Effects of consonant environment on vowel formant patterns

James M. Hillenbrand[a] and Michael J. Clark
Speech Pathology and Audiology, Western Michigan University, Kalamazoo, Michigan 49008

Terrance M. Nearey
Department of Linguistics, University of Alberta, Edmonton, Alberta T6G 2E7, Canada

(Received 21 June 2000; accepted for publication 7 November 2000)

A significant body of evidence has accumulated indicating that vowel identification is influenced by spectral change patterns. For example, a large-scale study of vowel formant patterns showed substantial improvements in category separability when a pattern classifier was trained on multiple samples of the formant pattern rather than a single sample at steady state [J. Hillenbrand *et al.*, J. Acoust. Soc. Am. **97**, 3099–3111 (1995)]. However, in the earlier study all utterances were recorded in a constant /hVd/ environment. The main purpose of the present study was to determine whether a close relationship between vowel identity and spectral change patterns is maintained when the consonant environment is allowed to vary. Recordings were made of six men and six women producing eight vowels (/i,ɪ,ɛ,æ,ɑ,ʊ,u,ʌ/) in isolation and in CVC syllables. The CVC utterances consisted of all combinations of seven initial consonants (/h,b,d,g,p,t,k/) and six final consonants (/b,d,g,p,t,k/). Formant frequencies for F_1-F_3 were measured every 5 ms during the vowel using an interactive editing tool. Results showed highly significant effects of phonetic environment. As with an earlier study of this type, particularly large shifts in formant patterns were seen for rounded vowels in alveolar environments [K. Stevens and A. House, J. Speech Hear. Res. **6**, 111–128 (1963)]. Despite these context effects, substantial improvements in category separability were observed when a pattern classifier incorporated spectral change information. Modeling work showed that many aspects of listener behavior could be accounted for by a fairly simple pattern classifier incorporating $F0$, duration, and two discrete samples of the formant pattern. [DOI: 10.1121/1.1337959]

PACS numbers: 43.70.Bk, 43.71.An, 43.72.Ne, 43.70.Fq [KRK]

I. INTRODUCTION

A major focus of recent vowel perception research has been an examination of the relationship between formant-frequency movements and vowel identity. A good deal of evidence has accumulated implicating a secondary but quite important role for spectral change in vowel recognition. Reviews of this work can be found in Nearey (1989) and Strange (1989). Briefly, the evidence favoring this view includes the work of Strange, Jenkins, and Johnson (1983) and Nearey and Assmann (1986) showing high identification rates for ''silent-center'' stimuli in which vowel centers were gated out, leaving only brief onglides and offglides. Nearey and Assmann also reported a sharp decrease in identification rates for silent center signals in which onglides and offglides were played in reverse order (see also Jenkins, Strange, and Edman, 1983; Parker and Diehl, 1984; Andruski and Nearey, 1992; Jenkins and Strange, 1999). Further, several studies have reported relatively high identification error rates for both natural and synthetic vowels with static formant patterns (Fairbanks and Grubb, 1961; Hillenbrand and Gayvert, 1993a; Hillenbrand and Nearey, 1999). For example, Hillenbrand and Nearey asked listeners to identify naturally produced /hVd/ syllables and two different formant-synthesized versions. An ''original formant'' (OF) set of synthetic signals was generated using the measured formant contours, and a second set of ''flat formant'' (FF) signals was synthesized with formant frequencies fixed at the values measured at the steadiest portion of the vowel. The OF synthetic signals were identified with substantially greater accuracy than the FF signals. Finally, a number of pattern recognition studies have reported better classification accuracy and/or improved prediction of listener error patterns for pattern recognition models that incorporate spectral change as opposed to models that are driven by spectral measurements sampled at a single cross section of the vowel (Assmann, Nearey, and Hogan, 1982; Nearey and Assmann, 1986; Parker and Diehl, 1984; Zahorian and Jagharghi, 1993; Hillenbrand *et al.*, 1995). For example, Hillenbrand *et al.* trained a discriminant classifier on various combinations of fundamental frequency and formant measurements from /hVd/ syllables spoken by 45 men, 48 women, and 46 children. The pattern classifier was substantially more accurate when it was trained on two samples of the formant pattern (taken at 20% and 80% of vowel duration) than a single sample taken at the steadiest portion of the vowel.

An important limitation of the work conducted on this problem to date is the exclusive reliance on either isolated vowels or /hVd/ syllables. It is firmly established that vowel formant patterns are affected not only by the identity of the vowel, but also by consonant environment. In a classic study, Stevens and House (1963) reported formant measurements for eight vowels (/i,ɪ,ɛ,æ,ɑ,ʌ,ʊ,u/) spoken by three men. The vowels were produced in isolation, in /hVd/ syllables, and in

[a]Electronic mail: james.hillenbrand@wmich.edu

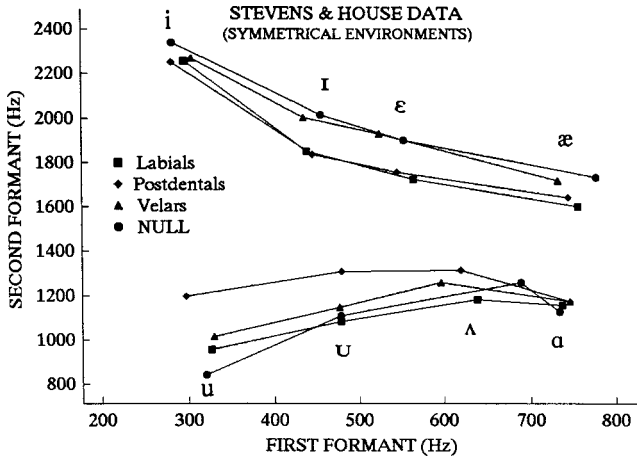

FIG. 1. Stevens and House (1963) data showing the effects of consonant environment on steady-state vowel formants.

symmetrical CVC syllables with 14 consonants (/p,b,f,v,θ,ð,s,z,t,d,ʧ,ʤ,k,g/). Effects of consonant context were examined by comparing the formant values in these 14 environments to formant values for the same vowels in isolation or /hVd/ context, which the authors referred to as ''null'' environments. Formant frequencies and bandwidths were measured for F_1-F_3 at the center of the vowel using a spectrum-matching technique. The most general summary of the Stevens and House findings is that the non-null consonant environments typically had the effect of shifting the formant frequencies—particularly F_2—toward more centralized values. Systematic effects were seen for the manner, voicing, and place of articulation of the flanking consonants. The place effects, which were easily the most important, are reproduced in Fig. 1. The effects of place on F_1 values tended to be small and rather consistent in magnitude from one vowel to the next. Place effects on F_2, on the other hand, were sometimes quite large and varied considerably in magnitude from one vowel to the next. The largest effect by far was an upward shift averaging about 350 Hz in F_2 for /u/ in the environment of postdental consonants; a shift averaging about 200 Hz was also seen for /ʊ/ in the environment of postdentals. There were also downward shifts in F_2 of some 100–200 Hz for front vowels (with the exception of /i/) in the environment of labial and postdental consonants. The effects of manner class and voicing were typically rather small. Vowels flanked by voiced consonants tended to be produced with slightly lower F_1 values as compared to the same vowels in the context of unvoiced consonants. Manner class had little effect on F_1 values, but vowels in stop consonant environments tended to have slightly higher F_2 values.

Stevens and House interpreted these varied findings in terms of a production undershoot model. The production system was assumed to be driven by targets corresponding to articulatory postures in null environments, but these idealized targets were purportedly not realized due to inertial constraints. Stevens and House (1963) also suggested that listeners make tacit use of knowledge of these context effects in recognizing vowels: ''The rules governing these deviations in the acoustic signal must, of course, be invoked in some way by the listener in order to make an identification of the signal'' (p. 122).

A problem that is presented by the findings discussed above is that there are clearly multiple influences on the detailed formant contours of even relatively simple citation-form CVC utterances. The two influences that are of interest in the present study are the consonant context effects described above and the ''vowel inherent spectral change'' patterns that have been observed in studies such as Nearey and Assmann (1986) and Hillenbrand *et al.* (1995) using isolated vowels or /hVd/ syllables. The primary question that is to be addressed is whether context effects such as those described by Stevens and House act to obscure or complicate the relationships between vowel identity and spectral change patterns that have been observed in previous studies using neutral contexts. Some preliminary evidence on this question comes from Zahorian and Jagharghi's (1993) study of 11 vowels in CVC context with 7 initial consonants and 6 final consonants. Zahorian and Jagharghi reported better pattern classification accuracy for feature sets incorporating spectral change than for static feature sets. However, no acoustic measurements were made of the coarticulatory patterns, making it impossible to relate either the pattern classification results or their listener data to specific context-conditioned effects.

The present study consisted of a replication and extension of Stevens and House, but with several differences in method. The most important of these were: (1) since consonant context effects are nearly certain to be more complex in the nonsymmetrical environments that typically prevail in natural speech, CVCs were recorded in both symmetrical and nonsymmetrical environments, and (2) since we were interested in studying the spectral change patterns for vowels, full format contours were measured rather than sampling the formant pattern once at steady state.

II. METHODS

A. Test signals

Talkers consisted of six men and six women between the ages of 25 and 64. Seven of the speakers were raised in Michigan; the others were from northern Illinois (2), upstate New York (1), Nebraska (1), and northern Ohio (1). All of the speakers were phonetically trained. The speech material consisted of isolated vowels and CVC syllables, only a subset of which was analyzed for the present study. The initial consonants consisted of /h,b,d,g,p,t,k,r,l,w/, the vowels consisted of /i,ɪ,ɛ,æ,ɑ,ʌ,ɔ,ʊ,u,ɝ/), and the final consonants consisted of /b,d,g,p,t,k,r,l/. The initial consonants, vowels, and final consonants were recorded in all combinations. Each of the ten vowels was also recorded in isolation, for a total of 9516 utterances (12 talkers×10 initial consonants×10 vowels×8 final consonants+10 isolated vowels, less 17 unpronounceable combinations, such as /rɝr/). For purely practical reasons, a subset of these recordings was selected for use in the present experiment. Selected for analysis were the eight vowels studied by Stevens and House (/i,ɪ,ɛ,æ,ɑ,ʌ,ʊ,u/) in isolation and in combination with seven initial consonants (/h,b,d,g,p,t,k/) and six final consonants (/b,d,g,p,t,k/).

Recordings were made in a small sound-attenuated booth using a Shure SM58 dynamic microphone. Signals were preamplified, low-pass filtered at 4.3 kHz, and directly digitized at a 10-kHz sample rate using a Tucker & Davis 16-bit A/D. Subjects read from word lists containing the phonetic transcriptions of the utterances to be read. The word lists were blocked by vowel and proceeded in an orderly way through the consonant environments. Subjects were given no special instructions regarding duration or intonation contour, except that they were urged try to avoid a drop in pitch at the end of each page of transcriptions. Individual syllables were later excised from the longer recordings. The signals were auditioned at this time, and if the experimenter noticed an obvious production error, the talker was brought back for another session to re-record the syllables that had been mispronounced. In a few cases the talker was no longer available, and the mispronounced utterances were simply deleted. Twenty-three utterances were deleted in this way, leaving a total of 4105 utterances.

B. Acoustic measurements

The formant estimation methods were similar to those described in Hillenbrand *et al.* (1995). Formant analysis began with the extraction of peaks from 14-pole, 128-point linear predictive coding (LPC) spectra every 5 ms using 25.6-ms Hamming-windowed segments. A graphical display of the spectral peaks was then overlaid on a gray-scale LPC spectrogram. Formant tracks for F_1-F_3 were determined by hand editing the spectral peaks, deleting spurious peaks in some cases, and interpolating through ''holes'' in the formant track in other cases. The number of LPC poles was occasionally increased to separate merged formants. In some cases—many of them involving formant mergers—it was judged that a formant could not be measured with confidence. In these cases, zeros were written into the formant slot, and the values for that formant were simply omitted from all subsequent analyses. Formants were edited only between the starting and ending times of the vowel, which were determined by visual inspection of the LPC spectrograms. Measures of vowel duration included the vocalic segment only and not the initial burst associated with consonant release.

Vowel ''steady-state'' times were determined automatically. We experimented with a number of algorithms and settled on a simple technique that seemed to show the best agreement with the visual inspection method that has been used in many previous studies. The vowel formant contour was reduced to an array of $\log F_2 - \log F_1$ values (Miller, 1989). The sum of differences between adjacent frames was then calculated for every sequence of five frames (35 ms) throughout the first 60% of the vowel.[1] Steady-state time was defined as the middle of the sequence of five frames showing the smallest absolute summed difference.

Fundamental-frequency contours were measured using a conventional autocorrelation pitch tracker (Hillenbrand, 1988), followed by hand editing using the tool described above. If there was any uncertainty about the F_0 contour, the experimenter examined the time waveform and a narrow band spectrogram.

C. Listening test

The test signals were presented for identification to 24 phonetically trained listeners. The listeners were first- and second-year graduate students in speech-language pathology. The listeners spoke the same Northern Cities dialect as the speakers, with roughly 80% of the listeners from Michigan, and the remainder from other areas of the upper Midwest, such as the northern parts of Indiana, Illinois, and Ohio. Subjects were tested individually in a quiet room in four sessions of about an hour each. Stimuli were scaled to maximum peak amplitude, low-pass filtered at 4.3 kHz at the output of a 16-bit D/A converter, amplified, and delivered to a single loudspeaker positioned about 1 m from the listener's head at an average intensity of approximately 77 dBA. Over the course of the four sessions, each listener identified one presentation of each of the 4105 signals. The order of presentation was fully randomized (i.e., not blocked by talker or context), and the presentation order was shuffled separately for each listener. Listeners responded by pressing one of eight keys on a computer keyboard that had been labeled with phonetic symbols for the vowels. The listening test was self-paced, and subjects could repeat a stimulus as many times as they wished before entering a response. Each listening test was preceded by a brief practice session to ensure that listeners understood the task and interpreted the phonetic symbols appropriately.

III. RESULTS

A. Listening test

The average identification rate for the test signals was 94.6%, with nearly identical rates for the male (94.5%) and female (94.8%) talker groups. The majority (61.7%) of the individual tokens were correctly identified by all 24 listeners, and 86% of the signals were identified as the intended vowel by at least 90% of the listeners. For 78 signals (1.9%) the label assigned by a plurality of the panel was a vowel other than that intended by the talker. The most common of these misidentifications consisted of tokens that were intended as /æ/ but heard as /ɛ/ (40% of the signals misidentified by a plurality of the listeners) and tokens that were intended as /ɛ/ but heard as /ɪ/ or /æ/ (23% of the signals misidentified by a plurality of the listeners). Average intelligibility for individual talkers varied from 88.7% to 98.0% (s.d.$=2.6$).

As seen in Fig. 2, intelligibility was higher for some vowels than others. As we will show below, the differences in identification rates across vowels are highly significant. Despite this, however, it is important to note that all individual vowels were well identified. Average rates varied from about 90% to 99%, with a standard deviation across vowels of only 3.4%.

Vowel intelligibility was also affected somewhat by consonant environment. Figure 3 shows identification rates, averaged across all vowels, as a function of both initial consonant and final consonant. (Labeling data for the isolated vowels are not shown in the figure.) It can be seen that the consonant environment effects are quite small in magnitude, with only 3.3% separating the most intelligible contexts from the least intelligible. The standard deviation computed across

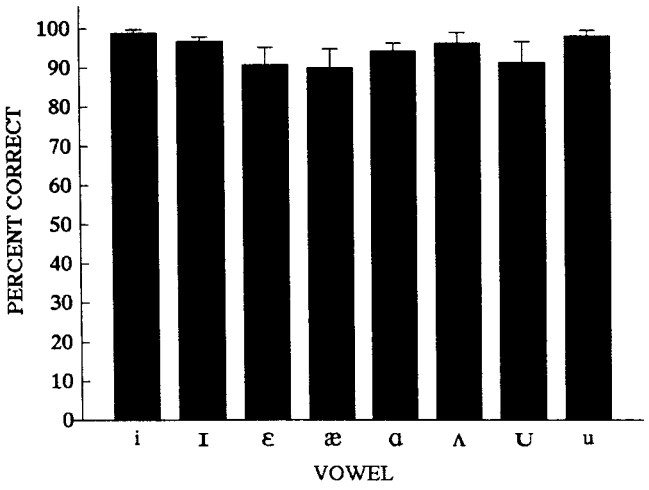

FIG. 2. Vowel intelligibility averaged across all consonant environments. Error bars show one standard deviation.

the seven average identification rates for initial consonants (i.e., the black bars in Fig. 3) is only 0.8%, and the standard deviation computed across the six average identification rates for final consonants (the open bars in Fig. 3) is only 1.3%.

1. Statistical treatment of labeling data

The two sources of individual variability in the perceptual responses that are reasonably viewed as random are speakers and listeners. Accordingly, repeated measures analyses of variance (ANOVAs) were run by listener (i.e., pooling syllable scores over talkers and treating listeners as random effect) and by talker (i.e., pooling syllable scores over listeners and treating talker as a random effect).[2] In each case, Studebaker's (1985) rationalized arcsine transformation was applied to the percent-correct values after pooling. Following a practice common in the psycholinguistics literature, we will report F ratios separately by listener (F_L) and by talker (F_T). We will also report the F'_{min} (Clark, 1973), which we will use to determine significance levels. In both cases, there are five treatment factors: initial consonant place, initial consonant voicing, vowel, final consonant place, and final consonant voicing.

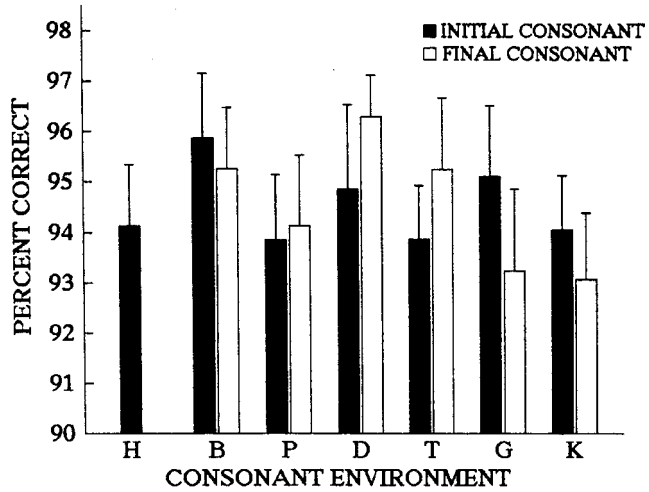

FIG. 3. Percent-correct vowel identification for each of seven initial consonant environments and six final consonant environments.

The usual procedures for *post hoc* comparison of means are clumsy in higher-order factorial analyses of variance. Consequently, we have chosen to follow up significant F tests with t tests derived from effect estimates in the linear model underlying the ANOVA. A significant t value for a specific coefficient of a main or interaction effect indicates that the coefficient in question was significantly different from the average of the entire family of coefficients for that effect. A significant t value can be also be interpreted as indicating that the cell mean associated with the coefficient in question is significantly different from the grand mean and any main effect's lower-order interaction terms. Significance levels of t tests are estimated using the Sidak approach to multiple comparisons with a family size equal to the number of effect coefficients for the main effect or interaction term in question.

The main effect for vowel was highly significant [$F_L(7,161)=30.1$, $F_T(7,77)=3.9$, $F'_{min}(7,97)=3.4$, $p<0.01$]. It is useful to have an indication of the relative contribution of phonetic factors. We will use *percent of total phonetic variation accounted for*, defined as the ratio of the sum of squares for a given main or interaction effect to the sum of squares of all phonetic factors. We base this measure on the *by-talkers* analysis. The vowel main effect was easily the most important of any of the phonetic effects that were observed, accounting for 41.2% of total phonetic variation (TPV). Sidak-corrected planned comparisons revealed that /u/ and /i/ were identified significantly better than average: /u/ by 1.1 percentage points [$t'_{min}(16)=4$, $p_{min}<0.01$], and /i/ by 1.2 percentage points [$t'_{min}(16)=7.1$, $p_{min}<0.000\,01$]. The main effect for final place of articulation was also highly significant [$F_L(2,46)=79.7$, $F_T(2,22)=12.2$, $F'_{min}(2,29)=10.6$, $p<0.001$] and accounted for about 5.7% of TPV. Sidak-corrected planned comparisons showed that vowels in the environment of final velars were identified less well than average by about 0.9 percentage points [$t'_{min}(14)=-3.4$, $p_{min}<0.01$], while those in final alveolar contexts were identified about 0.6 percentage points better than average [$t'_{min}(14)=4.0$, $p_{min}<0.001$]. No other main effects were significant.

Although they were all small in absolute magnitude, several interactions reached significance. The interaction pattern is displayed in Fig. 4, which shows percent correct as a function of the initial consonant, with the final consonant as the parameter. The vowel by final place interaction was highly significant [$F_L(14,322)=20.4$, $F_T(14,154)=4.4$, $F'_{min}(14,222)=3.6$, $p<0.0001$], accounting for about 9.6% of TPV. Sidak-corrected planned comparisons showed that syllables ending in /æ/+labial stops were identified about 2.1 percentage points better than average [$t'_{min}(18)=3.9$, $p_{min}<0.05$]. The vowel by final voicing interaction was also significant [$F_L(7,161)=36.8$, $F_T(7,77)=2.9$, $F'_{min}(7,89)=2.7$, $p<0.01$], accounting for about 4.9% of TPV. However, Sidak-adjusted comparisons failed to identify any specific effect as significantly above average.

The three-way initial voicing by initial place by vowel

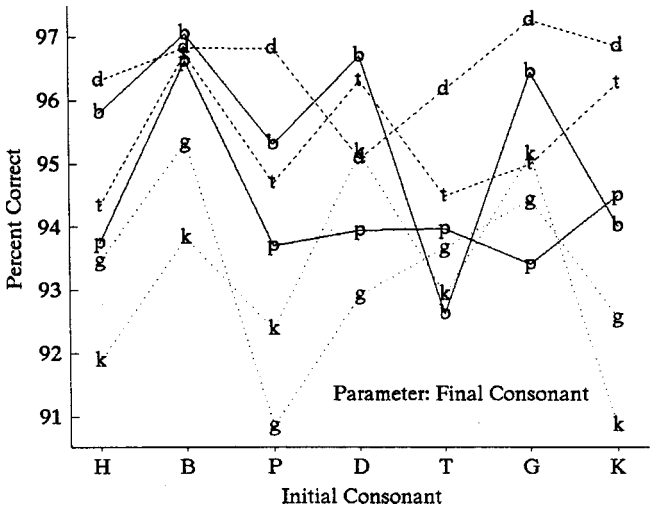

FIG. 4. Vowel intelligibility as a function of the initial consonant and final consonant.

interaction was highly significant [$F_L(14,322)=14.2$, $F_T(14,154)=3.8$, $F'_{min}(14,240)=3.0$, $p<0.001$], accounting for about 5.9% of TPV. The three-way interaction of initial voicing by final voicing by final place [$F_L(2,46)=21.6$, $F_T(2,22)=5.0$, $F'_{min}(2,32)=4.0$, $p<0.05$] and the four-way interaction of initial place by initial voicing by final place by final voicing [$F_L(4,92)=18.4$, $F_T(4,44)=3.4$, $F'_{min}(4,61)=2.8$, $p<0.05$] were also significant. These interactions accounted for less than 1% of the TPV. In no case for three- or four-way interactions did Sidak-adjusted comparisons identify specific factor combinations as significantly different from average for the family in question.

An important point which we hope does not get lost in the details of the ANOVA results reported above is that the influences of consonant environment on average vowel recognition rates are rather small in absolute magnitude. For example, as can be seen in Fig. 4, the full range of variation separating the most intelligible from the least intelligible contexts is only about 6%, and the standard deviation in average recognition rates computed over all 42 phonetic environments displayed in Fig. 4 (e.g., /hVb/, /hVd/, /hVp/,..., /kVk/) is a very modest 1.7%.

A final note on the listening test results concerns labeling data for /u/ and /ʊ/ in alveolar contexts. Recall that the largest context effects observed by Stevens and House (1963) consisted of a raising of F_2 for /u/ and /ʊ/ in the environment of alveolar consonants relative to the same vowels in null environments (Fig. 1). As will be discussed below, our formant measurements showed an even larger effect, averaging about 500–600 Hz for /u/ and about 200–300 Hz for /ʊ/. Our results further showed that the effect is conditioned primarily by the presence of a *syllable-initial* alveolar. It was therefore of some interest to determine whether there is any evidence that these context-conditioned shifts in formant values have an adverse effect on vowel intelligibility. The recognition rates for /u/ and /ʊ/ in the environment of syllable-initial alveolars turn out to be unremarkable. Average recognition rates, pooled across all contexts except initial alveolar, are 98.1% for /u/ and 91.5% for /ʊ/. These figures compare with nearly identical rates in initial alveolar contexts of 98.3% for /u/ and 92.5% for /ʊ/. There is, in short, no evidence that the largest of the context-conditioned shifts in vowel formants caused any difficulty for the listeners.

B. Acoustic measurements

1. Vowel duration

Vowel durations in various consonant voicing environments are shown in Table I. To simplify the examination of voicing effects, initial /h/ environments were excluded from the computation of the means reported in the first four columns of the table. The average durations associated with the eight vowels, pooled across all consonant environments, are strongly correlated with average durations from the /hVd/ data of Hillenbrand *et al.* (1995), and with the /tVp/ data of Black (1949). The widely observed increase in duration for vowels preceding voiced versus unvoiced stops is quite evident in our data (i.e., compare V–V with V–U and U–V with U–U). Also apparent in Table I is evidence for systematically longer vowels when *preceded by* voiced versus unvoiced stops (i.e., compare V–V with U–V and V–U with U–U in Table I). This effect, which averages some 20–40 ms, was confirmed by two separate ANOVAs for vowel and initial consonant voicing, one comparing durations for V–V environments with U–V environments (i.e., C_1=voiced/C_2=voiced versus C_1=unvoiced/C_2=voiced) and the second

TABLE I. Vowel durations in ms in different stop-consonant voicing environments, and in all consonant environments. Measurements for isolated vowels were excluded. Standard deviations shown in parentheses. (V–V=voiced initial consonant, voiced final consonant; U–V=unvoiced initial consonant, voiced final consonant; V–U=voiced initial consonant, unvoiced final consonant; U–U=unvoiced initial consonant, unvoiced final consonant.)

Vowel	V–V	U–V	V–U	U–U	All consonant environments
/i/	255.9 (46.8)	233.6 (48.3)	169.8 (32.8)	144.1 (34.1)	198.7 (61.3)
/ɪ/	190.6 (29.3)	174.2 (30.7)	137.3 (32.6)	116.7 (28.1)	153.1 (41.7)
/ɛ/	218.2 (28.3)	191.1 (29.8)	160.3 (31.6)	127.8 (27.9)	176.1 (44.7)
/æ/	331.8 (50.5)	286.0 (43.1)	254.0 (50.5)	214.0 (40.6)	266.6 (65.6)
/ɑ/	328.9 (57.3)	290.3 (41.9)	235.7 (50.1)	194.2 (45.2)	255.9 (73.8)
/ʌ/	215.0 (35.4)	178.8 (30.2)	146.6 (30.9)	118.6 (24.3)	162.9 (49.2)
/ʊ/	208.8 (35.2)	189.7 (31.2)	152.9 (36.0)	124.7 (31.9)	166.2 (46.6)
/u/	261.2 (46.5)	241.9 (44.5)	171.7 (37.1)	147.3 (31.3)	203.6 (66.6)
All vowels	251.8 (65.7)	223.5 (58.8)	177.2 (54.2)	146.3 (46.3)	198.0 (69.3)

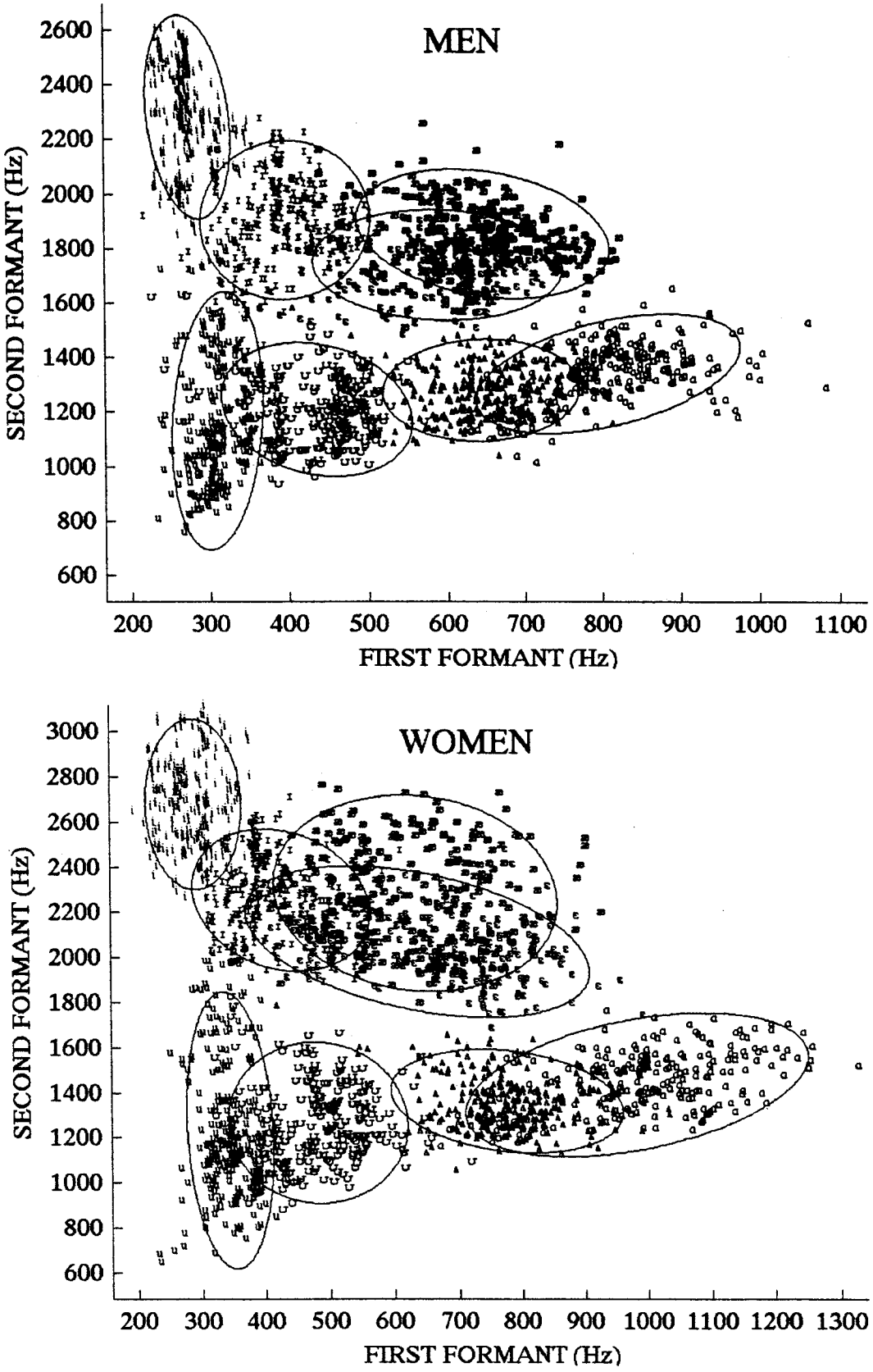

FIG. 5. Values of F_1 and F_2 measured at steady state for men (top panel) and women (bottom panel). Formant measurements are plotted for isolated vowels and for all consonant environments. Not displayed are measurements for signals with identification error rates of 15% or higher. To improve the clarity of the display the database was thinned by removing redundant data points, resulting in the display of measurements from about two-thirds of the well-identified tokens.

comparing V–U with U–U environments. Both ANOVAs showed highly significant effects for vowel and initial consonant voicing. This effect is consistent with Fischer-Jorgensen (1964) and Crystal and House (1988), but differs from the conclusions reached by Peterson and Lehiste (1960).

2. Average formant values

Values of F_1 and F_2 measured at steady state are displayed in Fig. 5. Formant measurements are plotted for isolated vowels and for all consonant environments. Not displayed in Fig. 5 are measurements for signals with identification error rates of 15% or higher. To improve the clarity of the display the database was thinned by removing redundant data points, resulting in the display of measurements from about two-thirds of the well-identified tokens. There is, of course, considerable variability in formant values within each vowel category, and a good deal of overlap among vowels.[3] A major goal of the present study was to determine what aspects of this variability were associated with consonant environment.

3. Effects of place of articulation

Figure 6 shows the effects of place of articulation on the frequencies of F_1 and F_2 for syllables that are symmetrical with respect to place of production (e.g., /bVb/, /bVp/, /pVp/, /pVb, /dVd/, /dVt/, etc.). Also plotted are formant values for isolated vowels and /hVd/ syllables, environments referred to as ''null'' by Stevens and House (1963). The general look of this figure is similar to the Stevens and House (SH) data (Fig. 1), which were based on strictly symmetrical syllables (/bVb/, /dVd/, /gVg/, etc.). As with SH, the largest effect by far is a raising of F_2 for /u/ in the environment of alveolar consonants. At about 500 Hz for the men and nearly 600 Hz for the women (relative to null environments), this upward shift is even larger than the roughly 350 Hz effect reported by SH.[4] Sizable upward shifts in F_2 for alveolar environments are seen for the remaining back/central vowels, especially /ʊ/, with shifts averaging 214 Hz for men and 281 Hz for women. Also seen for the back/central vowels was a fairly uniform upward shift in F_2 averaging 98 Hz for men and 117 Hz for women for the velar environments. For front vowels, the most consistent effect is a downward shift in F_2 of some 85–100 Hz for labial environments. As in the SH results, the effects of place on F_1 values tend to be rather small. The only moderately sizable effect that appears to be consistent across men and women is a downward shift in F_1 averaging some 50 Hz for /ɛ/ and /æ/ in the environment of alveolar and velar consonants.

The effects of place on formant values for syllables that are either symmetrical or nonsymmetrical with respect to place of articulation are shown in Figs. 7 and 8. Figure 7 shows the effects of initial consonant environment, while Fig. 8 shows the effects of final consonant environment. In Fig. 7, showing the effects of initial consonant place, averages that are plotted with the square symbols for labials, for example, were pooled over all syllables with C_1=/b,p/, regardless of the final consonant. Similarly, in Fig. 8, the square symbols represent averages pooled over all syllables with C_2=/b,p/, regardless of the initial consonant. Although there are many minor differences, the general look of Fig. 7 (initial environments) is quite similar to that of Fig. 6 (symmetrical environments only). However, there are some very important differences between Fig. 8 (final environments) and both Fig. 6 and Fig. 7. It can be seen, for example, that the upward shifts in F_2 for /u/ and /ʊ/ in alveolar environments are much smaller for final alveolars than initial alveolars. Similarly, the upward shifts in F_2 for /ʌ/ and /ɑ/ in alveolar and velar environments and the downward shifts in F_2 for front vowels in labial environments are much more pronounced when the relevant environments are initial rather than final. The conclusion from these comparisons is that the place-dependent effects for symmetrical environments seen in Fig. 1 from SH and Fig. 6 from the present study reveal primarily the effects of the place of articulation of the initial consonant rather than the final consonant.

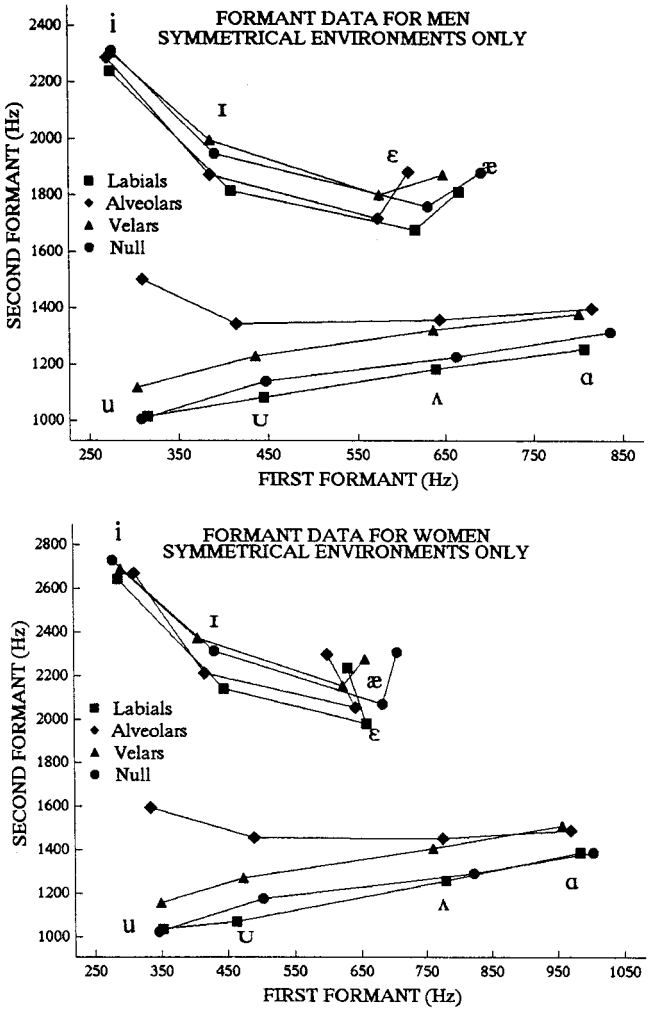

FIG. 6. Average formant frequencies at steady state as a function of the place of articulation of the surrounding consonants for men (top panel) and women (bottom panel). Data are shown for symmetrical environments only.

4. Effects of consonant voicing

Figure 9 shows the effects of consonant voicing for syllables that are symmetrical with respect to the voicing feature; for example, the data points identified as voiced (the

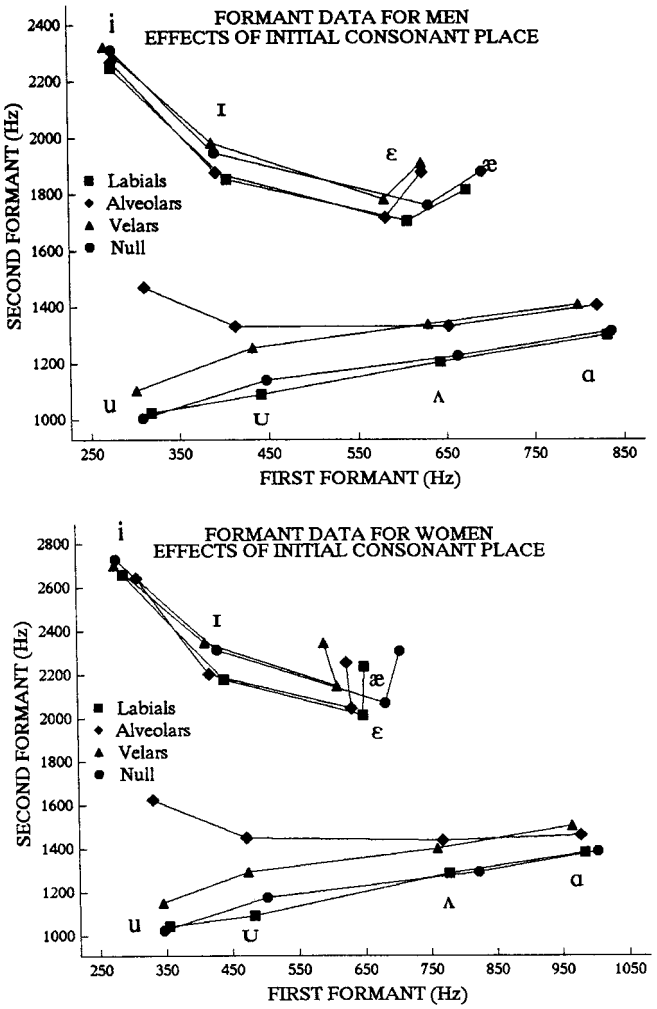

FIG. 7. Average formant frequencies at steady state as a function of the place of articulation of the initial consonant for men (top panel) and women (bottom panel).

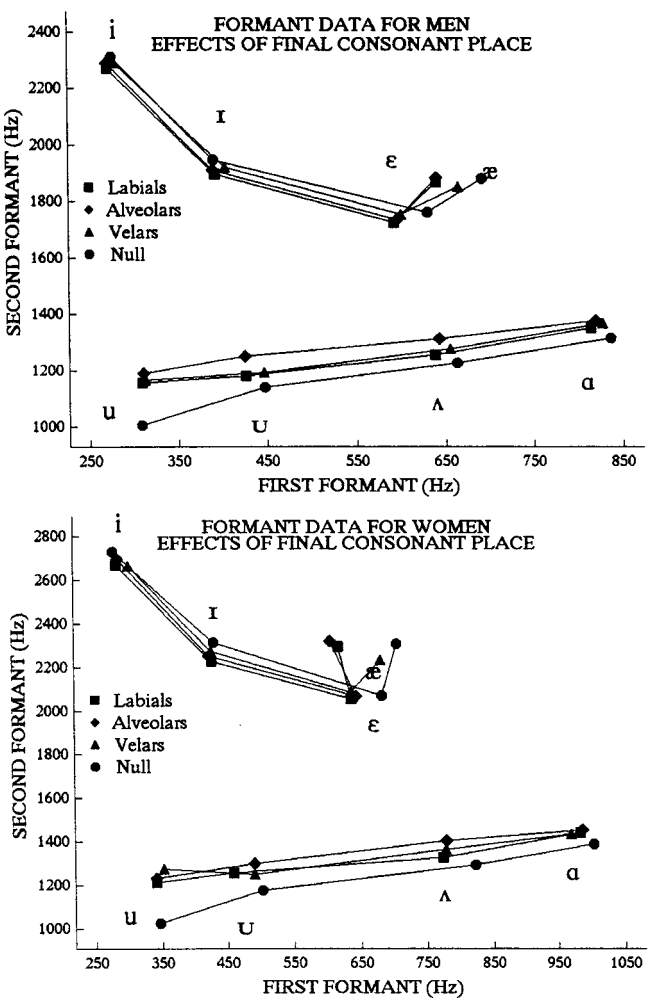

FIG. 8. Average formant frequencies at steady state as a function of the place of articulation of the final consonant for men (top panel) and women (bottom panel).

filled squares) represent averages pooled over all syllables with voiced initial and final consonants, regardless of place of production. For reference, formant values for the null environments are also shown. For the back/central vowels, the most consistent effect appears to be a tendency for the F_1 values of vowels flanked by voiced stops to be slightly lower than those in unvoiced-stop environments. For /ʌ/, the difference in F_1 between voiced and unvoiced environments is approximately 75 Hz for both men and women, but for the remaining back/central vowels the difference is quite small, typically averaging no more than 15–20 Hz. For the front vowels, the largest voicing-related differences are downward shifts in F_1 in voiced environments averaging about 90 Hz for /ɪ/, 90–120 Hz for /ɛ/, 70–100 Hz for /æ/, and negligible for /i/. In general, the tendency for F_1 values to be somewhat lower in the environment of voiced consonants is consistent with the findings of SH. We assume that these shifts in F_1 values are due at least in part to the slightly lower position of the larynx for voiced as compared to unvoiced consonants, with this difference carrying over into the vowel in the case of initial consonants and being anticipated in the case of final consonants.

Figures 10 and 11 show the effects of consonant voicing separately for initial and final consonant environments. For example, in Fig. 10, which shows initial environments, the values identified as unvoiced represent averages pooled over all non-null syllables with an unvoiced initial consonant, independent of either the place or voicing of the final consonant. Based strictly on visual inspection, Figs. 10 and 11 appear to show the same kinds of effects that were seen in symmetrical environments, but reduced in magnitude. The voicing effects, therefore, appear to derive approximately equally from initial and final consonants. The apparent attenuation of voicing effects in Figs. 10 and 11 is not surprising since half of the syllables whose formant values were pooled to calculate the means identified as voiced in Fig. 10, for example, were from syllables containing an unvoiced final consonant.

5. Spectral change patterns

Formant movement patterns, averaged across all phonetic environments, are shown in Fig. 12. The figure was created by connecting a line between the average formant values sampled at 20% of vowel duration and the average formant values sampled at 70% of vowel duration. The symbol for each vowel category is plotted at the location of the

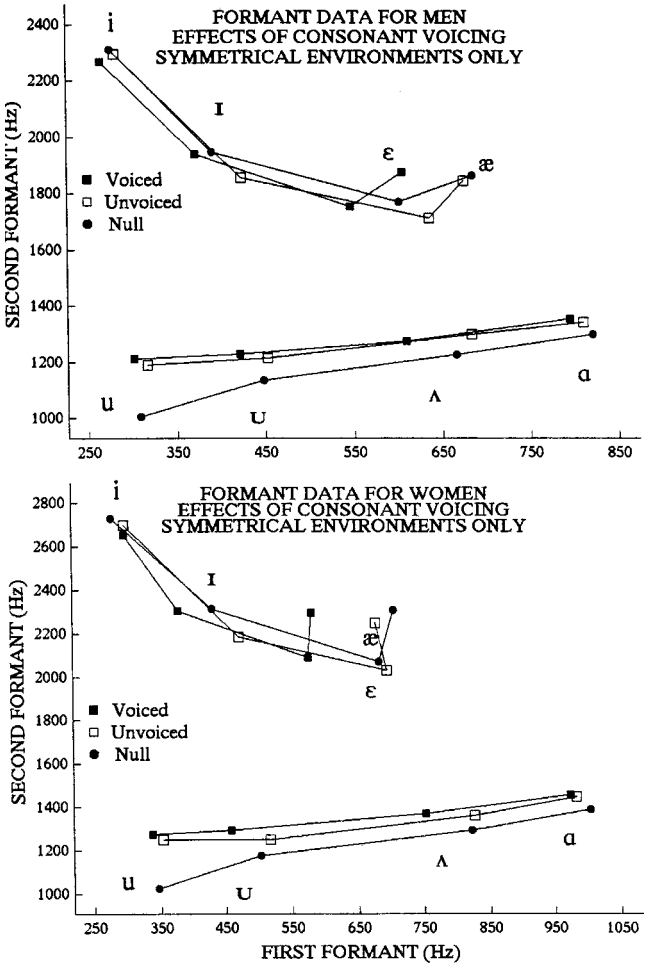

FIG. 9. Average formant frequencies at steady state as a function of the voicing of the surrounding consonants for men (top panel) and women (bottom panel). Data are shown for symmetrical environments only.

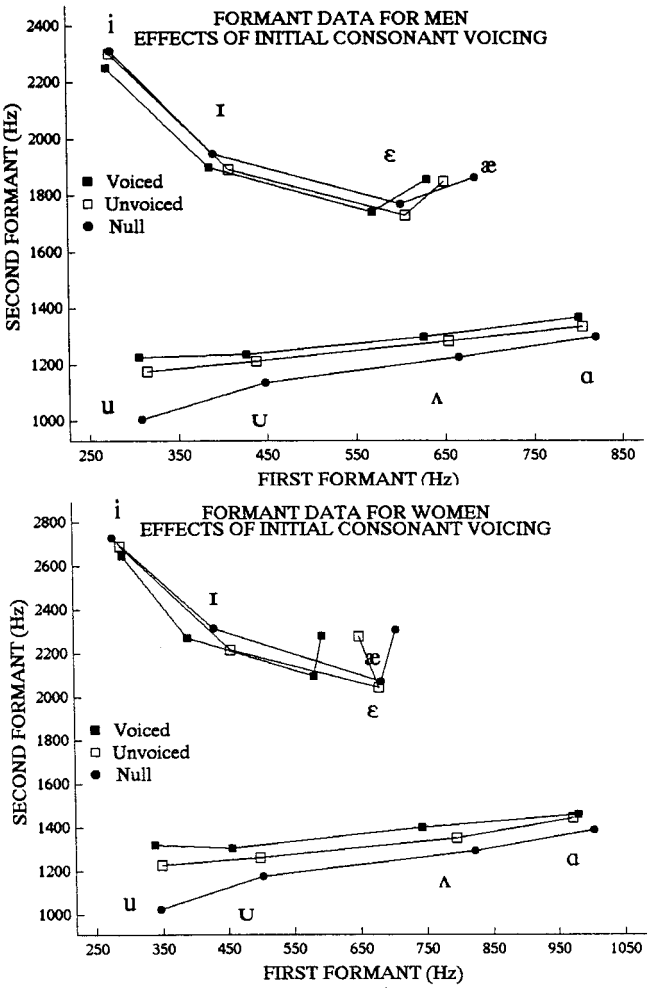

FIG. 10. Average formant frequencies at steady state as a function of the voicing of the initial consonant for men (top panel) and women (bottom panel).

second measurement; the larger symbols show formant values for the men. There are some similarities between these spectral change patterns and those observed in our earlier study of /hVd/ utterances (e.g., compare Fig. 12 with Fig. 1 of Hillenbrand and Nearey, 1999), but there are some important differences as well. Differences include: (a) the centralized offglide that was observed for /ʊ/ in the /hVd/ data is apparent in Fig. 12 as well, but the magnitude of the spectral movement is considerably attenuated; (b) the centralized offglide that was observed for /ʌ/ in the /hVd/ data is not evident in the present data; (c) the modest centralized offglide that was observed for /ɑ/ in the /hVd/ is not evident in the present data; in fact, a small movement toward the periphery is seen; and (d) the rather small centralized offglide that was observed for /ɛ/ in the /hVd/ is not evident in the present data; instead, a small movement toward the periphery was seen. Average spectral change patterns for /i/, /ɪ/, /æ/, and /u/ are grossly similar to those observed in the /hVd/ data.

6. Statistical treatment of acoustic data

A five-way factorial repeated measures analysis of variance was undertaken for the acoustic data for the stop-vowel-stop syllables.[5] The same treatment factors were considered here as in the previous analysis of the identification data, namely voicing and place of the initial consonant, vowel, and voicing and place of the final consonant. In all cases, formant frequencies were log transformed since there are clear indications that this improves the homogeneity of variance. Significance levels were determined using the Greenhouse–Geisser procedure for all F tests that involved more than one degree of freedom in the numerator. Many main effects and interactions turn out to be significant even by this conservative procedure. However, a substantial subset of the nominally significant effects accounts for a very small amount of the total phonetic variability (TPV). We have chosen to discuss only those significant interactions that account for at least 0.25% of the total variance due to all phonetic factors.

a. F_1. Significant main effects were found for all five phonetic factors: initial voicing, initial place, vowel, final voicing, and final place. (F values, F probabilities, and other numerical details concerning these and all other ANOVA results on the acoustic measurements can be found in the Appendix.) Not surprisingly, vowel effects were dominant in F_1, accounting for 97.6% of TPV. Initial and final voicing effects, though accounting for a very small proportion of TPV (0.2% and 0.3%, respectively), showed patterns consistent with previous observations. For initial consonants, voiced stops showed formant frequencies about 5.4% lower than voiceless. Initial and final place accounted for only

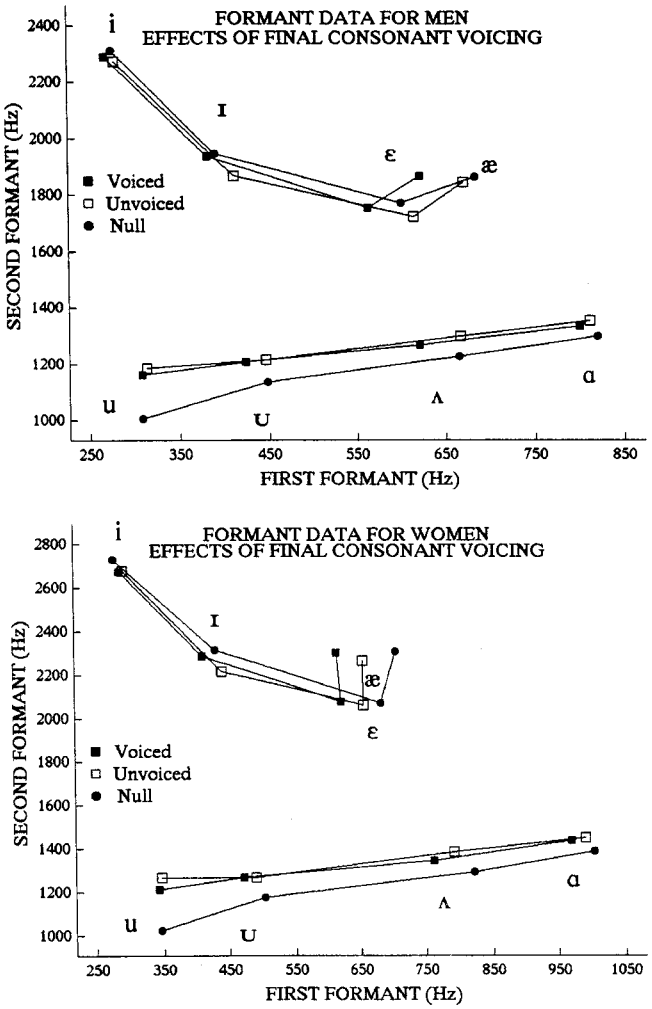

FIG. 11. Average formant frequencies at steady state as a function of the voicing of the final consonant for men (top panel) and women (bottom panel).

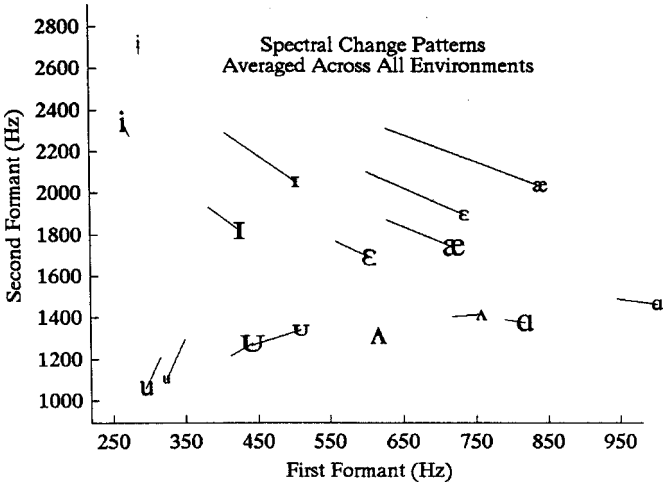

FIG. 12. Spectral change patterns for eight vowels averaged across all phonetic environments. The phonetic symbol identifying each vowel is plotted at the F_1-F_2 value for the second sample of the formant pattern (measured at 70% of vowel duration), and a line connects this point to the first sample (measured at 20% of vowel duration). The larger phonetic symbols designate formant values for men.

0.2% and 0.1% of phonetically induced variation in F_1. Sidak-corrected tests of contrasts showed that initial labials had F_1 values that were 2.4% above the mean, while velars were lower by about 2.1%. Final place effects were in the opposite direction, with final labials about 1.2% lower and final velars about 1.8% higher than average. There was only one significant interaction for F_1 that reached the variance-proportion criterion, namely initial voicing by vowel. This effect accounted for only about 0.3% of TPV. Sidak-corrected tests failed to yield any single contrast that was significantly different from zero.

b. F_2. Four of five main effects were significant for F_2. Vowels accounted for 91.2% of TPV (compared to over 97% for F_1). Initial place accounted for about 2.7% of TPV. Sidak-adjusted tests of contrasts revealed initial labial to be significantly (5.8%) lower, initial velars to be about 1.4% higher, and initial alveolars to be 4.7% higher than average. These place-induced deviations are consistent with the patterns noted by SH. Final place, while still significant, had substantially less effect on F_2 at steady state, accounting for only about 0.1% of TPV. Sidak tests showed that final labials were about 0.9% lower and final alveolars about 1.1% higher in frequency than average. Initial voicing accounted for about 0.1% of TPV, with initial voiceless stops showing F_2 values about 1.4% lower than voiced. Final voicing was not significant.[6]

By far the largest two-way interaction was initial place by vowel, which accounted for 4.8% of TPV. A large number of contrast coefficients were significant by the Sidak-adjusted test (see the Appendix). We will summarize the general findings here. The terms ''lower'' (or ''higher'') below can be interpreted as meaning that F_2 steady states for the initial place by vowel combination in question were lower (higher) than expected after adjusting for the main effects of the initial place and vowel in question. Significantly lower than expected were labial (by 9.7%) and velar (by 8.8%) before /u/; alveolars before front vowels (by 4.5% to 5.8%) and before /ɑ/ (by 2.8%). Significantly higher than expected were alveolars before /u/ (by an egregious 21.4%) and before /ʊ/ (by 5.8%); and labials before the front vowels /ɪ/, /ɛ/, and /æ/) (by 3.1% to 3.9%). Velars were also slightly higher than expected before /ʌ/ (by 1.6%) and before /ɪ/ (by 3.1%). Taken together with the main effects for place, the general trends can be viewed as being compatible with a degree of assimilation of the F_2 steady state towards roughly the expected F_2 locus for the initial consonant.

Although a number of other second- and higher-order interactions were significant, none accounted for more than about 0.2% of TPV, a criterion that corresponds to less than about 10% of the size of the main effect of initial place, or 5% of the initial place by vowel interaction.

c. F_3. The main effect of vowel accounted for 92.7% of TPV in F_3, with /u/ and /ʊ/ showing lower than average F_3 by about 8.5% and 5.6%, respectively. The vowels /i/ and /ɪ/ showed significantly higher than average F_3, by 15.8% and 2.1%, respectively. (We did not discuss the significance of contrast coefficients for vowel main effects in F_1 or F_2 because they are all significant and their pattern reflects the well-known expected locations of the vowel means in F_1-F_2 space.)

The main effect of initial place accounted for 0.9% of the TPV in F_3. Only initial velars showed a significant effect

size, being lower than average by 0.9%. The main effect of final voicing accounted for 0.8% of the TPV, with final voiceless consonants showing F_3 values about 1.2% lower than voiced. Final place of articulation accounted for about 0.3% of TPV. Sidak-adjusted tests of contrasts revealed final velars to be 0.4% lower and final alveolars to be 0.5% higher than average.

The initial place by vowel interaction was also significant, accounting for 2.7% of TPV (considerably more than the place main effect). Sidak comparisons showed initial alveolars and labials before /i/ and labials before /ɪ/ to be significantly lower than average (by 2.1% to 0.9%) while labials before /ʌ/ and /ɑ/ and velars before /i/ were higher than average (by about 1.1% to 3.2%). One additional two-way interaction, initial voicing by vowel, was also significant for F_3, accounting for 0.4% of TPV. However, Sidak tests of contrasts failed to identify any specific interaction coefficients that deviated significantly from average.

d. Summary. Statistical tests revealed a number of reliable effects of phonetic context on steady-state formant frequencies. By far the largest of these are associated with initial place of articulation in F_2. The general tendencies are in accord with preliminary observations of SH. Examination of the strong interaction effects of place with vowel also confirms that the effects of alveolars on the vowels /ʊ/ and especially /u/ are particularly strong. The large number of reliable effects in the production data contrasts with the paucity of context effects in perception. Furthermore, the largest effects in perception do not appear to correspond to those in the production data.

C. Discriminant analyses

Recall that in our earlier study of /hVd/ utterances (Hillenbrand *et al.*, 1995), discriminant analyses showed that vowels could be separated with substantially greater accuracy for pattern recognition models that incorporated spectral change as compared to otherwise comparable models that were trained on the formant pattern sampled at a single cross section of the vowel. The main purpose of the discriminant analyses reported here was to determine whether formant-frequency movements contribute to the separability of vowel categories for more complex CVC utterances in which formant movements are affected both by vowel identity and consonant environment. The pattern recognizer was a quadratic discriminant analysis technique (Johnson and Winchern, 1982) that was trained on various combinations of F_0, duration, and the three lowest formant frequencies. The formant values were sampled: (a) a single time at steady state, or (b) once at 20% of vowel duration and a second time at 70% of vowel duration.[7] For each parameter set, the pattern recognizer was run 12 separate times. On each run, the classifier was trained on 11 of the 12 talkers and tested on tokens from the single talker whose utterances had been omitted from the training. Excluded from both training and testing were: (a) tokens showing an unmeasurable formant in a formant slot that was included in the parameter list for a particular test, and (b) any token with a listener identification error rate of 15% or greater. In all cases linear frequencies in Hz were used.

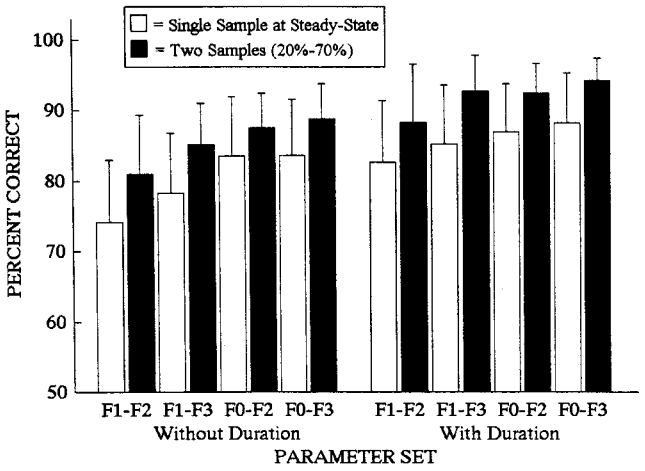

FIG. 13. Overall vowel classification accuracy for a quadratic discriminant classifier trained on various combinations of parameters.

Figure 13 shows recognition accuracy for the pattern classifier averaged across the 12 talkers for 16 different parameter sets, with the error bars showing the standard deviation calculated across the 12 talkers. It can be seen that the accuracy of the pattern classifier is higher when the model incorporates spectral change for all eight combinations of acoustic features. Averaged across the feature sets, classification accuracy was 6.1% higher for two samples of the formant pattern as compared with a single sample at steady state.[8] As shown in Table II, the improvement in classification accuracy for the two-sample models varies across vowels. Improvement with the addition of spectral change is the greatest for /ɪ/, /ɛ/, and /æ/, a cluster of vowels showing a good deal of overlap in static formant space (see Fig. 5).

As has been noted in other pattern recognition studies, there is also a substantial improvement in category separability with the addition of vowel duration. Averaged across the feature sets, classification accuracy was 6.3% higher with duration than without. As shown in Table II, very large improvements in classification accuracy averaging some 22%–25% were seen for /æ/ and /ɛ/. Substantial improvements of

TABLE II. Improvement in discriminant classification accuracy for each vowel with the addition of (a) spectral change (column 2) or (b) duration (column 3). Column 2 shows the average improvement in classification accuracy for two samples of the formant pattern as compared to a single sample. The averages were computed over the eight acoustic feature sets (i.e., F_1–F_2_NoDuration, F_1–F_3_NoDuration, F_0–F_2_NoDuration, F_0–F_3_NoDuration, F_1–F_2_Duration, etc.). Column 3 shows the average improvement in classification with the addition of duration as compared to otherwise identical parameter sets, averaged over the eight acoustic feature sets.

Vowel	Improvement with spectral change	Improvement with duration
/i/	2.5	0.4
/ɪ/	14.0	11.1
/ɛ/	15.2	24.5
/æ/	12.8	21.9
/ɑ/	9.8	12.9
/ʌ/	2.8	5.1
/ʊ/	3.5	0.6
/u/	4.2	0.2

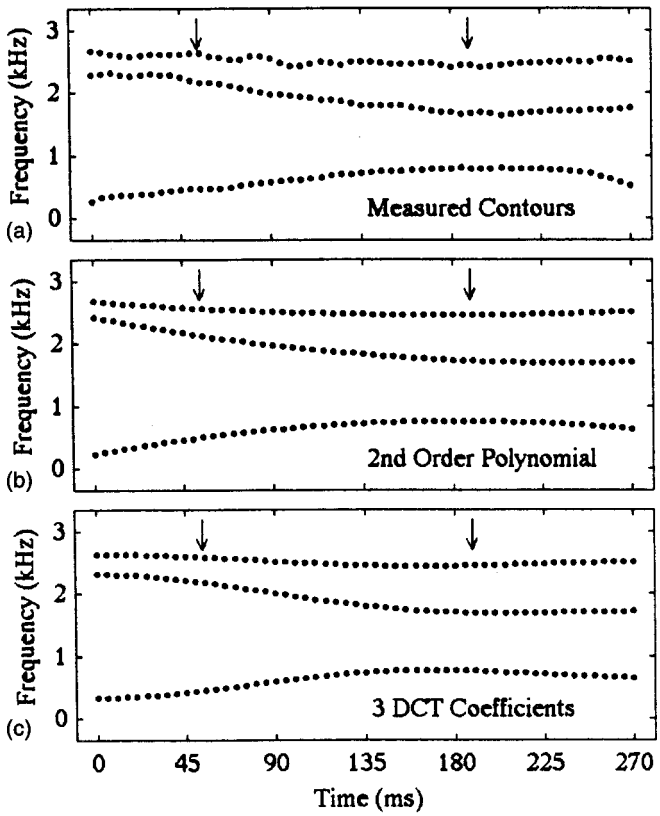

FIG. 14. From top to bottom: (a) the measured contours of F_1-F_3 for the vowel /æ/ in /gæd/; (b) a second-order polynomial fit to the contours; and (c) a three-coefficient discrete cosine transform fit to the contours. Arrows are drawn at 20% and 70% of vowel duration (see the text).

some 11%–13% were also seen for /ɪ/ and /ɑ/.

The two-sample method of capturing formant movements that was used in the present study, and in several previous studies, is not especially elegant and requires the more-or-less arbitrary selection of two discrete time points at which to sample the formant values. We experimented with two alternate schemes for capturing formant movements. One method involved fitting *n*th-order polynomials to the contours of F_1-F_3, followed by training and testing of the discriminant classifier on the coefficients of the polynomial. The curve fit was applied either to the full vowel or to the formant values from 20% to 70% of vowel duration. We experimented with several different orders of polynomial fits and different choices of sampling points. The second method involved the use of discrete cosine transform (DCT) coefficients to code the contours of the three lowest formants, as described by Zahorian and Jagharghi (1993). As with the polynomial method: (a) the coefficients were computed from the formant values of either the full vowel or the portion of the vowel from 20%–70% of vowel duration, and (b) we experimented with different numbers of DCT coefficients (see Fig. 14). The results from these two methods were not sufficiently promising to merit detailed description. Our main conclusion from this work is that both the polynomial and DCT method produced good classification results, but neither method was found to be superior to the simpler two-sample method that has been used in previous work.

For the listener data, we reported the results of statistical analyses showing the effects of consonant environment on recognition accuracy. This kind of statistical analysis is not possible with the pattern classification results for the simple reason that each token is given a single label by the pattern recognizer, providing no error term comparable to the variability in labeling responses across listeners. In the section below, we report the results of preliminary analyses that compare listener labeling responses with the output of the pattern classifier. However, there are two specific aspects of the pattern recognizer which warrant examination. As noted above, the largest context effect was a 500–600-Hz upward shift in F_2 for /u/ in the environment of initial-position alveolars. The obvious question is whether the pattern recognizer, trained on measurements from all phonetic contexts, would tend to misclassify /u/ in initial-position alveolar contexts. Using the best parameter set from those shown in Fig. 13 (duration, F_0, and two samples of F_1-F_3), the recognition rate for /u/ in the environment of initial-position alveolars was 98.3%, very similar to the 97.1% recognition rate for /u/ averaged across all contexts. The second-largest context effect was an upward shift of about 200 Hz in F_2 for /ʊ/ in the environment of initial-position alveolars. The recognition rate for /ʊ/ in initial-position alveolar environments was a respectable 85.1%, but lower than the recognition rate of 93.6% for /ʊ/ averaged across all contexts.

D. Correspondence between listener identification and discriminant analysis

As noted earlier, there is a disconnect of sorts between the listener data on the one hand and the formant measurements on the other. For example, the large number of statistically reliable effects of phonetic context on the formant frequencies contrasts sharply with the near uniformity in labeling accuracy across phonetic context. Additionally, the few reliable context effects that were observed in perception do not correspond with the largest of the effects found in the production data. At first glance, these findings would seem to discourage any consideration of a simple pattern recognition model that might account for the labeling behavior of listeners. However, as we have argued elsewhere (Hillenbrand and Nearey, 1999), it may not be adequate to compare variation in raw acoustic patterns to listeners' perception. The main reason for this is that even in the simplest model of categorization we can imagine, namely one based on minimum absolute distance of a token to a set of prototypes, more than the absolute location of a vowel token in pattern space must be considered. Specifically, the relative similarity to other category prototypes must also play a role. As noted in the example above, consider that tokens of /u/ following alveolars typically have much higher F_2 values than the population average of /u/. If those tokens have very low first formants, then they show second formants that are still substantially lower than /i/, the prototype of its nearest competitor category. Thus, it may still strike listeners as clearly more /u/-like than any other vowel, and hence the relatively large acoustic variation may produce little degradation of correct identification.

Discriminant analysis of the type reported above takes such factors of relative similarity directly into account and hence may provide perspective on the degree of correspon-

dence between classification by listeners and expected classification of the tokens based on the overall statistical properties of the distributions associated with each vowel. Here, we provide a brief analysis comparing aspects of discriminant analysis and listeners' categorization following methods described in detail in Hillenbrand and Nearey (1999). A quadratic discriminant analysis (QDA) was again performed using all 3390 tokens for which all of the following measurements were available: duration, F_0, and F_1-F_3 at 20% of the vowel duration and F_1-F_3 at 70% of the duration. When all the available tokens were used for both training and classification, 94.1% of the tokens were correctly classified. We define the modal response category of the panel of listeners as the category chosen by a plurality of listeners for each token. Under this hard-classification criterion, the panel showed a ''modally correct'' identification score of 98.1%, meaning that for all but about 2% of the tokens the label provided by a plurality of the panel agreed with the vowel intended by the talker. This is somewhat better than the QDA rate of 94.1%. There is substantial agreement between the QDA and the panel at the level of the individual token: 94.6% of tokens correctly identified by the panel were also correctly identified by QDA. There was also a reasonable level of agreement on the misidentified tokens: of the 64 tokens that were ''modally misidentified'' by the panel, the QDA chose the same incorrect category in 20 cases. Overall, the percent modal agreement between the QDA and the panel (i.e., where the panel of listeners agreed with the QDA on the best category, whether correct or incorrect) was 93.4%.

Following Hillenbrand and Nearey (1999), we also performed a *correct-response correlation analysis*, whereby the proportion of listeners' correct responses to each token was compared to the predicted *a posteriori* probability for the correct category from the QDA. The correct-response correlation r_c is defined as the Pearson correlation between the observed and predicted scores. The value of r_c will approach a maximum of 1.0 when variation in the relative probabilities of correct identification by listeners is matched by covariation in the predicted probabilities on a token-by-token basis. A modest but highly significant ($p<0.001$ by a randomization test) correlation of 0.28 was observed. This is generally similar to the value observed for the most similar model used in Hillenbrand and Nearey (1999, see Table VIII, model A). As in the case of the /hVd/ stimuli studied in Hillenbrand and Nearey, we have not yet matched listeners' performance in every respect. Nonetheless, the relatively simple discriminant analysis model adopted above seems to account well for the generally high identification rates by listeners. Similarly, it seems reasonable to suggest that lack of obvious correspondence between listeners' identification patterns and the magnitude of specific effects of context on acoustic properties is due in large measure to the relatively large degree of statistical separation among vowel classes. The reasonable success of the QDA demonstrates that the acoustic distinctiveness of the vowels is largely preserved despite variations in context.

IV. DISCUSSION

To summarize briefly, the primary purpose of this study was to evaluate the contribution of formant-frequency movements to the separability of vowel categories for CVC utterances involving variation in both the initial and final consonants; i.e., for utterances in which formant movements are influenced both by vowel identity and by coarticulatory phenomena. Recordings were made of six men and six women producing eight vowels (/i,ɪ,ɛ,æ,ɑ,ʌ,ʊ,u/) in isolation and in CVC syllables comprising all combinations of seven initial consonants (/h,b,d,g,p,t,k/) and six final consonants (/b,d,g,p,t,k/). A listening test showed high identification rates for the test utterances. More to the point, the effects of consonant environment on vowel intelligibility, while statistically significant in some cases, were quite small in magnitude, with only a few percent separating the most intelligible contexts from the least intelligible. In contrast to this perceptual stability, acoustic analysis showed a number of effects of consonant environment on vowel formant frequencies. The most important of these effects, which were generally consistent with SH, included: (1) a general tendency toward centralization for vowels produced in non-null environments; (2) large upward shifts in F_2 of 500–600 Hz for /u/ and 200–300 Hz for /ʊ/ in initial-position alveolar environments; (3) an upward shift of about 100 Hz in F_2 for /ɑ/ and /ʌ/ in initial-position alveolar environments; (4) an upward shift of about 100 Hz in F_2 for back vowels in initial-position velar environments; (5) a downward shift of about 85–100 Hz in F_2 for front vowels in initial-position labial environments; and (6) a tendency toward somewhat lower F_1 values for vowels in the environment of voiced consonants.

The central question that we sought to address was whether coarticulatory effects such as these would have the effect of obscuring the relationships between formant movement patterns and vowel identity that had been observed in several previous studies of isolated vowels or vowels in /hVd/ environments. Evidence from the pattern recognition studies reported in Sec. III C above is reasonably clear. Pattern recognition models that were trained on formant trajectories separated vowels with consistently greater accuracy than otherwise comparable models that were trained on static formant patterns. The improvement in overall classification accuracy with the addition of formant movement information was modest, averaging about 6 percentage points, but quite consistent across several combinations of parameters. Particularly large improvements averaging some 13–15 percentage points were seen for /ɪ/, /ɛ/, and /æ/.

As we have discussed elsewhere (e.g., Nearey, 1992; Hillenbrand and Nearey, 1999) pattern recognition evidence is not conclusive by itself for the simple reason that showing that a given feature improves category separability does not by itself prove that listeners make use of that feature in perception. There are, in fact, some clear examples in the literature of statistically based pattern classifiers greatly overestimating the perceptual importance of acoustic features (e.g., Hillenbrand and Gayvert, 1993b). As noted in the Introduction, in Hillenbrand and Nearey (1999) we were able to show that the pattern recognition evidence implicating an important role for spectral change in the classification of /hVd/

syllables was not, in fact, misleading. This was done by demonstrating that /hVd/ signals that had been resynthesized with flattened vowel formants were considerably less intelligible than otherwise comparable signals with formant movements matching the original utterances. In the absence of comparable perceptual evidence for the more complex CVC utterances studied here, we are forced to rely on a preliminary examination of the level of agreement between the listener data and the classification provided by the QDA. As discussed in Sec. III D, there are many—though by no means all—aspects of listener behavior that can be accounted for by a rather simple model incorporating $F0$, duration, and a very simple coding of spectral change consisting of two discrete samples of the formant pattern.

At first glance it might have been anticipated that the rather simple model of vowel classification that is embodied by the discriminant classifier would have been inherently incapable of accounting for the labeling behavior of the listener. As noted earlier, there are striking differences between the consonant-context effects that were observed in perception and those that were observed in the acoustic data. In our view, one of the most significant aspects of the listener data is the near uniformity in vowel intelligibility across different consonant environments. While there were a few statistically reliable effects of context on vowel intelligibility, these effects were small in absolute magnitude. This stands in contrast to the rather large number of reliable effects in production. Further, the largest of the effects that was observed in perception did not correspond in any obvious way to the largest context effects that were observed in perception. The clearest case of the apparently complex relationship between the production and perception data was the 500–600-Hz upward shift in F_2 values for /u/ following alveolar stops, which stands in contrast to the high intelligibility of postalveolar /u/, which was essentially indistinguishable from the intelligibility of /u/ in other environments. Taken together, these findings might well be seen as clear evidence favoring SH's conclusion that listeners internalize knowledge about the effects of context on vowel formants and invoke this knowledge in perception. While this remains a plausible account of the listener data, we do not believe that there is yet compelling evidence that a knowledge-based mechanism such as this is required. It should be kept in mind, for example, that the great majority of the statistically reliable context effects on the formant values are rather small in absolute terms. For example, the most consistent effect of context on F_1 values was a tendency toward lower frequencies in the context of voiced stops, but there was only one vowel (/ʌ/) for which the shift reached even a modest 75 Hz, and for the remaining vowels the shift averaged an auditorily undetectable 15–20 Hz. Further, of the rather large number of statistically reliable effects of context on F_2 values, the great majority averaged 100 Hz or less. There is, of course, the special case of the 500–600-Hz upward shift in F_2 for postalveolar /u/, but even for this very large context effect it is not obvious that a simple pattern recognition approach is incapable of accounting for the classification behavior of the listeners. As noted earlier, this shift has the effect of moving postalveolar /u/ upward into one of the few largely unoccupied regions of the crowded English vowel space where it remains unlikely to collide with other vowels. Our simple pattern classifier, in fact, recognized postalveolar /u/ with slightly greater accuracy than the recognition rate for /u/ averaged across all contexts.

Having made these arguments, we do not mean to imply that the variation in the acoustic properties of vowels induced by consonantal context is devoid of perceptual consequences. Specifically, we do not deny that such factors as (compensation for) consonant–vowel coarticulation (Strange, 1989; Nearey, 1989) or mechanisms of auditory contrast (e.g., Lotto and Kluender, 1998; Holt, Lotto, and Kluender, 2000) may play a role in listeners' identification of vowels in these stimuli. However, the preceding analyses indicate that even a simple two-target model, similar to that of Nearey and Assmann (1986), is adequate to account for a great deal of listeners' behavior. In future work, we plan to try to identify any systematic deviations from this baseline model and to study the relation of any such deviations to a wide range of hypotheses from the literature.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institutes of Health (No. 2-R01-DC01661) to Western Michigan University and by a grant from the Canadian Social Sciences and Humanities Research Council (No. 410-93-0053) to the University of Alberta.

APPENDIX: SIDAK-CORRECTED MULTIPLE COMPARISON TESTS OF EFFECT COEFFICIENTS

The formula for Sidak correction is $p_{\text{adj}} = 1 - (1 - p_{\text{nom}})^k$, where p_{adj} is the Sidak-adjusted probability value, p_{nom} is the nominal alpha level of a single t test, and k is the number of tests in a family. The value is always less than (and hence, significance tests are more powerful than), but often very close to, the simpler Bonferroni adjusted level p_{nom}/k. Please see Tables AI–AVI.

TABLE AI. ANOVA table for F_1. TPV is percent of total variance due to the effect in question compared to that of all phonetic factors. Only significant effects accounting for at least 0.25% of TPV are shown.

Effect	TPV	MS	df_{effect}	df_{error}	F	eps_GG1	p_GG
$\text{Voicing}_{\text{init}}$	0.52%	2.5058	1	11	22.71	1.000 000	0.000 584
Vowel	97.6%	66.849	7	77	266.35	0.000 000	0.000 000
$\text{Voicing}_{\text{fin}}$	0.31%	1.4872	1	11	51.58	1.000 000	0.000 018
$\text{Voicing}_{\text{init}}$ ×vowel	0.29%	0.1957	7	77	6.28	0.000 240	0.001 510

[1]Some vowels, especially tokens of /æ/, sometimes showed a centralized offglide with a formant pattern that was as steady as that in the vowel nucleus. The 60% criterion was intended to reduce the likelihood of identifying these offglides as steady-state points.

[2]More specifically, the procedure was equivalent to the following: For the ith subject (whether talker or listener), a linear model of the following form was estimated: $Y_{iwpvzq}=M_i+W_{iw}+P_{ip}+V_{iv}+Z_{iz}+Q_{iq}+WP_{iwp}+WV_{iwv}\cdots+QW_{iwq}+WPV_{iwpv}+\cdots+ZQ_{izq}+WPVZ_{iwpvz}+\cdots PVZQ_{ipvzq}+WPVZQ_{iwpvzq}$. Here, w is the index for initial voicing, p for initial place, v for vowel, z for final voicing, and q for final place. The single corresponding capital letters indicate main-effect coefficient estimates for the corresponding terms. Pairs and n-tuples indicate two- and n-way interaction effects, respectively, while the ellipses indicate the presence of additional n-way interaction terms. The t values for specific effects are calculated by taking the mean of the specific term in the formula as the numerator and the standard error of estimate (standard deviation divided by the number of subjects). For example, tests of the vowel by initial place effect would be based on $t=X_{pv}/S_{pv}$ where $X_{pv}=n^{-1}{}_{\square i}(PV_{ipv})$ and $S_{pv}=\{n^{-1}{}_{\square i}[PV_{ipv}-X_{pv}]^2\}^{1/2}$. This approach corresponds to one in which the ANOVA was run via a regression model using effects-coding repeated measures regression (Myers and Well, 1991). As recommended by Myers and Well, a unique error term is used for each contrast rather than using a pooled estimate. In the case of perceptual data, with random listener and random talker effects, a $t'_{\min}$ statistic was calculated from the t_L (by listeners) and t_T (by talkers) analysis. The magnitude of effects is always reported as a percentage above or below expected average effects for the family (main effect or interaction) in question. This was done using an approximate inverse of the Studebaker (1985) arcsine transformation.

[3]In examining Fig. 5, Ken Stevens, who provided a critique of this manuscript, noted the relative scarcity of tokens with F_2 values of about 1500 Hz for men and about 1700 Hz for women. Also noted for back vowels was a paucity of tokens with F_1 values of about 530 Hz for men and 600 Hz for women. Since these values correspond roughly with the lowest resonances of the subglottal system (Stevens, 1999), Stevens speculated that speakers may stay clear of these regions of formant space to avoid the alignment of supraglottal and subglottal resonances which would have the effect of shifting formant values away from the values for the uncoupled upper airway.

[4]The greater upward shift in $F2$ in initial alveolar environments relative to SH is difficult to interpret unambiguously since the SH data included results pooled from dental and prepalatal consonants in addition to the alveolar consonants used in the present study.

[5]Despite careful efforts to obtain fully balanced data sets, a small proportion of the target recordings later proved unsatisfactory, leading to a small number of missing values (23 of the original 4128 utterances were omitted due to pronunciation errors that were not noticed during the recording session). The problem of missing values in experiments with random factors is a complex one, with no single widely accepted solution. While we investigated the possibility of using more sophisticated measures, we could find none that would work with problems of the size of the data at hand. We opted for an extension of a method recommended by Myers and Well (1991). This is an iterative procedure which, on the first pass, substitutes the grand mean of the complete case observations for each missing value. On subsequent passes, the estimated value of the missing cases from the

TABLE AII. Main effect and interaction contrasts for F_1 significantly different from zero (by Sidak adjusted t test, $p_$adj) for each effect in Table AI.

Effect	Level	Size (%)	t	df	p	$p_$adj
$\text{Voicing}_{\text{init}}$	$[+\text{voice}]_{\text{init}}$	−2.7	−4.8	11	0.000 584	0.001 169
	$[-\text{voice}]_{\text{init}}$	2.7	4.8	11	0.000 584	0.001 169
Vowel	/i/	−43.7	−28	11	0.000 000	0.000 000
	/u/	−34.8	−21	11	0.000 000	0.000 000
	/ɪ/	−18.6	−11	11	0.000 000	0.000 002
	/ʊ/	−9.7	−4.5	11	0.000 847	0.006 758
	/ɛ/	19.9	7.8	11	0.000 008	0.000 064
	/æ/	25.1	7.4	11	0.000 013	0.000 107
	/ʌ/	40	19	11	0.000 000	0.000 000
	/ɑ/	76.3	23	11	0.000 000	0.000 000
$\text{Voicing}_{\text{fin}}$	$[+\text{voice}]_{\text{fin}}$	−2.1	−7.2	11	0.000 018	0.000 036
	$[-\text{voice}]_{\text{fin}}$	2.1	7.2	11	0.000 018	0.000 036

TABLE AIII. ANOVA table for F_2. TPV is percent of total variance due to the effect in question compared to that of all phonetic factors. Only significant effects accounting for at least 0.25% of TPV are shown.

Effect	TPV	MS	df	dfe	F	eps_GG1	p_GG
$\text{Place}_{\text{init}}$	2.69	3.3331	2	22	147.95	0.000 000	0.000 000
Vowel	91.2	32.2500	7	77	174.83	0.000 000	0.000 000
$\text{Place}_{\text{init}}\times\text{vowel}$	4.84	0.8558	14	154	63.01	0.000 000	0.000 000

TABLE AIV. Main effect and interaction contrasts for F_2 significantly different from zero (by Sidak adjusted t test, $p_$adj) for each effect in Table AIII.

Effect	Level	Size (%)	t	df	p	$p_$adj
$\text{Place}_{\text{init}}$	$[\text{lab}]_{\text{init}}$	−5.8	−16	11	0.000 000	0.000 000
$\text{Place}_{\text{init}}$	$[\text{vel}]_{\text{init}}$	1.4	4.1	11	0.001 804	0.005 402
$\text{Place}_{\text{init}}$	$[\text{alv}]_{\text{init}}$	4.7	12	11	0.000 000	0.000 000
Vowel	/u/	−26.9	−9.5	11	0.000 001	0.000 010
Vowel	/ʊ/	−24.1	−14	11	0.000 000	0.000 000
Vowel	/ʌ/	−19.5	−15	11	0.000 000	0.000 000
Vowel	/ɑ/	−15.4	−11	11	0.000 000	0.000 003
Vowel	/ɛ/	14.6	9.6	11	0.000 001	0.000 009
Vowel	/æ/	24.2	11	11	0.000 000	0.000 003
Vowel	/ɪ/	24.7	17	11	0.000 000	0.000 000
Vowel	/i/	49.0	23	11	0.000 000	0.000 000
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{lab}]_{\text{init}}\times$/u/	−9.7	−8.3	11	0.000 004	0.000 107
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{vel}]_{\text{init}}\times$/u/	−8.8	−9	11	0.000 002	0.000 050
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{alv}]_{\text{init}}\times$/ɪ/	−5.8	−12	11	0.000 000	0.000 003
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{alv}]_{\text{init}}\times$/ɛ/	−5.4	−10	11	0.000 000	0.000 012
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{alv}]_{\text{init}}\times$/i/	−5	−12	11	0.000 000	0.000 002
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{alv}]_{\text{init}}\times$/æ/	−4.5	−8.1	11	0.000 006	0.000 147
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{alv}]_{\text{init}}\times$/ɑ/	−2.8	−6.6	11	0.000 040	0.000 967
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{vel}]_{\text{init}}\times$/ʌ/	1.6	4.7	11	0.000 687	0.016 366
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{vel}]_{\text{init}}\times$/ɪ/	3	9.4	11	0.000 001	0.000 031
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{lab}]_{\text{init}}\times$/ɪ/	3.1	8.3	11	0.000 005	0.000 112
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{lab}]_{\text{init}}\times$/æ/	3.7	7.5	11	0.000 011	0.000 274
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{lab}]_{\text{init}}\times$/ɛ/	3.9	13	11	0.000 000	0.000 001
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{lab}]_{\text{init}}\times$/i/	5.2	9.7	11	0.000 001	0.000 025
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{alv}]_{\text{init}}\times$/ʊ/	5.8	5.6	11	0.000 152	0.003 646
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{alv}]_{\text{init}}\times[\text{u}]$	21.4	12	11	0.000 000	0.000 003

TABLE AV. ANOVA table for F_3. TPV is percent of total variance due to the effect in question compared to that of all phonetic factors. Only significant effects accounting for at least 0.25% of TPV are shown.

Effect	TPV	MS	df	dfe	F	eps_GG1	p_GG
$\text{Place}_{\text{init}}$	0.89	0.070 345	2	22	5.80	0.009 465	0.010 973
Vowel	92.7	2.082 7	7	77	62.84	0.000 000	0.000 000
$\text{Voicing}_{\text{fin}}$	0.76	0.120 22	1	11	35.66	1.000 000	0.000 093
$\text{Place}_{\text{fin}}$	0.31	0.024 569	2	22	13.30	0.000 382	0.000 712
$\text{Voicing}_{\text{init}}\times\text{vowel}$	0.45	0.010 091	7	77	3.04	0.011 446	0.030 239
$\text{Place}_{\text{init}}\times\text{vowel}$	2.7	0.030 276	14	154	7.35	0.000 000	0.000 033

TABLE AVI. Main effect and interaction contrasts for F_3 significantly different from zero (by Sidak-adjusted t test, $p_$adj) for each effect in Table AV.

Effect	Level	Size (%)	t	df	p	$p_$adj
$\text{Place}_{\text{init}}$	$[\text{vel}]_{\text{init}}$	−0.9	−3.7	11	0.003 627	0.010 840
Vowel	/u/	−8.5	−9.3	11	0.000 001	0.000 012
Vowel	/ʊ/	−5.6	−7.9	11	0.000 008	0.000 061
Vowel	/ɪ/	2.1	4.8	11	0.000 578	0.004 612
Vowel	/i/	15.8	12	11	0.000 000	0.000 001
$\text{Voicing}_{\text{fin}}$	$[-\text{voice}]_{\text{fin}}$	−0.6	−6	11	0.000 093	0.000 186
$\text{Voicing}_{\text{fin}}$	$[+\text{voice}]_{\text{fin}}$	0.6	6	11	0.000 093	0.000 186
$\text{Place}_{\text{fin}}$	$[\text{vel}]_{\text{fin}}$	−0.4	−3.1	11	0.009 882	0.029 354
$\text{Place}_{\text{fin}}$	$[\text{alv}]_{\text{fin}}$	0.5	5.4	11	0.000 210	0.000 629
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{alv}]_{\text{init}}\times$/i/	−2.1	−5.1	11	0.000 365	0.008 734
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{lab}]_{\text{init}}\times$/ɪ/	−1.3	−4.4	11	0.001 019	0.024 169
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{lab}]_{\text{init}}\times$/i/	−0.9	−4.3	11	0.001 184	0.028 022
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{lab}]_{\text{init}}\times$/ʌ/	1.1	4.3	11	0.001 186	0.028 072
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{lab}]_{\text{init}}\times$/ɑ/	1.4	4.9	11	0.000 494	0.011 783
$\text{Place}_{\text{init}}\times\text{vowel}$	$[\text{vel}]_{\text{init}}\times$/ɪ/	3.2	7.8	11	0.000 008	0.000 193

previous iteration is substituted. The estimates in question are based on all main effects and interactions except the highest-order interaction. This process is repeated until there is negligible change in any estimate on subsequent iteration. Statistics are based on standard methodology for fully balanced design except that degrees of freedom for error terms in ANOVA are based on the number of nonmissing cells associated with those terms. It is those adjusted degrees of freedom that are reported throughout the text and this Appendix.

[6]The absence of an effect for final voicing may have some relevance to SH's suggestion that the effects of consonant environment on vowel formants can be attributed primarily to articulatory undershoot. Since vowels preceding unvoiced stops are shorter on average than those preceding voiced stops (see Table I), one would presumably expect that the purely inertial effects associated with articulatory undershoot would be greater for the shorter-duration vowels preceding unvoiced stops. The absence of a final voicing effect would seem to call this interpretation into question.

[7]Pilot testing showed that the performance of the classifier was not strongly affected by different sampling points, making the choice of 20%–70% somewhat arbitrary. The 20%–70% sampling points performed slightly better than the 20%–80% points used in our earlier /hVd/ study.

[8]An anonymous reviewer suggested that the two-sample pattern classifier may have outperformed the one-sample classifier simply by reducing sampling error; i.e., by improving the odds that a representative sample was obtained of the effectively steady-state formant pattern. As a quick test of this possibility, we trained the pattern classifier on the same sets of parameters that were used in the tests that are summarized in Fig. 13, but tested the recognition model on parameter sets in which the ordering of the 20% and 70% samples was reversed. We reasoned that if the issue is simply sampling error, then the ordering of the samples will be unimportant. However, if it is the formant trajectory that is being captured by the two-sample method, as we argue, then the ordering of the samples will make a great deal of difference. The results showed that the ordering of the samples matters. Averaged across all parameter sets, classification accuracy was 10.6 percentage points higher for natural order than reverse order. Classification accuracy for the two-sample reverse-order features was also 4.3 percentage points lower on average than for a single sample of the formant pattern. See Nearey and Assmann (1986) for comparable tests with human listeners.

Andruski, J. E., and Nearey, T. M. (**1992**). ''On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables,'' J. Acoust. Soc. Am. **91**, 390–410.

Assmann, P., Nearey, T., and Hogan, J. (**1982**). ''Vowel identification: Orthographic, perceptual, and acoustic aspects,'' J. Acoust. Soc. Am. **71**, 975–989.

Black, J. W. (**1949**). ''Natural frequency, duration, and intensity of vowels in reading,'' J. Speech Hear. Disord. **14**, 216–221.

Clark, H. (**1973**). ''The language-as-fixed-effect fallacy,'' J. Verbal Learn. Verbal Behav. **12**, 335–339.

Crystal, T. H., and House, A. S. (**1988**). ''Segmental durations in connected-speech signals: Current results,'' J. Acoust. Soc. Am. **83**, 1553–1573.

Fairbanks, G., and Grubb, P. (**1961**). ''A psychophysical investigation of vowel formants,'' J. Speech Hear. Res. **4**, 203–219.

Fischer-Jorgensen, E. (**1964**). ''Sound duration and place of articulation,'' Zeitschrift Sprachwisse Phonetik **17**, 175–207.

Hillenbrand, J. (**1988**). ''MPITCH: An autocorrelation fundamental-frequency tracker,'' [Computer Program], Western Michigan University, Kalamazoo, MI.

Hillenbrand, J. M., and Gayvert, R. T. (**1993a**). ''Identification of steady-state vowels synthesized from the Peterson and Barney measurements,'' J. Acoust. Soc. Am. **94**, 668–674.

Hillenbrand, J. M., and Gayvert, R. T. (**1993b**). ''Vowel classification based on fundamental frequency and formant frequencies,'' J. Speech Hear. Res. **36**, 694–700.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Hillenbrand, J. M., and Nearey, T. N. (**1999**). ''Identification of resynthesized /hVd/ syllables: Effects of formant contour,'' J. Acoust. Soc. Am. **105**, 3509–3523.

Holt, L. L., Lotto, A. J., and Kluender, K. L. (**2000**). ''Neighboring spectral content influences vowel identification,'' J. Acoust. Soc. Am. **108**, 710–722.

Jenkins, J. J., and Strange, W. (**1999**). ''Perception of dynamic information for vowels in syllable onsets and offsets,'' Percept. Psychophys. **61**, 1200–1210.

Jenkins, J. J., Strange, W., and Edman, T. R. (**1983**). ''Identification of vowels in 'vowelless' syllables,'' Percept. Psychophys. **34**, 441–450.

Johnson, R. A., and Winchern, D. W. (**1982**). *Applied Multivariate Statistical Analysis* (Prentice-Hall, Englewood Cliffs, NJ).

Lotto, A. J., and Kluender, K. R. (**1998**). ''General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification,'' Percept. Psychophys. **60**, 602–619.

Miller, J. D. (**1989**). ''Auditory-perceptual interpretation of the vowel,'' J. Acoust. Soc. Am. **85**, 2114–2134.

Myers, J. L., and Well, A. D. (**1991**). *Research Design and Statistical Analysis* (HarperCollins, New York).

Nearey, T. M. (**1989**). ''Static, dynamic, and relational properties in vowel perception,'' J. Acoust. Soc. Am. **85**, 2088–2113.

Nearey, T. M. (**1992**). ''Applications of generalized linear modeling to vowel data,'' in *Proceedings of ICSLP 92*, edited by J. Ohala, T. Nearey, B. Derwing, M. Hodge, and G. Wiebe (University of Alberta, Edmonton, Alberta, Canada), pp. 583–586.

Nearey, T. M., and Assmann, P. (**1986**). ''Modeling the role of vowel inherent spectral change in vowel identification,'' J. Acoust. Soc. Am. **80**, 1297–1308.

Parker, E. M., and Diehl, R. L. (**1984**). ''Identifying vowels in CVC syllables: Effects of inserting silence and noise,'' Percept. Psychophys. **36**, 369–380.

Peterson, G. E., and Lehiste, I. (**1960**). ''Duration of syllable nuclei in English,'' J. Acoust. Soc. Am. **32**, 693–703.

Stevens, K. N. (**1999**). *Acoustic Phonetics* (The MIT Press, Cambridge, MA).

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.

Strange, W. (**1989**). ''Dynamic specification of coarticulated vowels spoken in sentence context,'' J. Acoust. Soc. Am. **85**, 2135–2153.

Strange, W., Jenkins, J. J., and Johnson, T. L. (**1983**). ''Dynamic specification of coarticulated vowels,'' J. Acoust. Soc. Am. **74**, 695–705.

Studebaker, G. A. (**1985**). ''A 'rationalized' arcsine transform,'' J. Speech Hear. Res. **28**, 455–462.

Zahorian, S., and Jagharghi, A. (**1993**). ''Spectral shape features versus formants as acoustic correlates for vowels,'' J. Acoust. Soc. Am. **94**, 1966–1982.