Interaction between duration, context, and speaking style in English stressed vowels

Seung-Jae Moon and Björn Lindblom

Citation: The Journal of the Acoustical Society of America **96**, 40 (1994); doi: 10.1121/1.410492

View online: https://doi.org/10.1121/1.410492

View Table of Contents: https://asa.scitation.org/toc/jas/96/1

Published by the Acoustical Society of America

JASA
THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA
Special Issue:
Additive Manufacturing and Acoustics
Read Now!

Interaction between duration, context, and speaking style in English stressed vowels

Seung-Jae Moon[a] and Björn Lindblom[b]
University of Texas at Austin, Austin, Texas 78712

(Received 21 December 1992; revised 5 January 1994; accepted 16 March 1994)

Acoustic observations are reported for English front vowels embedded in a /w—l/ frame and carrying constant main stress. The vowels were produced by five speakers in clear and citation-form styles at varying durations but at a constant speaking rate. The acoustic analyses revealed (i) that formant patterns were systematically displaced in the direction of the frequencies of the consonants of the adjacent pseudosymmetrical context; (ii) that those displacements depended in a lawful manner on vowel duration; (iii) that this context and duration dependence was more limited for clear than for citation-form speech, and that the smaller formant shifts of clear speech tended to be achieved by increases in the rate of formant frequency change. The findings are compatible with a revised, and biomechanically motivated, version of the vowel undershoot model [Lindblom, J. Acoust. Soc. Am. **35**, 1773–1781 (1963)] that derives formant patterns from numerical information on three variables: The "locus-target" distance, vowel duration, and rate of formant frequency change. The results further indicate that the "clear" samples were not merely louder, but involved a systematic, undershoot-compensating reorganization of the acoustic patterns.

PACS numbers: 43.70.Bk, 43.70.Fq, 43.70.Hs

INTRODUCTION

A. Vowel reduction

In the acoustic phonetic literature of the past few decades, "vowel reduction" is often described as a process of centralization since the formant patterns of reduced vowels tend to be displaced toward the center of the acoustic vowel chart. Vowel reduction as acoustic centralization has been observed in several languages. For instance, comparing American English vowels spoken in isolation with stressed and unstressed variants, Tiffany (1959) reported a decrease in the size of the acoustic vowel diagram. Unstressed vowels were closer to a "neutral, or at least central point" on the chart. In his classical (1948) monograph Joos stated that, on the average, consonants have a "centralizing effect" on English vowels "both vertically and horizontally." Stålhammar *et al.* (1973) and Karlsson (1992) compared Swedish vowels produced in isolation, in an /hVl/ context (V=vowel) and in read connected speech. The formant frequencies for unstressed vowels were displaced more toward a "neutral" vowel. For Dutch, Koopmans-van Beinum (1980) reported data on 12 vowels which were observed in isolation, in isolated words and under stressed and unstressed conditions in a read passage, in a short story retold, and in free conversation. All speakers revealed shrinking vowel spaces going from the isolated conditions to the free conversation. And there was centralization, that is a shift of each vowel toward the center of the vowel triangle. The degree of this shift differed for stressed and unstressed vowels.

Why reduced vowels tend to shift toward the center of the F_1/F_2 diagram has been explained along two major lines. The traditional view is that vowels that centralize do so because they become articulatorily more similar to schwa. For instance, Stetson (1951) observed that "in English it is possible to note a regular series of reduced "values" of the vowel which ends in schwa. With the increase in rate all vowels in unstressed syllables arrive at the common schwa."[1] The study of Stålhammar *et al.* (1973) provides some acoustic evidence for such a process. Significantly, the F_1 values of /ʉ/ and /ʊ/ (see their Fig. I-A-3) were higher under unstressed than stressed conditions. That would seem to indicate a change toward greater articulatory opening (cf. F_1=500 of a uniformly open vocal tract) rather than a consonant context effect which would tend to lower the F_1 of /ʉ/ and /ʊ/.

A second line of argument equates vowel reduction with contextual assimilation. The assumption is that, as a vowel assimilates (becomes articulatorily more similar) to its consonant environment, its acoustic properties approach those of the surrounding consonants. Thus if a place of articulation is characterized by formant frequencies that fall within the central range of the vowel formants, the "contextual assimilation" account predicts that the corresponding vowel formants should be displaced toward those consonant values. Accordingly they should centralize.

Lindblom (1963) investigated the formant patterns of eight Swedish short vowels embedded in /bVb/, /dVd/, and /gVg/ frames and varied in duration by the use of carrier phrases with different stress patterns. For both F_1 and F_2 the measurements showed examples of the so-called "undershoot" effect also described by Stevens and House (1963). That is, systematic shifts away from hypothetical target values were observed. A mathematical model was fitted to the data demonstrating that the magnitude of those displacements depended on only two things (i) the duration of the vowel and (ii) the extent of the CV formant transition (the "locus[2]-target" distance). The largest shifts were thus asso-

[a] Currently at English Department, Ajou University, Suwon, Korea.
[b] Currently at Department of Linguistics, Stockholm University, Sweden.

0001-4966/94/96(1)/40/16/$6.00

ciated with the short durations and the large "locus-target" distances. The data contained cases for which the centralization and the assimilation hypotheses made different predictions. For instance, in /bɩb/, /dɩd/, and /gɩg/, a movement toward schwa would make the F_1 of /ɩ/ higher, whereas assimilation would lower it. These and other similar cases provided no evidence for a change toward schwa.

Nord (1975, 1986) analyzed vowels in meaningful Swedish words having either /'CVCVC/ or /CV'CVC/ structure. In his experiment two of the manipulated variables were syllable position (initial or final) and stress (stressed or unstressed). Vowels would, or would not, undergo final lengthening depending on their position. They would furthermore be stressed or unstressed also depending on position. The results showed that, for comparable "locus-target" distances, vowel duration was not the only factor determining reduction: While vowels prolonged by final lengthening still showed more reduction than initial vowels, unstressed vowels showed formant displacements regardless of their duration. He observed that, while an unstressed vowel in initial position was coarticulated with adjacent consonants, in final position, its formants seemed to move, not toward its value for the adjacent consonants, but toward a schwa-like pattern. However, favoring a compromise position on the centralization versus assimilation issue, he concluded that, if the articulatory state after the /CVCVC/ sequence is assumed to approach a rest position, then final unstressed vowels could be said to have undergone assimilation. His account which is close to our own position suggests that reduction is contextual assimilation, but that centralization effects are to be expected if the adjacent context contains schwa-like elements.

Another issue raised by vowel reduction work concerns the relative roles played by duration and stress. Lindblom (1963) compared the results obtained by means of stress variations with those of a parallel experiment involving identical test syllables and tempo-controlled vowel durations. Finding similar undershoot effects, he concluded that duration seemed to be the primary determinant of vowel reduction. However, the Nord studies demonstrate a certain duration independence of reduction effects.

There are several other studies that contradict Lindblom's claim about the primacy of duration. In data from four speakers, Gay (1978) found no undershoot effects when vowels became shorter at a faster speaking rate. Kuehn and Moll (1976) investigated articulatory behavior at different speaking rates. VC and CV trajectories showed a tempo-dependent reduction of transition times, but there were interspeaker differences: Whereas some talkers reduced movement velocity as the speaking rate became faster—thereby producing decreases in articulatory movement and hence undershoot—others achieved less undershoot by increasing movement velocity. Similar findings were reported by Flege (1988). In "glossometry" experiments he found that undershoot was more closely related to peak velocity of tongue movement than to duration. He argued that style of movement, rather than timing per se, is the important variable. In Engstrand's (1988) x-ray study there is additional evidence that the timing and the style of movement can be controlled independently. The measurements exhibited formant differences between stressed and unstressed contexts, but showed no formant displacement as a function of tempo-induced durational variation. The author attributed the absence of undershoot to the possibility that the subject had used greater articulatory "precision" under the fast and stressed speaking conditions. In Engstrand and Krull (1989), data are presented from a sample of spontaneous Swedish. The extent of the F_3 transition was measured in the vowel of /vi:s/ when the sequence occurred in the adverbial phrase "på något vis" ("in a way") and when it occurred as a content word in other phrases. When plotted against the duration of /i:/ (their Fig. 2), both sets of data show a clear duration-dependent undershoot effect, but, in the adverbial phrase, the extent of the transition is reduced. In other words, a more elaborate pronunciation style was used when the vowel occurred in the, presumably, semantically more significant content words.

Van Son and Pols (1990, 1992) examined "static" and "dynamic" aspects of the formant patterns of seven Dutch vowels read at normal and (maximally) fast rates by an experienced male newscaster. Vowels were 15% shorter in fast speech. Median formant values were not very different at the two tempos. The average F_1 was higher in the fast rate implying that more open vowel articulations may have been used. No significant differences were observed for F_2. The correlation between the duration and the formant values was so small that duration was dismissed as a determinant of formant variations. Durations changed with speaking rate but without accompanying formant changes. In the (1992) study, formant track dynamics were examined by measuring formants at 16 temporally equidistant points during a vowel. Once a time normalization had been performed, the shape of the vowel formant tracks turned out to be near invariant despite the tempo-dependent changes of vowel duration.

As pointed out by Fourakis (1991), the early work on formant undershoot has been frequently interpreted as stating that "as the syllable (hence, vowel nucleus) decreased in duration, the vowel became reduced, or more schwa-like in character (Miller, 1981)" But, according to the undershoot model, no formant displacements would be expected for adjacent vowels and consonants with identical, or closely similar, formant values, and, accordingly, undershoot presupposes both short duration and a sizable "locus-target" distance.

Fourakis (1991) undertook an investigation whose focus was the roles of tempo and stress as determinants of vowel reduction in American English. Nine vowels were examined in two environments, /hVd/ and /bVd/, chosen because they were assumed "to interfere minimally with vowel articulation" (p. 1825). The results revealed that both stress- and tempo-induced formant shifts were minimal.

That result is fully compatible with the predictions of the duration- and context-dependent undershoot model in Lindblom (1963). Conceivably then, the lack of substantial duration-dependent formant displacement effects reported by Gay (1978), van Son and Pols (1990, 1992), Engstrand (1988), and Fourakis (1991) could be associated with the fact that, by and large, the test syllables had transitions which covered primarily moderate "locus-target" distances. However, that account must be rejected in the case of the data sets from Nord (1986) and Engstrand and Krull (1989) which both demonstrate the phenomenon of undershoot, but, importantly also the possibility of different degrees of undershoot at given vowel durations. The Lindblom (1963) model also ignores the fact that undershoot in articulatory movements depends on how a speaker decides to style a given movement, for instance with respect to its velocity (Kuehn and Moll, 1976; Flege, 1988).

In the preceding review we have made two distinctions: That between phonological reduction (which changes unstressed English vowels to schwa as in telegraphy versus telegraphic) and two types of phonetic reduction: "centralization" (vowels becoming more schwa-like) and "contextual assimilation" (vowels becoming more influenced by context). The present study is concerned with phonetic reduction and is focused on "formant undershoot" investigated in the context of clear speech.

B. Previous work on "clear speech"

Previous work on "clear speech" reveals that its acoustic characteristics are measurably different from those of "normal speech." Everyday observation indicates that speakers are in general able to adapt their style of talking to the needs of the situation. To be heard in noise, speakers automatically increase their vocal effort, a phenomenon known as the "Lombard effect" (Hanley and Steer, 1949; Draegert, 1951; Lane and Tranel, 1971). People may adopt a style sometimes referred to as "Foreignese" (Freed, 1978) when speaking to non-native speakers with limited comprehension skills. "Baby Talk," or "Motherese" is the "simplified register" (Ferguson, 1977) that mothers use in communicating with infants.

Chen and collaborators used a feedback method to elicit clearly spoken syllables (Chen, 1980; Chen *et al.*, 1983). They reported that one of their two subjects improved his intelligibility significantly when speaking clearly and showed an expanded vowel space and a tighter clustering within that space for each vowel category. In addition, in the clear speech, the tempo was slower and formant transition durations were increased. Both subjects showed more distinct VOT distributions for voiced and voiceless consonants. Similar findings were obtained by Moon *et al.* (1988) who observed that in clear speech vowels were longer and voiced and voiceless VOT values were more widely separated. Picheny *et al.* (1986) pointed out that clear speech tends to exhibit less vowel reduction, and that lax vowels are more susceptible to speaking mode than tense vowels. They also found that all word-final stops were released in clear but not in conversational speech and that a lengthening of all segments occurred in clear speech. In a study of segment durations, Moon (1990) observed more releases of word-final stops and a general prolongation of segments for clear tokens.

Ladefoged *et al.* (1976) studied six monosyllabic /bV/ words produced in seven different speaking styles varying from "completely informal conversation" to list reading. No systematic vowel formant differences were reported as a function of speaking style. Summers *et al.* (1988) found increases in amplitude, duration and pitch as well as formant displacements in speech produced in noise and, thus presumably, in a "clear" mode. A perception test indicated that, although normalized in terms of S/N ratio, the speech produced in noise was the more intelligible.

Uchanski *et al.* (1985) examined phoneme and pause duration in clear speech. They found that clear speech showed longer durations and that its articulation rate (syllables per second excluding pauses) was reduced by roughly a factor of 2.

Using a contrastive sentence frame "I said X not Y," Clark *et al.* (1987) examined clear-speech transforms in terms of suprasegmental dimensions. Their clear speech samples were louder, had higher pitch and had longer sentence duration. They also noticed that the X word was more stressed and was more peripheral than the Y word in vowel quality. However, they failed to find voicing contrast enhancement in stops.

C. Purpose

The specific aim of this study is to collect data on formant undershoot in an experiment in which the variables identified as the most important determinants of vowel reduction are carefully controlled. One of the reasons for choosing "clear speech" was the assumption that, in enunciating carefully, speakers make systematic changes in their pronunciation and try to speak more intelligibly. Would there in fact be such changes? If so, what would they be like acoustically? Would formant undershoot be observed? If present, would formant undershoot be different for clear and citation-form speech? If dependent, or independent, of style, how would the observed formant undershoot patterns bear on the issues just reviewed? These are the questions to be answered by the present study.

I. METHOD

A. Subjects

The subjects were five male students without known speech or hearing disorders whose dialects showed no marked regional characteristics but conformed with general American English. They were recruited from an undergraduate class. They were not informed of the goal of the experiment.

B. Speech samples

In the design of the speech materials an attempt was made to meet the following conditions:

(i) Test words should exhibit maximally large consonant-vowel formant transitions; (ii) the test vowels should carry the same degree of stress, preferably primary lexical stress; (iii) vowel duration should vary systematically over a sufficiently large range.

The first objective was reached by selecting sequences with front vowels occurring between labio-velar and velarized consonants, as in *wheel*, *will*, *well*, and *wail*. These words are characterized by a back-front-back tongue body movement and their spectrographic patterns typically display large formant excursions. Huang (1991) found that in En-

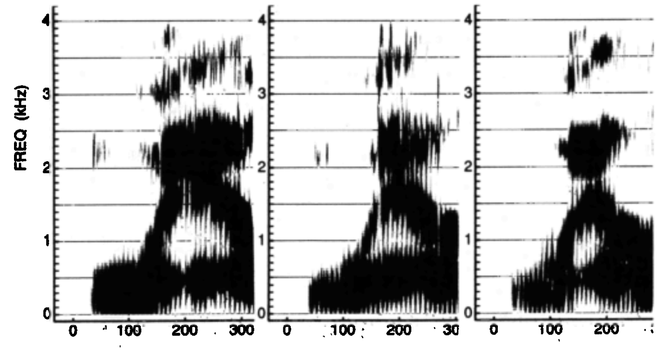

FIG. 1. Spectrograms of the initial syllables of will, will(ing), and Will(ingham).

glish /w/, /r/, and /l/, the F_2 of front vowels, especially lax front vowels, was considerably lowered as compared with their frequencies in /b/, /d/, and /g/ contexts. Our choice was motivated by the assumption that the greater a formant movement between a consonant and a vowel, the more sensitive it is likely to be as an indicator of "target undershoot" (Stevens and House, 1963), should that phenomenon occur. Another reason for the selection of the /wVl/ frame was that it provides an acoustically almost symmetrical CVC environment. Both /w/ and /l/ are characterized by low F_2 values (Lehiste, 1964). In the presentation and the analyses of the data below, we have introduced the simplification of quantifying context in terms of the F_2 of the preceding consonant, leaving the question of the right-to-left or left-to-right effects of coarticulation (Ohde and Sharf, 1975) unaddressed for the moment.

Condition two and three were satisfied by using the so-called "word-length effect." In English and other languages this is the durational shortening observed in the stressed vowel of an initial syllable as more and more unstressed syllables are added, as in e.g., *speed*, *speedy*, *speedily* (Ladefoged, 1975; Lehiste, 1974, 1975). The use of different word lengths insured that durational variations would be achieved neither by speakers subjectively controlling their speaking tempo (cf. Gay, 1978), nor by means of stress differences in the test words (cf. Lindblom, 1963). Thus with the present design, stress was ruled out as determinant of undershoot. (The trisyllabic words had to be invented. Subjects were asked to think of them as place or family names.[3])

In Fig. 1 the initial syllables of *will*, *willing*, and *Willingham* are illustrated. Note the wide excursion of F_2 (and, to some extent of, F_3). As vowel duration decreases the turning point of the parabola-like F_2 movement is shifted down in frequency.

In addition to the experimental productions, four more words were recorded in citation-form style: *heed*, *hid*, *head*, and *hayed*. The /hVd/ frame was used as a "null-context" condition (cf. Stevens and House, 1963). In the quantitative analyses to be reported below, they served as estimates of formant target values. All test words were spoken in isolation.

C. Procedures

The subjects were seated in front of a cardioid microphone placed 25 cm from the speaker's lips. The recording was done in the soundproof booth of the UT Phonetics Laboratory. The speech signal was recorded into the left channel of a Tandberg TD20A reel-to-reel tape recorder at 3 $\frac{3}{4}$ ips. The input level was continuously monitored and adjusted to obtain a maximum gain without exceeding the dynamic range of the recording system. Written records were continually made of the degree of attenuation introduced. On the average the clear speech (CS) tended to be 3–5 dB more intense than the citation form (CF).

Three blocks of recording sessions were arranged for each subject: One for the null-context words, one for the test words produced as "citation forms", and one for the same items spoken in the style of "clear speech," to be defined below. All recording sessions were monitored through an intercom by the experimenter in an adjacent room. Before the actual recording, a short practice session occurred to familiarize the subject with the materials.

Five instances of all test words were written on separate 3×5 index cards. This yielded 20 tokens for the /hVd/ words, and 60 tokens each for the CF and CS conditions. Cards were randomized. A different order was used for the CF and CS recordings.

For the null-context and CF recordings, subjects were asked to read the items on the cards at a comfortable rate and loudness. They were told to pause between tokens so that the noise made by flipping the cards would not interfere with the recorded signal. Otherwise, no specific instructions were given.

For the CS condition, the procedure was the same, except that subjects were asked to "overarticulate," that is to read the words as clearly as they could. This instruction was given in spoken and written form. They were asked to read the instruction aloud. During the CS recording, at unpredictable moments, the subject was interrupted through the intercom by the experimenter who would pretend that the token just pronounced had not been understood, and would ask for a repetition. This procedure had the effect of helping subjects maintain the clear mode throughout this condition. The tokens repeated at the request of the experimenter will be called "extra clear" tokens. For each subject 72 CS tokens were recorded (five repetitions of 12 test words plus one extra clear exemplar of each test word).

D. Data analysis

The audio signals were analyzed in four steps: Digitization, waveform editing, printouts of spectrograms, and spectral displays at selected time points in the waveforms. These analyses were performed using the MIT SpeechVax software originally developed by Dennis Klatt and adapted for the UT Phonetics Lab by Gerald Lame.

The audio signals were fed into a VaxStation II/GPX using a 10-kHz sampling rate and low-pass filtering at 4 kHz. The speech samples were examined with a waveform editor. A separate waveform file was created for each utterance. Wideband spectrograms were produced by computing, every 3.2 ms, a 256-point discrete Fourier magnitude spectrum using a 6.4-ms Hamming window.

The spectrograms were carefully inspected to determine time points for spectral analysis. At least four points were picked for each waveform. They included the middle of the /w/ constriction, the maximum F_2 frequency of the vowel, and a sample in the /l/ segment.

For each time point two spectral displays were produced. Both included harmonic spectra obtained from 512-point discrete Fourier transforms (DFTs). A selected portion of the waveform is first preemphasized (by representing it in terms of the first differences between successive samples, first difference coefficient=0.85) and then multiplied by a 299-point Hamming window. In one display, called the broadband spectral envelope display, an envelope produced by averaging harmonics within a 300-Hz window was superimposed on the DFT. The other display, the LPC spectral display, showed the DFT and an envelope derived from an LPC analysis (14 coefficients) of the preemphasized waveform.

E. Formant frequency and vowel duration measurements

The first three formants were measured at the steady states of the /hVd/ vowels. In the CF and CS tokens, measurements of the first three formants were obtained at two points: One at the maximum constriction of /w/ and the other at the frequency maximum of the second formant. To determine formant values, three spectral representations of identical time points were compared: A harmonic spectrum (DFT), and LPC analysis and a broadband spectral envelope. Formant estimates based on LPC analysis are sensitive to the configuration of harmonics within a formant as well as to various spectral irregularities. Therefore they were used only as supplementary information. The DFT displays were the primary source of formant frequency estimates. The spectrogram which displays the temporal continuity of formants most clearly, was examined to confirm the values derived by means of these three methods. For each vowel and each speaker, the average formant value of five repetitions was calculated. "Extra-clear" tokens were pooled with the other CS tokens, since they did not differ significantly in either durations or formant frequencies.

Vowel duration measurements turned out to be more problematic. As the test syllables consisted only of sonorant sounds, spectrographic boundaries clearly delimiting the vowel segment were not evident. An *ad hoc* method was used. For each speaker and vowel, the second formants in all tokens were traced onto a single sheet of paper. This included items from both styles and all word lengths. After a brief segment of initial coincidence, these tracings tended to diverge. The frequency of this point of divergence was chosen as the criterion for the /w/-vowel boundary. Analogously, the point of formant convergence was chosen as the vowel-/l/ boundary. These frequency thresholds will be called "crossover" values.

For two speakers the same crossover was used throughout the measurements. For the others, different values were selected for the initial and final boundaries and for the different vowel categories. However, for each individual vowel of any given speaker, the crossover criteria were kept constant and independent both of speaking style and word length. While this method did not allow us to compare vowel durations across speakers, it did enable us to make within-speaker durational comparisons across styles and word lengths.

II. RESULTS

A. Formant displacement: Direction and magnitude

This section presents measurements of vowel duration and second formant frequency values. In Table I the rows represent the individual speakers. The numbers in parentheses pertain to the CS condition. D is vowel duration averaged, for each speaker and style, across three word lengths and five or six repetitions (that is, over at least 15 tokens). As mentioned earlier, the criterion for measuring vowel durations was independent of style and word length, but not of speaker and vowel identity. Consequently, durations can only be compared within speakers and vowels, not across them. ΔD is the difference in duration between each speaker's clear and citation form values: D(CS)-D(CF). CS durations are larger, especially for tense /i/ and /eɩ/.

The column labeled F_{2T}^* is an average of five measurements from the /hVd/ context. F_{20} is a mean value computed for each speaker and style across all word lengths and repetitions (⇒15). Also listed are $F_{20} - F_{2T}^*$, and F_{2L}, the average F_2 of /w/, which was calculated analogously with D and F_{20}. t-tests were performed to ascertain whether the two $F_{20} - F_{2T}^*$ values were statistically significant or not. The # symbol indicates a level of significance better than $p=0.05$ and ## corresponds to $p<0.001$.

The information of Table I is displayed graphically in Fig. 2 in which $(\bar{F}_{20} - \bar{F}_{2T}^*)$, the grand average across all speakers, is plotted as a function of vowel for the two styles. As can be seen, this measure is consistently negative. Picturing that finding in terms of the shape of an average F_2 contour, we realize that F_2 begins low in /w/ and then rises, but does not reach as high as in /hVd/, before it changes direction and begins its descent for /l/ (cf. Fig. 1). If, following Stevens and House (1963), we take F_{2T}^* to be an estimate of the "context-free" value of F_2, the F_2 "target," then the fact that $(\bar{F}_{20} - \bar{F}_{2T}^*)$ exhibits large negative numbers, implies that there is a strong "undershoot" effect in the present data. From Fig. 2 and Table I, we see that, in terms of this measure of formant displacement, there is undershoot for all vowels and consistently more of it for CF and for CS.

The above statements describe trends characteristic of the present speakers as a group, but they are true also for speakers individually as demonstrated in Fig. 3. It compares CF (x axis) and CS (y axis) in terms of $|F_{20} - F_{2T}^*|$ for each speaker. There are four data points per speaker corresponding to the four vowels. The CS data fall below the line of identity except for one case, the /i/ of speaker R. Speaker W deviates from the general pattern in showing exceptionally small $|F_{20} - F_{2T}^*|$ values for CS. As will appear from Table III and the curve fitting exercise below, this is due to W's use of higher F_2 targets in CS. For the rest of the speakers, F_2 displacement for CS is roughly 60% of that for CF.

TABLE I. Average vowel durations and F_2 data for /i/, /ɩ/, /ɛ/ and /eɩ/ from five speakers. The rows represent the individual speakers. The numbers in parentheses pertain to the CS condition. D is vowel duration averaged, for each speaker and style, across three word lengths and five or six repetitions. ΔD is the difference in duration between each speaker's clear and citation form values: D(CS)-D(CF). F_{2L} is the average F_2 of /w/. The F_{20} column contains the average of at least 15 measurements from the /wVl/ context computed for each speaker and style across all word lengths and repetitions. F_{2T}^* is an average of five repetitions of /hVd/.

i	D	ΔD	F_{2L}	$F_{20}-F_{2T}^*$	F_{20}	F_{2T}^*
D	169 (222)	53	638 (634)	−200 (−180)	2224 (2244)	2424
G	174 (211)	37	652 (706)	−94 (−64)$^{\#}$	2083 (2113)	2177
R	174 (201)	27	636 (657)	−45 (−128)$^{\#}$	2177 (2094)	2222
S	163 (256)	93	507 (559)	−322 (−196)$^{\#\#}$	1981 (2107)	2303
W	138 (201)	63	620 (619)	−334 (−11)$^{\#\#}$	1801 (2124)	2135
			Grand average	−199 (−116)	$t=4.96$ ($p<0.001$)	
			Std.	130 (78)		
ɩ	D	ΔD	F_{2L}	$F_{20}-F_{2T}^*$	F_{20}	F_{2T}^*
D	85 (104)	19	638 (634)	−502 (−326)$^{\#\#}$	1543 (1719)	2045
G	112 (117)	5	652 (706)	−305 (−205)$^{\#}$	1560 (1660)	1865
R	83 (115)	32	639 (657)	−631 (−428)$^{\#\#}$	1301 (1504)	1932
S	100 (160)	60	507 (559)	−676 (−396)$^{\#\#}$	1301 (1581)	1977
W	67 (110)	43	620 (619)	−438 (−80)$^{\#\#}$	1322 (1680)	1760
			Grand average	−510 (−287)	$t=9.29$ ($p<0.001$)	
			Std.	150 (144)		
ɛ	D	ΔD	F_{2L}	$F_{20}-F_{2T}^*$	F_{20}	F_{2T}^*
D	103 (121)	18	638 (634)	−343 (−238)$^{\#}$	1545 (1650)	1888
G	142 (151)	9	652 (706)	−198 (−130)$^{\#}$	1517 (1585)	1715
R	112 (128)	16	639 (657)	−309 (−244)$^{\#}$	1506 (1571)	1815
S	156 (224)	68	507 (559)	−524 (−268)$^{\#\#}$	1187 (1443)	1711
W	73 (138)	65	620 (619)	−373 (−8)$^{\#\#}$	1242 (1607)	1615
			Grand average	−349 (−178)	$t=9.66$ ($p<0.001$)	
			Std.	118 (109)		
eɩ	D	ΔD	F_{2L}	$F_{20}-F_{2T}^*$	F_{20}	F_{2T}^*
D	195 (248)	53	638 (634)	−331 (−177)$^{\#\#}$	2021 (2175)	2352
G	194 (229)	35	652 (706)	−294 (−140)$^{\#\#}$	1888 (2042)	2182
R	159 (202)	43	639 (657)	−126 (−110)	2089 (2105)	2215
S	155 (264)	109	507 (559)	−405 (−192)$^{\#\#}$	1745 (1958)	2150
W	142 (224)	82	620 (619)	−422 (−89)$^{\#\#}$	1542 (1875)	1964
			Grand average	−316 (−142)	$t=12.68$ ($p<0.001$)	
			Std.	118 (43)		

In terms of their frequency extent, the formant shifts reported here are large compared with the findings of other investigations. Values of several hundred Hz are seen in Table I, those of the lax vowels being particularly large. Although there are individual differences in this respect, large formant displacements nonetheless occur for all speakers.

For each speaker and vowel, an analysis of variance was performed with formant displacement ($n\Rightarrow 15$) as the dependent variable and speaking style and word length as the independent two- and three-level variables. These analyses are summarized in Table II. They show highly significant effects for both style and word length for the lax /ɩ/ and /ɛ/ of all five speakers. Tense vowel tests resulted in fewer, but nevertheless a majority of significant cases. No significant interactions were seen except for the /ɛ/ of G and S and the /i/ of S and W ($p<0.05$).

Summarizing the results presented so far, we have demonstrated undershoot effects of considerable magnitude for all speakers and for all vowels in both clear and citation-form speech. Moreover, there is significantly less undershoot in clear than in citation-form tokens for all speakers.

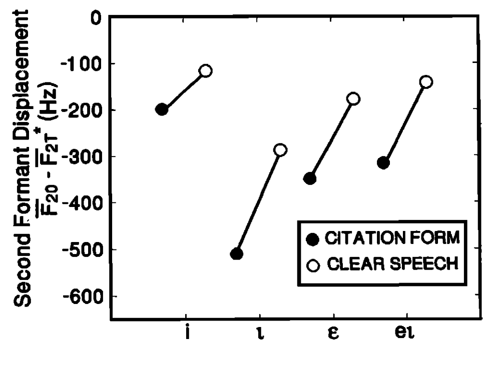

FIG. 2. Contextually induced displacement of the second formant. The measure along the ordinate is the difference between F_2 in the /wVl/ context and in the corresponding /hVd/ words. The $\bar{F}_{20}$ and $\bar{F}_{2T}^*$ values are grand averages of data from five speakers. For each speaker, F_{20} was sampled at the point of zero rate change in the vowel and an average was computed over at least 15 tokens (three word lengths and five or six repetitions) of each vowel. Those averages were then used to obtain F_{20}. Similarly, for each speaker, F_{2T}^* was measured in five repetitions of each vowel in /hVd/ words. The averages from the five speakers were then combined to form an $\bar{F}_{2T}^*$ value for each vowel.

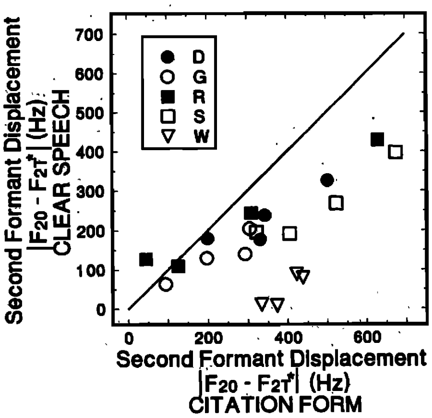

FIG. 3. Second formant displacement in clear (CS, y axis) and citation-form speech (CF, x axis). Each data point refers to a given vowel from a given speaker. Both axes show $|F_{20} - F_{2T}^*|$. This measure was obtained for each speaker and vowel by means of the averaging procedures described in Fig. 2.

B. Formant undershoot and duration

Above it was noted that CS vowel durations were longer, and that CS displayed less average undershoot than CF. In this section the relationship between duration and formant displacement is examined in a more detailed manner. [Figure 4(a)–(e) present the raw data on F_{20} and vowel duration for all speakers and vowels and for both styles. The diagrams all have the same format. F_{20}—which refers to F_2 sampled at the point of zero rate of change in the vowel—is plotted against duration for each token. Each plot has 15 CF points (solid circles) and 15 to 18 CS points (open circles). An arrow labeled /w/ indicates the average F_2 value of that segment (F_{2L} of Table II). The arrow marked /hVd/ shows F_{2T}^*, the "context-free," or "null-context," value of F_2. For the diphthong /eι/ in the /hVd/ environment, there are two arrows: One for the vowel nucleus and the other for the glide. Note that, for the /eι/ of the /weιl/ sequences, the same criterion was used as for the other vowels: The frequency was measured only at the F_2 maximum.]

In most of the panels, we see an increase in F_2 as durations get longer, or, equivalently, in shorter vowels F_2 is displaced further and further from the /hVd/ value and approaches that of /w/ more and more closely. This displacement is evident for both CS and CF. There is also another pattern with data points forming a more or less horizontal cluster with considerable overlap between CF and CS measurements. Examples are the /i/ data of speakers D, G, and R. Descriptively, we can say that, in the former—but not in the

TABLE II. Results of an analysis of variance with formant displacement as the dependent variable and speaking style (ST) and word length (WL) as the independent variables. The degrees of freedom for ST and WL are (1,26) and (2,26), respectively.

	i		ι		ε		eι	
speaker: D	ST	WL	ST	WL	ST	WL	ST	WL
F ratio	1.32	4.04	69.70	34.97	33.14	30.48	90.87	3.81
P	0.2657	0.0343	0.0001	0.0001	0.0001	0.0001	0.0001	0.0355
	i		ι		ε		eι	
speaker: G	ST	WL	ST	WL	ST	WL	ST	WL
F ratio	3.84	0.72	22.68	65.48	16.15	30.35	47.23	2.61
P	0.0609	0.4979	0.0001	0.0001	0.0004	0.0001	0.0001	0.0929
	i		ι		ε		eι	
speaker: R	ST	WL	ST	WL	ST	WL	ST	WL
F ratio	2.92	0.13	98.09	100.5	7.90	6.74	0.67	74.93
P	0.0993	0.8810	0.0001	0.0001	0.0091	0.0042	0.4213	0.0001
	i		ι		ε		eι	
speaker: S	ST	WL	ST	WL	ST	WL	ST	WL
F ratio	21.03	7.39	68.84	15.88	151.1	11.44	70.47	2.51
P	0.0001	0.0029	0.0001	0.0001	0.0001	0.0003	0.0001	0.1004
	i		ι		ε		eι	
speaker: W	ST	WL	ST	WL	ST	WL	ST	WL
F ratio	246.1	49.63	188.5	33.60	421.4	37.01	194.1	7.31
P	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0030

latter—cases, F_2 shows duration dependence.

The horizontal pattern appears especially for /i/ but also for /eɩ/ which are both tense vowels. Lax /ɩ/ and /ɛ/, however, show considerable duration-dependent formant displacement. A similar observation was made by Picheny *et al.* (1986) who found that, in conversational speech, the formant frequencies varied more in lax than tense vowels.

Secondly, the degree of undershoot is talker specific. Some speakers show very strong effects, whereas others show a smaller trend. Speaker W shows particularly well-

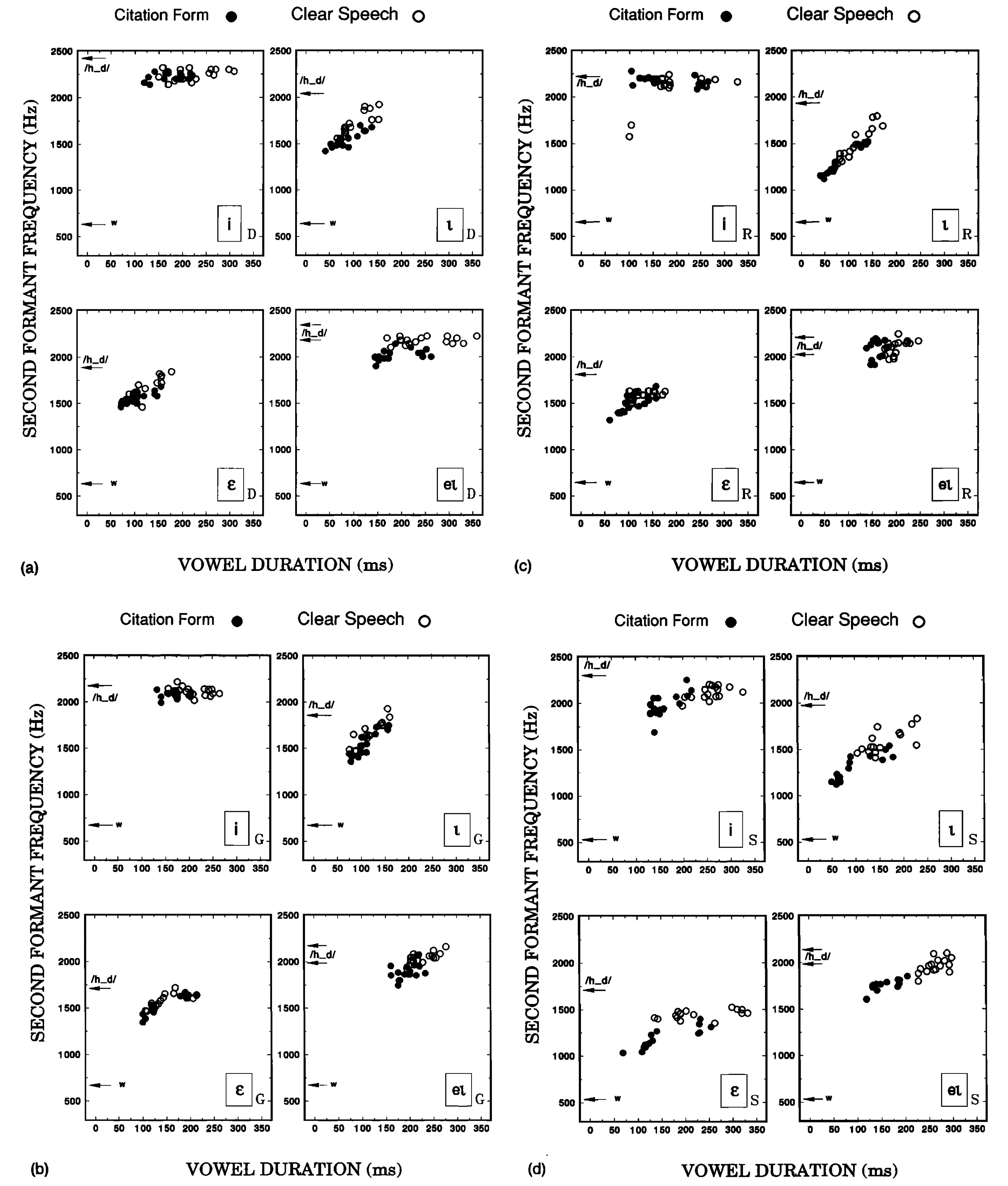

FIG. 4. (a)–(e) Second formant data plotted against vowel duration for four vowels as produced by five speakers. Arrows indicate F_2 positions in the /h—d/ context and in the /w/ preceding the vowel in the test word. Solid circles: CF, open circles: CS.

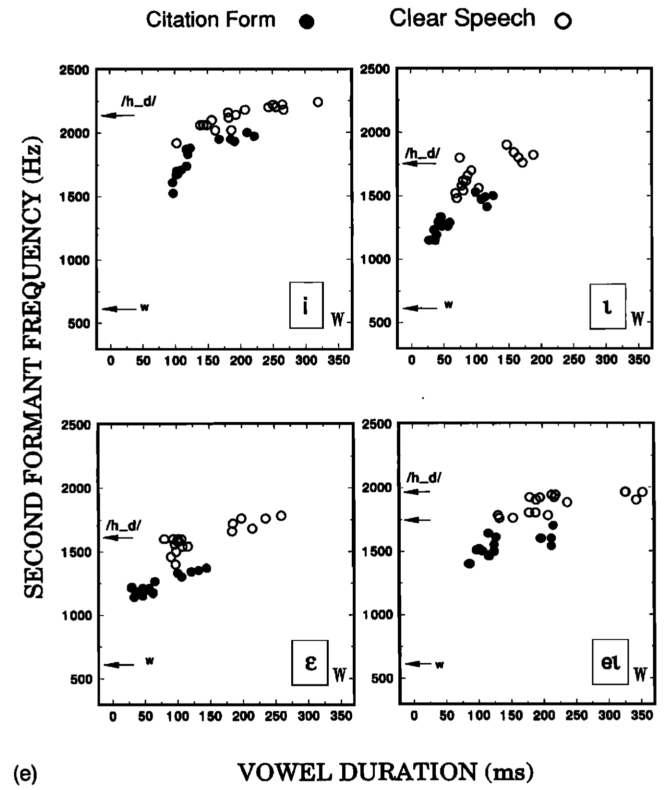

FIG. 4. (*Continued.*)

defined undershoot patterns for all vowels. Conceivably, this talker dependence may be attributed to how "clearly" the CF lists were spoken. Accordingly, for some speakers, the CF tokens may have been produced in a mode similar to the CS and may therefore show a ceiling effect.

The tense data for speaker R reveal an interesting pattern. In monosyllables R showed dark /l/'s with low F_2 values. So did the other speakers in that context. But in bi- and trisyllabic words, the high F_2 value of /l/ indicates that he used a "clear" variant [except in the case of the two outliers in Fig. 4(c)]. As the vowel became shorter, speaker R seems to have "sacrificed" the velar quality of /l/ rather than the vowel. Palatalizing the /l/ enabled him to reach the intended vowel target more easily and to minimize undershoot.

An attempt was made to quantify the data in terms of the mathematical model of a decaying exponential used in Lindblom (1963):

$$F_{20}=a^*(F_{2L}-F_{2T})e^{-\alpha D}+F_{2T}, \qquad (1)$$

where F_{20} is F_2 at the point of minimum rate of change (F_2 contour turning point); a and α are constants; F_{2L} is the value of F_2 in /w/; F_{2T} is the asymptote of the curve, in other words, the underlying "target;" D is vowel duration. Equation (1) can be rewritten as

$$\ln[(F_{20}-F_{2T})/(F_{2L}-F_{2T})]=\ln(a)-\alpha D. \qquad (2)$$

The values of F_{20}, F_{2L}, and D are known from the measurements. Unknown are F_{2T}, a and α. However, as Eq. (2) indicates, if an estimate of F_{2T} were available, the left-hand side of the expression is a linear function of D. Hence it should be possible to derive, for each vowel, style and speaker, $\ln(a)$ as the intercept, and $-\alpha$ (the slope), from a linear regression analysis.

Whenever possible, the F_2 values in the /hVd/ context were used to estimate F_{2T}, the target value. Lindblom (1963) assumed one target pattern per vowel. In the present investigation, it was not always possible to assign a single style-independent value to each vowel. For example, whenever the F_2 was higher in clear than in the null-context speech, $[(F_{20}-F_{2T})/(F_{2L}-F_{2T})]$ would be negative and so the use of Eq. (2) would become impossible. In such cases, the maximum F_2 value observed plus 50 Hz was taken to approximate the target. Accordingly, it was only when F_{20} measurements were below the F_{2T} defined as the null context of a given vowel that there was a single target value per vowel. Otherwise, two separate target values, one for each style, were used. For each speaker's F_{2L}, two values were used (CF and CS). It was defined as the F_2 value of /w/ sampled at the maximum constriction point, and averaged over all CF and CS tokens.

The data from the individual speakers are examined below in terms of this quantification. Estimates of the constants for all speakers are given in Table III. For each vowel, the F_{2T} (the target), $-\alpha$ (slope), the correlation coefficient r, and a [$\ln(a)$=intercept] are listed. The star indicates that the F_2 value of the null context was used as the target estimate. Otherwise the target is defined as ($F_{20\,\max}$+50), that is as the highest observed F_2 plus 50 Hz.

Fitting Eq. (1) to the data produces a descriptively acceptable quantification in many cases. For /i/ the model was not applied to D, G, and R who show pronounced duration-independent patterns. In the case of S and W, its application is more justified as seen from the r scores. When the CF and the CS data from S are pooled, a=0.46, $-\alpha$=−6.094, and r=−0.655. The lax vowels have generally high r scores for all speakers. The CS data for the /ε/ of S shows the lowest r as might be expected from their horizontally oriented appearance in Fig. 4(d). For the /eɩ/ of D and W flatter configurations are also seen especially for CS. When the CF and the CS data are pooled, for G's /eɩ/ and for S's /eɩ/, scores of r=−0.756 (G) and −0.809 (S) are obtained. As R's /eɩ/ gets shorter F_2 increases. A possible reason for this reversed pattern might be that, at shorter durations, the F_{20} estimate became more and more influenced by the glide than by the nucleus. A related phenomenon is discussed below in the presentation of the F_2 rate of change measurements.

An examination of the a and the α of Table III reveals covariation of these constants. Pooling vowels, speakers and styles, we find a significant correlation between $-\alpha$ and $\ln(a)$, (r=−0.90). Such a result indicates that long horizontal curve segments (high $-\alpha$'s) go with fast rises [large $\ln(a)$ values]. In Table III the $-\alpha$'s of CS tend to be numerically larger in the majority of the cases which means that, everything else being equal, CS generally shows less duration dependence.

In summary, the tense /i/ and /eɩ/ tended to show limited duration-dependent formant displacement, especially in CS. The data points were found to arrange themselves more or less horizontally in the F_2 vs duration diagrams. Those patterns occurred with the longest vowel durations. For lax

TABLE III. Results of fitting Eq. (1) to the data. $F_{2L}-F_{2T}$ is the frequency distance between "locus" and "target." The definition of this term is described in the text. The r is the correlation coefficient and a and α are constants. The numbers in parentheses pertain to the CS condition.

i	a	$F_{2L}-F_{2T}$	$-\alpha$	F_{2T}	r
D	0 (0)	−1786 (−1790)	− (−)	2424* (2424*)	n/a (n/a)
G	0 (0)	−1525 (−1562)	− (−)	2177* (2268)	n/a (n/a)
R	0 (0)	−1690 (−1565)	− (−)	2329 (2222*)	n/a (n/a)
S	1.85 (0.46)	−1796 (−1744)	−15.2 (−5.8)	2303* (2303*)	−0.78 (−0.51)
W	0.78 (0.57)	−1515 (−1671)	−9.9 (−9.3)	2135* (2290)	−0.92 (−0.92)
ι	a	$F_{2L}-F_{2T}$	$-\alpha$	F_{2T}	r
D	0.56 (0.79)	−1407 (−1411)	−5.5 (−12.6)	2045* (2045*)	−0.89 (−0.83)
G	0.79 (1.85)	−1213 (−1271)	−11.7 (−18.4)	1865* (1977)	−0.62 (−0.83)
R	0.86 (1.59)	−1293 (−1275)	−7.1 (−14.5)	1932* (1932*)	−0.99 (−0.90)
S	0.70 (0.74)	−1470 (−1418)	−4.2 (−6.6)	1977* (1977*)	−0.88 (−0.69)
W	0.65 (0.56)	−1140 (−1331)	−8.5 (−10.7)	1760* (1950)	−0.92 (−0.72)
ε	a	$F_{2L}-F_{2T}$	$-\alpha$	F_{2T}	r
D	0.48 (0.56)	−1250 (−1254)	−5.7 (−19.8)	1888* (1888*)	−0.88 (−0.89)
G	1.23 (0.59)	−1063 (−1063)	−14.2 (−8.9)	1715* (1769)	−0.94 (−0.68)
R	0.64 (0.39)	−1176 (−1158)	−8.3 (−5.1)	1815* (1815*)	−0.70 (−0.61)
S	0.72 (0.30)	−1204 (−1152)	−3.2 (−1.3)	1711* (1711*)	−0.86 (−0.50)
W	0.49 (0.56)	−995 (−1211)	−4.5 (−9.3)	1615* (1830)	−0.72 (−0.90)
eι	a	$F_{2L}-F_{2T}$	$-\alpha$	F_{2T}	r
D	0.26 (0)	−1714 (−1718)	−1.6 (−)	2352* (2352*)	−0.37 (n/a)
G	0.66 (2.88)	−1530 (−1476)	−6.5 (−15.6)	2182* (2182*)	−0.43 (−0.70)
R	0 (0)	−1576 (−1639)	− (−)	2215* (2296)	n/a (n/a)
S	0.51 (2.32)	−1643 (−1591)	−4.6 (−11.7)	2150* (2150*)	−0.80 (−0.52)
W	0.46 (0.33)	−1344 (−1391)	−2.8 (−6.1)	1964* (2010)	−0.68 (−0.74)

vowels, much stronger duration- and context-dependent shifts of F_{20} were observed.

Returning for a moment to the data of Table I and Fig. 2, we found more average undershoot in CF than in CS. Since longer durations were observed for CS, the question is whether those durational increases are sufficient to account for the more limited formant shifts in CS. The answer is provided by inspecting the plots of Fig. 4 and the numbers of Table III. For D [Fig. 4(a)] there are regions where the CF and CS points overlap durationally but where the two sets appear distinctly separated in frequency (/ι/, /ε/, and /eι/). Similar cases are G's /ι/, S's /ι/, and /ε/ and W's /ι/, /ε/, and /eι/. Such observations show that duration is not sufficient to account for the present undershoot patterns. In Table III those differences are paralleled by the behavior of α and F_{2T}. For CS, α, and occasionally also F_{2T}, tend to be higher. In response, we thus find that, while durational factors are responsible for some of the undershoot differences between CS and CF, they cannot be the only determinants.

C. Formant undershoot and F_2 rate of change

Since, as indicated by the literature review, articulatory undershoot has been examined in the light of movement velocity (Kuehn and Moll, 1976; Flege, 1988), supplementary information was obtained on the acoustic patterns of the present speech samples. In /wVl/ sequences, F_2 reflects primarily two articulators: The anterior–posterior movements of the tongue and the action of the lips. Although the relationship between F_2 and those two dimensions is not a linear one, for the purposes of the present discussion, F_2 velocity was considered to provide an acceptable approximate index of the underlying articulatory activity.

The velocity of the F_2 transitions in all vowels was estimated in the following way. With the aid of a Kay 5500 Sona-graph, the range of F_2 (ΔF_2) was measured from the /w/ constriction to the highest F_2 value. Two points were selected which were 25% down from the F_2 peak and 25% up from its value in /w/. The interval defined by those points (ΔT) was taken as the duration of the F_2 movement. Average velocity was calculated as $\Delta F_2/\Delta T$. This definition represents the average velocity during 50% of the total F_2 change. Estimates of peak velocities could not be reliably made.

Table IV lists the average F_2 velocity values by speaker, vowel, and style. In Fig. 5 this information is displayed as a function of $|F_{2L} - F_{2T}^*|$, the frequency difference between "locus" and "target." All speakers show a tendency toward a parallel between progressively larger velocities and greater $|F_{2L} - F_{2T}^*|$ values in the series /ε/–/ι/–/i/. Hence, for those vowels, F_2 rate of change is, to a certain extent, determined by the locus-target distance. However, there is a definite style dependence (more on that below). Also the /eι/ of most speakers presents a deviation from the main trend. The vowel /eι/'s special status can be accounted for as follows.

In the /hVd/ words, two sample points (nucleus and glide) were selected (cf. Fig. 4). In the /wVl/ tokens, they merged and were not easily identified so that F_2 had to be measured at a single point, the frequency maximum. The

TABLE IV. F_2 rate of change data for the five speakers (D,G,R,S,W), the four vowels and the two styles (CF—citation form; CS—clear speech). The values are averages calculated over at least 15 tokens (three word lengths and five or six repetitions) of each test word.

	i		ι		ε		eι	
	CF	CS	CF	CS	CF	CS	CF	CS
D	25.44	27.39	13.06	18.13	12.76	14.96	13.14	14.01
G	21.06	24.21	13.40	15.60	11.85	12.83	12.89	12.19
R	21.30	20.08	11.28	12.07	11.78	11.89	15.16	14.24
S	18.44	24.83	9.88	13.11	8.11	11.17	10.26	12.11
W	16.93	26.65	12.03	16.68	12.11	14.46	12.29	16.71

question arises whether that sample is representative of the nucleus, or if it is influenced by the palatal glide. Support for the latter possibility comes from comparing the nucleus estimates in the /hVd/ context with the F_{2T} values used to fit the exponential model to the data. [The difference in Hz between the nucleus of /heɩd/ and the F_{2T} values are 168 (CF and CS) for D; 188 (CF and CS) for G; 182 (CF) and 263 (CS) for R; 156 (CF and CS) for S; and 212 (CF) and 258 (CS) for W.]

When pooled data were used omitting [eɩ], $|F_{2L} - F_{2T}^*|$ and velocities showed strong linearity: For CS, velocity was equal to $23.8^*|F_{2L} - F_{2T}^*| - 14.2$ (r^2=0.97). In the case of CF, velocity could be described by $19.57^*|F_{2L} - F_{2T}^*| - 12.02$ (r^2=0.93).

It is clear that, for the majority of cases in Fig. 5, the CS observations exhibit larger average velocity values than those for CF. That is brought out even more clearly in Fig. 6 which presents the average CS-CF F_2 velocity difference in histogram form. In all cases except the /i/ of R and the /eɩ/ of G and R, F_2 appears to move faster in the CS condition.

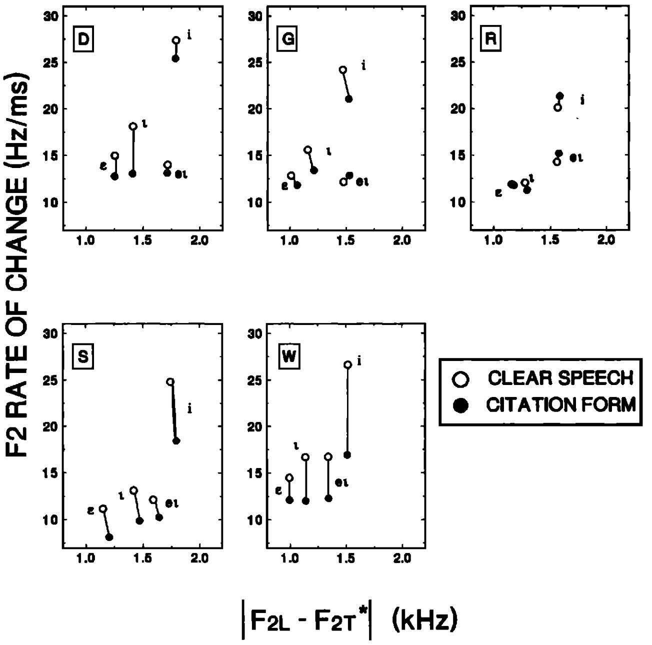

FIG. 5. The second formant rate of change is plotted against the frequency difference between "locus" and "target" ($|F_{2L} - F_{2T}^*|$) for all vowels and for all five speakers.

III. DISCUSSION

The selection of the speech samples was made so that the stress placed on all test syllables would be constant and could be categorized as "main" or "primary" stress. They were also chosen with a view toward producing sufficient durational variations. To avoid subjective judgements of speaking tempo, the "word length effect" was used to achieve that goal. As Fig. 4(a)–(e) make evident, vowel durations did indeed exhibit considerable variations. For most groups of CF and CS data in Fig. 4, vowels undergo shortening by 40%–60%. Consequently, the present formant undershoot (cf. Figs. 2 and 3, Table I) was observed under conditions of constant stress and constant speaking tempo.

As mentioned, the magnitude of the formant displacements is greater than the effects that have been reported in the literature. What accounts for these discrepancies? One difference between the present study and other investigations

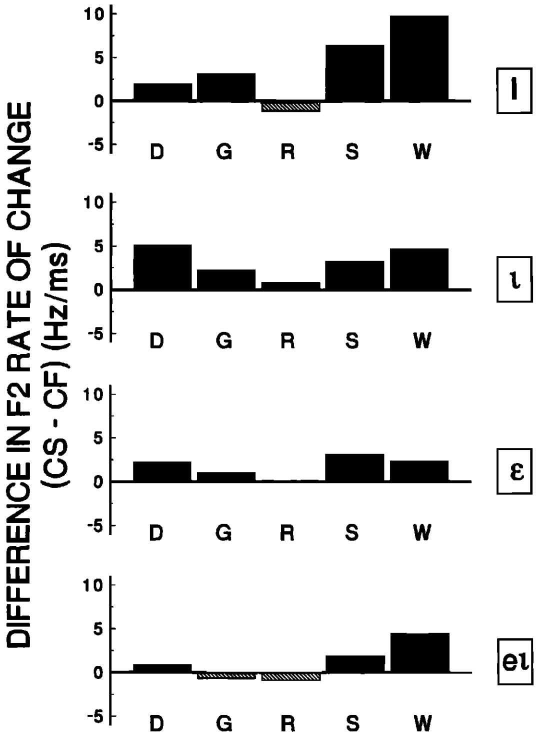

FIG. 6. Second formant velocity differences between CS and CF are plotted for each vowel and for each speaker. Positive values are indicated by solid bars, and negative ones by shaded bars.

is the deliberate use of the large "locus-target" distances of the /wVl/ context. The motivation for that choice was that large formant excursions would make reduction in the form of formant undershoot easy to observe and measure, should it occur. From the view point of the *average* effect that consonants have on vowel formant patterns in English, such transitions are somewhat atypical. However, there is no evidence to suggest that there is anything exceptional about their production. Consider the following two arguments.

One line of reasoning says that vowel reduction (formant undershoot) works in exactly the same way whatever the identity of the interacting consonants and vowels. Accordingly, there is nothing anomalous about the manner in which the gestures of /wVl/ syllables are coordinated. The only unusual thing about that context is that, when combined with front vowels, it gives rise to large formant transitions. However, that circumstance offers the advantage of putting a magnifying glass on the acoustic correlates of reduction, amplifying the undershoot process and providing larger, more easily interpreted effects. Another possibility is to say that vowel reduction works in different ways depending on the identities of the interacting segments and that the /wVl/ context is therefore governed by special coarticulation rules.

The position taken here is the first one, the more parsimonious one: The /wVl/ sequence, which obviously occurs frequently in real English words, has large formant transitions because it exploits a region of the phonetic space where the acoustic response to articulatory movement is particularly sensitive. However, that fact in no way justifies assuming that the manner in which the constituent gestures are coarticulated is exceptional and differs from that of other CVC syllables.

The choice of the /wVl/ frame was based on previous results demonstrating that one of the determinants of undershoot is "locus-target" distance. Accordingly, Eq. (1), adopted from Lindblom (1963), contains that parameter. Its value is used to control the magnitude of the predicted effect. We suggest that the present large undershoot effects are fully compatible with quantifying degree of context dependence in terms of "locus-target" distance.

A further property of Eq. (1) is that it treats formant displacement as contextual assimilation, not as a change toward schwa. The lax vowels of all speakers provide tests of that description and, as indicated in Fig. 4, also evidence supporting it. We note how, as durations shorten, formants move away from, rather than come closer to, a nominal schwa-like neutral value of 1500 Hz. That conclusion is not surprising in view of the choice of primary-stressed test vowels which excludes the phonological process of reduction that changes vowels to schwa in English.

Another controversial issue is whether formant undershoot depends on vowel duration. In Lindblom (1963), there were several CVC words for which formant frequencies were more or less constant and independent of duration. The reason why Lindblom's data could nevertheless be justifiably described by the exponential model of Eq. (1) was that the cases where duration did not seem to be a factor, coincided with small "locus-target" distances.

In the present study, a similar situation appears to prevail. The trend towards horizontal patterns in the F_{20} vs duration diagrams (Fig. 4) is most pronounced for the conditions that give, relatively speaking, the longest durations, that is for the tense vowels and for the CS condition. In this study, instances both with and without duration dependence can thus be identified. However, when the durational facts just mentioned are taken into consideration, the present data appear compatible with the interaction that Eq. (1) postulates between duration-dependence and "locus-target" distance. It is worth noting that it is never the case that horizontal formant-duration patterns occur at the shortest durations. Hence it appears justified to suggest that one of the reasons why several plots in Fig. 4—e.g., the /i/ vowels of D, G, and R—do not show formant undershoot is durational. In our interpretation, those patterns are compatible with models postulating duration dependence.

The conclusion to be drawn from the discussion so far is that support has been adduced for modeling phonetic vowel reduction (formant undershoot) as a duration-dependent and contextually determined process. However, it has also brought forth facts about formant displacement that are at variance with such models. For instance, differences were revealed between CF and CS with respect to degree of undershoot. As established in Sec. I, some of the style-dependent undershoot differences can be accounted for by invoking duration dependence. Some, but not all, speakers show several cases in which the CF and CS measurements overlap durationally and form distinct F_2 distributions. Those cases refute, or necessitate revisions of, models such as Eq. (1).

The data on F_2 rate of change offers an opportunity for developing such revisions. To provide some background for the discussion of the velocity data we shall first make some remarks about the biomechanics of articulatory movement. In the past, describing speech movements, phoneticians have noted that articulatory trajectories tend to resemble the output of damped mechanical systems. Several production models have been developed from the assumption that speech motor mechanisms can be seen as functionally equivalent to second-order mechanical systems (Henke, 1966; Lindblom, 1967; Öhman, 1967; Coker, 1976; Perkell, 1974; Fujisaki, 1983; Saltzman and Munhall, 1989).

For such a system with mass (M), friction (B) and elasticity (K), critical damping is said to occur when

$$B=2(KM)^{1/2}. \qquad (3)$$

The conditions under which positional "undershoot" would be observed in such a system are as follows. We shall assume that a force is applied whose time variations are rectangularly shaped. The response of the system—that is, $x(t)$, its displacement as a function of time—would be given by

$$x(t)=Au(t)[1-\alpha te^{-\alpha t}-e^{-\alpha t}]-Au(t-D) \times[1-\alpha(t-D)e^{-\alpha(t-D)}-e^{-\alpha(t-D)}]. \qquad (4)$$

This formulation assumes that the movement is from zero to A, where A is the full extent of the movement reached under asymptotic conditions. A is related to the magnitude of the applied force, $u(t)$ represents the unit step of

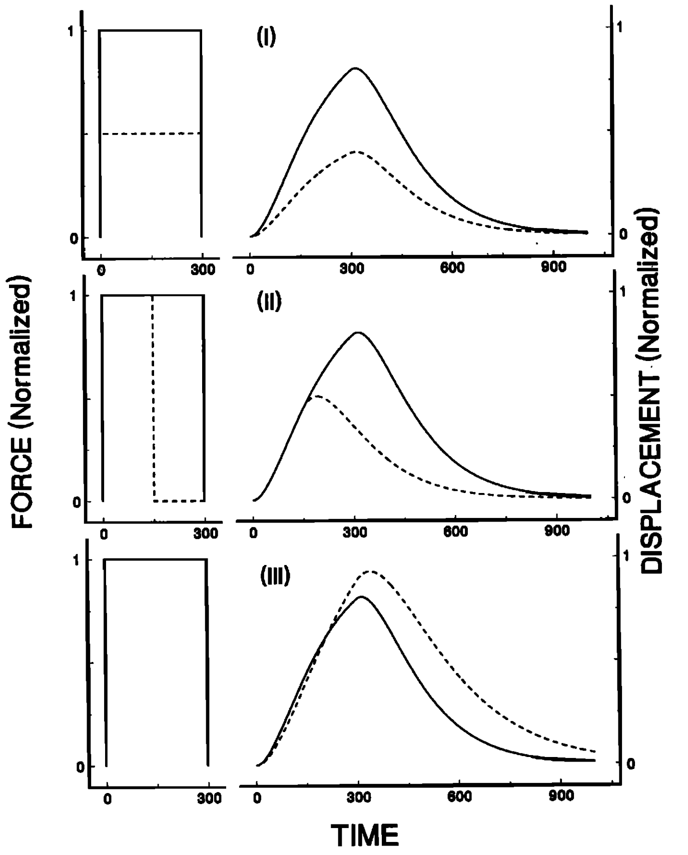

FIG. 7. Responses of a critically damped system (right set of curves) to stepwise varying input force (left set). In the left section of panel (I), a large (solid line) and a smaller force amplitude (dashed) of equal duration are presented. The associated displacement curves are shown on the right. The smaller force produces the more limited response. In panels (II) and (III), the solid lines of the force and displacement curves are identical with those of (I). In (II), the duration of the applied force is varied. The shorter input yields the smaller movement (dashed). In (III), the stiffness of the system is reduced, while the amplitude and duration of the input force are left unchanged. The lower stiffness leads to a more extensive displacement (dashed).

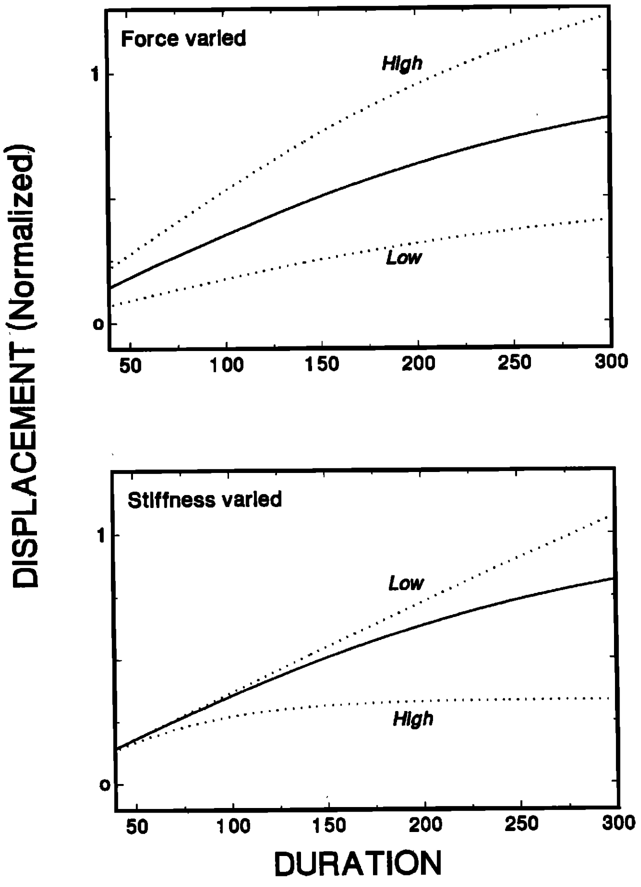

FIG. 8. Duration-dependent undershoot curves for the critically damped system of Fig. 7. The diagrams show the effects of varying input force (top) and system tuning (bottom). The x axis refers to duration of input force. The solid curves are identical in the two panels and were calculated according to the conditions of panel (II) of Fig. 7. In the upper panel, the input force was halved (lower curve) and increased by 50% (upper curve). In the lower panel, the system response was varied by making α three times smaller (upper curve, stiffness low) and three times larger (lower curve, stiffness high).

the rectangular forcing function, and $\alpha=(K/M)^{1/2}$.

Figure 7 gives a few examples of such critically damped responses. Three situations are shown. For simplicity, the time variations of the input force have the same rectangular "on–off" shape in all cases. In panel (I), this input force parameter has two values, large and small. In panel (II), force stays large, but the duration of the pattern is varied. In panel (III), force stays large, but the system tuning, or "stiffness," is varied. The mass of the system remains unchanged, but we should note that, for the system to remain critically damped, changing stiffness (=spring constant), always implies changing the friction value too [cf. Eq. (3)].

Also shown in Fig. 7 are the displacement curves derived for the three conditions. Decreasing input force or duration of input, or increasing system stiffness (while force remains the same), is associated with a reduction of the response, or "undershoot." The change in input amplitude in (I) leads to a scaling of the responses, the smaller force producing a smaller movement. In (II), there is not enough time for the system to complete its initial response, and, in (III), decreasing the stiffness, while keeping input force the same, produces a larger movement and therefore reduces undershoot.

The examples of Fig. 7 demonstrate three different ways of obtaining "undershoot." (II) is analogous to what we have called "duration-dependent undershoot." (I) can be termed "force-dependent" undershoot and, in (III), undershoot is "system dependent," that is, it is produced by a change in the time constant of the system.

It can be shown that $x_{\max}$, the maximum excursion of the above system, will vary according to:

$$x_{\max}=A\{\exp[\alpha D/(1-e^{\alpha D})]-\exp[\alpha De/(1-e^{\alpha D})]\}. \tag{5}$$

This expression tells us that, as D, the duration of the applied force, becomes larger, $x_{\max}$ will increase and eventually approach A asymptotically. We first note that $x_{\max}$ depends on three factors: the extent of the movement: A; the speed of the system response: α; and the duration of the input force: D.

In Fig. 8 we plot $x_{\max}$ as a function of D for the system configurations of Fig. 7. The solid curves of the two panels refer to the condition in B: The input force amplitude and the tuning of the system ("stiffness") remain the same, but duration of the input is varied. In the top diagram, the dashed curves differ from the solid one in that the force amplitude was either increased (top) or decreased (below) by a certain value. In the bottom case, input force was maintained constant while "stiffness" was either decreased (top) or increased (below) by a fixed value. The lesson taught by Fig. 8 is that, to compensate for undershoot, several strategies are available: Undershoot can be reduced by increasing the input force, by making the duration longer, or by speeding up the time constant of the system.

To gain some further insight into the undershoot phenomenon, we can make a direct analogy between the behavior of an elementary mechanical system and the speech production system, more specifically, between $x_{\max}$ and the maximum excursion of F_2—that is, F_{20} as measured in the present study. That such an analogy is both meaningful and instructive should be evident for the following reasons. As remarked earlier, there is an approximate, but straightforward, relationship between F_2 variations and tongue/lip movements in the present study. Second, several successful attempts have been made in the past to model articulatory movement in terms of critically damped second-order system (Henke, 1966; Lindblom, 1967; Öhman, 1967; Coker, 1976; Perkell, 1974; Fujisaki, 1983; Saltzman and Munhall, 1989). The analogy implies that A corresponds to "locus-target" distance, the F_2 rate of change is parallel to the velocity of the system response as quantified by α, and that the duration of the input force, D, bears a direct relation to vowel duration. Consequently, we conclude that, from a biomechanical point of view, formant undershoot ought to be a function of "locus-target" distance, vowel duration and F_2 rate of change.

There is evidence from this study as well as from other investigations that both "locus-target" distance and vowel duration play an important role in determining formant undershoot. Moreover, we find that, of the mechanisms for undershoot compensation inferred from the biomechanical considerations, the speakers invoked all three in the CS condition. First, the longer CS durations produced less undershoot. Second, for some speakers (Table III) the F_{2T} for CS is higher which indicates that a more "palatal" articulation was invoked compensatorily. Third, the biomechanical model also suggests that, everything else being equal, a faster transition entails less undershoot. The present data agree with that prediction. There is less undershoot in CS (Fig. 2) and, by and large, that style also shows the greater velocity values (Figs. 5 and 6). In further agreement with that expectation, the CS and the CF of speaker W have both the largest undershoot differences and the larger velocity discrepancies. On the other hand, speaker R, who does not increase CF velocities, shows no major differences between CF and CS with respect to undershoot. The argument further suggests that undershoot and velocity should be related not only at the level of average results, but for individual measurements. However, investigating that question runs into the difficulty of satisfying the condition of "everything else being equal." For any given CF or CS data point, a duration, an F_{20}, and a velocity is available, but, as evident from Fig. 4, a given CF, or CS, duration is only rarely matched by an identical or similar value for the other style. Nevertheless, this issue was addressed, in a preliminary fashion, by quantizing the CS and CF durations in terms of common 25-ms bins. The values falling within those bins were regarded as durationally equivalent and the associated CS and CF formant and velocity data were averaged. This procedure produced 28 durationally equivalent comparisons from all speakers except S whose data called for unacceptably large bin sizes and were therefore excluded. Also omitted were all the /eι/ measurements because of the above-mentioned difficulties of obtaining reliable estimates of the F_2 target for the /eι/ nucleus. Mean F_{20} and velocity values were calculated on the basis of an average of four data points per bin. The data were normalized with respect to "locus-target" distance to permit pooling. "Undershoot decrease" was defined as $\{[F_{20}\,(CF) - F^*_{2T}\,(CF)]\,/\,[F_{2L}\,(CF) - F^*_{2T}\,(CF)] - [F_{20}(CS) - F^*_{2T}(CS)]/[F_{2L}(CS) - F^*_{2T}(CS)]$, "velocity increase" as $[\mathrm{VEL(CS)}/(F_{2L}(\mathrm{CS}) - F^*_{2T}(\mathrm{CS})] - \mathrm{VEL(CF)}/[F_{2L}(\mathrm{CF}) - F^*_{2T}(\mathrm{CF})]\}$. A regression analysis indicated an r score of 0.77, a result which lends unequivocal support to invoking formant rate of change along with duration and context dependence in accounting for the present vowel undershoot data.

The present data demonstrate consistently smaller undershoot effects in CS that could possibly have been produced either by increasing force, or by lowering stiffness, or by both mechanisms. There is no way of distinguishing between the two cases, but, since all clear tokens were 3–5 dB more intense than the citation forms, there is some basis for assuming that clear forms were also articulated somewhat more forcefully. More rapid formant transitions suggest faster articulatory gestures which is compatible with the velocity-based measure of "articulatory effort" proposed and explored by Nelson (1983) and Nelson *et al.* (1984).

In summary, this research was undertaken to investigate the acoustic variations of the English front vowels /i/, /ι/, /ε/, and /eι/ when they were embedded in a /wVl/ frame, carried constant main stress and were produced at varying durations in clear and casual style and at a constant speaking rate. Acoustic analyses revealed: (i) that formant patterns were systematically displaced in the direction of the frequencies of the consonants of the adjacent pseudosymmetrical context (context dependence); (ii) that those displacements depended in a lawful manner on vowel duration (duration dependence); (iii) that this context and duration dependence was more limited for clear than for citation-form speech and that the smaller formant shifts of clear speech tended to be achieved by increases in the rate of formant frequency change ("style" or "effort" dependence).

The above findings are compatible with a revised, and biomechanically motivated, version of the vowel undershoot model (Lindblom, 1963) that derives formant patterns from numerical information on three (rather than two) variables: The *"locus-target" distance* (to capture "context dependence"), *vowel duration*, and rate of *formant frequency change* (the new variable and an indirect index of "articulatory effort").

Why does the present experiment produce evidence of strong context and duration dependence, whereas, in other projects, such effects have been weak, or negligible? What accounts for those discrepancies and how are the conflicting claims in the literature to be reconciled? The proposed model carries the implication that the absence or presence of formant undershoot effects would best be resolved, if experiments examined, not only durations and formant frequencies, but all three of the above variables. In other words, in response to the question, we suggest that current uncertainties about the status of formant undershoot effects may be dispelled once a more complete sampling of the variables that underlie and shape them is undertaken.

Finally, it also seems important to note that the present samples of clear speech were not merely amplified, louder versions of normal speech. As shown, their formant patterns differed in systematic ways from those of the citation forms. It appears reasonable to attribute those differences to the perceptual function of clear speech. As indicated by the introductory literature review, there is evidence suggesting that clear speech is more intelligible than casual speech. Accordingly, there is some justification for proposing that the present clear speech tokens involved an active output-oriented reorganization of phonetic gestures adaptively tuned to the purpose of attenuating, and to some extent compensating for, formant displacements and undershoot effects.

ACKNOWLEDGMENTS

This research was developed from the first author's doctoral dissertation and supported by a grant from the Advanced Research Program of the Texas Board of Coordination and grant No. BNS-9011894 from the NSF. We thank Dr. Jerald Lame at the University of Texas at Austin for his many valuable contribution to this project.

Chen, F. R. (**1980**). "Acoustic Characteristics and Intelligibility of Clear and Conversational Speech at the Segmental Level," Master's thesis, MIT, Cambridge, MA.

Chen, F. R., Zue, V. W., Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1983**). "Speaking Clearly: Acoustic Characteristics and Intelligibility of Stop Consonants," 1–8 in Working Papers II, Speech Communication Group, MIT.

Chomsky, N., and Halle, M. (**1968**). *Sound Pattern of English* (Harper & Row, New York).

Clark, J. E., Lubker, J. F., and Hunnicutt, S. (**1987**). "Some Preliminary Evidence for Phonetic Adjustment Strategies in Communication Difficulty," in *Language Topics: Essays in Honour of Michael Halliday*, edited by Steele and Threadgold (Benjamins, Amsterdam, The Netherlands).

Coker, C. (**1976**). "A model of articulatory dynamics and control," Proc. IEEE **64**(4), 452–460.

Delattre, P., Liberman, A. M., and Cooper, F. S. (**1955**). "Acoustic loci and transitional cues for consonants," J. Acoust. Soc. Am. **27**, 769–773.

Draegert, G. L. (**1951**). "Relationships between voice variables and speech intelligibility in high level noise," Speech Monogr. **18**, 272–278.

Engstrand, O. (**1988**). "Articulatory Correlates of Stress and Speaking Rate in Swedish VCV Utterances," J. Acoust. Soc. Am. **83**, 1863–1875.

Engstrand, O., and Krull, D. (**1989**). "Determinants of Spectral Variation in Spontaneous Speech," in *Proc. of Speech Research '89* (Budapest), pp. 84–87.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Ferguson, C. A. (**1977**). "Baby talk as a simplified register," in *Talking to Children*, edited by C. E. Snow and C. A. Ferguson (Cambridge U.P., Cambridge).

Flege, J. E. (**1988**). "Effects of Speaking Rate on Tongue Position and Velocity of Movement," J. Acoust. Soc. Am. **84**, 901–916.

Fourakis, M. (**1991**). "Tempo, Stress, and Vowel Reduction in American English," J. Acoust. Soc. Am. **90**, 1816–1827.

Freed, B. F. (**1978**). "Foreign Talk: A Study of Speech Adjustments Made by Native Speakers of English in Conversation with Non-Native Speakers," unpublished Ph.D. dissertation, University of Pennsylvania.

Fujisaki, H. (**1983**). "Dynamic characteristics of voice fundamental frequency in speech and singing," in *The Production of Speech*, edited by P. F. MacNeilage (Springer-Verlag, New York).

Gay, T. (**1978**). "Effect of Speaking Rate on Vowel Formant Movements," J. Acoust. Soc. Am. **63**, 223–230.

Hanley, T. D., and Steer, M. D. (**1949**). "Effect of level of distracting noise upon speaking rate, duration and intensity," J. Speech Hear. Disord. **14**, 363–368.

Henke, W. L. (**1966**). "Preliminaries to speech synthesis based upon an articulatory model," Proc. IEEE Conf. Speech Commun. Process, 170–182.

Huang, C. B. (**1991**). "An acoustic and perceptual study of vowel formant trajectories in American English," Doctoral dissertation, MIT.

Joos, M. (**1948**). *Acoustic Phonetics* (LSA, Waverly, Baltimore).

Karlsson, I. (**1992**). "Analysis and Synthesis of Different Voices with Emphasis on Female Speech," Doctoral dissertation, Royal Institute of Technology, Stockholm.

Koopmans-van Beinum, F. J. (**1980**). "Vowel Contrast Reduction, an acoustic and perceptual study of Dutch vowels in various speech conditions," Ph.D. thesis, University of Amsterdam, The Netherlands. Academische Pers B. V., Amsterdam.

Kuehn, D. P., and Moll, K. L. (**1976**). "A Cineradiographic Study of VC and CV Articulatory Velocities," J. Phon. **4**, 303–320.

Ladefoged, P. (**1975**). *A Course in Phonetics* (Harcourt Brace Jovanovich, New York).

Ladefoged, P., Kameny, I., and Brackenridge, W. (**1976**). "Acoustic effects of styles of speech," J. Acoust. Soc. Am. **59**, 228–231.

Lane, H. L., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.

Lehiste, I. (**1964**). *Acoustical Characteristics of Selected English Consonants,* IJAL **30**(3).

Lehiste, I. (**1974**). "Duration of Syllable Nuclei as a Function of Word Length and Stress Pattern," in *Proc. of the VIIIth ICA* (London), p. 300.

Lehiste, I. (**1975**). "Some Factors Affecting the Duration of Syllable Nuclei in English," Salzburger Beiträge zur Linguistik **1**, 81–104.

Lindblom, B. (**1963**). "Spectrographic Study of Vowel Reduction," J. Acoust. Soc. Am. **35**, 1773–1781.

Lindblom, B. (**1967**). "Vowel duration and a model of lip/mandible coordination," STL/QPSR **4**/1967 1–29.

Miller, J. L. (**1981**). "Some effects of speaking rate on phonetic perception," in *Perspectives on the Study of Speech*, edited by P. D. Eimas and J. L. Miller (LEA, Hillsdale, NJ), pp. 39–79.

Moon, S-J. (**1990**). "Durational aspects of clear speech," unpublished Master's Report, University of Texas at Austin.

Moon, S-J., Bonaventura, P., and Kelly, M. (**1988**). "Patterns of phonetic constancy and variation in three speaking styles," unpublished manuscript, University of Texas at Austin.

Nelson, W. L. (**1983**). "Physical principles for economics of skilled movements," Biol. Cybernet. **46**, 135–147.

Nelson, W. L., Perkell, J. S., and Westbury, J. R. (**1984**), "Mandible move-

[1] That interpretation suggests a continuous phonetic mechanism and should be distinguished from accounts that attribute reduction to a phonological redefinition of a vowel's "target" from a full vowel to a schwa, cf., the English rule proposed by Chomsky and Halle (1968) saying that a [-tense, -stress] vowel turns into schwa. [See also Fourakis (1991).]

[2] This usage is adopted for lack of better terminology. It deviates from the original meaning of the term "locus" as proposed by the Haskins group (Delattre *et al.*, 1955). It is more in the line with Stevens and House (1956) and Fant (1960, p. 25), who suggested that the term "be reserved for *F* patterns of specific sounds, primarily for the limiting positions of the articulators, e.g., at the state of maximum closure for consonants..."

[3] Speech materials consisting of mono-, bi-, or trisyllabic /wvl/ words. According to the duration of the initial stressed vowel, which is determined by the number of syllables per word, the test words are as follows: *long duration*—wheel, will, well, and wail; *short duration*—wheeling, willing, welling, and wailing; and *shorter duration*—Wheelingham, Willingham, Wellingby and Wailingby.

ments during increasingly rapid articulations of single syllables: Preliminary observations," J. Acoust. Soc. Am. **75**, 945–951.

Nord, L. (**1975**). "Vowel Reduction—Centralization or Contextual Assimilation?," in *Proceedings of the Speech Communication Seminar*, edited by G. Fant (Almqvist & Wiksell, Stockholm), Vol. 2, pp. 149–154.

Nord, L. (**1986**). "Acoustic Studies of Vowel Reduction in Swedish," in STL-QPSR 4/1986, Dept of Speech Communication, RIT, Stockholm, pp. 19–36.

Ohde, R. N., and Sharf, D. J. (**1975**). "Coarticulatory effects of voiced stops on the reduction of acoustic vowel targets," J. Acoust. Soc. Am. **58**(4), 923–927.

Öhman, S. (**1967**). "Word and sentence intonation: A quantitative model," STL/QPSR **2–3**/1967, 20–54.

Perkell, J. S. (**1974**). "A physiologically oriented model of tongue activity during speech production," Ph.D. thesis, MIT.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking Clearly for the Hard of Hearing II: Acoustic Characteristics of Clear and Conversational Speech," J. Speech Hear. Res. **29**(4), 434–446.

Saltzman, E. L., and Munhall, K. G. (**1989**). "A dynamical approach to gestural patterning in speech production," Ecol. Psychol. **1**(4), 333–382.

Stålhammar, U., Karlsson, I., and Fant, G. (**1973**). "Contextual effects on vowel nuclei," STL-QPSR, No. 4, pp. 1–18.

Stetson, R. H. (**1951**). *Motor Phonetics* (North-Holland, Amsterdam).

Stevens, K. N., and House, A. S. (**1956**). "Studies of formant transitions of using a vocal tract analog," J. Acoust. Soc. Am. **28**, 578–585.

Stevens, K. N., and House, A. S. (**1963**). "Perturbation of Vowel Articulations by Consonantal Context: An Acoustical Study," J. Speech Hear. Res. **6**, 111–128.

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effects of Noise on Speech Production: Acoustic and Perceptual Analyses," J. Acoust. Soc. Am. **84**, 917–928.

Tiffany, W. R. (**1959**). "Non-Random Sources of Variation in Vowel Quality," J. Speech Hear. Res. **2**, 305–317.

Uchanski, R. M., Reed, C. M., Durlach, N. I., and Braida, L. D. (**1985**). "Analysis of phoneme and pause durations in conversational and clear speech," J Acoust. Soc. Am. Suppl. 1 **77**, S54.

Van Son, R. J. J. H., and Pols, L. C. W. (**1990**). "Formant frequencies of Dutch vowels in a text, read at normal and fast rate," J. Acoust. Soc. Am. **88**, 1683–1693.

Van Son, R. J. J. H., and Pols, L. C. W. (**1992**). "Formant movements of Dutch vowels in a text, read at normal and fast rate," J. Acoust. Soc. Am. **92**, 121–127.