept of phonological form and the activities of speech production and speech perception. Journal of Phonetics, 10, 37-72.
56. Lisker, L., Liberman, A.M., Erickson, D.M., Dechovitz, D., and Mandler, R. (1977). On pushing the voice-onset-time (VOT) boundary about. Language and Speech, 20, 209-216.
57. MacKain, K.S., Best, C.T. and Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. Applied Psycholinguistics, 2, 369-390.
58. Macmillan, N.A., Kaplan, H.L., and Creelman, C.D. (1977). The psychophysics of categorical perception. Psychological Review, 84, 452-471.
59. Marks, L.E. (1978). The unity of the senses. New York: Academic.
60. Marslen-Wilson, W.D., and Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. Cognitive Psychology, 10, 29-63.
61. Massaro, D.W. (in press). Categorical partition: A fuzzy logical model of categorization behavior. In S.N. Harnad (Ed.), Categorical Perception. New York: Cambridge University Press.
62. Massaro, D.W., and Cohen, M.M. (1983a). Evaluation and integration of visual and auditory information in speech perception. Journal of Experimental Psychology: Human Perception and Performance, 9, 753-771.
63. Massaro, D.W., and Cohen, M.M. (1983b). Categorical or continuous speech perception: a new test. Speech Communication, 2, 15-35.

64. Massaro, D.W., and Oden, G.C. (1980a). Speech perception: A framework for research and theory. In N.J. Lass (Ed.), Speech and Language: Advances in research and practice. Vol.3. New York: Academic. Pp. 129-165.
65. Massaro, D.W., and Oden, G.C. (1980b). Evaluation and integration of acoustic features in speech perception. Journal of the Acoustical Society of America, 67, 996-1013.
66. Mattingly, I.G. (1972). Reading, the linguistic process, and linguistic awareness. In J.F. Kavanagh and I.G. Mattingly (Eds.), Language by ear and eye: The relationships between speech and reading. Cambridge, MA: MIT Press. Pp. 133-148.
67. McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264, 746-748.
68. Meltzoff, A.N., and Moore, M.K. (1985). Cognitive foundations and social functions of imitation and intermodal representation in infancy. In J. Mehler and R. Fox (Eds.), Neonate cognition: Beyond the blooming buzzing confusion. Hillsdale, NJ: Erlb Pp. 139-156.
69. Miller, G.A., and Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 27, 338-352.
70. Morais, J., Cary, L. Alegria, J., and Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? Cognition, 7, 323-331.
71. Neisser, U. (1976). Cognition and reality: Principles and implications of cognitive psychology. San Francisco: Freeman.
72. Ohala, J.J. (1981). The listener as a source of sound change. In C.S. Marek, R.A. Hendrick, and M.F. Miller (Eds.), Papers from the parasession in language and behaviour. Chicago: Chicago Linguistic Society. Pp. 178-203.
73. Ohala, J.J. (1983). The origin of sound patterns in vocal tract constraints. In P.F. MacNeilage (Ed.), The production of speech. New York: Springer-Verlag. Pp. 189-216.
74. Pisoni, D.B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. Perception & Psychophysics, 13, 253-260.
75. Pisoni, D.B., and Lazarus, J.H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. Journal of the Acoustical Society of America, 55, 328-333.
76. Remez, R.E., Rubin, P.E., Pisoni, D.B., and Carrell, T.D. (1981). Speech perception without traditional speech cues. Science, 212, 947-950.
77. Repp, B.H. (1981). On levels of description in speech research. Journal of the Acoustical Society of America, 69, 1462-1464.
78. Repp, B.H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. Psychological Bulletin, 92, 81-110.
79. Repp, B.H. (1983). Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. Speech Communication, 2, 341-362.
80. Repp, B.H. (1984). Categorical perception: Issues, methods, findings. In N.J. Lass (Ed.), Speech and language: Advances in Research and practice. Vol.10. New York: Academic. Pp. 243-335.
81. Repp, B.H., and Liberman, A.M. (in press). Phonetic category boundaries are flexible. In S.N. Harnad (Ed.), Categorical Perception. New York: Cambridge University Press.

82. Repp, B.H., and Williams, D.R. (1985). Categorical trends in vowel imitation: Preliminary observations from a replication experiment. Speech Communication, 4, 105-120.

83. Rosen, S., and Howell, P. (in press). Auditory, articulatory, and learning explanations of categorical perception in speech. In S.N. Harnard (Ed.), Categorical Perception. New York: Cambridge University Press.

84. Rosner, B.S. (1984). Perception of voice-onset-time continua: A signal detection analysis. Journal of the Acoustical Society of America, 75, 1231-1242.

85. Rubin, P., Baer, T., and Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. Journal of the Acoustical Society of America, 70, 321-328.

86. Sachs, R.M., and Grant, K.W. (1976). Stimulus correlates in the perception of voice onset time (VOT): II. Discrimination of speech with high and low stimulus uncertainty. Journal of the Acoustical Society of America, 60, (Suppl. No. 1), S91. (Abstract)

87. Samuel, A.G. (1977). The effect of discrimination training on speech perception: Noncategorical perception. Perception & Psychophysics, 22, 321-330.

88. Samuel, A.G. (1982). Phonetic prototypes. Perception & Psychophysics, 31, 307-314.

89. Schroeder, M.R., and Strube, H.W. (1979). Acoustic measurements of articulator motions. Phonetica, 36, 302-313.

90. Shepard, R.N. (1980). Psychophysical complementarity. In M. Kubovy and J.R. Pomerantz (Eds.), Perceptual organization. Hillsdale, NJ: Erlbaum. Pp. 279-341.

91. Shepard, R.N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. Psychological Review, 91, 417-447.

92. Sidwell, A., and Summerfield, Q. (1985). The effect of enhanced spectral contrast on the internal representation of vowel-shaped noise. Journal of the Acoustical Society of America, 78, 495-506.

93. Silman, S., Gelfand, S.A., and Silverman, K.E.A. (1984). Late-onset auditory deprivation: Effects of monaural versus binaural hearing aids. Journal of the Acoustical Society of America, 76, 1357-1362.

94. Soli, S.D. (1983). The role of spectral cues in discrimination of voice onset time differences. Journal of the Acoustical Society of America, 73, 2150-2165.

95. Stevens, K.N., and Blumstein, S.E. (1978). Invariant cues for place of articulation in stop consonants. Journal of the Acoustical Society of America, 64, 1358-1368.

96. Stevens, K.N., and Blumstein, S.E. (1981). The search for invariant acoustic correlates of phonetic features. In P.D. Eimas and J.L. Miller (Eds.), Perspectives in the study of speech. Hillsdale, NJ: Erlbaum. Pp. 1-38.

97. Studdert-Kennedy, M. (1982). On the dissociation of auditory and phonetic perception. In R. Carlson and B. Granström (Eds.), The representation of speech in the peripheral auditory system. Amsterdam: Elsevier. Pp. 9-26.

98. Studdert-Kennedy, M. (1985). Perceiving phonetic events. In W.H. Warren and R.E. Shaw (Eds.), Persistence and change: Proceedings of the first international conference on event perception. Hillsdale, NJ: Erlbaum.

99. Studdert-Kennedy, M., Liberman, A.M., Harris, K.S., and Cooper, F.S. (1970). Motor theory of speech perception: A reply to Lane's critical review. Psychological Review, 77, 234-249.

100. Summerfield, Q. (1979). Use of visual information for phonetic perception. Phonetica, 36, 314-331.
101. Summerfield, Q. (in press). Preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd and R. Campbell (Eds.), Hearing by eye. Hilldale, NJ: Erlbaum.
102. Tallal, P., and Stark, R. (1981). Speech acoustic-cue discrimination abilities of normally developing and language-impaired children. Journal of the Acoustical Society of America, 69, 568-574.
103. Tatham, M.A.A. (1980). Phonology and phonetics as part of the language encoding/decoding system. In N.J. Lass (Ed.), Speech and language: Advances in research and practice. Vol.3. New York: Academic. Pp. 35-73.
104. Titchener, E.B. (1909). Lectures on the elementary psychology of the thought-process.
105. Toulmin, S. (1972). Human understanding: The collective use and evolution of concepts. Princeton: Princeton University Press.
106. Tuller, B., and Fowler, C.A. (1980). Some articulatory correlates of perceptual isochrony. Perception & Psychophysics, 27, 277-283.
107. Tyler, R.S., Summerfield, Q., Wood, E.J. and Fernandes, M.A. (1982). Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners. Journal of the Accoustical Society of America, 72, 740-752.
108. Vinegrad, M.D. (1972). A direct magnitude scaling method to investigate categorical versus continuous modes of speech perception. Language and Speech, 15, 114-121.
109. Warren, R.M. (1981). Chairman's comments. In T. Myers, J. Laver, and J. Anderson (Eds.), The cognitive representation of speech. Amsterdam: North Holland. Pp. 34-37.
110. Watson, C.S., Kewley-Port, D., and Foyle, D.C. (1985). Temporal acuity for speech and nonspeech sounds: The role of stimulus uncertainty. Journal of the Acoustical Society of America, 77, (Suppl. No. 1), S27. (Abstract)
111. Welsh, L.W., Welsh, J.J., and Healy, M.P. (1983). Effect of sound deprivation on central hearing. Laryngoscope, 93, 1569-1575.
112. 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.
113. Werker, J.F., Gilbert, J.H.V., Humphrey, K., and Tees, R.C. (1981). Developmental aspects of cross-language speech perception. Child Development, 52, 349-355.
114. Wood, C.C. (1976). Discrimination, response bias, and phoneme categories in discrimination of voice onset time. Journal of the Acoustical Society of America, 60, 1381-1389.
115. Yates, J. (1985). The content of awareness is a model of the world. Psychological Review 92, 249-284.