

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 21 (1996-1997)
Indiana University

**Some Factors Affecting Recognition of Spoken Words
by Normal Hearing Adults¹**

Ann R. Bradlow,² Gina M. Torretta,³ and David B. Pisoni

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

¹ This research was supported by NIH-NIDCD Training Grant DC-00012 and by NIH-NIDCD Research Grant DC-00111 to Indiana University. We are grateful to Luis Hernandez for technical support.

² Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL

³ Central Institute for the Deaf, St. Louis, MO

Some Factors Affecting Recognition of Spoken Words by Normal Hearing Adults

Abstract. An analysis of intelligibility data from a carefully constructed database of recorded speech was conducted in order to investigate the combined effects of various talker-, listener-, and item-related characteristics that contribute to variability in intelligibility of isolated words. Materials came from the Indiana Multi-Talker Word Database, which consists of a set of recorded words from multiple talkers at three speaking rates along with intelligibility data in the form of transcriptions by a large number of native English listeners. Results showed a strong effect of lexical discrimination (“easy” words had higher intelligibility scores than “hard” words), and a strong effect of speaking rate (slow and medium rate words had higher intelligibility scores than fast rate words). Furthermore, we observed a complex relationship between the various factors whereby the difficulties imposed by one factor, such as a fast speaking rate or an inherently difficult lexical item, could be overcome by the advantage gained by the listener’s experience with the speech of a particular talker. Implications of these findings for the development of spoken language processing assessment instruments and assistive devices for “special listener populations” are discussed.

Introduction

It is well known that even under ideal listening and speaking conditions, transmission accuracy of the speaker’s intended message to the listener often varies greatly. Recent work in our laboratory has focused on some of the factors that contribute to the observed variability in normal speech intelligibility. To date, several factors have been shown to directly influence overall speech intelligibility. First, the degree of variability in the stimulus materials has been shown to have a major impact on the listener’s speech recognition accuracy. For example, word recognition accuracies decrease and response times increase when listeners are presented with spoken word lists that incorporate a high-degree of stimulus variability due to the presence of multiple talkers and speaking rates, relative to spoken word lists in which such stimulus variability is minimized (Mullennix et al., 1989; Sommers et al., 1994). Second, familiarity on the part of the listener’s with the talker’s voice and articulatory characteristics enhances word recognition accuracy under difficult listening conditions. For example, Nygaard, Sommers and Pisoni (1994) showed that listeners were more accurate at identifying words in noise when spoken by a familiar talker than when spoken by a novel talker. Third, lexical characteristics of the particular words in a stimulus set exert a strong influence on overall intelligibility. Several studies have shown that “easy” words (i.e., words with few phonetically similar “neighbors” with which they could be confused) have a distinct intelligibility advantage over “hard” words (i.e., highly confusable words with many phonetically similar neighbors) (Pisoni et al., 1985; Luce, 1986; Luce et al., 1990). Finally, in a recent study of the talker-specific acoustic-phonetic characteristics that correlate with inter-talker intelligibility differences, Bradlow et al. (1997) showed that talkers who exhibited a high-degree of “articulatory precision” in their speech generally had higher overall speech intelligibility scores than talkers who tended to produce more “reduced” speech. Taken together, these studies demonstrate the range of stimulus-, listener- and talker-related factors that combine to result in the observed variability in normal speech intelligibility.

The present study continues this line of research by investigating the combined effects of various talker-, listener-, and item-related characteristics that contribute to overall intelligibility of isolated words in

a carefully constructed database of recorded speech. Materials for this study came from the Indiana Multi-Talker Word Database (Torretta, 1995), which consists of a set of recorded words from multiple talkers at three speaking rates along with intelligibility data in the form of transcriptions by a large number of native English listeners. This intelligibility data provided us with the means to assess the combined effects of speaking rate, lexical discrimination, and listener-talker adaptation on isolated word intelligibility by native listeners. By directly examining the effects of these characteristics on native-language word intelligibility, we hoped to obtain a baseline measure of normal variability in isolated word recognition that could then serve as a basis for comparison of word recognition performance by a variety of listeners under various presentation conditions. For example, non-native and hearing-impaired listeners appear to be particularly sensitive to stimulus variability and adverse listening conditions, such as in the presence of multiple-talkers or background noise. Thus, knowledge about the factors that affect normal speech intelligibility by normal listeners may be particularly useful for the development of spoken language processing assessment instruments and assistive devices for "special listener populations."

Method

The "Easy" and "Hard" Word Lists

An "easy" list and a "hard" list of words (75 items each) were compiled such that the two lists differed in terms of three lexical characteristics (Pisoni et al., 1985; Luce, 1986; Luce et al., 1990; Luce and Pisoni, 1997). First, using the word frequency counts provided by the Brown Corpus of printed text (Kucera and Frances, 1967), the words were selected such that the mean word frequency of the easy list was substantially higher than that of the hard list (309.7 vs. 12.2 per million). Second, using an on-line version of Webster's Pocket Dictionary (20,000 entries) in conjunction with a custom-designed lexical search program, words were selected such that the neighborhood density (the number of phonetic "neighbors") of the easy list was lower than that of the hard list (13.5 vs. 26.6). In these neighborhood density counts, a neighbor of a given word was defined as any word that differed from the target word by a one phoneme addition, substitution or deletion in any position. Third, the two word lists were constructed such that the neighborhood frequency (the mean frequency of the neighbors) of the easy list was much lower than that of the hard list (38.3 vs. 282.2 per million). The net result of these three word selection criteria, was that the "easy" list consisted of a set of words that are frequent in the language, and that have few phonetically-similar, low-frequency neighbors with which they could be confused. In contrast, the "hard" list consisted of words with many neighbors that are high in frequency relative to the target word. Easy words "stick out" from sparse neighborhoods; hard words are "swamped" by dense neighborhoods. Finally, in order to ensure that subjects would be familiar with all of the words in both lists, all words were judged as highly familiar by normal-hearing adults, i.e., received a familiarity rating of 6.7 or higher on a 7 point scale where 1 indicated the lowest and 7 indicated the highest degree of familiarity (Nusbaum et al., 1984).

Digital Speech Recordings

Ten talkers (five males and five females) were recorded producing both the easy and the hard word lists at three different speaking rates (fast, medium, and slow), giving a total of 4500 tokens (150 words x 3 speaking rates x 10 talkers). None of the talkers had any known speech or hearing impairments at the time of recording, and all were native speakers of General American English. The talkers were told in advance that they would be asked to produce three word lists of 150 words each at three different speaking rates. Each individual talker was allowed to regulate his/her own speaking rate, so long as the three rates were distinct. An analysis of the word durations for each talker at each of the three rates, confirmed that each

talker successfully produced the three lists with three distinct speaking rates. The mean durations were 809 ms (range 576-1030 ms), 525 ms (range 466-579 ms), and 328 ms (range 264-413 ms) for the slow, medium, and fast words, respectively.

All 150 words (75 easy plus 75 hard) were presented to the talkers in random order on a CRT monitor in a sound-attenuated booth (IAC 401A). The stimuli were transduced with a Shure (SM98) microphone, and digitized on-line (16-bit analog-to-digital converter (DSC Model 240) at a 20 kHz sampling rate). The recordings were all live-monitored by an experimenter for gross misarticulations and hesitations. Each individual digital file was then edited by hand to remove the silent portions at the beginning and end of each word file. The average root means square amplitude of each of the digital speech files was then equated. Finally, the files were converted to PC WAV format for presentation to listeners using a PC-based perceptual testing system (Hernandez, 1995).

Speech Intelligibility Tests

Speech intelligibility scores were collected from independent groups of ten normal-hearing listeners, each of whom transcribed the full set of 150 words from one talker at one speaking rate, for a total of thirty groups of ten listeners (10 talkers x 3 speaking rates). The words were presented to the listeners in random order over matched and calibrated DT-100 headphones via a PC-based perceptual testing system (Hernandez, 1995). The words were presented in the clear (no background noise was added) at a comfortable listening level (75 dB/SPL). On each trial, the listeners heard the word and then typed in the response on the computer keyboard. In the data scoring, a word was counted as correct if all of the letters were present and in the correct order, if all the letters were present but not in the correct order, or if the transcribed word was a homophone of the intended word.

These transcription scores provided a means of investigating the effects of speaking rate (fast vs. medium vs. slow) and lexical discrimination (easy vs. hard) on isolated word intelligibility. Additionally, since each group of listeners transcribed the full set of 150 words by a single talker at a single rate in a single transcription session, we could also use these intelligibility data to investigate whether listeners adapted to talker-specific characteristics to the extent that the intelligibility scores improved from the beginning to the end of the transcription session. We hypothesized that this kind of listener-talker "attunement" on the part of the listener, which occurs over the course of exposure to the speech of a particular talker, would interact with the lexical (easy vs. hard) and speaking-rate (fast vs. medium vs. slow) factors such that there would be a greater listener-talker adaptation effect as the other factors increased in difficulty. Such a finding would indicate that listener-talker familiarity can compensate for the word recognition difficulties associated with increased speaking rate and easily-confused lexical items.

Results

Figure 1 shows the overall percent correct transcription scores across all talkers and listeners for the easy and hard word lists at each of the three speaking rates. As expected based on earlier investigations of the effects of these lexical characteristics on speech perception (Pisoni et al., 1985; Luce, 1986; Luce et al., 1990), the easy word lists were generally more accurately transcribed than the hard word lists. As shown in Table I, the higher transcription accuracy for the easy list relative to the hard list held true for almost all speakers at all three speaking rates. The exception were for Talkers 1, 5, 6 and 9 at the slow rate, where there was no easy-hard difference, and for Talker 6 at the medium rate where there was a very small advantage for the hard word list. Thus, the word identification advantage for easy words over hard words is a highly robust effect that generalizes across multiple talkers and speaking rates.

 Insert Figure 1 about here

Figure 1 also shows a substantial decline in transcription accuracy for the fast rate relative to the medium and slow rates for both the easy and the hard word lists. However, there was little difference in transcription accuracy between the slow and medium rate words. This pattern of results was somewhat surprising in view of the fact that, on average, the slow words were about 54% longer than the medium words. Thus, it appears that isolated word intelligibility is not enhanced by slowing the speaking rate.

These findings were all confirmed by a repeated-measures ANOVA (nested design) with both rate (fast, medium, slow) and lexical category (easy, hard) as within subject variables, and the intelligibility scores for each talker in each condition averaged across all ten listeners as the dependent variable (see Table I). There was a main effect of rate ($F(2,18)=7.456, p=.0013$), and a main effect of lexical category ($F(1,18)=20.111, p=.0015$). An examination of the contrasts showed a significant difference (at the $p<.005$ level) between the fast and medium rates for both the easy and the hard words, but no difference between the medium and slow rates for either the easy or the hard words. Furthermore, at all three rates, the easy vs. hard difference was significant at the $p<.005$ level.

TABLE 1.

Mean intelligibility scores across all ten listeners for the easy and hard word lists by each talker at each speaking rate.

Talker	Easy			Hard		
	Slow	Medium	Fast	Slow	Medium	Fast
1	91.07	92.40	86.13	82.67	81.20	72.27
2	94.40	95.47	94.27	94.80	94.40	89.33
3	94.67	94.00	94.93	88.93	89.60	92.53
4	92.40	96.00	88.27	88.67	87.20	78.00
5	94.00	94.40	86.27	89.47	91.33	75.47
6	92.93	93.87	91.87	92.80	90.40	89.73
7	90.67	89.20	89.47	91.07	90.26	87.87
8	94.93	96.27	92.93	93.60	88.40	89.47
9	95.07	96.67	95.73	92.40	92.13	84.40
10	95.07	98.40	96.27	94.93	95.46	90.67
mean	93.52	94.67	91.61	90.93	90.04	84.97

The next step in our analysis of these intelligibility data was to investigate whether isolated word intelligibility can be enhanced as the listener becomes accustomed to the talker's voice. In particular, we wondered whether hard words that were presented later in a transcription session would be more accurately transcribed than hard words presented earlier in the session. In other words, we were interested in seeing whether listener-talker adaptation might compensate for the processing difficulties introduced by the lexical confusability factor.

Figure 2 shows the percent correct transcription scores for the easy and hard words in the first quartile (Q1) and fourth quartile (Q4) of the transcription sessions at the fast (upper panel), medium (middle panel), and slow (bottom panel) speaking rates. In each case the first and fourth quartiles were

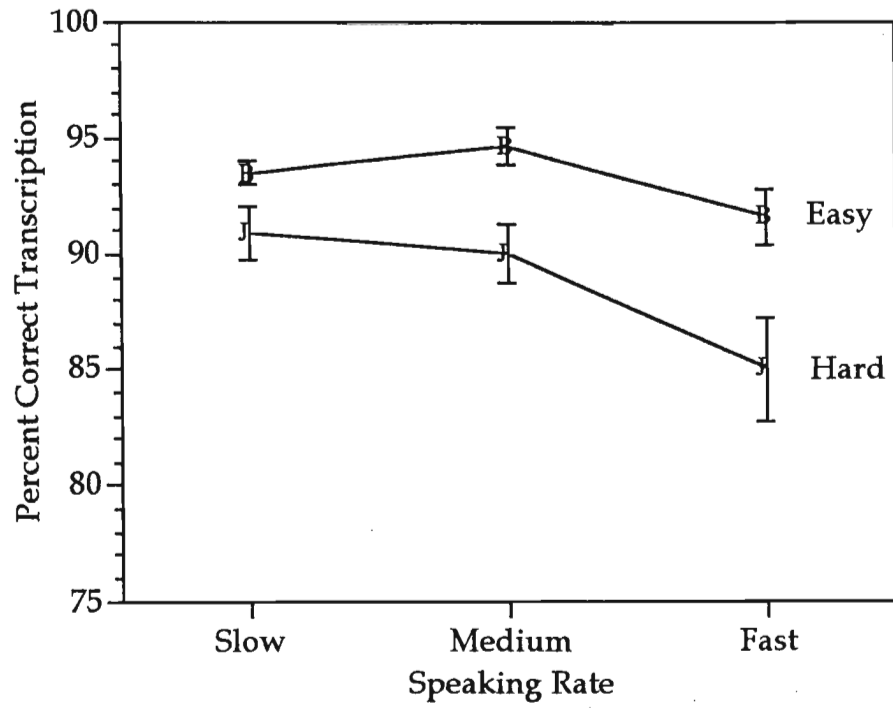


Figure 1. Transcription accuracy for the easy and hard word lists at slow, medium, and fast speaking rates.

taken as the first and last 38 words presented to the listeners, respectively. As shown in Figure 2, hard words presented in the last quartile were generally more accurately transcribed than hard words presented in the first quartile at all three speaking rates. In contrast, there was no noticeable difference between easy words presented in the first and fourth quartiles at all three speaking rates. Separate ANOVA's for each speaking rate showed that for all three rates there was a main effect of quartile, such that the Q4 intelligibility scores were higher than the Q1 intelligibility scores. There was also a main effect of lexical discrimination, such that easy words had higher intelligibility scores than hard words. Furthermore, the quartile by lexical category interaction was significant. Post-hoc tests showed that at all three speaking rates the Q4-Q1 difference was significant for the hard words, but not for the easy words.

 Insert Figure 2 about here

These data indicate that as the listener becomes accustomed to the talker's voice and articulatory patterns, the intelligibility difficulty introduced by the lexical characteristics of hard words relative to easy words is "neutralized" to a large extent. Furthermore, a comparison of the first and fourth quartile intelligibility scores across the three speaking rates (see Table 2) showed that the intelligibility of fast rate words in the fourth quartile (mean = 89.67%) approached the intelligibility scores for the slow and medium rate words in the first quartile (means = 90.80% and 90.05%, respectively). In other words, the listener's experience with the talker's speech compensated for the intelligibility difficulty introduced by the fast speaking rate. In general, this pattern of results suggests that listener-talker adaptation is an important factor that interacts with other talker- and item-related factors, such as speaking rate and lexical characteristics, in determining the overall intelligibility of normal speech by normal listeners.

TABLE 2.

**Mean intelligibility scores for each speaking rate
 in the first and fourth quartile.**

	First Quartile	Fourth Quartile
Slow	90.80	92.90
Medium	90.05	93.04
Fast	85.98	89.67

Discussion

The primary goal of this study was to examine the combined effects of various talker-, item-, and listener-related factors on normal speech intelligibility by normal listeners. Results showed a strong effect of lexical discrimination (easy words had higher intelligibility scores than hard words), and a strong effect of speaking rate (slow and medium rate words had higher intelligibility scores than fast rate words). Furthermore, we observed a complex relationship between the various factors whereby the difficulties imposed by one factor, such as a fast speaking rate or an inherently difficult lexical item, could be overcome by the advantage gained by the listener's experience with the speech of a particular talker.

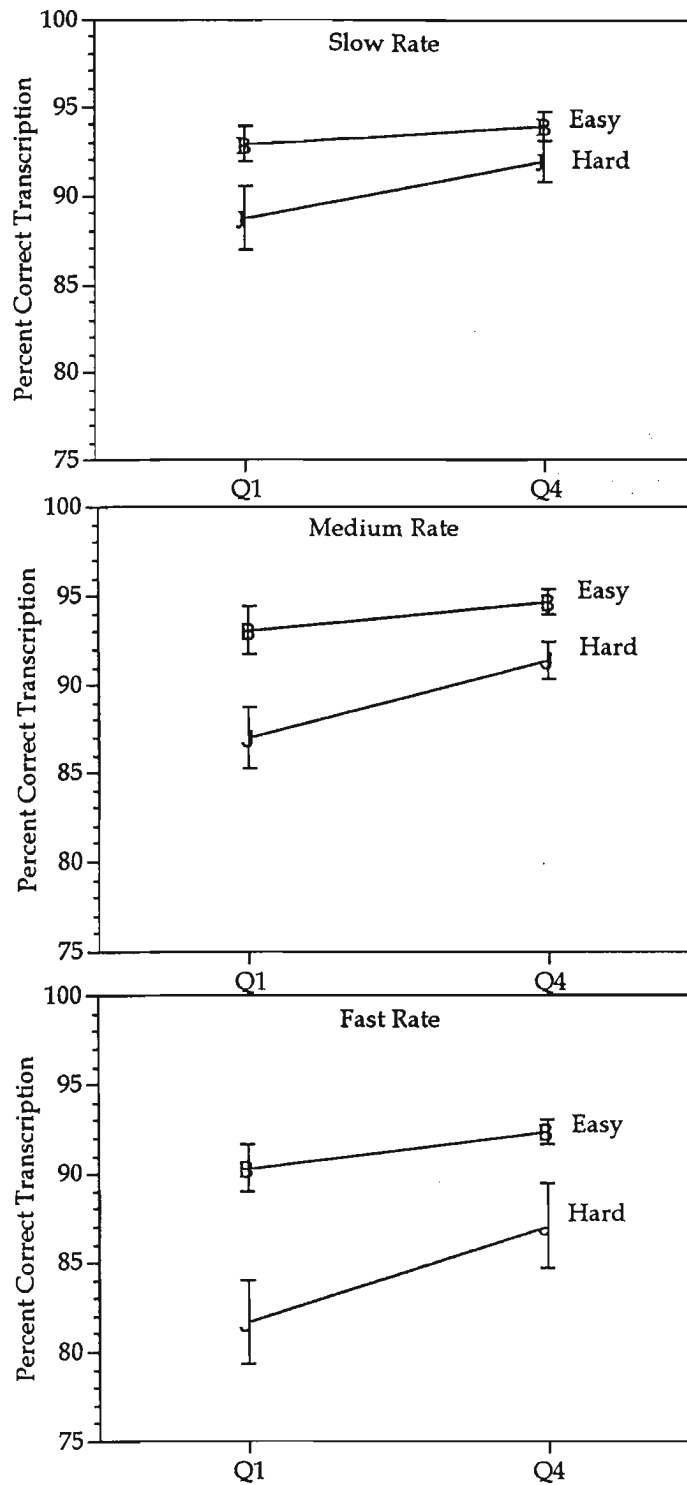


Figure 2. Comparison of transcription accuracies for the easy and hard words presented in the first (Q1) and fourth (Q4) quartile of the transcription sessions at the fast (upper panel), medium (middle panel), and slow (lower panel) speaking rates.

In our investigation of the factors that affect normal speech intelligibility, the present study focused on factors that are related to capabilities unique to speech perception and spoken-language processing. For example, the acoustic-phonetic changes that typically occur as a consequence of a change in speaking rate are directly related to the sound system of the language which imposes limits on the relative expandability and compressibility of various acoustic-phonetic elements. Similarly, the distinction between easy and hard words is directly related to the structure of the lexicon as a whole. Finally, the type of perceptual adaptation that we observed on the part of the listener to the speech patterns of the talker is directly related to the listener's knowledge of the range of possible within-talker variability given the phonetic requirements of the language. Thus, while we do not mean to minimize the importance of basic psychoacoustic capabilities for auditory perception, including speech perception, our focus has been on the higher-level cognitive and linguistic capabilities that are essential for "robust" speech perception and spoken language processing. This focus reflects our concern with the vulnerability of these capabilities in other populations, such as hearing impaired children and adults, non-native listeners, and the elderly.

As we gain a deeper understanding of the operations that are involved in normal speech perception we can begin to develop new tests, and ultimately new training procedures, that focus directly on the complex cognitive and linguistic capabilities that are critical for robust speech perception. Based on the findings of the present study in conjunction with those of other studies reviewed in the introduction, we can delineate several factors that a sensitive robust test of speech perception and spoken language processing should attempt to address. First, the test should examine how listeners cope with stimulus sets that incorporate a high degree of variability due to, for example, multiple talkers and speaking rates. Second, the test should investigate the extent to which listeners perceive words in the context of other words in the lexicon, that is, the extent to which listeners display evidence of having developed a phonetically structured lexicon in long-term memory. Finally, the test should assess the listeners ability to compensate for stimulus-related difficulties by taking advantage of consistent aspects of the speech signal, such as the listener-talker adaptation that we observed in the present data. Such tests are already under development for use with various clinical populations (Kirk et al., 1995; Sommers et al., 1997; Sommers, 1997; Kirk et al., in press). We expect that further development of speech perception assessment instruments that take these factors into account will find a wide range of highly beneficial clinical and research applications.

References

- Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, 20, 255-272.
- Hernandez, L. R. (1995). Current computer facilities in the Speech Research Laboratory. *Research on Spoken Language Processing, Progress Report*, 20, 389-394, Indiana University, Bloomington, IN.
- Kirk, K. I., Pisoni, D. B., & Miyamoto, R. C. (In press). Effects of stimulus variability on speech perception in listeners with hearing impairment. *Journal of Speech, Language, and Hearing Research*.
- Kirk, K. I., Pisoni, D. B., & Osberger, M. J. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear and Hearing*, 16, 470-481.
- Kucera, F. & Francis, W. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

- Luce, P. A. (1986). Neighborhoods of words in the mental lexicon. *Research on Speech Perception, Technical Report No. 6*, Indiana University, Bloomington, IN.
- Luce, P. A., Pisoni, D. B., and Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistics and computational perspectives*. Cambridge, MA: MIT Press.
- Luce, P. A. & Pisoni, D. B. (In press). Recognizing spoken words: the Neighborhood Activation Model. *Ear and Hearing*.
- Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, 85, 365-378.
- Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research in Speech Perception, Progress Report* , 10, 357-376, Indiana University, Bloomington, IN.
- Nygaard, L. C., Sommers, M. C., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5, 42-46.
- Pisoni, D. B., Nusbaum, H. C., Luce, P. A. and Slowiaczek, L. M. (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication*, 4, 75-95.
- Sommers, M. S. (1997). Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment. *Journal of the Acoustical Society of America*, 101, 2278-2288.
- Sommers, M. S., Kirk, K. I., & Pisoni, D. B. (1997). Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format. *Ear and Hearing*, 18, 89-99.
- Sommers, M. S., Nygaard, L. C. & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition: I. Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America*, 96, 1314-1324.
- Torretta, G. M. (1995). The easy-hard word multi-talker speech database: An initial report. *Research on Spoken Language Processing, Progress Report* , 20, 321-334, Indiana University, Bloomington, IN.