The Effects of Speaking Rate and Speaking Mode on Intelligibility

by

Jean Christine Krause

B.S.E.E., Georgia Institute of Technology (1993)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1995

© Jean Christine Krause, MCMXCV. All rights reserved.

The author hereby grants to MIT permission to reproduce and distribute publicly paper and electronic copies of this thesis document in whole or in part, and to grant others the right to do so.

Author ....
Department of Electrical Engineering and Computer Science
August 28, 1995

Certified by ....
Louis D. Braida
Henry E. Warren Professor of Electrical Engineering
Thesis Supervisor

Accepted by ....
Frederic R. Morgenthaler
Chairman, Departmental Committee on Graduate Students

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

NOV 0 2 1995

LIBRARIES

The Effects of Speaking Rate and Speaking Mode on Intelligibility

by

Jean Christine Krause

Submitted to the Department of Electrical Engineering and Computer Science
on August 28, 1995, in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

In adverse listening conditions, large and robust increases in intelligibility can be achieved by speaking clearly. The most striking differences between clear and conversational speech are associated with differences in speaking rate. To understand these differences, the intelligibility of speech in a variety of speaking modes was investigated at three different speaking rates. Talkers with significant speaking experience were asked to produce clear and conversational speech at slow, normal, and quick rates. Previous studies show that the speaking rate for clear speech (100 words-per-minute) is roughly one-half that of normal rates for conversational speech. Therefore, during training, the talkers were given feedback on their intelligibility in order to elicit the clearest possible speech at each speaking rate. Talkers also recorded sentences in several other speaking modes such as soft, loud, and conversational with pauses inserted, as required for input to some automatic speech recognition systems.

All speech materials used for intelligibility tests were nonsense sentences which provided no semantic context to aid listeners in identifying key words. The results of the tests for normal hearing listeners in a background of wide-band noise indicated that soft and loud modes, as well as conversational speech with pauses inserted, did not provide as much of an intelligibility advantage as clear speech. The results also showed that the intelligibility advantage of clear speech can be extended to faster speaking rates. After training, talkers successfully produced a form of clear speech at nearly 200 words-per-minute. Moreover, the intelligibility of slow conversational speech was less than the intelligibility of clear speech produced at roughly the same speaking rate. These results suggest that acoustical factors other than reduced speaking rate are responsible for the high intelligibility of clear speech.

Thesis Supervisor: Louis D. Braida
Title: Henry E. Warren Professor of Electrical Engineering

Acknowledgments

Regarding the details of the thesis, I extend thanks to my advisor, Prof. Louis Braida, for his intellectual support and guidance. I also thank Rosalie Uchanski for her suggestions. Many, many thanks to FRED AND BECKY whose tireless work allowed me to finish on time even after the hard disk failure. I am grateful to my talkers, EK, JM, MI, RG, and SA for their hard work, patience, and enthusiasm. I am also extremely grateful to all of my listeners, especially the eight who participated in formal testing. It was a long and tedious experiment, and CA, AC, AG, FK, AM, JP, JS, and MS bore it very good-naturedly. Finally, I am thankful for the NIH grant and the ONR-NDSEG Fellowship which supported this work.

I thank my family for their love, respect, and trust. I especially appreciate my parents' ability to display a genuine interest in my life without ever intruding upon my decisions. They and my brother taught me early in life what to expect of true friends. In fact, I cannot imagine a better childhood companion and friend than my brother, Brian.

My other friends may not be members of my family, but I could feel no closer to them if they were. They are beautiful people, whom I love and admire. I thank them all for letting me into their lives. Yolanda and her wisdom never cease to amaze me; she is the strongest person I know. Nicole's laugh is refreshing (and contagious), and her ability to think deeply is equally well-developed. Nancy's intellectual presence, as well as her sharp sense of humor, make for endless hours of stimulating conversation. In addition, I have been greatly influenced by SAH, whose kindness and awareness of others is a model for all to follow. Another source of influence and strength is "the group", who served as my Georgia Tech family. I initially liked Walter for his quick wit, but I have come to appreciate even more the high priority he places on his friendships. I thank Dave for his caring, his patience, and his undoubtedly high phone bill. Jeff's loyalty and sensitivity are a rare find — I wish him and Sandra all the happiness in the world. I appreciate Kelly for the symphony, his German, his sense-of-humor, and his interest in contributing to society. I have come to value Hengame for her outlook on life, her generosity, and her endless energy. Studying with her, Jeff, and Dave also helped me to learn more than I ever could have alone. Also, thanks to Jon, John, Mona, and Shelley for all of their help and support. In addition, I would like to thank Tony, who reminded me how fun life can be, and James, whose wonderful email stories kept me smiling even in the darkest hours of grad school. Finally, I thank Aaron for his constant support, his neverending patience, and for making my first two years at MIT both fun and fulfilling.

Lastly, I would like to thank the members of the lab for creating a sense of community. It has been fun getting to know all of you. In particular, I must thank Lou, one of the most kind and generous people I have ever known. Also, thanks to Joe for his persistence in organizing social events and for instigating many interesting discussions. I appreciate Dave Lum for his friendly, open nature and his company during off-peak hours in the lab earlier in the year. I am grateful to Jay for countless lunches, dinners, and heartfelt conversations. Thanks to Phil, whose stories are sure to make me smile and whose desk is the finest in the land. Thanks to Matt for whistling while he works and for giving hours of his time to my thesis for no reason at all. Thanks to Stasie, who always had just the right words of encouragement. I also must thank Paul D., Julie, Karen, Lorraine, Ann, Nora, Louise, and Jhy-Shing for being helpful and friendly. Most especially, I thank Maroula. Without her friendship, "Club Theses", and *Grease*, I could never have found the strength to finish.

Contents

1 Introduction **11**

1.1 Introduction 11

1.2 Background 12

1.2.1 Intelligibility Differences 12

1.2.2 Acoustical Differences 13

1.2.3 Speaking Rate Changes and Effects on Intelligibility 14

1.2.4 Slow Speech 16

1.3 Overview 16

2 Selecting Talkers **18**

2.1 Recruiting Talkers 18

2.2 Preliminary Intelligibility Assessment 19

2.2.1 Participants 19

2.2.2 Methods 20

2.2.3 Results 21

2.3 Final Selection 24

3 Acquiring Data **25**

3.1 Eliciting Speech 25

3.1.1 Speaking Rate 26

3.1.2 Speaking Mode 26

3.2 Recording Speech 30

3.2.1 Recording Sessions 30

3.2.2 Recording Setup 32
3.3 Processing Data 32

4 Intelligibility Tests and Results **34**
4.1 Methods for Testing Intelligibility 34
4.1.1 Listeners 34
4.1.2 Presentation Sessions 35
4.1.3 Presentation Setup 35
4.2 Results of Intelligibility Tests 36
4.2.1 Speaking Mode Results 36
4.2.2 Speaking Rate Results 41
4.3 Summary of Results 49

5 Discussion **50**
5.1 Goals of Intelligibility Experiment 50
5.2 Evaluation of Results 52
5.3 Comparison with Previous Data 53
5.4 Summary 56

6 Conclusion **57**
6.1 Talker Selection and Training 57
6.2 Intelligibility Results 58
6.3 Suggestions for Future Work 58

A Sentence Lists **63**

B Listener Audiograms **65**

C Key-word Scores **69**

List of Figures

2-1 Intelligibility increase from conversational to clear speech for each talker. 21

2-2 Intelligibility vs. speaking rate for conversational and clear speech of each talker. 23

2-3 Intelligibility vs. speaking rate (excluding pauses) for conversational and clear speech of each talker. 23

2-4 Intelligibility and rate data (excluding pauses) for selected talkers. . . 24

3-1 Interactive setup for providing talker feedback on intelligibility of clear speech during practice and recording sessions. The talker and the first three listeners were stationed alone in four sound-treated rooms. The fourth listener and the experimenter remained in a quiet part of the lab. 28

4-1 Average key-word scores versus speaking condition. Data represent an average over talker and listener. 38

4-2 Average key-word scores for each talker versus speaking condition. Data represent an average over listener. 38

4-3 Average speaking rates versus speaking condition. Data represent an average over talker and listener. 42

4-4 Average speaking rates for each talker versus speaking condition. Data represent an average over listener. 42

4-5 Average key-word scores versus speaking rate. Data represent an average over talker and listener. 43

4-6 Key-word scores, averaged over talker and listener, versus speaking rate. Conversational speech (SLOW, CONV, and QUICK) is indicated with squares, and clear speech (CLEAR, FAST_CLEAR, and QUICK_CLEAR) is indicated with triangles. 43

4-7 Average key-word scores for T1, averaged over listener, versus speaking rate. Dotted lines represent instances where intelligibility was improved without a reduction in speaking rate. 45

4-8 Average key-word scores for T2, averaged over listener, versus speaking rate. Dotted lines represent instances where intelligibility was improved without a reduction in speaking rate. 45

4-9 Average key-word scores for T3, averaged over listener, versus speaking rate. Dotted lines represent instances where intelligibility was improved without a reduction in speaking rate. 46

4-10 Average key-word scores for T4, averaged over listener, versus speaking rate. Dotted lines represent instances where intelligibility was improved without a reduction in speaking rate. 46

4-11 Average key-word scores for T5, averaged over listener, versus speaking rate. Dotted lines represent instances where intelligibility was improved without a reduction in speaking rate. 47

5-1 Average key-word scores for clear speech at slow, normal and quick speaking rate for all five talkers. Solid lines represent $m1$ and $m2$ from Table 5.1. 54

5-2 Average key-word scores for conversational speech at slow, normal and quick speaking rate for all five talkers. Solid lines represent $m1$ and $m2$ from Table 5.1. 54

List of Tables

2.1 Talkers who participated in the preliminary screening and their speaking experience. 19

3.1 Procedure for recording stimuli. Note that each stimulus list was recorded in exactly two speaking modes. 31

4.1 Talker identification labels for the five talkers used in intelligibility tests. 36

4.2 Percent correct key-word scores (I) and speaking rates in wpm (r) for each of the five talkers. Key-word scores are averaged across all eight normal hearing listeners. 39

4.3 Analysis of variance of the increase in intelligibility between each test mode and CONV mode for four modes (SOFT, LOUD, CLEAR, and CONV+PAUSE). Factors which are significant at a 0.01 level are indicated by asterisks. 40

4.4 Analysis of variance of the increase in intelligibility between clear and conversational speech at three speaking rates. Factors which are significant at a 0.01 level are indicated by asterisks. 48

5.1 Results of $m1$ and $m2$ averaged across talker. For both modes, slopes are calculated by the equations $m1 = (I_{normal} - I_{slow})/(r_{normal} - r_{slow})$ and $m2 = (I_{quick} - I_{normal})/(r_{quick} - r_{normal})$. Subscripts represent nominal speaking rate. Any positive slope data was considered an anomaly and omitted from the calculation. Thus, T1 and T2 were excluded from conversational mode calculations, and T1 and T3 were excluded from clear mode calculations. 53
5.2 Results of $m1$ and $m2$ averaged across talker for each study. The arcsine transform was used to normalize the variance from each study. Slopes are scaled by a factor of 100 for ease of display. 55

A.1 Sentence lists recorded by each talker for formal intelligibility tests. SP and LST corresponds to Picheny's notation for describing the corpus in Appendix B of his thesis[15]. 64

B.1 Audiograms for the four listeners who provided talker T1 with feedback on the intelligibility of her clear speech. Numbers reflect hearing level in dB for the ear used during the experiment. 66
B.2 Audiograms for the four listeners who provided talker T2 with feedback on the intelligibility of her clear speech. Numbers reflect hearing level in dB for the ear used during the experiment. 66
B.3 Audiograms for the four listeners who provided talker T3 with feedback on the intelligibility of her clear speech. Numbers reflect hearing level in dB for the ear used during the experiment. Subject JL replaced subject JG in the second session. 66
B.4 Audiograms for the four listeners who provided talker T4 with feedback on the intelligibility of her clear speech. Numbers reflect hearing level in dB for the ear used during the experiment. 67
B.5 Audiograms for the four listeners who provided talker T5 with feedback on the intelligibility of her clear speech. Numbers reflect hearing level in dB for the ear used during the experiment. 67

B.6 Audiograms for the eight listeners who participated in the intelligibility tests. Numbers reflect hearing level in dB. 68

C.1 Raw and percent correct key-word scores for T1. 70
C.2 Raw and percent correct key-word scores for T1. 74
C.3 Raw and percent correct key-word scores for T3. 78
C.4 Raw and percent correct key-word scores for T4. 82
C.5 Raw and percent correct key-word scores for T5. 86

Chapter 1

Introduction

1.1 Introduction

In difficult communication situations, many talkers adopt a speaking style that permits them to be understood more easily. Recent studies have demonstrated that this altered speaking style, known as clear speech, is significantly more intelligible than conversational speech for both hearing impaired listeners[16] and normal hearing listeners in noise[21] and reverberation[14]. Furthermore, the intelligibility advantage is independent of listener, presentation level, and frequency-gain characteristic[16]. These results suggest that signal processing schemes that convert conversational speech to a sufficiently close approximation of clear speech could improve speech intelligibility in many situations.

In order to implement such signal processing schemes, however, it is first necessary to identify the acoustical factors responsible for the high intelligibility of clear speech. While many acoustical differences between clear speech and conversational speech have been described[17], the specific characteristics of clear speech responsible for its high intelligibility have not yet been isolated. Specifically, the role of speaking rate in highly intelligible, naturally produced clear speech has not been fully determined. The contribution of speaking rate to intelligibility is particularly important for hearing aid applications, since audio and visual signals must remain synchronized for maximum benefit to the listener.

Unfortunately, the relationship between speaking rate and intelligibility may be quite complex. Recent studies indicate that straightforward manipulations of the speech waveform, such as a uniform alteration of speaking rate[18] and a non-uniform alteration of speaking rate[21], cannot account for the intelligibility difference between clear and conversational speech. In addition, Uchanski's[20] preliminary attempt to obtain naturally produced clear speech at a normal speaking rate from a professional talker was not successful. Since speakers vary in their ability to produce highly intelligible speech, however, further work in this area is justified. This thesis describes a method for eliciting both clear and conversational speech at a variety of speaking rates and implements a series of intelligibility tests designed to evaluate the effects of speaking rate on intelligibility.

1.2 Background

Research on naturally produced clear speech dates back several decades. In recent years, however, the focus of this research has shifted from investigating intertalker differences to investigating intratalker differences between clear and conversational speech. Reports by Picheny *et al.*[16, 17, 18], Uchanski[21], Chen[2], and Payton[14] establish that, independent of the talker, clear speech is significantly more intelligible than conversational speech both for hearing impaired listeners in quiet and for normal hearing listeners in noise and reverberation. In addition, these reports describe both the acoustical differences and the speaking rate differences between the two modes of speaking. The major results of these and other related studies are summarized below.

1.2.1 Intelligibility Differences

In a series of studies, Picheny *et al.* investigated the differences between clear and conversational speech. The first study[15, 16] tested five hearing impaired listeners on sets of 50 nonsense sentences spoken by three male talkers in both conversational and clear speaking modes. The sentences were presented at three different presentation levels using two distinct frequency-gain characteristics. Intelligibility scores for key-words were found to be 17 points higher for clear speech than for conversational speech on average. Moreover, this intelligibility difference was independent of talker, listener, presentation level, and frequency-gain characteristic, at least to a first approximation. This intelligibility advantage of clear speech over conversational speech was verified by Uchanski[20, 21] and extended to include normal hearing listeners in noise.

In a related study, Chen[2] investigated the intelligibility of conversationally and clearly spoken consonant-vowel (CV) syllables. The CV's were formed from one of the six stop consonants (/p/, /t/, /k/, /b/, /d/, /g/) followed by one of the three point vowels (/i/, /a/, /u/). Each CV was spoken both clearly and conversationally by three male talkers and presented to three normal hearing listeners in noise. On average, the CV identification score for clear speech was 22 percentage points higher than for conversational speech.

More recently, Payton *et al.*[14] examined the effects of noise and reverberation on intelligibility. In this study, nonsense sentences spoken clearly and conversationally were presented in various environments to ten normal hearing and two hearing impaired listeners. The environments were combinations of three levels of reverberation and four levels of noise, although not every environment was presented to every listener. On average, the scores for clear speech were 20 points higher than conversational speech for normal hearing listeners and 26 points higher for hearing impaired listeners. In addition, this advantage depended only on the intelligibility score for conversational speech; it was independent of listener and environment to a first approximation.

1.2.2 Acoustical Differences

After establishing the high intelligibility of clear speech, Picheny *et al.*[15, 17] went on to study the acoustical differences between conversational and clear speaking modes. They performed an acoustical analysis of 50 nonsense sentences spoken clearly and conversationally by three male talkers. Substantial acoustical differences between each talker's clear and conversational speech were observed for articulation rate, number of pauses, and number of phonological modifications. Short-term spectra of consonant and vowels as well as relative intensities of plosives and fricatives were also found to differ between clear and conversational speech. Although this study identified many acoustical differences between clear and conversational speech, it did not attempt to determine which differences were responsible for the high intelligibility of clear speech.

An acoustical analysis of clear and conversational speech was also performed by Chen[2] in the CV-syllable intelligibility study. Acoustic measurements of the CV's used in the study demonstrated that clearly spoken syllables exhibited a significantly longer voice onset time for voiceless consonants. Also, the formant frequencies of vowels were found to cluster more tightly in clear speech, suggesting that the formants more closely approximated their target values. Other measurements revealed that clear speech exhibited a larger vowel triangle, a larger consonant-to-vowel ratio, and longer formant-transition durations.

1.2.3 Speaking Rate Changes and Effects on Intelligibility

One of the most striking differences between clear and conversational speech lies in speaking rate; the typical speaking rate for clear speech (100 words per minute) is roughly half that of conversational speech[15, 17]. As a result, several studies have attempted to determine whether a reduced speaking rate is essential to highly intelligible speech. For example, Picheny *et al.*[15, 18] conducted a probe experiment to investigate the effect of overall speaking rate on intelligibility. Using Malah's algorithm[12], one male talker's clear sentences were uniformly time-compressed to conform to a normal conversational speaking rate of 200 wpm, and his conversational sentences were uniformly expanded to typical clear speaking rates of 100 wpm. After this time-scaling of the waveforms, the processed sentences were presented to five hearing-impaired listeners. In both cases, the processed speech was less intelligible than the unprocessed speech. In a later study[20, 21], Uchanski *et al.* used a non-uniform time-scaling method, the Griffin-Lim[9] algorithm, to process the sentences in order to determine the contribution of segmental-level durational differences between clear and conversational speech. Both hearing impaired listeners in quiet and normal hearing listeners in noise found the processed sentences to be less intelligible than the unprocessed sentences. Although neither time-scaling procedure produced fast, clear speech that was more intelligible than unprocessed conversational speech, non-uniform time-scaling was generally less harmful to intelligibility than uniform time-scaling. In both cases, intelligibility tests were also performed on sped clear speech which was slowed to clear speaking rates again and on slowed conversational speech which was sped to conversational rates. Percent-correct key word scores for these twice-processed materials were similar to those scores obtained for the unprocessed materials, indicating that most of the decrease in intelligibility was not due to signal processing artifacts.

In addition to studies of time-scaled speech, a substantial amount of work has focused of the role of pauses in clear speech. More frequent and longer pauses, in conjunction with lengthened speech sounds, are responsible for the reduced speaking rate of clear speech[17]. Investigating the effects of pauses on intelligibility, Choi[3] measured the intelligibility of pause-processed sentences. Her results indicate that adding pauses to conversational speech does not improve its intelligibility and deleting pauses from clear speech does not decrease its intelligibility. This data is supported by a similar study by Uchanski[20], which found that key words excised from clearly spoken sentences had nearly the same intelligibility as the same words in sentence context.

While most clear speech research has focused on signal processing techniques to achieve clear speech at normal speaking rates, a preliminary experiment by Uchanski[21, 20] sought to elicit fast clear speech naturally. In this experiment, a professional talker attempted to produce clear speech at a variety of rates. Two hearing impaired listeners in quiet and two normal hearing listeners in noise were tested. The results of the intelligibility tests suggested that the talker could not improve his intelligibility without slowing down. Other talkers, however, may have different strategies for speaking clearly. Therefore, more work in this area must be completed before any conclusions regarding naturally produced clear speech at normal speaking rates are justified.

1.2.4 Slow Speech

To date, clear speech at normal speaking rates has not been achieved, whether through artificial or natural means. Another body of research has focused on speaking rate alone, without considering intelligibility. Various acoustical differences between normal and slow speech have been established. Crystal and House[5, 6], for instance, examined acoustical differences between the fastest and the slowest talkers in a group reading from the same script. Han[10] also observed that an increase in speaking rate was achieved mostly by deleting pauses, rather than shortening speech sounds. These studies, however, are not useful for understanding clear speech without corresponding measurements of intelligibility. It is imperative to determine whether slow speech, without emphasis on clarity, has comparable intelligibility to clear speech. Some linguists hypothesize that clarity is independent of speaking rate[22]. If slow speech can indeed be less intelligible than clear speech, then the acoustical differences between the two speaking styles could help identify which acoustic factors are responsible for the high intelligibility of clear speech.

1.3 Overview

Previous studies have not identified the specific characteristics responsible for highly intelligible speech. In particular, the effect of speaking rate on the intelligibility of clear speech remains unresolved. While a significant body of research has investigated signal-processing schemes for manipulating rate, little research has been dedicated to eliciting clear speech naturally. Uchanski's[20] preliminary attempt to elicit clear speech at normal rates from a professional talker was unsuccessful. This talker, however, stated that he had emphasized speed rather than clarity in his professional training. Moreover, talkers vary in their strategies for manipulating clarity and rate, so further work in this area is justified. This study defines a more structured way than previous studies for eliciting clear speech at normal speaking rates. In particular, significant effort was devoted to both selecting and training talkers. The talkers in this intelligibility experiment were selected from a large pool of talkers because they demonstrated unusual characteristics in intelligibility and/or speaking rate. After selection, each talker practiced his/her clear speech with feedback on intelligibility from listeners. Talkers were encouraged to experiment with different speaking strategies during the practice sessions. These procedures are described in more detail in Chapters 2 and 3. The testing procedures and intelligibility results are presented in Chapter 4. Finally, a discussion of results and suggestions for future work are included in Chapters 5 and 6.

Chapter 2

Selecting Talkers

In order to improve the chances of obtaining clear speech at faster speaking rates naturally, much attention was given to selecting talkers for the experiment. Only talkers with significant speaking experience were considered. All potential talkers participated in a preliminary screening, and the five participants with the highest potential for producing fast clear speech were selected.

2.1 Recruiting Talkers

Talkers were recruited throughout the New England area. Advertisements were posted at local colleges with programs in television and radio broadcasting, public speaking, and other communications disciplines. In addition, a description of the experiment was provided to local radio stations as well as the New England Speakers Bureau (NESB). Everyone who responded was interviewed in order to verify the extent of his/her speaking experience. Those talkers with at least two years of speaking experience were asked to participate in a preliminary study to evaluate their intelligibility.

Table 2.1: Talkers who participated in the preliminary screening and their speaking experience.

Subject ID	Talker	Gender	Speaking Experience	Years
S1	DF	M	High School and College Debate Team	5
S2	RG	F	High School and College Debate Team	6
S3	RT	F	High School and College Debate Team	4
S4	SS	F	High School and College Debate Team	4
S5	SA	M	High School and College Debate Team	7
S6	EK	F	College Television and Radio, Public Speaking	5
S7	DC	F	Professional Speaker, NESB	2
S8	GS1	F	Professional Speaker, NESB	2
S9	JM	F	Professional Speaker, NESB	5
S10	TG	M	Radio Broadcasting Student	2
S11	MI	F	Radio Broadcasting Student	2
S12	GS2	M	Professional Radio Broadcaster	2
S13	EP	M	Radio Broadcasting Student	3
S14	TW	M	Radio Broadcasting Student	3
S15	DN	F	Radio Broadcasting Student	2

2.2 Preliminary Intelligibility Assessment

After the initial interviews, fifteen talkers were selected for the preliminary intelligibility screening. Talkers were recorded in both conversational and clear speaking modes. To assess the intelligibility of each talker, the recordings were presented in the presence of wide-band noise to normal hearing listeners. The speaking rates of the talkers were also examined. Five talkers who exhibited potential for producing fast clear speech were asked to participate in the experiment.

2.2.1 Participants

The talkers selected to participate in the preliminary intelligibility test all possessed a minimum of two years of speaking experience. The group was comprised primarily of local students and professionals whose work required attention to speaking skills. A description of the talkers and their speaking experiences is summarized in Table 2.1.

2.2.2 Methods

The preliminary screening was designed to evaluate the intelligibility of many talkers in a reasonably short time period. Therefore, the methods used were quick and straightforward. As a result, the results provide only a crude indication of each talker's ability to produce clear speech.

Eliciting and Recording Speech

In order to obtain the clearest possible speech from each talker with minimal training, the talkers were familiarized with the characteristics of clear speech. The talkers listened to samples of both conversational and clear speech, and differences between the two speaking modes were discussed. The talkers were asked to mimic the clear speaking styles which had been presented, and they were given feedback on both rate and clarity. The goal of obtaining clear speech at normal speaking rates was explained, but each talker was instructed not to increase speaking rate at the expense of clarity. The talker was then given one hour to practice his/her clear speech. A list of 50 sentences was then recorded clearly, and another list of 50 sentences was recorded conversationally. Each talker recorded a unique set of 100 sentences. The speech materials used for recording were obtained from the corpus of nonsense sentences described by Picheny *et al.*[16]. The speech was recorded at a 48kHz sampling rate, using a SONY 59ES Digital Audio Tape Deck.

Evaluating Intelligibility

The recorded materials were copied to disk using the CardDPlus digital recording utility. The files were then decimated to 20kHz, a sampling rate compatible with the hardware available for playing waveforms. The overall rms power in each sentence was calculated, and each set of 50 sentences was normalized so that each sentence had an rms equal to the average of the set. A stereo waveform was then created, with the normalized speech on the left channel, and wide-band noise of equal rms on the right channel. To evaluate the intelligibility of each talker, the speech and noise were presented monoaurally to normal hearing listeners at a signal-to-noise ratio of –4 decibels (dB). Intelligibility scores for key words were determined using the grading rules described by Picheny *et al.*[16]. The nonsense sentences used as stimuli in the intelligibility tests provide no semantic context which could aid the listener in identifying key words.

2.2.3 Results

When speaking clearly, each talker achieved some improvement in intelligibility over his/her conversational speech. This increase in intelligibility can be seen in Figure 2-1.

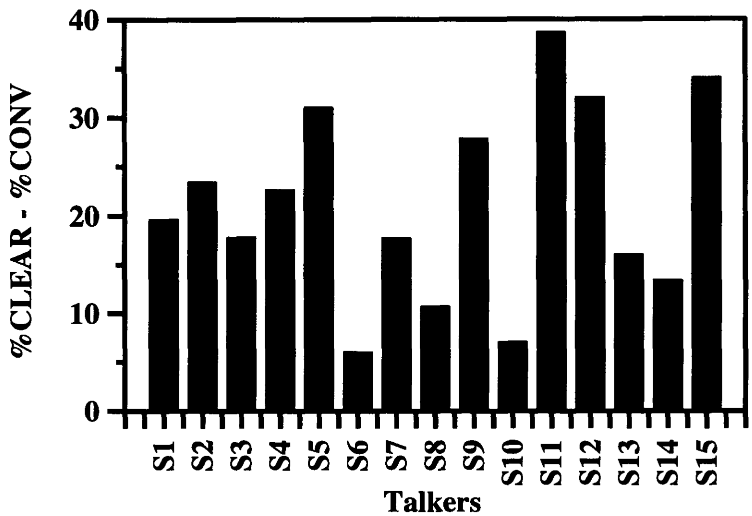

Figure 2-1: Intelligibility increase from conversational to clear speech for each talker.

In addition to measuring intelligibility, each talker's speaking rate was estimated by calculating the average duration of the recorded sentences. This average was converted to a words-per-minute figure. Figure 2-2 demonstrates that all talkers slowed down in order to increase their intelligibility. Since the goal was to identify talkers who could not only produce clear speech but also demonstrate a high potential for producing clear speech at normal speaking rates, it was important to identify whether the talker's slower rate was due to increased pauses or slower articulation. Therefore, pause durations for each talker were estimated by a simple threshold test.

Since both Choi[3] and Uchanski[20] concluded that the addition or deletion of pauses does not affect intelligibility, the speaking rate for each talker, excluding pauses, was then calculated. The resulting relationship between intelligibility and speaking rate for each talker is plotted in Figure 2-3. It is interesting to note that the lines for each talker are more clustered than in Figure 2-2. Moreover, with the exception of Talker EK, the lines for each talker have similar slopes, suggesting that a direct relationship between articulation rate and intelligibility may exist. This relationship can be approximated by the equation $I_j = a_j - mR$ where I represents intelligibility; R represents speaking rate excluding pauses; and a and m are positive constants. The scores for each talker depicted in Figure 2-3 were the primary criterion for selecting talkers to participate in further intelligibility experiments.

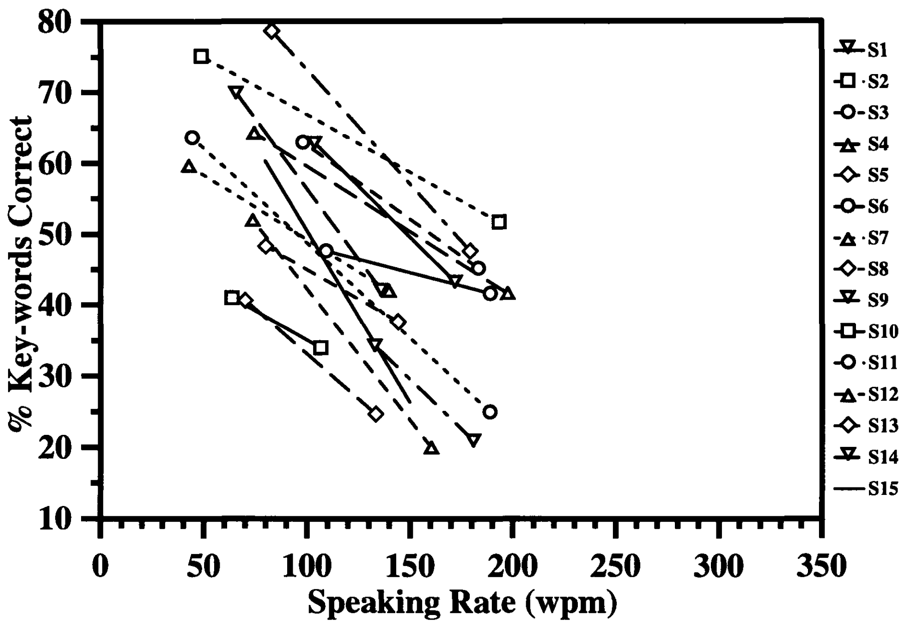

Figure 2-2: Intelligibility vs. speaking rate for conversational and clear speech of each talker.

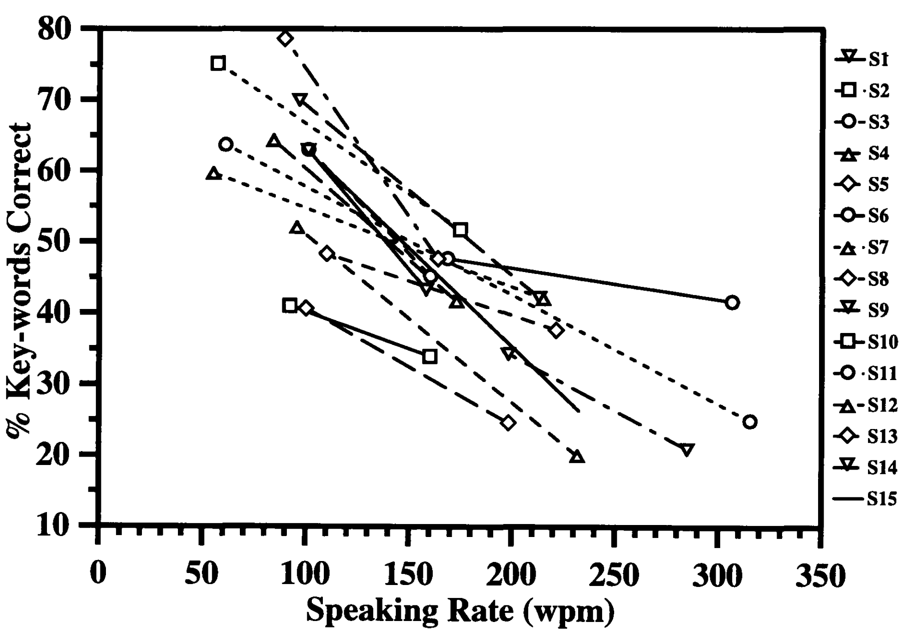

Figure 2-3: Intelligibility vs. speaking rate (excluding pauses) for conversational and clear speech of each talker.

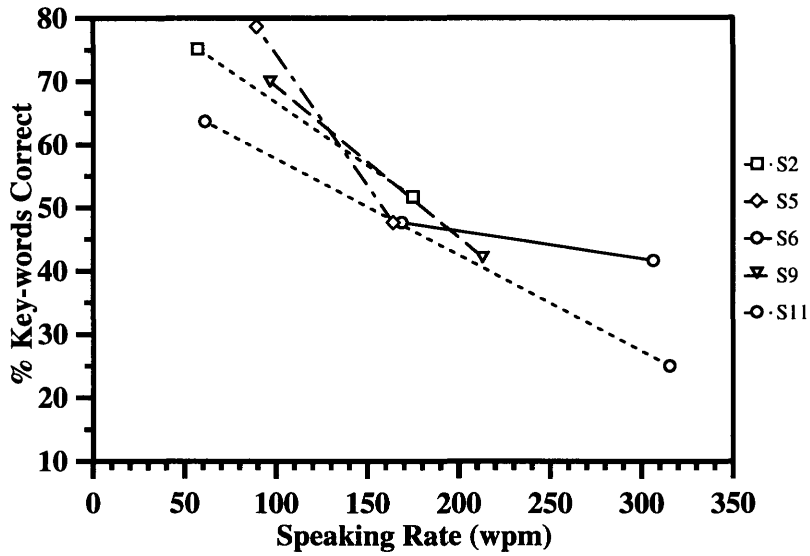

Figure 2-4: Intelligibility and rate data (excluding pauses) for selected talkers.

2.3 Final Selection

Five talkers with different strengths were selected in order to improve the chances of finding at least one talker who could produce clear speech at normal speaking rates. The data for the five selected talkers is shown in Figure 2-4. Talker S5 was selected because his clear speech had the overall highest intelligibility at 79 percent. Talker S11 was selected because she had the greatest increase in intelligibility between conversational and clear speech. She also demonstrated the ability to change her speaking rate significantly, from 61 wpm in clear mode to 315 wpm in conversational mode. S6 was selected for her ability to speak at a higher rate than most other talkers, both in conversational (307 wpm) and clear (169 wpm) modes. In addition, she exhibited an unusually low absolute value of the slope parameter m. It was hoped this absolute value of m could be increased with training. Talkers S2 and S9 were selected because their overall intelligibility in both modes was higher than most of the other talkers at similar speaking rates. These five talkers began training with feedback on rate and intelligibility. The training sessions are described in Section 3.1.

Chapter 3

Acquiring Data

In the preliminary intelligibility tests, the selected talkers exhibited some ability to control clarity and/or speaking rate. Before recording the stimuli, however, it was important for each talker to obtain feedback on his/her own speech. Therefore, talkers received training and recorded sentences using an interactive paradigm similar to the method previously described by Chen[2]. All sentences used both for practice and for formal intelligibility tests were from the corpus of nonsense sentences described by Picheny *et al.*[16]. The specific lists used for testing are described in Appendix A. These sentences provide no semantic context to aid the listeners in identifying key words.

3.1 Eliciting Speech

Clear and conversational speech were elicited from the talkers at three different relative speaking rates: slow, normal, and quick. The interactive recording setup described in Section 3.1.2 provided feedback to the talkers on both speaking rate and intelligibility. Talkers were encouraged to experiment with several different speaking strategies, using the feedback to determine which strategies were most successful. Speech in several other speaking modes was also elicited from talkers.

3.1.1 Speaking Rate

Each talker chose his/her own slow, normal, and quick speaking rates. To determine the normal rate, each talker was instructed to read 200 sentences at a rate (s)he considered appropriate for normal conversation. For the quick rate, the talkers were instructed to read 50 sentences as rapidly as possible. After each talker had practiced speaking clearly, (s)he was instructed to produce 100 clear sentences, with interactive feedback from listeners on intelligibility. These sentences were used to designate the talker's slow speaking rate. Clear speech was used to establish the slow rate for two reasons. First, this method imposed no rate constraints on the production of clear speech. Secondly, this method required the talker to produce slow conversational speech at a clear speaking rate in a later recording session, which allowed for a direct comparison of the intelligibility of clear and conversational modes at the same rate.

Throughout the experiment, the speaking rate of talkers was specified by a metronome. In each case, the speaking rate was calculated by dividing the total number of words read by the duration of the sentences. The average number of words per sentence was also calculated, and the metronome was set to click once at the beginning and at the end of each sentence. Setting the metronome to the appropriate sentence rate rather than word or syllable rate allowed the talker freedom to determine the duration of individual speech segments. Both in training and recording sessions, the appropriate speaking rate was communicated to the talker by presenting the output of the metronome over headphones.

3.1.2 Speaking Mode

At each of the three speaking rates, several speaking modes were elicited from the talkers. Four speaking modes were obtained at the normal rate: SOFT, LOUD, conversational (CONV), and clear (denoted by FAST_CLEAR since it was recorded at a rate faster than typical clear rates). Two speaking modes were obtained at the quick rate: conversational (QUICK) and clear (QUICK_CLEAR); three modes were elicited at the slow rate: conversational (SLOW), CLEAR, and conversational with pauses between the words (CONV+PAUSE), as if speaking to an automatic speech recognition system. For all speaking modes except conversational, the talker was given objective feedback on his/her speech. Talkers were given the opportunity to practice with this feedback until they were comfortable speaking in a particular mode. The methods for eliciting each mode are described below. Conversational speech was elicited simply by instructing the talkers to speak sentences as they would in normal conversation.

CLEAR, FAST_CLEAR, AND QUICK_CLEAR Modes

Clear speech was elicited from the talker using an interactive paradigm derived from the method described by Chen[2] for eliciting clear speech with syllables. In Chen's experiment, a talker repeated a syllable until the listener perceived it correctly in the presence of masking noise. While this method could be used with nonsense sentences, its disadvantage is that repetition of sentences increases their intelligibility[20]. To avoid the intelligibility benefit of repetition, four normal hearing listeners were employed to provide the talker with feedback on the intelligibility of his/her speech. The talker's speech was distorted by multiplicative noise[19] monaurally over headphones to each of the listeners in turn. Multiplicative noise was used to make the intelligibility tests difficult, because it maintains a constant signal-to-noise ratio (SNR). Thus, the talkers could not increase intelligibility simply by speaking more loudly. A SNR was determined experimentally at the beginning of the session. Initially, the SNR = 0 dB, and it was decreased in increments of 0.2 dB until the subjects received on average no more than one key word correctly from the talker's first utterance of the sentence.

After the SNR was selected, the talker and the listeners were separated (see Figure 3-1). Each listener could hear the talker only when (s)he was addressed. The talker could hear the four listeners at all times. The experimenter and the talker also had the freedom to communicate at any time throughout the session.

The procedure for eliciting clear speech required the talker to repeat a sentence with increased emphasis on articulation until it was perceived correctly by a listener. This procedure was invoked for both practice and recording sessions. In every session,

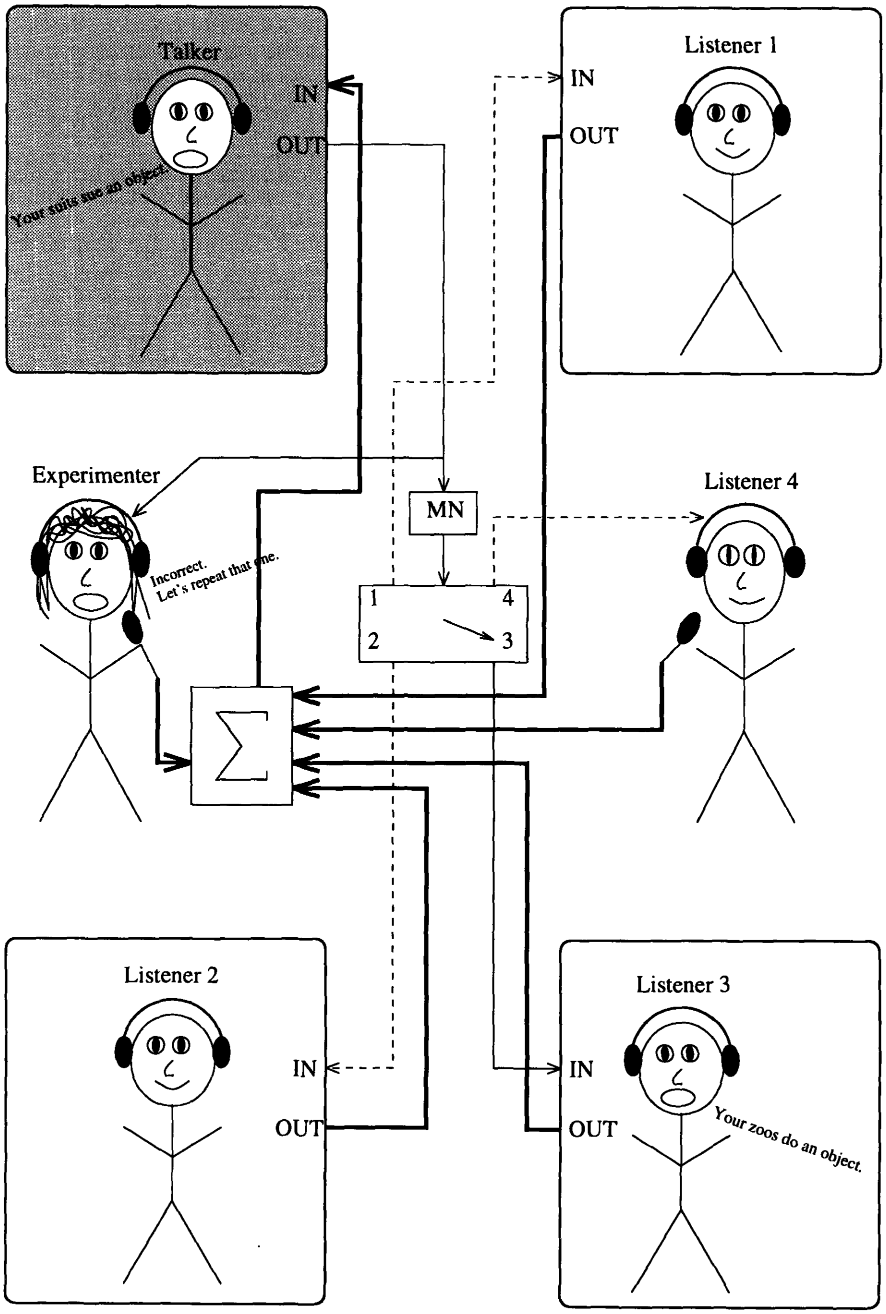

Figure 3-1: Interactive setup for providing talker feedback on intelligibility of clear speech during practice and recording sessions. The talker and the first three listeners were stationed alone in four sound-treated rooms. The fourth listener and the experimenter rem ined in a quiet part of the lab.

an order was established for the talker to address each of listeners in turn. The talker read each sentence, and the designated listener responded verbally with the sentence heard. The listener's response was regarded as correct if more than half of the key words were correctly identified. If the response was incorrect, the talker repeated the sentence to the next listener; if the response was correct, the talker presented a new sentence to the next listener. The four listeners were not given feedback on whether or not the response was correct. A sentence was not repeated additional times after it had been presented to all four listeners. If three sentences in a row were presented to all four listeners without a correct response, the SNR was increased by 0.2 dB.

During practice sessions, the talkers were encouraged to experiment with different speaking strategies and allowed to practice as much as they felt necessary. In addition to feedback from the listeners, the experimenter also provided instruction, reminding the talker to adhere to the timing cues provided by the metronome (for FAST_CLEAR and QUICK_CLEAR modes) and pointing out patterns of mistakes among the listeners. In addition, the experimenter served as a judge of the listener's responses and decided whether or not the talker should repeat a sentence.

Additional Speaking Modes

In addition to clear and conversational speech, the talkers also produced three other speaking modes: SOFT, LOUD, and CONV+PAUSE. SOFT and LOUD speech were elicited from the talker with the use of a sound-level meter. A Realistic Digital Sound Level Meter, located approximately 2 1/2 feet from the talker's mouth, was set to measure the maximum A-weighted sound pressure level in a sentence. The talker was instructed to read ten sentences in a conversational manner, and the sound-level was measured for each sentence. The largest and smallest of these levels were noted. For SOFT speech, the talker was instructed to speak at sound-levels at least 15 dB below the largest level measured during his/her conversational speech. The level on the meter was reported to the talker after each sentence was read, and (s)he repeated the sentence if necessary. LOUD speech was elicited in a similar manner, except the talker was instructed to exceed the smallest level measured during conversational speech by at least 15 dB. Finally, the the CONV+PAUSE mode was elicited by setting the metronome to click approximately once for each word spoken. The talker was instructed to produce each word in a conversational manner, beginning each word as the metronome clicked. The metronome rate was set so that the speaking rate in wpm would be approximately equal to the talker's slow speaking rate.

3.2 Recording Speech

Each talker recorded 700 nonsense sentences in a sound-treated room over four recording sessions. These sentences consisted of seven sentence lists, each containing 50 sentences. Each of the talkers recorded seven unique lists. Every sentence list was recorded in two speaking modes; one of the two speaking modes was always conversational. For a given stimulus list, the two speaking modes were generally recorded at the same speaking rate in order to facilitate intelligibility comparisons without speaking rate as a factor. One list, however, was recorded at two different rates. It was recorded once in CONV mode at the normal rate and once in CLEAR mode at the slow rate. A summary of the modes and rates corresponding to the seven stimulus lists is shown in Table 3-1. The list recorded at two different speaking rates was *normal_rate4*.

3.2.1 Recording Sessions

Each talker participated in four recording sessions. The first and fourth sessions were two hours in length, and the second and third sessions were three hours. In the first session, talkers spent approximately one hour practicing conversational speech and becoming familiar with the sentences and the recording equipment. In the next half-hour, lists *normal_rate1* through *normal_rate4* were recorded in CONV mode. In the final half-hour, list *quick_rate1* was recorded in five sets of ten sentences. Recording QUICK mode in sets of ten sentences helped talkers avoid misreading sentences. After the first session, each talker's normal and quick speaking rates were calculated.

For the second and third recording sessions, the interactive setup shown in Fig-

Table 3.1: Procedure for recording stimuli. Note that each stimulus list was recorded in exactly two speaking modes.

Stimulus List	Speaking Mode	Speaking Rate
normal_rate1	CONV	normal
	SOFT	normal
normal_rate2	CONV	normal
	LOUD	normal
normal_rate3	CONV	normal
	FAST_CLEAR	normal
normal_rate4	CONV	normal
	CLEAR	slow
quick_rate1	QUICK	quick
	QUICK_CLEAR	quick
slow_rate1	SLOW	slow
	CLEAR	slow
slow_rate2	SLOW	slow
	CONV+PAUSE	slow

ure 3-1 was used to elicit clear speech from the talker. In the second session, the talker spent up to two hours experimenting with different speaking strategies, using listener feedback to settle on a strategy. After the talker felt comfortable producing clear speech, lists *normal_rate4* and *slow_rate1* were recorded in CLEAR mode. From these lists, the talker's slow speaking rate was determined. In the third session, the metronome signal was presented over headphones, as the talker attempted to produce FAST_CLEAR and QUICK_CLEAR speech, conforming to the speaking rates from Session 1. In all cases, the experimentally determined SNR was higher for the third session than for the second session. The talker spent up to two hours practicing clear speech at the two rates, and then lists *normal_rate3* and *quick_rate1* were recorded. In the final recording session, each talker recorded one sentence list in SOFT, LOUD, CONV+PAUSE modes, and two lists in SLOW mode. Before recording each list, the talker practiced the appropriate speaking mode, with feedback, for 20-25 minutes.

The recording sessions were broken into smaller subsessions to reduce fatigue among the talker and listeners. The first and fourth sessions were divided into two

55 minute subsessions, with several short breaks within each subsession. Each talker was also given water to drink and encouraged to rest when fatigued. The second and third sessions were less structured; listeners were given breaks approximately once every half-hour. Again, the talker was instructed to rest briefly anytime (s)he felt necessary.

3.2.2 Recording Setup

All recording sessions took place with the talker seated in a sound-treated room. The sentences were placed in front of him/her on a stand to prevent paper rattling. A Sennheiser MD 421 cardioid microphone was positioned approximately 6 inches in front of the talker's mouth. The rolloff filter on the microphone was adjusted to the speech setting. The microphone output was amplified using a Symetrix SX202 Dual Microphone Preamplifier. The amplifier output was then recorded digitally on a personal computer (PC) disk, using a DAL card with a 20kHz sampling rate. The recording function on the PC was controlled by commands from a DIGITAL VAX workstation. For backup purposes, all sessions were also recorded to digital audio tape at a 48kHz sampling rate, using a SONY 59ES Digital Audio Tape Recorder.

In the first and fourth recording sessions, the experimenter sat in the booth with the talker, providing instruction when necessary. While in the booth, the experimenter also entered start and stop commands for recording into a terminal of the VAX. These commands signalled the talker when to begin each sentence. In the second and third sessions, however, the experimenter was outside the booth entering commands into a terminal of the VAX. In this case, the talker received instructions to start each sentence over headphones. In addition, the metronome output was presented to the talker over headphones throughout the final two recording sessions.

3.3 Processing Data

All sentences recorded in the four sessions were used as stimuli for the intelligibility tests. Before processing, sentences were checked for errors such as mispronounced words or clipped signal waveforms. When possible, these sentences were recorded again to eliminate such errors. In a few cases, mispronounced words were noted so that responses could be graded accordingly during intelligibility tests. In addition, extraneous keyboard sounds were edited from sentences recorded during the first and fourth sessions.

After recording the stimuli, the first stage of processing was normalization of amplitude. In previous studies[20, 14], the quantity used for normalization was long-term average root-mean-square (rms) power of the sentence. This method of normalization, however, is only appropriate for sentences which have comparable pause durations. In this experiment, pause durations varied greatly since talkers spoke at three different speaking rates. Therefore, the rms amplitude of each sentence without pauses was used for normalization purposes. Each sentence was normalized to have a rms of 2185 digital units (maximum amplitude = 32767). A stereo waveform was then created with normalized speech on the left channel and wide-band noise of the same rms level on the right channel.

Chapter 4

Intelligibility Tests and Results

After processing the recorded materials, experiments were performed to assess the intelligibility of the various speaking modes. The goal was to determine whether normal hearing listeners could derive intelligibility benefits from more than one speaking mode at a given speaking rate. In particular, it was hoped that highly intelligible speech could be found at normal and quick as well as slow speaking rates.

4.1 Methods for Testing Intelligibility

All sentences recorded by the five talkers were used in the intelligibility experiments. Normal hearing listeners were employed to evaluate the intelligibility of the speech in the presence of additive wide-band noise. The materials were presented to the listeners over the course of 16 two-hour sessions. Intelligibility scores were based on the percentage of key words correct, using the scoring rules developed by Picheny *et al.*[16].

4.1.1 Listeners

Eight normal hearing listeners (four males, four females) were obtained from the MIT community. The listeners were all native speakers of English who possessed at least a high school education. They ranged in age from 18 to 29 years. The results of each listener's hearing test is listed in Appendix B. Listeners were tested monoaurally over headphones in a sound-treated room. Each listener selected which ear would receive the stimuli and was encouraged to switch the stimulus to the other ear when fatigued.

4.1.2 Presentation Sessions

Listeners were tested in 16 two-hour sessions over the course of approximately eight weeks. The amount of time between the first and second presentations of a sentence list was at least two weeks in order to minimize learning effects. Listeners responded by writing their answers on paper. They were given as much time as needed to respond but were presented each sentence only once.

Listeners were presented a total of 70 sentence lists (5 talkers x 7 lists/talker x 2 modes/list). In each session, listeners were tested on 4-5 sentence lists. Every session included a five-minute break after the presentation of each list. In addition, a 10-minute break was given near the halfway point of each session. Listeners were also encouraged to rest briefly as necessary.

4.1.3 Presentation Setup

The processed waveforms were stereo signals with speech on one channel and speech-shaped noise of the same rms on the other channel. The speech-shaped noise samples were originally developed for the Hearing in Noise Test described by Nilsson *et al.*[13]. The waveforms were played from a PC through a DAL card. The PC was controlled by one of the listeners, who was seated at a terminal of the VAX. The speech was attenuated by 1.8 dB and added to the speech-shaped noise. The resulting signal (SNR = –1.8 dB) was presented to the listeners monoaurally over headphones. All listeners were seated in a sound-treated room. The eight listeners were divided into two separate testing groups. Although the groups met at different times, both groups heard the lists in the same order.

Table 4.1: Talker identification labels for the five talkers used in intelligibility tests.

Talker	Subject ID	Talker ID
EK	S6	T1
JM	S9	T2
MI	S11	T3
RG	S2	T4
SA	S5	T5

4.2 Results of Intelligibility Tests

To simplify notation, the five talkers selected for the intelligibility experiment are designated T1 through T5 (see Table 4.1). For each of the talkers, the intelligibility scores (I), averaged across listeners, and the speaking rates (r), including pauses, are presented in Table 4.2. The intelligibility results for individual listeners are listed in Appendix C.

As mentioned previously, all clipped waveforms were identified and rerecorded. Unfortunately, some of these newly recorded waveforms were lost when a hard disk failed. As a result, several of the sentence lists used in the intelligibility experiments contained a few clipped sentences. In most cases, the clipping was very minor, affecting at most a few hundred out of roughly 250 thousand samples. For sentences recorded by T5 in CLEAR mode, however, the clipping is more severe. Consequently, his key-word score in CLEAR mode of 77 percent may not be an accurate reflection of his intelligibility. It is not likely that the score is greatly elevated, however, since he achieved a score of 73 percent for his FAST_CLEAR speech, which was not clipped. All other modes recorded by T5 were unaffected.

4.2.1 Speaking Mode Results

The effect of speaking mode on intelligibility is displayed in Figure 4-1. CLEAR mode was most intelligible at 63 percent key-words correct, followed in order of decreasing intelligibility by FAST_CLEAR (59%), CONV+PAUSE (58%), SLOW (56%), LOUD (53%),

QUICK_CLEAR (46%), CONV (45%), QUICK (27%), SOFT (26%) modes. The 18 percent advantage for CLEAR relative to CONV mode is consistent with Picheny (15%)[16] and Uchanski (17%)[20].

The effect of the factor talker is shown in Figure 4-2. The QUICK, CLEAR, and SOFT speaking modes exhibited the largest talker effect, with scores ranging over 46, 43, and 52 points, respectively, across talkers.

Of the speaking modes tested, none provided as much of an intelligibility advantage over CONV speech as the CLEAR mode. CONV+PAUSE exhibited a 14 point increase relative to the CONV condition, but on average this mode was 2 points less intelligible than SLOW speech. For all the talkers, LOUD speech was more intelligible than CONV speech, but on average the advantage was less for LOUD speech than for CLEAR speech. SOFT speech was less intelligible on average than CONV speech.

For four of the seven sentence lists recorded by each talker, an analysis of variance was performed on the intelligibility increase over CONV mode after an arcsine transformation ($\arcsin\sqrt{I_j/100}$) to normalize the variances. Table 4.3 shows the results of this analysis with the factors talker, listener, and speaking mode. All three main factors as well as the talker x mode and the listener x talker x mode interactions were significant at the 0.01 level. The values of the F-distribution used for the F-tests were obtained from Bennett and Franklin[1]. When necessary, values were interpolated using the reciprocal of the degrees of freedom. The speaking modes included in this analysis were SOFT, LOUD, CLEAR, and CONV+PAUSE. Only these speaking modes were analyzed because the other modes also included an additional experimental factor, speaking rate. A separate ANOVA was performed for modes including the rate factor (see Section 4.2.2).

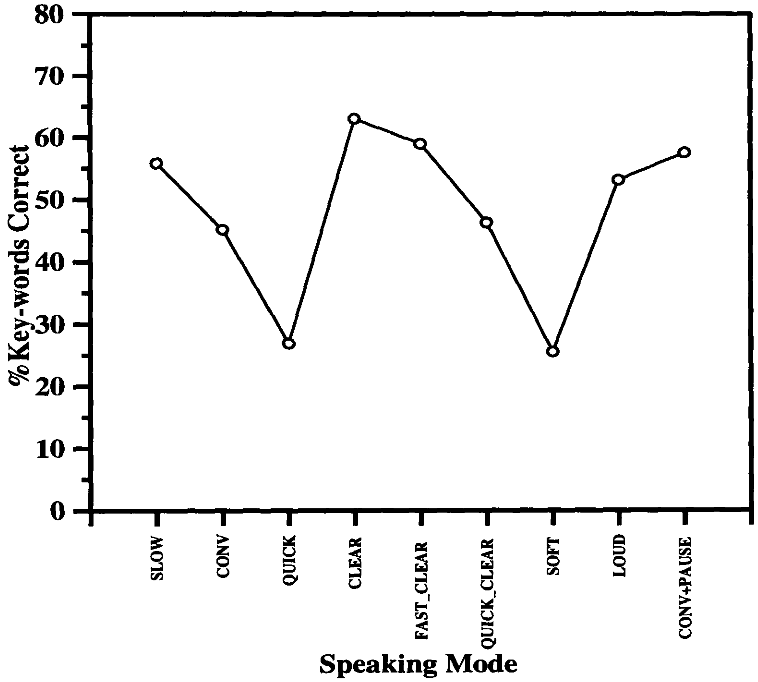

Figure 4-1: Average key-word scores versus speaking condition. Data represent an average over talker and listener.

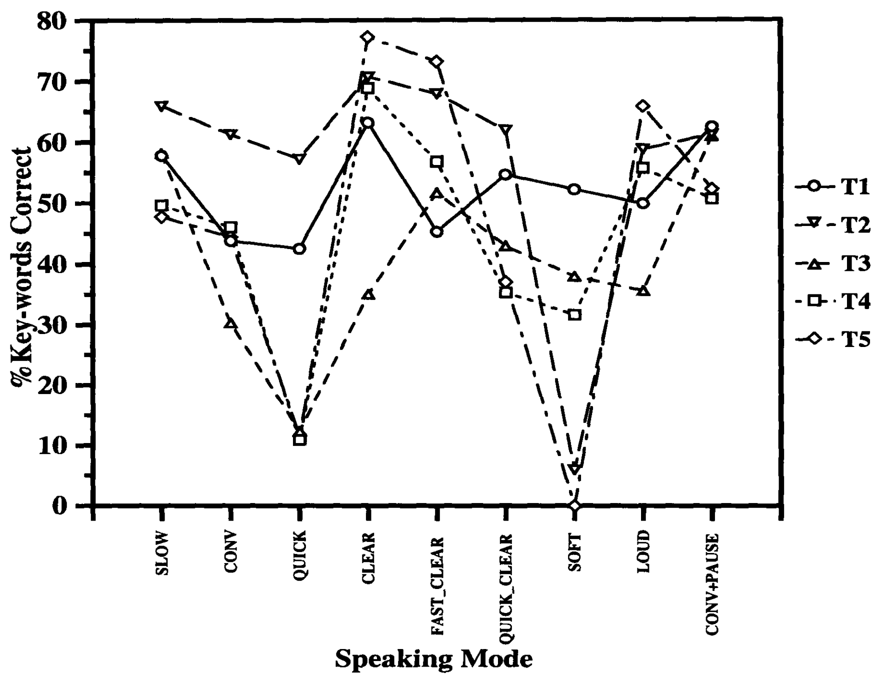

Figure 4-2: Average key-word scores for each talker versus speaking condition. Data represent an average over listener.

Table 4.2: Percent correct key-word scores (I) and speaking rates in wpm (r) for each of the five talkers. Key-word scores are averaged across all eight normal hearing listeners.

List	Mode		Talker T1	T2	T3	T4	T5	AVG
normal_rate1	CONV	I	44.8	63.1	25.1	47.1	45.6	45.1
		r	196	143	196	190	173	180
	SOFT	I	52.2	5.9	37.9	31.6	0.0	25.5
		r	168	145	190	184	170	171
normal_rate2	CONV	I	38.1	57.7	40.7	49.0	40.6	45.2
		r	186	140	204	191	174	179
	LOUD	I	49.9	58.8	35.5	55.7	65.8	53.1
		r	162	141	179	206	162	170
normal_rate3	CONV	I	51.7	66.6	28.6	46.3	51.8	49.0
		r	186	149	201	196	169	179
	FAST_CLEAR	I	45.3	67.9	51.7	56.8	73.2	59.0
		r	187	144	200	186	146	174
normal_rate4	CONV	I	40.8	57.7	26.7	42.0	39.9	41.4
		r	185	144	191	189	169	175
	CLEAR	I	59.8	71.1	33.3	67.7	76.0	61.6
		r	78	129	68	47	123	89
quick_rate1	CONV	I	42.6	57.2	12.4	10.9	11.6	26.9
		r	242	193	275	324	312	269
	QUICK_CLEAR	I	54.6	61.9	42.9	35.3	37.0	46.3
		r	205	172	228	254	230	218
slow_rate1	SLOW	I	51.6	64.3	48.2	43.1	46.4	50.7
		r	90	144	94	59	127	103
	CLEAR	I	66.6	70.3	36.8	69.9	78.5	64.4
		r	86	128	71	46	116	89
slow_rate2	SLOW	I	64.0	67.6	68.1	56.2	49.2	61.0
		r	87	142	94	56	133	102
	CONV+PAUSE	I	62.5	61.2	60.9	50.6	52.3	57.5
		r	75	117	74	55	117	87

Table 4.3: Analysis of variance of the increase in intelligibility between each test mode and CONV mode for four modes (SOFT, LOUD, CLEAR, and CONV+PAUSE). Factors which are significant at a 0.01 level are indicated by asterisks.

Sum of Squares	Mean Square	Degrees of Freedom	F-ratio	$\%\omega^2$	FACTOR
0.1279	0.0320	4	-	0.2	REPS (R)
0.3877	0.0554	7	3.9	0.5	* LISTENER (L)
0.3602	0.0129	28	-	0.5	LxR
5.1575	1.2893	4	65.0	7.3	* TALKER (T)
0.5287	0.0330	16	-	0.7	TxR
0.5559	0.0199	28	1.4	0.8	TxL
1.2051	0.0108	112	-	1.7	TxLxR
25.4495	8.4832	3	350	36.0	* MODE (M)
0.1591	0.0132	12	-	0.2	MxR
0.5090	0.0242	21	1.6	0.7	MxL
1.0049	0.0120	84	-	1.4	MxLxR
27.6538	2.3045	12	103.5	39.1	* MxT
1.8867	0.0393	48	-	2.7	MxTxR
1.8709	0.0222	84	1.6	2.6	* MxTxL
3.8723	0.0115	336	-	5.5	MxTxLxR
70.7293	0.0885	799			TOTAL
9.1450	0.0143	640			Residual (Error term)

4.2.2 Speaking Rate Results

The effect of speaking rate on intelligibility is displayed in Figure 4-3. QUICK mode was the most rapid at 269 words-per-minute, followed in order of decreasing rate by QUICK_CLEAR (218 wpm), CONV (178 wpm), FAST_CLEAR (174 wpm), SOFT (171 wpm), LOUD (170 wpm), SLOW (103 wpm), CLEAR (89 wpm), CONV+PAUSE (88 wpm) modes. The rates for CLEAR and CONV modes are consistent with the speaking rates reported by Picheny[17].

The effect of the factor talker is shown in Figure 4-4. The QUICK speaking mode exhibited the largest talker effect, with scores ranging 131 wpm across talkers.

Figure 4-5 shows the intelligibility data as a function of speaking rate. Since only conversational modes (SLOW, CONV, and QUICK) and clear modes (CLEAR, FAST_CLEAR, and QUICK_CLEAR) had speaking rate as a factor, the intelligibility results for these modes as a function of speaking rate are also plotted separately in Figure 4-6. At an average speaking rate of 174 wpm, the key-word score for the FAST_CLEAR mode was 14 points higher than for the CONV speech at nearly the same speaking rate (178 wpm).

The effect of the talker factor is shown in Figures 4-7 through 4-11. Instances where a talker achieved an intelligibility benefit without significantly changing his/her speaking rate are indicated with dotted lines. Each talker obtained an increase in intelligibility between CONV and FAST_CLEAR modes. Talkers also obtained an increase in intelligibility between QUICK and QUICK_CLEAR modes, although every talker also reduced his/her speaking rate in order to achieve this increase. For both T4 and T5, key-word scores for SLOW speech were only four percentage points higher than for CONV speech. For these talkers, CLEAR speech was much more intelligible than conversational speech at the nearly the same speaking rate (18 point higher key-word scores for T4 and 28 point higher key-word scores for T5). Trends for the other three talkers are less clear. T3 failed to produce highly intelligible speech at the slow speaking rate, although her FAST_CLEAR and QUICK_CLEAR modes were more intelligible than her conversational speech at similar speaking rates. T2 varied her speaking rate the least of all the talkers and reported having difficulty adhering to

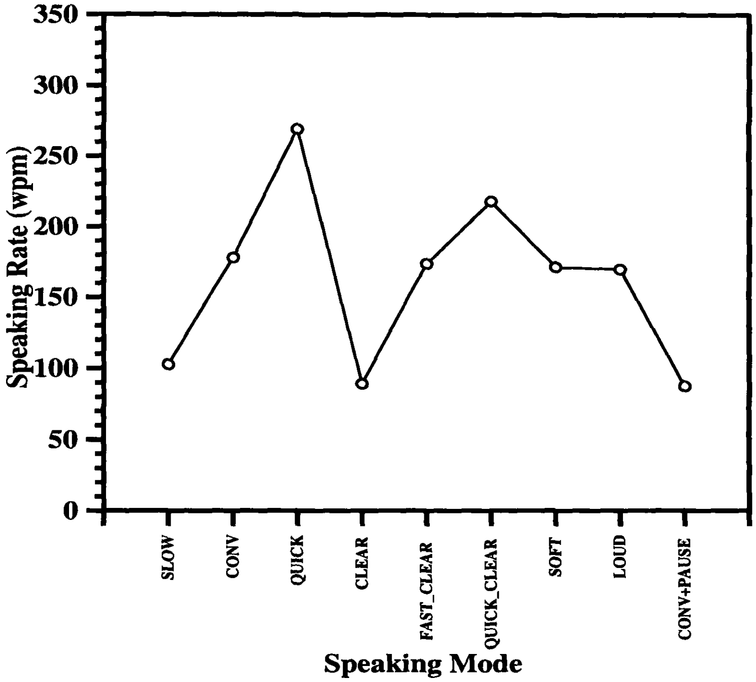

Figure 4-3: Average speaking rates versus speaking condition. Data represent an average over talker and listener.

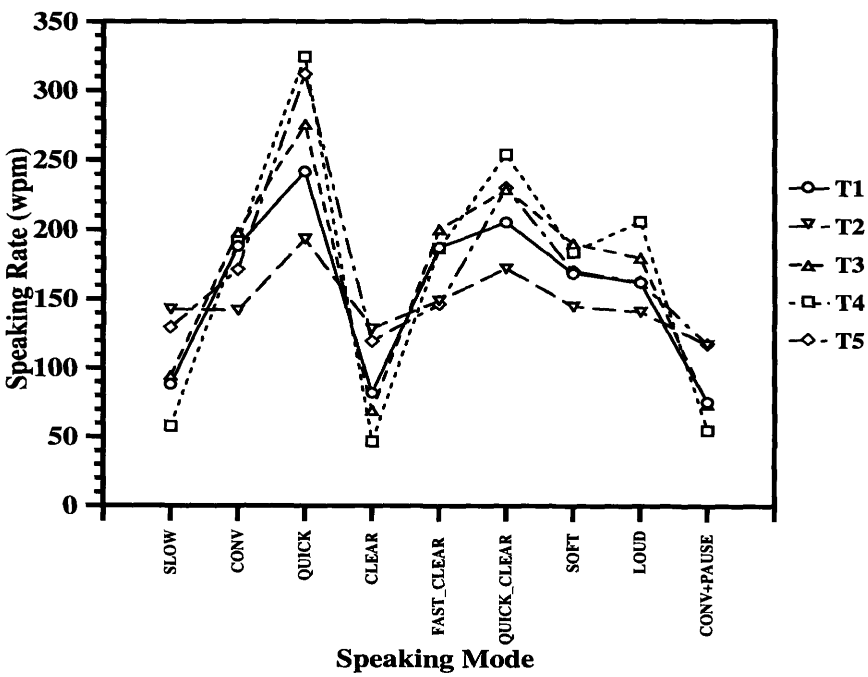

Figure 4-4: Average speaking rates for each talker versus speaking condition. Data represent an average over listener.

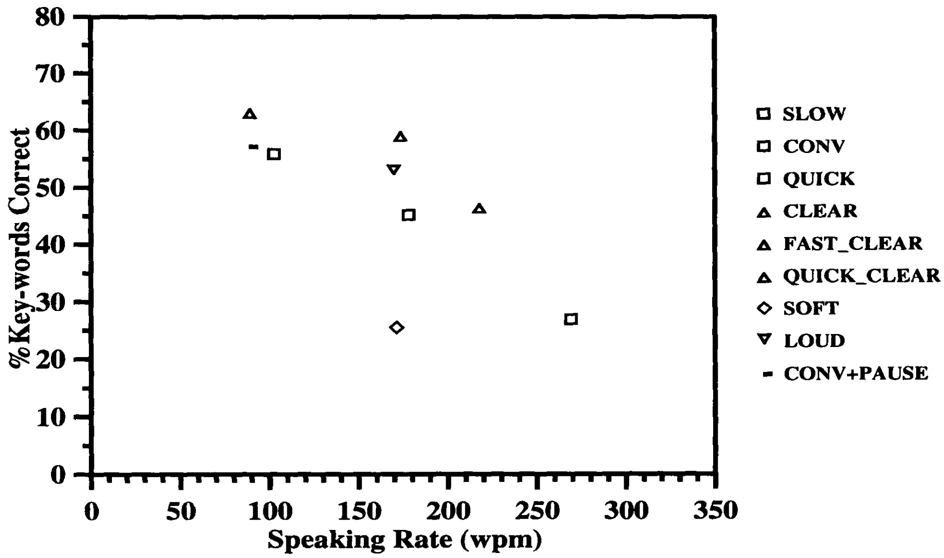

Figure 4-5: Average key-word scores versus speaking rate. Data represent an average over talker and listener.

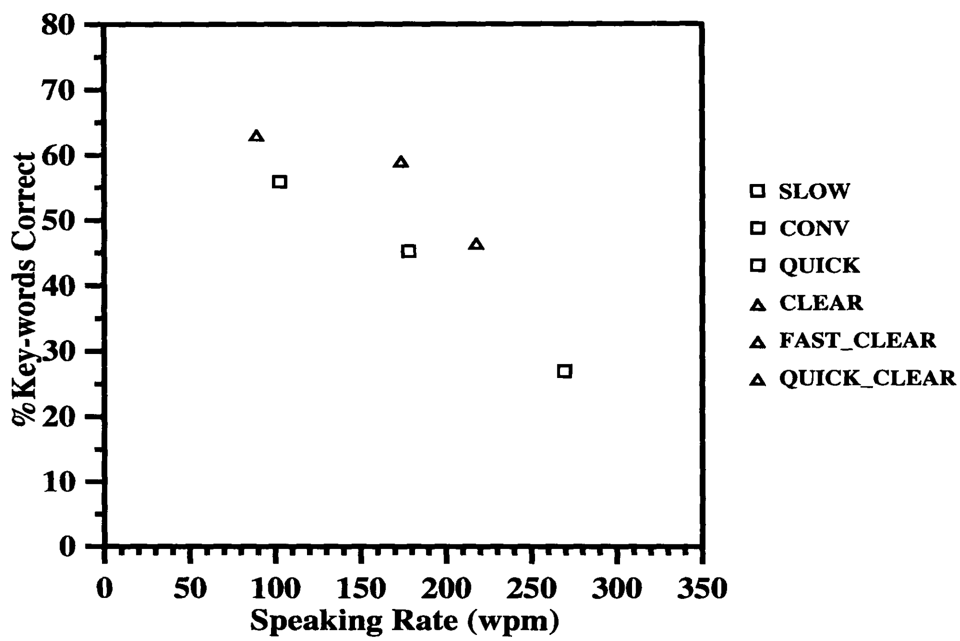

Figure 4-6: Key-word scores, averaged over talker and listener, versus speaking rate. Conversational speech (SLOW, CONV, and QUICK) is indicated with squares, and clear speech (CLEAR, FAST_CLEAR, and QUICK_CLEAR) is indicated with triangles.

the metronome. Her intelligibility drops off quickly at speaking rates above 150 wpm. T1 reported that she preferred speaking quickly, which may partly explain the high key-word score obtained for her QUICK_CLEAR mode.

For the three sentence lists recorded by each talker in conversational and clear modes at different speaking rates, an analysis of variance was performed on the intelligibility increase over conversational mode after an arcsine transformation to normalize the variances. Table 4.4 shows the results of this analysis with the factors talker, listener, and speaking rate. The speaking rates used in the analysis were the nominal rates slow, normal, and quick. The rate factor, listener x rate, talker x rate, and listener x talker x rate interactions were significant at the 0.01 level.

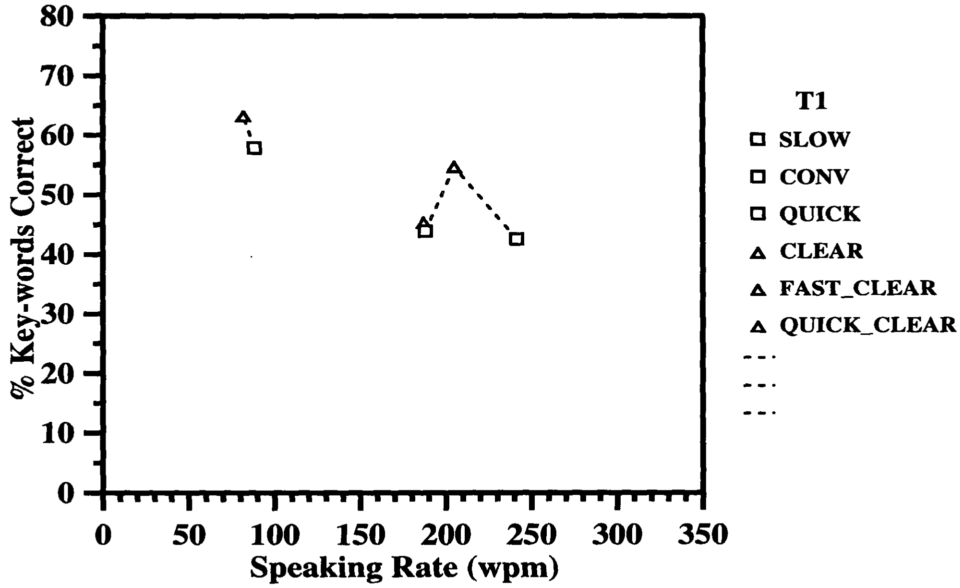

Figure 4-7: Average key-word scores for T1, averaged over listener, versus speaking rate. Dotted lines represent instances where intelligibility was improved without a reduction in speaking rate.

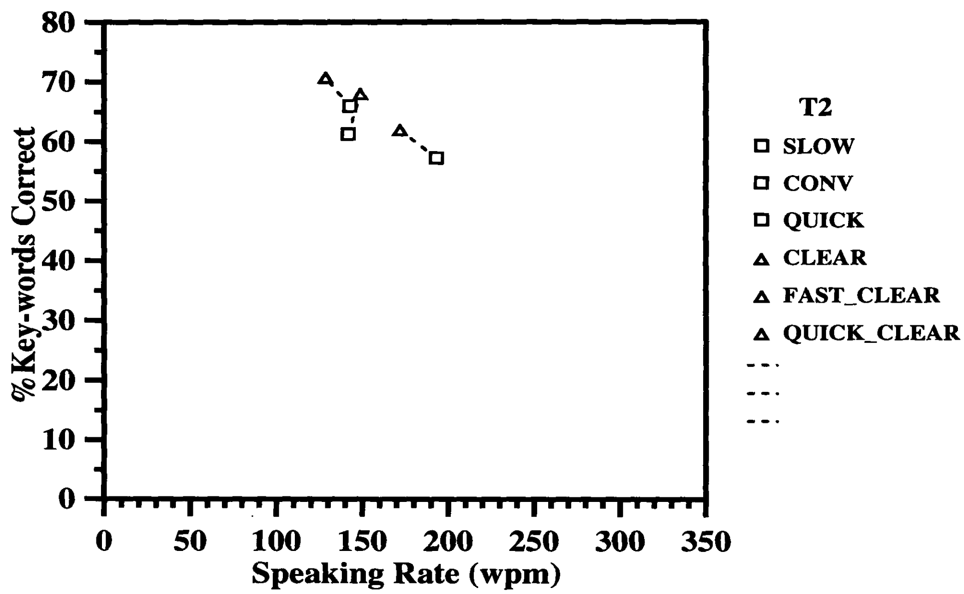

Figure 4-8: Average key-word scores for T2, averaged over listener, versus speaking rate. Dotted lines represent instances where intelligibility was improved without a reduction in speaking rate.

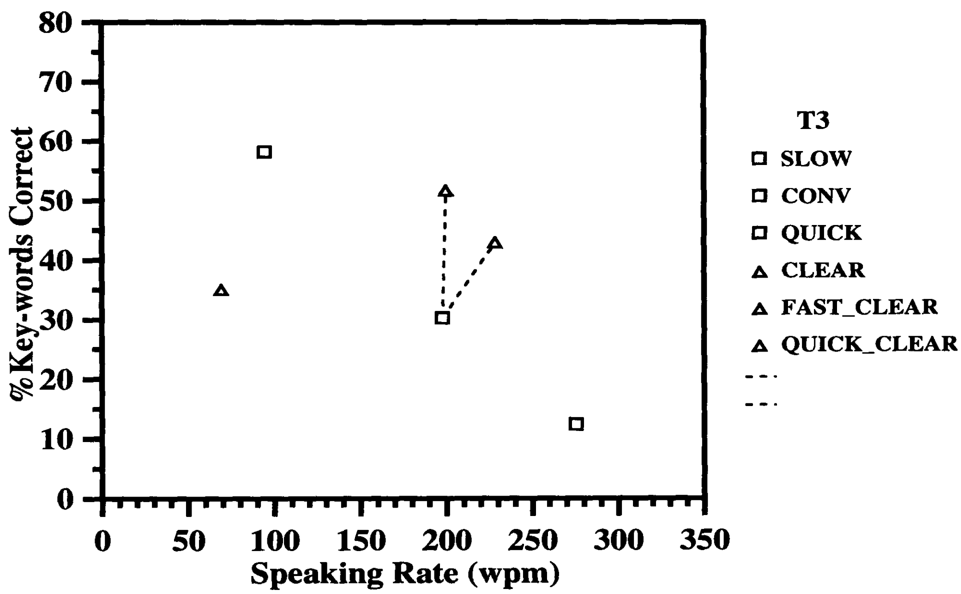

Figure 4-9: Average key-word scores for T3, averaged over listener, versus speaking rate. Dotted lines represent instances where intelligibility was improved without a reduction in speaking rate.

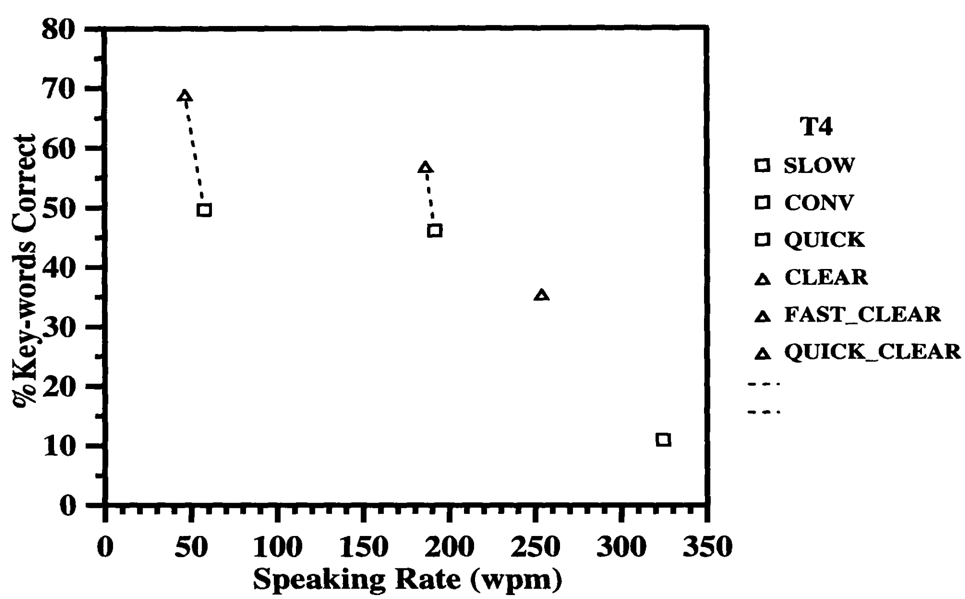

Figure 4-10: Average key-word scores for T4, averaged over listener, versus speaking rate. Dotted lines represent instances where intelligibility was improved without a reduction in speaking rate.

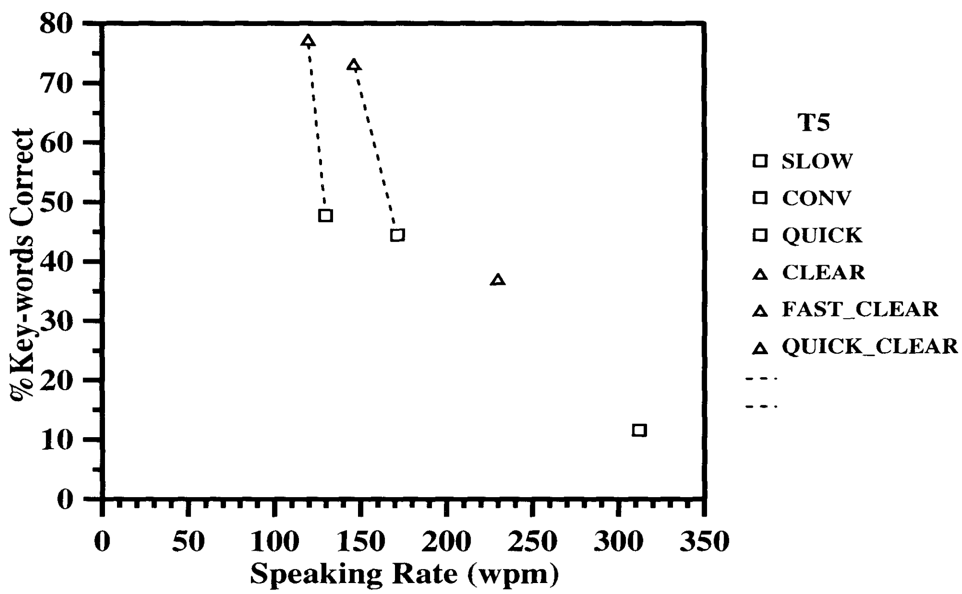

Figure 4-11: Average key-word scores for T5, averaged over listener, versus speaking rate. Dotted lines represent instances where intelligibility was improved without a reduction in speaking rate.

Table 4.4: Analysis of variance of the increase in intelligibility between clear and conversational speech at three speaking rates. Factors which are significant at a 0.01 level are indicated by asterisks.

Sum of Squares	Mean Square	Degrees of Freedom	F-ratio	$\%\omega^2$	FACTOR
0.0096	0.0024	4	-	0.0	REPS (R)
0.8101	0.1157	7	2.3	3.4	LISTENER (L)
0.2195	0.0078	28	-	0.9	LxR
5.4125	1.3531	4	1.9	22.9	TALKER (T)
0.4204	0.0263	16	-	1.8	TxR
0.8281	0.0296	28	0.9	3.5	TxL
1.0770	0.0096	112	-	4.5	TxLxR
1.6473	0.8236	2	16.1	7.0	* SPEAKING RATE (SR)
0.2111	0.0264	8	-	0.9	SRxR
0.7152	0.0511	14	3.7	3.0	*SRxL
0.7953	0.0142	56	-	3.4	SRxLxR
5.8538	0.7317	8	23.4	24.7	* SRxT
1.3720	0.0429	32	-	5.8	SRxTxR
1.7522	0.0313	56	2.3	7.4	* SRxTxL
2.5564	0.0114	224	-	10.8	SRxTxLxR
23.6805	0.0395	599			TOTAL
6.6612	0.0139	480			Residual (Error term)

4.3 Summary of Results

Of the additional speaking modes tested, none provided intelligibility benefits as great as CLEAR speech. SOFT mode was on average less intelligible than CONV speech. Although the CONV+PAUSE mode exhibited a 14 point higher key-word score than the CONV mode, it was 2 points less intelligible on average than SLOW speech. LOUD mode on average exhibited eight percent higher key-word scores than CONV, but this advantage is less than the advantage provided by both CLEAR and FAST_CLEAR modes.

All talkers succeeded at achieving an intelligibility advantage over conversational speech in both CLEAR and FAST_CLEAR mode. For all talkers except T3, CLEAR speech was also more intelligible than SLOW speech. These results support the hypothesis that there are acoustic properties of clear speech other than speaking rate which are responsible for its high intelligibility. In addition, all talkers exhibited higher key-word scores for QUICK_CLEAR mode than for QUICK mode. To achieve this increase, however, every talker reduced his/her speaking rate.

Chapter 5

Discussion

In previous studies, highly intelligible speech was not achieved at normal speaking rates through artificial or natural means. Consequently, an important objective of this study was to elicit clear speech at normal speaking rates. In this chapter, the results of the intelligibility experiments are evaluated in the context of the goals of the thesis. In addition, the results are compared to data obtained in other experiments by Picheny[15], Uchanski[20], and Payton[14].

5.1 Goals of Intelligibility Experiment

Previous studies had not identified the contribution of speaking rate to the high intelligibility of clear speech. This thesis was designed to investigate the relationship between speaking rate and intelligibility in two ways. First, a method was designed for eliciting clear speech at normal speaking rates (FAST_CLEAR speech). This method emphasized selecting talkers with previous speaking experience and providing them with additional training. During training, the setup for eliciting FAST_CLEAR speech allowed talkers to receive interactive feedback on both intelligibility and rate. With this feedback, all talkers achieved an intelligibility increase over CONV mode. If the increase was at least five percentage points and the corresponding speaking rate was within 15% of his/her normal rate, then the talker succeeded at producing a form of clear speech at near-normal rates. Talkers T2, T3, T4, and T5 met this criterion, producing FAST_CLEAR speech that was more intelligible than CONV speech at nearly the same speaking rate.

A second measure for investigating speaking rate effects was evaluating the intelligibility of SLOW speech for each of the talkers. In this case, the goal was to determine whether slow speech, without emphasis on clarity, has comparable intelligibility to clear speech. If so, then it would be reasonable to attribute the high intelligibility of clear speech to its slow rate. On the other hand, if slow speech was less intelligible than clear speech, then the acoustical differences between the two speaking styles could help identify which factors are responsible for the high intelligibility of clear speech. As with the criterion for FAST_CLEAR speech, if the difference in intelligibility between CLEAR and SLOW speech was at least five percentage points and the corresponding speaking rates were within 15% of his/her slow rate, the talker may be considered to have successfully produced CLEAR speech by manipulating factors other than speaking rate. Talkers T1, T2, T4, and T5 produced CLEAR speech which met this objective.

In addition to comparing intelligibility within speaking rates, it is also useful to compare intelligibility across speaking rates. In several instances, a talker's FAST_CLEAR speech was more intelligible than his/her SLOW speech. Similarly, in some cases, QUICK_CLEAR speech was more intelligible than the CONV mode, which was spoken at a slower rate. These results (marked with dotted lines in Figures 4-7 through 4-11) again support the hypothesis that the talker manipulated factors other than speaking rate to improve his/her intelligibility, which was the primary goal of the experiment.

Another goal of this thesis was to investigate the intelligibility of several additional speaking modes. In particular, the intelligibility of SOFT, LOUD, and CONV+PAUSE modes were examined. In each case, the objective was to obtain an additional speaking mode which was at least as intelligible as clear speech at the corresponding speaking rate.

5.2 Evaluation of Results

The attempt to elicit additional speaking modes that were at least as intelligible as clear speech was not successful. Overall, the key-word scores for SOFT, LOUD, and CONV+PAUSE modes were lower than the scores for clear speech at the corresponding speaking rate. T3 was an exception, however, achieving a score of 61% for her CONV+PAUSE mode, which was higher than her scores for both CLEAR (35%) and SLOW (58%) modes. The only other exception was T1, whose scores for SOFT (52%) and LOUD (50%) modes exceeded her score for FAST_CLEAR (45%) mode. In QUICK_CLEAR mode, however, her 55% key-word score was higher than scores for all three of the conversational modes. Thus, in nearly every case, the scores for clear speech at the appropriate speaking rate were higher than for any of the other speaking modes.

Figures 4-7 through 4-11 show, however, that the results of the speaking rate experiments were more promising. Dotted lines indicate instances where talkers improved intelligibility without reducing speaking rate. Talkers T1, T2, T4, and T5 successfully produced CLEAR speech which was more intelligible than SLOW speech at roughly the same rate, and talkers T2, T3, T4, and T5 successfully produced FAST_CLEAR speech which was more intelligible than CONV speech at the same rate. At the quick rate, talkers T1 and T2 met the criterion for success, producing QUICK_CLEAR speech which was more intelligible than QUICK speech.

To investigate the relationship between rate and intelligibility further, the key-word scores for each of the talkers were plotted as a function of speaking rate. Figure 5-1 depicts the scores for clear speech (CLEAR, FAST_CLEAR, and QUICK_CLEAR modes), and Figure 5-2 shows the scores for conversational speech (SLOW, CONV, and QUICK modes). In both graphs, the intelligibility function can be roughly described by two linear segments which meet at 200 wpm. In general, both graphs exhibit a small negative slope for speaking rates less than 200 wpm ($m1$) followed by a more negative slope for rates greater than 200 wpm ($m2$). Slopes $m1$ and $m2$ were calculated and for each of the modes and averaged across talkers. The results (represented by solid lines in Figures 5-1 and 5-2) are summarized in Table 5.1. Slope $m1$ is roughly the same for

Table 5.1: Results of $m1$ and $m2$ averaged across talker. For both modes, slopes are calculated by the equations $m1 = (I_{normal} - I_{slow})/(r_{normal} - r_{slow})$ and $m2 = (I_{quick} - I_{normal})/(r_{quick} - r_{normal})$. Subscripts represent nominal speaking rate. Any positive slope data was considered an anomaly and omitted from the calculation. Thus, T1 and T2 were excluded from conversational mode calculations, and T1 and T3 were excluded from clear mode calculations.

Mode	$m1$	$m2$
conversational	-0.053	-0.24
clear	-0.053	-0.34

both modes, indicating that on average the intelligibility benefit of clear modes over conversational modes is robust for speaking rates up to nearly 200 wpm. As rates get faster, however, clear speech intelligibility drops off more quickly ($m1_{clear} < m1_{conv}$), which could be a reflection of physical limitations on articulation at very high rates.

5.3 Comparison with Previous Data

Direct comparisons of intelligibility scores from previous studies are difficult for a number of reasons. First, Uchanski[20], Picheny[15] and Payton[14] used hearing impaired subjects in many of their intelligibility tests. Although Uchanski and Payton also used normal hearing listeners, the listening conditions were different from the conditions used in this study. Secondly, the methods used for processing sentences varied in each of the studies. Uchanski normalized the long-term rms, including pauses, of sentences and presented the signal to the listeners at SNR = –4 dB. Payton tested at several different signal-to-noise ratios, using both speech-shaped and white noise. Because of these experimental differences, direct comparisons of intelligibility scores are meaningless. A comparison of the increase in intelligibility as a function of speaking rate, however, lends insight into the relationship between intelligibility and rate. To perform this comparison, an arcsine transform ($\arcsin\sqrt{I_j/100}$) of the percent correct key-word scores from each study was computed to normalize variance in scores. The slopes $m1$ and $m2$ were then calculated for the transformed data

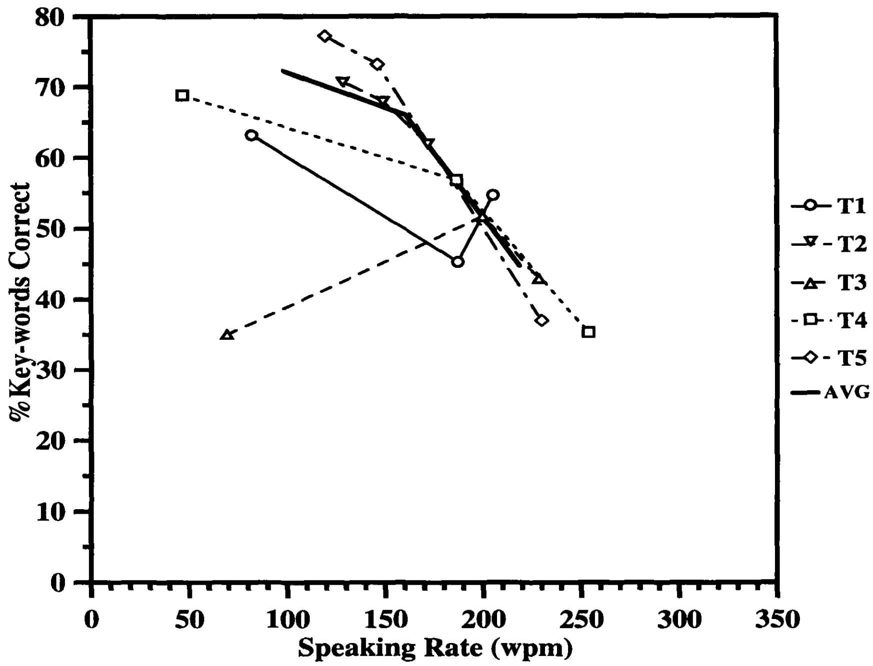

Figure 5-1: Average key-word scores for clear speech at slow, normal and quick speaking rate for all five talkers. Solid lines represent $m1$ and $m2$ from Table 5.1.

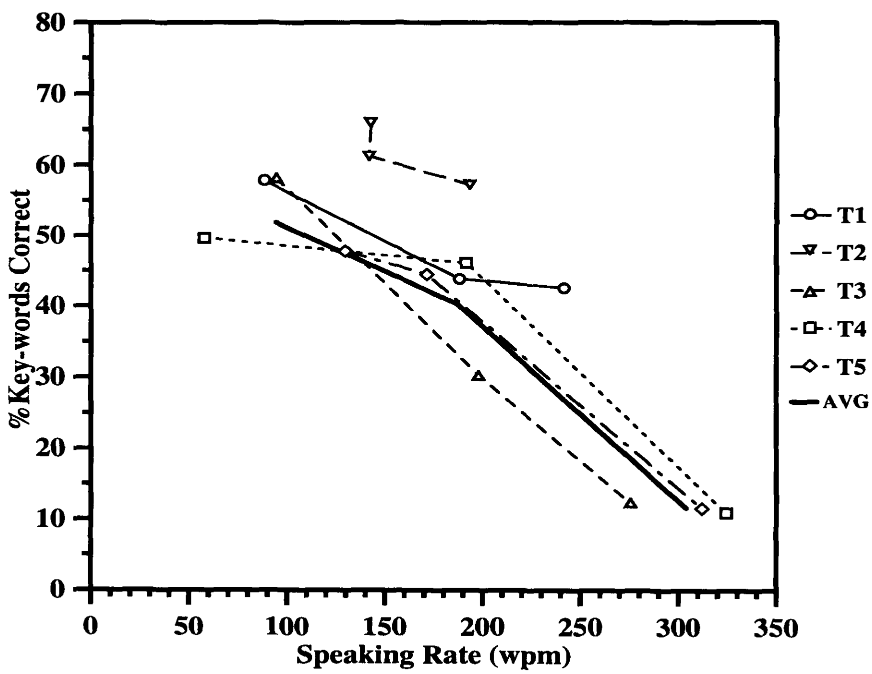

Figure 5-2: Average key-word scores for conversational speech at slow, normal and quick speaking rate for all five talkers. Solid lines represent $m1$ and $m2$ from Table 5.1.

Table 5.2: Results of $m1$ and $m2$ averaged across talker for each study. The arcsine transform was used to normalize the variance from each study. Slopes are scaled by a factor of 100 for ease of display.

Mode	Study	Talker(s)	Listener(s)	Scaling	$100m1$	$100m2$
conversational	present	T3, T4, T5	normal	natural	-0.13	-0.22
clear	present	T2, T4, T5	normal	natural	-0.14	-0.35
clear	Picheny	MM	impaired	uniform	-0.31	NA
clear	Uchanski	MM	impaired	non-uniform	-0.29	NA
clear	Uchanski	JM	normal	natural	-0.25	-0.31
clear	Uchanski	JM	impaired	natural	-0.24	-0.29

from each study. The results are summarized in Table 5.2, which compares $m1$ and $m2$ values from this study with the uniform time-scaling experiment performed by Picheny[18], the non-uniform time-scaling experiment performed by Uchanski[21], and the preliminary natural elicitation experiment with professional speaker, John Moschitta, Jr. (JM), performed by Uchanski[20].

Table 5.2 shows that in all other studies considered, intelligibility of clear speech decreases more rapidly as a function of speaking rate than intelligibility of conversational speech. Consequently, none of the other studies succeeded in obtaining highly intelligible FAST_CLEAR speech at normal speaking rates. These studies, however, did not focus on talker selection and training. None of the other studies provided the talkers with training, and the only talker selected for his speaking ability was JM in Uchanski's experiment. In the present study, focusing on talker selection and training greatly improved FAST_CLEAR scores. As a result, the intelligibility benefit of clear speech was maintained from slow through normal speaking rates. This benefit can be seen in Table 5.2, where $m1$ is nearly the same for both conversational and clear modes (-0.13 and -0.14, respectively).

5.4 Summary

The relationship between intelligibility and speaking rate may be quite complex. The results of this study show that with proper training, some talkers can obtain clear speech at normal speaking rates. Therefore, properties other than speaking rate must account for the high intelligibility of clear speech. Yet even talkers who routinely practice quick speech are less intelligible at high rates, as shown by the fact that $m2 < m1$ for all talkers. Thus, it seems likely that physical limitations on articulation may reduce intelligibility at high rates. For the talkers in this study, the ability to preserve an intelligibility benefit above conversational speech decreases quickly at rates above 200 wpm. With more practice, it is possible that this boundary could be increased. Regardless, the rates that have been achieved will be useful for making acoustical comparisons between FAST_CLEAR and CONV speaking modes. Such analysis may help identify which factors are responsible for the high intelligibility of clear speech.

Chapter 6

Conclusion

This thesis was designed to study the effects of speaking mode and speaking rate on intelligibility. Procedures were developed for eliciting CONV+PAUSE, SOFT, and LOUD modes from talkers. In addition, a structured method was defined for eliciting clear speech naturally at both slow and normal speaking rates. This method for eliciting clear speech emphasized talker selection and training.

6.1 Talker Selection and Training

Fifteen talkers with significant speaking experience were recruited in the first phase of this study. During preliminary intelligibility tests, the five selected talkers demonstrated unusual characteristics in intelligibility and/or speaking rate. These talkers practiced producing clear speech at all three speaking rates for approximately four hours with feedback on intelligibility from listeners. The talkers were encouraged to experiment with different speaking strategies during practice sessions.

Talkers reported that the listener feedback was very helpful for developing clear speech. In particular, one talker noted that trends in listener responses raised his awareness of common phoneme confusions. He reported that this information was useful in deciding which phonemes to emphasize. Other talkers expressed interest in listening to speech distorted by multiplicative noise in order to gain information on how to speak more clearly. This request suggests that some talkers believe they have natural strategies for speaking clearly in difficult communication situations. Moreover, these strategies may differ depending on the nature of the distortion.

After practicing with feedback, all talkers except T1 produced FAST_CLEAR which was a form of clear speech at near normal speaking rates. The talkers also developed QUICK_CLEAR speech which was more intelligible than QUICK speech. Although talkers T3, T4, and T5 slowed their speaking rates significantly to achieve this intelligibility benefit, it is possible that with more training the talkers may have attained quick speaking rates.

6.2 Intelligibility Results

The results of the intelligibility experiments confirmed the roughly 15 point intelligibility advantage of CLEAR speech over conversational speech reported by Picheny[15] and Uchanski[20]. Furthermore, for every talker except T3, CLEAR speech was also more intelligible than SLOW speech. Since these two modes were recorded at approximately the same speaking rate, this result suggests that acoustical properties other than speaking rate may be responsible for the high intelligibility of clear speech. The successful elicitation of FAST_CLEAR speech at normal speaking rates further supports this hypothesis. On average FAST_CLEAR speech was 14 points more intelligible than CONV mode. Thus, the intelligibility benefit of clear speech over conversational speech was preserved for speaking rates up to approximately 200 wpm. With more practice, it is possible that talkers could learn to produce highly intelligible clear speech at faster speaking rates as well.

6.3 Suggestions for Future Work

The intelligibility results have established that clear speech exists for speaking rates up to nearly 200 wpm. This finding suggests that acoustical factors other than speaking rate may be responsible for the high intelligibility of clear speech. Therefore, the acoustical differences between CLEAR and SLOW speech and between FAST_CLEAR and CONV speech should be examined. In addition to acoustic differences, a more detailed analysis of rate differences would also be helpful. One method for studying such differences would be to analyze pause durations and segmental level durations for each of the speaking modes. This analysis is essential for further understanding of the contribution of speaking rate to intelligibility.

Relating word-level errors and phoneme-level errors to acoustical differences could help identify which factors are correlated with an increase in intelligibility. This analysis may prove difficult, however, due to word and phoneme omissions by listeners. If so, a fixed set of nonsense syllables could be recorded in a variety of speaking modes and rates. The syllables could be elicited using techniques similar to those methods described in this study. With a fixed set of syllables, word and phoneme omissions would be less frequent, and the relationship between perceptual errors and acoustical differences between modes could be fully explored.

Bibliography

[1] Carl A. Bennett and Norman L. Franklin. *Statistical Analysis in Chemistry and the Chemical Industry*, chapter 7, pages 319–468, 708–709. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, Inc., London, 1954.

[2] F. R. Chen. Acoustic characteristics and intelligibility of clear and conversational speech. Master's project, Massachusetts Institute of Technology, Research Laboratory of Electronics, May 1980.

[3] S. Choi. The effect of pauses on the intelligibility of sentences. Bachelor's project, Massachusetts Institute of Technology, Research Laboratory of Electronics, September 1987.

[4] Sung C. Choi. *Introductory Applied Statistics in Science*, chapter 9, pages 191–220. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1978.

[5] T. H. Crystal and A. S. House. Segmental durations in connected-speech signals: Preliminary results. *Journal of the Acoustical Society of America*, 72:705–716, 1982.

[6] T. H. Crystal and A. S. House. Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America*, 83(4):1553–1573, April 1988.

[7] Olive Jean Dunn and Virginia A. Clark. *Applied Statistics: Analysis of Variance and Regression*, chapter 10, pages 218–232. Wiley Publications in Statistics. John Wiley and Sons, Inc., New York, second edition, 1987.

[8] F. Goldman Eisler. *Psycholinguistics: Experiments in Spontaneous Speech.* Academic Press, New York, New York, 10 January 1968.

[9] D. W. Griffin and J. S. Lim. Signal estimation from modified short-time fourier transform. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 32:236–243, 1984.

[10] M. S. Han. Acoustic-phonetic study on speech tempo. *Stud. Sounds*, 12:70–83, 1966.

[11] M. S. Harris and N. Umeda. Effect of speaking mode on temporal factors in speech: Vowel duration. *Journal of the Acoustical Society of America*, 56(3):1016–1018, September 1974.

[12] D. Malah. Time-domain algorithms for harmonic bandwidth reduction and time scaling of speech signals. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 27(2):121–133, 1979.

[13] Michael Nilsson, Sigfrid D. Soli, and Jean A. Sullivan. Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, 95:1085–1099, February 1994.

[14] K. L. Payton, R. M. Uchanski, and L. D. Braida. Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *Journal of the Acoustical Society of America*, 95(3):1581–1592, March 1994.

[15] M. A. Picheny. *Speaking Clearly for the Hard of Hearing.* PhD dissertation, Massachusetts Institute of Technology, Research Laboratory of Electronics, June 1981.

[16] M. A. Picheny, N. I. Durlach, and L. D. Braida. Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, 28:96–103, March 1985.

[17] M. A. Picheny, N. I. Durlach, and L. D. Braida. Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, 29:434–446, December 1986.

[18] M. A. Picheny, N. I. Durlach, and L. D. Braida. Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech and Hearing Research*, 32:600–603, September 1989.

[19] M. R. Schroeder. Reference signal for signal quality studies. *Journal of the Acoustical Society of America*, 44:1735–1736, 1968.

[20] R. M. Uchanski. *Spectral and Temporal Contributions to Speech Clarity for Hearing Impaired Listeners.* PhD dissertation, Massachusetts Institute of Technology, Research Laboratory of Electronics, May 1988.

[21] R. M. Uchanski, S. Choi, L. D. Braida, C. M. Reed, and N. I. Durlach. Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech and Hearing Research, submitted for publication.*, 1993.

[22] A. M. Zwicky. On casual speech. *Papers from the Eighth Regional Meeting, Chicago Linguistic Society*, pages 607–615, 1972.

Appendix A

Sentence Lists

All sentences used in intelligibility tests were from the corpus of sentences described by Picheny *et al.*[15]. Table A.1 explains which sentences were used in each of the sentence lists.

Table A.1: Sentence lists recorded by each talker for formal intelligibility tests. SP and LST corresponds to Picheny's notation for describing the corpus in Appendix B of his thesis[15].

List	Talker				
	T1	T2	T3	T4	T5
normal_rate1	SP1, LST1	SP1, LST8	SP2, LST1	SP2, LST8	SP3, LST1
normal_rate2	SP1, LST2	SP1, LST9	SP2, LST2	SP2, LST9	SP3, LST2
normal_rate3	SP1, LST3	SP1, LST10	SP2, LST3	SP2, LST10	SP3, LST3
normal_rate4	SP1, LST4	SP1, LST11	SP2, LST4	SP2, LST11	SP3, LST4
quick_rate1	SP1, LST5	SP1, LST13	SP2, LST5	SP2, LST12	SP3, LST5
slow_rate1	SP1, LST6	SP1, LST12	SP2, LST6	SP2, LST13	SP3, LST5
slow_rate2	SP1, LST7	SP1, LST14	SP2, LST7	SP2, LST14	SP3, LST7

Appendix B

Listener Audiograms

The hearing levels of all the listeners who participated in the interactive setup for eliciting clear speech described in Section 3.1 are listed in Tables B.1 through B.5. The hearing levels of the eight listeners who participated in the intelligibility tests are listed in Table B.6.

Table B.1: Audiograms for the four listeners who provided talker T1 with feedback on the intelligibility of her clear speech. Numbers reflect hearing level in dB for the ear used during the experiment.

Listener	250 Hz	500 Hz	1000Hz	2000 Hz	4000 Hz
JD	0.0	6.0	6.0	12.0	17.5
TT	5.0	-1.0	5.0	22.5	25.5
CT	10.5	10.5	-19.0	-5.0	-2.0
JD	20.5	15.5	5.0	12.5	4.0

Table B.2: Audiograms for the four listeners who provided talker T2 with feedback on the intelligibility of her clear speech. Numbers reflect hearing level in dB for the ear used during the experiment.

Listener	250 Hz	500 Hz	1000Hz	2000 Hz	4000 Hz
RH	14.0	5.0	0.0	6.0	10.0
ES	6.0	0.0	6.0	12.0	6.0
LB	12.0	12.0	0.0	12.0	6.0
SS	6.0	6.0	6.0	6.0	0.0

Table B.3: Audiograms for the four listeners who provided talker T3 with feedback on the intelligibility of her clear speech. Numbers reflect hearing level in dB for the ear used during the experiment. Subject JL replaced subject JG in the second session.

Listener	250 Hz	500 Hz	1000Hz	2000 Hz	4000 Hz
RM	0.0	6.0	6.0	12.0	17.5
JG	17.5	12.0	6.0	6.0	6.0
JL	6.0	9.5	-1.5	1.5	10.5
AN	0.0	0.0	0.0	0.0	-6.0
MK	10.0	10.0	5.0	10.0	40.0

Table B.4: Audiograms for the four listeners who provided talker T4 with feedback on the intelligibility of her clear speech. Numbers reflect hearing level in dB for the ear used during the experiment.

Listener	250 Hz	500 Hz	1000Hz	2000 Hz	4000 Hz
DH	0.0	0.0	6.0	17.5	6.0
AS	6.0	0.0	-6.0	0.0	6.0
BB	0.0	0.0	6.0	17.5	-12.0
JB	6.0	0.0	0.0	12.0	6.0

Table B.5: Audiograms for the four listeners who provided talker T5 with feedback on the intelligibility of her clear speech. Numbers reflect hearing level in dB for the ear used during the experiment.

Listener	250 Hz	500 Hz	1000Hz	2000 Hz	4000 Hz
JR	12.0	12.0	12.0	17.5	6.0
JS	17.5	12.0	0.0	0.0	0.0
CN	17.5	12.0	12.0	6.0	12.0
JD	12.0	6.0	6.0	-6.0	-12.0

Table B.6: Audiograms for the eight listeners who participated in the intelligibility tests. Numbers reflect hearing level in dB.

Listener		250 Hz	500 Hz	1000Hz	2000 Hz	4000 Hz
CA	R	0.0	0.0	0.0	12.0	6.0
	L	17.5	6.0	17.5	17.5	12.0
AC	R	6.0	6.0	0.0	0.0	6.0
	L	6.0	12.0	6.0	6.0	6.0
AG	R	-6.0	0.0	12.0	6.0	6.0
	L	0.0	0.0	12.0	12.0	-6.0
FK	R	0.0	0.0	12.0	6.0	0.0
	L	0.0	6.0	6.0	6.0	0.0
AM	R	12.0	12.0	12.0	17.5	6.0
	L	12.0	12.0	12.0	17.5	6.0
JP	R	12.0	12.0	17.5	17.5	-6.0
	L	6.0	12.0	6.0	17.5	17.5
JS	R	17.5	17.5	0.0	12.0	6.0
	L	12.0	6.0	-6.0	12.0	17.5
MS	R	-6.0	0.0	0.0	0.0	6.0
	L	-6.0	0.0	6.0	0.0	-6.0

Appendix C

Key-word Scores

The raw scores and corresponding percent correct key-words scores for all the listeners who participated in the intelligibility experiment described in Chapter 4 are listed in Tables C.1 through C.5. A description of which sentences were used for each sentence list can be found in Appendix A.

Table C.1: Raw and percent correct key-word scores for T1.

Talker EK, List *normal_rate1*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	20	21	13	12	13	19	16	24
sublist2	15	18	13	10	8	9	15	19
sublist3	17	15	15	13	13	13	18	15
sublist4	20	17	14	15	13	14	16	20
sublist5	16	20	19	14	17	16	18	25
total	88	91	74	64	64	71	83	103
% correct	49.4	51.1	41.6	36.0	36.0	38.9	46.6	57.9
Soft	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	23	17	21	15	22	22	18	25
sublist2	19	17	15	10	16	19	21	15
sublist3	19	17	17	11	25	16	23	26
sublist4	21	21	15	9	19	11	23	20
sublist5	17	16	22	14	20	16	25	25
total	99	88	90	59	102	84	110	111
% correct	55.6	49.4	50.6	33.1	57.3	47.2	61.8	62.4
Talker EK, List *normal_rate2*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	11	10	15	9	7	10	13	9
sublist2	12	13	13	7	15	14	14	15
sublist3	19	18	18	10	16	12	15	15
sublist4	15	16	15	8	12	16	18	18
sublist5	15	8	14	8	10	12	14	15
total	72	65	75	42	60	64	74	72
% correct	41.9	37.8	43.6	24.4	34.9	37.2	43.0	41.9
Loud	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	17	13	15	11	11	14	16	22
sublist2	19	21	17	13	18	13	19	25
sublist3	25	19	22	8	17	14	21	21
sublist4	18	18	18	13	13	20	20	22
sublist5	17	15	23	13	21	13	19	24
total	96	86	95	58	80	74	95	114
% correct	54.9	49.1	54.3	33.1	45.7	42.3	54.3	65.1

Talker EK, List *normal_rate3*							
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	17	13	12	11	22	14	19	17
sublist2	14	22	15	8	17	20	17	24
sublist3	15	26	18	18	22	18	17	22
sublist4	16	20	12	10	20	14	15	19
sublist5	21	17	22	22	23	22	19	25
total	83	98	79	69	104	88	87	107
% correct	48.0	56.6	45.7	39.9	60.1	50.9	50.3	61.8
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	10	10	21	12	17	17	17	17
sublist2	10	8	15	6	16	17	14	21
sublist3	12	18	11	10	20	15	14	17
sublist4	14	17	15	11	17	14	16	22
sublist5	18	19	19	10	18	18	19	24
total	64	72	81	49	88	81	80	101
% correct	37.6	42.4	47.6	28.8	51.8	47.6	47.1	59.4

Talker EK, List *normal_rate4*							
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	14	10	11	9	10	13	17	15
sublist2	9	10	12	14	12	14	18	17
sublist3	13	12	13	12	14	14	21	17
sublist4	21	15	19	11	13	18	15	20
sublist5	24	15	11	16	13	12	19	18
total	81	62	66	62	62	71	90	87
% correct	45.5	34.8	37.1	34.8	34.8	39.9	50.6	48.9
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	20	22	16	19	16	22	18	20
sublist2	20	22	21	21	27	25	20	22
sublist3	26	27	26	19	24	26	27	25
sublist4	22	23	16	20	20	24	19	23
sublist5	17	19	15	23	16	18	21	24
total	105	113	94	102	103	115	105	114
% correct	59.0	63.5	52.8	57.3	57.9	64.6	59.0	64.0

Talker EK, List *quick_rate1*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	19	12	17	10	16	16	13	18
sublist2	12	16	14	7	13	8	12	19
sublist3	15	14	11	8	19	9	13	17
sublist4	18	10	12	7	18	9	19	18
sublist5	24	20	16	17	19	12	18	24
total	88	72	70	49	85	54	75	96
% correct	50.9	41.6	40.5	28.3	49.1	31.2	43.4	55.5
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	15	15	19	13	17	20	18	19
sublist2	15	20	20	13	17	21	20	22
sublist3	12	20	16	13	14	17	21	20
sublist4	17	21	21	15	18	14	17	24
sublist5	26	25	30	17	24	19	24	27
total	85	101	106	71	90	91	100	112
% correct	49.1	58.4	61.3	41.0	52.0	52.6	57.8	64.7

Talker EK, List *slow_rate1*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	15	23	16	25	11	15	18	27
sublist2	17	23	16	22	17	14	13	25
sublist3	11	20	11	15	17	14	11	25
sublist4	18	21	16	22	20	13	15	23
sublist5	17	19	16	19	16	13	16	21
total	78	106	75	103	81	69	73	121
% correct	45.6	62.0	43.9	60.2	47.4	40.4	42.7	70.8
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	23	21	24	21	27	22	22	25
sublist2	21	29	25	20	25	20	24	27
sublist3	23	21	21	16	19	19	21	23
sublist4	26	25	22	18	22	26	21	25
sublist5	26	24	25	21	20	24	23	24
total	119	120	117	96	113	111	111	124
% correct	69.6	70.2	68.4	56.1	66.1	65.0	65.0	72.5

Talker EK, List *slow_rate2*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	23	20	21	20	23	18	22	24
sublist2	20	20	21	19	18	21	18	21
sublist3	22	23	26	20	26	24	25	25
sublist4	21	25	24	21	29	27	24	25
sublist5	22	19	20	16	27	21	23	22
total	108	107	112	96	123	111	112	117
% correct	62.4	61.8	64.7	55.5	71.1	64.2	64.7	67.6
ASR	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	27	21	20	21	20	23	14	22
sublist2	23	18	22	15	24	24	20	22
sublist3	21	24	26	18	24	25	23	24
sublist4	26	18	19	24	27	19	22	26
sublist5	23	17	20	15	22	20	20	26
total	120	98	107	93	117	111	99	120
% correct	69.4	56.6	61.8	53.8	67.6	64.2	57.2	69.4

Table C.2: Raw and percent correct key-word scores for T1.

Talker JM, List *normal_rate1*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	21	19	23	18	17	25	16	21
sublist2	20	23	15	13	18	18	22	22
sublist3	20	23	21	19	22	24	22	24
sublist4	22	23	26	21	23	21	23	24
sublist5	21	22	21	16	23	21	23	22
total	104	110	106	87	103	109	106	113
% correct	62.7	66.3	63.9	52.4	62.0	65.7	63.9	68.1
Soft	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	3	3	3	2	2	2	4	7
sublist2	2	1	4	1	1	2	0	2
sublist3	4	2	1	3	2	3	2	2
sublist4	1	2	1	1	3	1	3	1
sublist5	1	1	1	0	0	0	3	2
total	11	9	10	7	8	8	12	14
% correct	6.6	5.4	6.0	4.2	4.8	4.8	7.2	8.4

Talker JM, List *normal_rate2*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	20	18	18	14	19	14	20	18
sublist2	22	15	16	12	18	18	21	20
sublist3	17	22	14	15	14	15	15	20
sublist4	22	24	24	19	23	19	22	23
sublist5	26	24	26	23	25	24	27	28
total	107	103	98	83	99	90	105	109
% correct	62.2	59.9	57.0	48.3	57.6	52.3	61.0	63.4
Loud	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	17	15	15	14	17	17	18	24
sublist2	20	14	17	11	19	12	20	15
sublist3	22	18	15	13	20	17	17	21
sublist4	28	29	29	20	25	23	25	26
sublist5	21	30	26	19	28	23	24	25
total	108	106	102	77	109	92	104	111
% correct	62.8	61.6	59.3	44.8	63.4	53.5	60.5	64.5

Talker JM, List *normal_rate3*							
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	25	29	20	26	30	27	24	31
sublist2	21	25	22	24	23	17	21	23
sublist3	20	21	26	18	23	20	20	28
sublist4	28	27	24	17	29	21	23	30
sublist5	23	19	20	15	27	19	20	26
total	117	121	112	100	132	104	108	138
% correct	66.9	69.1	64	57.1	75.4	59.4	61.7	78.9
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	26	26	22	22	25	25	20	27
sublist2	20	23	29	23	19	24	21	27
sublist3	25	25	22	24	27	23	25	29
sublist4	24	27	20	22	25	24	24	30
sublist5	23	24	20	19	23	22	20	25
total	118	125	113	110	119	118	110	138
% correct	67.4	71.4	64.6	62.9	68	67.4	62.9	78.9

Talker JM, List *normal_rate4*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	26	23	22	21	21	21	25	18
sublist2	26	25	25	23	20	24	24	20
sublist3	23	21	17	19	21	21	22	23
sublist4	16	18	15	17	18	17	22	17
sublist5	20	17	17	14	23	20	17	22
total	111	104	96	94	103	103	110	100
% correct	62.4	58.4	53.9	52.8	57.9	57.9	61.8	56.2
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	32	27	30	24	25	26	29	24
sublist2	28	24	26	21	27	27	26	30
sublist3	28	25	22	26	23	25	23	23
sublist4	30	26	23	19	28	25	24	29
sublist5	26	24	19	20	23	24	25	27
total	144	126	120	110	126	127	127	133
% correct	80.9	70.8	67.4	61.8	70.8	71.3	71.3	74.7

Talker JM, List *quick_rate1*							
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	18	22	20	18	20	11	26	25
sublist2	18	25	19	18	16	11	25	21
sublist3	20	16	18	20	24	14	26	22
sublist4	26	26	22	16	21	14	23	24
sublist5	23	16	14	17	22	15	29	25
total	105	105	93	89	103	65	129	117
% correct	59.7	59.7	52.8	50.6	58.5	36.9	73.3	66.5
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	24	22	26	20	27	24	26	30
sublist2	21	22	21	16	23	17	17	23
sublist3	22	21	23	12	24	20	22	25
sublist4	24	22	21	19	25	18	22	27
sublist5	21	18	20	12	20	22	25	28
total	112	105	111	79	119	101	112	133
% correct	63.6	59.7	63.1	44.9	67.6	57.4	63.6	75.6

Talker JM, List *slow_rate1*							
Conv	CA	AC	AG	FK	AM	JP	JS	MS
	21	20	18	20	21	23	21	24
sublist2	27	22	21	18	25	19	23	23
sublist3	22	20	18	21	23	23	22	27
sublist4	20	24	21	18	25	20	25	27
sublist5	22	29	16	26	29	27	30	29
total	112	115	94	103	123	112	121	130
% correct	63.3	65.0	53.1	58.2	69.5	63.3	68.4	73.4
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	21	25	23	23	22	21	23	25
sublist2	21	28	24	20	25	20	22	26
sublist3	27	27	24	20	23	21	27	33
sublist4	27	25	21	14	24	21	24	31
sublist5	29	31	24	22	30	25	29	30
total	125	136	116	99	124	108	125	145
% correct	71.8	78.2	66.7	56.9	71.3	62.1	71.8	83.3

Talker JM, List *slow_rate2*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	23	25	28	26	25	27	30	31
sublist2	21	21	23	20	26	21	25	24
sublist3	26	21	24	23	24	22	22	26
sublist4	25	20	20	21	25	17	24	25
sublist5	20	19	19	23	24	25	20	25
total	115	106	114	113	124	112	121	131
% correct	66.5	61.3	65.9	65.3	71.7	64.7	69.9	75.7
ASR	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	20	24	25	17	22	21	18	22
sublist2	27	28	25	17	26	26	20	26
sublist3	20	22	19	16	20	20	15	24
sublist4	23	19	17	18	19	21	14	24
sublist5	25	19	26	16	19	18	23	26
total	115	112	112	84	106	106	90	122
% correct	66.5	64.7	64.7	48.5	61.3	61.3	52.0	70.5

Table C.3: Raw and percent correct key-word scores for T3.

Talker MI, List *normal_rate1*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	7	11	4	5	8	11	11	9
sublist2	11	11	11	9	15	9	9	11
sublist3	11	5	6	2	10	4	8	9
sublist4	4	7	4	2	11	3	10	14
sublist5	7	14	11	5	14	9	12	11
total	40	48	36	23	58	36	50	54
% correct	23.3	27.9	20.9	13.4	33.7	20.9	29.1	31.4
Soft	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	16	8	11	5	14	10	14	14
sublist2	16	17	13	7	17	12	14	18
sublist3	14	15	17	4	16	10	13	17
sublist4	8	10	7	5	14	11	9	18
sublist5	19	18	16	9	21	11	14	20
total	73	68	64	30	82	54	64	87
% correct	42.4	39.5	37.2	17.4	47.7	31.4	37.2	50.6

Talker MI, List *normal_rate2*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	17	11	13	9	15	12	7	22
sublist2	11	14	11	13	10	6	12	17
sublist3	16	18	14	11	9	11	9	19
sublist4	15	13	15	16	16	12	16	22
sublist5	18	12	16	15	17	14	21	22
total	77	68	69	64	67	55	65	102
% correct	44.3	39.1	39.7	36.8	38.5	31.6	37.4	58.6
Loud	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	14	15	12	8	12	8	13	12
sublist2	9	14	11	12	15	10	10	12
sublist3	16	9	12	11	14	11	12	15
sublist4	18	19	15	8	16	15	21	13
sublist5	9	10	7	13	7	11	13	13
total	66	67	57	52	64	55	69	65
% correct	37.9	38.5	32.8	29.9	36.8	31.6	39.7	37.4

Talker MI, List *normal_rate3*							
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	8	8	8	8	13	9	8	12
sublist2	13	9	10	7	10	8	8	9
sublist3	9	8	4	5	12	4	7	15
sublist4	9	14	10	7	7	9	14	19
sublist5	11	10	11	10	14	15	7	17
total	50	49	43	37	56	45	44	72
% correct	28.9	28.3	24.9	21.4	32.4	26.0	25.4	41.6
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	19	16	17	12	20	13	20	22
sublist2	20	19	17	17	21	20	22	24
sublist3	15	19	18	17	21	14	17	20
sublist4	18	15	16	9	21	19	23	22
sublist5	12	16	19	13	22	12	14	24
total	84	85	87	68	105	78	96	112
% correct	48.6	49.1	50.3	39.3	60.7	45.1	55.5	64.7

Talker MI, List *normal_rate4*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	10	7	6	13	5	6	6	16
sublist2	6	16	9	10	6	7	6	17
sublist3	9	9	9	6	5	6	11	19
sublist4	13	10	5	6	5	6	8	13
sublist5	8	9	5	10	8	7	9	10
total	46	51	34	45	29	32	40	75
% correct	27.9	30.9	20.6	27.3	17.6	19.4	24.2	45.5
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	11	16	14	10	10	11	11	14
sublist2	17	12	13	12	7	18	12	18
sublist3	10	10	9	8	4	4	10	9
sublist4	11	14	13	10	8	10	11	16
sublist5	17	12	13	7	9	10	10	14
total	66	64	62	47	38	53	54	71
% correct	38.6	37.4	36.3	27.5	22.2	31.0	31.6	41.5

Talker MI, List *quick_rate1*							
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	1	3	4	3	6	6	1	6
sublist2	6	8	7	2	5	4	1	8
sublist3	3	4	4	4	7	2	3	9
sublist4	4	3	2	2	5	3	2	5
sublist5	6	2	5	4	6	4	2	9
total	20	20	22	15	29	19	9	37
% correct	11.6	11.6	12.7	8.7	16.8	11.0	5.2	21.4
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	23	14	18	14	17	10	17	20
sublist2	19	17	12	8	14	8	17	17
sublist3	18	16	15	12	17	9	18	21
sublist4	18	16	15	7	9	5	14	21
sublist5	20	15	12	11	13	10	19	18
total	98	78	72	52	70	42	85	97
% correct	56.6	45.1	41.6	30.1	40.5	24.3	49.1	56.1

Talker MI, List *slow_rate1*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	18	18	12	16	23	15	10	24
sublist2	13	22	13	16	19	10	12	23
sublist3	14	23	15	19	20	13	15	21
sublist4	10	22	13	19	23	15	17	23
sublist5	13	20	9	21	17	8	13	24
total	68	105	62	91	102	61	67	115
% correct	39.1	60.3	35.6	52.3	58.6	35.1	39.0	66.1
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	10	7	10	11	6	9	10	12
sublist2	11	19	14	7	10	15	7	14
sublist3	15	17	14	13	9	13	16	16
sublist4	11	16	11	6	8	11	12	15
sublist5	20	14	19	9	16	19	18	22
total	67	73	68	46	49	67	63	79
% correct	38.5	42.0	39.1	26.4	28.2	38.5	36.2	45.4

Talker MI, List *slow_rate2*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	25	23	22	26	22	25	21	20
sublist2	27	24	22	21	28	19	17	24
sublist3	29	22	23	25	21	25	21	28
sublist4	22	22	21	14	23	22	20	27
sublist5	24	23	22	21	25	21	20	23
total	127	114	110	107	119	112	99	122
% correct	76.0	68.3	65.9	64.1	71.3	67.1	59.3	73.1
ASR	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	21	18	26	19	31	22	21	28
sublist2	18	15	18	17	22	21	19	26
sublist3	24	18	14	19	22	19	19	21
sublist4	20	16	22	15	23	24	17	23
sublist5	20	20	17	18	20	20	18	23
total	103	87	97	88	118	106	94	121
% correct	61.7	52.1	58.1	52.7	70.7	63.5	56.3	72.4

Table C.4: Raw and percent correct key-word scores for T4.

Talker RG, List *normal_rate1*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	21	19	18	14	19	18	20	22
sublist2	16	15	10	11	11	12	12	12
sublist3	17	19	17	8	19	15	21	21
sublist4	17	18	15	15	21	17	11	18
sublist5	16	15	16	14	20	21	21	21
total	87	86	76	62	90	83	85	94
% correct	49.4	48.9	43.2	35.2	51.1	47.2	48.3	53.4
Soft	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	13	9	11	6	12	13	7	9
sublist2	15	5	9	4	13	7	11	11
sublist3	15	9	9	6	14	7	12	13
sublist4	15	16	16	6	19	12	14	14
sublist5	12	14	10	7	15	12	12	11
total	70	53	55	29	73	51	56	58
% correct	39.8	30.1	31.3	16.5	41.5	29.0	31.8	33.0

Talker RG, List *normal_rate2*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	19	13	18	12	16	20	19	21
sublist2	17	18	12	11	12	12	11	15
sublist3	20	19	15	14	20	13	14	25
sublist4	19	18	19	15	16	12	16	23
sublist5	21	18	18	12	20	13	14	22
total	96	86	82	64	84	70	74	106
% correct	56.8	50.9	48.5	37.9	49.7	41.4	43.8	62.7
Loud	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	24	21	16	12	14	16	24	16
sublist2	22	21	21	15	23	22	20	23
sublist3	20	19	17	15	17	14	20	16
sublist4	22	18	20	16	18	19	21	23
sublist5	20	16	21	15	21	19	17	19
total	108	95	95	73	93	90	102	97
% correct	63.9	56.2	56.2	43.2	55.0	53.3	60.4	57.4

Talker RG, List *normal_rate3*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	15	24	16	12	22	15	17	19
sublist2	18	24	18	9	23	18	21	22
sublist3	20	12	18	7	13	16	16	13
sublist4	8	9	13	10	15	16	14	15
sublist5	15	11	15	11	20	15	17	21
total	76	80	80	49	93	80	85	90
% correct	44.4	46.8	46.8	28.7	54.4	46.8	49.7	52.6
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	20	14	20	17	21	20	18	26
sublist2	18	21	20	16	21	19	16	27
sublist3	20	23	20	14	19	24	22	21
sublist4	20	18	14	12	17	18	21	25
sublist5	17	17	20	15	22	23	19	22
total	95	93	94	74	100	104	96	121
% correct	55.5	54.4	55.0	43.3	58.5	60.8	56.1	70.8

Talker RG, List *normal_rate4*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	16	16	18	12	11	11	15	20
sublist2	11	15	16	9	15	10	17	20
sublist3	17	19	19	13	17	12	19	21
sublist4	19	15	16	15	16	10	14	15
sublist5	13	15	10	11	9	9	9	20
total	76	80	79	60	68	52	74	96
% correct	43.7	46.0	45.4	34.5	39.1	29.9	42.5	55.2
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	29	21	21	22	22	25	23	28
sublist2	28	25	25	23	22	28	24	22
sublist3	28	26	27	26	18	25	26	26
sublist4	26	24	26	17	19	23	21	23
sublist5	18	25	23	25	17	18	22	26
total	129	121	122	113	98	119	116	125
% correct	74.1	69.5	70.1	64.9	56.3	68.4	66.7	71.8

Talker RG, List *quick_rate1*							
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	6	9	5	3	3	2	3	7
sublist2	6	5	3	2	3	1	4	6
sublist3	4	5	2	3	4	4	4	4
sublist4	3	6	4	1	4	3	3	6
sublist5	6	7	2	1	2	1	4	3
total	25	32	16	10	16	11	18	26
% correct	14.2	18.2	9.1	5.7	9.1	6.3	10.2	14.8
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	16	20	13	7	15	10	17	16
sublist2	15	17	8	7	12	7	13	13
sublist3	10	12	15	8	13	8	7	19
sublist4	14	15	9	14	12	9	15	18
sublist5	9	18	13	9	7	8	16	13
total	64	82	58	45	59	42	68	79
% correct	36.4	46.6	33.0	25.6	33.5	23.9	38.6	44.9

Talker RG, List *slow_rate1*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	11	16	7	10	17	8	11	22
sublist2	16	17	14	17	23	11	18	24
sublist3	12	18	13	19	20	13	13	20
sublist4	13	13	8	11	15	6	7	24
sublist5	17	16	16	13	17	13	12	22
total	69	80	58	70	92	51	61	112
% correct	40.1	46.5	33.7	40.7	53.5	29.7	35.5	65.1
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	28	25	20	20	24	26	28	27
sublist2	22	30	26	26	28	26	26	32
sublist3	24	17	25	22	22	23	19	27
sublist4	20	24	26	22	25	25	29	27
sublist5	20	21	21	21	19	20	22	27
total	114	117	118	111	118	120	124	140
% correct	66.3	68.0	68.6	64.5	68.6	69.8	72.1	81.4

Talker RG, List *slow_rate2*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	18	20	21	14	29	20	26	26
sublist2	14	19	17	22	19	19	15	18
sublist3	17	16	19	17	22	17	19	21
sublist4	20	20	22	23	21	24	18	21
sublist5	18	19	17	21	24	19	22	22
total	87	94	96	97	115	99	100	108
% correct	49.2	53.1	54.2	54.8	65.0	55.9	56.5	61.0
ASR	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	11	17	19	14	18	13	13	19
sublist2	15	15	21	17	16	16	17	15
sublist3	11	16	19	17	21	14	16	21
sublist4	15	24	25	23	24	25	20	24
sublist5	18	17	18	18	19	19	19	18
total	70	89	102	89	98	87	85	97
% correct	39.5	50.3	57.6	50.3	55.4	49.2	48.0	54.8

Table C.5: Raw and percent correct key-word scores for T5.

Talker SA, List *normal_rate1*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	15	15	21	7	17	12	16	21
sublist2	12	13	10	7	19	6	21	19
sublist3	18	18	13	9	18	14	13	20
sublist4	17	19	16	10	16	14	21	26
sublist5	15	17	19	11	23	16	23	21
total	77	82	79	44	93	62	94	107
% correct	44	46.9	45.1	25.1	53.1	35.4	53.7	61.1
Soft	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	0	0	0	0	0	0	0	0
sublist2	0	0	0	0	0	0	0	0
sublist3	0	0	0	0	0	0	0	0
sublist4	0	0	0	0	0	0	0	0
sublist5	0	0	0	0	0	0	0	0
total	0	0	0	0	0	0	0	0
% correct	0	0	0	0	0	0	0	0

Talker SA, List *normal_rate2*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	10	14	14	11	14	10	12	16
sublist2	15	18	12	14	11	16	18	19
sublist3	14	18	10	8	11	13	16	18
sublist4	15	18	9	13	12	14	14	21
sublist5	16	22	13	14	14	13	16	22
total	70	90	58	60	62	66	76	96
% correct	39.3	50.6	32.6	33.7	34.8	37.1	42.7	53.9
Loud	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	23	22	24	18	19	27	25	24
sublist2	27	20	20	20	24	20	26	26
sublist3	22	21	19	15	22	19	24	23
sublist4	23	26	27	22	21	27	29	29
sublist5	25	29	27	17	25	27	28	25
total	120	118	117	92	111	120	132	127
% correct	67.4	66.3	65.7	51.7	62.4	67.4	74.2	71.3

Talker SA, List *normal_rate3*							
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	20	23	15	10	22	17	23	20
sublist2	13	14	12	10	15	23	13	20
sublist3	18	18	15	12	21	19	22	22
sublist4	22	13	19	17	21	20	23	26
sublist5	18	15	18	12	19	23	18	20
total	91	83	79	61	98	102	99	108
% correct	52.3	47.7	45.4	35.2	56.3	58.6	56.9	62.1
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	25	27	25	23	28	22	27	32
sublist2	27	24	21	20	26	20	23	27
sublist3	22	24	23	20	24	24	25	27
sublist4	28	28	33	26	29	23	32	31
sublist5	25	23	31	19	25	23	29	28
total	127	126	133	108	132	112	136	145
% correct	73.0	72.4	76.4	62.1	75.9	64.9	78.2	83.3

Talker SA, List *normal_rate4*							
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	18	16	6	6	10	6	18	25
sublist2	15	18	13	16	11	12	16	23
sublist3	13	15	11	10	11	17	19	19
sublist4	11	22	14	8	11	11	16	24
sublist5	10	19	10	13	13	11	19	19
total	67	90	54	53	56	57	88	110
% correct	37.2	50	30	29.4	31.1	31.7	48.9	61.1
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	26	28	26	22	25	24	22	26
sublist2	22	26	28	25	27	24	26	26
sublist3	26	31	31	26	29	29	32	30
sublist4	27	29	27	25	31	31	27	27
sublist5	32	30	29	23	27	30	31	31
total	133	144	141	121	139	138	138	140
% correct	73.9	80	78.3	67.2	77.2	76.7	76.7	77.8

Talker SA, List *quick_rate1*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	1	4	1	3	4	6	3	4
sublist2	2	4	6	5	5	5	2	9
sublist3	4	5	5	2	1	5	1	7
sublist4	2	5	3	2	7	2	4	11
sublist5	5	4	3	3	3	1	3	12
total	14	22	18	15	20	19	13	43
% correct	7.9	12.4	10.2	8.5	11.3	10.7	7.3	24.3
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	10	13	6	6	14	12	9	11
sublist2	17	15	13	9	10	10	14	14
sublist3	17	16	14	9	18	17	15	17
sublist4	14	10	14	11	18	13	16	14
sublist5	11	17	14	10	16	11	10	19
total	69	71	61	45	76	63	64	75
% correct	39.0	40.1	34.5	25.4	42.9	35.6	36.2	42.4

Talker SA, List *slow_rate1*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	18	23	10	16	17	18	16	22
sublist2	14	18	8	16	18	11	12	17
sublist3	21	20	12	17	22	18	17	24
sublist4	13	17	13	16	17	16	15	23
sublist5	9	19	5	18	19	10	12	22
total	75	97	48	83	93	73	72	108
% correct	42.9	55.4	27.4	47.4	53.1	41.7	41.1	61.7
Clear	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	29	32	29	29	31	30	28	31
sublist2	23	26	25	24	29	28	23	26
sublist3	25	29	29	20	27	26	27	28
sublist4	22	27	27	23	29	30	25	27
sublist5	28	27	31	26	29	32	32	30
total	127	141	141	122	145	146	135	142
% correct	72.6	80.6	80.6	69.7	82.9	83.4	77.1	81.1

Talker SA, List *slow_rate2*								
Conv	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	24	23	20	14	19	22	22	19
sublist2	14	16	12	10	14	11	15	15
sublist3	19	19	19	17	22	16	22	19
sublist4	16	13	15	10	18	9	20	16
sublist5	14	18	19	13	19	18	17	19
total	87	89	85	64	92	76	96	88
% correct	50.6	51.7	49.4	37.2	53.5	44.2	55.8	51.2
ASR	CA	AC	AG	FK	AM	JP	JS	MS
sublist1	18	26	23	18	20	20	18	28
sublist2	11	9	10	8	9	17	15	24
sublist3	20	17	21	16	20	20	26	23
sublist4	18	15	12	8	18	17	14	20
sublist5	22	23	17	15	23	19	21	20
total	89	90	83	65	90	93	94	115
% correct	51.7	52.3	48.3	37.8	52.3	54.1	54.7	66.9