



Published in final edited form as:

*Speech Commun.* 2008 April 1; 50(4): 323–336. doi:10.1016/j.specom.2007.11.001.

## A new method for eliciting three speaking styles in the laboratory

James D. Harnsberger<sup>1</sup>, Richard Wright<sup>2</sup>, and David B. Pisoni

Speech Research Laboratory, Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana 47405

### Abstract

In this study, a method was developed to elicit three different speaking styles, reduced, citation, and hyperarticulated, using controlled sentence materials in a laboratory setting. In the first set of experiments, the reduced style was elicited by having twelve talkers read a sentence while carrying out a distractor task that involved recalling from short-term memory an individually-calibrated number of digits. The citation style corresponded to read speech in the laboratory. The hyperarticulated style was elicited by prompting talkers (twice) to reread the sentences more carefully. The results of perceptual tests with naïve listeners and an acoustic analysis 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. In addition, all talkers consistently produced sentences in the citation and hyperarticulated styles. In the second set of experiments, the reduced style was elicited by increasing the number of digits in the distractor task by one (a heavier cognitive load). The procedures for eliciting citation and hyperarticulated sentences remained unchanged. Ten talkers were recorded in the second experiment. The results showed that six out of ten talkers differentiated all three styles as predicted (70% of all sentences recorded). In addition, all talkers consistently produced sentences in the citation and hyperarticulated styles. Overall, the results demonstrate that it is possible to elicit controlled sentence stimulus materials varying in speaking style in a laboratory setting, although the method requires further refinement to elicit these styles more consistently from individual participants.

### Keywords

Speaking styles; Speech perception; Speaking rate; Vowel dispersion

## 1. 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 audio 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 any results 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

<sup>1</sup>Now at the University of Florida, Gainesville, FL.

<sup>2</sup>Now at the University of Washington, Seattle, WA.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

other speaking styles that occur naturally, such as more reduced styles of speech that can be observed in unmonitored conversations (Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988; Picheny, Durlach, & Braidá, 1989; Byrd, 1994) and hyperarticulated or clear speech, whether directed to normal and hearing-impaired listeners (Picheny, Durlach, & Braidá, 1985; Picheny, Durlach, & Braidá, 1986; Picheny, Durlach, & Braidá, 1989; Payton, Uchanski, & Braidá, 1994; Uchanski, Choi, Braidá, & Durlach, 1996), infants (Fernald, Taeschner, Dunn, Papousek, de Boysson-Bardies, & Fukui, 1989; Kuhl, Andruski, Chistovich, Chistovich, Kozhevnikova, Ryskina, Stolyarova, Sundberg, & Lacerda, 1997), non-native speakers (Bradlow & Bent, 2002; Uther, Knoll, & Burnham, 2007), or automatic speech recognizers (Oviatt, MacEachern, & Levow, 1998; Oviatt, Levow, Moreton, & MacEachern, 1998). Differences between reduced and laboratory citation speaking styles 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.

While citation speech is fairly uniform in its characteristics from word to word and repetition, conversational speech is typified by a wide range of hyperarticulation and reduction at the word and syllable levels (Krull, 1989; Duez, 1992, Wassink, Wright, & Franklin, 2007). The probability of any one word being reduced (*hypoarticulated*) or hyperarticulated is widely thought to be related to its informational load (i.e., Lindblom, 1996; Jurafsky, *et al*, 2001; Aylett and Turk, 2004); the less important the word is to a particular utterance's meaning and the more predicable it is from context, the more hypoarticulated it typically is. Likewise, if a word is particularly important to the meaning of an utterance and if it is not predictable from the preceding set of words, it is likely to be hyperarticulated. Thus, the use of conversational speech in controlled experiments poses another problem to the experimenter: It is difficult to control the degree of reduction or hyperarticulation on any particular word. Like reduced speech, hyperarticulated speech displays features that distinguish it from citation speech, including vowel space expansion, a decrease in speaking rate, increase in the number and duration of pauses, and increase in energy in the mid-frequency range of the spectrum, and a decrease in disfluencies, among others. Such hyperarticulated speech effects have been observed primarily in American English, but a few studies observe similar patterns in other languages (e.g., Croatian – Smiljanic & Bradlow, 2005; German – Köster, 2001; Jamaican – Wassink, Wright, & Franklin, 2007). It is a style adaptively employed by speakers to enhance recognition and comprehension by the listener given the listening environment at hand (Lindblom, 1990; Moon & Lindblom, 1994; Lindblom, 1996; Jurafsky, *et al*, 2001; Aylett & Turk, 2004). It also occurs frequently in human-computer interaction when automatic speech recognizers produce errors; speakers will shift to a hyperarticulated speaking style which, while helpful in human speech communication, can actually limit the success of human-computer interaction given the automatic speech recognizers are not commonly trained on this style (Hirschberg, Litman, & Swerts, 2004).

The differences between natural speaking styles and speech elicited in laboratory reading tasks pose a problem for theories of speech perception and spoken word recognition by human listeners, most of which have been formulated from studies using controlled speech materials: To what extent do these findings generalize to the speaking styles that people produce and perceive in natural settings? 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 the setting of reading aloud in a laboratory, although these issues have begun to be addressed in computer/machine word recognition (Ostendorf, Byrne, Bacchiani, Finke, Gunawardana, Ross, Roweis, Shriberg, Talkin, Waibel, Wheatley, & Zeppenfeld, 1996; Shriberg, 2001; Liu, Shriberg, Stolcke, Hillard, Ostendorf, & Harper, 2006; Bates, Ostendorf, & Wright, 2007). Other types of “nonlinguistic” variability have been shown 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 episodic 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 highly 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 speaking styles has been a methodological one: How does a researcher elicit different speaking styles, including more reduced and naturalistic speech, while at the same time 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 specific aspects of discourse structure (Swerts & Collier, 1992; Speer, Sokol, & Schafer, 1999). However, none of these methods can guarantee the elicitation of specific sentences. What is needed is a technique to elicit a variety of speaking styles found outside the laboratory, particularly more hypoarticulated styles, while controlling for linguistic content. Speech samples elicited by such a method would represent an important point on the continuum between experimental control and ecological validity.

To address this research need, a method has been developed for eliciting sentences in different speech styles in the laboratory while controlling for the particular sentence materials used. The range of speech styles 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) elicited reduced speech by having subjects read a sentence while engaging in a concurrent digit span task that involved 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. It was chosen as the distractor task because, in piloting, it was successful in producing the desired speech style while minimizing talker disfluencies. The logic behind the use of any distractor task to elicit a reduced speaking style was based on two sets of observations in the literature. First, multiple studies have documented limited resources in working memory and, with an increase in cognitive load, fewer processing resources are allocated to the motor tasks involved in producing speech (Baddeley, Thomson, & Buchanan, 1975; Oberauer & Kliegl, 2006). It was expected that the decrease in processing resources would result in decreased gestural displacement of the speech articulators as is normally seen in hypoarticulated speech. Second, it is a common observation in the sociolinguistic literature that the self-consciousness of the speaker results in a formality of speech style (Labov, 1984). The digit span task was conceived of as load that could be optimized to produce reduction without a dramatic increase in speech errors and therefore minimize the high degree of self-monitoring, and the resulting formal style that typically occurs in word reading tasks.

Citation speech was elicited by simply having listeners read single sentences presented on a computer screen. Hyperarticulated speech was collected 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, fundamental frequency (f<sub>0</sub>) 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.<sup>3</sup> 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 potentially 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 the current 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 immediate 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, it was 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 (Experiment 1), by 25 untrained listeners (Experiment 2), and in an acoustic analysis (Experiment 3) to determine if the individual calibration method was successful. Three other experiments (Experiments 4 – 6) were also conducted, involving a further modification

---

<sup>3</sup>Speaker MD's reduced sentences were also perceptually distinguishable from his/her citation sentences in a 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.

to cognitive load method of eliciting reduced speech, namely, the use of a more challenging memory span task that was still individually calibrated. Experiments 4, 5, and 6 parallel the design of the first three experiments in the use of phonetically trained judges, naïve listeners, and acoustic analysis to evaluate this particular method of eliciting a variety of speaking styles in the laboratory.

## 2. Experiment 1

### 2.1 Methods

**2.1.1 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.

**2.1.2 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) 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.

**2.1.3 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.

**2.1.3.1 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 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.

**2.1.3.2 Reduced condition:** The Reduced condition was similar to the same condition described by Brink et al. (1998), and consisted of 136 trials, four trials for each of the 34 SPIN sentences, with a 1 second (s) intertrial interval (ITI). 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. The trial order varied randomly for each participant.

**2.1.3.3 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.

Following the Citation condition, the Hyperarticulation condition was presented 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 by a sentence printed on the computer screen that read: "Please read the sentence more clearly." After responding, they were prompted again to read the sentence more clearly. Only the final reading was taken to be the example of the "hyperarticulated" reading of the sentence for subsequent analysis. Over the course of the Hyperarticulation condition, 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.

## 2.2 Results and Discussion

In an earlier test of this method of eliciting different speaking styles (Brink et al., 1998), the recorded 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 revealed no significant differences between the reduced and citation sentences of five of the six talkers recorded. Given this initial failure and the fact that this experiment represented 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 classification of these materials by the judge was conducted without any information with regards to the target (that is, presumed) speaking style. 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 style<sup>4</sup>.

<sup>4</sup>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 reduced 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 most 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.

The apparent success in Experiment 1 of the individually-calibrated load method for eliciting a reduced speaking style required further evaluation since only impressionistic judgments were used from a single phonetically-trained listener. In Experiment 2, perceptual judgments of the elicited stimuli were collected from a group of naïve listeners in a paired comparison task to affirm the results of Experiment 1.

### 3. Experiment 2

#### 3.1 Methods

**3.1.1 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.

**3.1.2 Stimulus Materials**—The stimulus materials consisted of 26 to 34<sup>5</sup> 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.

**3.1.3 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 (ISI). 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.

---

<sup>5</sup>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.

### 3.2 Results and Discussion

The results of the Paired Comparison Task are shown 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 ( $r = 0.77, p < 0.05$ ), confirming the judgments of the phonetically-trained listener.

To provide further evidence that the new method was successful in eliciting the three intended speaking styles, the stimulus materials recorded in Experiment 1 were also acoustically analyzed. The particular acoustic measures taken were a subset of those used by Brink et al., ones that were the most successful in differentiating the three speaking styles elicited from the single talker who produced a consistent reduced-citation style contrast. The measures included the duration of the sentences, the duration of sentence keywords, and the relative size of the talker's vowel space. The latter measure was based on measurements of the first and second formants of vowels in keywords in the sentences.

## 4. Experiment 3 – Acoustic analysis

### 4.1 Methods

**4.1.1 Participants and Stimulus Materials**—The stimulus materials consisted of 26 to 34 reduced, citation, and hyperarticulated sentences from the twelve talkers recorded in Experiment 1.

**4.1.2 Procedures**—The recorded sentences were analyzed acoustically for the duration of the sentences as well as three to four keywords. All of the keywords were content words and commonly appeared in one of three positions within the sentences: (1) near the beginning (usually the participant noun), (2) near the middle (usually the main verb) and (3) in the final position (usually the main object of the verb or of a preposition). Duration was measured directly from the waveforms with accompanying wide band spectrograms for reference using Cool Edit 2000 software.

The keywords in each sentence, in all three styles, were also analyzed acoustically for *vowel dispersion*, defined as the average Euclidian distance in Barks of keyword vowels from the center of an individual's vowel space (Bradlow, Torretta, and Pisoni, 1996). Vowel formant measures were made from an overlaid LPC-FFT display. The LPC employed 12-16 coefficients (based on the participant) and a 25 ms frame size. The FFT used a 1024-point window. A wide-band spectrogram was used for reference. The formant measures were made at the point of maximal displacement of F1 and F2 (as described in Wright, 2003). The results of the acoustic analyses were used to examine the differences between the reduced, citation, and hyperarticulated styles of individual participants.

### 4.2 Results

**4.2.1 Duration measures**—Figures 1 and 2 display the mean keyword and sentence durations for each participant, respectively. The mean durations varied between 250 – 750 ms, with the longest durations used in the hyperarticulated speaking style. A comparison of the



results across participants permitted a common observation: keyword duration changed the most for the hyperarticulated style, while the shift from reduced to citation involved much more modest lengthening (or even none at all). This observation can easily be extended to the sentence duration data as well.

The mean duration measures for each participant were submitted to separate 3 (Style: Reduced, Citation, Hyperarticulated)  $\times$  2 (Unit of Analysis: Keyword, Sentence) repeated measures ANOVAs. For every participant, there were significant main effects of Style and Unit of Analysis, as well as a significant Style by Unit of Analysis interaction. Appendix B lists the results of the statistical analysis by participant. Post hoc analyses (Tukey t-tests) showed that seven participants produced positive, significant differences between the reduced and citation styles in whole sentences (specifically, participants #1, 2, 5, 8, 10, 11, 12). In contrast, only one participant (#2) produced positive, significant differences between the two styles in keyword duration. Both the sentences and keywords read in the hyperarticulated style differed significantly from those read in the reduced and citation styles for every participant, as predicted. The differences in duration between the sentences read in the hyperarticulated style and those in the reduced and citation styles were much greater in magnitude than the differences in duration between the reduced and citation styles. Overall, seven out of twelve participants differentiated the three styles by manipulating some aspect of the temporal properties of the sentence.

**4.2.2 Vowel dispersion**—The vowel dispersion measures for each participant and style are given in Figure 3. The vowels produced in different styles generally varied in dispersion in the manner predicted, that is, in their distance from the center of the vowel space. Thus, vowels produced in a reduced speech style resulted in modestly smaller vowel spaces than those produced in a citation style. Furthermore, vowels produced in a hyperarticulated style resulted in more expanded vowel spaces. An example of a vowel space differing in degree of dispersion appears in Figure 4, which shows Participant 2's vowel spaces computed from the keyword vowels in each style. Figure 4 demonstrates that, as participants articulate in speaking styles that increase in articulatory precision (i.e., from reduced to hyperarticulated, in order of increasing precision), the corresponding vowel spaces expand.

To compare differences in vowel space size for individual talker's stimuli, *style difference scores* for individual vowels were calculated. Style difference scores are the difference in Euclidean distance from the center of the vowel space to the center of an individual vowel for the same talker in two different styles. The center of the vowel space corresponded to the mean F1 and F2 values across all of a talker's vowels in a given speaking style. The center of an individual vowel corresponded to the mean F1 and F2 values across all tokens of that vowel. For each talker, 27 style difference scores were calculated (9 vowels, 3 pairs of styles). An example of an individual style difference score would be the reduced-citation score for the /i/ vowel from talker 1. The equation for calculating this example difference score appears in (1),

$$\left( \sqrt{(x_{ic} - x_{cc})^2 + (y_{ic} - y_{cc})^2} \right) - \left( \sqrt{(x_{ir} - x_{cr})^2 + (y_{ir} - y_{cr})^2} \right) \quad (1)$$

where,  $x_{ic}$  and  $y_{ic}$  are the mean F1 and F2, respectively, for the talker's /i/ vowel in the citation style;  $x_{cc}$  and  $y_{cc}$  are the mean F1 and F2, respectively, across all vowel tokens from talker 1 in the citation style;  $x_{ir}$  and  $y_{ir}$  are the mean F1 and F2, respectively, for the talker's /i/ vowel in the reduced style; and  $x_{cr}$  and  $y_{cr}$  are the mean F1 and F2, respectively, across all vowel tokens from talker 1 in the reduced style.

To determine if the influence of speaking style involved significant changes in vowel dispersion for individual talkers, the least squares means of these style difference scores were compared to zero in post hoc tests of a mixed model ANOVA with Participant and Style as factors. In this analysis, Participant ( $F(11,96) = 2.4, p < 0.01$ ), Style ( $F(2,16) = 23.9, p < 0.05$ ), and their interaction ( $F(22,176) = 2.0, p < 0.01$ ) were all significant. For all but one of the participants, the difference scores involving the hyperarticulated style were significantly different. In contrast, the comparison between the reduced and citation styles was only significant for two of the twelve participants (i.e., participants 1 and 2). Overall, vowel space dispersion was primarily a factor only in the hyperarticulated speaking style. It appeared to play less of a role in the perception of all three speaking styles for even those speakers who produced distinguishable reduced-citation pairs based on the results of the first two experiments (e.g., participants 9 and 12).

### 4.3 Discussion

The goal of Experiment 3 was to use acoustic analysis techniques to evaluate the success of the calibrated cognitive load method in eliciting three distinct speech styles. The acoustic analysis, particularly the duration measures, showed that the revised procedure was successful in eliciting reduced speech from a majority of the talkers, although large individual differences were observed. With just a fixed cognitive load, only one of six participants produced reliable differences between the reduced and citation styles based on an extensive acoustic analysis of their utterances (Brink et al., 1998). Thus, the results of Experiments 1, 2, and 3 suggest that individually calibrating the cognitive load for the individual participant results in a more consistent elicitation procedure for a reduced style of speech.

While a success rate of seven out of twelve participants represents a marked improvement over the earlier results reported by Brink et al. (1998), the time and effort required by this procedure to elicit the style differences for just 34 different sentences necessitated changes in the experimental procedure to reduce the range of individual differences. One possible improvement would be to increase the cognitive load that the distractor task (i.e., the digit span task) imposes. In Experiment 1, the distractor task was calibrated to the individual's digit span. Individual trials had digit sequences equal to the individual span, one digit longer, or one digit shorter. Thus, in many trials, listeners had to recall a relatively short digit sequence, or one matched to the individual's span. Only in a subset of trials was the digit sequence longer than the individual's measured span. A heavier cognitive load was considered to be potentially a more effective distractor task to elicit reduced speech. Specifically, the cognitive load could be increased by one digit more than the participant's individual digit span, to insure that the task is sufficiently demanding for the listener as they produce the sentences. Experiments 4, 5, and 6 tested the use of a heavier cognitive load in the Reduced condition. As in the first three experiments, the efficacy of the elicitation method was evaluated by a phonetically-trained judge (Experiment 4), by a group of naïve listeners (Experiment 5), and, finally, using acoustic analysis techniques (Experiment 6).

## 5. Experiment 4

### 5.1 Methods

**5.1.1 Participants**—Ten native speakers of American English, seven females and three males ranging in age from 18 to 20, 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.

**5.1.2 Stimulus Materials**—The stimulus materials were the same as those used in Experiment 1 (section 2).

**5.1.3 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.

**5.1.3.1 Digit span task:** The digit span task was the same as the one used in Experiment 1 (section 2).

**5.1.3.2 Reduced condition:** The Reduced condition was identical to the Reduced condition of Experiment 1 (section 2), with one exception. In Experiment 1, the length of the digit sequence that served as the distractor task in Reduced condition was the same as the span score in the digit span task (plus or minus one digit). In Experiment 4, the digit sequence length in the Reduced condition was one digit plus the span score of the digit span task. Individual trials were calibrated with this heavier load, plus or minus one digit. For example, if a subject had a span of 7 in the digit span task, the digit sequences in the Reduced condition were calibrated to a span score of 8. Thus, the digit sequences of individual trials could be 7, 8, or 9 digits in length.

**5.1.3.3 Citation and hyperarticulation conditions:** The Citation and Hyperarticulation conditions were identical to those described in Experiment 1 (section 2).

## 5.2 Results and Discussion

As in Experiment 1, the elicited sentences were evaluated impressionistically by a single phonetically-trained judge. The results of this evaluation appear in Table 3. Each percentage in the three sentence pair columns represents the percentage of sentence pairs judged to be qualitatively different in terms of speaking style<sup>6</sup>. An examination of the impressionistic results shows that the sentence pairs that included hyperarticulated sentences were as differentiable as they were in Experiment 1. For the reduced-citation pairs, eight of the ten participants (#13, 15, 16, 17, 18, 20, 21, and 22) produced qualitative differences in 50% or more of their hypoarticulated and citation sentence pairs. The percentage of pairs judged to be different still varied by individual talker, from as low as 21% to as high as 100%. Across all of the participants, 69% of the reduced-citation pairs differed in the predicted manner, as compared with 52% of the reduced-citation pairs elicited in Experiment 1. Thus, the heavy cognitive load of the Reduced condition was effective for most of the talkers tested.

## 6. Experiment 5

### 6.1 Methods

**6.1.1 Participants**—Twenty-seven native speakers of American English, twenty females and seven males ranging in age between 18 and 26, 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.

**6.1.2 Stimulus Materials**—The stimulus materials consisted of 26 to 34 hypoarticulated, citation, and hyperarticulated sentences from the four talkers, namely subjects 16, 17, 20, and

<sup>6</sup>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.

21 from Experiment 4, whose reduced-citation sentence pairs were frequently judged to be qualitatively different in Experiment 4.

**6.1.3 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 ISI. 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. The 27 participants were divided into four groups of four 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.

## 6.2 Results

The results of the Paired Comparison Task are shown in Table 4, 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, 72% - 100% of trials, indicating that the method was again successful in eliciting citation and hyperarticulated sentences. In the reduced-citation pairs, percent correct scores were similar to those of the phonetically-trained judge in Experiment 4, ranging between 73% and 96%. Overall, the percent correct scores in the Paired Comparison Task showed a significant correlation with the corresponding percentages of the phonetically-trained listener shown in Table 4 ( $r = 0.66, p < 0.05$ ).

## 7. Experiment 6 – Acoustic analysis

### 7.1 Methods

**7.1.1 Participants and Stimulus Materials**—The stimulus materials consisted of reduced, citation, and hyperarticulated sentences from the ten talkers recorded in Experiment 4.

**7.1.2 Procedures**—The procedures used in Experiment 6 were the same as those in Experiment 3. The stimulus materials were measured in terms of sentence duration, keyword duration, and vowel space size.

### 7.2 Results and Discussion

**7.2.1 Duration measures**—Figures 5 and 6 display the mean keyword and sentence durations for each participant, respectively. The mean keyword durations were comparable to those observed in Experiment 3, although citation style keywords were longer (by about 10%) in this experiment. The sentence durations were also similar to those observed in the first twelve participants, although larger differences were observed with these speakers between their reduced and citation sentences. However, the most robust effects of speaking style still appear in the shift to the hyperarticulated style.

The mean duration measures for each participant were submitted to separate 3 (Style: Reduced, Citation, Hyperarticulated)  $\times$  2 (Unit of Analysis: Keyword, Sentence) repeated measures ANOVAs. For every participant, there were significant main effects of Style and Unit of

Analysis, as well as a significant Style by Unit of Analysis interaction. Appendix C lists the results of the statistical analysis by participant. Post hoc analyses (Tukey t-tests) showed that six participants (13, 16, 17, 18, 20, 21) produced positive, significant differences between the reduced and citation styles in whole sentences, while only two participants (17, 21) produced positive differences between the two styles in keyword duration. Both the sentences and keywords read in the hyperarticulated style differed significantly from those read in the reduced and citation styles for every participant, as predicted.

**7.2.2 Vowel dispersion**—Figure 7 shows the differences in vowel dispersion for each individual participant between the three styles, with greater dispersion corresponding to a larger vowel space. Reduced, citation, and hyperarticulated (Hyper) styles were predicted to differ in increasing order in degree of vowel dispersion. Figure 7 demonstrates that, as participants articulate in speaking styles that increase in articulatory precision (i.e., from reduced to hyperarticulated, in order of increasing precision), the corresponding vowel spaces expand accordingly/respectively.

The vowel dispersion data for this modified elicitation method was analyzed in the same manner as in section 4.2.2.2. That is, style difference scores were computed and their least squares means were compared to zero in post hoc tests of a mixed model ANOVA, with Participant and Style as factors. In this analysis, Participant ( $F(9,80) = 5.6, p < 0.01$ ), Style ( $F(2,16) = 13.4, p < 0.01$ ), and their interaction ( $F(18,144) = 3.9, p < 0.01$ ) were all significant. In examining individual participant scores, only two speakers showed significant differences across all three styles (participant 17 and 21). Three speakers showed no significant differences whatsoever (18, 19, 22). More commonly, the vowel space of the hyperarticulated style was larger than either just the reduced style (participant 20) or both reduced and citation (participants 13, 14, 15, and 16). In summary, fewer participants showed significant differences in vowel dispersion than in sentence duration.

## 8. General Discussion and Conclusions

The results of six experiments 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 in up to 80% of subjects, in a large majority of those subjects' sentences. This elicitation method appeared to primarily affect overall speaking rate: Differences in style were typically observed in sentence durations rather than keyword durations or vowel dispersion. This pattern indicates that subjects responded to this laboratory procedure by shortening their function words and interword pauses when shifting between all three styles. Vowel dispersion differences were observed in the hyperarticulated style for a majority of subjects, but for the crucial reduced-citation distinction, only a total of four subjects out of 22 across both sets of experiments modified their vowel spaces. The use of a heavier cognitive load in Experiments 4 – 6 did increase the proportion of speakers who produced a reduced style that was distinctive from their citation style. However, the overall durational and vowel dispersion differences were still quite modest.

While the individually-calibrated cognitive load method is clearly effective in many instances, the large minority of subjects still failed to produce reduced sentences that were perceptually distinct from their citation sentences. This may in part be due to the restrictive criteria used for selecting reduced and citation sentences. The individual participants' success rates were based on comparisons of one example of each sentence in each speech style, out of a total of two to four elicited. 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 were 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, does not ensure that all subjects will produce a reduced-citation distinction. One way to account for these residual speakers may involve differences among individuals in terms of their performance on the Digit Span Task versus their ability to correctly recall digits in the Reduced condition. That is, the distractor task could disproportionately affect some listeners more than others in span recall. If true, such individual differences would be reflected in the percentage of digit sequences correctly recalled in the Reduced condition. Table 5 lists these percentages for the twelve participants in Experiments 1 - 3, who are grouped in terms of their digit span. As Table 5 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 are related to issues 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 will ignore the load task and simply read the sentence in citation style. If the load is not sufficiently demanding of attentional resources, then no pressure will exist to induce reduction. 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 "operational" span may have been much lower than that indicated in the simple digit span task, making the calibrated load too difficult to induce reduction. 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 an attempt to improve the method, it may also be worthwhile to modify other aspects of the Reduced condition. For example, a different cognitive load could be employed, such as a sequence of familiar words, unfamiliar words, nonsense words, or perhaps anomalous sentences. Such sequences may constitute a more effective load due to their morphological, lexical, or syntactic similarity to the target sentences and greater uncertainty. However, such similarity could also have the undesirable effect of generating more disfluencies in the target sentences. In addition, an adaptive algorithm could be used throughout the elicitation of reduced sentences. Currently, a fixed range of loads is used in the elicitation procedure, one that has been calibrated to the individual participant in an immediate serial recall digit span task (the calibration task). However, due to changes in attention or fatigue, a participant's operational digit span could change over the course of the elicitation procedure and, thus, could be higher or lower than that measured in the calibration task. One way to address this possibility would be to adjust the cognitive load adaptively over the course of the elicitation procedure, increasing the load when participants continue to perform well (i.e., recall the digit sequence correctly), and decreasing the load when participants fail to correctly recall a sequence in order.

Of course, it is possible that all of these proposed 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 only 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.

## Stimulus Sentences

### Appendix A

The farmer harvested his crop.	The hockey player scored a goal.
His boss made him work like a slave.	How long can you hold your breath?
He caught the fish in his net.	At breakfast he drank some juice.
Close the window to stop the draft.	The king wore a golden crown.
The beer drinkers raised their mugs.	He got drunk in the local bar.
I made the phone call from a booth.	The doctor prescribed the drug.
The cut on his knee formed a scab.	The landlord raised the rent.
The railroad train ran off the track.	Playing checkers can be fun.
They drank a whole bottle of gin.	Throw out all this useless junk.
The airplane dropped a bomb.	Her entry should win first prize.
I gave her a kiss and a hug.	The stale bread was covered with mold.
The soup was served in a bowl.	I ate a piece of chocolate fudge.
The cookies were kept in a jar.	The story had a clever plot.
How did your car get that dent?	He's employed by a large firm.
The baby slept in his crib.	The mouse was caught in the trap.
The cop wore a bullet-proof vest.	I've got a cold and a sore throat.
No one was injured in the crash.	The judge is sitting on the bench.

## ANOVAs for individual subjects (Experiment 3)

### Appendix B

Subject	Style	Unit of Analysis	Interaction
1	$F(2,189) = 229.5, p < 0.0001$	$F(1,189) = 1857.3, p < 0.0001$	$F(2,189) = 123.8, p < 0.0001$
2	$F(2,180) = 324.1, p < 0.0001$	$F(1,180) = 3217.4, p < 0.0001$	$F(2,180) = 194.5, p < 0.0001$
3	$F(2,189) = 169.1, p < 0.0001$	$F(1,189) = 3527.6, p < 0.0001$	$F(2,189) = 86.9, p < 0.0001$
4	$F(2,192) = 339, p < 0.0001$	$F(1,192) = 3553.4, p < 0.0001$	$F(2,192) = 192, p < 0.0001$
5	$F(2,192) = 346.2, p < 0.0001$	$F(1,192) = 2114.6, p < 0.0001$	$F(2,192) = 202.2, p < 0.0001$
6	$F(2,195) = 168.5, p < 0.0001$	$F(1,195) = 2235.2, p < 0.0001$	$F(2,195) = 96.2, p < 0.0001$
7	$F(2,189) = 339.8, p < 0.0001$	$F(1,189) = 2963.9, p < 0.0001$	$F(2,189) = 213.3, p < 0.0001$

Subject	Style	Unit of Analysis	Interaction
8	$F(2,180) = 146.7, p < 0.0001$	$F(1,180) = 2459.3, p < 0.0001$	$F(2,180) = 78.2, p < 0.0001$
9	$F(2,165) = 205.3, p < 0.0001$	$F(1,165) = 1908.4, p < 0.0001$	$F(2,165) = 109.6, p < 0.0001$
10	$F(2,186) = 127.9, p < 0.0001$	$F(1,186) = 2490.2, p < 0.0001$	$F(2,186) = 65.3, p < 0.0001$
11	$F(2,156) = 65.5, p < 0.0001$	$F(1,156) = 4213.4, p < 0.0001$	$F(2,156) = 38.6, p < 0.0001$
12	$F(2,138) = 187.7, p < 0.0001$	$F(1,138) = 1318.2, p < 0.0001$	$F(2,138) = 110.3, p < 0.0001$

## ANOVAs for individual subjects (Experiment 6)

### Appendix C

Subject	Style	Unit of Analysis	Interaction
13	$F(2,192) = 214.8, p < 0.0001$	$F(1,192) = 2337.8, p < 0.0001$	$F(2,192) = 133.4, p < 0.0001$
14	$F(2,198) = 285.8, p < 0.0001$	$F(1,198) = 3496.8, p < 0.0001$	$F(2,198) = 162, p < 0.0001$
15	$F(2,180) = 109.1, p < 0.0001$	$F(1,180) = 2967.8, p < 0.0001$	$F(2,180) = 57.2, p < 0.0001$
16	$F(2,198) = 96, p < 0.0001$	$F(1,198) = 2648.7, p < 0.0001$	$F(2,198) = 50.4, p < 0.0001$
17	$F(2,198) = 367.6, p < 0.0001$	$F(1,198) = 4271.3, p < 0.0001$	$F(2,198) = 243.2, p < 0.0001$
18	$F(2,186) = 87.5, p < 0.0001$	$F(1,186) = 1771.5, p < 0.0001$	$F(2,186) = 47, p < 0.0001$
19	$F(2,198) = 186.9, p < 0.0001$	$F(1,198) = 2346.3, p < 0.0001$	$F(2,198) = 100.2, p < 0.0001$
20	$F(2,198) = 253.1, p < 0.0001$	$F(1,198) = 2776.1, p < 0.0001$	$F(2,198) = 140.5, p < 0.0001$
21	$F(2,180) = 326.2, p < 0.0001$	$F(1,180) = 2557.9, p < 0.0001$	$F(2,180) = 179.6, p < 0.0001$
22	$F(2,198) = 139.1, p < 0.0001$	$F(1,198) = 3158, p < 0.0001$	$F(2,198) = 80.5, p < 0.0001$

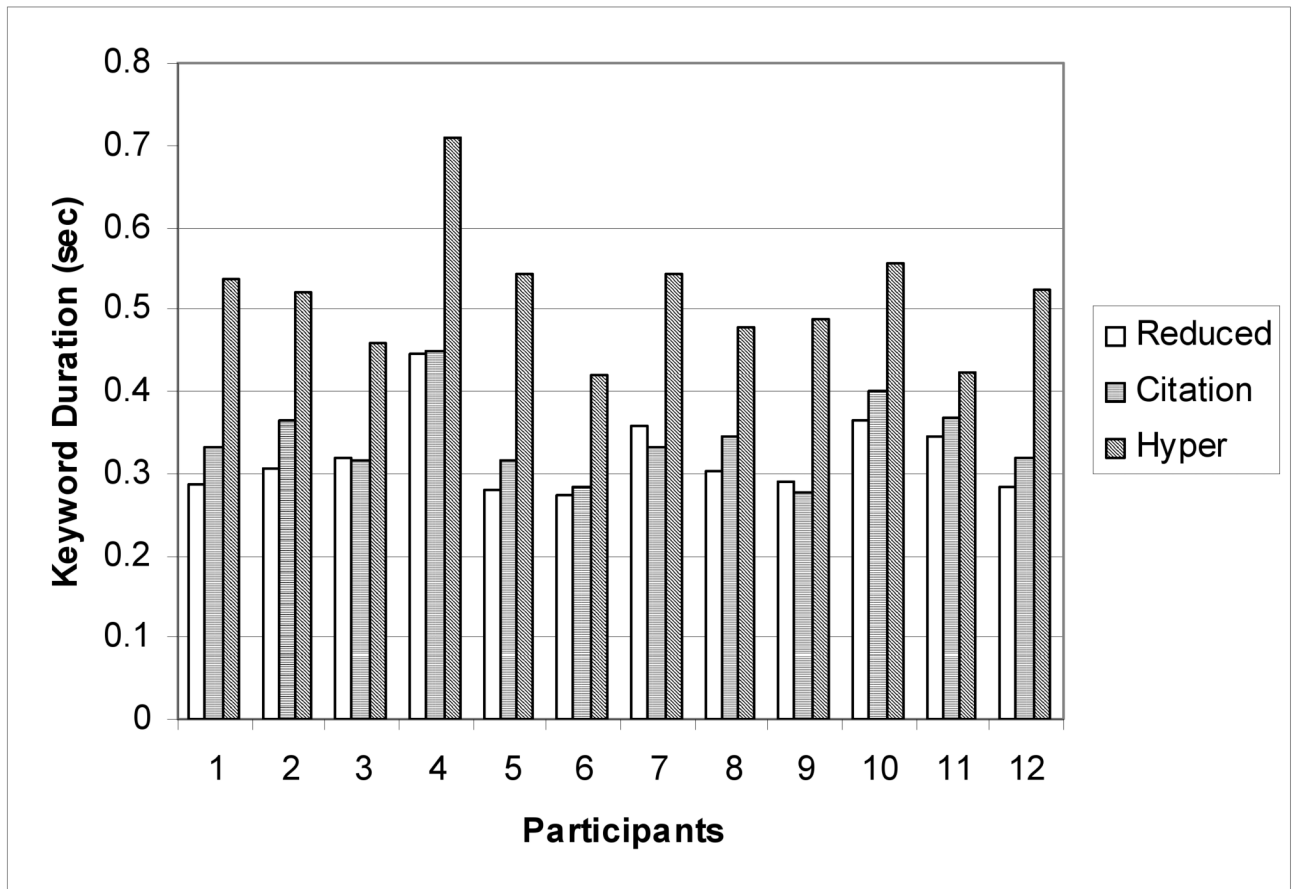
## References

- Aylett M, Turk A. The smooth signal hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Lang Speech* 2004;47:31–56. [PubMed: 15298329]
- Baddeley AD, Thomson N, Buchanan M. Word length and structure of short-term-memory. *J Verbal Learning Verbal Behavior* 1975;14:575–589.
- Bates R, Ostendorf M, Wright R. Symbolic phonetic features for modeling of pronunciation variation. *Speech Commun* 2007;49:83–97.
- Bradlow AR, Bent T. The clear speech effect for non-native listeners. *J Acoust Soc Am* 2002;112:272–284. [PubMed: 12141353]
- Bradlow AR, Nygaard LC, Pisoni DB. Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Percept Psychophys* 1999;61:206–219.
- Bradlow AR, Torretta GM, Pisoni DB. Intelligibility of normal speech. I. Global and fine-grained acoustic-phonetic talker characteristics. *Speech Commun* 1996;20:255–272.
- Brink, J.; Wright, R.; Pisoni, DB. Research on Spoken Language Processing Progress Report No 22. Speech Research Laboratory, Indiana University; Bloomington, IN: 1998. Eliciting speech reduction in the laboratory: Assessment of a new experimental method; p. 396–420.
- Byrd D. Relations of sex and dialect to reduction. *Speech Commun* 1994;15:39–54.

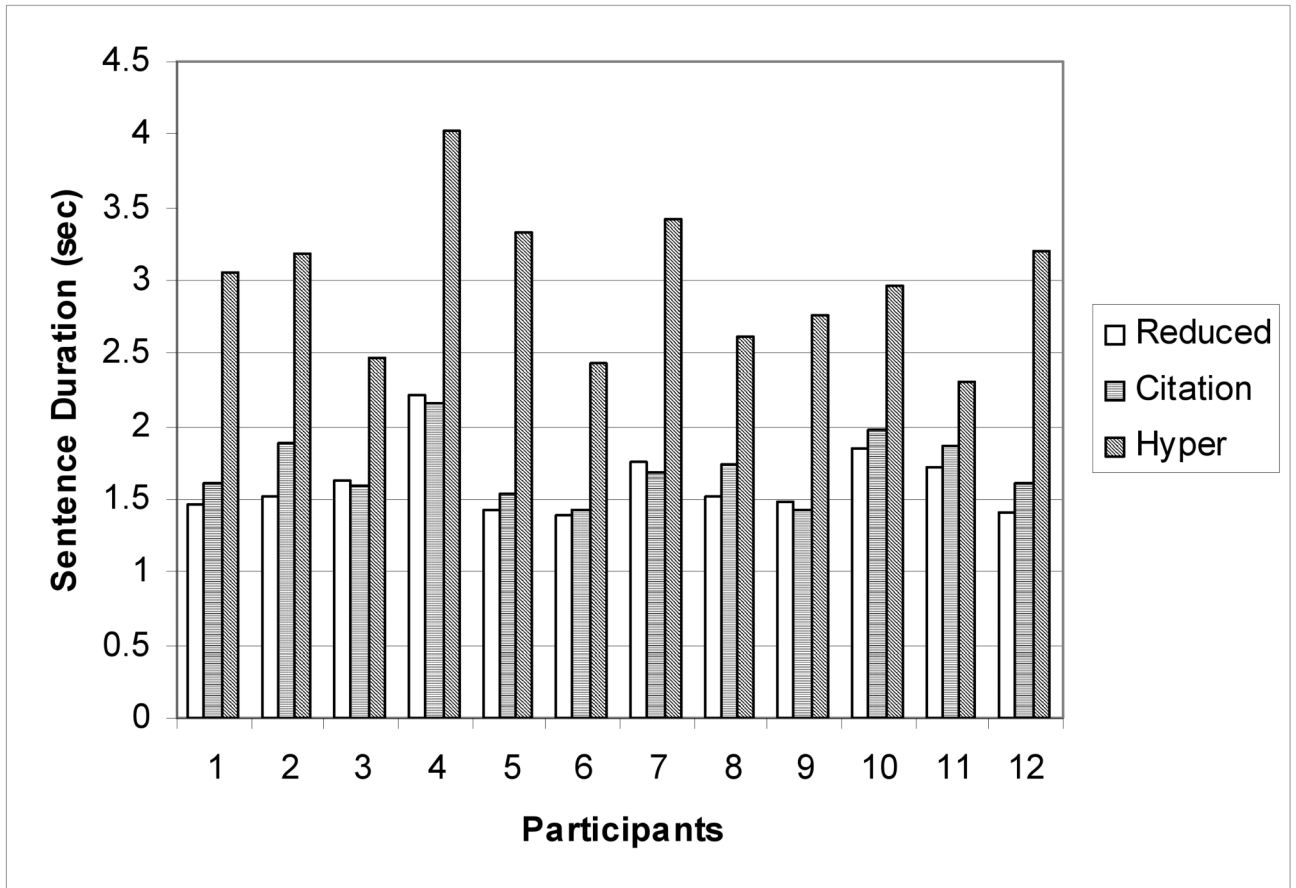


- Cavanagh JB. Relation between the immediate memory span and the memory search rate. *Psych Rev* 1972;79:525–530.
- Duez D. Second formant locus-nucleus patterns: An investigation of spontaneous French speech. *Speech Commun* 1992;11:417–427.
- Fernald A, Taeschner T, Dunn J, Papousek M, de Boysson-Bardies B, Fukui I. A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *J Child Lang* 1989;16:477–501. [PubMed: 2808569]
- Hirschberg, J.; Nakatani, CH. *Proc Association for Computational Linguistics*. Santa Cruz, CA: 1996. A prosodic analysis of discourse segments in direction-giving monologues; p. 286-293.1996
- Hirschberg J, Litman D, Swerts M. Prosodic and other cues to speech recognition failures. *Speech Commun* 2004;43:155–175.
- Johnson K, Flemming E, Wright R. The hyperspace effect: Phonetic targets are hyperarticulated. *Language* 1993;69:505–528.
- Jurafsky, D.; Bell, A.; Gregory, M.; Raymond, W. Probabilistic relations between words: Evidence from reduction in lexical production. In: Bybee, J.; Hopper, P., editors. *Frequency and the emergence of linguistic structure*. Amsterdam, The Netherlands: John Benjamins; 2001. p. 229-254.
- Kalikow DN, Stevens KN, Elliot LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Amer* 1977;61:1337–1351. [PubMed: 881487]
- Köster S. Acoustic-phonetic characteristics of hyperarticulated speech for different speaking styles. *ICASSP IEEE Int Conf Acoust Speech Signal Process Proc* 2001;2:873–876.
- Krull, D. *Phonetic Experimental Research*. Vol. PERILUS X. Institute of Linguistics, Stockholm University; 1989. Consonant-vowel coarticulation in spontaneous speech and reference words; p. 101-105.
- Kuhl PK, Andruski JE, Chistovich IA, Chistovich LA, Kozhevnikova EV, Ryskina VL, Stolyarova EI, Sundberg U, Lacerda F. Cross-Language analysis of phonetic units in language addressed to infants. *Science* 1997;277:684–686. [PubMed: 9235890]
- Labov, W. *Sociolinguistic patterns*. University of Pennsylvania Press; Philadelphia: 1972.
- Labov, W. Field methods of the Project on Linguistic Change and Variation. In: Baugh, J.; Sherzer, J., editors. *Language in Use: Readings in Sociolinguistics*. Englewood Cliffs, NJ: Prentice Hall; 1984. p. 28-53.
- Levitt H. Transformed up-down methods in psychoacoustics. *J Acoust Soc Amer* 1971;49:467–477. [PubMed: 5541744]
- Lindblom, B. Explaining phonetic variation: A sketch of the H and H theory. In: Hardcastle, W.; Marchal, A., editors. *Speech Production and Speech Modeling*. Dordrecht, Netherlands: Kluwe Academic Publishers; 1990. p. 403-439.
- Lindblom B. Role of articulation in speech perception: Clues from production. *J Acoust Soc Amer* 1996;99:1683–1692. [PubMed: 8819859]
- Liu Y, Shriberg E, Stolcke A, Hillard D, Ostendorf M, Harper M. Enriching speech recognition with automatic detection of sentence boundaries and disfluencies. *IEEE Transactions Audio Speech Lang Proc* 2006;14:1526–1540.
- Moon SJ, Lindblom B. Interaction between duration, context, and speaking style in English stressed vowels. *J Acoust Soc Amer* 1994;96:40–55.
- Milroy, L. *Observing and analyzing natural language*. Basil Blackwell; Oxford: 1987.
- Mullennix JW, Pisoni DB. Stimulus variability and processing dependencies in speech perception. *Percept Psychophysics* 1990;61:206–219.
- Nygaard LC, Sommers MS, Pisoni DB. Effects of stimulus variability on perception and representation of spoken words in memory. *Percept Psychophysics* 1995;57:989–1001.
- Oberauer K, Kliegl R. A formal model of capacity limits in working memory. *J Memory Lang* 2006;55:601–626.
- Ostendorf M, Byrne B, Bacchiani M, Finke M, Gunawardana A, Ross K, Roweis S, Shriberg E, Talkin D, Waibel A, Wheatley B, Zeppenfeld T. Modeling systematic variations in pronunciation via a

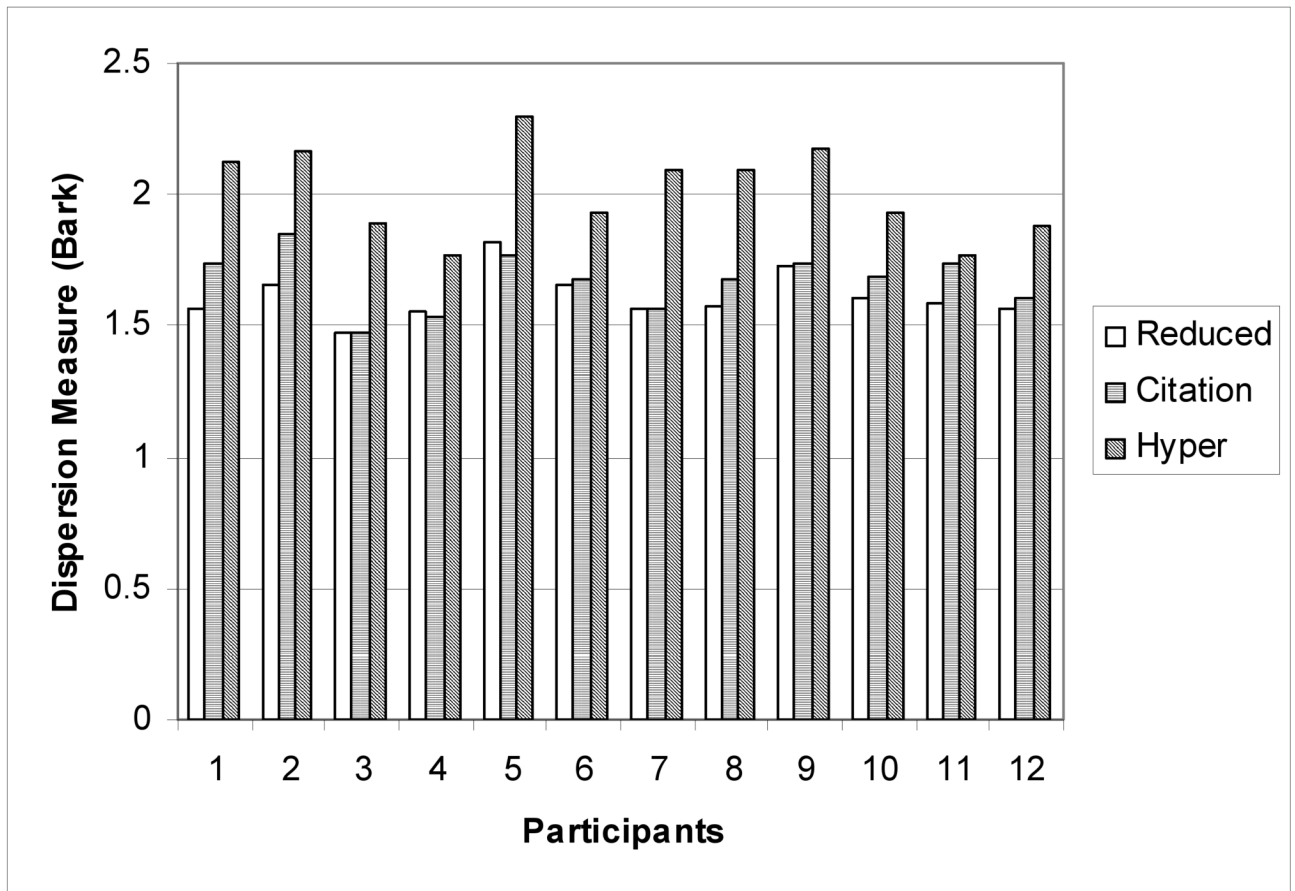
- language-dependent hidden speaking mode. Report on 1996 CLSP/JHU Workshop on Innovative Techniques for Large Vocabulary Continuous Speech Recognition. 1996
- Oviatt SL, Levow G, Moreton E, MacEachern M. Modeling global and focal hyperarticulation during human-computer error resolution. *J Acoust Soc Amer* 1998;104:3080–3098. [PubMed: 9821350]
- Oviatt SL, MacEachern M, Levow G. Predicting hyperarticulate speech during human-computer error resolution. *Speech Commun* 1998;24:87–110.
- Payton KL, Uchanski RM, Braida LD. Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *J Acoust Soc Amer* 1994;95:1581–1592. [PubMed: 8176061]
- Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *J Speech Hear Res* 1985;28:96–103. [PubMed: 3982003]
- Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *J Speech Hear Res* 1986;29:434–446. [PubMed: 3795886]
- Picheny MA, Durlach NI, Braida LD. 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. *J Speech Hear Res* 1989;32:600–603. [PubMed: 2779204]
- Schriberg E. To 'errr' is human: ecology and acoustics of speech disfluencies. *J Int Phonetic Association* 2001;31:153–169.
- Smiljanić R, Bradlow AR. Production and perception of clear speech in Croatian and English. *J Acoust Soc Amer* 2005;118:1677–1688. [PubMed: 16240826]
- Speer SR, Sokol SB, Schafer AJ. Prosodic disambiguation of syntactic ambiguity in discourse context. *J Acoust Soc Amer* 1999;106:2275.
- Swerts M, Collier R. On the controlled elicitation of spontaneous speech. *Speech Commun* 1992;11:463–468.
- Summers W, Pisoni DB, Bernacki RH, Pedlow RI, Stokes MA. Effects of noise on speech production: Acoustic and perceptual analyses. *J Acoust Soc Amer* 1988;84:917–928. [PubMed: 3183209]
- Uchanski RM, Choi SS, Braida LD, Durlach NI. Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *J Speech Hear Res* 1996;39:494–509. [PubMed: 8783129]
- Uther M, Knoll MA, Burnham D. Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Commun* 2007;49:2–7.
- Wassink AB, Wright R, Franklin A. Intraspeaker variability in vowel production: an investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. *J Phonetics* 2007;35:353–379.
- Wright, R. Factors of lexical competition in vowel articulation. In: Local, John; Ogden, Richard; Temple, Rosalind, editors. *Phonetic Interpretation: Papers in Laboratory Phonology VI*. Cambridge: CUP; 2003. p. 75-87.



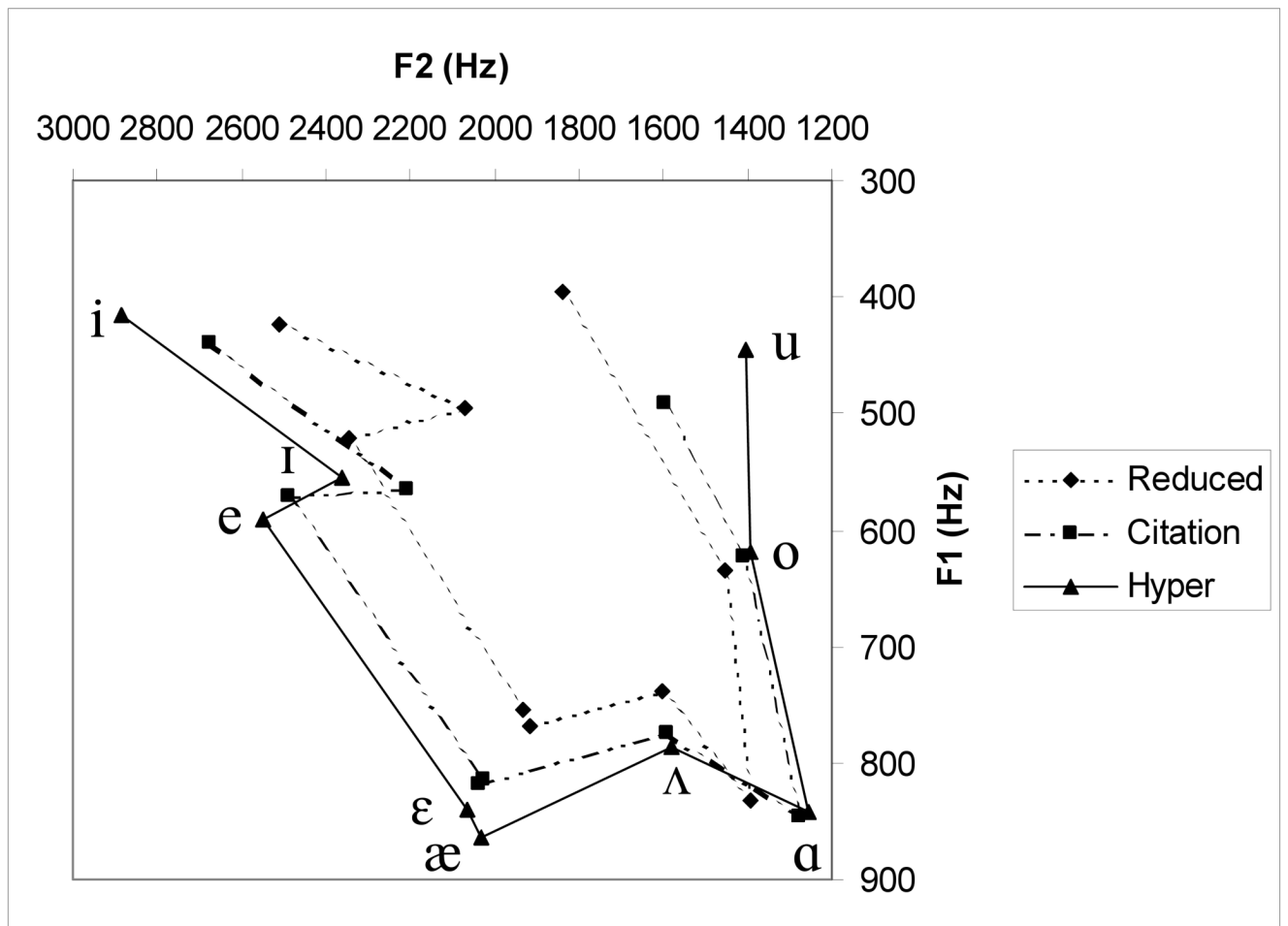
**Fig. 1.**  
Mean keyword durations of each style for each participant.



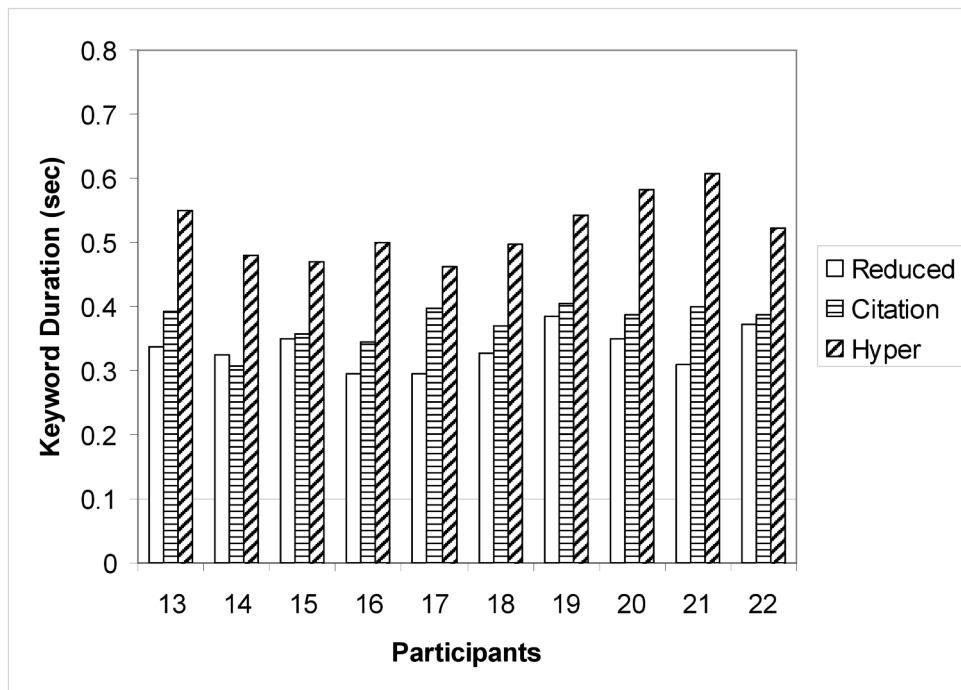
**Fig. 2.**  
Mean sentence durations of each style for each participant.



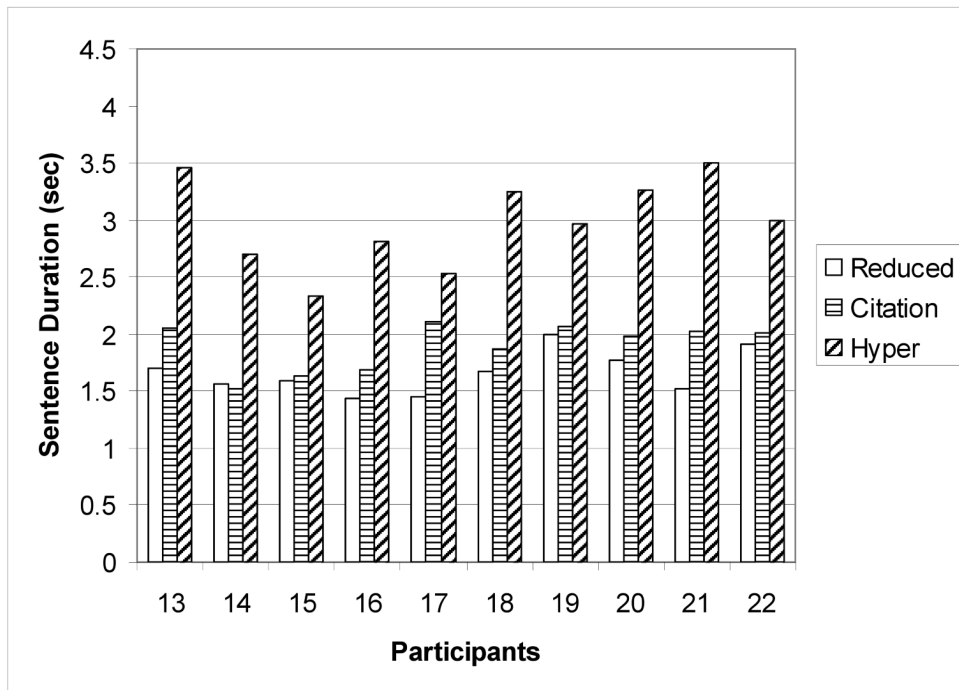
**Fig. 3.**  
Vowel dispersion measures of each style for each participant.



**Fig. 4.** The vowel spaces of participant 2 for keyword vowels in Reduced, Citation, and Hyperarticulated (Hyper) styles.

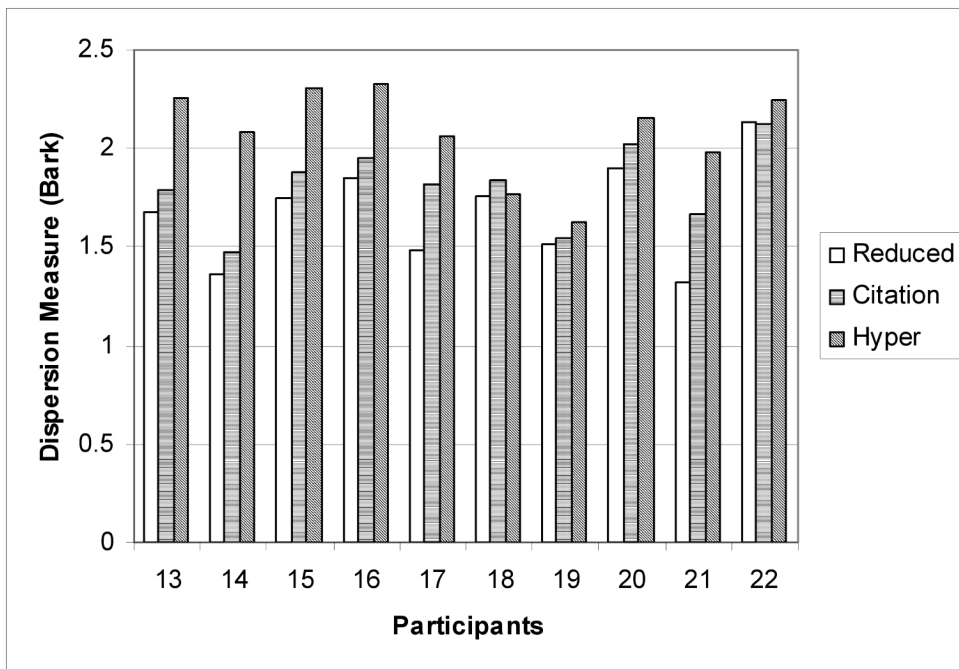


**Fig. 5.**  
Mean keyword durations of each style for each participant.



**Fig. 6.**  
Mean sentence durations of each style for each participant.





**Fig. 7.**  
Vowel dispersion measures of each style for each participant.

**Table 1**

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

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 2**

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

Talker	Reduced-Citation	Reduced-Hyperarticulated	Citation-Hyperarticulated
2	88%	100%	99%
3	78%	93%	85%
9	71%	86%	96%
12	73%	96%	98%

**Table 3**

The percentage of sentence pairs judged to be qualitatively different in speech style in Experiment 4.

Subject	Reduced-Citation	Reduced-Hyperarticulated	Citation-Hyperarticulated
13	85%	100%	94%
14	21%	100%	100%
15	74%	100%	100%
16	76%	100%	100%
17	100%	100%	82%
18	68%	100%	100%
19	32%	100%	100%
20	79%	100%	100%
21	100%	100%	100%
22	59%	100%	100%

**Table 4**

The percentage of sentence pairs judged correctly by the untrained listeners in Experiment 5.

Talker	Reduced-Citation	Reduced-Hyperarticulated	Citation-Hyperarticulated
16	85%	88%	72%
17	73%	98%	95%
20	87%	95%	83%
21	96%	100%	86%

**Table 5**

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

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