

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 23 (1999)
Indiana University

**Eliciting Speech Reduction in the Laboratory II:
Calibrating Cognitive Loads for Individual Talkers¹**

James D. Harnsberger and David B. Pisoni

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

¹ This work was supported by NIH-NIDCD Training Grant DC00012 and NIH-NIDCD Research Grant R01-DC00111 to Indiana University. We would like to thank Jim Brink for his technical assistance and Richard Wright for his help and advice.

Eliciting Speech Reduction in the Laboratory II: Calibrating the Cognitive Load for Individual Talkers

Abstract. This experiment extended work done previously in our laboratory to develop a method to elicit from talkers three different speaking styles, reduced, citation, and hyperarticulated, using controlled materials in a laboratory setting. In the initial experiment, the reduced style was elicited by having subjects read a sentence while carrying out a distractor task that involved recalling a fixed number of digits from short-term memory. The original experiment was clearly limited in its success at eliciting a reduced style of speech: Only one of the six talkers showed significant differences between reduced and citation speech based on an acoustic analysis of the sentences. In this study, we chose to calibrate the distractor memory task to an individual's short-term memory span as measured by a simple digit span task. That is, the number of digits to be recalled after reading aloud a test sentence was determined by each individual's digit span. Twelve talkers were recorded in this experiment. The results showed that six of the twelve talkers produced a reduced style of speech for the test sentences in the distractor task relative to the same sentences in the citation style condition, as determined by a phonetically-trained judge. This initial evaluation was confirmed in a perceptual test using a pairwise comparison task in which normal-hearing, untrained listeners were presented with two sentences varying in speech style and were asked to choose the most carefully pronounced sentence. The results showed that 71% - 88% of the sentence sets tested were correctly differentiated by speech style, indicating that the individual calibration method was a substantial improvement over the elicitation method of the original experiment.

Introduction

Traditionally in studies of speech production and perception that use natural speech, utterances are recorded under highly controlled conditions in a laboratory setting. Control over the recording conditions and the nature of the materials recorded (particular syllables, words, sentences) serves to limit sources of error in the data collection process, or to avoid particular confounds that might render the results uninterpretable. Control over the quality and structure of the materials also insures that an experiment can be replicated in other laboratories, a key aspect of any experiment. However, it has long been recognized that the style of speaking elicited from talkers reading linguistic material aloud in a laboratory setting differs systematically from more reduced styles of speech that can be observed in unmonitored conversations outside of the laboratory (Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988; Picheny, Durlach, & Braida, 1989; Byrd, 1994). These differences can include the duration of the utterance and its constituent words, pausing, and the degree of centralization in the quality of vowels, to name a few. Such differences pose a problem for theories of speech perception and spoken word recognition, most of which were formulated from studies using controlled speech materials: To what extent do these findings generalize to the speech styles that people produce and perceive in a natural setting? The perception of variability that exists among speech styles has not been studied in detail, no doubt due to the problem of eliciting naturalistic speech in the decidedly unnatural manner and setting of reading aloud in a laboratory. Other types of "nonlinguistic" variability have proven to have an effect on speech perception and spoken word recognition, including talker, rate, and stimulus variability (Mullennix & Pisoni, 1990; Nygaard, Sommers, & Pisoni, 1995; Bradlow, Nygaard, & Pisoni, 1999). These studies suggest that listeners encode in long-term memory significant details and properties of speech signals that they encounter, and that these details

influence the subsequent perception and recognition of speech. If listeners are sensitive to detailed, episodic properties of speech, then variation in those properties due to speaking style differences may also play an important role in speech processing, one that has thus far been neglected in speech perception and spoken word recognition research.

One factor limiting the study of the perception of speech styles has been a methodological one: how does a researcher elicit different speech styles, including more reduced and naturalistic speech, while controlling for the particular syllables, words, or sentences to be studied? Recording natural conversation, or guided conversation on a particular topic, has been used in the study of sociolinguistic variability in speech production, namely, in the elicitation of stigmatized, or less prestigious, sounds, words, or syntactic structures of a dialect (Labov, 1972; Milroy, 1987). In other methods, subjects have been asked to participate in and narrate a task (Hirschberg & Nakatani, 1996). Such procedures have been useful in eliciting particular intonational forms and in studying particular aspects of discourse structure (Swerts & Collier, 1992; Speer, Sokol, & Schafer, 1999). However, none of these methods can guarantee the elicitation of specific sentences.

In our laboratory, we have attempted to develop methods for eliciting sentences in different speech styles in the laboratory while controlling for the particular sentence materials used. The range of speech styles being studied includes a reduced, or hypoarticulated style, that should more closely resemble the speech style employed in natural settings in conversation than laboratory read speech. The first version of this method was described by Brink, Wright, and Pisoni (1998). Brink et al. (1998) attempted to elicit three speaking styles, namely, reduced, or hypoarticulated speech; citation, or read speech, a style that is normally used in reading controlled materials in a laboratory setting; and hyperarticulated speech. Each style was elicited in a separate condition of the experiment.

Brink et al. (1998) attempted to elicit reduced speech by having subjects read a sentence while engaging in a concurrent task, specifically, remembering a digit sequence of five to seven digits in length that was presented immediately prior to the sentence. After reading the sentence, subjects were asked to recall the digit sequence in the same order in which it was presented. The digit span task was a distractor task, chosen to place the subject under a cognitive load while reading a sentence. The digit span task was chosen as the distractor task because, in piloting, it was successful in producing the desired speech style while minimizing talker disfluencies. Citation speech was elicited by simply having listeners read single sentences presented on a computer screen. Hyperarticulated speech was elicited in an experimental condition quite similar to the Citation speech condition. Subjects were asked to read single sentences presented on a computer screen. Over the course of the condition, they were prompted in a subset of trials to repeat the sentence “more clearly.” After responding to that prompt, subjects were given the same prompt a second time, and the second reading was chosen to represent hyperarticulated speech. This procedure had been used successfully in an earlier study by Johnson, Flemming, and Wright (1993).

Brink et al. (1998) tested this method with six talkers, all native speakers of English, and evaluated its success in a detailed acoustic analysis, examining properties of the sentence, as well as key words in the sentence, in terms of duration, f_0 range, absolute RMS energy, energy range, degree of vowel centralization, and degree of vowel dispersion. The results of the acoustic analysis showed that the method was successful in eliciting a hyperarticulated speech style that was highly distinct from the citation style, a result that was found for all six talkers. The duration, vowel centralization, and vowel dispersion measures showed the most consistent differences across talkers in speech production. However, the method failed to elicit significant differences between the reduced and citation sentences for five of the six talkers. Only one talker, MD, produced reduced speech that was distinguishable from citation speech in the acoustic

analysis.² Interestingly, MD also had the highest error rate for correctly recalling digit sequences during the Reduced condition, indicating that the fixed load of five to seven digits was sufficiently challenging for MD. For the other five talkers, digit sequences of a length of five to seven may not have been sufficiently demanding for them, particularly given that the average immediate memory span for digits is 7.7 (Cavanagh, 1972). Individual differences in digit span may complicate the use of a fixed range of digit sequence lengths since, for a given individual, the degree of cognitive load that a subject is placed under may either be too great or little, depending on the subject's own digit span. If the load is too great, then the possibility exists that a subject will simply ignore the distractor task, making the Reduced condition in essence the same as the Citation condition. If the distractor task is too easy, a subject may be able to allocate sufficient attentional resources to the task of reading the sentence, a result that also effectively eliminates any differences between reduced and citation speech. One solution to the problem of using fixed sequence lengths would be to measure an individual's digit span and use that value to "calibrate" the Reduced condition, making the task difficult enough to draw attentional resources away from reading, but not so difficult so as to be ignored or induce disfluencies.

The goal of this study was to test an individually-calibrated cognitive load method of eliciting reduced speech. In the experiments reported here, the cognitive load task was calibrated to the individual talker's memory span in a forward digit span task administered prior to the Reduced condition. A talker's digit span, as measured in this task, was the value used to determine the range in the length of digit sequences in the Reduced condition of this study. The Citation and Hyperarticulation conditions of Brink et al. (1998) remained unchanged in this study. With the addition of individual calibration, we predicted that all of the talkers recorded would produce a reduced speech style that was perceptually distinct from the citation and hyperarticulated speech styles. Twelve talkers were recorded in this revised procedure for the Reduced condition, and in the original Citation and Hyperarticulation conditions. The elicited sentences were then evaluated by a phonetically-trained judge and by 25 untrained listeners to determine if the individual calibration method was a success.

EXPERIMENT 1

Methods

Participants

Twelve native speakers of American English, seven females and five males ranging in age between 18 and 30, participated in this study. Participants received \$15 total compensation for participating in two one-hour sessions. None of the subjects reported any history of speech or hearing disorders at the time of testing.

Stimulus Materials

The participants read 34 sentences from the 200 sentences comprising the SPIN set (Kalikow, Stevens, & Elliot, 1977). The SPIN sentences are short sentences, five to eight words in length, ending in a high frequency monosyllabic noun. The 34 SPIN sentences selected for this study are listed in Appendix A. The recording took place in a sound-attenuated chamber (IAC Audiometric Testing Room, Model 402)

² Speaker MD's reduced sentences were also perceptually distinguishable from his/her citation sentences in a pilot Paired Comparison task with three native speakers of English. These native speakers successfully picked the citation sentences as "more carefully pronounced" in reduced-citation sentence pairs, on 89% of test trials. For a detailed description of the Paired Comparison task, see Experiment 2 for a study using the same methodology.

using a head-mounted Shure (SM98) microphone positioned one inch away from the subject's chin. The recordings were digitized at 22.05 kHz (16 bit sampling) using a Tucker-Davis Technologies System II and stored on an IBM-PC 486 computer.

Procedures

The participants carried out four tasks over two test sessions. In the first session, participants were administered a simple forward digit span task (see Digit Span Task) and were recorded reading sentences in the Reduced condition. In the second session, which took place within seven days of the first session, participants were recorded reading sentences in the Citation and Hyperarticulation conditions.

Digit Span Task. In the digit span task, participants were presented with a sequence of single digits (0 - 9) on a computer screen inside of the sound-attenuated chamber, and asked to recall the sequence correctly in the order in which it was presented. The participants' responses were digitized and played via headphones to the experimenter, who sat outside of the booth and scored the responses. The responses themselves were not stored to disk as sound files. The length of the digit sequence that was presented started at four, and then increased or decreased via an adaptive staircase algorithm (Levitt, 1971). The algorithm increased the sequence length by one digit for every two sequences at a given length that were successfully recalled by the participant. Whenever the participant responded to a sequence incorrectly, the sequence length was reduced by one digit on the following trial. Over the course of the 25 trials of the task, the sequence length for individual participants increased until the sequence length began eliciting errors. Thus, by the end of the task, participants were “oscillating” between the sequence length that they could consistently recall, and a longer sequence that induced errors. The longest sequence length that was consistently recalled was taken to be the participant's digit span. This value was then used to calibrate the cognitive load in the Reduced condition.

Reduced Condition. The Reduced condition was similar to the Reduced condition described by Brink et al. (1998), and consisted of 136 trials, four trials for each of the 34 SPIN sentences, with a 1 s inter-trial interval (ITI). The order of the blocks of four trials varied randomly for each participant. Each trial consisted of four parts: initially, participants were presented with a digit sequence, which remained on the screen for 2 s; then, after a 2.5 s interval, a sentence was displayed on the computer screen for the participant to read; next, the participant's response was recorded over a 6 s window; finally, participants were prompted to recall the digit sequence in the correct order. The length of the digit sequence was based on the participant's digit span as measured in the Digit Span Task. The length of the digit sequence in a given trial was either the same as the span score, or plus/minus one digit. For example, if a participant had a span of seven in the digit span task, he/she would be presented with digit sequences ranging in length from six to eight. The same sentence, embedded in the digit span task, was presented four times, with the fourth reading taken as the reduced sentence for subsequent analysis. Before the recording began for the Reduced condition, participants were told that they would be participating in a short-term memory experiment. Participants were instructed to focus on the digit span task in the Reduced condition, in the hope that they would be less careful in monitoring their production of the test sentences.

Citation and Hyperarticulation Conditions. The Citation and Hyperarticulation conditions were identical to those described earlier by Brink et al. (1998). In the Citation condition, participants were prompted to read aloud a sentence that appeared on the computer screen. Each sentence was presented once, for a total of 34 trials, with a 1 s ITI. The order in which the sentences were presented was randomized for each participant.

The Hyperarticulation condition was similar to the Citation condition, and consisted of two types of trials. The first trial type, the “citation cycle,” was identical to a Citation condition trial. In the second trial type, the “hyperarticulation cycle,” participants were also prompted to read aloud a sentence appearing on the computer screen. After reading this sentence, participants were then prompted to “Please read the sentence more clearly.” After responding, they were asked again to read the sentence more clearly. Thus, for the hyperarticulation cycle, the same sentence was read three times, with the third reading taken to be the example of the “hyperarticulated” reading of the sentence for subsequent analysis. The 34 sentences each appeared in three citation cycles and one hyperarticulation cycle. The program controlling the experiment was designed to insure that the Hyperarticulation condition began with a citation cycle, and that hyperarticulation cycles were separated by at least two citation cycles.

Results and Discussion

In an earlier test of this method of eliciting different speaking styles (Brink et al., 1998), the elicited sentences were initially evaluated using a detailed acoustic analysis of those cues that have been commonly cited in prior work as important ones in differentiating speech styles. The cues measured by Brink et al. included the duration, RMS energy, and energy range of each sentence and of three “key” words in each sentence (usually the subject, verb, and object of each sentence); the sentence f0 range; the degree of vowel centralization of vowels in key words, and the degree of vowel dispersion of vowels in key words. The results of this lengthy acoustic analysis revealed no significant differences between the reduced and citation sentences of five of the six talkers recorded. Given this failure and the fact that the present experiment represents an “exploratory” phase in the development of a method of eliciting speech styles in the laboratory, a less time-consuming method of evaluating the results was chosen, namely, an impressionistic evaluation of the sentences by a single phonetically-trained judge. The results of this evaluation appear in Table 1. Each percentage in the three sentence pair columns represents the percentage of sentence pairs judged to be qualitatively different in terms of speaking style³.

Subject	Reduced-Citation	Reduced-Hyperarticulated	Citation-Hyperarticulated
1	67%	100%	100%
2	88%	100%	100%
3	91%	100%	100%
4	6%	100%	100%
5	24%	100%	100%
6	41%	100%	100%
7	18%	100%	100%
8	41%	100%	100%
9	76%	100%	100%
10	41%	100%	100%
11	50%	100%	100%
12	82%	100%	100%

Table 1: The percentage of sentence pairs judged to be qualitatively different in speech style.

³ The number of sentence pairs that each percentage represented varied slightly due to the fact that individual subjects occasionally produced disfluencies in their readings of a particular sentence. Sentence pairs were excluded from evaluation in cases in which one of the sentences involved a disfluency or disfluencies.

An examination of the impressionistic results shows that the sentence pairs that included hyperarticulated sentences were clearly differentiable, just as they were in the acoustic analysis of Brink et al. (1998). The critical sentence pairs for this study were the Reduced-Citation pairs, given the failure of the earlier method of Brink et al. to elicit measurable differences between these two styles. In the individual calibration method of this study, only half of the participants (1, 2, 3, 9, 11, and 12) produced qualitative differences in 50% or more of their hypoarticulated and citation sentence pairs. The percentage of pairs judged to be different varied widely by individual talker, from as low as 6% to as high as 91%. Thus, the method of individually calibrating the cognitive load of the Reduced condition was effective for a subset of the talkers tested. This result is clearly an improvement over the method of Brink et al. that successfully elicited reduced sentences from only one participant out of six.

EXPERIMENT 2

Methods

Participants

Twenty-five native speakers of American English, seventeen females and eight males ranging in age between 18 and 21, participated in this study. For participating in a single one-hour session, the participants received either \$7.50 or one credit towards their research requirement if they were enrolled in an undergraduate psychology class. None of the subjects reported any history of speech or hearing disorders at the time of testing.

Stimulus Materials

The stimulus materials consisted of 26 to 34⁴ hypoarticulated, citation, and hyperarticulated sentences from the four talkers, namely subjects 2, 3, 9, and 12 from Experiment 1, whose reduced-citation sentence pairs were most frequently judged to be qualitatively different in Experiment 1.

Procedures

A trial in the Paired Comparison Task consisted of two different readings of each sentence from each talker. The two readings were presented in pairs, with a 1 s interstimulus interval. Participants were asked to choose which sentence was read more carefully by using a mouse to press on one of two buttons on a computer screen, denoting the first or the last sentence of the pair. The sentence pairs differed only in terms of the speaking style in which they were produced, resulting in three types of pairs: reduced-citation, reduced-hyperarticulated, and citation-hyperarticulated. The sentence pairs always involved the same sentence produced by the same talker. An example of a Paired Comparison trial would be a reduced “The farmer harvested the crop,” produced by Talker 2, paired with a hyperarticulated “The farmer harvested the crop,” also produced by Talker 2. The 25 participants were divided into four groups of three to eight participants each. Each group listened to the sentence pairs of a single talker. The sentence pairs were presented in both orders (i.e., citation-hyperarticulated and hyperarticulated-citation). Thus, each participant responded to 156 to 204 trials, depending on which talker they were randomly assigned to.

⁴ The full set of 34 sentences were not used for each talker because, in a limited number of cases, some talkers produced sentences with disfluencies.

Results

The results of the Paired Comparison Task appear in Table 2, which lists the percentage of sentences judged correctly for the three types of sentence pairs for each talker. In this table, “Sentences” refers to the number of different SPIN sentences from each talker, while “Listeners” refers to the number of participants that judged the sentence pairs of a given talker. As expected, listeners correctly chose the hyperarticulated sentence as the one that was read “more carefully” in a high percentage of sentence pairs, 85% - 100% of trials, indicating that the method was successful in eliciting citation and hyperarticulated sentences. In the critical test pairs, the Reduced-Citation pairs, percent correct scores were the same or slightly lower than those of the phonetically-trained judge in Experiment 1, ranging between 71% and 88%. Overall, the percent correct scores in the Paired Comparison Task showed a significant correlation with the corresponding percentages in Table 1 ($r = 0.77$, $p \leq 0.05$), confirming the judgments of the phonetically-trained listener.

Talker	Sentences	Listeners	Reduced-Citation	Reduced-Hyperarticulated	Citation-Hyperarticulated
2	31	8	88%	100%	99%
3	34	7	78%	93%	85%
9	30	7	71%	86%	96%
12	26	3	73%	96%	98%

Table 2: The percentage of sentence pairs judged correctly by the untrained listeners.

General Discussion and Conclusions

The results of Experiments 1 and 2 showed that the individual calibration method of eliciting reduced speech, in conjunction with the Citation and Hyperarticulated conditions, was successful in producing three distinct speech styles from half of the subjects, in a large majority of those subjects' sentences (71 - 88%). However, the results did not match the predicted success rate because half of the subjects did not produce reduced sentences that were perceptually distinct from their citation sentences. One possible reason for this failure may be the restrictive criteria used for selecting reduced and citation sentences. The individual participants' success rates were based on comparisons of one example, out of two to four elicited, of each sentence in each speech style. For instance, the fourth sentence in a block of four in the Reduced condition was taken to be the “reduced” example of that sentence. Each sentence also has three citation readings, one from the Citation condition and two from the Hyperarticulated condition. This strategy of sentence selection may have unduly biased the outcome by restricting the capacity of the method to elicit different speech styles. The other readings of each sentence in each style condition may have been equally good, or better, representatives of each style. For instance, instead of comparing only the fourth reading of a sentence in the Reduced condition to the first reading of a sentence in the Citation condition, the third Reduced sentence could be compared to the second Citation sentence, to see if discernible differences exist between the two. A preliminary examination of all of the readings of each style of each sentence by the phonetically-trained judge from Experiment 1 indicates that in the case of one talker, Participant 6, the percentage of his/her perceptually distinct Reduced-Citation pairs increased to the range of talkers 1, 2, 3, 9, 11, and 12 when other readings are considered (see Table 1). Thus, relaxing the

criteria of which sentence readings are chosen to represent reduced- and citation-style sentences may give a more accurate analysis of the relative success of the method.

Even using relaxed criteria, however, one third of the subjects still failed to produce a Reduced-Citation distinction. Moreover, the success rate across sentences for participants that were judged as producers of the distinction was less than perfect. Apparently, other factors are at work in determining whether or not a sufficient cognitive load will induce reduction in the sentences of a given participant. One possible factor may be differences among individuals in terms of their performance on the Digit Span Task versus their ability to correctly recall digits in the Reduced condition. Individuals may differ in their capacity to recall digits while engaging in a distractor task that would not be reflected in their performance on a simple digit span task. Such individual differences would be reflected in the percentage of digit sequences correctly recalled in the Reduced condition. Table 3 lists these percentages for the twelve participants, who are grouped in terms of their digit span.

Participant	% Correct	Span
10	62	5
12	38	6
11	60	6
1	68	6
2	75	6
5	78	6
6	21	7
7	41	7
3	42	7
8	53	7
9	26	8
4	32	8

Table 3. The percentage of trials in which digit sequences were correctly recalled in the reading portion of the Reduced condition.

As Table 3 shows, participants who had performed the same on the simple digit span task were not equivalent in their ability to recall digit sequences in the Reduced condition. Participants with a span of six ranged from 38% - 78% correct, averaging 67% correct. Participants with spans of 7 also varied widely, and participants with spans of 7 and 8 generally did worse in recalling digit sequences with a distractor sentence than participants with spans of 5 or 6. The implications of these differences lie in the issue of cognitive load and its capacity to induce reduction in sentences. The calibration of the cognitive load is crucial to the method's success. If the cognitive load is too great, participants may ignore the load task and simply read the sentence in citation style. If the load is not sufficiently demanding of attentional resources, then participants may find it too easy to both read the distractor sentence and recall a digit sequence, rendering the distractor sentence a citation, rather than a reduced, sentence. In this experiment, some participants may have been highly successful in recalling digit sequences in the simple digit span task. However, when asked to recall digit sequences in the context of an intervening reading task, their effective span may have been much lower than that indicated in the simple digit span task, making the calibrated load too difficult to induce reduction. Thus, participants may differ in terms of their capacity to cope with a distractor task in recalling digit sequences. If this is indeed a possible source of error in the experimental

method, one solution may be to redesign the Digit Span Task to match the Reduced condition. In future work, we intend to test such a revised method, with a greater number of subjects representing each span length (i.e., 5, 6, 7, and 8) to attempt to calculate the necessary correlations to support this hypothesis.

Of course, it is possible that small changes in the Reduced condition will still fail to produce the desired reduced-citation style difference for all talkers. This method may not represent the optimal solution to the problem of eliciting reduced speech in a controlled experiment. Ultimately, the success of the method outlined here, or any variant of it, must be judged not only in simple perceptual experiments, such as the Paired Comparison task, which simply indicates that different speech styles were elicited on a “carefulness” continuum. The acoustic properties of the elicited sentences must also be measured, and their differences correlated with those differences measured in studies of speech styles in natural speech. Only if the reduced sentences elicited by the method described here display the properties of naturally-occurring reduced speech can the method be judged a success. Thus, acoustic analysis, following successful perception tests, should form the basis of method evaluation in our subsequent work.

In summary, our study successfully tested a revised method to elicit three different speaking styles, reduced, citation, and hyperarticulated, by calibrating the task for eliciting reduced speech to the individual participant’s short-term memory span. Relative to the original experiment, the calibration method was a substantial improvement, eliciting reduced sentences in 71 – 88% of test trials from half of the talkers recorded as determined by a phonetically-trained judge as well as a group of naïve listeners. Using more relaxed criteria for selecting “reduced” sentences out of the set of four recorded from each talker, the success rate of the revised method increased to seven of the twelve talkers tested. Of course, this new scoring method still resulted in five talkers from whom reduced sentences were not successfully elicited. To improve on the current method, a new individual calibration technique was proposed, involving a digit sequence distractor task that is continuously calibrated over the course of the Reduced condition.

References

- Bradlow, A.R., Nygaard, L.C., & Pisoni, D.B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception and Psychophysics*, *61*, 206-219.
- Brink, J., Wright, R., & Pisoni, D.B. (1998). Eliciting speech reduction in the laboratory: Assessment of a new experimental method. In *Research on Spoken Language Processing Progress Report No. 22* (pp. 396-420). Bloomington, IN: Speech Research Laboratory, Indiana University.
- Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, *15*, 39-54.
- Cavanagh, J.B. (1972). Relation between the immediate memory span and the memory search rate. *Psychological Review*, *79*, 525-530.
- Hirschberg, J. & Nakatani, C.H. (1996). A prosodic analysis of discourse segments in direction-giving monologues. ACL-96.
- Kalikow, D.N., Stevens, K.N., & Elliot, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, *61*, 1337-1351.

- Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, *69*, 505-528.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*, 467-477.
- Milroy, L. (1987). *Observing and analyzing natural language*. Oxford: Basil Blackwell.
- Mullennix, J.W. & Pisoni, D.B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics*, *61*, 206-219.
- Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception and Psychophysics*, *57*, 989-1001.
- Picheny, M.A., Durlach, N.I., & Braida, L.D. (1989). 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.
- Speer, S.R., Sokol, S.B., & Schafer, A.J. (1999). Prosodic disambiguation of syntactic ambiguity in discourse context. *Journal of the Acoustical Society of America*, *106*, 2275.
- Swerts, M. & Collier, R. (1992). On the controlled elicitation of spontaneous speech. *Speech Communication*, *11*, 463-468.
- Summers, W., Pisoni, D.B., Bernacki, R.H., Pedlow, R.I., & Stokes, M.A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *Journal of the Acoustical Society of America*, *84*, 917-928.

Appendix 1: Stimulus Sentences

The farmer harvested his crop.
 His boss made him work like a slave.
 He caught the fish in his net.
 Close the window to stop the draft.
 The beer drinkers raised their mugs.
 I made the phone call from a booth.
 The cut on his knee formed a scab.
 The railroad train ran off the track.
 They drank a whole bottle of gin.
 The airplane dropped a bomb.
 I gave her a kiss and a hug.
 The soup was served in a bowl.
 The cookies were kept in a jar.

How did your car get that dent?
 The baby slept in his crib.
 The cop wore a bullet-proof vest.
 No one was injured in the crash.
 The hockey player scored a goal.
 How long can you hold your breath?
 At breakfast he drank some juice.
 The king wore a golden crown.
 He got drunk in the local bar.
 The doctor prescribed the drug.
 The landlord raised the rent.
 Playing checkers can be fun.
 Throw out all this useless junk.

Her entry should win first prize.
The stale bread was covered with mold.
I ate a piece of chocolate fudge.
The story had a clever plot.

He's employed by a large firm.
The mouse was caught in the trap.
I've got a cold and a sore throat.
The judge is sitting on the bench.

