



Published in final edited form as:

J Speech Lang Hear Res. 1997 December ; 40(6): 1395–1405.

Effects of Stimulus Variability on Speech Perception in Listeners With Hearing Impairment

Karen Iler Kirk, David B. Pisoni, and R. Christopher Miyamoto*
Indiana University School of Medicine Indianapolis

Abstract

Traditional word-recognition tests typically use phonetically balanced (PB) word lists produced by one talker at one speaking rate. Intelligibility measures based on these tests may not adequately evaluate the perceptual processes used to perceive speech under more natural listening conditions involving many sources of stimulus variability. The purpose of this study was to examine the influence of stimulus variability and lexical difficulty on the speech-perception abilities of 17 adults with mild-to-moderate hearing loss. The effects of stimulus variability were studied by comparing word-identification performance in single-talker versus multiple-talker conditions and at different speaking rates. Lexical difficulty was assessed by comparing recognition of “easy” words (i.e., words that occur frequently and have few phonemically similar neighbors) with “hard” words (i.e., words that occur infrequently and have many similar neighbors). Subjects also completed a 20-item questionnaire to rate their speech understanding abilities in daily listening situations. Both sources of stimulus variability produced significant effects on speech intelligibility. Identification scores were poorer in the multiple-talker condition than in the single-talker condition, and word-recognition performance decreased as speaking rate increased. Lexical effects on speech intelligibility were also observed. Word-recognition performance was significantly higher for lexically easy words than lexically hard words. Finally, word-recognition performance was correlated with scores on the self-report questionnaire rating speech understanding under natural listening conditions. The pattern of results suggest that perceptually robust speech-discrimination tests are able to assess several underlying aspects of speech perception in the laboratory and clinic that appear to generalize to conditions encountered in natural listening situations where the listener is faced with many different sources of stimulus variability. That is, word-recognition performance measured under conditions where the talker varied from trial to trial was better correlated with self-reports of listening ability than was performance in a single-talker condition where variability was constrained.

Keywords

hearing loss; speech perception; adults; stimulus variability

Human listeners typically display a great deal of perceptual robustness. They are able to perceive and understand spoken language very accurately and efficiently over a wide range of transformations that modify the acoustic-phonetic properties of the signal in many complex ways (Pols, 1986; Pisoni, 1993). Among the conditions that introduce variability in the speech signal are differences that are due to talkers, speaking rates, and dialects, as well as a variety of signal degradations—such as noise, reverberation, and the presence of

©1997, American Speech-Language-Hearing Association

Contact author: Karen Iler Kirk, PhD, Department of Otolaryngology-Head and Neck Surgery, Indiana University School of Medicine, 702 Barnhill Drive, Indianapolis, IN 46202. kkirk@iupui.edu.

*Currently affiliated with the University of Cincinnati

competing voices in the environment. To account for this apparent perceptual constancy in listeners with normal hearing, it is generally assumed that spoken word recognition includes some type of perceptual normalization process in which listeners compensate for acoustic-phonetic variability in the speech signal to derive an idealized symbolic internal representation (Pisoni 1993). Spoken words are then identified by matching these normalized perceptual representations to neural representations stored in long-term lexical memory (Pisoni & Luce, 1986; Studdert-Kennedy, 1974).

Over the last few years, Pisoni and his colleagues have carried out a number of studies designed specifically to investigate the effects of stimulus variability on the spoken word-recognition performance of listeners with normal hearing (Mullennix, Pisoni, & Martin, 1989; Mullennix & Pisoni, 1990; Nygaard, Sommers, & Pisoni, 1992; Sommers, Nygaard, & Pisoni, 1994). The results have demonstrated that several sources of stimulus variability affect spoken word-recognition performance in a variety of tasks. For example, when random variations in the talker's voice occurred across successive tokens, word-recognition accuracy decreased and response latency increased compared to performance in a homogeneous single-talker condition (Mullennix et al., 1989; Mullennix & Pisoni, 1990; Nygaard et al., 1992). Similar results have been obtained when speaking rate is varied. Sommers et al. (1994) examined the effects of variations in talker characteristics, speaking rate, and presentation levels. When talker and speaking rate were varied independently, both sources of stimulus variability influenced performance. Again, spoken word-recognition performance was better in the single-talker than in the multiple-talker condition. Also, word recognition scores were higher when a specific speaking rate was held constant in a block of trials than when different speaking rates were combined in a block of trials. Simultaneous variations in both talker and speaking rate yielded even greater decrements in word-recognition performance than varying either of the dimensions alone. However, the effects of stimulus variability were not observed for all stimulus dimensions. Sommers et al. (1994) found that variations in overall presentation level across a 30-dB range had no effect on spoken word recognition. They suggested that normalization of speech signal variability occurs only for those dimensions that directly affect the perception of a given phonetic contrast. Taken together, the results of these studies on stimulus variability support the proposal that increasing acoustic-phonetic variability in a perceptual test or block of trials consumes processing resources required for the perceptual normalization process (Mullennix et al., 1989). Thus, whereas listeners with normal hearing can maintain perceptual constancy in the presence of acoustic-phonetic variability, normalization occurs at a cost to other perceptual and cognitive processes used in spoken-language understanding (Luce, Feustel, & Pisoni, 1983; Rabbitt, 1966, 1968).

Several other factors as well have been shown to influence spoken-word recognition in listeners with normal hearing. For example, the lexical properties of words, such as word frequency (i.e., the frequency of occurrence of words in the language) and lexical similarity, affect the accuracy and speed of spoken-word-recognition performance (Elliott, Clifton, & Servi, 1983; Luce, Pisoni, & Goldinger, 1990; Treisman, 1978a, 1978b). One measure of lexical similarity is the number of lexical neighbors or words that differ by one phoneme from the target word (Greenburg & Jenkins, 1964; Landauer & Streeter, 1973). For example, the words *bat*, *cut*, *scat*, and *at* are all neighbors of the target word *cat*. Words with many lexical neighbors are said to come from dense neighborhoods, whereas those with few neighbors come from sparse neighborhoods (Luce, 1986). Luce et al. (1990) have shown that words from lexically dense neighborhoods are harder to identify than words from sparse neighborhoods.

The Neighborhood Activation Model (NAM) was developed by Luce (1986) to account for differences in word-recognition performance (Hood & Poole, 1980). The model provides a

two-stage account of how the structure and organization of the sound patterns of words in lexical memory contributes to the perception of spoken words. According to the NAM, a stimulus input activates a set of similar acoustic-phonetic patterns in memory within a multidimensional acoustic-phonetic space. Activation levels are assumed to be proportional to the degree of similarity to the target word (see Luce et al., 1990). Over the course of perceptual processing, the sound pattern corresponding to the input receives successively higher activation levels, whereas the activation levels of similar patterns are attenuated. This initial activation stage is then followed by lexical selection among a large number of potential candidates that are consistent with the acoustic-phonetic input. Word frequency is assumed to act as a biasing factor in this model by adjusting multiplicatively the activation levels of the acoustic-phonetic patterns. In the process of lexical selection, the activation levels are then summed, and the probabilities of choosing each pattern are computed on the basis of the overall activation level (Luce, 1986). Word recognition occurs when a given acoustic-phonetic representation is chosen on the basis of these computed probabilities.

In a series of behavioral experiments with adult listeners with normal hearing, Luce and his colleagues (Cluff & Luce, 1990; Luce, 1986; Luce et al., 1990) showed that three independent factors—word frequency (Kucera & Francis, 1967), neighborhood density, and neighborhood frequency (i.e., the average frequency of words in a lexical neighborhood)—all influence spoken-word perception. Easy words (high-frequency words from sparse neighborhoods) were recognized faster and with greater accuracy than hard words (low-frequency words from dense neighborhoods), because easy words have less competition from other phonetically similar words in their neighborhoods. The overall pattern of findings from these studies demonstrates that words are recognized in the context of other phonetically similar words and that lexical discrimination is controlled by several factors that influence identification performance.

Relatively little is known at present about how stimulus variability and lexical information or knowledge influence speech perception in listeners with hearing impairment (HI). When traditional speech-discrimination tests were first developed for clinical use in the late 1940s, most audiologists and experimental psychologists assumed that stimulus variability was a source of noise in the signal that had to be reduced or eliminated so that the listener could accurately perceive the message (i.e., recognize the word that was hidden in the speech signal). Speech-discrimination performance in listeners with and without hearing impairment was frequently assessed with phonetically balanced monosyllabic word lists such as the NU-6 (Tillman & Carhart, 1966). These traditional speech-discrimination tests consist of words spoken by a single talker at a single speaking rate. Although the lists were equated for word frequency, they were not balanced for lexical difficulty of the test words (see Pisoni, Nusbaum, Luce, & Slowiaczek, 1985). Indeed, at the time no one even knew about the effects of phonetic similarity or, lexical knowledge on speech perception or the lexicon. Thus, the more central cognitive processes that may account, in part, for substantial individual variations in word-recognition performance were not incorporated in the design of speech discrimination tests of this kind. A number of recent theoretical developments in perception, memory, and learning have suggested that stimulus variability in speech perception should not be viewed as a source of noise; rather it is lawful and informative and may actually help the listener in carrying out perceptual analysis to recover the intended message (Pisoni, 1993). For example, although introducing stimulus variability may reduce word-identification accuracy, it has been shown to increase the sequential recall of words—perhaps because tokens produced by different talkers demand attention and processing resources (Goldinger, Pisoni, & Logan, 1991).

There are several reasons why assessing the speech-perception performance of listeners with hearing impairment with materials that introduce stimulus variability or vary lexical

difficulty might be beneficial. First, it is of interest theoretically. These listeners receive an impoverished acoustic-phonetic input, and therefore they represent the opportunity to generalize the principles that have been proposed recently concerning the nature of the underlying perceptual mechanisms used in speech perception—particularly the claims about encoding of variability in speech (Goldinger et al., 1991). If listeners with hearing impairment have broad phonetic categories, we would expect to see large differences between their ability to recognize lexically easy and hard words, as the latter require the listener to make finer phonetic discriminations. Furthermore, traditional tests do not measure any of the perceptual abilities used in normalization, and therefore they may not accurately predict speech-perception abilities in more natural listening conditions, where many sources of variability are encountered. Studying the role of different sources of variability in speech discrimination tests, and the relationship between performance on these tests and estimates of real-world listening abilities, might provide new insights into the speech-perception differences noted among individual subjects and might enable us to identify factors that are predictive of performance.

The purposes of the present investigation were (a) to examine the effects of stimulus variability and lexical difficulty on spoken-word recognition by listeners with mild-to-moderate sensorineural hearing loss, and (b) to compare the spoken word-recognition performance of these listeners under conditions involving stimulus variability with their perceived communication abilities in natural listening situations.

Method

Subjects

Subjects were recruited from among patients seen for audiological assessment in the Department of Otolaryngology–Head and Neck Surgery at Indiana University Medical Center. Adults between the ages of 18 and 66 years were eligible to participate. The upper age limit was set to minimize aging effects on word recognition. All subjects had a mild-to-moderate bilateral sensorineural hearing loss and showed speech discrimination scores of 80% or higher on recorded NU -6 word lists (Tillman & Carhart, 1966). Seventeen individuals met these selection criteria and participated as subjects; 9 were men and 8 were women. Their ages ranged from 29 to 66 years, with an average of 55 years. All subjects had a postlingually acquired hearing loss; the etiology was Meniere's disease for one subject, and chronic otitis media for a second subject. The remaining subjects had hearing losses of unknown origin. Table 1 presents mean unaided auditory thresholds and NU-6 scores for all the subjects. Nine of the subjects used amplification. All subjects were paid for their participation in this study.

Stimulus Materials

The subjects' speech-discrimination performance was assessed with the NU-6 test. This test consists of 50 phonetically balanced, monosyllabic words. A recorded version produced by Auditec of St. Louis, in which the stimulus items were produced by a male talker, was administered to all subjects.

Talker-Variability Materials—The stimulus materials used to assess the effects of talker variability were those used by Goldinger et al. (1991). The stimuli consisted of monosyllabic words selected from the Modified Rhyme Test (MRT) developed by House, Williams, Hecker, and Kryter (1965). Three hundred MRT words were originally recorded by 10 male and 10 female talkers for a total of 6000 recorded items. The words were produced in isolation, recorded on audiotape, and then digitized by a 12-bit analog-to-digital converter using a 10 kHz sampling rate. The stimuli were low-pass filtered at 4.8 kHz and stored as

digital files. The RMS amplitude of the words was equated across all tokens using a signal-processing software package (Luce & Carrell, 1981).

A subset of 100 items consisting of 50 easy and 50 hard words was selected from this master digital database using the following criteria: All words were rated as highly familiar by listeners with normal hearing using a 7-point rating scale (Nusbaum, Pisoni, & Davis, 1984). The stimuli were selected on the basis of their frequency of occurrence in the language and the number of phonemically similar words that could be generated by adding, deleting, or substituting one phoneme from the target word (Greenburg & Jenkins, 1964). Lexical statistics were obtained on these words from a computerized version of *Webster's Pocket Dictionary* (Luce, 1986; Luce et al., 1990.) The median word frequency counts for the easy and hard words were 57 and 5 per 1 million words, respectively. The median number of phonemically similar neighbors was 14 for the easy words and 35 for the hard words. The single-talker list was created by selecting half of the easy and hard words (25 of each) produced by one of the male talkers in the database. The multi-talker list included the remaining 50 words. In this test the talker for each item was randomly selected from among 5 male and 5 female talkers (the male talker from the single-talker condition was not used). All of the stimulus items in the database were previously shown to be highly intelligible to a group of listeners with normal hearing (Goldinger et al., 1991).

Speaking Rate Materials—The test stimuli were drawn from a database of 450 words (nine 50-item phonetically balanced word lists) produced by a male talker. Each word was presented to the talker embedded in the carrier phrase, “Please say the word _____.” The sentences were presented to the talker on a CRT screen placed inside a sound attenuated IAC booth. The talker was instructed to produce the sentences at three distinct speaking rates (fast, medium, and slow), but was not given any other specific instructions. Sentences were recorded in blocks, with speaking rate held constant for each block. The talker was given 10 practice sentences at the beginning of each block to become familiar with the speaking rate required for that block. The talker's productions were monitored by an experimenter from a speaker placed outside the sound booth so that sentences that included mispronunciations could be re-recorded immediately. The stimuli were digitized online with a 12-bit analog-to-digital converter using a 10 kHz sampling rate. The stimuli were low-pass filtered at 4.8 kHz and stored as digital files. Target words were edited from the sentences, and the RMS amplitude was equated across tokens using the procedures described above.

Two conditions were created for this study: a single-speaking-rate condition and a mixed-speaking-rate condition. The single-speaking-rate condition involved 50 items produced at a medium speaking rate, whereas the mixed-speaking rate involved 150 items (50 each at fast, medium, and slow rates). For the latter condition, speaking rate varied randomly from trial to trial on the test. Because of an error, one word was repeated in the single-speaking-rate and the mixed-speaking-rate conditions. The average duration of words presented in single-rate conditions was 519 ms. For the mixed-speaking-rate condition, the average durations of fast, medium, and slow speaking-rate stimuli were 392 ms, 522 ms, and 1290 ms, respectively.

Self-Assessment of Communication Abilities—A 20-item questionnaire was used to assess the subject's perceived speech-understanding abilities in daily listening situations (see Appendix). The majority of items on this questionnaire were drawn from the Profile of Hearing Aid Performance (PHAP) developed by Cox and Gilmore (1990). This is a 66-item inventory intended to quantify communication performance with or without the use of hearing aids. Test items were drawn from four subscales on the PHAP: (1) Familiar Talkers, (2) Reduced Cues, (3) Background Noise, and (4) Distortion of Sound. Two additional subscales, (5) Gender and (6) Speaking Rate, were created for the present investigation.

Each question presented a statement describing communication abilities in a typical listening situation (e.g., “I can understand conversation when I am walking with a friend through a quiet park” or “I find that most people speak too rapidly”). Subjects responded by using a 7-point rating scale with which they indicated the percentage of time they believed the statement to be true—where *A* indicated that the statement was always true (99%) and *G* indicated that the statement was never true (1%).

Procedures

Subjects were tested individually in a sound-attenuated IAC booth. The stimuli were presented monaurally under earphones (Telephonics, TDH-39P) to the better ear. If both ears were the same audio logically, the preferred ear was used. Stimuli were presented to the right ear for 9 subjects and to the left ear for 8 subjects. The tests were presented at each subject's maximum comfort level (MCL), which was determined before testing commenced. The mean presentation level ranged from 70 dB to 100 dB HL, with an average of 81.4 dB HL. Stimulus presentation was blocked by test. That is, the two talker conditions and the two speaking-rate conditions were presented in succession. The order of test presentation and the order of conditions within a block were randomized across subjects. Subjects responded by repeating aloud the word they heard while the examiner recorded their responses in specially prepared booklets. Responses were scored by the percentage of words correctly identified. To be counted as correct, a response had to be identical to the target word. For example, the plural form of a singular target word was counted as an incorrect response.

The self-assessment questionnaire was administered following completion of the word-recognition tests. The subjects' responses on this questionnaire were then correlated with their word-recognition scores on the talker and rate tests described above. The correlational analysis examined whether the relationship between objective measures of performance and the subjects' perceived listening skills differed as a function of stimulus variability or speaking rate.

Results

Figure 1 presents the mean percent of easy and hard words correctly identified in the single- and multiple-talker test conditions. On average, these subjects correctly identified approximately 74% (range= 44%–100%) of the easy words when they were presented by a single talker, and about 62% (range= 25%–83%) of the easy words when the talker varied from trial to trial. The trend was similar for hard words, although overall word-recognition performance was poorer than when easy words were presented. The subjects recognized about 62% of the hard words in the single-talker condition (range = 33%–88%), but only about 43% of the hard words in the multiple-talker condition (range = 12%–72%).

A two-way analysis of variance with repeated measures was computed for the percent-correct scores, with talker condition and lexical difficulty as the independent variables. The results demonstrated that both factors reliably influenced word-recognition performance. Performance was significantly better in the single-talker condition than in the multiple-talker condition, $F(1, 16) = 36.5, p < .0001$. There was also a highly significant effect of lexical difficulty, $F(1, 16) = 64.5, p < .0001$. Lexically easy words were recognized better than lexically hard words. No interaction was observed between the talker conditions and lexical difficulty.

Word-recognition performance in the single-speaking-rate condition and for each of the three speaking rates used in the mixed-speaking-rate condition is shown in Figure 2. The percent of words correctly identified in the single-speaking-rate condition and in the mixed-

speaking-rate condition when words were produced at a medium speaking rate was the same (65%). The ranges were similar as well (38%–86% for the single-rate medium words and 41%–91% for the mixed-rate medium words). However, within the mixed-speaking-rate condition, there was approximately a 10% decline in word-recognition performance when speaking rate increased. Thus, the subjects correctly identified an average of 75% of the slow words (range = 52%–98%) but only 55% of the fast words in the mixed-speaking-rate condition (range = 38%–85%). A one-way analysis of variance was computed with the four speaking rates (single-rate medium, mixed-rate slow, mixed-rate medium, and mixed-rate fast) as the independent variables. The results showed a highly significant effect of speaking rate, $F(3, 13) = 32.9, p < .0001$. Post hoc tests (Tukey HSD) revealed that word-recognition performance was significantly better in the slow-speaking-rate condition and significantly poorer in the fast-speaking-rate condition ($p < .05$). As noted above, there was no reliable difference for words produced in the two medium-speaking-rate conditions.

Self-Assessment of Communication Skills

Individual responses to the self-assessment questionnaire were analyzed as follows: First, the response to each question was transformed so that “1” always represented the poorest speech understanding and “7” represented the best understanding for the situation described. Next, for each subscale, responses were summed across all questions and then averaged to obtain a subscale score. The six subscale scores were then summed to obtain a total score for each subject (maximum total score = 42). The Appendix presents a list of the questions used for each subscale. Table 2 shows the individual subjects’ scores obtained on the self-assessment questionnaire for the six subscales, along with their mean total score.

To examine the relationships between the effects of stimulus variability and lexical difficulty and the subjects’ perceived communication abilities in daily listening situations, each subject's word-recognition scores on the talker-variability and speaking-rate-variability tests were correlated with their subscale scores and total score on the self-assessment questionnaire. These correlations are presented in Table 3. The subjects’ total score on the questionnaire was moderately correlated with scores obtained in the multiple-talker and mixed-speaking-rate conditions, but not with conditions in which stimulus variability was minimized (i.e., single-talker and single-speaking-rate conditions). Across all subscales, the correlations with word-recognition performance were generally better for the high-stimulus variability conditions (i.e., multiple-talker and mixed-speaking-rate conditions). That is, spoken-word-recognition performance under perceptually robust conditions involving stimulus variability was significantly correlated with the subjects’ mean total score on the questionnaire and on several of the subscales. However, spoken-word-recognition performance under the single-talker and single-speaking conditions was significantly correlated only with scores on the Distortion of Sound subscale. The subjects’ rating of their spoken-word recognition in situations with background noise showed significant correlations ($p < .025$) in two out of the eight test conditions (multiple talker/easy words and multiple talker/hard words).

Discussion

The purpose of the present study was to examine the effects of talker variability and speaking rate on spoken-word recognition by listeners with hearing impairment—two factors that are known to reduce speech intelligibility in listeners with normal hearing. In addition, we examined the role of lexical difficulty in spoken-word identification by listeners with hearing loss. Overall, the results of the present investigation indicated that word-recognition performance of listeners with hearing loss is affected by talker variability and speaking rate, as well as by lexical difficulty. Although controls with normal hearing were not included in the present study, the pattern of performance demonstrated by our

listeners with hearing impairment is similar to that reported in the literature for adults with normal hearing (Mullennix et al., 1989; Mullennix & Pisoni, 1990; Nygaard et al., 1992; Sommers et al., 1994). That is, when stimulus variability was introduced by varying the number of talkers or when the speaking rate or lexical difficulty was increased, word-identification performance declined significantly. In considering the pattern of results obtained in the talker and speaking-rate conditions, it is important to keep in mind that all of the listeners scored very well on the NU-6 test (mean = 92% correct). Indeed, they were selected originally to have high scores on this measure. Using traditional clinical criteria, these listeners with hearing impairment displayed good-to-excellent spoken-word-recognition performance. However, their performance on the talker and speaking-rate conditions revealed a somewhat different picture of their speech-perception abilities. When tested under more robust or demanding listening conditions, their word-recognition performance was much lower. Although these listeners showed highly consistent and highly significant differences on all three of the independent variables manipulated on each of these tests—specifically, talker variability, speaking rate, and lexical difficulty—the nature of these effects differed.

In the talker conditions, the listeners' performance was lower in the multiple-talker condition than in the single-talker condition. Furthermore, their recognition of lexically difficult words was significantly poorer than their recognition of lexically easy words. This pattern was highly consistent across subjects. Both of these factors (talker variability and lexical difficulty) produced independent effects on performance, as shown by the absence of any statistical interactions. These results are similar to those reported by Sommers et al. (1994) for listeners with normal hearing who were tested in noise. Their listeners also identified words with significantly greater accuracy when in single-talker than in multiple-talker conditions. Our present findings suggest that although listeners with hearing loss may have reduced frequency selectivity and audibility, and thus a reduced ability to make fine phonetic discriminations, they do perceive words in the context of other phonetically similar words in their lexicons. Their performance on these tests revealed the same overall pattern for lexically easy and hard words that has been reported for listeners with normal hearing on comparable tests of speech perception (Mullennix et al., 1989; Mullennix & Pisoni, 1990). However, because word frequency and lexical density co-varied in this study, future investigations will be needed to separately determine the effects of each variable on speech perception by listeners with hearing impairment.

When speaking rate was manipulated, the listeners with hearing impairment consistently displayed their best spoken-word-recognition performance for slow words and their worst performance for fast words. Words produced in the medium speaking-rate condition were recognized between these extremes. This pattern of results is similar to that found for listeners with normal hearing. Picheny, Durlach, and Braida (1985a; 1985b) found that when talkers were instructed to speak as if they were addressing a listener with hearing loss, they altered their speaking style in several consistent ways, including a reduction in the rate of speech. These alterations in speaking style yielded substantial increases in speech intelligibility. The present results provide further evidence that such alterations do yield higher levels of speech intelligibility in listeners with hearing loss. One possible interpretation of this result is that the subjects were sensitive to the talker's speaking rate and were responding to the degree of articulatory precision and phonetic contrast in these test items.¹

¹However, as pointed out by an anonymous reviewer, it may be that superior recognition of words produced at a slow speaking rate resulted from an increase in the time available for perceptual processing.

Introducing stimulus variability by varying speaking rate did not seem to influence performance of listeners with hearing impairment in the same way as did varying the number of talkers. That is, words produced at a medium speaking rate were identified equally well in the single-speaking-rate and the mixed-speaking-rate conditions. This result differs from that of Sommers et al. (1994), who tested listeners with normal hearing in noise and found that both talker and speaking-rate variability significantly reduced spoken-word recognition—as compared to conditions in which stimulus variability was constrained. However, our procedures differed from those of Sommers et al. (1994) as well. They tested three separate groups of listeners in the single-rate condition, one at each speaking rate (fast, medium, or slow), and compared the performance of these groups to that of another group that heard the multiple-speaking-rate condition. In our study, only the medium speaking rate was presented in the single-speaking-rate condition. It may be that the effects of variability are most evident under difficult listening conditions, such as when the talker is speaking at a fast rate (or when the listeners with normal hearing are tested under noise). Future investigations should include three different single-speaking-rate conditions (fast, medium, and slow) to further examine the effect of speaking-rate variability on spoken-word recognition by listeners with hearing loss.

Given that the NU-6 speech discrimination scores for these listeners were very good to begin with, why did we observe such a wide range of variability on the talker and speaking-rate conditions, all of which employed very familiar, high-frequency words (Nusbaum et al., 1984)? We suggest that these new speech-discrimination tests are measuring several important abilities used in normal speech perception that are not typically assessed by traditional low-variability tests such as the NU-6. The specific perceptual attributes that we are measuring in these perceptually robust tests assess the listener's ability to compensate for stimulus variability in their listening environment, an important characteristic of normal, everyday speech-perception performance.

Perhaps the most salient and potentially important attribute of normal speech perception is the ability to perceptually normalize and compensate for changes in the source characteristics of the talker's voice and associated changes in the talker's speaking rate. In the talker-variability conditions, we measured the difficulty listeners have when the talker's voice changes from trial to trial during the test sequence. Under these high-variability testing conditions, a listener must be able quickly to readjust and to retune his or her perceptual criteria from one trial to the next. This process evidently requires additional processing resources and attentional capacity of the listeners (Goldinger et al., 1991; Rabbit, 1966, 1968). These somewhat unusual high-variability testing conditions may be contrasted with more traditional speech-discrimination tests used routinely in clinical audiology. These tests have typically employed only one talker, who produced a single token of each test item. Similarly, in the speaking-rate conditions, we measured the difficulty listeners have when the stimulus items were produced at different speaking rates. Again, intraditional speech-discrimination tests, the test items are not only produced at only one speaking rate, but they are also very clearly articulated by the talker to achieve and maintain maximum levels of speech intelligibility throughout the test.

We believe that traditional low-variability speech-discrimination tests such as the NU-6 may provide inflated estimates of a listener's true underlying perceptual capabilities to recognize speech in real-world environments where there is an enormous amount of stimulus variability. Indeed, our analyses of the self-assessment questionnaire showed that despite very good NU-6 scores, these listeners still reported a wide variety of perceptual difficulties in everyday communication environments. We recognize that self-report measures should be considered cautiously because they are based on the observer's explicit reports of his/her own conscious awareness, which often can be unreliable. However, from the pattern of

responses obtained, it appears that our listeners have specific perceptual difficulties in adverse listening conditions, and they are aware of these specific conditions. These adverse listening conditions include recognizing speech in noise, listening to several competing voices in their environment, or dealing with other factors related to processing stimulus variability in their immediate listening environment.

The subject's perceived communication abilities were more highly correlated with performance under conditions involving stimulus variability than under those in which variability was minimized. That is, moderately high correlations were found between scores on the multi-talker and mixed-speaking-rate conditions and several of the subscales adapted from the PHAP. The subscale most strongly correlated with performance was the Distortion of Sound subscale ($r = .65, p < .005$). This result must be interpreted with some caution, as the Distortion of Sound subscale contained only one item ("People's voices sound unnatural to me."). However, scores on other subscales were also significantly correlated with performance on the multi-talker and mixed-speaking-rate conditions, including Familiar Talkers ($r = .44, p < .05$), Reduced Cues ($r = .56, p < .01$), and Background Noise ($r = .48, p < .03$). Interestingly, performance on the speaking-rate conditions was not correlated significantly with the Speaking Rate subscale. All but one of the questions on this subscale probed how well these listeners could understand speech produced at a fast speaking rate, which is something that is difficult for many listeners with hearing loss. In fact, the mean score for the Speaking Rate subscale was much lower than the scores for all of the other subscales except for Background Noise. In retrospect, including questions that assessed understanding of speech produced at a variety of speaking rates might have yielded a better estimate of performance, one that correlated more strongly with the objective measures of spoken-word recognition.

Because a wider range of scores was observed on the perceptually robust tests than on the NU-6, it appears that these new measures of speech discrimination may better assess precisely the specific nature of spoken-word-recognition deficits that accompany a hearing loss. Furthermore, it appears that these new perceptually robust tests are measuring several underlying aspects of speech perception in the laboratory that appear to generalize to conditions encountered in everyday listening environments. Perceptual tests that specifically manipulate and assess different sources of stimulus variability may therefore be more predictive of actual speech-perception performance in natural communication situations than the traditional low-variability speech-discrimination tests that have been used routinely for many years in the clinic.

The present results have several important theoretical implications as well. The methods described here can be used to assess the indirect or implicit effects of encoding episodic details in speech, such as speaking style or voice quality—a topic that has been virtually ignored in the past because of the emphasis on abstractionist symbolic coding in memory (Pisoni, 1993). The speech signal simultaneously carries information in parallel about the linguistic content of the speaker's message as well as the source characteristics of the talker who produced the signal. In past accounts of speech perception, these two properties of speech, the linguistic message and the indexical form, have been clearly dissociated from each other under the assumption that they are processed independently by the nervous system (Ladefoged, 1975; Laver & Trugill, 1979). Recent behavioral findings have shown that this assumption is incorrect (Pisoni, 1993; Jerger, Martin, Pearson, & Dinh, 1995). The linguistic coding of speech is, in fact, affected by the personal characteristics of the speaker. Moreover, the indexical information in the signal is not lost or discarded during perceptual analysis. Listeners with hearing loss typically have a loss of audibility and reduced frequency selectivity (Florentine, Buus, Scharf, & Zwicker, 1980; Tyler, Wood, &

Fernandes, 1983). Despite this, they do appear to perceive spoken words in the context of other words. The difficulty appears to be in *accessing* the lexicon, not the lexicon itself.

Research using the perceptually robust tests (PRT) of speech perception may help us go beyond assessments in the laboratory and clinic to better predict performance in everyday listening conditions where there are many sources of stimulus variability affecting the listeners' (HI) performance. Under most conditions, listeners with normal hearing are able to ignore these sources of stimulus variability and recover the speaker's intended message, despite degradations in the signal or changes in speaking rate. However, listeners with hearing impairment often anecdotally report that they have difficulty understanding speech in background noise, or an unfamiliar talker, or one who speaks too quickly. Future investigations concerning stimulus variability and lexical difficulty effects on speech perception should directly compare the performance of listeners with normal hearing and listeners with hearing impairment to determine whether the two groups are similarly affected by these factors.

We believe that the ability to deal with different sources of stimulus variability in the speech signal may also be an important diagnostic criterion for predicting success with hearing aids or cochlear implants, and may provide a new perspective on dealing with the individual differences that are observed among patients using these devices. Past results obtained from listeners with normal hearing and now from listeners with hearing impairment (the present study) suggest that acoustic-phonetic variability in the speech signal is not lost as a result of early stages of perceptual analysis. Stimulus variability should not be thought of as simply a source of noise that must somehow be reduced or eliminated during perceptual analysis. Instead, stimulus variability plays an important role in speech perception and spoken-word recognition; it is informative to the listener. With the development of new, perceptually robust tests, the clinician may be provided with important and potentially useful information about the underlying perceptual capabilities of listeners with hearing impairment—information that may not have been considered in the past because of the exclusive reliance on traditional low-variability speech discrimination tests in clinical audiology.

Acknowledgments

This research was supported by NIDCD Research Grants DC-00064, DC-00111, and DC-00126 and NIDCD Training Grant DC-00012 to Indiana University. We thank Julia Renshaw, Molly Pope, and Michelle Wagner-Escobar for their help and assistance in recruiting the patients used in this study. We also thank Mitchell Sommers, Scott Lively, John Karl, and Luis Hernandez for technical assistance at various stages of the project. The assistance of Darla Sallee and Linette Caldwell in the preparation of this manuscript is also acknowledged.

Appendix

Questions used to assess the subject's perceived communication abilities under some natural listening conditions (adapted from Cox and Gilmore, 1990).

Familiar Talker Subscale

Question 1. I can understand others in a small group situation if there is no noise.

Question 2. I can understand speech when I am in a department store talking with the clerk.

Question 3. I can understand conversation when I am walking with a friend through a quiet park.

Question 4. I can understand my family when they talk to me in a normal voice.

Question 7. I can understand conversation during a quiet dinner with my family.

Question 9. I understand the newscaster when I am watching TV news at home alone.

Question 10. When I am in face-to-face conversation with one member of my family, I can easily follow along.

Background Noise

Question 6. When I am in a restaurant and the waitress is taking my order, I can comprehend her question.

Question 19. I can understand a speaker in a small group, even when those around us are speaking softly to each other.

Question 20. I can understand conversations even when several people are talking.

Reduced Cues Subscale

Question 5. When I'm talking with the teller at the drive-in window of my bank, I understand the speech coming from the loudspeaker.

Question 8. When I am listening to the news on my car radio, and the car windows are closed, I understand the words.

Speaking Rate Subscale

Question 11. I understand speech when I am talking to a bank teller, and I am one of a few customers at the bank.

Question 12. I can understand speech when someone is talking at a fast rate.

Question 13. I find that most people speak too rapidly.

Question 14. I can understand my family when they speak rapidly to me.

Question 15. When I am in a quiet restaurant, I can understand people who speak fast.

Gender Subscale

Question 16. I have trouble understanding female voices.

Question 17. I have trouble understanding male voices.

Distortion of Sound

Question 18. People's voices sound unnatural.

References

- Cluff MS, Luee PA. Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception & Performance.* 1990; 16:551–563. [PubMed: 2144570]
- Cox RM, Gilmore C. Development of the Profile of Hearing Aid Performance (PHAP). *Journal of Speech and Hearing Research.* 1990; 33:343–357. [PubMed: 2359275]

- Elliott LL, Clifton LAB, Servi DG. Word frequency effects for a closed-set word identification task. *Audiology*. 1983; 22:229–240. [PubMed: 6870665]
- Florentine M, Buus S, Scharf B, Zwicker E. Frequency selectivity in normally hearing and hearing-impaired observers. *Journal of Speech and Hearing Research*. 1980; 23:647–669.
- Goldinger SD, Pisoni DB, Logan JS. On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning Memory & Cognition*. 1991; 17:152–162.
- Greenburg JH, Jenkins JJ. Studies in the psychological correlates of the sound system of American English. *Word*. 1964; 20:157–177.
- Hood JD, Poole JP. Influence of the speaker and other factors affecting speech intelligibility. *Audiology*. 1980; 19:434–455. [PubMed: 7436861]
- House AS, Williams CE, Itecker MHL, Kryter KD. Articulation-testing methods: Consonantal differentiation with a closed-response set. *Journal of the Acoustical Society of America*. 1965; 3:158–166. [PubMed: 14265103]
- Jerger S, Martin R, Pearson DA, Dinh T. Childhood hearing impairment: Auditory and linguistic interactions during multidimensional speech processing. *Journal of Speech and Hearing Research*. 1995; 38:930–948. [PubMed: 7474984]
- Kucera, F.; Francis, W. *Computational analysis of present-day American English*. Brown University Press; Providence, RI: 1967.
- Ladefoged, P. *A course in phonetics*. Harcourt Brace Jovanovich; New York: 1975.
- Landauer TK, Streeter LA. Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning & Verbal Behavior*. 1973; 12:119–131.
- Laver, J.; Trudgill, P. *Phonetic and linguistic markers in speech..* In: Scherer, KR.; Giles, H., editors. *Social markers in speech*. Cambridge University Press; Cambridge, UK: 1979. p. 1-31.
- Luce, P. *Research in Speech Perception Technical Report No. 6*. Speech Research Laboratory, Indiana University; Bloomington: 1986. *Neighborhoods of words in the mental lexicon..*
- Luce, PA.; Carrell, TD. *Research in Speech Perception, Progress Report No. 7*. Speech Research Laboratory, Indiana University; Bloomington: 1981. *Creating and editing waveforms using WAVES..*
- Luce PA, Feustel TC, Pisoni DB. Capacity demands in short-term memory for synthetic and natural speech. *Human Factors*. 1983; 25:17–32. [PubMed: 6840769]
- Luce, PA.; Pisoni, DB.; Goldinger, SD. *Similarity neighborhoods of spoken words..* In: Altmann, GTM., editor. *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. MIT Press; Cambridge, MA: 1990. 1990
- Mullennix JW, Pisoni DB. Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*. 1990; 47:379–390. [PubMed: 2345691]
- Mullennix JW, Pisoni DB, Martin CS. Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*. 1989; 85:365–378. [PubMed: 2921419]
- Nusbaum, HC.; Pisoni, DB.; Davis, CK. *Research on Speech Perception Progress Report No. 10*. Indiana University; Bloomington: 1984. *Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words..*
- Nygaard, LC.; Sommers, MS.; Pisoni, DB. *Effects of speaking rate and talker variability on the representation of spoken words in memory..* *Proceedings of the 1992 International Conference on Spoken Language Processing; Banff, Canada*. 1992.
- Picheny M, Durlach N, Braida L. Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*. 1985a; 28:96–103. [PubMed: 3982003]
- Picheny M, Durlach N, Braida L. Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*. 1985b; 29:434–446. [PubMed: 3795886]
- Pisoni DB. Long-term memory in speech perception: Some new findings on talker variability, speaking rate, and perceptual learning. *Speech Communication*. 1993; 4:75–95.

- Pisoni DB, Luce PA. Speech perception: Research, theory and the principal issues. *Pattern Recognition by Humans & Machines. Speech Perception.* 1986; 1:1–50.
- Pisoni DB, Nusbaum HC, Luce PA, Slowiaczek LM. Speech perception, word recognition and the structure of the lexicon. *Speech Communication.* 1985; 4:75–95.
- Pols, LCW. Variation and interaction in speech.. In: Perkell, JS.; Klatt, DH., editors. *Invariance and variability in speech processes.* Lawrence Erlbaum and Associates; Hillsdale, NJ: 1986. p. 140-162.
- Rabbitt P. Recognition memory for words correctly heard in noise. *Psychonomic Science.* 1966; 6:383–384.
- Rabbitt P. Channel-capacity, intelligibility and immediate memory. *Quarterly Journal of Experimental Psychology.* 1968; 20:241–248. [PubMed: 5683763]
- Sommers MS, Nygaard LC, Pisoni DB. Stimulus variability and spoken word recognition I: Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America.* 1994; 96:1314–1324. [PubMed: 7962998]
- Studdert-Kennedy, M. The perception of speech.. In: Sebeok, TA., editor. *Current trends in linguistics.* Vol. 2. Academic Press; New York: 1974. p. 1-62.
- Tillman, TW.; Carhart, R. U.S.A.F. School of Aerospace Medicine; Brooks Air Force Base, TX: 1966. An expanded test for speech discrimination utilizing CNC monosyllabic words (Technical Report No. SAM-TR-66-55)..
- Treisman M. Space or lexicon? The word frequency effect and the error response frequency effect. *Journal of Verbal Learning & Verbal Behavior.* 1978a; 17:35–59.
- Treisman M. A theory of the identification of complex stimuli with an application to word recognition. *Psychological Review.* 1978b; 85:525–570.
- Tyler RS, Wood EJ, Fernandes M. Frequency resolution and discrimination of constant and dynamic tones in normal and hearing-impaired listeners. *Journal of the Acoustical Society of America.* 1983; 74:1190–1199. [PubMed: 6643841]

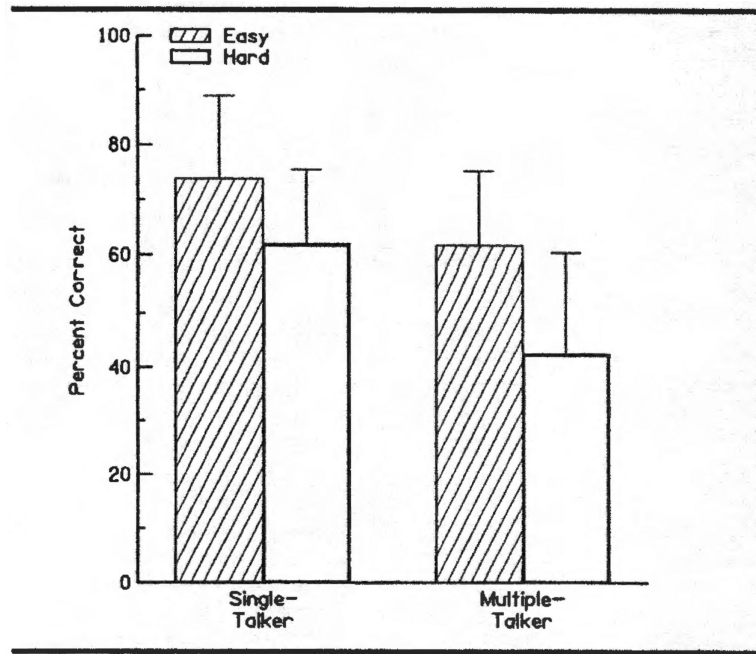


Figure 1. Percent of easy and hard words correctly identified in the single-talker and multiple-talker conditions. Error bars represent + 1 standard deviation.

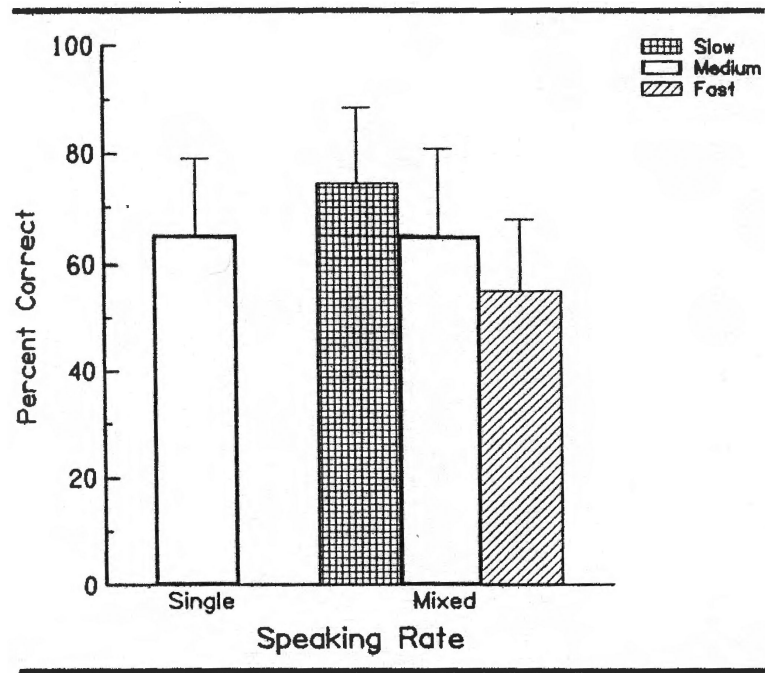


Figure 2. Percent of words correctly identified in the single-speaking-rate (medium rate) condition, and for the slow, medium, and fast words in the mixed-speaking-rate condition. Error bars represent + 1 standard deviation.

Table 1

Mean unaided auditory thresholds and speech discrimination scores for the 17 subjects.

	Auditory thresholds (dB HL)			NU-6 Scores (% correct)	
	500 Hz	1000 Hz	2000 Hz		4000 Hz
<i>M</i>	26.5	33.8	44.4	51.7	92.5
Range	5-60	10-65	30-75	25-70	82-100
<i>SD</i>	16.4	15.1	12.7	12.2	7.2

Table 2

Individual subscale scores and total score obtained on the speech perception self-assessment questionnaire. Scores of “1” and “7” represent the poorest and best speech understanding, respectively.

Subject #	Familiar talker	Background noise	Reduced cues	Rate	Gender	Distortion	Total score
01	6	3	6	4	5	5	29
02	7	5	6	6	5	6	35
03	6	4	6	5	5	6	32
04	5	3	5	4	4	5	26
05	6	5	6	5	5	5	32
06	6	4	5	5	5	6	31
07	5	5	5	5	3	4	27
08	3	2	2	3	6	4	20
09	6	4	5	5	5	7	32
10	6	5	6	5	5	6	33
11	7	5	6	6	6	7	37
12	6	6	6	5	6	7	36
13	6	4	6	4	4	5	29
14	5	4	5	4	5	5	28
15	4	3	4	3	5	5	24
16	5	4	5	4	4	3	25
17	5	2	3	4	4	1	19
Total	94	68	87	77	82	87	495
Mean	5.53	4.0	5.12	4.53	4.82	5.12	29.12

Correlations of talker-variability and speaking-rate word-recognition performance with scores on the speech perception self-assessment questionnaire.

Table 3

	Talker-variability test						Speaking-rate test					
	Single-talker			Multiple-talker			Single rate			Mixed rate		
	Easy	Hard		Easy	Hard		Med	Slow	Med	Fast	Med	Fast
Total Score	.01	.37		.58 ^{***}	.49 ^{**}		.27	.44 [*]	.42 [*]	.49 ^{**}		
Subscales												
Familiar Talkers (FT)	-.06	.22		.44 [*]	.28		.07	.33	.34	.43 [*]		
Reduced Cues (RC)	-.04	.34		.56 ^{***}	.37		.21	.43 [*]	.40 [*]	.40 [*]		
Background Noise (BN)	-.11	.28		.48 [*]	.51 ^{**}		.15	.28	.27	.22		
Distortion of Sound (DS)	.21	.56 ^{***}		.65 ^{***}	.57 ^{***}		.55 ^{***}	.59 ^{***}	.60 ^{***}	.65 ^{***}		
Gender (G)	.16	-.06		.12	.18		.12	.06	.00	.16		
Speaking Rate (SR)	.20	.19		.36	.28		-.01	.21	.19	.30		

Significance Levels:

* $p < .05$

** $p < .025$

*** $p < .01$

**** $p < .005$