This book was manufactured in the United States of America.

Typography by The Type Shoppe, Inc.
Printing and binding by BookCrafters.
Cover design by Joseph Dieter, Jr.
Book design by Sheila Stoneham.

Library of Congress Cataloging-in-Publication Data
Speech perception and linguistic experience : issues in cross-language research / edited by Winifred Strange.
p. cm.
Experiences of participants in a conference entitled "Workshop on Cross-Language Speech Perception", held at the University of South Florida in 1992.
Includes bibliographical references and index.
ISBN 0-912752-36-X
1. Speech perception—Congresses. 2. Second language acquisition—Congresses. 3. Psycholinguistics—Methodology—Congresses. I. Strange, Winifred. II. Workshop on Cross-Language Speech Perception (1992 : University of South Florida)
P37.5.S68S68 1995 95-43317
401'.93—dc20 CIP

YORK PRESS BALTIMORE

Chapter • 15

Variability and Invariance in Speech Perception
A New Look at Some Old Problems in Perceptual Learning

David B. Pisoni and
Scott E. Lively

From the earliest days of modern cognitive psychology, theorists have devoted much of their research to abstractionist accounts of perception, learning, and memory. In keeping with the Zeitgeist, most researchers assumed that the stimulus environment was impoverished and that the perceiver engaged in a great deal of constructive processing in order to make sense of the chaotic world (Neisser 1967). Perhaps one of the best examples of this approach in cognitive psychology can be found in the field of speech perception, which has always relied very heavily on the abstractionist views derived from formal linguistic theory to define the units of perceptual analysis. By viewing language as an idealized symbolic system consisting of discrete context-free elements, linguists could focus their research efforts on more abstract theoretical issues, such as phonology and syntax, without having to worry about how speech is perceived or how it is represented in the mind of the listener. The following statements from Chomsky about the role of idealized forms of language in linguistic theory, and his remarks about the competence–performance distinction are well known to linguists and psycholinguists and epitomize the abstractionist viewpoint:

> Linguistic theory is concerned primarily with an ideal speaker–listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance. (Chomsky 1965)

> We thus make a fundamental distinction between competence (the speaker-hearer's knowledge of his language) and performance (the actual use of language in concrete situations). Only under the idealization set forth in the preceding paragraph is performance a direct reflection of competence. (Chomsky 1965)

One consequence of this formal approach to language has been the almost total disregard for a variety of important problems that deal with stimulus variability and acoustic–phonetic invariance in speech perception. Even when attention has been directed to acoustic and perceptual analyses of the speech signal, most researchers have been content in occupying themselves with the search for acoustic invariance that corresponded in a one-to-one manner with a set of classically defined static perceptual categories such as phonemes or phonetic segments (Stevens and Blumstein 1978, 1980). However, the concept of the phoneme as an abstract idealized linguistic unit has always remained problematical to engineers, perceptual psychologists and speech scientists, who have spent many years trying to find first-order acoustic invariants for phonemes in the speech waveform. The continued search for acoustic–phonetic invariance is surprising given the fact that there has always been a great deal of disagreement, even in linguistics, about precisely what phonemes are and how they should be defined within a particular linguistic theory.

From the very earliest days of modern speech research going back to the invention of the sound spectrograph, it became apparent that such linguistic units as phonemes were not discrete elements in the speech waveform (Fant 1973). Instead, they turned out to be highly context-dependent units that could be affected by a wide variety of factors that modulated their physical realization in the speech waveform. Numerous perceptual experiments in the early 1950s revealed the existence of multiple acoustic cues to almost every phonetic contrast. In many cases, these cues were acoustically quite diverse, overlapping in time and often highly redundant so that the listener could reliably perceive a particular phonetic distinction despite noise or degradation in the signal.

The traditional abstractionist view of speech as an idealized sequence of discrete symbolic units has had a profound and long-lasting influence in the field of speech perception. Much of the early work on speech cues carried out at Haskins Laboratories in the 1950s was initially concerned with identifying acoustic invariants in the speech signal that corresponded uniquely to such linguistic units as phonemes of the linguistic message. However, within a short time researchers discovered that the acoustic cues to many speech sounds were influenced in systematic ways by the surrounding phonetic context (Cooper et al. 1952). These findings suggested that a search for a set of first-order acoustic–

phonetic invariances, that is, simple one-to-one correspondences between speech cues and successive phonemes of the perceived linguistic message, was unlikely to be very successful. The conclusions that Cooper et al. arrived at in 1952 are a good example of the thinking about the problem at the time:

> ...the important point, however phrased, is a caution that one may not always be able to find the phoneme in the speech wave, because it may not exist there in free form; in other words, one should not expect always to be able to find acoustic invariants for the individual phonemes. (Cooper et al. 1952 p. 605)

Despite these conclusions made over 40 years ago, the abstractionist assumptions about the role of phonemes and discrete segmental representations in speech perception have been maintained over the years, and the search for acoustic–phonetic invariants has continued even up to the present time, although these views are currently framed within the context of spoken word recognition and lexical access (Stevens 1993) or neurobiological accounts that employ neurally inspired recognition algorithms (Sussman, McCaffrey, and Matthews 1991). The consequences of these views about speech have been quite substantial and wide reaching in terms of both theory and research, as well as experimental methodology in a number of different areas, such as speech recognition, speech synthesis, infant perceptual development, clinical audiology and cross-language studies of speech perception.

Considered against this historical background, a number of recent findings have raised important questions about the traditional metatheory and formalization of language that has been assumed by most speech researchers since the late 1940s. In particular, the issue of stimulus variability has come to the forefront in recent discussions of perception, learning, and memory (Brooks 1978; Elman and McClelland 1986). The heart of the problem deals with the mapping of highly variable context-sensitive speech signals onto sequences of discrete context-free perceptual categories that the listener is assumed to construct as the end product of the perceptual process.

There are good reasons for believing that the major problems of variability in speech perception can be accounted for by several recent proposals concerning categorization, classification, and concept learning (Medin and Barsalou 1987). Along with recent developments in the field of categorization, there is also a growing body of research that provides evidence for the encoding of specific episodic information in memory along with the details of perceptual analysis (Schacter and Church 1992; Goldinger 1992; Kolers 1973). These findings from studies of "nonanalytic cognition" have raised a number of additional questions about the traditional views surrounding the nature of perception and memory and the more general claims for the primacy of abstractionist

symbolic representations in cognition (Jacoby and Brooks 1984). If we consider the problems of variability in speech perception to be a special case of the more general problems of categorization and classification, then it seems appropriate to examine how recent models of categorization might contribute to the solution of several long-standing problems in speech perception (Medin and Barsalou 1987). We attempt to do that here.

This chapter is divided into three major sections. In the first section, we consider whether the properties of speech are compatible with the criteria proposed in recent studies of nonanalytic cognition. Despite the long history of abstractionist or symbolic accounts of speech perception, there is evidence that the details of specific instances are also encoded in memory and affect subsequent perceptual processing and retention. In the second section, we summarize several recent studies on variability in speech perception and spoken word recognition. These studies demonstrate that stimulus variability is not lost as a consequence of perceptual processing and may be useful and informative to listeners in a variety of perceptual and memory tasks. Finally, in the third section, we describe the results of several recent laboratory training studies on the acquisition of English /ɹ/ and /l/ by Japanese listeners. Our findings on perceptual learning of novel linguistic contrasts demonstrate that under certain experimental conditions, where there is high stimulus variability, Japanese listeners can learn to perceive novel linguistic contrasts in a robust manner. We also show that this knowledge generalizes to new words containing /ɹ/ and /l/ and to novel tokens produced by new talkers. In addition, we have found that the perceptual learning and knowledge acquired under these particular high-variability training conditions appears to be retained over time, even without additional exposure to these contrasts in the linguistic environment.

ABSTRACTIONIST VERSUS EPISODIC APPROACHES TO SPEECH PERCEPTION

A number of recent studies on categorization and memory have provided evidence for the encoding and retention of episodic information and the details of perceptual analysis (Jacoby and Brooks 1984; Brooks 1978; Tulving and Schacter 1990; Schacter 1990). According to this approach, stimulus variability is considered to be "lawful" and informative to perceptual analysis (Elman and McClelland 1986). Memory involves encoding specific instances, as well as the processing operations used during recognition (Kolers 1973; Kolers 1976). The major emphasis of this view of cognition is on particulars, rather than abstract generalizations or symbolic coding of the stimulus input into idealized categories. Thus, the problems of variability and invariance found in speech perception can be approached in a fundamentally different way by nonanalytic or instance-based accounts of perception and memory.

We believe that the findings from studies on nonanalytic cognition are directly relevant to theoretical questions about the nature of perception and memory for speech and to assumptions about abstractionist representations based on formal linguistic analyses. When the criteria used for postulating episodic or nonanalytic representations are examined carefully, it becomes clear that speech signals display a number of distinctive properties that make them excellent candidates for this approach (Jacoby and Brooks 1984; Brooks 1978). These criteria are summarized below.

High Stimulus Variability

Speech signals display a great deal of physical variability primarily because of factors associated with the production of spoken language. Among these factors are within- and between-talker variability, changes in speaking rate and dialect, differences in social contexts, syntactic, semantic, and pragmatic effects, and emotional state, as well as a wide variety of effects caused by the ambient environment such as background noise, reverberation, and microphone characteristics (Klatt 1986). These diverse sources of variability produce large changes in the acoustic–phonetic properties of speech, and they need to be accommodated in theoretical accounts of the categorization process in speech perception.

Complex Category Relations

The use of phonemes as perceptual units in speech perception entails a set of complex assumptions about category membership. These assumptions are based on linguistic criteria involving such principles as complementary distribution, free variation, and phonetic similarity. In traditional taxonomic linguistics, for example, the concept of a phoneme is used in a number of different ways, as shown by the following definitions from Gleason (1961):

> The phoneme is the minimum feature of the expression system of a spoken language by which one thing that may be said is distinguished from any other thing which might have been said.
>
> A phoneme is a class of sounds...There is no English phoneme which is the same in all environments, though in many phonemes the variation can easily be overlooked, particularly by a native speaker.
>
> A phoneme is a class of sounds which: (1) are phonetically similar and (2) show certain characteristic patterns of distribution in the language or dialect under consideration.
>
> A phoneme is one element in the sound system of a language having a characteristic set of interrelationships with each of the other elements in that system.

> The phoneme cannot, therefore, be acoustically defined. The phoneme is instead a feature of language structure. That is, it is an abstraction from the psychological and acoustical patterns which enables a linguist to describe the observed repetitions of things that seem to function within the system as identical in spite of obvious differences...The phonemes of a language are a set of abstractions...

Thus, speech sounds display complex category relations that place a number of strong constraints on the class of models that can account for these operating principles.

Incomplete Information

Spoken language is a highly redundant symbolic system that has evolved to maximize transmission of linguistic information. In the case of speech perception, research has demonstrated the existence of multiple speech cues for almost every phonetic contrast. Although these speech cues are, for the most part, highly context-dependent, they also provide information that can facilitate comprehension of the intended message when the signal is presented under degraded conditions. This feature of speech perception permits very high rates of information transmission, even under poor listening conditions.

High Analytic Difficulty

Speech is inherently multidimensional in nature. As a consequence, many quasi-independent articulatory attributes can be mapped onto the phonological categories of a specific language. Because of the complexity of speech and the high acoustic–phonetic variability, the category structure of speech is not amenable to simple hypothesis testing. As a result, it has been extremely difficult to formalize a set of explicit rules that can successfully map speech cues onto discrete phoneme categories. The perceptual units of speech are also highly automatized. The underlying category structure of a language is learned in a tacit and incidental way by young children.

Relations Among Perception, Production, and Acoustics

Among category systems, speech appears to be unique because of the close relations between production and perception. Speech exists simultaneously in three very different domains: the acoustic domain, the articulatory domain, and the perceptual domain. Although the relations among these three domains are complex, they are not arbitrary. The sound contrasts used in a language function within a common linguistic system that is assumed to encompass both production and perception. Thus, the phonetic contrasts generated in speech production by the vocal tract are precisely the same acoustic differences that are distinctive in perceptual analysis (Stevens 1972). As a result, any theoretical account of speech perception must also take into consideration aspects of speech production and acoustics.

In learning the sound system of a language, children must not only develop abilities to discriminate and identify sounds, but they must also be able to control the motor mechanisms used in articulation to generate precisely the same phonetic contrasts in speech production that they have become attuned to in perception. One reason that the developing perceptual system might preserve very fine phonetic details, as well as the specific characteristics of the talker's voice, would be to allow young children to imitate accurately and reproduce speech patterns heard in their surrounding language-learning environment (Studdert-Kennedy 1983). This skill would provide children with an enormous benefit in acquiring the phonology of the local dialect from speakers they are exposed to early in life.

In summary, when properties of speech are examined closely, it becomes plausible to assume that very detailed information about specific instances in speech perception might be stored in memory. In contrast to a symbolic rule-based approach, listeners may store a very large number of instances and then use them in an analogical rather than analytic way to categorize novel stimuli (Brooks 1978; Whittlesea 1987). Recent findings from studies on talker variability in speech perception support this conclusion.

TALKER VARIABILITY IN SPEECH PERCEPTION

We have carried out a number of experiments to study the effects of different sources of variability on speech perception and spoken word recognition (Pisoni 1990). Instead of reducing or eliminating variability in the stimulus materials, as most speech researchers have routinely done, we specifically introduced variability from different talkers to study the effects of these variables on perception (Pisoni 1992a). Our research on this problem began with the observations of Mullennix, Pisoni, and Martin (1989) who found that the intelligibility of isolated spoken words presented in noise was affected by the number of talkers that were used to generate the test words in the stimulus ensemble. In one condition, all the words in a test list were produced by a single talker; in another condition, the words were produced by 15 different talkers, including male and female voices. The results were very clear. Across three different signal-to-noise ratios, identification performance was always better for words that were produced by a single talker than words produced by multiple talkers. Trial-to-trial variability in the speaker's voice apparently affected recognition performance. These findings replicated results originally reported by Peters (1955) and Creelman (1957) and suggested that the perceptual system must engage

in some form of "recalibration" each time a novel voice is encountered during the set of test trials using multiple voices.

In a second experiment, we measured naming latencies to the same words presented in both test conditions (Mullennix, Pisoni, and Martin 1989). We found that subjects were not only slower to name words presented in multiple-talker lists but they were also less accurate when their performance was compared to words from single-talker lists. Both sets of findings were surprising at the time, because all the test words used in the experiment were highly intelligible when presented under quiet listening conditions. The intelligibility and naming data immediately raised a number of additional questions about how the various perceptual dimensions of the speech signal are processed and encoded by the human listener. At the time, we naturally assumed that the acoustic attributes used to perceive voice quality were independent of the linguistic properties of the signal. However, no one had ever tested this assumption directly.

In another series of experiments we used a speeded classification task to assess whether attributes of a talker's voice were perceived independently of the phonetic form of the words (Mullennix and Pisoni 1990). Subjects were required to attend selectively to one stimulus dimension (e.g., voice) while simultaneously ignoring another stimulus dimension (e.g., phoneme). Across all conditions, we found increases in interference from both perceptual dimensions when the subjects were required to attend selectively to only one of the stimulus dimensions. The pattern of results suggested that words and voices were processed as integral dimensions; that is, the perception of one dimension (e.g., phoneme) affects classification of the other dimension (e.g., voice) and vice versa, and subjects cannot selectively ignore irrelevant variation on the nonattended dimension. If both perceptual dimensions were processed separately, as we originally assumed, we should have observed little, if any, interference from the nonattended dimension. Not only did we find mutual interference, suggesting that the two dimensions, voice and phoneme, were perceived in a mutually dependent manner, but we also found that the pattern of interference was asymmetrical. It was easier for subjects to ignore irrelevant variation in the phoneme dimension when their task was to classify the voice dimension than it was to ignore the voice dimension when they had to classify the phonemes.

The results from these perceptual experiments were surprising given our assumption that the indexical and linguistic properties of speech are perceived independently. To study this problem further, we carried out a series of memory experiments to assess the neural representation of speech in long-term memory. Experiments on serial recall of lists of spoken words by Martin et al. (1989) and Goldinger et al. (1991) demonstrated that specific details of a talker's voice are also encoded into long-term memory. Using a continuous recognition memory procedure,[1] Palmeri et al. (1993) found that detailed episodic information about a talker's voice is also encoded in memory and is available for explicit judgments, even when a great deal of competition from other voices is present in the test sequence.

Finally, in another set of experiments, Goldinger (1992) found very strong evidence of implicit memory for attributes of a talker's voice that persists for a relatively long time after perceptual analysis has been completed. He also showed that the degree of perceptual similarity between voices affects the magnitude of the repetition effect in several implicit memory tasks. For example, he found that subjects identified spoken words more accurately when they were repeated using the same voice they had originally been presented in than when they were repeated in a different voice. These findings suggest that the perceptual system encodes very detailed talker-specific information about spoken words in episodic memory representations.

Taken together, our findings on the effects of talker variability in perception and memory tasks provide support for the proposal that detailed perceptual information about a talker's voice may be preserved in some type of perceptual representation system (PRS) (Schacter 1990) and that these attributes are encoded implicitly into long-term memory. At the present time, it is not clear whether there is one composite representation in memory or whether these different attributes are encoded in parallel in separate representations (Eich 1982; Hintzman 1986). It is also not clear whether spoken words are encoded and represented in memory as a sequence of abstract symbolic phoneme-like units along with much more detailed episodic information about specific instances and the processing operations used in perceptual analysis. These are important questions for future research on the internal representation of speech in memory.

These recent findings on talker variability have encouraged us to examine more carefully the tuning or adaptation that occurs when a listener becomes familiar with the voice of a specific talker (Nygaard, Sommers, and Pisoni 1994). This particular problem has not received very much attention despite the obvious relevance to problems of speaker normalization, acoustic–phonetic invariance, and the potential

[1]The continuous recognition memory task is an adaptation of the standard recognition memory paradigm. It differs from the standard paradigm in that there is no discrete study phase or test phase. Instead, subjects in a continuous recognition memory task decide on every trial whether a particular item has been presented before during the experiment. Typically, items are repeated at several different fixed intervals or "lags." Accuracy and latency are the dependent variables and are measured as a function of lag (Shepard and Teghtsoonian 1961).

application to automatic speech recognition and speaker identification (Kakehi 1992; Fowler 1990). Our search of the research literature on talker adaptation revealed only a small number of behavioral studies on this topic, and all of them appeared in obscure technical reports from the mid-1950s.

To determine how familiarity with a talker's voice affects the perception of spoken words, we had two groups of listeners learn to identify explicitly a set of unfamiliar voices over a 9-day period using common names (i.e., Bill, Joe, Sue, Mary). After the subjects learned to recognize the voices, we presented them with a set of novel words mixed in noise at several signal-to-noise ratios; one group heard the words produced by talkers on whom they were previously trained, the other group heard the same words produced by new talkers to whom they had not been previously exposed. In this phase of the experiment, which was designed to measure speech intelligibility, subjects were required to identify the words rather than recognize the voices, as they had done in the first phase of the experiment.

The results of the intelligibility experiment are shown in Figure 1 for the two groups of subjects. We found that identification performance for the trained group was reliably better than the control group at each of the signal-to-noise ratios tested. The subjects who had heard novel words produced by familiar voices were able to recognize words more accurately than subjects who received the same novel words produced by unfamiliar voices. Two other groups of subjects were also tested in the intelligibility experiment as controls; however, these subjects did not receive any training in recognizing the voices and were, therefore, not exposed to any of the stimuli prior to listening to the same set of words in noise. One control group received the set of words presented to the trained experimental group; the other control group received the words that were presented to the trained control subjects. The performance of these two control groups was not only the same, but was also equivalent to the intelligibility scores obtained by the trained control group. Thus, only the subjects in the experimental group who were explicitly trained on the voices showed an advantage in recognizing novel words produced by familiar talkers.

The findings from this perceptual learning experiment demonstrate that exposure to a talker's voice facilitates subsequent perceptual processing of novel words produced by the same talker. Thus, speech perception and spoken word recognition draw on highly specific perceptual knowledge about a talker's voice that is obtained in an entirely different experimental task— explicit voice recognition as compared to a speech intelligibility test.

What kind of perceptual knowledge does a listener acquire when he listens to a speaker's voice and is required to carry out an explicit

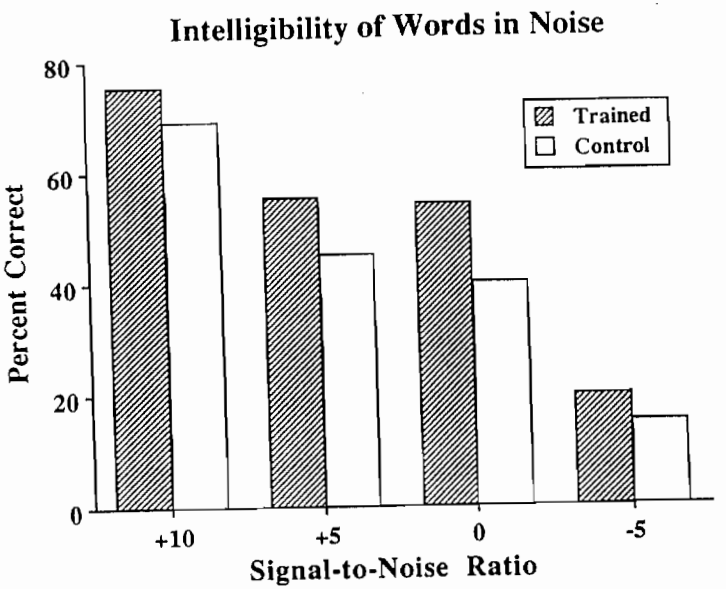

Figure 1. Percentage correct word recognition (intelligibility) as a function of signal-to-noise ratio for the trained and control subjects on the transfer task administered after voice recognition training was completed (from Nygaard, Sommers, and Pisoni 1994).

name recognition task as our subjects did in this experiment? One possibility is that the procedures or perceptual operations (Kolers 1973) used to recognize the voices are retained in some type of "procedural memory," and these analysis routines are reinvoked when the same voice is encountered in a subsequent intelligibility test. This kind of procedural knowledge might increase the efficiency of the perceptual analysis for novel words produced by familiar talkers, because detailed analysis of the speaker's voice would not have to be carried out over and over again as each new word was encountered. Another possibility is that specific instances—perceptual episodes or exemplars of each talker's voice—are stored in memory and then later retrieved during the process of word recognition when new tokens from a familiar talker are presented (Jacoby and Brooks 1984).

Whatever the exact nature of this knowledge turns out to be, the important point to emphasize here is that prior exposure to a talker's voice facilitates subsequent recognition of novel words produced by the same talkers. Such findings demonstrate a form of implicit memory for a talker's voice that is distinct from the retention of the individual items used and the specific task that was employed to familiarize the listeners with the voices (Schacter and Church 1992; Roediger 1990).

These findings provide additional support for the view that the neural representation of spoken words encompasses both a phonetic description of the utterance *and* information about the structural description of the source characteristics of the specific talker. Thus, speech perception appears to be carried out in a "talker-contingent" manner; indexical and linguistic properties of the speech signal are apparently closely interrelated and are not dissociated in perceptual analysis.

ROLE OF STIMULUS VARIABILITY IN PERCEPTUAL LEARNING OF /ɹ/ AND /l/

Developments in nonanalytic approaches to cognition and exemplar-based approaches to categorization have led us to reconsider a number of issues related to the acquisition of new phonetic categories. In particular, we have become interested in examining the contribution of stimulus variability to the formation of robust perceptual categories. In this section, we summarize the results of two experiments designed to investigate several issues in perceptual learning of /ɹ/ and /l/ by Japanese listeners. First, we examined the role of talker variability in training. To study this, we compared the performance of listeners who were trained with only a single voice to the performance of listeners who were trained with several talkers. Second, we examined the retention of new phonetic categories over time. The issue of retention without additional training is important, because it allowed us to assess the ultimate success of our training procedure by examining how new information is retained in memory.

Several aspects of our general training strategy are important to emphasize before discussing any results. One of the goals of cross-language training experiments is to facilitate acquisition of robust new phonetic categories. Two criteria are important in defining robust perceptual categories: First, the categories must be applied across a wide variety of new talkers and new phonetic environments. This means that listeners must demonstrate generalization both to new talkers and to novel words. Second, the new categories must be stable over time. In other words, if listeners form robust categories during training, then their performance should be above baseline levels after extended intervals without any further training.

Another important aspect of our approach to perceptual learning concerns how new categories are formed. The traditional approach to cross-language speech perception training has been to use synthetic stimuli that vary only in the critical cues used by native speakers of the language (see, however, Yamada and Tohkura 1992). Although this approach has been successful in some cases in modifying speech perception (see Pisoni et al. 1982; McClaskey, Pisoni, and Carrell 1983), it overlooks the rich diversity of acoutic cues that are present in natural

speech. In both of the experiments we describe below, listeners were trained with natural speech tokens. We hypothesized that the richness and diversity of cues found in natural speech would aid listeners in forming robust new perceptual categories that could be generalized to new phonetic environments and new talkers (see also Kuhl 1983). We also assumed that by training some listeners with tokens from multiple talkers, we would maximize the number of cues that listeners had available to them in recognizing new words.

Task variables also play an important role in training non-native listeners to perceive new phonetic contrasts (Jenkins 1979; see also Logan and Pruitt this volume). Two types of tasks, discrimination and identification, have traditionally been used during training. In discrimination training, listeners are presented with sequences of stimuli and are asked to decide if the stimuli are the same or different or if a member of the stimulus ensemble is unique. The assumption is that listeners will focus their attention on cues that contrast new perceptual categories. The drawback of this approach is that listeners' attention may be focused too sharply on any stimulus differences they can detect. Instead of responding to higher, more abstract category-level differences, subjects may respond to subtle, within-category changes that differentiate *stimuli* rather than *categories*. Thus, discrimination training encourages listeners to attend to small, within-category differences and does not promote the formation of new perceptual categories that are robust to the variability in the natural environment (Carney, Widin, and Viemeister 1977; Jamieson and Morosan 1986, 1989; Liberman et al. 1961; Pisoni 1973; Strange and Dittmann 1984; Werker and Logan 1985; Werker and Tees 1984).

In identification training, on the other hand, subjects are asked to identify explicitly a single stimulus on each trial. Uncertainty in the task is controlled by restricting the number of possible response alternatives. Whereas discrimination tasks require listeners to attend to small within-category differences, identification training encourages subjects to group perceptually similar objects into the same category (Lane 1965, 1969). Thus, discrimination training promotes "acquired distinctiveness"; whereas, identification training promotes "acquired equivalence" (Lawrence 1949, 1950; Gibson and Gibson 1955; Liberman et al. 1961).

In both of the experiments described below, we employed a pretest–posttest design that was identical to the tests used by Strange and Dittmann (1984). Training was conducted over a 15-day period using a two-alternative forced-choice identification training procedure. Immediate feedback was given only during training. All training and testing was conducted using natural speech. In the first experiment, subjects were trained with only a single talker. In the second experiment,

subjects were trained with five different talkers. After the completion of training, subjects were given two tests of generalization. One test consisted of new words produced by a familiar talker; the other test consisted of novel words produced by an unfamiliar talker.

Training with a Single Talker

The motivation for training listeners with a single talker comes from recent findings of Logan, Lively, and Pisoni (1991). In their experiment, Japanese listeners were trained to perceive English /ɹ/ and /l/ using a two-alternative forced-choice identification procedure. Subjects were trained using five different talkers who produced English words containing /ɹ/ and /l/ in five different phonetic environments. The authors found that listeners' accuracy improved by 5%–7% from the pretest to the posttest, as well as during training. In addition, they found a marginal difference in accuracy between talkers during the tests of generalization. The familiar talker was responded to slightly more accurately than the unfamiliar talker, although the generalization results were obtained with only three subjects. Based on these preliminary results, Logan, Lively, and Pisoni (1991) concluded that the high-variability identification paradigm was effective in training Japanese listeners to acquire the /ɹ/–/l/ contrast.

Logan, Lively, and Pisoni's training study included two sources of variability. First, /ɹ/ and /l/ appeared in several phonetic environments. We assumed that variability within a single phonetic environment might not be sufficient to foster generalization to /ɹ/s and /l/s in other phonetic environments (Jamieson and Morosan 1986, 1989). Second, we presented listeners with tokens from multiple talkers during training. We assumed that this would provide listeners with a rich set of cues to the new contrast. Moreover, this procedure would prevent listeners from becoming attuned too closely to a particular voice (Goto 1971). Training with multiple talkers was also thought to encourage generalization to new voices.

Because we included several sources of variability in the training stimuli, it is not clear what the relative contributions of each source of variability were to the observed pattern of results (Logan, Lively, and Pisoni 1993; Pruitt 1993). Recently, we conducted an experiment to determine more precisely how talker variability affected performance during training and generalization to new words and new talkers (Lively, Logan, and Pisoni 1993). We trained a group of six Japanese listeners in a pretest–posttest design with the same two-alternative forced-choice identification paradigm used in the earlier study by Logan et al. The only difference was that listeners were trained with only a single talker, rather than five different talkers. Subjects were trained for 15 days on the same set of 136 words that contained /ɹ/ or /l/ in five phonetic environments (initial singleton, initial consonant clusters, intervocalic, final consonant clusters, and final singleton positions). Generalization to new words and to a new talker were also assessed at the conclusion of training.

We predicted that the reduction in talker variability should have several consequences on identification performance. First, increases in accuracy and decreases in response latency should be observed during training. This result would not be surprising, given that we had already observed changes in performance with a much more variable stimulus set. Second, generalization should be adversely affected by the reduction in talker variability. If listeners become attuned to the specific characteristics of a particular voice during perceptual learning (Goto 1971), then they would not be expected to generalize very well to a new talker used in the generalization tests. Moreover, if subjects are learning about specific stimuli rather than general cues or rules to the new contrast, then they would not be expected to generalize very well to novel stimuli produced by a familiar talker.

The results of the single-talker experiment confirmed our predictions. As shown in Figure 2, listeners' accuracy increased significantly from the pretest to the posttest for these contrasts in most phonetic environments. Response times also decreased significantly from the pretest to the posttest for phonemes in all phonetic environments. During training, subjects' accuracy increased and response latencies decreased from Week 1 of training to Week 2 of training. No significant changes in performance were observed between Week 2 and Week 3 of training. Accuracy improved the most for /ɹ/s and /l/s in initial consonant clusters and intervocalic position.

The results of the tests of generalization which are shown in Figure 3 revealed the limitations of the single-talker training procedure. Listeners responded more accurately to words produced by the familiar talker when the /ɹ/s or /l/s occurred in initial singleton or intervocalic positions. A trend was also observed for better performance with the familiar talker when /ɹ/s or /l/s were in initial consonant clusters. Responses also tended to be faster to the familiar talker. However, absolute level of performance on both tests of generalization was relatively low. Mean accuracy with the familiar talker was equivalent to the level of performance observed during the first week of training. Similarly, mean accuracy with the unfamiliar talker was worse than performance during the first week of training for contrasts in initial singleton, initial consonant clusters, and intervocalic environments. These findings demonstrate that when listeners are trained with only a single talker they do not generalize very well to new words produced by a new talker or to new words produced by the old talker used in training.

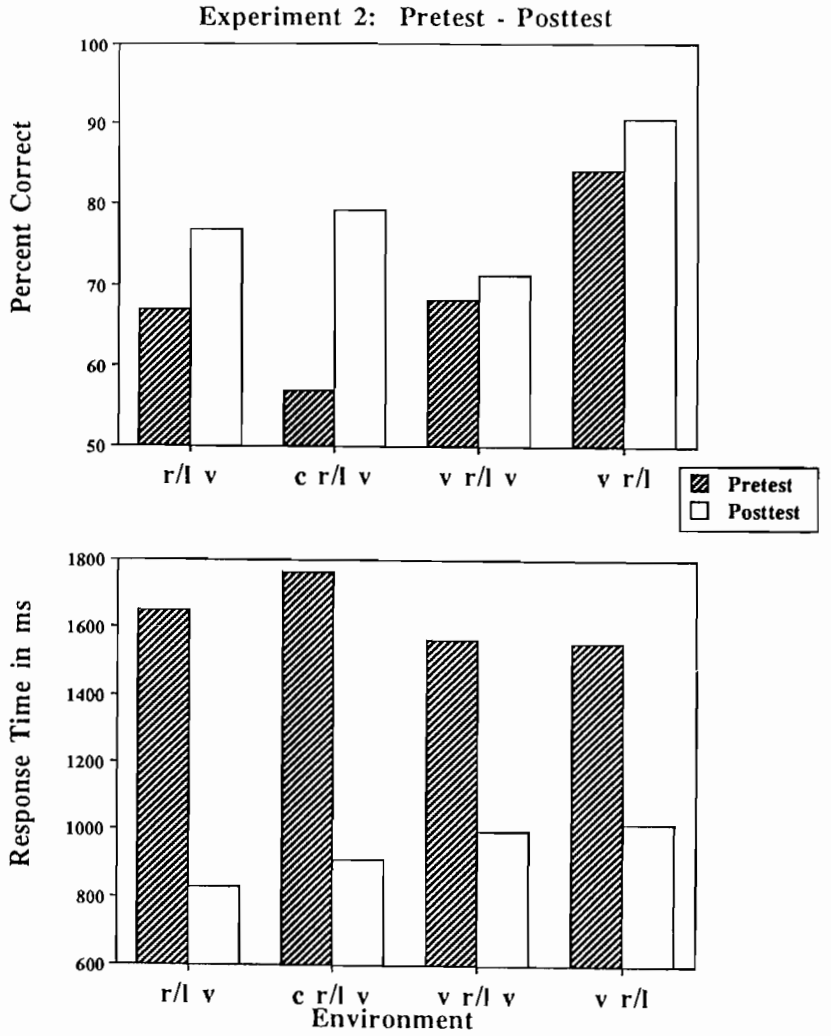

Figure 2. The upper panel shows percentage correct identification from the single-talker training condition on the pretest and posttest. The filled bars show the results from the pretest. The open bars show the results from the posttest. The lower panel shows response times (from Lively, Logan, and Pisoni 1993).

Taken together, the results of this experiment indicate that the single-talker training paradigm was generally less effective in facilitating robust acquisition of /ɹ/ and /l/. Although subjects encoded some stimulus-specific knowledge, they did not seem to be able to apply this knowledge in the generalization tests with novel words produced by novel talkers. The results support Goto's (1971) observations that Japanese

listeners become attuned to a small set of voices when they acquire English and that more extensive training with different voices is required for robust generalization to new words produced by new talkers.

Long-term Retention of New Phonetic Categories

Our initial training study demonstrated that Japanese listeners could be trained in the laboratory to perceive the English /ɹ/–/l/ contrast. Moreover, the results demonstrated the importance of talker variability to generalization to new words and new voices. In the experiment using only a single talker, we found that an absence of talker variability in the stimulus set used during training was detrimental to generalization performance. The outcome of these two experiments jointly satisfy one criterion for a successful training paradigm. Trained listeners should demonstrate generalization to both new voices and new words. Logan, Lively, and Pisoni's results suggest that training with moderate amounts of talker variability in the stimulus ensemble encourages generalization; whereas, training with only a single talker does not (see also Lively et al. 1994, Exp. 1).

Neither Logan, Lively, and Pisoni's (1991) investigation, nor the single-talker experiment described above addressed the second criterion for a successful training procedure. Any robust training paradigm should also encourage the long-term retention of new phonetic contrasts.

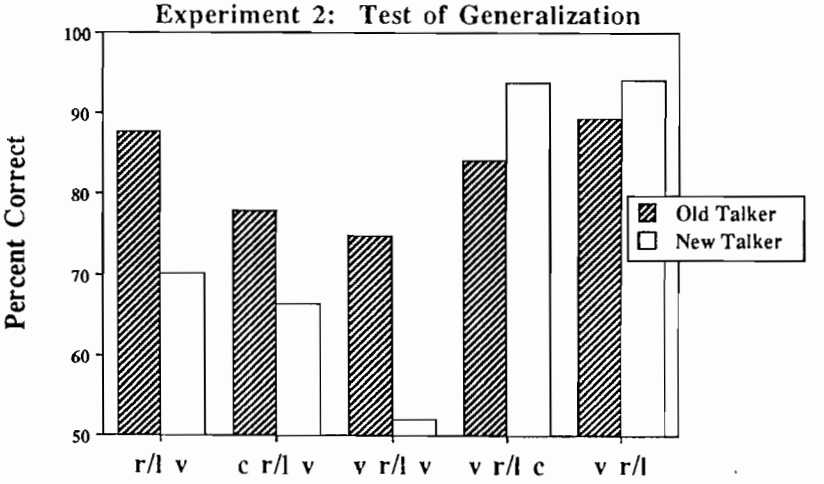

Figure 3. Percentage correct identification from the single-talker training condition on the test of generalization is shown as a function of phonetic environment. The filled bars show the results for novel words produced by a familiar talker; the open bars show results for a new talker. The data also reveal a significant interaction between talker and phonetic environment (from Lively, Pisoni, and Logan 1993).

It is important to determine whether short-term laboratory-based procedures produce only temporary reorganization of a listener's perceptual capabilities or whether these changes are more permanent. Two predictions concerning the retention of new phonetic categories can be made. First, it is possible that changes observed in the laboratory may be short-lived. If listeners are living in a monolingual speaking environment in which the new linguistic contrast is rarely encountered, then subjects might be expected to return to baseline levels of performance without any further training or exposure. On the other hand, if our training procedures encourage the development of long-term changes in perceptual organization, then listeners might be expected to retain much of what they learned during training without additional training or feedback.

We recently assessed these predictions in collaboration with researchers at the ATR laboratories in Kyoto, Japan (Lively et al. 1994). Nineteen monolingual speakers of Japanese were trained using a slightly modified version of Logan, Lively, and Pisoni's high-variability training procedure. Subjects were trained with exactly the same stimuli used by Logan and colleagues. Five talkers produced the /r/–/l/ contrast in five different phonetic environments. Subjects were tested in a pretest–posttest design. Training lasted for 15 days, and listeners heard one training talker each day. By the end of training, listeners had heard each talker three times. Following the conclusion of training, listeners were given the same tests of generalization described earlier. Three months after the conclusion of training, subjects returned to the laboratory and were given a follow-up posttest and the two tests of generalization again. As in our previous experiments, a two-alternative forced-choice identification task was used throughout the entire experiment and feedback was given only during training.

For the most part, the overall pattern of results replicated those obtained in our original study (Logan, Lively, and Pisoni 1991). Subjects improved from the pretest to the posttest by 12% overall. During training, listeners' accuracy increased by an average of approximately 11%, and response times decreased by approximately 600 ms. The increases in accuracy were almost twice as large as those obtained by Logan, Lively, and Pisoni. The difference in the size of the training effects between the two studies may be because in our earlier study, we tested subjects who were living in the United States at the time of the experiment and were also enrolled in English classes. Thus, our subjects may have received extensive exposure to the /ɹ/ and /l/ contrast outside of the laboratory before the training procedures began, and this may have affected their ability to show additional improvements in the laboratory environment. In contrast, the subjects in the present study were living in a monolingual Japanese-speaking environment, and it is highly unlikely that they received any exposure to

this contrast in their immediate surroundings. Given that pretest levels of performance were generally higher for Logan, Lively, and Pisoni's subjects, it is unlikely that they could have observed an improvement as large as the one obtained in the present experiment.

The results of the tests of generalization showed that familiarity with the talker producing the stimuli facilitated identification performance. Subjects were significantly more accurate when words were produced by a talker used in training than when the words were produced by an unfamiliar talker. In terms of absolute level of performance, generalization accuracy was quite good. Average performance with the familiar talker was equivalent to mean accuracy during the third week of training. Similarly, accuracy with the unfamiliar talker was equivalent to mean performance during the second week of training.

Because we observed large differences in performance among talkers used during training and we had selected the most intelligible talker to use during the tests of generalization, we wanted to assess the differences in base-line intelligibility as the source of the results obtained during generalization. We tested this hypothesis by having an additional 14 naive Japanese listeners perform the tests of generalization without any prior training. No significant differences in intelligibility were observed: Mean accuracy with the "familiar" talker was 71%; whereas, mean accuracy with the "unfamiliar" talker was 70%. These results rule out simple differences in intelligibility between the familiar and unfamiliar talkers as the source of our generalization results.

The most interesting data from the retention experiment come from the follow-up tests given 3 months after the conclusion of the original training. In these tests, 16 of the original 19 subjects returned and were given the posttest and the two tests of generalization again. The posttest results are shown in Figure 4. Surprisingly, mean accuracy decreased only 2% from the posttest given at the end of training to the follow-up posttest given three months later. This decrease was not statistically significant. A similar pattern was obtained in the follow-up tests of generalization shown in Figure 5. Mean accuracy decreased only 1.5% from the original generalization tests to the follow-up tests. Interestingly, the effect of talker was still significant, even after a 3-month interval. Words produced by a familiar talker, the talker used during training, were identified more accurately than words produced by an unfamiliar talker.

The results of the retention experiment are theoretically important for several reasons. First, the present findings demonstrate that the high-variability identification training paradigm meets our second criterion for successful training procedures: Listeners show long-term retention of new perceptual categories without any further training. To our knowledge, these results are the first demonstration of long-term retention of new phonetic categories acquired in a laboratory training experiment.

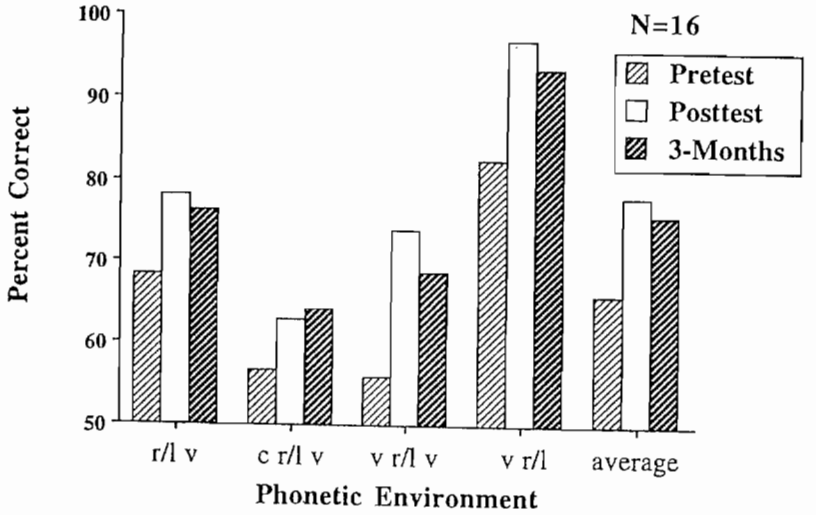

Figure 4. Accuracy scores for monolingual Japanese listeners for the pretest, posttest, and 3-month follow-up as a function of phonetic environment (from Lively et al. 1994).

Second, our findings provide support for the nonanalytic approach to cognition outlined above. We suggest that listeners encode detailed representations of spoken words into long-term memory and that these representations are used to facilitate the later recognition of new items. These representations are assumed to include attributes of a talker's voice. During the tests of generalization, subjects were more accurate in responding to words produced by a talker used in training than to words produced by an unfamiliar talker. Moreover, these differences were still evident 3 months after the conclusion of training. The pattern of results cannot be accounted for by differences in base-line intelligibility between the two talkers. Precisely what characteristics of the stimulus materials are retained or how long this information is preserved remains an important question for future research.

Taken together, the results of the training experiments described above suggest several important methodological and theoretical conclusions. First, the present findings indicate that adult Japanese listeners can be trained to identify the English /ɹ/–/l/ contrast. Second, stimulus variability, particularly talker variability, appears to be an important factor in promoting robust generalization. When listeners were trained with a single talker, generalization to new words and a new talker was poor. In contrast, when listeners were trained with a more variable stimulus set that included several talkers, generalization improved substantially. Third, training with a high-variability, two-alternative forced-choice training paradigm meets both of the criteria we outlined

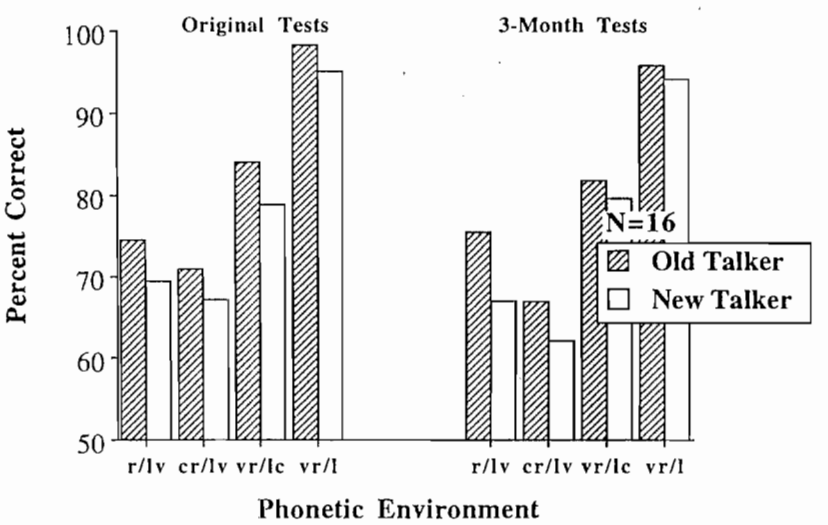

Figure 5. Percentage correct identification on the tests of generalization. The left panel shows the results immediately after training; the right panel shows the results after a 3-month retention interval from the follow-up tests (from Lively et al. 1994).

for successful training procedures. Subjects show generalization to new words and new talkers, and they demonstrate retention of the new phonetic contrast over time. It should be noted, however, that generalization was incomplete. We still observed differences among talkers and some loss over time. It is possible that other training procedures may be more effective in promoting long-term retention and generalization of new phonetic contrasts. Finally, we suggest that the results of our high-variability laboratory-based training procedures provide evidence for the encoding of talker-specific information in speech perception and that this information is used to facilitate the recognition of new words. Thus, for our listeners, variability was highly informative in helping them to develop robust perceptual categories (Elman and McClelland 1986).

SUMMARY AND CONCLUSIONS

We have covered a lot of ground in this chapter in our attempts to bring together a number of different but closely related areas of research that bear on several long-standing theoretical issues dealing with categorization in speech perception. Here we summarize the major findings and draw several conclusions.

One of the most salient findings to emerge from this research is the importance of stimulus variability in perceptual learning of novel

linguistic contrasts. In contrast with other investigations that have used a low-variability stimulus set during training, we found moderate but highly consistent changes in performance over the course of training. Moreover, our subjects displayed generalization to new words and to tokens produced by a new talker. We suggest that these findings are attributable to the high-variability training procedure that promotes the development of robust perceptual categories (Jamieson and Morosan 1986, 1989).

Another important finding was the observation that listeners acquire information about these perceptual categories by encoding specific instances or exemplars from the stimulus ensemble, including details of individual voices. The subjects in our experiments did not appear to acquire abstract context-independent categories for /ɹ/ and /l/ that were invariant across different talkers. The perceptual learning that took place using these laboratory-training procedures was stimulus-specific, although evidence for generalization was found, possibly via analogy, when the stimulus set contained a great deal of variability.

These findings raise a number of important theoretical issues in speech perception that go well beyond the specific demonstration that Japanese listeners can be trained to perceive English /ɹ/ and /l/ reliably. Our findings from this training study and our other experiments on talker variability show that several long-held views about speech perception may be incorrect (see also Elman and McClelland 1986; Schacter and Church 1992). The emphasis on mapping of speech onto discrete symbolic units has drawn many researchers away from the fundamental question of how the perceptual categories of speech and their mental representations should be conceptualized theoretically, given the enormous variability in the physical signal. The traditional idealized view of speech has also encouraged an approach and methodology for conducting research that is designed to deliberately reduce or eliminate as many sources of variability as possible in the stimulus materials in order to obtain reliable perceptual data from listeners in a wide variety of experimental paradigms. The underlying assumption of this approach to research in speech is that stimulus variability is a source of noise—something to be eliminated from the signal so that the perceptual invariants for the idealized linguistic categories would somehow emerge.

The present results on perceptual learning demonstrate that variability is lawful and informative. Listeners encode and retain detailed stimulus information from the signal. When viewed within the context of nonanalytic approaches to cognition, variability in speech is simply a natural consequence of the complex category structure of spoken language. Attempts to reduce or eliminate stimulus variability in perceptual and memory experiments on spoken language over the last 40

years may have provided a misleading or distorted picture of the underlying perceptual process, which appears to be able to cope quite well with these diverse sources of variability in the speech signal.

Our findings also raise several theoretical issues about the neural representation of speech and the types of information in the signal that listeners preserve. Several results show that detailed information about a talker's voice is encoded into long-term memory and is used in speech perception and spoken word recognition. In past accounts, indexical information about the speaker's voice was clearly dissociated from properties of the linguistic message (Liberman and Mattingly 1985). The present findings demonstrate that some sources of indexical information are encoded into memory and do become part of the representation of spoken words (Laver and Trudgill 1979).

The present findings are also relevant to several long-standing assumptions about the perceptual normalization for speech, particularly claims about the loss of stimulus-specific information. The evidence we have obtained in our variability and perceptual learning experiments suggests that the process of talker normalization may not be carried out automatically without cost and that information may not be lost as a consequence of perceptual analysis and categorization. The locus of the talker normalization process, if one actually exists, may not be in the auditory periphery, as many researchers have assumed in the past, but may be more centrally located in the recognition process itself, which draws heavily on specific knowledge in long-term memory for categorization.

In summary, we suggest that the traditional approach to speech perception has been somewhat misguided with regard to the nature of the perceptual operations that occur when listeners process spoken language. Variability may not be noise. Rather, it appears to be informative to perception. We have briefly reviewed the results of several studies that have demonstrated the encoding and retention of talker-specific details in speech perception. We believe that these studies point to important new directions in speech perception research in which variability, rather than invariance, is regarded as an important problem for study. This approach to speech perception leads to the view that the perceptual categories in speech must be adaptive, dynamic, and extremely flexible in order to accommodate the changing stimulus environment that is one of the most distinctive characteristics of speech production and perception.

ACKNOWLEDGMENTS

This research was supported in part by NIH Research Grant DC-00111-15 and in part by NIDCD Research Training Grant DC00012-14 to Indiana University in Bloomington, IN.

REFERENCES

Barsalou, L. W. 1993. Flexibility structure and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In *Theories of Memories*, eds. A. C. Collins, S. E. Gathercole, M. A. Conway, and P. E. M. Morris. Hillsdale, NJ: Erlbaum.

Brooks, L. 1978. Nonanalytic concept formation and memory for instances. In *Cognition and Categorization*, eds. E. Rosch and B. Lloyd. Hillsdale, NJ: Erlbaum.

Carney, A., Widin, G., and Viemeister, N. 1977. Noncategorical perception of stop consonants varying in VOT. *Journal of the Acoustical Society of America* 62:961–70.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J. 1952. Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America* 24:597–606.

Creelman, C. D. 1957. Case of the unknown talker. *Journal of the Acoustical Society of America* 29:655.

Eich, J. E. 1982. A composite holographic associative memory model. *Psychological Review* 89:627–61.

Elman, J. L., and McClelland, J. L. 1986. Exploiting lawful variability in the speech wave. In *Invariance and Variability in Speech Processes*, eds. J. S. Perkell and D. H. Klatt. Hillsdale, NJ: Erlbaum.

Fant, G. 1973. *Speech Sounds and Features*. Cambridge, MA: MIT Press.

Fowler, C. A. 1990. Listener-talker attunements in speech. Haskins Laboratories Status Report on Speech Research SR-101/102:110–29.

Gibson, J. J., and Gibson, E. J. 1955. Perceptual learning: Differentiation or enrichment? *Psychological Review* 62:32–41.

Gleason, H. A. 1961. *An Introduction to Descriptive Linguistics*. New York: Holt, Rinehart & Winston.

Goldinger, S. D.1992. Words and voices: Implicit and explicit memory for spoken words. *Research on Speech Perception Technical Report No. 7*, Indiana University, Bloomington, IN.

Goldinger, S. D., Pisoni, D. B., and Logan, J. S. 1991. On the locus of talker variability effects in recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17:152–62.

Goto, H. 1971. Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia* 9:317–23.

Hintzman, D. L. 1986. Schema abstraction in a multiple-trace memory model. *Psychological Review* 93:411–23.

Jacoby, L. L. and Brooks, L. R. 1984. Nonanalytic cognition: Memory, perception, and concept learning. In *The Psychology of Learning and Motivation*, ed. G. Bower. New York: Academic Press.

Jamieson, D., and Morosan, D. 1986. Training non-native speech contrast in adults: Acquisition of English /ð/–/θ/ contrasts by francophones. *Perception & Psychophysics* 40:205–15.

Jamieson, D., and Morosan, D. 1989. Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology* 43:88–96.

Jenkins, J. J. 1979. Four points to remember: A tetrahedral model of memory experiments. In *Levels of Processing Human Memory*, eds. L. S. Cermak and F. L. M. Craik. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kakehi, K. 1992. Adaptability to differences between talkers in Japanese monosyllabic perception. In *Speech Perception, Production and Linguistic Structure*, eds. Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka. Tokyo, Japan: Ohmsha Publishing Co. Ltd.

Klatt, D. H. 1979. Speech perception: A model of acoustic–phonetic analysis and lexical access. *Journal of Phonetics* 7:279–312.

Klatt, D. H. 1986. The problem of variability in speech recognition and in models of speech perception. In *Invariance and Variability in Speech Processes*, eds. J.S. Perkell and D. H. Klatt. Hillsdale, NJ: Erlbaum.

Kolers, P .A. 1973. Remembering operations. *Memory and Cognition* 1:347–55.

Kolers, P .A. 1976. Pattern analyzing memory. *Science* 191:1280–81.

Kuhl, P .K. 1983. Perception of auditory equivalence classes for speech in early infancy. *Infant Behavioral Development* 6:263–85.

Kuhl, P. K. 1991a. Human adults and human infants show a "perceptual magnet effect" for the prototype of speech categories, monkeys do not. *Perception & Psychophysics* 50:93–107.

Kuhl, P. K. 1991b. Speech prototypes: Studies on the nature, function, ontogeny and phylogeny of the "centers" of speech categories. In *Speech Perception, Production and Linguistic Structure*, eds. Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka. Tokyo: OHM Publishing Co. Ltd.

Lane, H. 1965. The motor theory of speech perception: A critical review. *Psychological Review* 7:275–309.

Lane, H. 1969. A behavioral basis for the polarity principle in linguistics. In *Research in Verbal Behavior and Some Neurological Implications*, eds. K. Salzinger and S. Salzinger. New York: Academic Press.

Laver, J., and Trudgill, P. 1979. Phonetic and linguistic markers in speech. In *Social Markers in Speech*, eds. K. R. Scherer and H. Giles. Cambridge: Cambridge University Press.

Lawrence, D. H. 1949. Acquired distinctiveness in cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology* 39:770–84.

Lawrence, D. H. 1950. Acquired distinctiveness in cues: II. Selective association in a constant stimulus situation. *Journal of Experimental Psychology* 40:175–88.

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. *Journal of Experimental Psychology* 61:379–88.

Liberman, A. M., and Mattingly, I. G. 1985. The motor theory of speech perception revised. *Cognition* 21:1–36.

Lively, S. E., Pisoni, D. B., and Logan, J. S. 1992. Some effects of training Japanese listeners to identify English /r/ and /l/. In *Speech Perception, Production and Linguistic Structure*, ed. Y. Tohkura. Tokyo: Ohmsha Publishing Co. Ltd.

Lively, S. E, Logan, J. S., and Pisoni, D. B. 1993. Training Japanese listeners to identify English /r/ and /l/ II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America* 94:1242–55.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. 1994. Training Japanese listeners to identify English /r/ and /l/ III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America* 96:2076–87.

Logan, J. S., Lively, S. E. and Pisoni, D. B. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America* 89:874–86.

Logan, J. S., Lively, S. E., and Pisoni, D. B. 1993. Training listeners to perceive novel phonetic categories: How do we know what is learned? *Journal of the Acoustical Society of America* 94:1148–51.
Martin, C. S., Mullennix, J. W., Pisoni, D. B., and Summers, W. V. 1989. Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory and Cognition* 15:676–84.
McClaskey, C., Pisoni, D., and Carrell, T. 1983. Transfer of training to a new linguistic contrast in voicing. *Perception &Psychophysics* 34:323–30.
Medin, D. L., and Barsalou, L. W. 1987. Categorization processes and categorical perception. In *Categorical Perception: The Groundwork of Cognition*, ed. S. Harnad. Cambridge: Cambridge University Press.
Mullennix, J. W., and Pisoni, D. B. 1990. Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics* 47:379–90.
Mullennix, J. W., Pisoni, D. B., and Martin, C. S. 1989. Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America* 85:365–78.
Neisser, U. 1967. *Cognitive Psychology*, New York: Appleton-Century-Crofts.
Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. 1994. Speech perception as a talker-contingent process. *Psychological Science* 5:42–46.
Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. 1993. Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19:1–20.
Peters, R. W. 1955. The relative intelligibility of single-voice and multiple-voice messages under various conditions of noise. *Joint Project Report No. 56, U.S. Naval School of Aviation Medicine*, pp. 1–9. Pensacola, FL.
Pisoni, D. B. 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics* 13:253–60.
Pisoni, D. B. 1978. Speech perception. In *Handbook of Learning and Cognitive Processes*, vol. 6, ed. W. K. Estes. Hillsdale, NJ: Erlbaum.
Pisoni, D. B. 1990. Effects of talker variability on speech perception: implications for current research and theory. *Proceedings of 1990 International Conference on Spoken Language Processing*. Kobe, Japan.
Pisoni, D. B. 1992a. Some comments on invariance, variability and perceptual normalization in speech perception. *Proceedings 1992 International Conference on Spoken Language Processing* Banff, Canada, 12–17 October 1992.
Pisoni, D. B. 1992b. Some comments on talker normalization in speech perception. In *Speech Perception, Production and Linguistic Structure*, eds. Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka. Tokyo, Japan: Ohmsha Publishing Co. Ltd.
Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. 1982. Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance* 8:297–314.
Pruitt, J. S. 1993. Comments on "Training Japanese listeners to identify English /r/ and /l/: A first report." *Journal of the Acoustical Socieity of America* 94:1146–47.
Roediger, H. L. 1990. Implicit memory: Retention without remembering. *American Psychologist* 45:1043–56.
Rosch, E. 1975a. Cognitive reference points. *Cognitive Psychology* 7:532–47.
Schacter, D. L. 1990. Perceptual representation systems and implicit memory: Toward a resolution of the multiple memory systems debate. In *Development and Neural Basis of Higher Cognitive Function*, ed. A. Diamond. *Annals of the New York Academy of Sciences*, Vol. 608:543–71.
Schacter, D. L. 1992. Understanding Implicit Memory: A Cognitive Neuroscience Approach. *American Psychologist* 47:559–69.
Schacter, D. L. and Church, B. A. 1992. Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:915–30.
Shepard, R. N., and Teghtsoonian, M. 1961. Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology* 62:302–309.
Smith, E., and Medin, D. 1981. *Categories and Concepts*. Cambridge, MA: Harvard University Press
Stevens, K. N. 1971. Sources of inter- and intra-speaker variability in the acoustic properties of speech sounds. *Proceedings of the Seventh International Congress of Phonetic Sciences*. The Hague: Mouton.
Stevens, K. N. 1972. The quantal nature of speech: Evidence from articulatory acoustic data. In *Human Communication: A Unified View*, eds. E. E. David, Jr., and P .B. Denes. New York: McGraw-Hill.
Stevens, K. N. 1993. Lexical access from features. In *Speech Technology for Man-Machine Interaction*, eds. P. V. S. Rao and B. Kalia. Tata, New Delhi: McGraw-Hill.
Stevens, K. N., and Blumstein, S. E. 1978. Invariant cues for place of articulation in stop-consonants. *Journal of the Acoustical Society of America* 64:1358–68.
Stevens, K. N., and Blumstein, S. E. 1980. The search for invariant acoustic correlates of phonetic features. In *Perspectives on the Study of Speech*, eds. P. D. Eimas and J. L. Miller. Hillsdale, NJ: Erlbaum.
Strange, W., and Dittmann, S. 1984. The effects of discrimination training on the perception of /r–l/ by Japanese adults learning English. *Perception & Psychophysics* 32:131–45.
Studdert-Kennedy, M. 1974. The perception of speech. In *Current Trends in Linguistics*, ed. T. A. Sebeok. The Hague: Mouton.
Studdert-Kennedy, M. 1983. On learning to speak. *Human Neurobiology* 2:191–95.
Sussman, H. M., McCaffrey, H. A., and Matthews, S. A. 1991. An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America* 90:1309–25.
Tulving, E., and Schacter, D. L. 1990. Priming and human memory systems. *Science* 247:301–306.
Werker, J., and Logan, J. 1985. Cross-language evidence for three factors in speech perception. *Perception & Psychophysics* 37:35–44.
Werker, J., and Tees, R. 1984. Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America* 75:1866–78.
Whittlesea, B. W. A. 1987. Preservation of specific experiences in the representation of general knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13:3–17.
Yamada, R. A., and Tohkura, Y. 1991. Perception of American English /r/ and /l/ by native speakers of Japanese. In *Speech Perception, Production and Linguistic Structure*, eds. Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka. Tokyo: Ohmsha Publishing Co. Ltd.
Yamada, R., and Tohkura, Y. 1992. The effects of experimental variables in the perception of American English /r,l/ by Japanese listeners. *Perception & Psychophysics* 52:376–92.