

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 19 (1993-1994)
Indiana University

**On the Contribution of Instance-Specific Characteristics
to Speech Perception¹**

A. R. Bradlow, L. C. Nygaard, and D. B. Pisoni

*Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405*

¹ This research was supported by NIDCD Research Grant DC-00111 and NIDCD Research Training Grant DC-00012 to Indiana University, Bloomington, IN 47405.

Abstract

Utterance-, speaker-, and listener-related correlates of speech intelligibility were investigated using data from the Indiana Multi-Talker Sentence Database and from a talker identification study. The sentence database consists of 100 Harvard sentences produced by 20 speakers, as well as intelligibility data in the form of sentence transcriptions from 10 listeners per talker. This database provided us with the means to investigate some of the sentence- and talker-specific correlates of speech intelligibility. Results showed that talker-related characteristics such as gender and individual differences in phonetic implementation were correlated with the observed variability in talker intelligibility. Specifically, the data showed that female talkers, who tend to exhibit fewer instances of phonological reduction phenomena, were generally more intelligible than male talkers, who may be less precise in their phonetic implementation of phonological forms. An investigation of the sentence-related factors that correlate with variability in overall sentence intelligibility revealed that the number and type of words that comprise a particular sentence were important factors controlling intelligibility. Sentences with relatively high overall intelligibility were generally shorter (had fewer words), and had more easy-to-recognize words (i.e. shorter, more frequent, and more phonetically distinctive words) than sentences with relatively low overall intelligibility. The talker identification study trained listeners to identify 10 talkers by name over a period of 9 days, after which the listeners performed a word identification task with novel words spoken by the now familiar voices as well as by novel voices. These data provided us with the means to investigate the effect of talker familiarity on speech perception, and the relationship between talker identifiability and talker intelligibility. Results of this investigation showed that listeners who learned to identify the voices showed an advantage in the word identification task with words spoken by familiar talkers relative to their performance in the task with unfamiliar talkers. These data also showed that talkers who were easily identified within the group of 10 talkers were not the most intelligible talkers as measured by the word identification task, implying that talker distinctiveness and talker intelligibility are not necessarily related.

On the Contribution of Instance-Specific Characteristics to Speech Perception

Introduction and Background

The role of variability in the listener's interpretation of the speech signal has been the topic of extensive research, and in general, it has been treated as a source of "noise" to be separated from the meaningful, abstract, symbolic units of speech [1,2]. For example, the general approach of many studies of speech acoustics has been to perform various measurements of speech sounds as produced various talkers in various phonetic and/or prosodic environments, e.g. [3-5]. The data are then used to derive generalizations about the nature of speech sounds and their contextual variation, which can then be used to investigate the acoustic cues to the related perceptual contrasts. An explicit assumption of this approach is that the variability inherent in the speech signal presents an "obstacle" to the listener that needs to be removed, or "stripped away", from the signal to facilitate perception of the underlying abstract linguistic units. Accordingly, the driving force behind this general research agenda has been the specification of the principles that underlie the observed variability in the speech signal so that underlying linguistic units can be perceptually "recovered" by the listener.

In contrast, our theoretical approach treats variability of the speech signal as a useful source of information that is available to listeners at all stages in their interpretation of the speech signal [6-8]. For example, this approach predicts that listeners will be sensitive to inter-talker differences; and that, rather than removing this source of variability from the signal as a consequence of perceptual analysis, listeners use this information as a basis for identifying talker characteristics that can aid in the interpretation of the linguistic message. Accordingly, in our acoustic analyses of sentences produced by multiple talkers we have deliberately avoided averaging across many talkers to derive summary generalizations about speech production; rather, we focus on inter-talker differences and try to correlate these differences with differences in listener responses. In general, our approach contrasts markedly with the traditional, "abstractionist approach" to speech because we focus on instance-specific variation, as opposed to the traditional emphasis on instance-independent generalizations about idealized, abstract symbolic forms [9,10].

In keeping with this general theoretical orientation, the research presented in this report is motivated by three observations regarding variability in speech perception. First, we observe variability in the intelligibility of different sentences across many talkers and listeners. Second, we observe variability in the intelligibility of different talkers across many sentences and listeners. And third, we observe variability in the perceptual strategies used by different listeners in learning to identify the voices of different talkers, and in their use of this talker-related information in speech perception. In other words, we observe that some sentences are more intelligible than others, that some talkers are more intelligible than others, and that some listeners make better use of instance-specific information in speech perception than others. The findings reported here represent an attempt to identify some of the specific utterance-, talker-, and listener-related correlates of speech perception.

Two sources of data provide the basis for our analyses. The first set of data comes from a multi-talker sentence database which includes 100 sentences produced by 20 talkers (ten females and ten males) giving a total of 2000 recorded sentences [11]. All sentences were taken from the list of Harvard sentences [12], and consist of one main clause, five keywords, and any number of additional function words. None of the talkers had any known speech or hearing impairments, and all recordings were screened for misarticulations. (Any misarticulated sentences were re-recorded.) This database also includes intelligibility

scores for each sentence and talker. These scores were obtained from listening tests in which 200 listeners (ten per talker) transcribed each of the 100 sentences. These sentence transcriptions were then scored according to a criterion that counted a sentence as correctly transcribed if, and only if, all five keywords were correctly transcribed. None of the listeners had any known speech or hearing impairments, and the sentences were presented to the listeners over headphones in the clear, that is, without any signal distortions. Examination of these intelligibility data revealed considerable variability in the intelligibility of individual sentences and individual talkers.

The second set of data comes from a talker identification study [13], in which listeners were trained over a period of several days to identify the voices of ten talkers (five females and five males). The stimuli were recordings of isolated monosyllabic words produced by the ten talkers; nineteen listeners were trained over a nine-day period to identify the talkers by name. On the tenth day, subjects participated in two test phases: the first was a talker identification task in which subjects were required to explicitly identify the now "familiar" voices producing a new set of words; the second was a speech intelligibility task in which subjects identified a new set of words produced by either the old, familiar talkers or by a new set of ten unfamiliar talkers [13]. The results of this study provide information about the relationship between talker distinctiveness (that is, talker identifiability) and talker intelligibility, as well as data concerning the variability in the performance of different listeners in these two types of perceptual tasks.

Taken together, the results from analyses of the multi-talker sentence database and the talker identification study provided us with a rich set of data that we used to explore instance-specific correlates of speech intelligibility.

SENTENCE-RELATED CORRELATES OF SPEECH INTELLIGIBILITY

We begin with an analysis of the specific sentence-related characteristics that correlate with variability in sentence intelligibility. The intelligibility tests using the 100 sentences from the talker variability database showed that the sentence intelligibility scores across all talkers and listeners ranged from 54% to 98% correct transcription, with a mean and standard deviation of 87.7% and 8.7%, respectively. In order to examine the sentence-related correlates of this variation in intelligibility, a set of high-intelligibility sentences was selected for comparison with a set of low-intelligibility sentences. The high-intelligibility set consisted of all sentences with greater than 95% correct transcription ($n=14$); the low-intelligibility set consisted of all sentences with less than 75% correct transcription ($n=9$). Since all of these sentences are similar with respect to clause structure, our comparisons of the sets of high- versus low-intelligibility sentences focused on characteristics such as sentence length and the lexical characteristics of the individual keywords.

Our first finding in comparing the high-intelligibility sentences and low-intelligibility sentences was that the high-intelligibility sentences have fewer words on average (7.2 versus 8.2 words per sentence, $p(21)=0.03$ by an unpaired 2-tailed t-test). This count of words includes all words in the sentences, even though the sentence intelligibility scores are based on the correct transcription of only the five keywords in each sentence. The results suggest that the number of words surrounding the keywords has an effect on the overall sentence intelligibility: Words in longer sentences are more susceptible to error than words in shorter sentences. Furthermore, an examination of the repeated transcription errors for both sets of sentences showed that, although in both cases the vast majority of errors can be thought of as low-level perceptual errors, a larger proportion of the listener errors for the low-intelligibility sentences can be thought of as higher-level "memory" errors (14% versus 7%). For example, a repeated, low-level perceptual error in the high-intelligibility sentences was found in the first word of the sentence, "Kick the

ball straight and follow through,” which was transcribed as “keep” more than once. Clearly, these two words are very close phonetically, and both are semantically compatible with the rest of the sentences. In contrast, a common, higher-level “memory” error in the low-intelligibility sentences was the interchange of “strong” and “firm” in the transcription of the sentence, “The heart beat strongly and with firm strokes.” In this case, the source of the error appears to be a memory confusion rather than a misperception. Thus, based on the higher proportion of such “memory” errors for the low- as opposed to the high-intelligibility sentence groups, it seems plausible that longer sentences simply have more transcription errors due, in part, to the higher memory load.

The second finding from our comparison of high- and low-intelligibility sentences concerned the keyword characteristics. Across all Harvard sentences, the majority of the keywords were content words, that is, words that can be morphologically complex such as nouns, verbs, adjectives, and adverbs; however, in many cases the five keywords of a sentence included one or more function words, that is, words that are morphologically simplex such as auxiliaries, prepositions, pronouns, and demonstratives. A comparison of the keywords in the high- and the low-intelligibility sentences showed that the high- intelligibility sentences had a higher proportion of function keywords (21.4%) than the low-intelligibility sentences (11.1%). Since function words generally have a much higher frequency of occurrence in the language than content words, the higher proportion of function keywords in the high-intelligibility sentences leads to a higher mean word frequency for the keywords in the high- compared to the low-intelligibility sentences (1064 versus 152 occurrences per one million words of printed text, $p(113)=0.05$)². Similarly, since function words are generally shorter than content words, the mean word length for the high-intelligibility sentences was shorter than for the low-intelligibility sentences (3.6 versus 4.1 phonemes per word, $p(113)=0.025$). These analyses suggest that overall sentence intelligibility is related to the mean word frequency and length of the individual words in the sentence, which are, in turn, related to their lexical status (that is, function versus content word).

Another difference between the high- and low-intelligibility sentences is related to the neighborhood characteristics of the keywords [15]. The “similarity neighborhood” of a word is the set of words that differ from the target word by a one-phoneme substitution, deletion, or addition in any position [15]. The “lexical density” of a neighborhood is equal to the number of such neighbors, and the mean neighborhood frequency is the mean word frequency of all the words in a lexical neighborhood. Using these neighborhood characteristics we can describe a word as “easy” if it comes from a “sparse” neighborhood, and/or its frequency is higher than the mean neighborhood frequency of other phonetically similar words. Such a word has been shown to be more accurately and quickly identified than a “hard” word, that is, one that comes from a “dense” neighborhood, and/or does not occur with a higher frequency than its neighbors [15-17]. Using a computerized version of Webster’s Pocket Dictionary, which is based on 20,000 entries, the neighborhood characteristics for the keywords in the high- and low-intelligibility sentences that appear in this dictionary were found and analyzed³.

As shown in Figure 1, for the high-intelligibility sentences the mean difference between keyword frequency (1140 per million) and mean neighborhood frequency (185 per million) is quite large (955 per million); whereas, for the low-intelligibility sentences the difference is 59 per million (209 - 150). Additionally, a higher percentage of the keywords in the high-intelligibility sentences have higher frequencies than the mean frequency of the other words in their similarity neighborhood. In terms of mean

² These word frequency counts are based on the Brown Corpus of printed text [14].

³ Of the high intelligibility sentence keywords, 59 out of 70 (84.3%) appeared in this online dictionary; of the low intelligibility sentence keywords, 43 out of 45 (95.6%) were in this dictionary.

neighborhood density, however, the high- and low-intelligibility sentence keywords come from equally dense neighborhoods (13.6 versus 13.3 neighbors per keyword). Thus, based on these analyses, the high-intelligibility sentences contain keywords that are more distinctive from their similarity neighborhoods in terms of word frequency, and they are therefore “easier” to recognize than the low-intelligibility sentence keywords. In other words, these words are more perceptually salient, and therefore less confusable with other phonetically similar words.

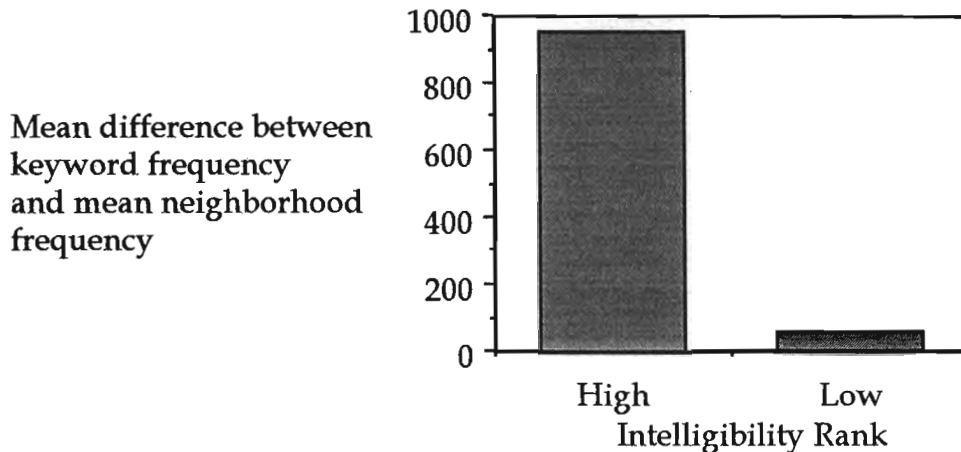


Figure 1. The mean difference between keyword frequency and mean neighborhood frequency for the high- and low-intelligibility sentences.

In summary, we have found that the number and nature of words that comprise a sentence have an effect on the overall intelligibility of the sentence, as measured by listener transcriptions. Specifically, words in longer sentences are more vulnerable to transcription errors than those in shorter sentences. Additionally, the lexical and neighborhood characteristics of the individual words that comprise a sentence, such as word frequency and mean neighborhood frequency, correlate with its overall intelligibility. Specifically, on average, the high-intelligibility sentences have more function keywords than the low-intelligibility sentences, resulting in words that are generally more frequent and shorter in length. Furthermore, the keywords in the high-intelligibility sentences are perceptually more distinctive relative to other phonetically similar words in their lexical neighborhoods than the keywords in the low-intelligibility sentences. Earlier work has shown that such lexical and neighborhood characteristics are determining factors in the speed and accuracy of isolated word recognition [15-17]; the present results extend this finding to words in sentences by demonstrating that these same lexical characteristics play an important role in overall sentence intelligibility.

TALKER-RELATED CORRELATES OF SPEECH INTELLIGIBILITY

We now turn to a discussion of variability in talker intelligibility. The mean intelligibility scores across all 100 sentences for the 20 individual talkers in the talker variability database ranged from 81% to 93% correct transcription, with a mean and standard deviation of 87.9% and 3.1%, respectively. Many talker-related, or “indexical,” factors might be expected to correlate with talker intelligibility, such as gender, overall speaking rate, dialect, fundamental frequency, vocal tract length and other details of speech production that can vary idiosyncratically from one speaker to another. In this section, our aim is to identify some of the talker-related factors that may affect speech intelligibility. We focus here on gender

and overall speaking rate, as well as on two cases that examine talker-related details of speech timing in order to understand their perceptual consequences.

In a recent study of the TIMIT multi-talker database [18], Byrd [19] found that the prevalence of reduction phenomena, such as increased speech rate, reduced frequency of stop releases, alveolar flapping, and vowel centralization were more prevalent among male speakers than female speakers. Based on this result, one might expect that the more carefully articulated speech of females would lead to higher intelligibility scores for females compared to males. In fact, a comparison of the intelligibility scores for the female and male talkers in our database showed that the females have generally higher intelligibility scores than the males (89.4% versus 86.3% correct transcription, $p(18)=0.02$). Furthermore, all three of the talkers with intelligibility scores above 90% are female, and all three talkers with intelligibility scores below 85% are male. Thus, this correlation of gender and intelligibility in our database is consistent with the higher incidence of reduction phenomena for male talkers than for female talkers in the TIMIT database [19]. Taken together, these two results suggest that male and female talkers differ in the precision of articulation, and that this difference has an effect on overall speech intelligibility. However, a direct connection between speech articulation and intelligibility for different talkers still remains to be made from the same source of data.

Overall speaking rate has been shown to be the primary factor that distinguishes carefully articulated speech from reduced speech, since many other reduction phenomena can be directly related to it [20,21]. Thus, we began by examining this factor in our attempt to explore the connection between reduction phenomena and overall speech intelligibility for male and female talkers. A comparison of the sentence durations for all 100 sentences for the three talkers with the highest intelligibility scores with those for the three talkers with the lowest intelligibility scores in the talker variability database revealed that, indeed, the former are longer than the latter (2054 versus 2008 milliseconds, $p(598)=0.03$). This suggests that overall speaking rate and gender are factors that distinguish the most from the least intelligible talkers. However, we also found that the mean sentence durations for all ten males were longer than for all ten females (2155 versus 2085 milliseconds, $p(18)<0.001$), and that for all 20 talkers, mean sentence duration did not correlate with mean talker intelligibility ($r = 0.073$). Thus, although the most and least intelligible talkers in this sample can be distinguished by both gender and speech rate, when the whole set of speakers is included in the analysis, the correlations between gender and rate, and intelligibility and rate no longer hold.

In summary, it appears that gender may indeed correlate with talker intelligibility; however, it is not immediately apparent that, for all speakers, this correlation is due to overall speaking rate. This result leads us to suspect that, although speech rate may play a role in determining the intelligibility of a talker (as shown by the rate difference between the three highest and the three lowest intelligibility scorers in our talker variability database), there are additional factors that can vary independently from overall rate and that contribute to overall talker intelligibility.

In order to investigate the fine-grained variability in timing details that may contribute to talker intelligibility, we present two cases of consistent listener errors that shed light on the perceptual consequences of some idiosyncratic timing differences between talkers. The first case comes from the phrase, "The walled town..." which was often transcribed by listeners as "The wall town ...". This error constituted 90% of the transcription errors for this sentence. In order to determine the acoustic factors that correlate with /d/ recognition in this phrase, various portions of the speech signal for each speaker were measured and then correlated with the rate of /d/ recognition for that speaker. The vowel-to-vowel durations, that is, the portion between the /a/ of "wall" and the /a^u/ of "town," was measured from the

point at which there was a marked decrease in amplitude and change in waveform shape as the preceding vowel-sonorant sequence ends, until the onset of periodicity for the following vowel. In almost all cases, this portion consisted of a single /d/-like closure portion and a single /t/-like release portion. Most talkers (18/20) did not release the /d/ and then form a second closure for the /t/. Furthermore, the /d/ closure portion generally consisted of a period with very low amplitude, low frequency vibration, followed by a silent portion and then the /t/-like release burst and aspiration periods. Separate acoustic measurements of all of these components of the vowel-to-vowel period were taken, as well as the duration of the preceding /wal/ sequence.

Rank order correlations of these measures with the rate of /d/ recognition for each talker showed that the total vowel-to-vowel duration correlated quite highly with /d/ detection (Spearman rho = 0.702); however, an even higher correlation was found with the duration of voicing during the /d/ closure (Spearman rho = 0.744). In fact, this correlation between the absolute amount of voicing during the /d/ closure and the rate of /d/ detection for the individual talkers was stronger than any other proportional measure of this period. For instance, the rank order correlations with rate of /d/ detection for the proportion of voicing during closure to the total closure duration, and to the duration of the preceding word were only -0.412 and 0.033, respectively. In other words, the duration of voicing during closure, in an absolute sense, appears to be the most reliable cue to the presence of a voiced consonant in this phonetic environment. Inter-talker variability in voicing during voiced stop closure is a well-documented phenomenon in the production of American English, e.g. [22]. The present correlation of the talker intelligibility data with the acoustic data provides a direct perceptual correlate of this source of variability and shows that listeners are, indeed, sensitive to this acoustic-phonetic variation, and use this information as a cue to the presence or absence of a segment.

The second case of a consistent listener error occurred in the phrase "the play seems," which was often transcribed by listeners as "the place seems." This error constituted 70% of all the transcription errors for that sentence. In this case, we examined the timing details of the acoustic signal in order to see what determined the syllable affiliation of the medial /s/. Measurements were taken of the duration of the /s/ (marked by the high frequency, high amplitude turbulent waveform) and of the preceding and following syllables (/plej/ and /simz/ respectively). Results showed that the absolute duration of the /s/ does not correlate very strongly with the rate of "play seems" transcription (Spearman rho = -0.254); whereas, when taken as a proportion of the /plej/ duration, that is, as a proportion of the preceding word, the rank order correlation with rate of "play seems" transcription is quite strong (Spearman rho = -0.653). In other words, the longer the /s/ relative to "play," the more likely it is to be syllabified by a listener as both the coda of the preceding word, and the onset of the following word. Thus, in this case the listeners appear to draw on more global information about the speaking rate of the talker in deciding on the placement of the word boundary (see [23,24]).

Furthermore, in this case, there appears to be a gender-related factor in the timing relationship between the medial /s/ and the preceding word, "play." In general, the duration of the /s/ relative to the preceding word was shorter for the female talkers than for the male talkers; and consequently, the female talkers' renditions of this phrase were more often correctly transcribed. Thus, in this case, the female talkers as a group appear to be more precise with respect to controlling this timing relationship than their male counterparts. Although this case is not a matter of reduction (in fact, the correct form is shorter in duration), the apparent gender difference in speech production, which is, in turn, correlated with rate of correct transcription, is consistent with the hypothesis that the more carefully articulated speech of female talkers is also more intelligible. Moreover, this case provides an example that explains why overall speech rate is not the only, or even the primary, talker characteristic that determines talker intelligibility: finer

acoustic-phonetic details of the timing relations between phonetic segments in an utterance make an important contribution to overall speech intelligibility.

LISTENER-RELATED CORRELATES OF SPEECH INTELLIGIBILITY

Information about the variability in speech intelligibility due to listener-related factors was obtained from the talker identification training studies, in which the listeners could be divided into two groups based on their performance during training [13]. In this study, nineteen listeners were trained to explicitly identify by name the voices of ten talkers producing isolated, mono-syllabic words. By the ninth day of training, nine listeners were able to identify the talkers with greater than 70% accuracy (the "good" listeners); whereas, the remaining ten listeners failed to reach this level of accuracy (the "poor" listeners). In order to investigate this difference in listener performance, confusion matrices that counted the number of times listeners confused each voice with each of the other nine voices were generated. Separate confusion matrices were generated for the "good" and "poor" listeners, which were then subjected to two separate multi-dimensional scaling solutions [25]. In this way, the psychological dimensions that differentiate the talkers' voices for the "good" and the "poor" learners could be explored.

Figure 2 shows the scaling solutions of the confusion matrices for the two groups of listeners at the end of the training period. Both groups of listeners were successful in separating male from female talkers along dimension 1 (not shown in the two-dimensional plots of Figure 2). However, by the end of the nine-day training period the "good" listeners seem to use dimension 3 to distinguish the female talkers and dimension 2 to distinguish the male talkers. In contrast, the "poor" listeners seem to use both dimensions to distinguish all ten talkers, and, consequently are less successful at the talker identification task than the "good" listeners. Thus, these scaling solutions demonstrate that listeners differ in the strategies they use to learn to explicitly identify different talkers, and that talkers differ in their distinctiveness. This finding raises two issues. First, does the fact that a talker's voice is clearly identified help the listener to recognize words produced by this talker? And second, is talker distinctiveness related to overall talker intelligibility?

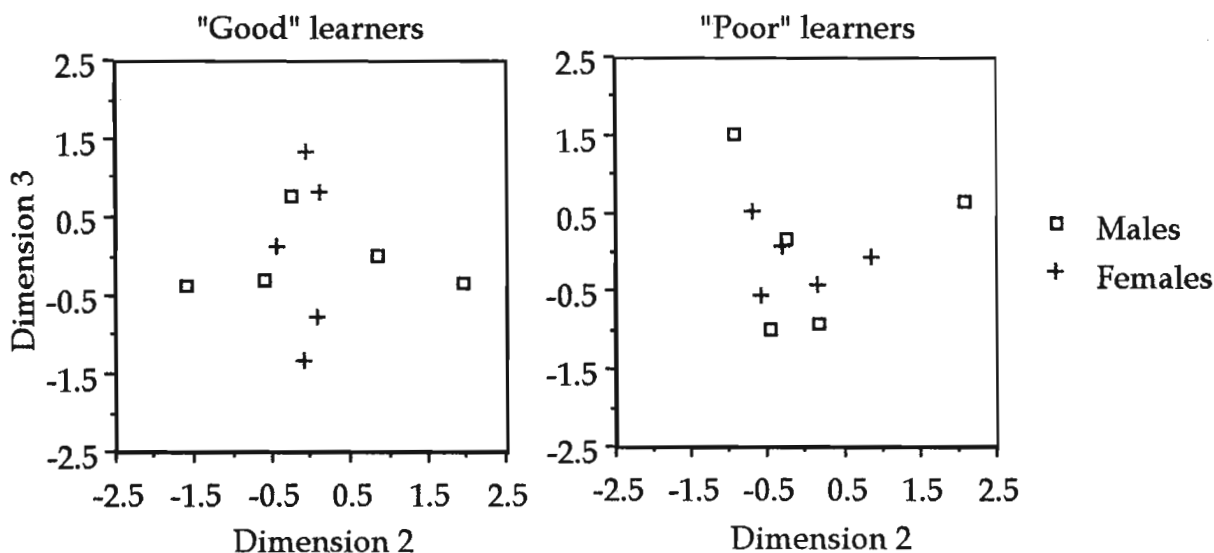


Figure 2. Multi-dimensional scaling solutions for the "good" and "poor" listeners in the talker identification training study (from [13, 25]).

In response to the first issue, we found that in the test phase of the study, the "good" listeners showed an advantage in the word recognition task for novel words produced by familiar voices over novel words produced by unfamiliar voices; whereas, the "poor" listener group did not show any difference due to voice familiarity in the word recognition task. Thus, the "good" listeners apparently use their knowledge about a talker's voice such that their performance on a word recognition task is enhanced relative to the "poor" listeners. This result suggests that listeners differ in their ability to learn to identify talkers' voices and that these differences in perceptual learning do indeed affect speech perception.

We have seen that from the listener's point of view, individual voice identification and word recognition interact, producing an advantage in the recognition of novel words spoken by familiar voices relative to unfamiliar voices (see also [26-28]). A related question is whether talker distinctiveness and talker intelligibility are correlated; in other words, is the most distinctive voice also the most intelligible voice? It is clear from the data in the talker variability database that some talkers are more intelligible than others. Furthermore, it is clear from the talker identification training study that some talkers' voices are more distinctive than others; for example, as shown in the scaling plots in Figure 2, some talkers are more easily distinguished from the other nine talkers by both the "good" and "poor" groups of listeners. However, the data from the talker identification training study indicate that the overall rank order correlation for the ten talkers' identifiability and intelligibility scores is quite low (Spearman rho = -0.143), indicating that voice intelligibility and identifiability are not well correlated. Thus, it would appear that from both the listener's and the talker's points of view, individual voice identifiability and speech intelligibility are separate factors that, although not correlated, can interact to the extent that instance-specific characteristics are employed in the general processes of speech communication.

CONCLUDING REMARKS

The findings reviewed in this report suggest that the "indexical" [29] or "personal" properties of speech may play an important role in speech perception by placing constraints on phonetic and lexical interpretation. Human listeners apparently do not discard the fine instance-specific phonetic details that are encoded in the speech signal. As we have seen from two separate sets of analyses, these acoustic-phonetic details are preserved in memory and provide a rich source of information to assist in speech perception. Specifically, the results of these investigations provide a clear demonstration of the relationship between variation in speech intelligibility and variation of the speech signal due to sentence- and talker-related characteristics. The results also show that the lexical and neighborhood characteristics of the words that comprise a sentence correlate with its overall intelligibility, implying that lexical characteristics that determine isolated word intelligibility operate at the sentence-level as well. Additionally, we found a correlation between inter-talker differences and overall talker intelligibility, suggesting that listeners are sensitive to the fine-grained acoustic-phonetic details that distinguish the speech of one talker from another, and that these differences contribute to a specific talker's overall intelligibility. Taken together, the correlation between word-level characteristics and overall sentence intelligibility, and the correlation between fine-grained acoustic-phonetic differences and overall talker intelligibility, demonstrate the important role that variability plays in controlling speech intelligibility. Thus, the pattern of results that emerges is one in which seemingly small, detailed effects are retained throughout the process of speech perception. The results indicate that, rather than being normalized to fit an abstract, idealized symbolic representation of the meaningful units of speech, these sources of low-level variability in the acoustic signal "propagate up" to higher levels of processing to modulate speech intelligibility.

The results of the talker-identification training study provide a direct demonstration of listener-related differences and the effect these strategies have on speech perception. The data also show that a

listener's ability to learn to identify talkers' voices transfers to the recognition of new words produced by the familiar talkers. Thus, listeners apparently retain "talker-specific" information in memory and make use of this stored information in speech perception and spoken word recognition. This study suggests that speech perception is a "talker-contingent process," and that the talker-specific, indexical properties of speech may not be clearly dissociated from the abstract, linguistic properties; rather, listeners appear to be sensitive to both types of information in the speech signal, and knowledge about a talker's specific vocal tract properties may assist in the perception of that talker's speech.

We interpret these results as providing a demonstration of the contribution of instance-specific information to speech perception. Rather than viewing the inherent variability of the acoustic speech signal as "noise" that is somehow filtered out, or "normalized," by the processes of speech perception, we consider instance-specific variability as information in the stimulus that is directly encoded in the neural representation of speech, and is operative throughout the processes of speech perception and spoken word recognition.

References

1. D. Shankweiler, W. Strange, & R. Verbrugge (1977). Speech and the problem of perceptual constancy. In R. Shaw & J. Bransford (Eds.), *Perceiving, Acting and Knowing*. Potomac, MD: L. Erlbaum.
2. M. Studdert-Kennedy (1974). The perception of speech. In T. A. Sebeok (Ed.), *Current Trends in Linguistics*. The Hague: Mouton.
3. T. Crystal and A. House (1982). Segmental durations in connected-speech signals: Preliminary results. *Journal of the Acoustical Society of America*, **72**, 705-716.
4. T. Crystal & A. House (1988). Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America*, **83**, 1553-1573.
5. V. Zue & M. Laferriere (1979). Acoustic study of medial /t, d/ in American English. *Journal of the Acoustical Society of America*, **66**, 1039-1050.
6. D. Pisoni (1992). Some comments on invariance, variability and perceptual normalization in speech perception. In J. Ohala, T. Nearey, B. Derwing, M. Hodge, & G. Wiebe (Eds.), *Proceedings 1992 International Conference on Spoken Language Processing*. Alberta, Canada: University of Alberta.
7. M. Sommers, L. Nygaard, & D. Pisoni (1992). Stimulus variability and the perception of spoken words: Effects of variations in speaking rate and overall amplitude. In J. Ohala, T. Nearey, B. Derwing, M. Hodge, & G. Wiebe (Eds.), *Proceedings 1992 International Conference on Spoken Language Processing*. Alberta, Canada: University of Alberta.
8. J. Elman & J. McClelland (1986). Exploiting lawful variability in the speech wave. In J. Perkell and D. Klatt (Eds.), *Invariance and Variability in Speech Processes*. Hillsdale, NJ: L. Erlbaum.
9. L. Brooks (1978). Nonanalytic concept formation and memory for instances. In E. Rosch and B. Lloyd (Eds.), *Cognition and Categorization*. Hillsdale, NJ: Erlbaum.
10. L. Jacoby & L. Brooks (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. Bower (Ed.), *The Psychology of Learning and Motivation*. New York, NY: Academic Press.
11. J. Karl & D. Pisoni (1994). The role of talker-specific information in memory for spoken sentences. *Journal of the Acoustical Society of America*, **95**, 2873.
12. IEEE (1969). IEEE recommended practice for speech quality measurements. *IEEE Report No. 297*.
13. L. Nygaard, M. Sommers & D. Pisoni (1994). Speech perception as a talker-contingent process. *Psychological Science*, **5**, 42-46.
14. F. Kucera & W. Francis (1967). *Computational Analysis of Present Day American English*. Providence, RI: Brown University Press.

15. P. Luce (1986). Neighborhoods of words in the mental lexicon. *Research on Speech Perception, Technical Report No. 6*, Indiana University.
16. D. Pisoni, H. Nusbaum, P. Luce, & L. Slowiaczek (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication*, 4, 75-95.
17. P. Luce, D. Pisoni & S. Goldinger (1990). Similarity neighborhoods of spoken words. In G. Altmann (Ed.), *Cognitive Models of Speech Processing: Psycholinguistics and Computational Perspectives*. Cambridge, MA: MIT Press.
18. V. Zue, S. Seneff & J. Glass (1990). Speech database development at MIT: TIMIT and beyond. *Speech Communication*, 9, 351-356.
19. D. Byrd (1992) Sex, dialects, and reduction. In J. Ohala, T. Nearey, B. Derwing, M. Hodge, & G. Wiebe (Eds.), *Proceedings 1992 International Conference on Spoken Language Processing*. Alberta, Canada: University of Alberta.
20. M. Picheny, N. Durlach, & L. Braida (1989). Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech and Hearing Research*, 32, 600-603.
21. R. Uchanski, K. Miller, C. Reed, & L. Braida (1992). Effects of token variability on resolution for vowel sounds. In M. E. H. Schouten (Ed.), *The Auditory Processing of Speech: From Sounds to Words*. New York, NY: Mouton de Gruyter.
22. J. E. Flege & W. S. Brown Jr. (1982). The voicing contrast between English [p] and [b] as a function of stress and position-in-utterance. *Journal of Phonetics*, 10, 335-345.
23. J. L. Miller (1987). Rate-dependent processing in speech perception. In A. Ellis (Ed.), *Progress in the Psychology of Language*. Hillsdale, NJ: Erlbaum.
24. J. L. Miller, F. Grosjean, & C. Lomanto (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica*, 41, 215-225.
25. L. Nygaard & M. Kalish (1994). Modeling the effect of learning voices on the perception of speech. *Journal of the Acoustical Society of America*, 95, 2873.
26. D. Van Lancker (1991). Personal relevance and the human right hemisphere. *Brain and Cognition*, 17, 64-92.
27. D. Van Lancker, J. Kreiman, & K. Emmorey (1985). Familiar voice recognition: Patterns and parameters. Part I: Recognition of backward voices. *Journal of Phonetics*, 13, 19-28.
28. D. Van Lancker, J. Kreiman, & T. Wickens (1985). Familiar voice recognition: Patterns and parameters. Part II: Recognition of rate-altered voices. *Journal of Phonetics*, 13, 39-52.
29. J. Laver & P. Trudgill (1979). Phonetic and linguistic markers in speech. In K. Scherer and H. Giles (Eds.), *Social Markers in Speech*. Cambridge, UK: Cambridge University Press.