RESEARCH ON SPOKEN LANGUAGE PROCESSING

Progress Report No. 19 (1993-1994)

Indiana University

Sources of Variability in Speech Perception by Hearing-Impaired Listeners[1]

David B. Pisoni, Chris Miyamoto[2], Karen Iler Kirk[3], Mitchell S. Sommers[4] and Mary J. Osberger[5]

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

[1] This research was supported by NIDCD Research Grants DC-0064 and DC-00111 and NIDCD Training Grant DC-00012 to Indiana University. We thank Julia Renshaw, Molly Pope and Michelle Wagner for their help and assistance in recruiting the patients used in this study. We also thank Scott Lively, John Karl and Luis Hernandez for technical assistance in various stages of data analysis.

[2] Also at Indiana University School of Medicine, Indianapolis, IN.

[3] Also at Indiana University School of Medicine, Indianapolis, IN.

[4] Now at Washington University, St. Louis, MO.

[5] Now at Advanced Biomedics, San Francisco.

Abstract

Traditional word recognition tests used to measure the speech perception abilities of adults have typically used phonetically balanced word lists produced by one talker at one speaking rate. Such measures may not adequately evaluate the perceptual processes used to perceive speech under more natural listening conditions, where many sources of stimulus variability typically influence performance. The purpose of this study was to examine the influence of two factors, stimulus variability and lexical difficulty, on the speech perception abilities of adults with mild-to-moderate hearing loss. The effects of stimulus variability were examined by comparing word identification in single-talker vs. multiple-talker conditions, and in single-speaking-rate vs. multiple-speaking-rate conditions. The influence of 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"). The results demonstrated that both sources of stimulus variability, talker and speaking rate, produced significant effects on speech intelligibility. Identification scores were significantly poorer in the multiple-talker than in the single-talker conditions. Furthermore, word recognition scores decreased significantly as speaking rate increased. Lexical difficulty also influenced performance. Scores for "easy" items were significantly better than scores for "hard" items. The implication of these results for the development of new robust speech perception measures that assess the underlying perceptual processes will be discussed.

Sources of Variability in Speech Perception by Hearing-Impaired Listeners

Introduction

Human listeners typically display a great deal of "perceptual robustness." That is, 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. Among the conditions that introduce variability in the speech signal are differences due to talkers, speaking rates and dialects, as well as a variety of signal degradations such as noise, reverberation and the presence of competing voices. One of our long-term goals has been to understand how these different sources of variability affect speech perception and spoken word recognition in listeners with normal hearing (Pisoni, 1992, 1993). More recently, we have extended these investigations to listeners with hearing impairments (Kirk et al., 1993; Sommers et al., 1993). In the present study, we systematically examined the effects of talker variability and speaking rate on spoken word recognition by listeners with sensorineural hearing impairment.

It is generally assumed that spoken word recognition involves several underlying perceptual processes in which the speech signal is converted into an acoustic-phonetic representation, normalized for factors such as talker or speaking rate, and then identified by matching the transformed internal representation to items in long-term lexical memory (Pisoni & Luce, 1986; Studdert-Kennedy, 1974). In this last step, it has been shown that successful discrimination and identification of a particular target is influenced by several lexical factors, such as the frequency of occurrence of words in the lexicon and the number of other phonemically similar words with which the target can be confused (Luce, 1986). That is, words that occur frequently and have few phonemically similar neighbors (i.e., "easy" words) are identified with greater accuracy than words that have the opposite lexical characteristics (i.e., "hard" words).

Purpose

Little is currently known about the effects of stimulus variability or lexical factors on word recognition by listeners with hearing impairment. Traditional speech discrimination tests such as the NU-6 or the W-22 lists, which are widely used clinically, were designed to eliminate sources of variability which were viewed as undesirable sources of "noise" in the signal. As a result, performance on tests of this kind may not accurately reflect a listener's ability to cope with the stimulus variability encountered in daily communication. The purpose of the present investigation was:

- To examine the effects of stimulus variability and lexical difficulty on word recognition by listeners with a mild-to-moderate sensorineural hearing loss.

- To assess whether listeners with sensorineural hearing impairment recognize words relationally, that is, in the context of other phonetically similar words.

- To compare spoken word recognition performance of these listeners with their perceived communication abilities in natural listening situations.

Method

Subjects

Subjects were recruited from among patients seen for audiological assessment in the Dept. of Otolaryngology at Indiana University Medical Center. Adults between 18-66 years of age were eligible for participation in the study. The upper age limit was included to minimize aging effects on word recognition performance. All subjects had a mild-to-moderate bilateral sensorineural hearing loss and showed speech discrimination scores of 80% or higher on recorded NU-6 lists (Tillman & Carhart, 1966). Seventeen individuals met these criteria and participated as subjects. Nine of the subjects used amplification. More detailed subject information is presented in Table I.

Insert Table I about here

Talker Variability Test

The stimulus materials used to assess the effects of talker variability employed 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 female and 10 male talkers for a total of 6000 recorded items. A subset of 100 items consisting of 50 "easy" and 50 "hard" words was selected from this digital database using the following criteria. All words were rated as highly familiar by listeners with normal hearing using a seven-point rating scale (Nusbaum, Pisoni, and 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. These lexical statistics were obtained from a computerized version of Webster's Pocket Dictionary (Luce, 1986; Luce, Pisoni & Goldinger, 1990). The median word frequency counts for the "easy" and "hard" words were 57 and 5 per one-hundred thousand words, respectively. The median number of phonemically similar neighbors was 14 for the "easy" words and 35 for the "hard" words. Figure 1 shows the relationship among these two variables. The single-talker list was created using half of the "easy" and "hard" words (25 of each) produced by one of the male talkers. The multiple-talker list contained the remaining words with the talker for each item randomly selected from among 5 male and 5 female talkers. All of the stimulus items were highly intelligible to a group of normal-hearing listeners.

Insert Figure 1 about here

Speaking Rate Test

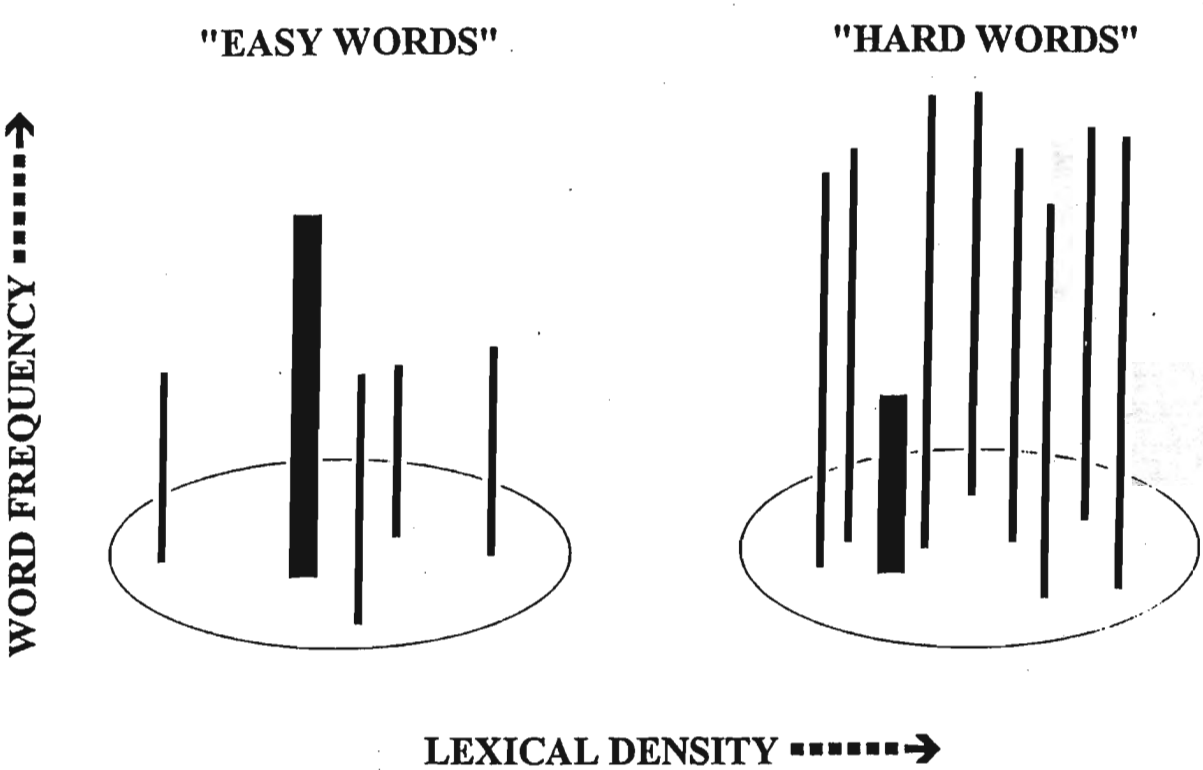
The stimulus items used to assess the effects of speaking rate variability were drawn from a digital database of 3000 words. The original database consisted of two 50-item phonetically balanced (PB) word lists produced by six male and four female talkers at three different speaking rates (fast, medium, and slow). The average durations of words in the fast, medium, and slow conditions were 375, 533, and 905 ms, respectively (Sommers et al. 1992). Tokens produced by a male talker were used to create two conditions: a single-speaking rate condition containing 50 items produced at a medium speaking rate and a mixed-speaking rate condition containing 150 items (50 items at each of three speaking rates presented in random order).

TABLE I

SUBJECT BACKGROUND AND SELECTION CRITERIA

	Mean	Range	Standard deviation
Age	55.3 years	29-66 years	11.83 years
Auditory Thresholds:			
500 Hz	26.5 dB HL	5-60 dB	16.4 dB
1000 Hz	33.8 dB HL	10-65 dB	15.1 dB
2000 Hz	44.4 dB HL	30-75 dB	12.7 dB
40000 Hz	51.76 dB HL	25-70 dB	12.2 dB
NU-6 Scores:	92.47 %	82-100 %	7.2 %

FIGURE 1



Self-Assessment of Communication Abilities

Subjects also completed a 20-item questionnaire designed to assess their speech understanding abilities in daily listening situations. 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: Familiar Talkers (FT), Reduced Cues (RC), Background Noise (BN), and Distortion of Sound (DS). Two additional subscales, Gender (G) and Speaking Rate (SR), were created for the present investigation. Each question presents 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."). Subjects respond using a seven-point rating scale indicating the percentage of time they believe the statement is true (e.g., "Always" [99%] to "Never" [1%]).

Procedures

All stimuli were presented to the better ear under headphones at the maximum comfort level (MCL) while subjects were seated inside a sound-treated room. The mean presentation level ranged from 70 to 100 dB HL, with an average of 81.4 dB HL. Subjects responded by repeating aloud the word they heard while the examiner recorded their responses in specially prepared booklets. Responses were scored by the percent 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. Subject responses were transformed so that 1 always represented the poorest speech understanding and 7 represented the best speech understanding. The average score for each subscale was summed to derive a mean total score.

Results

Talker Variability Test:

Figure 2 presents the mean percent of "easy" and "hard" words correctly identified in the single-talker and multiple-talker test conditions.

Insert Figure 2 about here

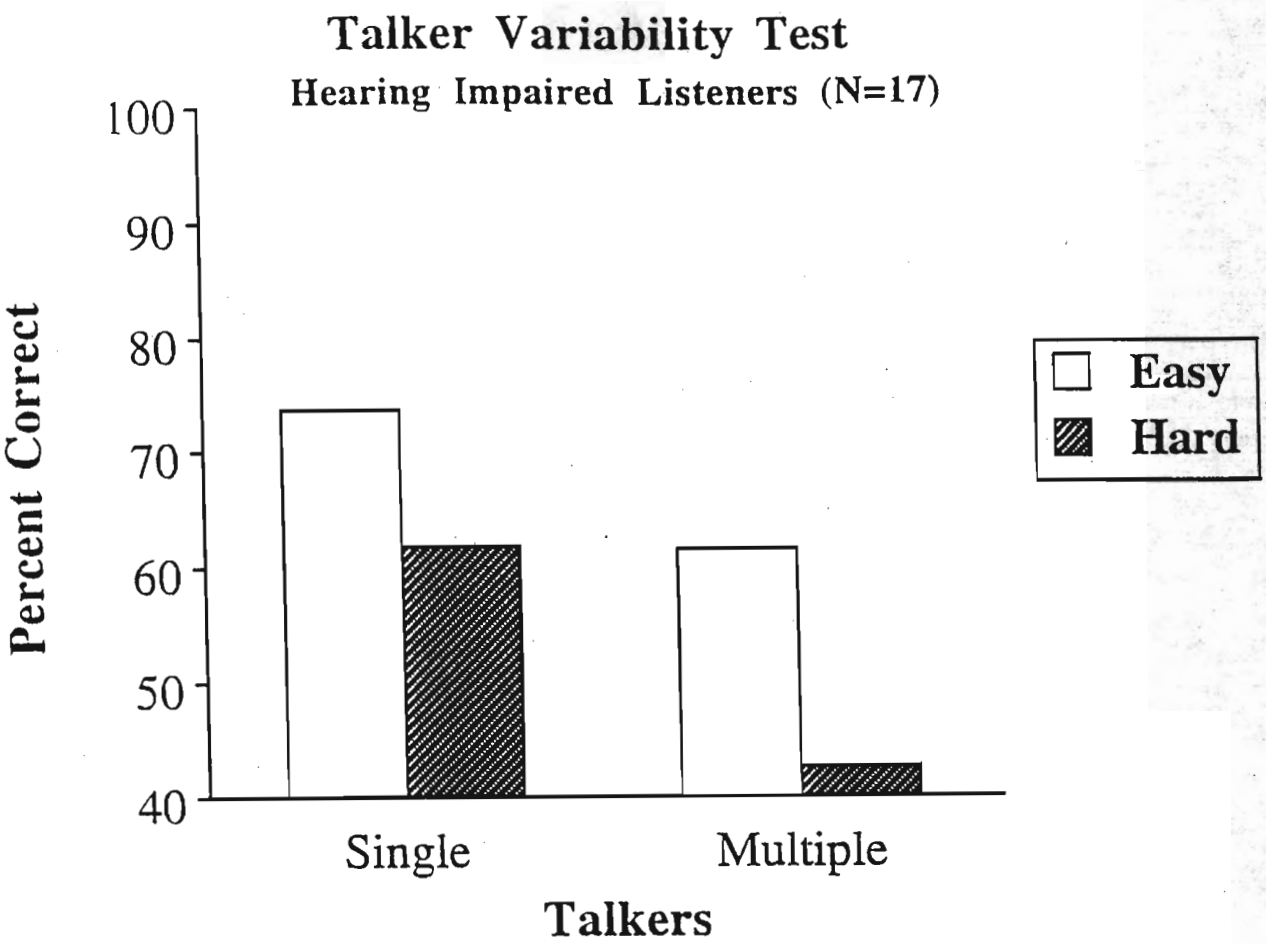
A two-way analysis of variance with repeated measures was computed for the percent correct scores with talker condition and lexical difficulty as the independent variables. The results showed:

- Performance was significantly better in the single-talker than in the multiple-talker condition ($F[1,16] = 36.5$; $p < .0001$).

- There was a highly significant effect of lexical density; "easy" words were identified more accurately than "hard" words ($F[1,16] = 64.50$; $p < .0001$).

- No interaction was observed between the talker conditions and lexical difficulty.

FIGURE 2


Talker Variability Test
Hearing Impaired Listeners (N=17)
Percent Correct
100
90
80
70
60
50
40
Single
Multiple
Talkers
Easy
Hard

Speaking Rate Test:

Figure 3 presents the mean percent of words correctly identified in the single-speaking-rate condition and in each of the three speaking rates used in the mixed-rate condition.

Insert Figure 3 about here

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:

- There was a highly significant effect of speaking rate (F=32.9; p<.0001).

- Post-hoc tests revealed that word recognition was best in the slow-speaking rate condition and worst in the fast-speaking rate condition. Performance on the two medium rate conditions fell in between (p<.05); there was no difference in performance for words produced at the medium rates for the single-rate and mixed-rate conditions.

Self-Assessment of Communication Skills:

Table II shows the mean scores obtained on the modified PHAP self-assessment questionnaire for the individual subjects on each of six subscales along with their mean total score. The maximum total score possible on this scale is 42.

Insert Table II about here

To examine the relationship 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 were correlated with their mean total score on the modified PHAP as well as with the individual subscale scores on the PHAP. These correlations are presented in Table III.

Insert Table III about here

The results of the correlational analyses for the two speech discrimination tests showed:

- Spoken word recognition scores under robust conditions containing stimulus variability (i.e., the multiple-talker and mixed-speaking rate conditions) were significantly correlated with the subjects' mean total score on the modified PHAP and several of the subscales. In particular, the subscale DS (Distortion of Sound) showed highly significant correlations (p<.005) in seven out of eight of the test conditions.

- Word recognition performance under conditions in which stimulus variability was reduced were not significantly correlated with the total score on the modified PHAP.

FIGURE 3

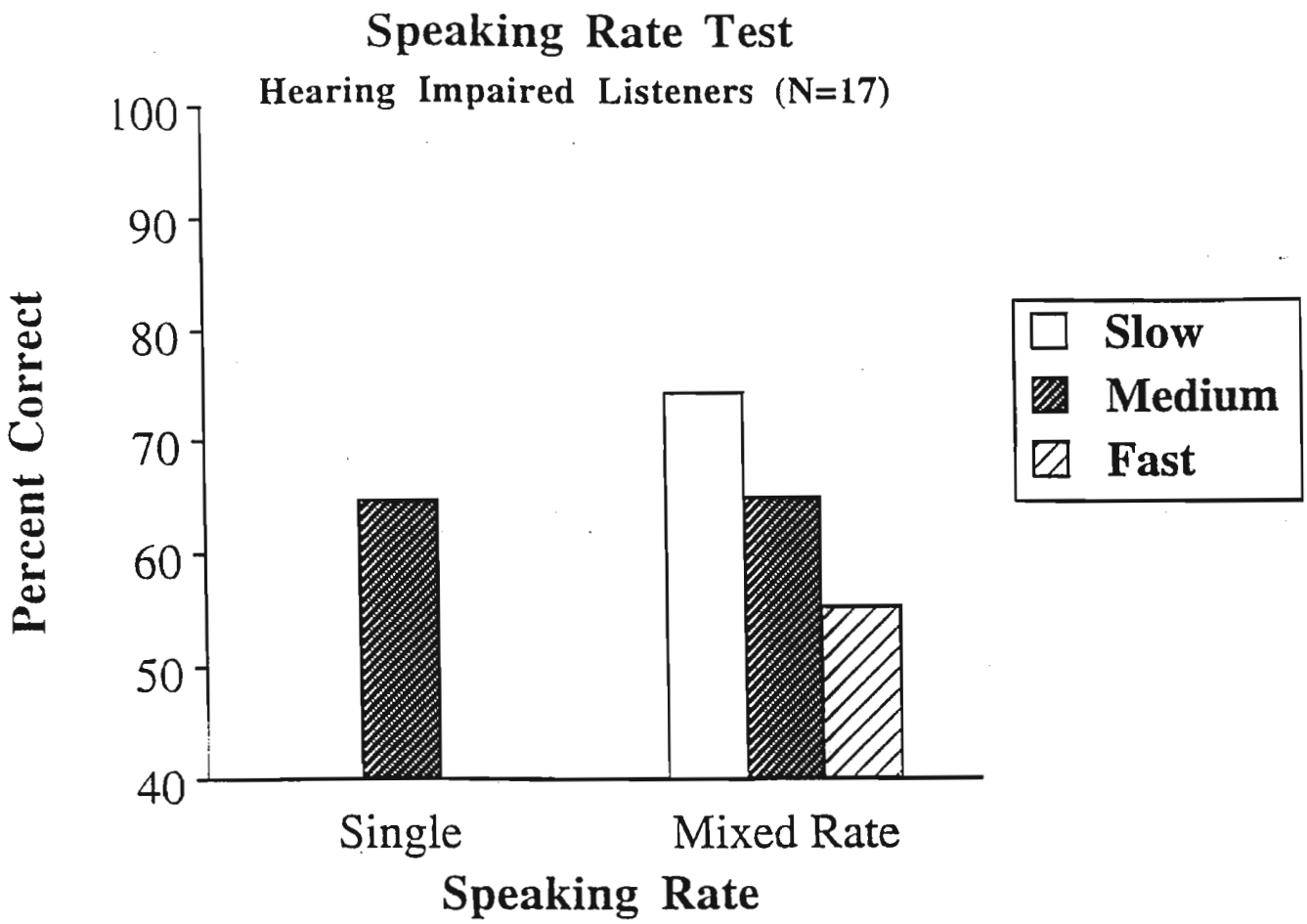

Speaking Rate Test
Hearing Impaired Listeners (N=17)
Percent Correct
100
90
80
70
60
50
40
Single
Mixed Rate
Speaking Rate
Slow
Medium
Fast

TABLE II

SPEECH PERCEPTION SELF-ASSESSMENT QUESTIONNAIRE

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	
Mean	5.53	4.0	5.12	4.53	4.82	5.12	

Table III

Correlations of Talker Variability & Speaking Rate Test Scores with PHAP Self-Assessment Questionnaire

	TALKER VARIABILITY TEST				SPEAKING RATE TEST			
	Single-talker		Multiple-talker		Single-Rate		Mixed-Rate	
	Easy Words	Hard Words	Easy Words	Hard Words	Medium	Slow	Medium	Fast
Total PHAP:	.01	.37	.58***	.49**	.27	.44*	.42*	.49**
Subscales:								
1. Familiar Talkers (FT)	-.06	.22	.44*	.28	.07	.33	.34	.43*
2. Reduced Cues (RC)	-.04	.34	.56***	.37	.21	.43*	.40*	.40*
3. Background Noise (BN)	-.11	.28	.48**	.51**	.15	.28	.27	.22
4. Distortion of Sound (DS)	.21	.56***	.65****	.57***	.55***	.59****	.60****	.65****
5. Gender (G)	.16	-.06	.12	.18	.12	.06	.00	.16
6. Speaking Rate (SR)	-.20	.19	.36	.28	-.01	.21	.19	.30

Significance Levels: * = p<.05; ** = p<.025; *** = p<.01; **** = p<.005

Discussion

Using traditional clinical criteria, these hearing-impaired listeners display "good" to "excellent" speech discrimination performance. Their NU-6 scores ranged from 82% to 100% correct. However, an examination of their performance on the talker variability and rate tests reveals a somewhat different picture of their speech perception abilities when assessed under more demanding conditions where specific sources of variability have been introduced into the testing situation. In each case, their performance was worse under the more perceptually demanding conditions. In the talker variability tests, for example, the listeners' performance was not only lower in the multiple-talker conditions compared to the single-talker conditions, but these listeners apparently had much more difficulty making fine acoustic-phonetic discriminations among lexically difficult words in the multiple-talker conditions. These findings suggest that while hearing-impaired listeners may have reduced frequency selectivity and audibility and generally display reduced abilities in making fine phonetic discriminations, they do perceive highly familiar words "relationally," that is, in the context of other phonetically similar words just as listeners with normal hearing do (Luce et al., 1990).

In the speaking rate tests, listeners consistently displayed their best recognition performance for "slow" words and their worst performance for "fast" words; words spoken at the "medium" rate were recognized in between these extremes. This pattern of results suggests that these listeners were not only sensitive to the talker's speaking rate used in producing these materials but were also responding to the degree of articulatory precision and acoustic-phonetic contrast present in these test items.

Traditional "low-variability" speech discrimination tests such as the NU-6 consistently 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 variability. Indeed, our analyses of the self-assessment questionnaire showed that despite good-to-excellent NU-6 scores, these same listeners reported several perceptual difficulties in everyday communication situations. The pattern of responses from the self-assessment questionnaire revealed that these listeners have specific perceptual difficulties when they have to recognize speech in noise or listen to several competing voices in the background or deal with other factors that require perceptual compensation for stimulus variability in their immediate listening environment. The moderately high correlations between scores on the talker variability and speaking rate tests and several subscales on the PHAP self-assessment questionnaire suggest that these new perceptually demanding tests are measuring several underlying aspects of speech perception in the laboratory that appear to generalize to conditions encountered in "real-world" listening environments.

Conclusions

The problem of "perceptual constancy" has been one of the major theoretical issues in the field of speech perception for over 45 years. Normal listeners are able to perceive and recognize speech at very high levels of accuracy despite substantial acoustic differences in the stimulus materials due to speaker, rate, dialect and surrounding phonetic context. Traditional speech discrimination tests are based on the classical "symbolic" view of human perception, memory and cognition. This theoretical approach assumes that the neural representation of speech is abstract, symbolic and highly idealized and that little, if any, of the fine details of the stimulus pattern or perceptual operations are preserved in memory after perceptual analysis is completed. Recent findings with normal listeners have shown that this view substantially underestimates the amount of stimulus information that is encoded during perceptual analysis (see Pisoni, 1993, for a review).

The results of the present investigation with hearing-impaired listeners suggest the following conclusions:

• New robust measures of speech perception performance which specifically include manipulations of different voices and speaking rates were found to be correlated with listeners' self-assessments of their own hearing impairment. Listeners know and display an awareness of the specific conditions under which they have difficulties in recognizing and understanding speech. Moreover, the overall pattern of performance on these new speech discrimination tests is similar to normal listeners' performance in response to these variables.

• While hearing-impaired listeners typically have some loss of audibility and reduced frequency selectivity, which may be partly responsible for their inability to deal effectively with different sources of variability in speech, they do appear to perceive spoken words "relationally" in the context of other phonetically similar words.

• There is a need for the development of new robust perceptual tests for use with hearing-impaired listeners that can predict speech perception performance outside of the laboratory or clinic in more naturalistic settings which include factors such as background noise, competing voices, different speaking rates and unfamiliar talkers. These new perceptual tests should be designed to measure a listener's abilities to compensate for different sources of variability and maintain perceptual constancy in speech under a variety of conditions.

• Finally, research on listeners with different kinds of hearing-impairments may provide important new information about the underlying perceptual processes and neural mechanisms used in speech perception and spoken word recognition. An examination of the initial sensory representation and neural coding of the speech signal by these listeners provides an opportunity to test and generalize several recent theoretical principles derived from studies of normal listeners who have much more robust and detailed auditory representations of speech signals. We are particularly interested in generalizing the recent claims surrounding the encoding of stimulus variability by the nervous system and the retention of highly detailed episodic information about speech in long-term memory to clinical populations that have difficulty perceiving speech under adverse listening conditions outside the laboratory and clinic.

References

Cox, R.M. & Gilmore, C. (1990). Development of the Profile of Hearing Aid Performance (PHAP). *Journal of Speech and Hearing Research*, **33**: 343-357.

House, A.S., Williams, C.E. Hecker, M.H.L. & Kryter, K.D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. *Journal of the Acoustical Society of America,* **35**: 158-166.

Kirk, K.I., Sommers, M.S., Pisoni, D.B. & Osberger, M.J. (1993). Auditory word recognition by pediatric cochlear implant users: Lexical and indexical influence. Paper presented at the *Conference on Implantable Auditory Prostheses*, Smithfield, RI, July, 1993.

Luce, P. (1986). Neighborhoods of words in the mental lexicon. *Research on Speech Perception Technical Report No. 6*. Bloomington, IN: Speech Research Laboratory, Indiana University.

Luce, P.A., Pisoni, D.B. & Goldinger, S.D. (1990). Similarity neighborhoods of spoken words. In G.T.M. Altmann (Ed.) *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*. Cambridge, MA: MIT Press.

Nusbaum, H.C., Pisoni, D.B. & Davis, C.K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report No. 10*. Bloomington, IN: Speech Research Laboratory, Indiana University.

Pisoni, D.B. (1992). Some comments on invariance, variability and perceptual normalization in speech perception. *Proceedings 1992 International Conference on Spoken Language Processing*, Banff, Canada, Oct. 12-16, 1992.

Pisoni, D.B. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech Communication,* **13**: 109-125.

Pisoni, D.B. & Luce, P.A. (1986). Speech perception: Research, theory and the principal issues. In E.C. Schwab and H.C. Nusbaum (Eds.), *Pattern Recognition by Humans and Machines: Speech Perception*, (Vol. 1, pp. 1-50). New York: Academic Press.

Sommers, M.S., Kirk, K.I. & Pisoni, D.B. (1993). Some new directions in evaluating the speech perception abilities of cochlear implant patients: A preliminary report. Paper presented at the *Conference on Implantable Auditory Prostheses,* Smithfield, RI, July, 1993.

Sommers, M.S., Nygaard, L.C. & Pisoni, D.B. (1992). Stimulus variability and the perception of spoken words: Effects of variations in speaking rate and overall amplitude. *Proceedings 1992 International Conference on Spoken Language Processing*, Banff, Canada, Oct. 12-16, 1992.

Studdert-Kennedy, M. (1974). The perception of speech. In T.A. Sebeok (Ed.) *Current Trends in Linguistics* (Vol. 2, pp. 1-62). New York: Academic Press.

Tillman, T.W. & Carhart, R. (1946). An expanded test for speech discrimination utilizing CNC monosyllabic words. *Northwestern University Auditory Test No. 6. Technical Report No. SAM-TR-66-55.* USAF School of Aerospace Medicine, Brooks Air Force Base, TX.