

---

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**  
Progress Report No. 20 (1995)  
*Indiana University*

**The “Easy-Hard” Word Multi-Talker Speech Database:  
An Initial Report<sup>1</sup>**

**Gina M. Torretta**

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

---

<sup>1</sup> This research was supported by NIH-NIDCD Research Grant DC00111 and NIH-NIDCD Training Grant DC-00012 to Indiana University-Bloomington. I gratefully acknowledge the assistance of Benedetta Mariotti and Melissa Kluck in preparing the stimulus materials, and Bill Svec for assistance with program and perceptual training system (PTS) development. I am indebted to Ann Bradlow for her generosity and patience in guiding data analyses and to Mitch Sommers for his input on data collection.

## The "Easy-Hard" Word Multi-Talker Speech Database: An Initial Report

**Abstract.** One research strategy used in our laboratory has focused on the development of large speech databases as general purpose resources for use in various experiments. The speech database described here incorporates several variables known to affect spoken word recognition. The database consists of 4500 digital speech files. Five male and five female talkers each produced a list of 150 words at three different speaking rates. The basic word list contains 75 lexically "easy" and 75 lexically "hard" words as defined by the Neighborhood Activation Model (NAM) of Luce (1986). This report provides a description of the database as well as some intelligibility data for the medium and fast rate tokens.

### Introduction

The development of large speech databases is one research strategy used in our laboratory. The goal of the present project was to develop a carefully constructed speech database that can be accessible in CD-ROM format for general use. There are three phases in the development of the database. Phase one is stimulus preparation, including word list formulation, speech digitization, and speech file preparation. Phase two is the collection of intelligibility data, and statistical analyses of listener responses to the utterances. Phase three will involve acoustic-phonetic analyses of a subset of the digital speech files. Currently, the project is in phase two. In this initial report, both phase one and phase two are described in detail, as well as the intended progression of phase three.

The Easy-Hard Word Database consists of 4500 digital speech files. The heart of the database is a set of 150 English words that vary in both word frequency and lexical similarity. The set of words was recorded from ten talkers, five males and five females, at three different speaking rates (150 words x 3 rates x 10 talkers = 4500 tokens). The database is intended to provide a set of speech materials for experimentation in areas of speech perception and spoken word recognition, as well as acoustic-phonetic analyses.

### Phase 1: Stimulus Preparation

Previous work in our laboratory has shown that lexical characteristics such as word familiarity, lexical frequency and lexical similarity have behavioral consequences in a variety of spoken word recognition tasks (Luce, 1986; Luce et al., 1990; Kirk et al., 1995; Sommers et al., 1995). *Word familiarity* is a subjective rating of familiarity, or perceived commonness of words (Nusbaum et al., 1984). *Lexical frequency* is defined as the average number of times a word occurs in printed text (Kucera & Francis, 1967). *Neighborhood density* is the number of "neighbors" or words that differ by one phoneme from the target word. An example of a lexical neighborhood for the target word "pat" are words such as "cat, bat, pit, pet, spat and pats." The word "pat" has more neighbors than, for example, the word "phone." The lexical neighborhood for "pat" would therefore be considered more "dense" than the lexical neighborhood for the word "phone." Figure 1 shows the relationships among these three stimulus variables.

-----  
Insert Figure 1 about here  
-----

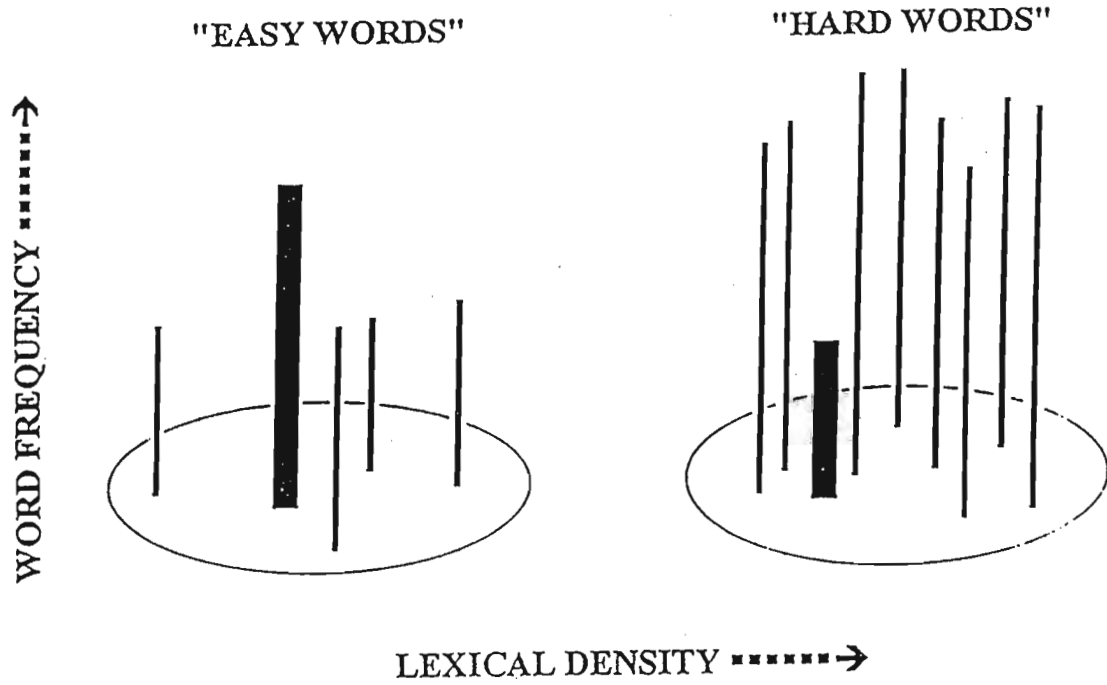


Figure 1: The Neighborhood Activation Model (Luce, 1986): Word frequency, familiarity and lexical density are characteristics of "lexical neighborhoods" that influence spoken word recognition.

In this figure, each bar represents a target word, and the lexical neighborhoods are enclosed by the ovals. Word familiarity is represented by bar width, word frequency is represented by bar height, and neighborhood density is represented by the number of bars within an oval. The Neighborhood Activation Model (NAM) of Luce (1986) provides a computational analysis of the relationship between these factors. According to NAM, the number of nearest neighbors can inhibit or enhance the correct selection of a target word from lexical memory. Lexically "easy" words come from "sparse" similarity neighborhoods, whereas lexically "hard" words come from "dense" similarity neighborhoods. Furthermore, "easy" words have a higher frequency than the mean frequency of other words in the neighborhood, whereas "hard" words have lower frequencies than the mean neighborhood frequency. In other words, "easy" words are perceptually distinctive and "stick out" from their similarity neighborhoods, whereas, "hard" words are "swamped" by their neighbors.

The word list for this database was constructed using a computer-readable version of Webster's Pocket Dictionary (Pisoni et al., 1985; Nusbaum et al., 1984). Using an in-house lexical search program (SRLEX), monosyllabic CVC words with high familiarity ratings (>6.7 on a 7 point scale, where 1 indicated the lowest and 7 indicated the highest degree of familiarity) were selected. Lexical frequency and neighborhood density varied in the list. The 75 "easiest" words and the 75 "hardest" words were then selected from the list. The 75 "easy" words are high frequency words (mean=310 occurrences per million) with relatively few neighbors (mean=14). These are words from "sparse" lexical neighborhoods. In contrast, the 75 "hard" words have relatively low frequency (mean=12 words per million), and more phonetically similar neighbors (mean=27). These words come from "dense" similarity neighborhoods. The mean word frequency for the "easy" words is higher than the mean neighborhood frequency (309.7 versus 38.3 occurrences per million), whereas the opposite is true for the "hard" words (12.2 versus 282.2 occurrences per million for the word frequency and neighborhood frequency, respectively). Table 1 gives the means for lexical frequency, familiarity, neighborhood density and neighborhood frequency for the subsets of 75 "easy" and 75 "hard" words.

The next step was to make audio recordings of this set of words spoken by various talkers. The talkers sat in a sound-attenuated IAC booth in front of a microphone and read a randomized word list displayed on a CRT monitor. Talkers were instructed to speak in a normal speaking voice, varying only the rate of speech as appropriate for each of the three different speaking rate conditions.

All speech samples were low-pass filtered at 10kHz and digitally sampled at 20kHz with 16-bit resolution using a DSC Model 240 analog-to-digital converter interfaced to a VAX station 3500. After digital sampling, the speech files were edited using a digitally controlled waveform editor to remove the silent portions on either side of the word and to visually check the waveform to ensure speech sample integrity. Rerecordings were performed in appropriate cases (e.g., speaking level too loud/ too soft) and retained for the final version of the speech database.

After segmentation, files were prepared for presentation in our newly implemented PTS lab (Hernandez, 1995). Overall RMS (root-mean-square) amplitude levels for each speech file were digitally equated to ensure equal presentation levels. Following leveling, files were converted to WAV format for presentation in the PC laboratory.

**Table 1****Means of lexical characteristics of words in the Easy-Hard Word Database.**

	<b>Easy</b>	<b>Hard</b>
<b>Frequency</b>	309.7	12.2
<b>Familiarity</b>	7.0	6.8
<b>Density</b>	13.5	26.6
<b>Neighborhood Frequency</b>	38.3	282.2

**Phase 2: Intelligibility Assessment**

The purpose of this new database is to provide researchers with a large sample of carefully selected spoken words by multiple talkers produced at several speaking rates. In addition to having the database of speech tokens, intelligibility data were also collected. The results of the intelligibility tests for both the medium and fast rate of speech are reported here as well.

**Subjects**

Two-hundred subjects from Indiana University participated in intelligibility data collection. All subjects received course credit for their participation in the experiment. All subjects had normal hearing and reported no history of speech or hearing disorders at the time of testing.

**Procedure**

An intelligibility testing program developed for the PTS system (Hernandez, 1995) was used to present the stimuli to subjects and record their responses. For both the medium and fast speaking rates, the 150 words from each of the 10 talkers were presented binaurally in the clear over matched and calibrated DT-100 Beyerdynamic headphones to 10 listeners at a comfortable listening level (75 dB/SPL). For each data collection session, a listener heard words spoken by only one talker, and no two listeners heard the same talker during the same session. All listeners received a set of typed instructions explaining the task. Listeners were informed that they would hear a list of English words, and, following each word, they were required to type the word they heard into the computer. All listeners were informed by the experimenter that there would be ample time to check any possible spelling mistakes and make corrections before proceeding with list presentation. Questions and instruction clarifications were handled by the experimenter prior to each data collection session. Following data collection, subjects were debriefed about the nature of the experiment they participated in, and any questions were answered.

## Data Analysis

All output data files were processed twice for error tabulations. The intelligibility data program matched the typed input of each listener to a template list of correct responses to provide an initial correct response assessment. This first pass over listener responses was intended to mark as correct all typed input that exactly matched the template list. Then each response marked as incorrect was examined individually by hand in order to accept or reject "incorrect" marked responses on a case-by-case basis. This second pass was intended to catch any homonyms or misspellings that still formed a possible correct response (e.g., 'mit' for 'mitt'), or miscellaneous responses and misspellings which were identifiable as the intended correct response (e.g., 'wrko' for 'work'; 'lvoe' for 'love'). All errors which were unidentifiable as either misspellings, typing errors, or incorrect responses remained marked as incorrect.

Database tables were compiled to represent the data for each speech rate in several ways. First, responses for all listeners were examined to assess overall performance. Following these analyses, the data were examined by the categories "easy" and "hard" in order to assess lexical as well as talker effects. Second, responses for each random presentation order were compiled so performance could be assessed over the course of the experiment to assess learning effects. Finally, an error database was created in order to compare errors across talkers and listeners.

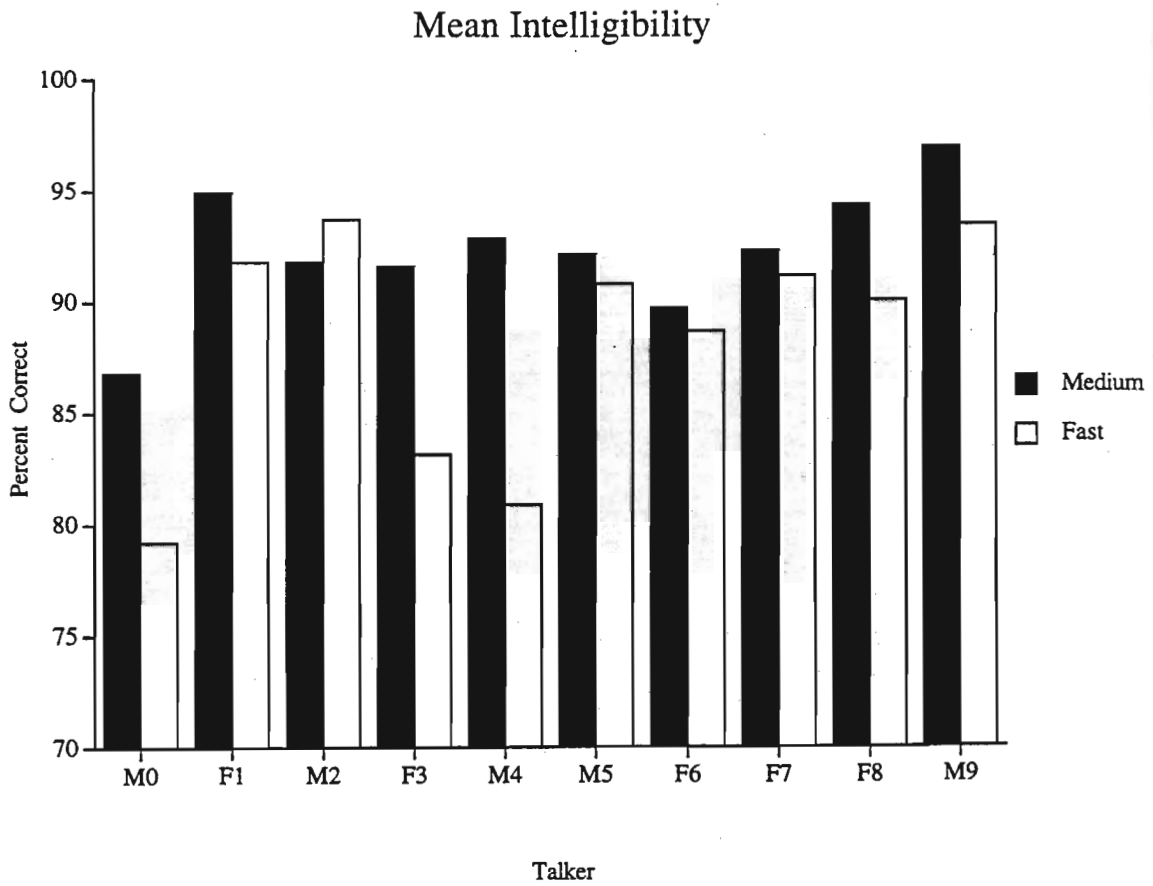
## Results

Overall intelligibility (in terms of percent correct transcription) was scored separately for each speaking rate and for each talker. Figure 2 shows each talker's mean intelligibility score across all 150 words and across all ten listeners for both speaking rates.

-----  
 Insert Figure 2 about here  
 -----

The mean intelligibility score for the medium rate was 92% correct with a range of 86.7% to 96.9% correct; the mean score for the fast rate was 88% with a range of 79.2% to 93.7% correct. A repeated measures ANOVA, with listener response as the repeated measure, indicated a main effect of talker for both the medium and fast rates ( $F(9,148)=7.41$ ,  $p<.0001$ , and  $F(9,148)=12.81$ ,  $p<.0001$ , respectively) across all 150 words. This pattern indicates significant differences in overall intelligibility across the ten talkers for both speaking rates.

Figure 3 shows the effect of lexical category ("easy" vs. "hard") on intelligibility scores for each talker at the medium (Figure 3a) and fast (Figure 3b) speaking rates. A repeated measures ANOVA, with listener response as the repeated measure, showed a main effect of lexical category for both rates ( $F(1,148)=13.19$ ,  $p=.0003$ , and  $F(1,148)=13.96$ ,  $p=.0004$ , for medium and fast rates respectively). There were also significant interactions between talker and lexical category ( $F(9,1332)=3.22$ ,  $p=.0007$ , and  $F(9,1332)=2.33$ ,  $p=.013$ , for medium and fast rates respectively). The listeners' responses across words were also assessed. A repeated measures ANOVA showed that individual listener performance did not vary significantly across words for either speaking rate.



**Figure 2:** The effect of speaking rate on mean intelligibility scores.

-----  
 Insert Figure 3 about here  
 -----

Changes in listener performance over the course of the experiment were also assessed for each speaking rate by comparing intelligibility scores for the first and last quartile ( $n=38$ ) of the randomly ordered word list presented in each data collection session. Figure 4 shows percent correct transcription scores for each quartile for each talker at the medium (Figure 4a) and fast (Figure 4b) speaking rates. A repeated measures ANOVA, with listener response as the repeated measure, showed a main effect of quartile ( $F(1,90)=18.61$ ,  $p<.0000$ , and  $F(1,90)=28.08$ ,  $p<.0000$ , for medium and fast rates respectively). Performance was consistently better in the last quartile than the first, and this effect was independent of the specific test item presentation orders, which differed from listener to listener.

-----  
 Insert Figure 4 about here  
 -----

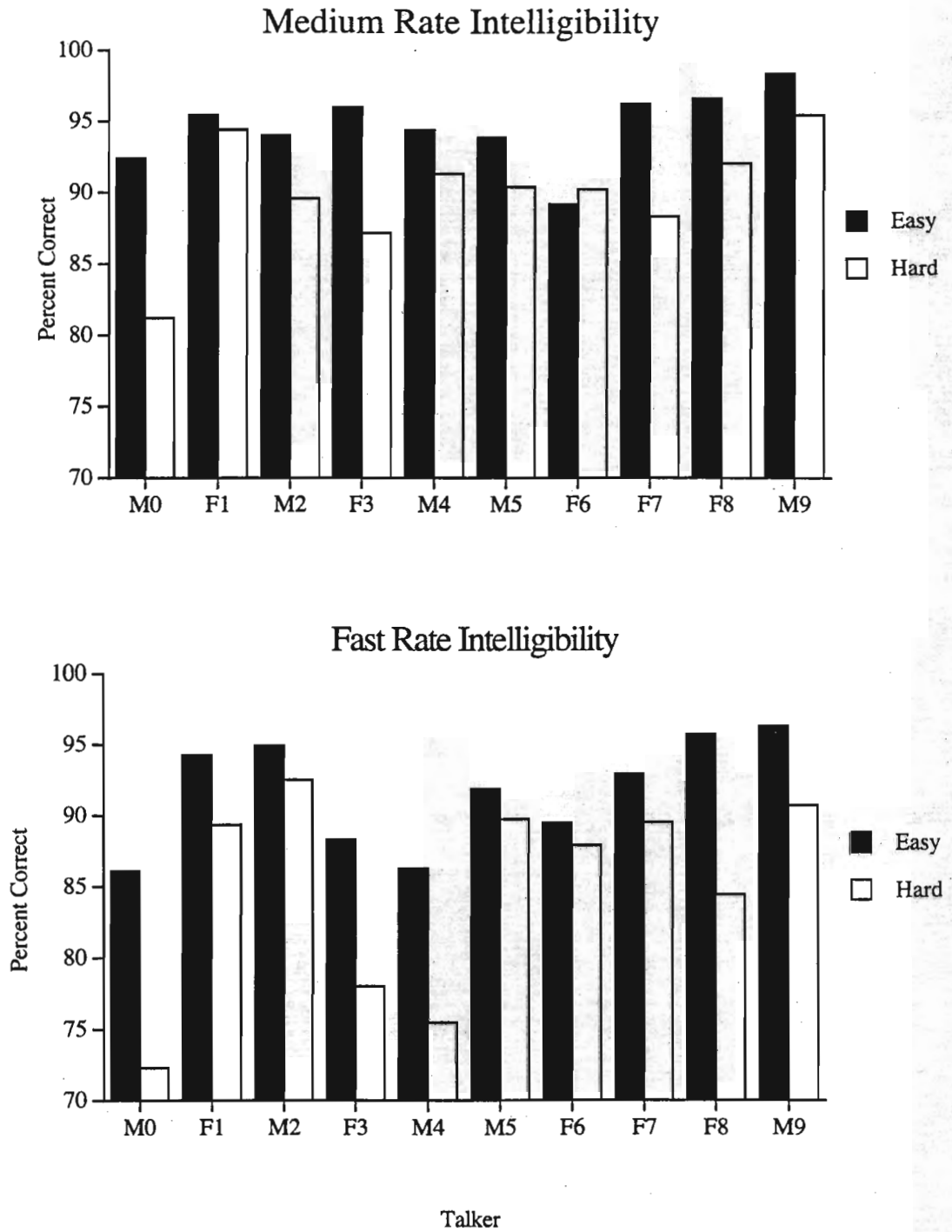
Taken together, the results shown in Figures 3 and 4 indicate that “easy” words were more accurately recognized than “hard” words (shown in Figure 3), and that listener performance improved from the first to fourth quartile of trials presented during a test session (shown in Figure 4). In addition to these main effects, there was also an interaction between quartile and lexical category for both medium and fast speaking rates ( $F(1,90)=5.53$ ,  $p<.0209$ , and  $F(1,90)=5.64$ ,  $p<.0196$ , respectively). Figure 5 shows the overall intelligibility scores for the “easy” and “hard” words in the first and fourth quartiles of the randomly ordered word lists at the medium (shown in Figure 5a) and fast (shown in Figure 5b) speaking rates. The graph demonstrates that the transcription scores improved more for the lexically “hard” words than for the lexically “easy” words.

-----  
 Insert Figure 5 about here  
 -----

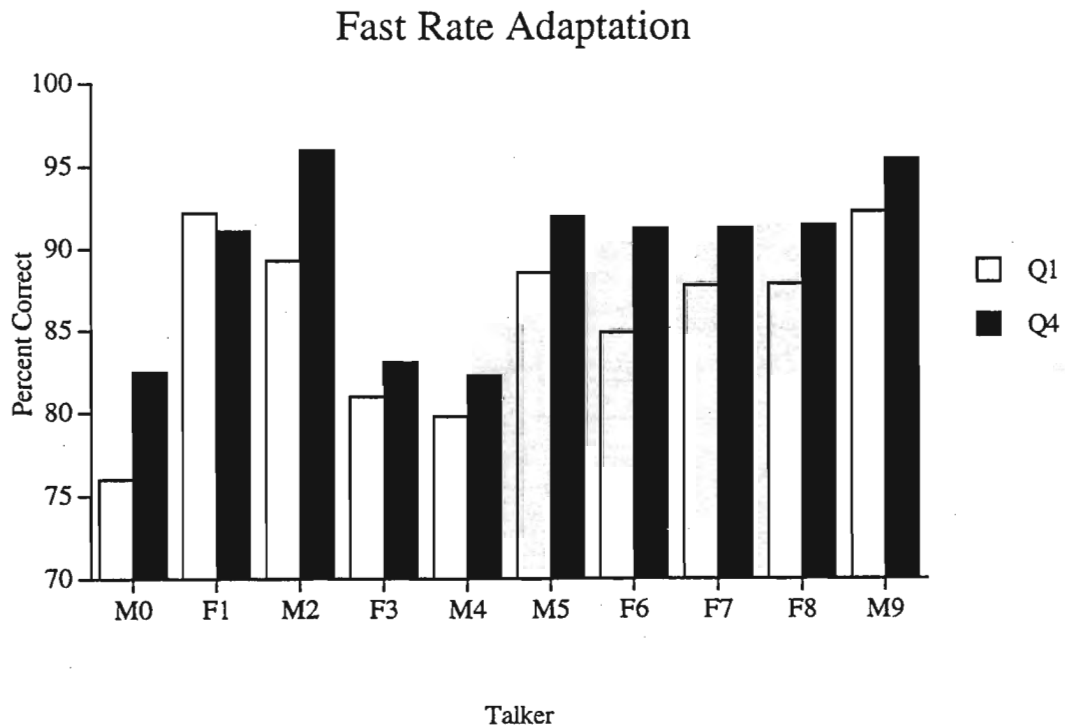
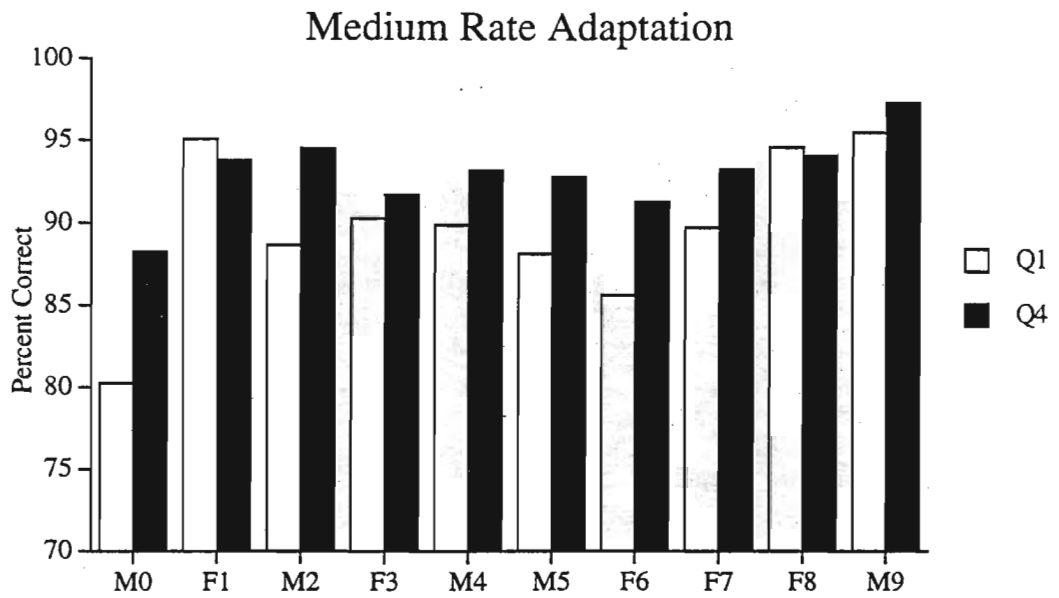
Additionally, difference scores between the first and fourth quartile for “easy” and “hard” words for each listener (fourth quartile minus first quartile percent correct transcription scores) were compared. A paired t-test on these difference scores showed a larger difference for the “hard” words than for the “easy” words. This pattern of results held for both the medium speaking rate ( $t(99) = -2.41$ ,  $p<.0178$ ) and the fast speaking rate ( $t(99) = -2.37$ ,  $p=.0198$ ). Figure 6 displays these difference scores, showing the interaction between lexical characteristics and presentation order for both medium and fast speaking rates.

-----  
 Insert Figure 6 about here  
 -----

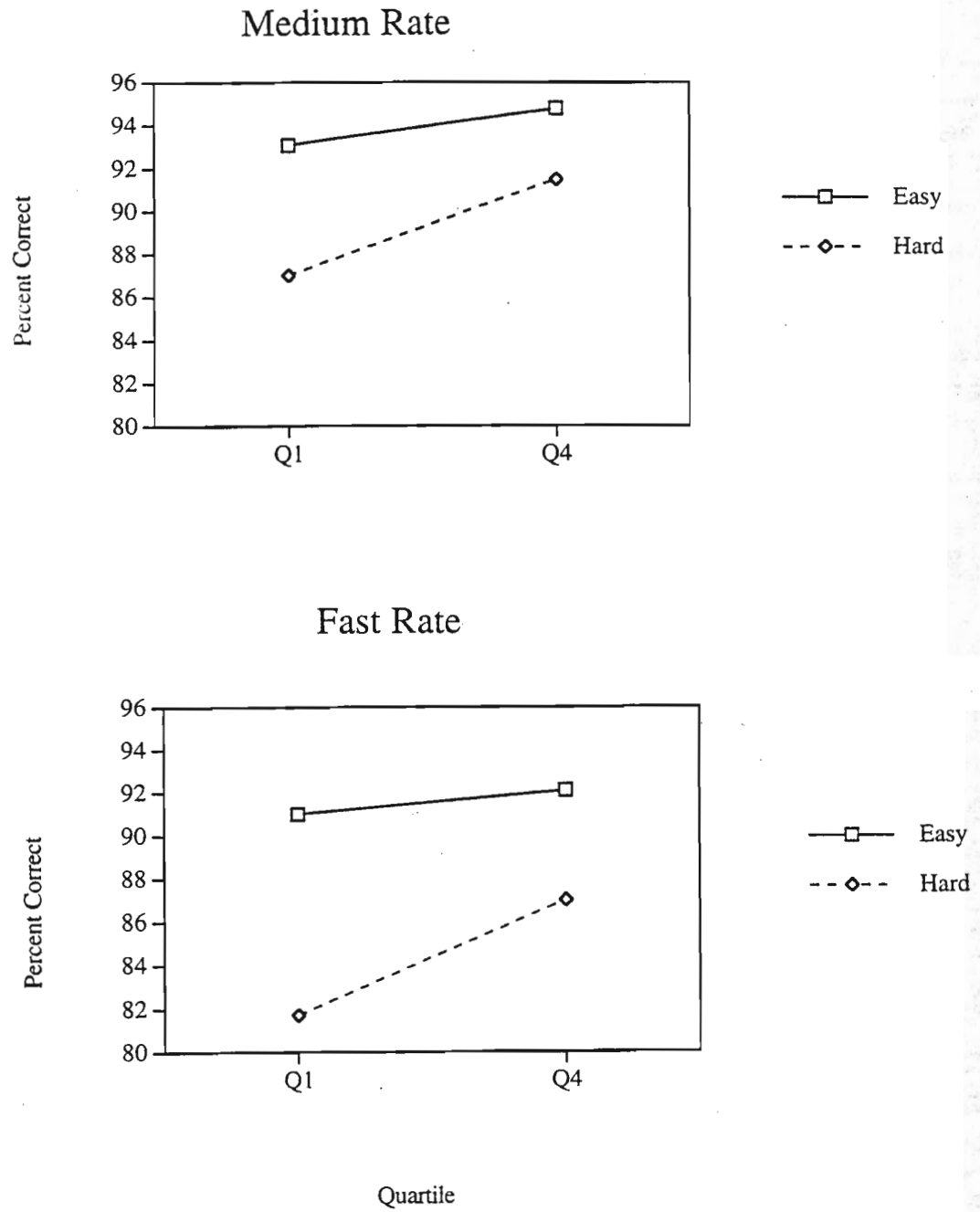
In summary, results of the intelligibility tests revealed three general patterns. First, “easy” words were more accurately recognized than “hard” words. This was expected based on previous research. Second, listeners’ word recognition accuracy improved from the first to the fourth quartile, demonstrating listener adaptation or “tuning” to the specific characteristics of an individual talker. Finally, the adaptation effect to the talker’s voice was larger for lexically “hard” words than for lexically “easy” words, indicating an interaction between the process of lexical discrimination and perceptual learning of a talker’s voice



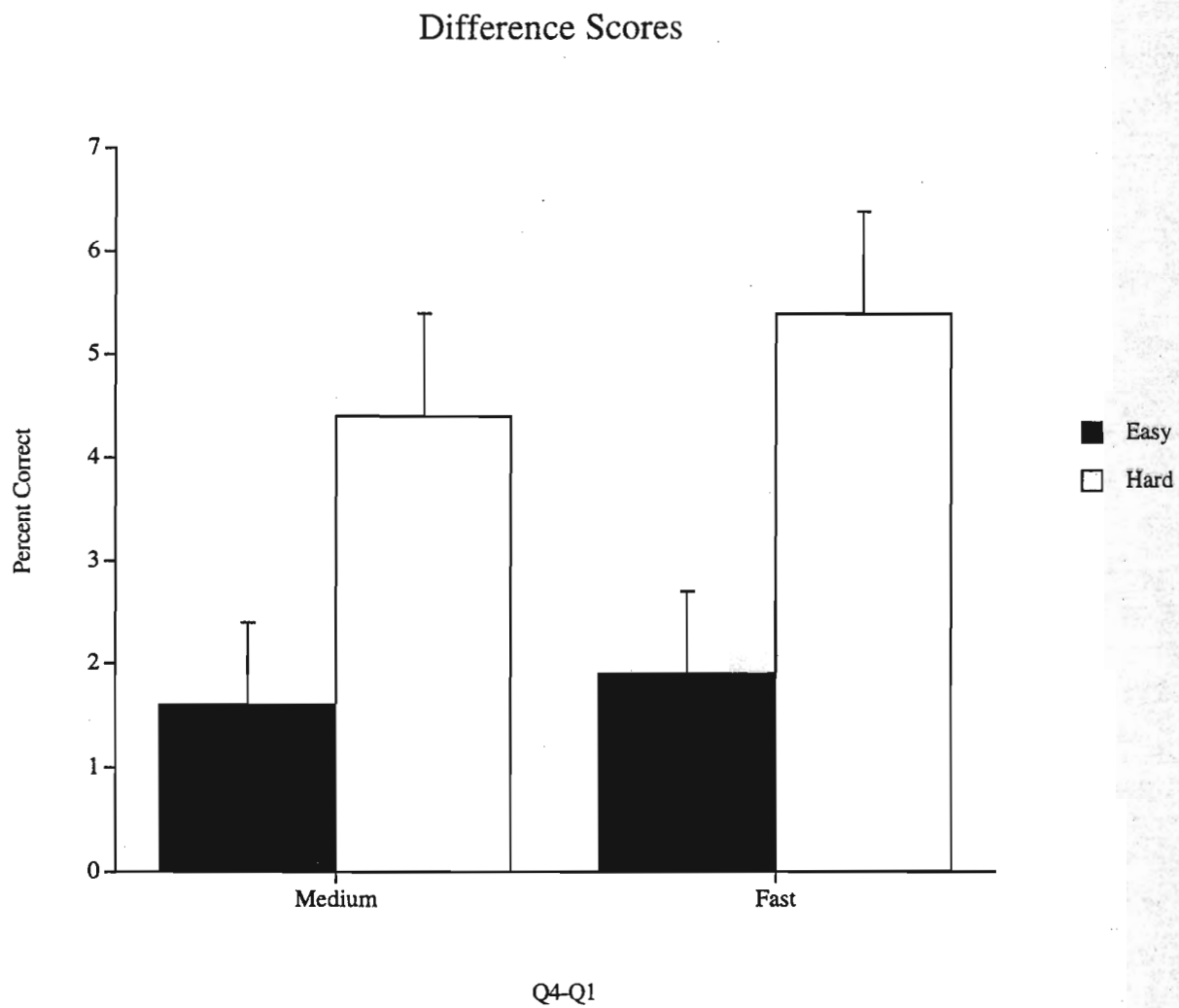
**Figure 3:** The effect of lexical characteristics on intelligibility scores for medium (top) and fast (bottom) rates of speech.



**Figure 4:** The effect of adaptation from first to fourth quartile of the word lists for medium (top) and fast (bottom) rates of speech.



**Figure 5:** The interaction between lexical characteristics and quartile of the word lists for medium (top) and fast (bottom) rates of speech.



**Figure 6:** Difference scores showing the interaction between lexical characteristics and quartile presentation order for medium and fast rates of speech.

during the course of the experiment. Larger learning effects for the "hard" lexical category may be due to either lexical complexity, neighborhood characteristics of these words, or ceiling effects of performance for "easy" words. A follow-up study, in which the stimulus tokens are presented in noise, should help to determine which of these accounts is correct.

### Phase Three: Phonetic Analysis

Phase three of this project involves acoustic-phonetic measurements of the individual talkers included in this database. A vowel-space measure will be obtained for each talker. A subset of words will be selected to represent a minimum of six tokens for each of the vowels /i, a, o/. First and second formant frequencies will be measured from the steady-state portions of the vowels, and individual vowel spaces will be plotted in order to compare vowel space area across talkers, speaking rates and lexical categories. A phonetic assessment of this type can provide a basis for investigating the relationship between speech production, lexical characteristics and overall intelligibility. This final phase will complete the initial project.

### Summary

The overall goal of this project was to provide researchers in our laboratory with a large multiple talker database of isolated spoken words that have specific lexical characteristics. The additional data provided with this digital audio speech database, such as the intelligibility measures and the acoustic-phonetic analyses, are useful data from which researchers may plan further experiments appropriately.

### References

- Hernandez S., L.R. (1995) Implementation of a PC-based perceptual testing system (PTS): A first milestone. *Research on Spoken Language Processing Progress Report No. 19*. Bloomington, IN: Speech Research Laboratory, Indiana University. Pp. 321-328.
- Kirk, K.I., Pisoni, D.B., & Osberger, M.J. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear & 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*. 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. Altmann (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* (pp. 122-147). 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. Pp. 357-376.

Pisoni, D.B., Nusbaum, H.C., Luce, P.A., & Slowiaczek, L.M. (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication*, **4**, 75-95.

Sommers, M.S., Kirk, K.I., & Pisoni, D.B. (1996). Some considerations in evaluating spoken word recognition by normal-hearing and cochlear implant listeners I: The effects of response format. *Research on Spoken Language Processing Progress Report No. 20* (this volume). Bloomington, IN: Speech Research Laboratory, Indiana University. Pp. 31-49.