

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**  
Progress Report No. 22 (1998)  
*Indiana University*

**Eliciting Speech Reduction in the Laboratory:  
Assessment of a New Experimental Method<sup>1</sup>**

**James Brink, Richard Wright<sup>2</sup> and David B. Pisoni**

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

---

<sup>1</sup> This work is supported by NIH-NIDCD Research Grant DC00111 and Training Grant DC00012 to Indiana University. We thank Jimmy Harnsberger for editorial help and comments on an earlier version of this paper.

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

## Eliciting Speech Reduction in the Laboratory: Assessment of a New Experimental Method

**Abstract.** The aim of this study was the development and assessment of a novel method for recording speech produced under varying degrees of articulatory precision in the laboratory. Three conditions were designed to elicit reduced, citation, and careful speaking styles. In the first condition, talkers read sentences while simultaneously carrying out a distractor task. In the second condition, talkers simply read the same set of sentences in the absence of any other task. In the third condition, talkers read the sentences but were frequently prompted by the computer to re-read the last sentence more carefully. Recordings from six talkers, three female and three male, were made and measurements were carried out along a number of dimensions that have been used in the literature to distinguish careful from reduced speech or that have been linked to differences in intelligibility between talkers. Using acoustic analysis, measures of word duration, sentence duration, overall sentence energy, word energy, sentence energy range, word energy range, pitch range of the sentence, and two measures of F1-F2 vowel space dispersion were obtained for all six talkers. Overall, we observed consistent effects of speaking mode across most talkers for many of the measured dimensions. Individually, there was variability between talkers as to which dimensions showed the greatest differences between conditions. The overall pattern of results indicates that these experimental procedures can be used to elicit varying degrees of articulatory control in a highly controlled laboratory environment.

### Introduction

The study of spoken language has traditionally used speech samples that have been recorded under highly controlled laboratory conditions. This is true of articulatory and acoustic studies of speech production as well as studies of perception, word recognition, and memory for spoken words. Some researchers refer to this speaking style as “laboratory speech.” There are very good reasons for an experimenter to record speech under controlled conditions, the most important being the well-founded desire to minimize noise in the data collection process. Most experimenters now recognize the importance of studying a wider variety of speech than can be observed in the traditional recording setup, though the study of systematic variability in spoken language is a topic that has received little attention until recently (See Stevens, 1996; Pisoni, 1997).

One exception to this generalization is the recent trend towards the study of inter-talker variability. For example, Byrd (1994) studied transcriptions of the TIMIT database and found gender and dialect differences in the degree of phonological reduction. Johnson, Ladefoged and Lindau (1993) conducted an articulatory study of a smaller, more homogeneous, group of talkers and found significant differences in rate and global patterns of articulator movement. Similarly, there is a relatively long history of studying the effect of inter-talker variability on speech intelligibility (see Bradlow, Torretta, & Pisoni, 1996 for a discussion). More recently, studies have demonstrated that stimulus variability affects speech perception, spoken word recognition and memory for spoken words (Nygaard & Pisoni, 1995, *inter alia*). There are far fewer studies of within-talker variability, due perhaps to the fact that, while it is relatively easy to record a wide range of talkers reading identical words or sentences in matched environments, it is difficult to elicit variability in speaking styles and speaking modes from any one talker or to elicit equivalent changes in style across different talkers.

One way of collecting different samples of speech from the same talker is to make naturalistic recordings of everyday conversations. This is problematic from the point of view of quantitative studies of spoken language for reasons that include the problem of the control of talker external variables such as background noise. Background noise is known to affect speech production in important ways but, when recorded in naturalistic settings, its presence would rule out using recordings for detailed acoustic analysis or as stimuli for perceptual experiments. This method also has the problem of capturing utterances with comparable segmental and lexical content under a variety of contexts across different talkers. One of the most commonly used methods for recording a variety of styles from the same talker while still controlling content and recording conditions is to direct the talker to read words or sentences at a variety of speaking rates. Another, less common, method is to use a metronome to ensure the same rate changes across talkers. While the speaking rate manipulation has proven useful, and while other methods have been developed, there is an obvious need for a greater variety of controlled recording methods.

One of the most difficult types of speech to elicit in a formal recording setup is reduced, or hypoarticulated, speech and yet one of the greatest areas of interest is the systematic differences between clear, or hyperarticulated, speech and reduced speech (Dalby, 1984). While it might seem at first that recording speech at a variety of rates would be sufficient for eliciting reduction in speech, reasons for wanting rate-independent methods becomes apparent if aspects of the articulatory system are considered. As with any other highly complex motor task, talking involves a dynamic interaction between highly coordinated gestures. A change in rate of articulator movement may decrease accuracy of movement or articulatory displacement, but it will also lead to changes in the coordinative complex that are not seen in slower but less carefully produced speech.

The primary goal of this study was to develop and assess a new experimental methodology for recording speech that is produced in a way that mirrors the types of rate-independent phonetic and phonological reduction seen in conversational speech. Ideally, this method would be computerized, reducing the possibility of experimenter contamination. Moreover, it would involve the controlled manipulation of known factors in the reduction of conversational speech, including social, cognitive (degree of load), and informational (given versus new information). Of these three general factors, social ones are likely the most difficult to manipulate in a controlled recording setup. The remaining two, cognitive and informational, are more likely candidates for successful manipulation. Past research has shown that as cognitive loads increase, speech becomes more reduced (Summers et al., 1988; Lively et al., 1993). In addition, reduction has also been frequently observed in circumstances where information in the utterance is highly redundant, which allows for talkers to economize effort while successfully communicating (for a discussion see Wright, 1997; Lindblom, 1990). Thus, if the experimenter can combine increased cognitive load and reduced informational load in a given task, then highly reduced speech could be elicited.

One problem with introducing a cognitive load is to find a procedure that does not introduce disfluencies into speech (Summers et al., 1988; Lively et al., 1993). A variety of cognitive loads were explored and the one that seemed subjectively to create the greatest degree of reduction without introducing unwanted disfluencies was a digit span task, one originally suggested by Richard Shiffrin. In this procedure, subjects read the target sentences as distractors. Each sentence was used repeatedly as a distractor thereby further reducing its informational load. To ensure that the distractor sentences truly represented "reduced" sentences in the talker's continuum of speaking styles, citation and careful speech must also be elicited. It is much easier to elicit citation and careful speech in a laboratory recording setup. The formality of the recording environment and especially the visual presence of a microphone tends to encourage a relatively careful (citation) style of speaking in most subjects. It is a little more difficult to elicit very careful speech especially if the participants in the study are familiar with the

recording material. However, a method that has been used successfully in the past to elicit careful speech is to repeatedly instruct participants to read the material more carefully (Johnson, Flemming, & Wright, 1993). This method has the advantage of being relatively easy to administer automatically by having a computer program simply present a visual prompt to the participant to read more carefully.

## **Method and Procedures**

### **Participants**

Six Indiana University undergraduates participated in this experiment, three males and three females. Participants were between the ages of 18 and 22, had normal hearing, and reported no history of speech or hearing disorders. Additionally, all subjects were born and raised in northern or central Indiana (Indiana north of and including Indianapolis). The subjects were paid a total of \$15 for their participation in two sessions, each lasting about one hour long.

### **Recording Stimuli**

The subjects read a subset of forty-six sentences from the set of 200 SPIN sentences (Kalikow, Stevens, & Elliot, 1977) with “highly predictable” last words. A complete list of stimulus sentences appears in Appendix 1.

### **Recording Apparatus**

Subjects read the stimulus sentences in a sound-attenuated single-walled anechoic recording chamber (Industrial Acoustics Company Audiometric Testing Room, Model 402) using a Shure (SM98) microphone. The recordings were made using 16-bit digital sampling at a rate of 22,050 Hz with a Tucker-Davis Technologies System II on an IBM-PC 486 computer. A program was designed to control the presentation of the instructions and stimuli on a CRT monitor in the recording chamber as well as record the subjects’ responses. An experimenter was outside of the chamber during the experiment to monitor the subjects’ responses. All recordings were digitized direct to disk for later analysis.

### **Procedure**

Subjects were given printed instructions as well as oral instructions before beginning each part of the task (see Appendix 2). Subjects were instructed to put on the microphone/headset, after which the position of the microphone was adjusted to be approximately 1 inch below the chin. The microphone distance was kept the same across talkers so that differences in amplitude could be controlled. Subjects were also asked not to adjust the headset during the task. The recording sessions were divided into two parts, as outlined below. The first part was completed 1-4 days before the second part. Following the completion of the second session, the subjects were debriefed on the design and purpose of both parts of the experiment.

#### *Part I (Reduced Condition)*

Before the recording session began, the subjects were told that they were participating in a short-term memory task (see instructions, Appendix 2), and were instructed to focus on memorizing a sequence of digits that were presented visually on a computer monitor. They were also told that there would be sentences to read as distractors. This session consisted of 184 trials (four trials for each sentence) and lasted about 55 minutes. At the beginning of each trial, the cue “Get ready for the next set of numbers” was displayed. After a 1000 ms interval, a list of four to seven digits was presented on the screen. The digits remained on the screen for 2000 ms. After a 2500 ms interval, one of the stimulus

sentences was presented on the screen which the subject read aloud. The computer recorded the subject's response for 6000 ms, ensuring that the complete response would be recorded. Finally, the cue "Repeat the numbers" was presented on the screen. The subject's responses were manually recorded by the experimenter. There was an ITI of 1000 ms before the next set of numbers was presented. The trials were ordered in blocks of four, with each trial in a given block presenting the same stimulus sentence. The order of the blocks was randomized for each subject, with a total of 46 blocks, each with a different stimulus sentence presented on each trial. There were no additional breaks or pauses between blocks. The subjects were unaware of the blocking, except for the presentation of the same stimulus sentence four times.

### *Part II (Citation and Hyperarticulation Conditions)*

The second session, which lasted about 50 minutes, was divided into two phases; the first phase consisted of 46 trials, and the second consisted of 230 trials. For the first phase (citation condition), each trial had the following format: first, a cue was presented on the screen ("Get ready for the next sentence"); then, after a 1000 ms interval, one of the stimulus sentences was presented on the screen, which the subject read aloud. The ITI was 1000 ms. The trials were presented in a different random order for each subject.

The second section (hyperarticulation condition) had a format similar to the first except that there were two trial types. Trial type 1 (citation cycle) was identical to that in the citation condition: the citation cue "Get ready for the next sentence" was presented, following a 10000 ms interval a stimulus sentence was presented on the CRT monitor. In trial type 2 (hyperarticulation cycle) the cue "Please read the sentence more clearly," was presented and after a 1000 ms interval a stimulus sentence was presented. While the method was being piloted, it was found that the maximum amount of hyperarticulation occurred after the first or second presentation of the hyperarticulation cue after which there was a marked decrease in the amount of effort subjects were willing to invest in making that particular stimulus sentence clearer. Therefore, the hyperarticulation cue was presented only twice in each hyperarticulation cycle. The control program was written so that a particular stimulus sentence could either begin a regular citation cycle (trial type 1) or begin a hyperarticulation cycle (trial type 2). The hyperarticulation cycle could not appear on the initial stimulus sentence in the recording session and there was a minimum of two citation cycles (trial type 1) before a hyperarticulation cycle could begin again. Each sentence went through the "hyperarticulation cycle" once and through the "citation cycle" three times. All trials were randomized for each subject.

## **Measurement Methods**

### **Assessment**

There are two main questions that we were interested in exploring in this study. First, to what extent is this new methodology successful at eliciting three styles of speech in an automated recording setup? Second, do individual talkers differ in degree of reduction/hyperarticulation and do they differ along different dimensions? It is obvious that the second question can only be answered if the method is shown to be successful overall, but since the same set of measures will answer both questions, it seems reasonable to look at individual variation as well.

### **Acoustic Measurements**

The sentences were measured along a variety of acoustic dimensions that have been shown to be correlated with both the carefulness of speech and with speech intelligibility (e.g., Picheny, Durlach, & Braida, 1986; Uchanski, Choi, Braida, Reed, & Durlach, 1996; Bradlow, et al., 1996; Moon & Lindblom,

1994; Wright, 1997). In addition to listening to each talker's recordings and making subjective judgements about how careful each subject sounded, five acoustic measures were obtained: sentence and key word duration, sentence and key word RMS energy, sentence and key word energy range, sentence fundamental frequency ( $f_0$ ) range, and centralization/dispersion of key word vowels. The word measures were taken on three key words per sentence. All of the key words were content words and appeared in three positions within the sentences: (1) near the beginning (usually the subject noun), (2) near the middle (usually the main verb) and (3) in the final position (usually the main object of the verb). All measurements were made using the Kay Multi-Speech program and were logged automatically for later statistical analysis. Duration measures were made directly from waveforms with accompanying wide band spectrograms for reference. Energy measures (RMS and difference) were made over the portions of the waveform that had been marked by the duration measures and were based on voltage displacement.  $F_0$  measures were made using an automatic pitch-tracking algorithm that was hand-corrected using a narrow-band spectrogram for reference. Vowel formant measures were made from an overlaid LPC-FFT display (LPC: 12-16 coefficients based on the talker, 25 ms frame size; FFT: 1024 point Blackman window) with a wide-band spectrogram for reference. The formant measures were made at the point of maximal displacement of F1 and F2 (see Wright, 1997). Vowel space centralization was based on F1-F2 measures and were calculated using two formulas: 1) overall vowel space expansion based on Euclidean distances of vowels from the center of the F1-F2 space (see Bradlow, et al., 1996), and 2) vowel space repulsive force based on the inverse sums of squared distances between vowels of differing categories in an F1-F2 space (see Wright, 1997; Liljencrants & Lindblom, 1972).

## Results and Discussion

The results are divided into four separate sections, one for each type of measurement: duration, energy,  $f_0$  range, and vowel space. A brief discussion of the results is provided after each type of measurement, and an overall conclusion section summarizes the results.

### *Measurements and Statistical Tests*

Data were collected from measurements of the forty-six sentences produced in each condition. For the duration, energy, and  $f_0$  measures, an analysis of variance (ANOVA) was performed on the data, split by the three conditions. Post hoc Fisher's PLSD t-tests were also performed to determine significance levels between groups. Statistical analyses could not be performed on the vowel force and dispersion measures because these measures summed over all vowels produced in a given sentence condition (reduced, citation, hyperarticulation). Thus, there were only six values for each sentence condition, corresponding to the six talkers recorded, representing too few values to enter into an ANOVA.

For each of the key word measurements, duration, RMS energy, energy range, vowel force, and vowel dispersion, analyses were also performed over different word positions in the sentence. These position-sensitive analyses demonstrate how the speaking rate measures varied as the sentences progressed. In these measurements, Word 1 was at or near the beginning of the sentence (usually the main subject), Word 2 was in the middle (usually the main verb), and Word 3 was always the last word in the sentence (usually the direct object).

### **Durations**

#### *Sentence Durations*

Figure 1a shows the mean sentence duration for each talker across the three conditions. From this figure, it is apparent that each subject produced much longer sentences in the hyperarticulation condition than in the citation or reduced conditions. T-tests confirmed that the mean duration in the

hyperarticulation condition was significantly greater than in the other two conditions for each talker at the  $p < 0.0001$  level. However, the reduced and citation conditions did not show a consistent pattern of differences: for five of the talkers, SF, LB, DS, JC, and SK, the citation and reduced conditions were not significantly different at the  $p < 0.01$  level. Only one talker (MD) displayed longer durations in the citation condition than in the reduced condition ( $p < 0.0001$ ).

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

#### *Key Word Duration*

Figure 1b shows the key word duration for each talker across the three conditions. The results here are virtually identical to the results at the sentence level: all six talkers had significantly longer hyperarticulated words at the  $p < 0.0001$  level. In contrast, only one of six talkers, MD, had citation words significantly longer ( $p < 0.0001$ ) than the reduced words.

#### *Word-Sentence Ratio*

Figure 1c shows the word-sentence ratio for each talker across the three conditions. The word-sentence ratio represents the durations of the three key words divided by the duration of the sentence, and is an indicator of the amount of pausing in a sentence, with lower ratios corresponding to more pausing. Three of the talkers, JC, LB, and DS, showed no significant differences between any of the conditions. Three of the talkers had significantly smaller ratios in the hyperarticulation condition. ( $p < 0.01$ ) None of the talkers displayed any significant differences between the citation and reduced conditions.

#### *By-Position Word Durations*

In Figure 2, the by-position measures for word position are shown for each talker. Again, in each talker and in most of the word positions, the hyperarticulated words were longer than the words in the other conditions. In the citation and reduced conditions, however, there was a consistent lengthening of the final word (Word 3), whereas in the hyperarticulated condition, there was a consistent shortening of the final word. This was observed in a number of talkers, such as DS, LB, JC, and SK, having much less of a difference between conditions in the final word. As in the other duration measures, only talker MD has reduced words consistently shorter than citation words across all of the word positions.

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

#### *Duration Discussion*

Overall, the duration measures showed consistently longer sentences and words in the hyperarticulation condition across all six talkers. The hyperarticulation condition was also consistently marked by a decrease in the duration of the last word in the sentence, whereas all of the other sentences showed an increase in the duration of the final word of the sentence. However, only talker MD demonstrated significant shortening of word and sentence length from the citation to the reduced condition. Thus, duration is an effective indicator of hyperarticulated sentences compared to the other two conditions, but it cannot reliably discriminate between the citation and reduced conditions using the cognitive load method.

## Overall Duration Measures Across Talkers

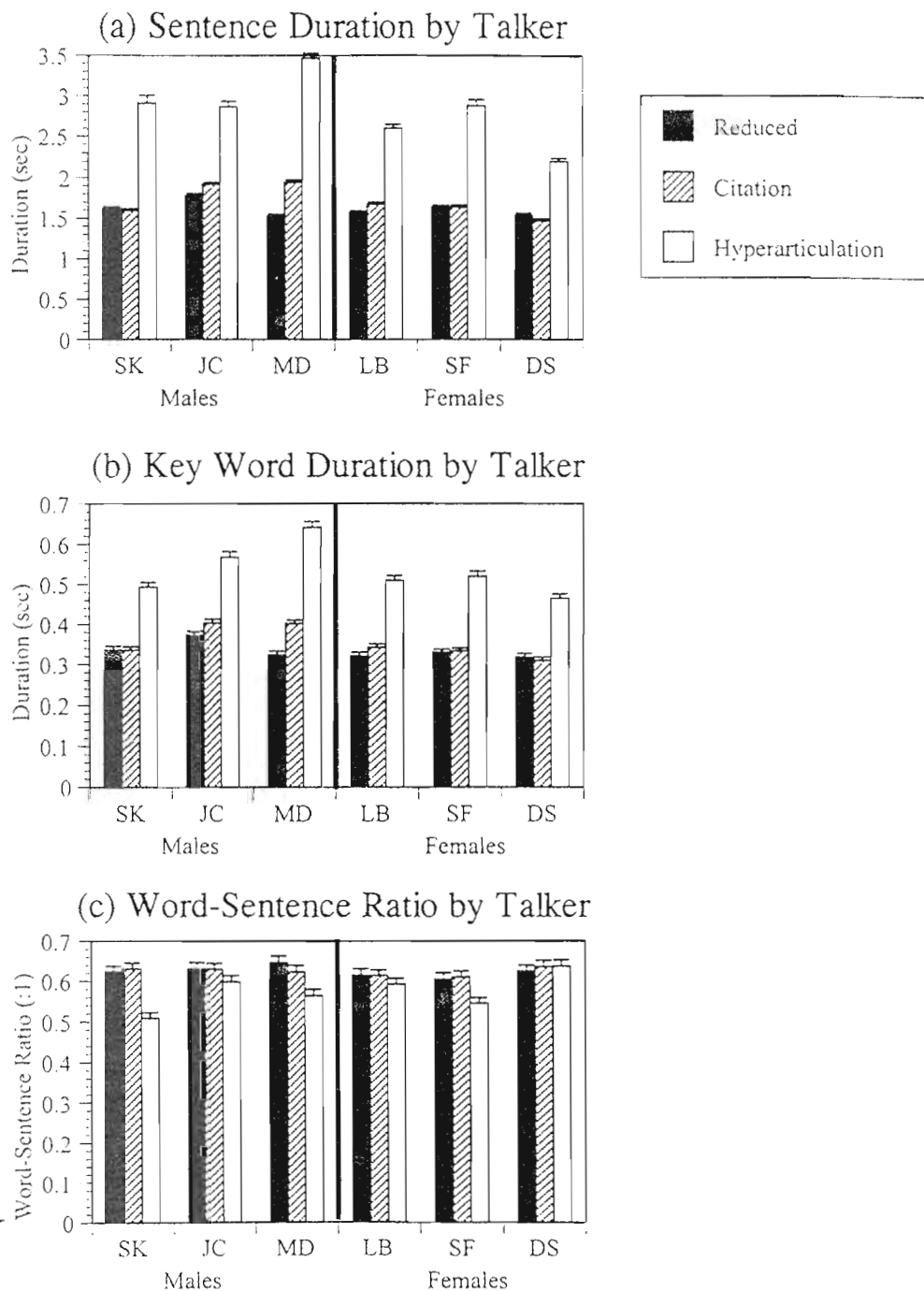


Figure 1: Duration measures across talkers - error bars indicate one standard error

### Key Word Duration Across Sentence Position

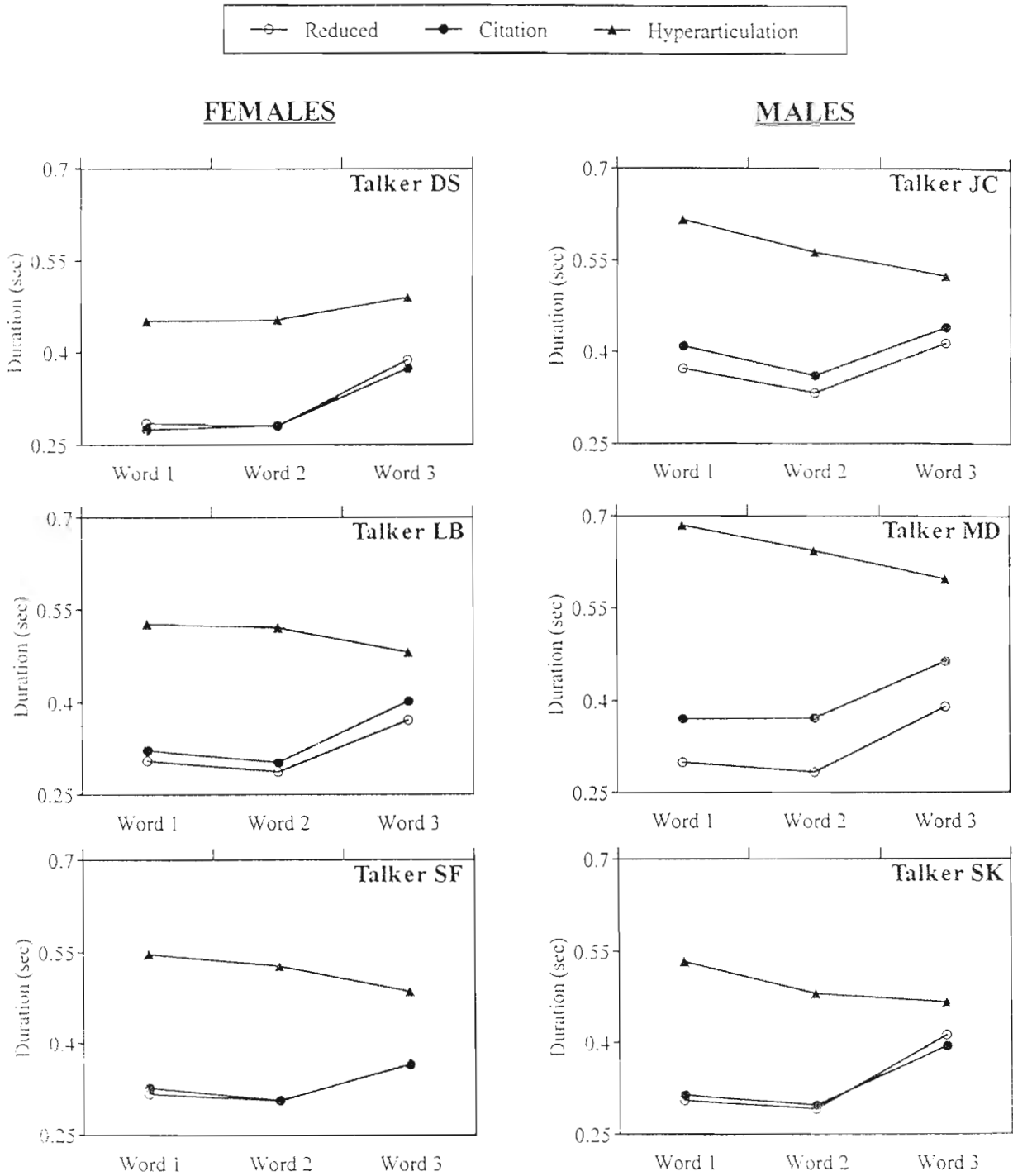


Figure 2: Key word duration for each talker across conditions, split by sentence position

## Energy

### *Sentence RMS Energy*

Figures 3 through 6 show the various RMS energy and energy range measurements for the six talkers. In Figure 3a, the sentence RMS energy is shown for each of the talkers across conditions. As shown in the graph, there were no trends across talkers. Some of the talkers showed significant differences between conditions; however, none of the talkers have the same relationships between conditions.

-----  
Insert Figures 3 through 5 about here.  
-----

### *Sentence Energy Range*

Figure 3b shows the mean energy range across the sentences for the six talkers in each of the conditions. For two of the talkers, JC and LB, the energy range for the hyperarticulation condition was significantly higher than the other two conditions ( $p < 0.05$ ), while for two other talkers, SF and DS, the energy range for the reduced condition was significantly smaller than the other two conditions ( $p < 0.05$ ). The other two talkers, SK and MD, showed no significant differences between conditions.

### *Key Word RMS Energy and Energy Range*

In Figure 3c, the overall RMS energy for the key words is shown for each of the talkers. The results were similar to those found for the sentence RMS energy. Once again, there were no consistent trends across talkers. Figure 3d shows the key word energy range across talkers. In all six of the talkers, the hyperarticulation key words had a significantly higher energy range than the other conditions at the  $p < 0.05$  level. However, only one of the talkers (DS) showed a significant difference between the citation and reduced conditions. The citation condition was significantly higher than the reduced condition.

### *By-Position Word RMS Energy*

The key word RMS energy measures for each talker across sentence position are shown in Figure 4. Four of the talkers, SF, JC, MD, and SK, showed no significant differences between any of the conditions in any of the positions, while talker LB had reduced energy measures consistently smaller than the other two. Surprisingly, talker DS had citation energy measures consistently larger than the other two. Across all of the talkers and all of the conditions, the key word RMS energy generally declined as the sentence progressed.

### *By-Position Word Energy Range*

Figure 5 shows the key word energy ranges across sentence position for the six talkers. As in the overall key word measure, the hyperarticulation energy ranges were consistently higher than the other conditions in all sentence positions across most of the talkers. Only talker DS demonstrated any difference between the citation and reduced conditions, with the citation condition having a significantly lower range ( $p < .05$ ) in words 2 and 3. In the citation and reduced conditions for some of the talkers, especially the males, the middle word generally had the lowest energy range, with the first and last key words having similar ranges; in all male talkers, the reduced energy range was significantly lower in the middle word than in the first and last words ( $p < .01$ ), and in male talkers JC and SK, the same effects hold in the citation condition. There are no significant differences in energy between positions in any of the talkers for the hyperarticulation energy ranges.

## Overall Energy Measures

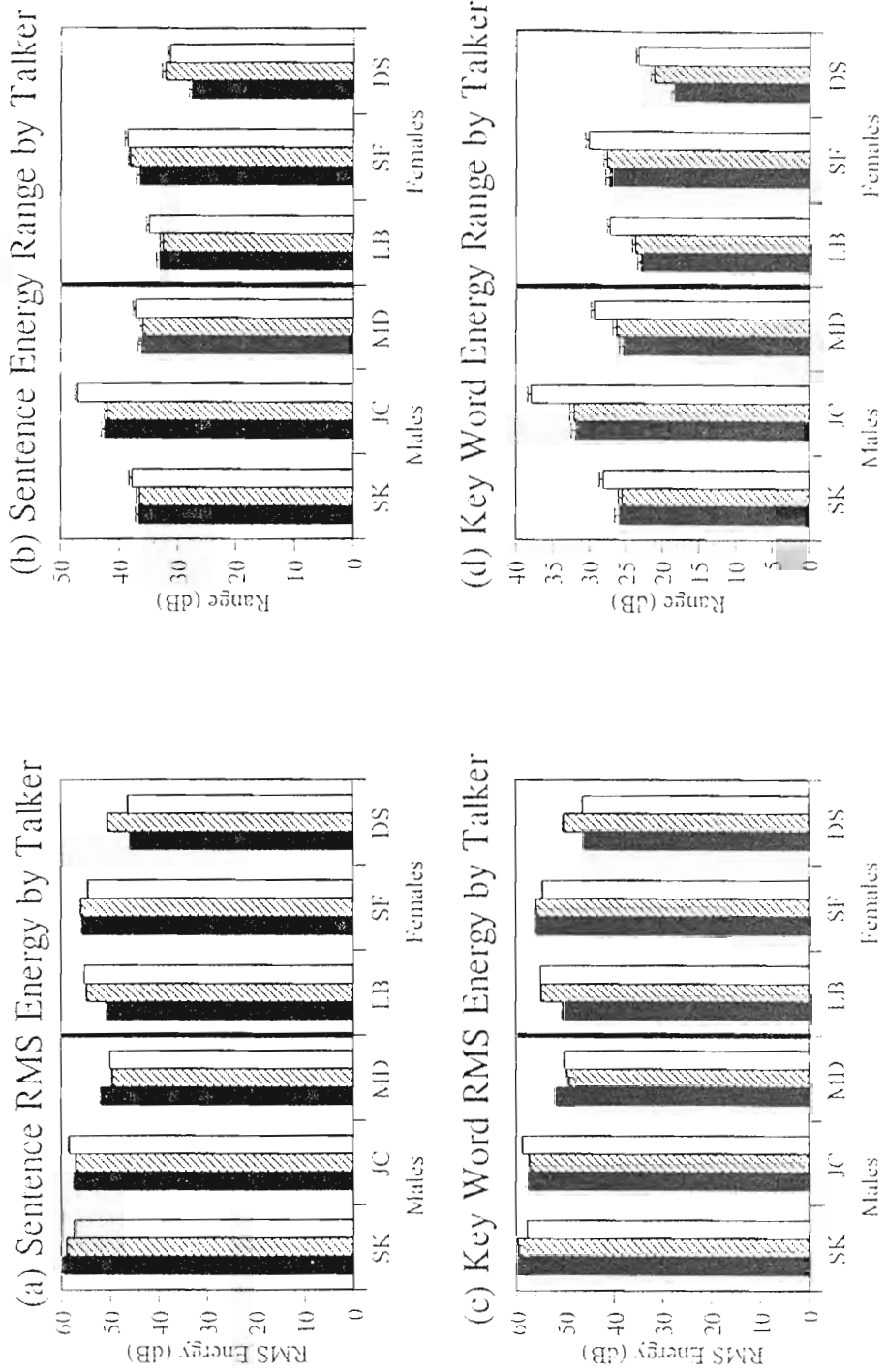
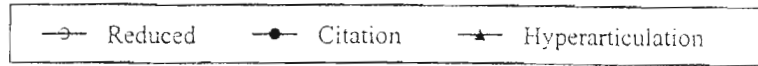


Figure 3: Overall RMS Energy and Energy Range measures across talkers - error bars indicate one standard error

### Key Word RMS Energy Across Sentence Position



#### FEMALES

#### MALES

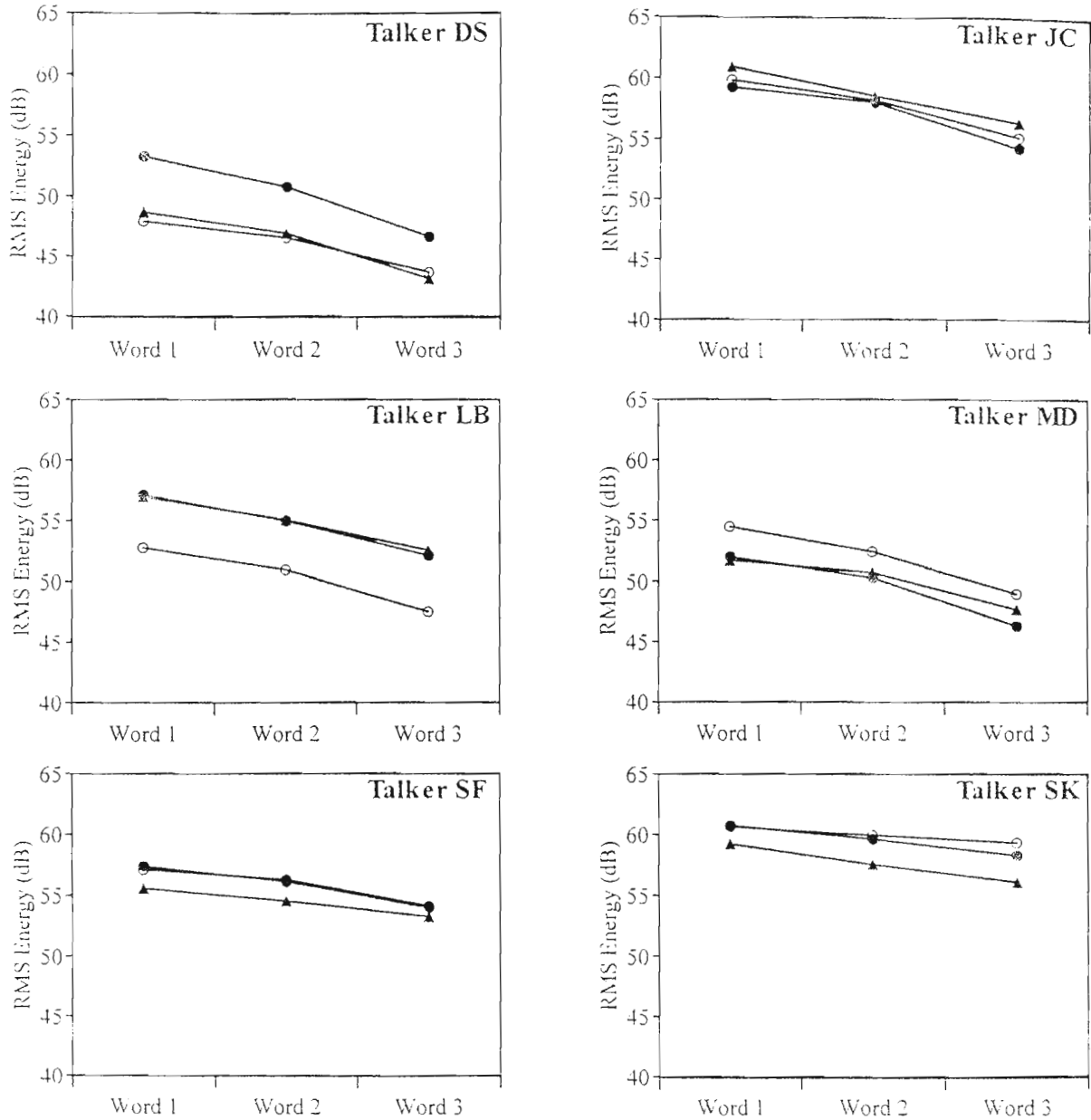
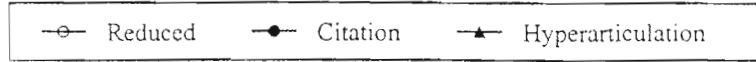


Figure 4: Key word RMS energy for each talker across conditions, split by sentence position

### Key Word Energy Range Across Sentence Position



#### FEMALES

#### MALES

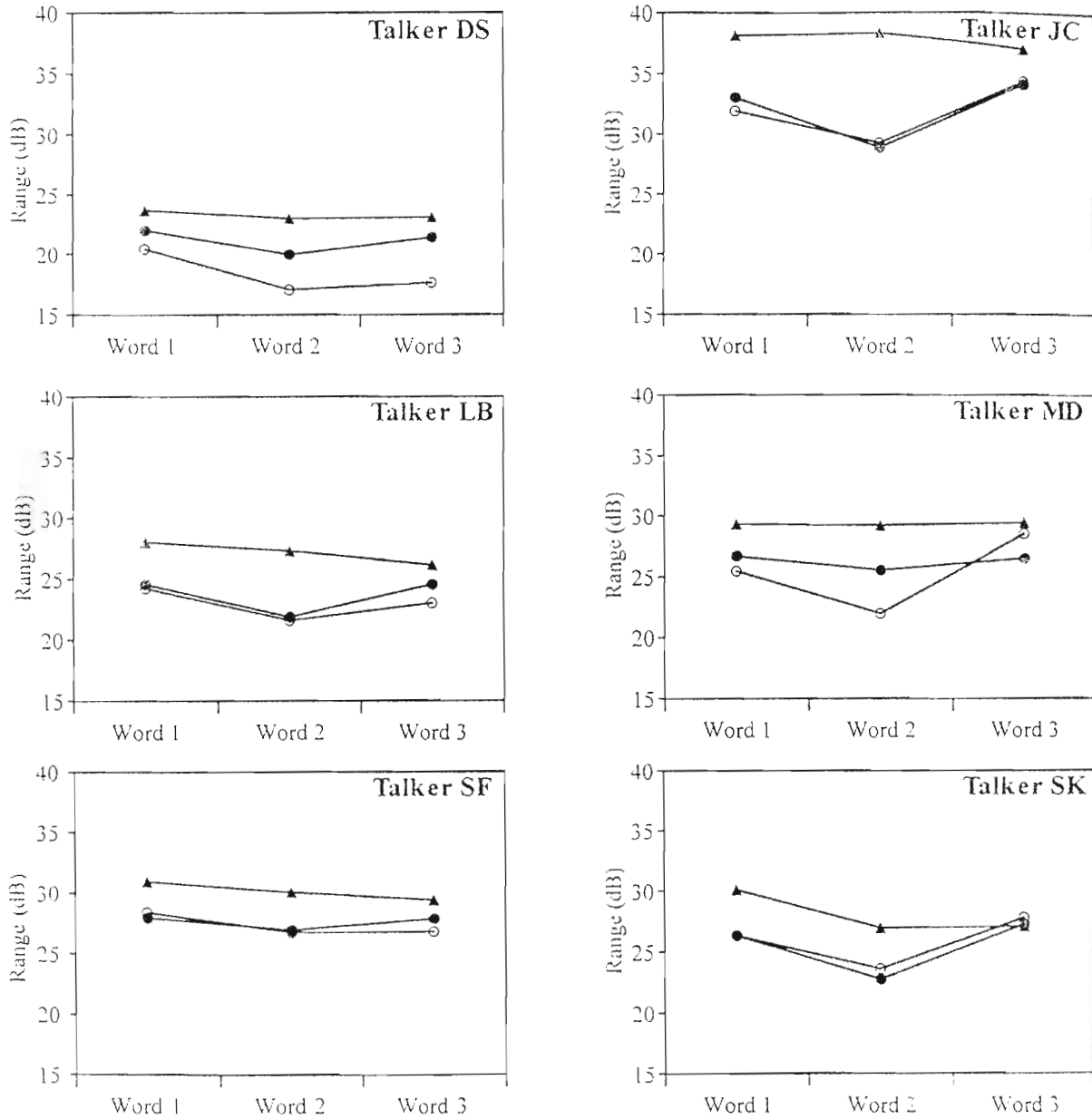


Figure 5: Key word energy range for each talker across conditions, split by sentence position

### *Energy Discussion*

RMS energy differences were not consistent across talkers between any of the conditions. Energy range did not indicate anything at the sentence level, but it was consistently higher at the word level in the hyperarticulation condition than in the other two conditions. The citation and reduced conditions did not show any consistent differences across talkers in any of the energy measurements.

### **F0 Range**

Figure 6 shows the f0 range of each of the six talkers across conditions. For two of the talkers, JC and LB, there was a measurable progression from the reduced to the hyperarticulation condition, with the reduced having the smallest range. However, only JC showed any significant differences. His hyperarticulation range was significantly higher than the range obtained in the other two conditions ( $p < 0.0001$ ). Two of the talkers, MD and SF, showed no significant differences in f0 range between conditions. The final two talkers, SK and DS, showed a different, unexpected pattern: reduced sentences had the smallest range, then hyperarticulation, and then citation with the largest range. Only the reduced and citation conditions were significantly different at the  $p < 0.01$  level for talker SK, and all of the conditions were significantly different at the  $p < 0.0001$  level for DS. In summary, no consistent differences in f0 range were observed across talkers for the three conditions recorded using the cognitive load method.

-----  
Insert Figures 6 through 9 about here.  
-----

### **Vowel Force and Dispersion**

#### *Vowel Force Measure*

Figure 7a shows the vowel force measure for all talkers across the three conditions. Four of the talkers (SK, MD, LB, and DS) showed the expected progression from the reduced to the hyperarticulation conditions, with the reduced condition having the greatest force measure (indicating a more compact vowel space), and the hyperarticulation condition having the lowest force measure. Two of the talkers varied from this trend: talker SF showed very similar reduced and citation force measures, while talker JC's lowest force measure was for the citation, rather than reduced, condition.

#### *Vowel Dispersion Measure*

Figure 7b shows the vowel dispersion measures for all of the talkers across the three conditions. Three of the talkers, MD, LB, and SF, showed a clear progression from the reduced to the hyperarticulation condition, with the reduced condition having the lowest dispersion (most compact vowel space) and the hyperarticulation condition having the highest. Two of the talkers (JC and DS) demonstrated a similar trend, though there was very little difference between the reduced and citation conditions. One talker, SK, had a citation dispersion measure that was lower than in the other two conditions.

#### *By-Position Vowel Force*

In Figure 8, the by-position measures for vowel force are shown for each talker across conditions. There were few consistent by-position effects, appearing only on a talker-by-talker basis. For two of the talkers, DS and SK, the vowel space of the reduced condition was much more compact in the final word of the sentence, as indicated by the higher force value. Also, in talker SF's results, the reduced force measure was much greater in the first key word of the sentence than in the rest of the sentence. Talker SK showed a unique change in his vowel space as the sentence progressed: in the

Sentence f0 Range by Talker

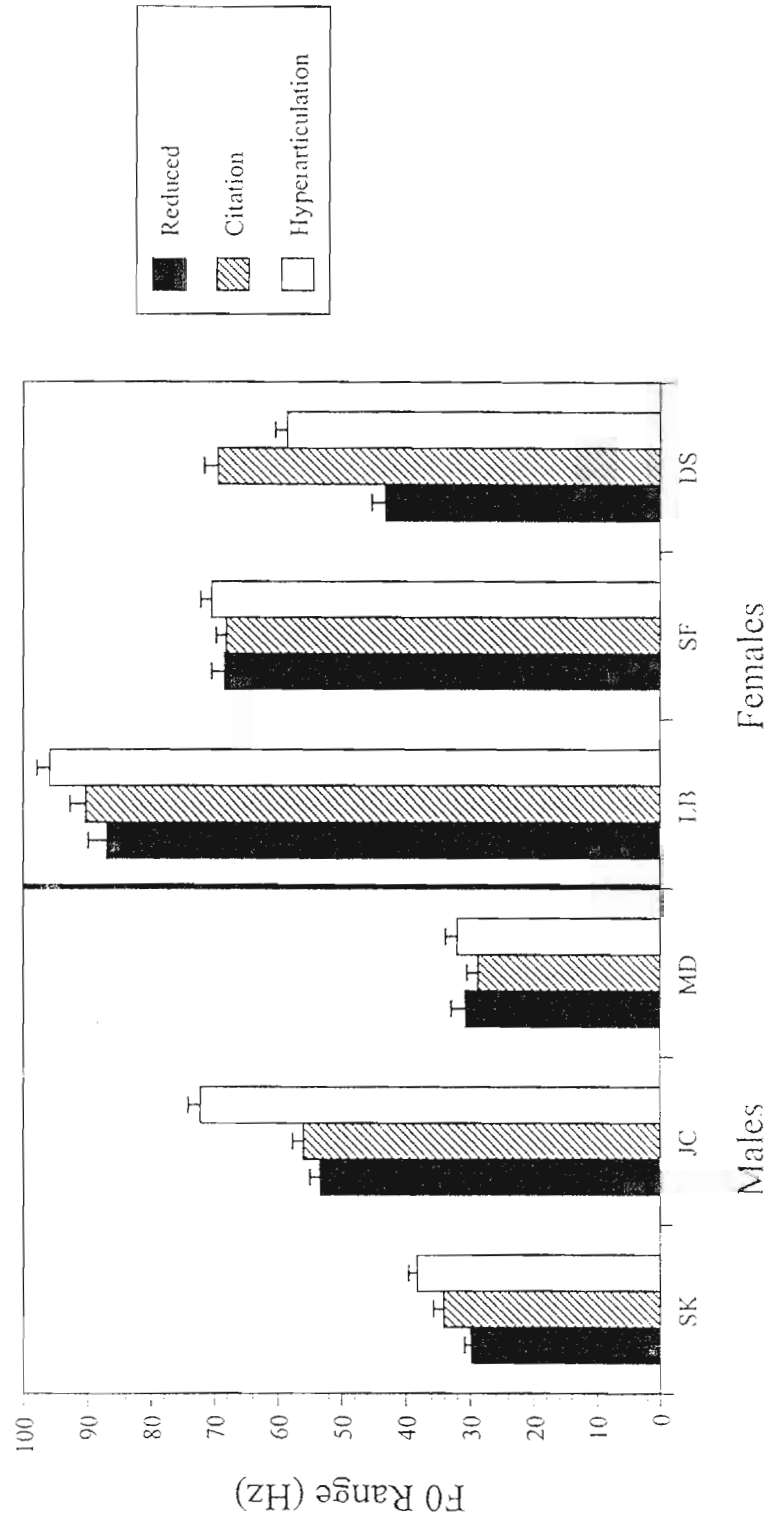
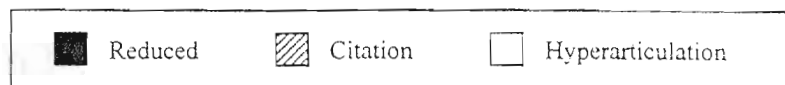
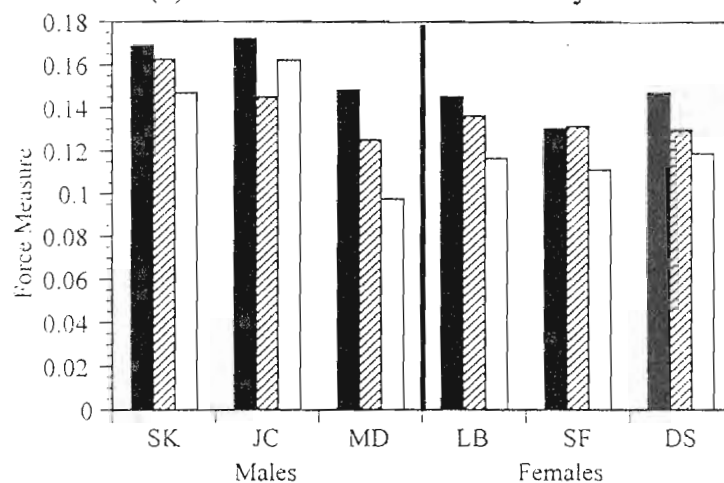


Figure 6: Sentence f0 range across talkers - error bars indicate one standard error

## Vowel Space Measures



(a) Vowel Force Measure by Talker



(b) Vowel Dispersion Measure by Talker

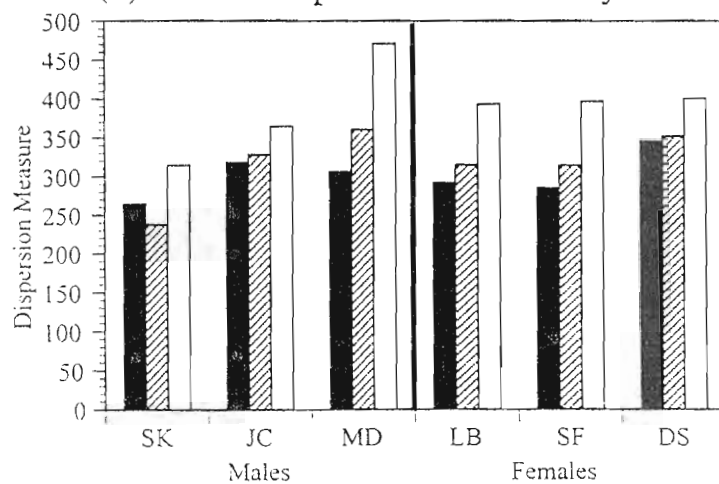
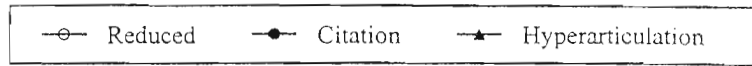


Figure 7: Vowel force and dispersion measures across talkers

### Vowel Force Measure Across Sentence Position



#### FEMALES

#### MALES

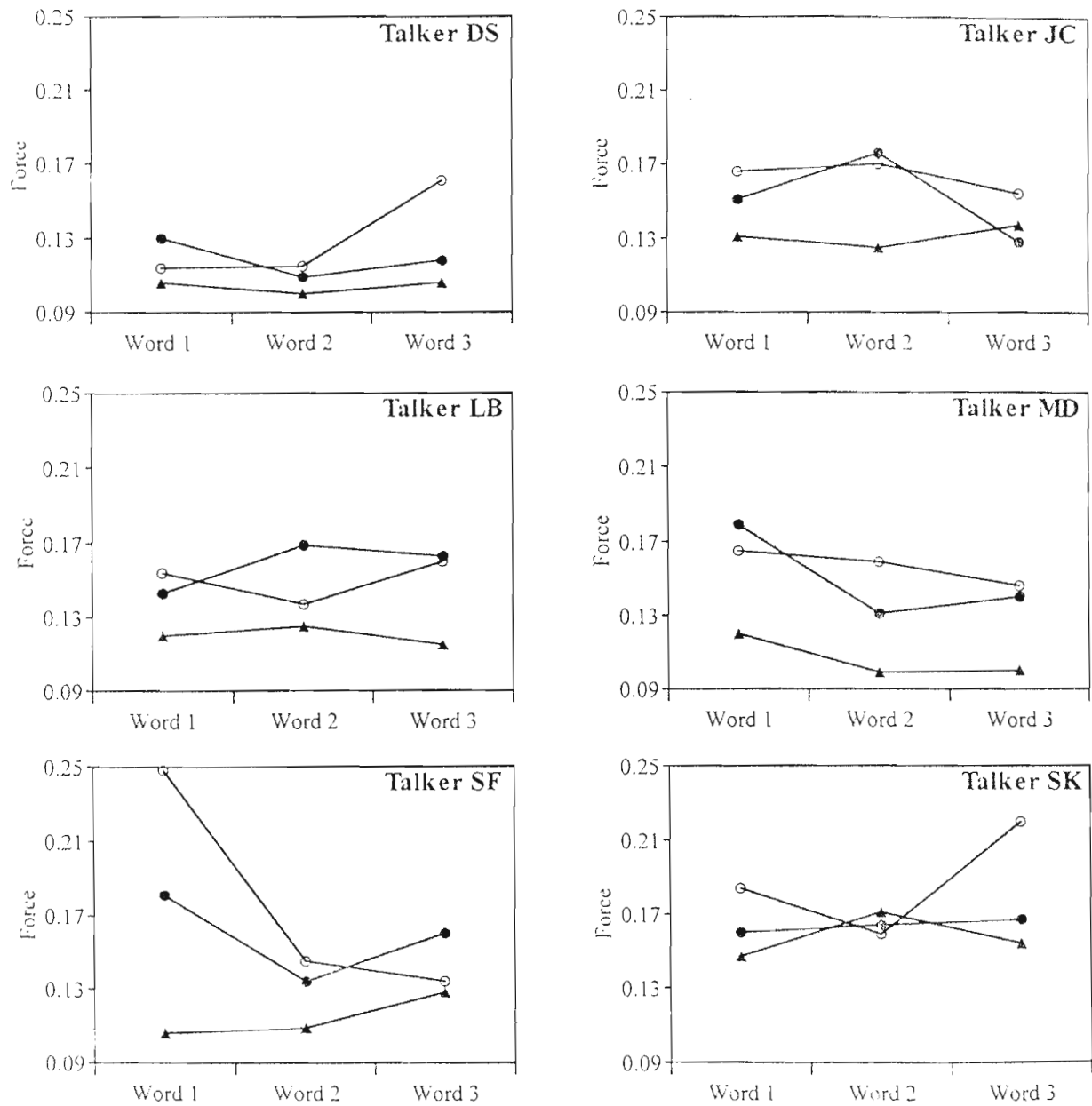
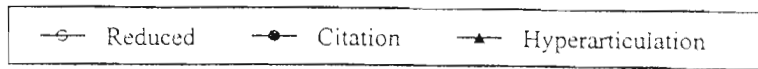


Figure 8: Force measures for each talker across conditions, split by sentence position

### Vowel Dispersion Across Sentence Position



#### FEMALES

#### MALES

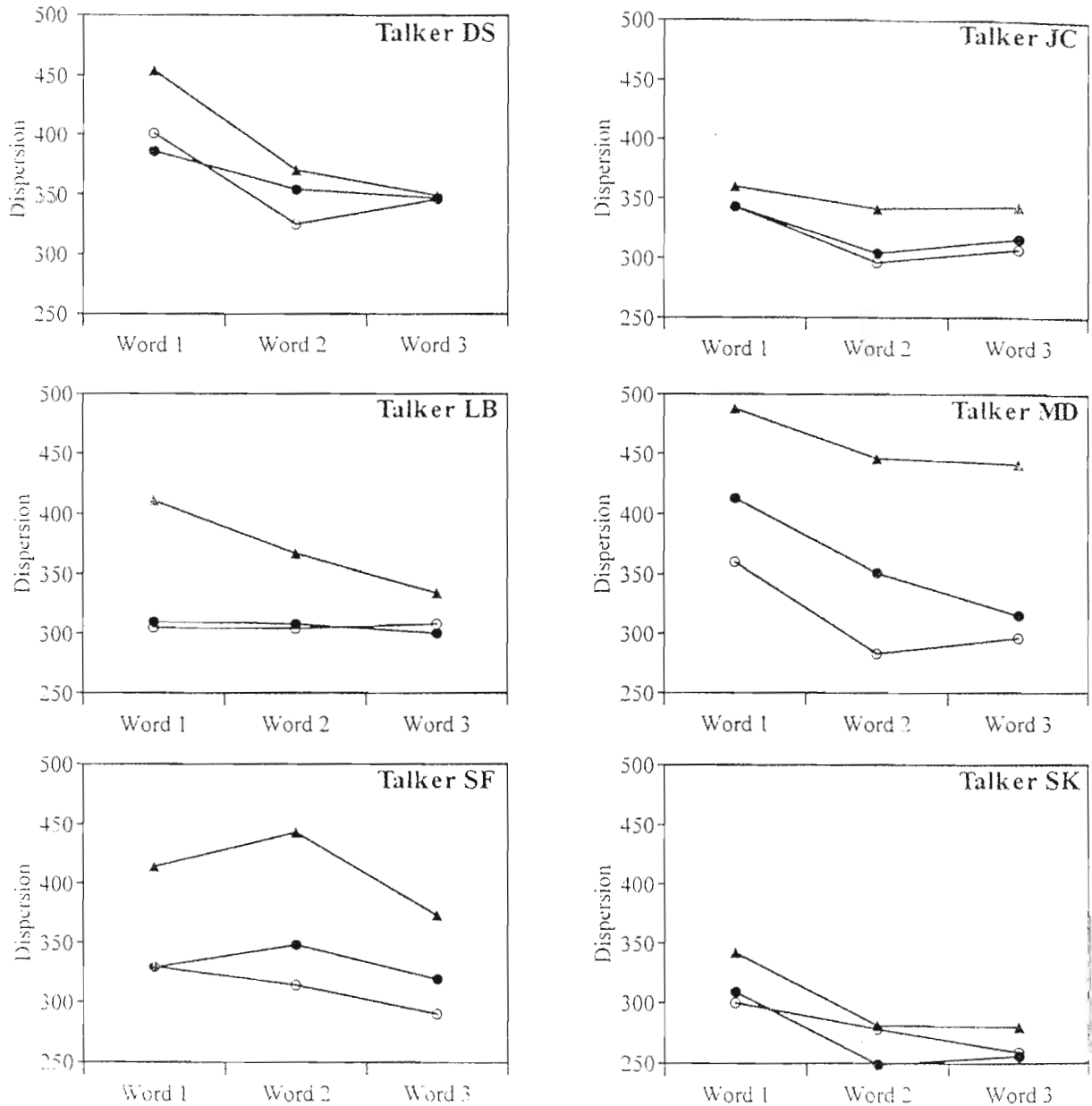


Figure 9: Vowel dispersion for each talker across conditions, split by sentence position

hyperarticulation and reduced conditions, the force measures were more extreme (lower in the hyperarticulation, higher in the reduced) in the first and last key words of the sentence, but the measures were very similar in all conditions in the middle of the sentence.

#### *By-Position Vowel Dispersion*

Figure 9 shows the by-position vowel dispersion measures for all six talkers across the three conditions. In almost all talkers and conditions, the dispersion measures decreased from the beginning to the end of the sentence, indicating that the talkers' vowel spaces became more compact as the sentence progressed. The only talker that did not follow this trend was SF, who had an increase in dispersion from word 1 to word 2 in the citation and hyperarticulation conditions, followed by a decrease in dispersion for word 3. Though dispersion generally decreased by position for all three conditions, there were no by-position effects that consistently distinguished any of the three conditions.

-----  
 Insert Figures 10 and 11 about here.  
 -----

#### *Vowel Space Plots*

Figures 10 and 11 show the F1 by F2 vowel plots of all six talkers in each of the three conditions. Some of the talkers, such as SK, demonstrated very little difference between the three conditions, while others, such as MD, showed fairly drastic differences across all three of the conditions. In general, however, the largest between-condition differences across the talkers were between the hyperarticulation condition and the other two conditions; the citation and reduced conditions generally produced similar vowel plots, with the exception of talker MD.

#### *Vowel Force and Dispersion Discussion*

The vowel space measures were the most consistent dimensions along which the talkers varied across all three conditions. All six talkers had lower force measures in the hyperarticulation condition than the other two conditions, and four of them had higher force measures in the reduced condition than the citation condition. Vowel dispersion measures demonstrated almost identical relationships between conditions. Although these results are generally consistent, more talkers need to be recorded and analyzed to determine if these differences are statistically significant and generalizable beyond this small sample.

## **Conclusions**

Overall, the acoustic measures seemed to accurately reflect the differences between the hyperarticulation condition and the other two speaking mode conditions. However, eliciting hyperarticulated speech was not the main focus of the study. Our main goal was to elicit casual speech in a laboratory environment. The results did not show consistent and significant differences between the citation and reduced conditions across all talkers. This does not indicate, however, a complete failure of the methodology. A few of the talkers showed significant differences between the reduced and citation conditions along certain dimensions, particularly the vowel force and dispersion measures. However, the magnitude of the differences between the citation and reduced conditions varied between talkers, and there were not enough talkers to determine if the differences were meaningful and generalizable.

Of the talkers who did show differences between the reduced and citation conditions, the most notable was talker MD, who produced significant differences in duration and the greatest differences in vowel force and dispersion between the two conditions. The sentences in the citation and reduced condition for MD were also subjectively very distinct, much more than any of the other talkers. Recall

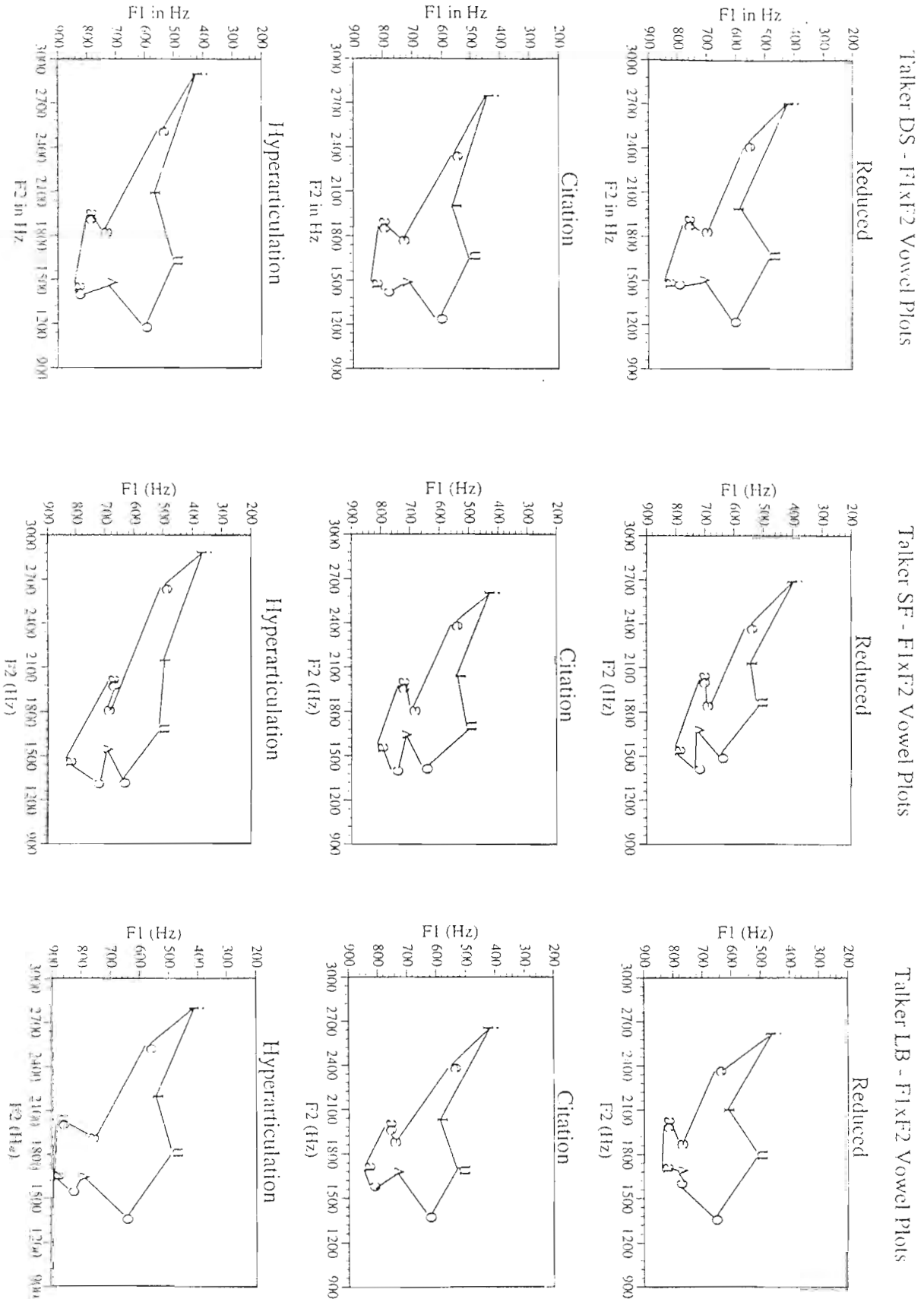


Figure 10: F1xF2 vowel space plots for talkers DS, SF and LB in each of the three conditions

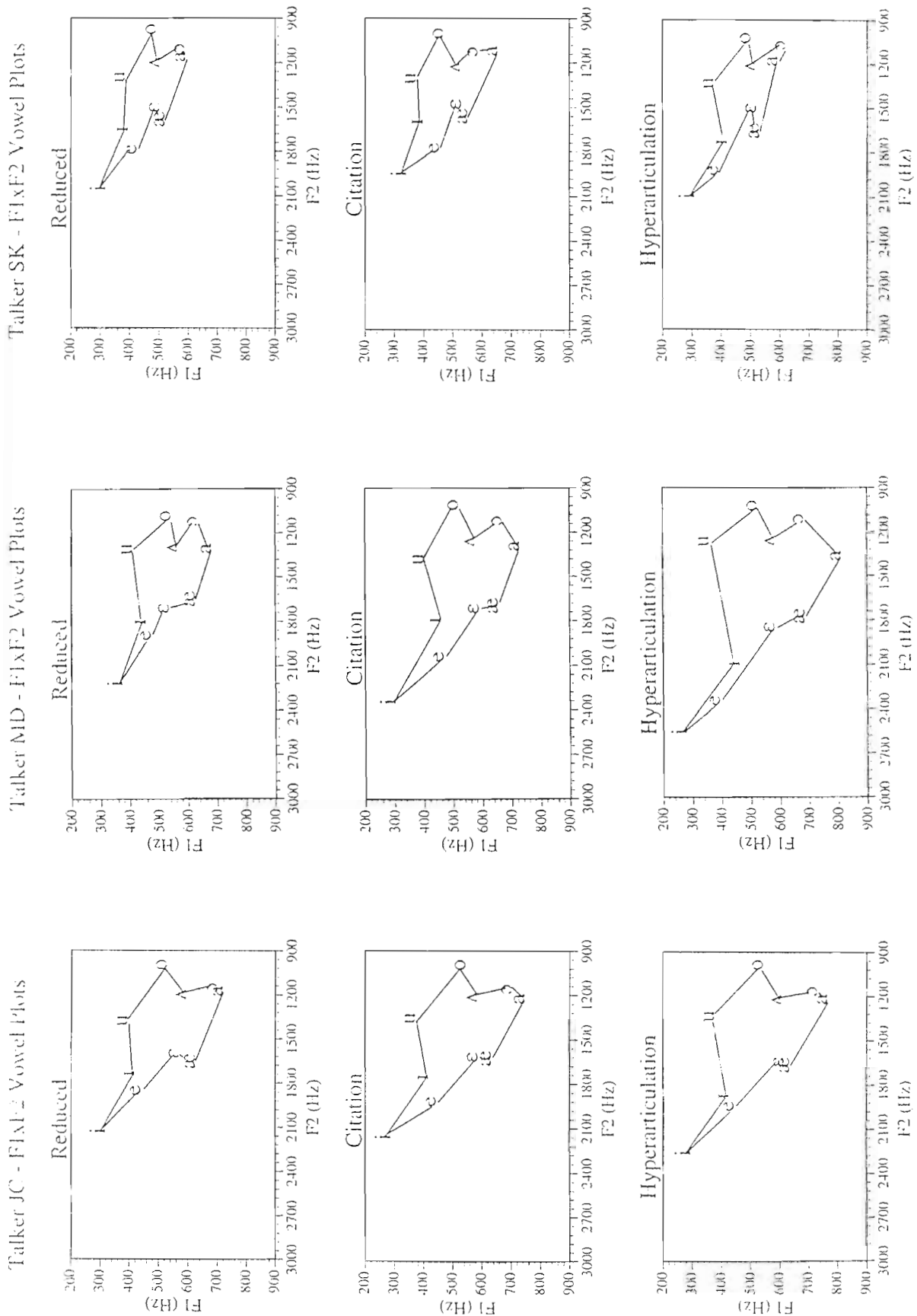


Figure 11: F1xF2 vowel space plots for talkers JC, MD and SK in each of the three conditions

that in the reduced condition, subjects were instructed to memorize a series of digits and recall them after reading the “distractor” sentence. On the sentences measured, MD recalled the correct series of digits on 21 out of the 46 sentences, or about 45% of the time. The other five subjects performed much better on the digit recall task, correctly recalling between 32 (70%) and 45 (98%) of the digit sequences correctly. Thus it seems that the length of the digit spans chosen, 5-7, was sufficient to make the task difficult for MD, but not for the other talkers. Therefore, the cognitive load was apparently not sufficient to elicit reduction for the majority of talkers measured in this study. Since MD was the only talker to demonstrate clear and consistent differences between the reduced and citation conditions, one would expect that if the other talkers had been given enough of a cognitive load, they may have demonstrated a similar separation of the reduced and citation conditions. In order to further test this hypothesis with the present data, additional analyses will be performed comparing the sentences for which the digits were correctly recalled to those for which the digits were incorrectly recalled.

Because the cognitive load we used in this experiment was not sufficient to elicit reduction for the majority of the talkers, slight changes will need to be made to the experimental methodology. One solution would be to simply increase the number of digits in the digit span task. This, however, might have the consequence of making the task too difficult for subjects; they may ignore the secondary task completely, and consequently have no cognitive load at all. A more effective solution would be to determine each subject’s digit span before the first part of the experiment. The length of the digit strings presented during the first part could then be adjusted to correspond to the digit spans of each individual subject, thus providing an appropriate level of difficulty of the task for each individual talker. We plan to implement this procedure in the next experiment on this project.

### References

- Bradlow, A.R., Torretta, G.M., & Pisoni, D.B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, **20**, 255-272.
- Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, **15**, 39-54.
- Dalby, J.M. (1984). Phonetic structure of fast speech in American English. Unpublished Ph.D. thesis. Indiana University.
- Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, **69**:3.
- Johnson, K., Ladefoged, P. & Lindau, M. (1993). Individual differences in vowel production. *Journal of the Acoustical Society of America*, **94**, 701-714.
- Kalikow, Stevens, & Elliot (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.
- Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, **48**, 839-862.
- Lindblom, B. (1990). Expelling variation: A sketch of H & H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling*. Pp. 403-439. Dordrecht: Kluwer.

- Ivily, S.E., Pisoni, D.B., Summers, W.V., & Bernacki, R.H. (1993) Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. *Journal of the Acoustical Society of America*, **93**, 2962-9273.
- Moon, S.J., & Lindblom, B. (1994). Interaction between duration, context and speaking style in English stressed vowels. *Journal of the Acoustical Society of America*, **96**, 40-55.
- Nygaard, L.C., & Pisoni, D.B. (1995). Speech perception: New directions in research and theory. In J.L. Miller & P.D. Eimas (Eds.), *Speech, language, and communication* (pp. 63-96).
- Picheny, M.A., Durlach, N.I., & Braida, L.D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, **29**, 434-446.
- Pisoni, D.B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J.W. Mullennix (Eds.), *Talker Variability in Speech Processing*. San Diego: Academic Press, Pp. 33-66.
- Stevens, K.N. (1996). Understanding variability in speech: A requisite for advances in speech synthesis and recognition. *Journal of the Acoustical Society of America*, **100**, 2634.
- Summers, W.V., 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.
- Uchanski, R.M., Choi, S., Braida, L. D., Reed, C.M., & Durlach, N.I. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech and Hearing Research*, **39**, 494-509.
- Wright, R. (1997). Lexical competition and reduction in speech: A preliminary report. In *Research on Spoken Language Processing Progress Report No. 21*. (pp. 471-486). Bloomington, IN: Speech Research Laboratory, Indiana University.

**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.  
Paul took a bath in the tub.  
The cabin was made of logs.  
Watermelons have lots of seeds.  
We swam at the beach at high tide.  
The boat sailed across the bay.  
The detectives searched for a clue.  
Please wipe your feet on the mat.  
The dealer shuffled the cards.  
The swimmer dove into the pool.  
The cigarette smoke filled his lungs.  
The bride wore a white gown.  
The nurse gave him first aid.

**APPENDIX 2:** Instructions to talkers**Part I:**

Welcome to the Speech Research Laboratory. We appreciate your participation in this experiment and also hope that you will find it interesting.

Today you will be participating in the first part of an experiment which will help determine the effects of distractor tasks on short-term memory. You will be presented with single-digit numbers, in sets of five to seven, and you will be given a few seconds to memorize these numbers. After this, a sentence will appear on the screen, which you must read out loud into a microphone. Finally, you will be prompted to repeat the set of numbers from memory. Do not feel discouraged if you are not always able to remember all of the numbers - our experiment is designed to be difficult; you should, however, do your best to remember the numbers as accurately as possible.

All of the numbers and sentences you say will be recorded by the computer and also monitored by the experimenter outside of the booth. You will be scored on remembering the correct numbers as well as the correct order of the numbers. Even if you are unable to remember the exact order of the digits, however, you should still repeat as many of the numbers as you can in any order.

This experiment should last from 45-60 minutes. Before you begin the experiment, be sure that the headset and microphone are snug but comfortable. We ask that you do not adjust the headset and/or microphone during the experiment, as this will cause the computer to make inaccurate recordings. If for any reason you would like to stop the experiment or need to leave the booth, alert the experimenter and the experiment will be stopped. If you have any questions, ask the experimenter now; if not, we will begin the experiment. Thanks again for your cooperation.

## **Part II:**

Welcome to the Speech Research Laboratory. Today you will be participating in the second part of our experiment. You will go through a series of trials and be asked to repeat the sentences which appear on the screen. You will be given the prompt: "Get ready for the next sentence" - after a brief pause, a sentence will appear on the screen, and your task is to read this sentence out loud into the microphone. Occasionally you will be prompted to read the sentence again more clearly; if so, be sure to concentrate on saying the sentence as clearly as possible. Unless specifically prompted for a clear reading, read each sentence casually.

This experiment should last from 45-60 minutes. Before you begin the experiment, be sure that the headset and microphone are properly adjusted and comfortable. We ask that you do not adjust the headset and/or microphone during the experiment, because this will cause the computer to make inaccurate recordings. If you have any questions, ask the experimenter now; if not, we are ready to begin the experiment. Thanks again for your cooperation.

## **Debriefing:**

Thank you for participating in this study. The purpose of the first part of this experiment was to obtain casual speech (normal, everyday, conversational speech) in a lab environment; when a subject is seated in a booth with a microphone and headset, however, they almost always speak more clearly and carefully. Since the subject has to concentrate on the memorization of numbers, they generally think the reading of the sentence is "unimportant" and therefore say it casually.

The sentences obtained in this part of the study will be compared with those obtained in the second part of the study, which was focused on obtaining careful speech. The recordings from both parts will be used a) to verify that this experiment was indeed able to obtain casual speech, and b) to see the effects of training with casual or careful sentences on a listener's ability to recognize a voice.