Amplitude variations in coarticulated vowels

Ewa Jacewicz[a] and Robert Allen Fox
Department of Speech and Hearing Science, The Ohio State University, Columbus, Ohio 43210-1002

(Received 19 February 2007; revised 13 February 2008; accepted 19 February 2008)

This paper seeks to characterize the nature, size, and range of acoustic amplitude variation in naturally produced coarticulated vowels in order to determine its potential contribution and relevance to vowel perception. The study is a partial replication and extension of the pioneering work by House and Fairbanks [J. Acoust. Soc. Am. **22**, 105–113 (1953)], who reported large variation in vowel amplitude as a function of consonantal context. Eight American English vowels spoken by men and women were recorded in ten symmetrical CVC consonantal contexts. Acoustic amplitude measures included overall rms amplitude, amplitude of the rms peak along with its relative location in the CVC-word, and the amplitudes of individual formants F1–F4 along with their frequencies. House and Fairbanks' amplitude results were not replicated: Neither the overall rms nor the rms peak varied appreciably as a function of consonantal context. However, consonantal context was shown to affect significantly and systematically the amplitudes of individual formants at the vowel nucleus. These effects persisted in the auditory representation of the vowel signal. Auditory spectra showed that the pattern of spectral amplitude variation as a function of contextual effects may still be encoded and represented at early stages of processing by the peripheral auditory system. [DOI: 10.1121/1.2897034]

PACS number(s): 43.70.Fq, 43.70.Bk [AL] Pages: 2750–2768

I. INTRODUCTION

This paper examines acoustic amplitude variation in naturally produced vowels as a function of vowel quality and its immediate consonantal context. Usually, formant frequencies are considered to be the primary factors determining vowel identity. However, although "secondary" acoustic cues such as duration, fundamental frequency (f_0), and amplitude have also been shown to contribute significantly to the phonetic identity of vowels, particularly little is known about the role of vowel amplitude. There is an extensive body of work supporting the linguistic use of vowel length distinctions and the perceptual importance of duration in vowel identification (e.g., Ainsworth, 1972, 1981; Mermelstein, 1978; Gottfried and Beddor, 1988; Whalen, 1989). The significance of f_0 to vowel quality has also been established (e.g., Nearey, 1989; Whalen and Levitt, 1995; Katz and Assmann, 2001). Yet, over the years, neither vowel amplitude nor the amplitudes of individual formants have been studied as systematically as have the other two secondary cues. Understandably, the literature on the subject, whether pertaining to vowel production or perception, is sparse.

Perhaps the best known set of studies exploring the role of formant amplitude comes from early speech perception research. Given the possibility of independent control over the formant amplitudes of static vowels using the parallel synthesis method, a number of experiments demonstrated that manipulations of the amplitudes of particular formants affected the perceived quality of the vowel (Lindqvist and Pauli, 1968; Carlson *et al.*, 1970; Ainsworth and Millar, 1972; Aaltonen, 1985; Schwartz and Escudier, 1987). If formant amplitudes indeed contribute to the perception of vowel quality, the question arises whether the auditory system is also sensitive to possible variations in formant amplitudes arising from coarticulation effects. However, before this question is addressed, we need to have a much better understanding of the nature, size, and range of such variation in naturally produced speech, especially in terms of the effect of immediate context. That is, given that naturally produced vowels are rarely static and dynamic cues such as formant transitions are coded primarily in the frequency domain, what is the nature and acceptable size of formant amplitude variation that may covary with dynamic frequency information? A better understanding of the nature and permissible range of such variation can help define more precisely the appropriate values in creating synthetic versions of vowels. In turn, by providing better controls, we may find that the contribution of formant amplitude to vowel perception may involve percepts other than phonetic quality, most likely vowel naturalness and timbre.

It needs to be pointed out that over the years, the primary interest in studying vowel amplitude was in determining its role in signaling linguistic prominence distinctions such as stress and accent (e.g., Fry, 1955; Lehiste and Peterson, 1959; Beckman, 1986; Sluijter and Van Heuven, 1996; Fant *et al.*, 2000a, b). However, this type of variation is fundamentally different in that it involves speaker-intended modification of vocal intensity. With regard to the basic amplitude variation, there is essentially one study which showed that vowel amplitude also varied as a function of its immediate consonantal context: House and Fairbanks (1953).

House and Fairbanks (1953) reported measurements of relative power for six American English vowels /i, e, æ, ɑ, o, u/ along with their duration and f_0. These vowels were produced in symmetrical stressed CVC syllables with twelve consonants /p, t, k, b, d, g, f, v, s, z, m, n/ differing in

[a]Electronic mail: jacewicz.1@osu.edu

 0001-4966/2008/123(5)/2750/19/$23.00

voicing, manner, and place of articulation. Averaged across all consonantal contexts, the vowels fell in two groups: greater in relative power (/o, e, u/) and lesser in relative power (/i, æ, ɑ/). The differences within each group were not significant. A surprising result was that low vowels /æ, ɑ/ were the weakest in the entire set, contrary to what might have been expected given the greatest openness of both vowels. Another striking finding was that consonant voicing had a substantial effect on the vowel's amplitude: Vowels embedded in voiced consonants were more than twice as strong as vowels in the context of voiceless consonants. Considerable differences were also found as a function of consonant manner: Vowels in the context of fricatives were about 60% stronger than vowels in the context of stops. The effects of consonant place were small, showing that vowels in the contexts of velars were only slightly weaker that those in the contexts of either labials or alveolars.

These findings motivated the present study of the variation in formant amplitude in consonantal context. If specific contexts introduce such considerable changes to the overall vowel amplitude, we may expect these changes to be manifested also in the vowel spectrum. Furthermore, they may still be present at the vowel nucleus and not only at the vowel margins. There is a lack of empirical acoustic data which would relate the variability in the overall vowel amplitude as a function of consonantal context to the corresponding changes in amplitudes of particular formants. It is also unknown how the context affects formant amplitude of vowels differing in their articulation with respect to the tongue body position and degree of openness. We may only infer that context effects may be different for different vowel spectra, introducing local changes that might enhance a particular portion of the spectrum to which the perceptual system might be sensitive and may utilize it in processing and recognizing the vowel at the lexical level.

The contribution of formant amplitudes to the perception of vowel quality was recently examined in terms of manipulation of the global spectral tilt of synthetic vowels, based on the rules of formant amplitude in cascade filter model of the vocal tract (Ito *et al.*, 2001; Kiefte and Kluender, 2005). However, as demonstrated by Kiefte and Kluender (2005), spectral tilt seems to be more effective for steady-state vowels than for vowels with inherent spectral change. It becomes even more difficult to specify the synthesis values and to predict the perceptual efficacy of the global spectral tilt that may additionally vary as a function of the immediate phonetic context of the vowel.

In acoustic phonetic theory, it is generally accepted that the spectrum of a vowel is the spectrum of the glottal source filtered by the vocal tract and the vocal tract acts as a composite filter of several bandpass filters, one for each formant (e.g., Stevens, 1998; Titze, 2000). Although the vowel-specific spectral slope can be predicted from the complex interaction between glottal source and vocal tract filtering (e.g., Fant, 1956; Fant *et al.*, 1963; Fant and Lin, 1987), the size of the filter-related size variation in formant amplitude such as that found in vowels in context has not yet been integrated into the models. Yet, such spectral details may bolster speech production models, leading also to improved speech recognition schemes.

The present acoustic study is a partial replication and extension of House and Fairbanks (1953) to individual formants. The purpose of the replication is to verify the general patterns of amplitude variation in that study using current speech analysis tools. The investigation of the amplitude of individual formants, extending the original House and Fairbanks study, aims to relate the variability in the vowel spectrum to the changes in the overall vowel amplitude in consonantal context. The overall goal of this study is to characterize the nature, size, and range of acoustic amplitude variation in coarticulated vowels in order to determine its contribution and relevance to vowel perception.

II. METHODS

A. Speakers

Twenty speakers (ten men and ten women) of Midwestern American English with no known history of speech disorders participated in the present study—fourteen were born and raised in Ohio, four in Michigan, and two in Wisconsin. They were students aged 18–36 years enrolled in a variety of majors at The Ohio State University and were phonetically untrained.

B. Speech materials

Eight relatively monophthongal American English vowels were selected: /i, ɪ, ɛ, æ, ɑ, ɔ, ʊ, u/. Four of the vowels /i, æ, ɑ, u/ were among those studied by House and Fairbanks (HF), and four /ɪ, ɛ, ɔ, ʊ/ were new. Since the set in HF did not include short vowels, the vowels /ɪ, ʊ/ were added as the lax counterparts to the tense /i, u/ and the two diphthongal vowels /e, o/ were replaced by the relatively monophthongal vowels /ɛ, ɔ/. Consistent with HF, each vowel was embedded in a symmetrical C_1VC_1 context. The selected consonant set consisted of ten oral consonants /p, t, k, b, d, g, f, v, s, z/ as in HF. The two nasal consonants /m, n/ were not included in the present set. We decided to concentrate on oral vowels only because vowels in a symmetrical nasal context would have been produced with a significant degree of nasalization which would have introduced both nasal formants and nasal zeroes into the vowel spectra. In the present study, the CVCs were produced as monosyllables in a stressed position in a short phrase "It's a _____" [ɪtsə_____] and not as disyllables as in HF, in which each CVC was prefixed by unstressed [hə] (e.g., *hupeep*, *hudeed*, *hukeek*, etc.). This modification was introduced merely for practical reasons as the phonetically untrained subjects found it difficult to read a nonsense word with the above-mentioned prefix. However, in both HF and the present study the stressed target syllable was preceded by an unstressed schwa. The complete set of stimulus materials in the orthographic form presented to the speakers is listed in Table I. Each speaker recorded 240 utterances in a single session lasting approximately 1 h (8 vowels × 10 consonants × 3 repetitions). The data for /ɔ/ from three subjects and for /ɑ/ from three other subjects were subsequently discarded because it was determined during the

TABLE I. Orthographic representation of the recorded stimulus set.

Context	/i/ (heat)	/ɪ/ (hit)	/ɛ/ (bed)	/æ/ (bad)	/ɑ/ (lot)	/ɔ/ (law)	/ʊ/ (good)	/u/ (who)
[p]	peep	pip	pep	pap	pop	pawp	poup	poop
[t]	teet	tit	tet	tat	tot	tawt	tout	toot
[k]	keek	kik	kek	kak	kock	kawk	kouk	kook
[f]	feef	fif	fef	faf	fof	fawf	fouf	foof
[s]	sees	sis	ses	sas	sos	saws	sous	soos
[b]	beeb	bib	beb	bab	bob	bawb	boub	boob
[d]	deed	did	ded	dad	dodd	dawd	doud	dood
[g]	geeg	gig	geg	gag	gog	gawg	goug	goog
[v]	veev	viv	vev	vav	vov	vawv	vouv	voov
[z]	zeez	ziz	zez	zaz	zoz	zawz	zouz	zooz

analysis phase that these speakers had a complete merger of /ɔ/ and /ɑ/ (this was not true for the other speakers).

C. Procedure

Recordings were controlled by a program in MATLAB. The speaker was seated in a sound-attenuated IAC booth and was facing a LCD monitor placed outside the booth's window. A head-mounted Shure SM10A dynamic microphone was used, positioned at a distance of 2 in. from the speaker's lips. Speech samples were recorded and digitized at a 44.1 kHz sampling rate directly onto a hard disc drive. A model example word containing the vowel of interest was first displayed on the screen (e.g., "heat"). Next to the word, the stimulus sentence appeared (e.g, "It's a **beeb**"). The speaker read the sentence placing stress on the CVC word and producing the vowel quality as in the displayed word. To overcome any orthographic uncertainties, the presentation was blocked by vowel type. The order of consonantal contexts for each vowel type and the order of vowels themselves were randomized for each speaker. There was a short practice set completed before the start of each vowel category to assure that the speaker was comfortable with the spelling of the CVCs. After recording each sentence, the experimenter either accepted and saved the sentence or re-recorded it in the case of any mispronunciations. Speakers took short breaks upon request and the position of the microphone was monitored by the experimenter throughout the recording session to assure its constant distance from the speaker's lips.

D. Acoustic measurements

The set of measurements included overall rms amplitude, amplitude of the rms peak along with its location in the vowel, the amplitudes of individual formants F1–F4 along with their frequencies, vowel duration, and f_0. Prior to acoustic analysis, the tokens were digitally filtered and downsampled to 11.025 kHz. Measurements of both vowel duration and the duration of the whole CVC word served as input for subsequent automated measurements of rms peak location and the entire set of amplitude and frequency measurements. Standard measures of vowel duration were used (Peterson and Lehiste, 1960; Hillenbrand *et al.*, 1995). Vowel onsets and offsets were located by hand, primarily on the basis of a wave form display with segmentation decisions checked against a spectrogram. For vowels in the context of fricatives, vowel onset was measured from the cessation of noise in the periodic wave form following a frication offset and vowel offset was determined by the onset of noise in the wave form signaling the start of final frication. For vowels in the context of stops, vowel onset was measured from onset of periodicity (at a zero crossing) following the release burst. Vowel offset for voiceless stops was defined at the point at which the amplitude of the vowel dropped to near zero (which was also coincident with elimination of all periodicity in the wave form). The vowel offset for voiced stops was defined as that point when the amplitude dropped to near zero and any periodicity in the wave form contained no high frequency components (as expected for voicing produced during the stop closure). Because vowel amplitude was the primary focus, we eliminated those portions of the vowel near offset that were produced with creaky voice. This happened in about 20% of productions from three speakers who had a tendency to use irregular phonation at word offsets. All segmentation decisions were later checked and corrected (and then rechecked) by both experimenters using a MATLAB program that displayed the segmentation marks superimposed over a display of the CVC's wave form (in two different views—a view that included the entire CVC and an expanded view that concentrated on the vowel portion only).

Two measures of vowel amplitude were used in this study, the peak rms and the overall rms. The peak rms measure estimates the peak energy of the vowel, i.e., the highest amplitude of the vocalic portion or syllable nucleus. The overall rms (the quadratic mean) is a measure of vowel amplitude calculated from vowel onset to vowel offset. The choice of both measures instead of one was dictated by our desire to replicate the HF results choosing the amplitude values that would correspond most closely to the "maximum level for each syllable." HF measured the maximum level of each stressed syllable, defined as the point of maximum intensity reached during the production of the stressed vowel as indicated on an analog sound level meter and recorded graphically on paper. However, in reporting their result, HF used another measure called "relative power," which was a manipulation "to facilitate arithmetic treatment" (p. 110). Namely, each measurement for each subject was expressed in decibels above the lowest value for that subject and the difference N between the two values ($N_{\max} - N_{\min}$) was con-

verted to relative power for this subject which was equal to $\text{antilog}_{10} N/10$. Calculated in this way, each of these relative power values is a simple ratio value and there is no specific measure unit that can be assigned to them. In reporting their results as relative power, HF did not use any specific unit as they did in reporting their duration (in seconds) and f_0 (in Hz) measurements (see, for example, their Table 5). Their vowels were simply comparatively greater or smaller in relative power, which might be thought of as greater or smaller in magnitude, as all values were reported in relative power.

From the HF report, we do not know the exact decibel values for their "syllable maximum" measurements. Since our present focus was to relate the measures of overall vowel amplitude and the amplitude of individual formants, we measured both the peak rms and the overall rms amplitude to observe their relationship for individual vowels and consonantal contexts. The amplitudes of individual formants were then measured at the temporal location of the rms peak, which gives the best estimate of how spectral energy is distributed in the most "intense" region in the vowel. We expected higher decibel values for the peak rms measure as compared to the overall rms, but whether the effects of consonantal context would be the same for both measures could not be determined without empirical investigation. Based on how the analyses were done using the experimental apparatus available to HF, the overall rms measure may correspond more closely to HF measurements. Yet, examination of the overall rms measure alone would be problematic for the present focus which investigates changes to the amplitude spectrum in addition to the more "global" amplitude variation as a function of consonantal context. It was therefore decided to report the results for the overall rms in addition to the results for rms peak amplitude.

To find the location and size of the amplitude peak, the rms values of a series of overlapping 25 ms windows were calculated whose number varied with the duration of the vowel. The temporal location of the center of the 25 ms window with the highest rms value was defined as the location of the rms peak. The location of the rms peak was expressed as its relative position (with values ranging from 0 to 100%) with regard to the duration of the word.[1]

Formant frequency values and amplitudes of the first four formant peaks based on 14-pole LPC analysis were extracted automatically using a MATLAB program which determined the location of the rms peak. The program displayed these values along with the FFT and LPC spectrum and provided a wideband spectrogram of the entire vowel. No preemphasis was applied. In some cases, the formant frequency values obtained were compared to the formant peaks identified using smoothed FFT spectra and from the wideband spectrograms with formant tracks displayed (using the program TF32, Milenkovic, 2003). Errors in formant estimation in LPC analysis were then hand-corrected. The effects of consonantal context on formant amplitude were assessed by first examining the variation of A1, the strongest formant of each vowel. The decibel values for A1 (and for the higher formants A2, A3, and A4) were measured relative to the reference level of the speech analysis program (written in MATLAB), which is the lowest nonzero instantaneous amplitude that is coded by the A/D conversion. The differences between the amplitude of each higher formant and that of A1 were then analyzed. Because amplitude spectrum varies as a function of vowel category, the amplitude variation of the higher formants became more apparent when reported relative to A1. The (negative) decibel differences reflect the strength of the higher formant relative to A1. Smaller negative decibel difference values here denote a stronger higher formant. Conversely, greater negative decibel difference values imply its relative weakness.

The f_0 of the rms peak window was computed using cepstral analysis (Oppenheimer and Willsky, 1977) and hand-corrected when the method measured the frequency of a higher harmonic.

E. Statistical treatment of the amplitude results

In the HF study, the amplitude measures for each subject were expressed relative to the lowest value for that subject. A primary reason for this expression was to control for individual variation (in vocal characteristics, general amplitude levels, etc.). However, as HF note, the results of their analysis of variance (ANOVA) tests indicated that subject variation had not been "completely obliterated." The current study took the approach of using within-subject (also called repeated-measures) designs for all inferential statistical tests. The advantage of a within-subject design is that idiosyncratic variations among subjects (which is commonly found in phonetic experiments but is of no interest here) can be controlled for. In particular, in within-subject designs the differences among subjects are measured and separated from the error terms in the calculation of the test statistic (e.g., t and F values).

We used within-subject ANOVAs to assess the significance of amplitude variation. For each amplitude-related measure (i.e., location of rms peak, amplitude of rms peak, overall amplitude, and amplitudes of formants) an overall ANOVA was first completed with the within-subject factors vowel and consonantal context. Speaker gender was included as a between-subjects factor. To further explore significant consonant effects for each measure, separate repeated-measures ANOVAs were completed on individual vowels for the factors manner, voicing and place for the /b, p, d, t, v, f, z, s/ set, and for the factors place and voicing for the /b, p, d, t, g, k/ set. In each of these additional analyses, speaker gender was included as a between-subjects factor. For all reported significant main effects and interactions, the degrees of freedom for the F-tests were Greenhouse–Geisser adjusted when there were significant violations of sphericity. In addition to the significance values, a measure of the effect size—partial eta squared is reported (η^2, whose value can range from 0.0 to 1.0, should be considered a measure of the proportion of variance explained by a dependent variable when controlling for other factors). Post hoc analyses were completed using either additional ANOVAs on selected subsets of the data (with appropriate F tests) or Bonferroni-adjusted t-tests.

The results for vowel duration were assessed by within-subject ANOVAs and Bonferroni-adjusted t-tests. f_0 results

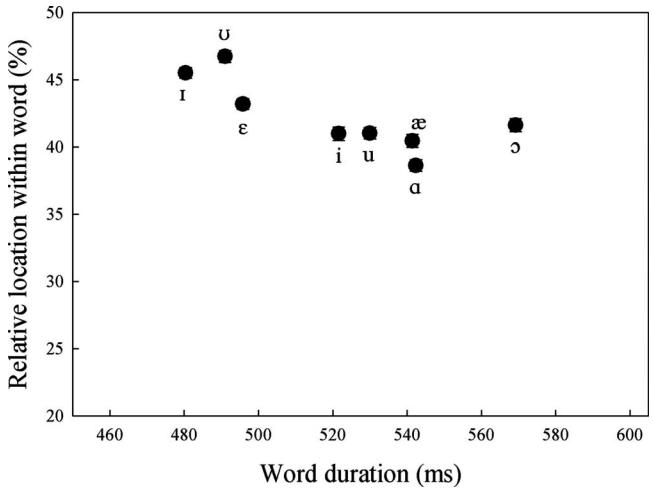

FIG. 1. Mean relative location of rms peak within a word as a function of each word's mean duration averaged across ten consonantal contexts.

were not assessed statistically because they are reported here merely to establish consistency with the HF data for all three secondary cues, including vowel amplitude, vowel duration, and f_0. At present, vowel duration and f_0 do not constitute a focus of the study.

III. RESULTS

A. Amplitude of rms peak and overall rms amplitude

The amplitude of rms peak was measured at its temporal location within the CVC word which varied for individual vowels. Despite the variation, the mean peak values for all vowels were still before vowel midpoint (Fig. 1). In general, the peaks of the short vowels /ɪ, ɛ, ʊ/ occurred later than the peaks of the remaining longer vowels.

The variation in the relative positions of rms peak within word as a function of consonantal context is summarized in Fig. 2. The peaks in the context of stops occurred later than in the context of fricatives and, statistically, the effect of manner was strong $[F(117)=306.16, p<0.001, \eta^2=0.947]$. The peaks were also significantly later in the context of

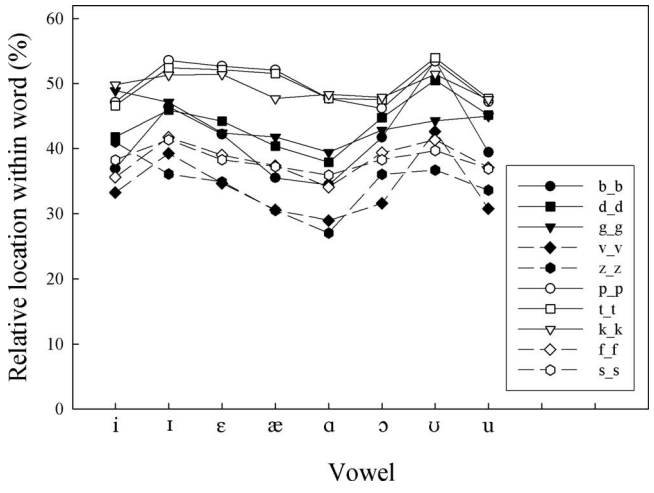

FIG. 2. Mean relative location of rms peak within a word shown for each vowel and for each individual consonantal context.

TABLE II. Mean vowel amplitude at rms peak and overall rms amplitude (in decibels) (s.d.) in consonantal contexts grouped by voicing, manner, and place of articulation. All vowels are pooled.

Grouped consonant environments	rms peak	Overall rms
Voicing		
Voiceless	−15.58 (3.89)	−18.26 (3.76)
Voiced	−15.45 (3.72)	−18.26 (3.62)
Manner of articulation		
Stop	−15.68 (3.79)	−18.28 (3.65)
Fricative	−15.27 (3.83)	−18.22 (3.74)
Place of articulation		
Labial	−15.01 (3.80)	−17.71 (3.69)
Alveolar	−15.79 (3.84)	−18.62 (3.72)
Velar	−15.97 (3.66)	−18.63 (3.50)

voiceless consonants than voiced $[F(1,17)=82.09, p<0.001, \eta^2=0.828]$. The effects of place were more variable and were mostly not significant.

Table II shows, following HF, means for amplitude of rms peak and overall rms amplitude grouped by consonant voicing, manner, and place across all vowels pooled. The mean amplitudes at rms peaks were consistently higher (2.74 dB) than the mean overall rms amplitudes. As is evident, however, neither amplitude values varied appreciably as a function of the consonantal features. This is in sharp contrast to HF results, particularly for voicing and manner.

Figure 3, drawn after HF, shows rms means for individual consonantal contexts (stops and fricatives) across all vowels pooled. Unlike the HF data, there is no sharp division

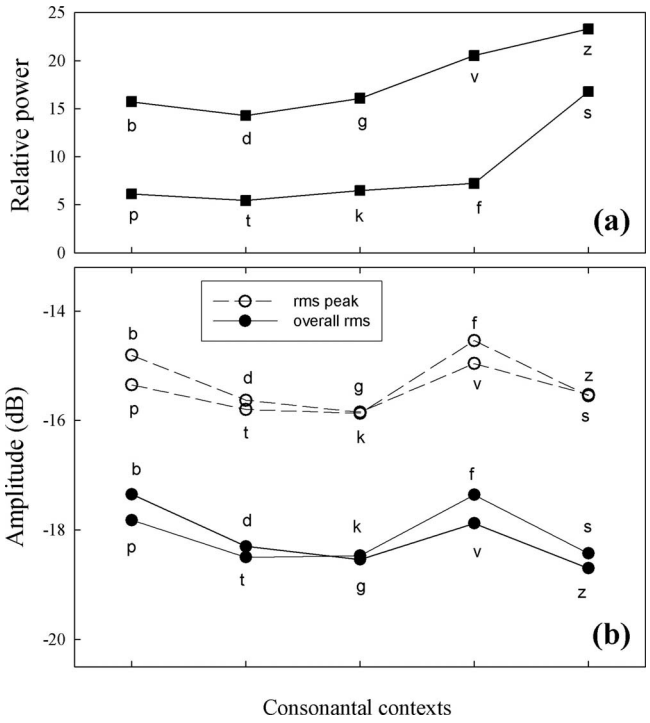

FIG. 3. Consonantal context effects on vowel amplitude pooled across all vowels. (a) Data from House and Fairbanks (1953) for their relative power measure (see the text). (b) Mean amplitudes (in decibels) of rms peak and overall rms in the present study.

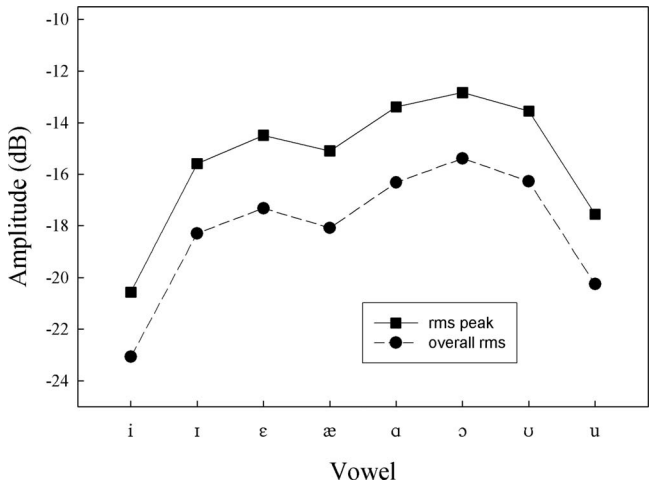

FIG. 4. Mean amplitudes of rms peak and overall rms for individual vowels averaged across all consonantal contexts.

between vowel amplitude in voiced and voiceless contexts. Although the general trends for all stops and the fricatives /v, f/ are consistent with HF, the results for the fricatives /z, s/ differ rather drastically. Unlike the HF results, the amplitudes of vowels in the environments of either /z/ or /s/ were not the greatest in the entire consonant set and did not differ appreciably from those in the context of alveolar stops. As shown in Table II, all amplitude values at rms peaks were about 3 dB higher than were overall rms amplitudes.

The amplitudes of individual vowels averaged across all consonantal contexts are shown in Fig. 4. Consistently with the trend already observed, amplitudes at rms peaks were about 3 dB higher than overall rms amplitudes. Back vowels had higher amplitude than front vowels. At rms peak, the overall mean for back vowels was −14.33 dB as opposed to −16.44 dB for front vowels. For overall rms, the mean values for back and front vowels were −17.06 and −19.19 dB, respectively. There was general regularity in the relationship between the degree of vowel openness and vowel amplitude: The two most closed vowels /i/ and /u/ had the lowest amplitudes while the most open back vowels /ɑ/ and /ɔ/ had the highest. However, the correspondence between greater openness of a vowel and its increased amplitude was less consistent for the remaining vowels. In particular, it was surprising that the amplitudes for both /ɪ/ and /ʊ/ were relatively high and comparable with those for /æ/ and /ɑ/, respectively. Paired two-tailed *t*-tests confirmed these observations, showing that there was no statistical difference in the amplitudes of the vowels /ɪ/ and /æ/ and the vowels /ʊ/ and /ɑ/. The difference was also not significant between /ʊ/ and /ɔ/. The differences for all remaining vowel pairs were significant for both rms peak and overall rms amplitude.

It may be recalled that the vowels /æ/ and /ɑ/ were drastically low in relative power in the HF study, approximating the relative power of the vowel /i/. Although various explanations for this unexpected outcome were offered, HF admitted that "a confident explanation of the atypical finding for

TABLE III. Summary of significant main effects from repeated measures ANOVAs for relative amplitude of rms peak and overall rms amplitude. Shown are partial eta squared values (η^2). (*) $p<0.050$, (**) $p<0.010$, (†) $p<0.001$; (---) not significant; fr=fricative, st=stop, vd=voiced, vl=voiceless, l=labial, alv=alveolar, vel=velar.

	/i/	/ɪ/	/ɛ/	/æ/	/ɑ/	/ɔ/	/ʊ/	/u/
rms peak								
/b, d, p, t, v, z, f, s/								
Manner	---	---	0.286* fr>st	0.452** fr>st	0.532** fr>st	---	---	---
Voicing	0.332** vd>vl	---	---	---	---	---	---	---
Place	0.414** l>alv	0.462** l>alv	0.503† l>alv	0.447** l>alv	0.438** l>alv	0.528** l>alv	0.311* l>alv	0.839† l>alv
/b, d, g, p, t, k/								
Voicing	0.337** vd>vl	---	---	---	---	---	---	---
Place	---	0.362† l>vel>alv	0.454† l>alv>vel	0.254** l>alv>vel	0.202* l>vel>alv	0.302** l>alv>vel	0.329** l>alv>vel	0.694† l>vel>alv
Overall rms								
/b, d, p, t, v, z, f, s/								
Manner	---	---	---	---	---	---	0.260* st>fr	0.201* st>fr
Voicing	---	---	---	---	---	0.375* vl>vd	---	---
Place	0.611† l>alv	0.699† l>alv	0.700† l>alv	0.565† l>alv	0.539** l>alv	0.689† l>alv	0.390** l>alv	0.850† l>alv
/b, d, g, p, t, k/								
Voicing	0.258* vl>vd	---	208* vd>vl	---	---	0.425** vl>vd	---	---
Place	---	0.512† l>alv>vel	0.610† l>alv>vel	0.388† l>alv>vel	---	0.353** l>alv>vel	0.407† l>alv>vel	0.730† l>vel>alv

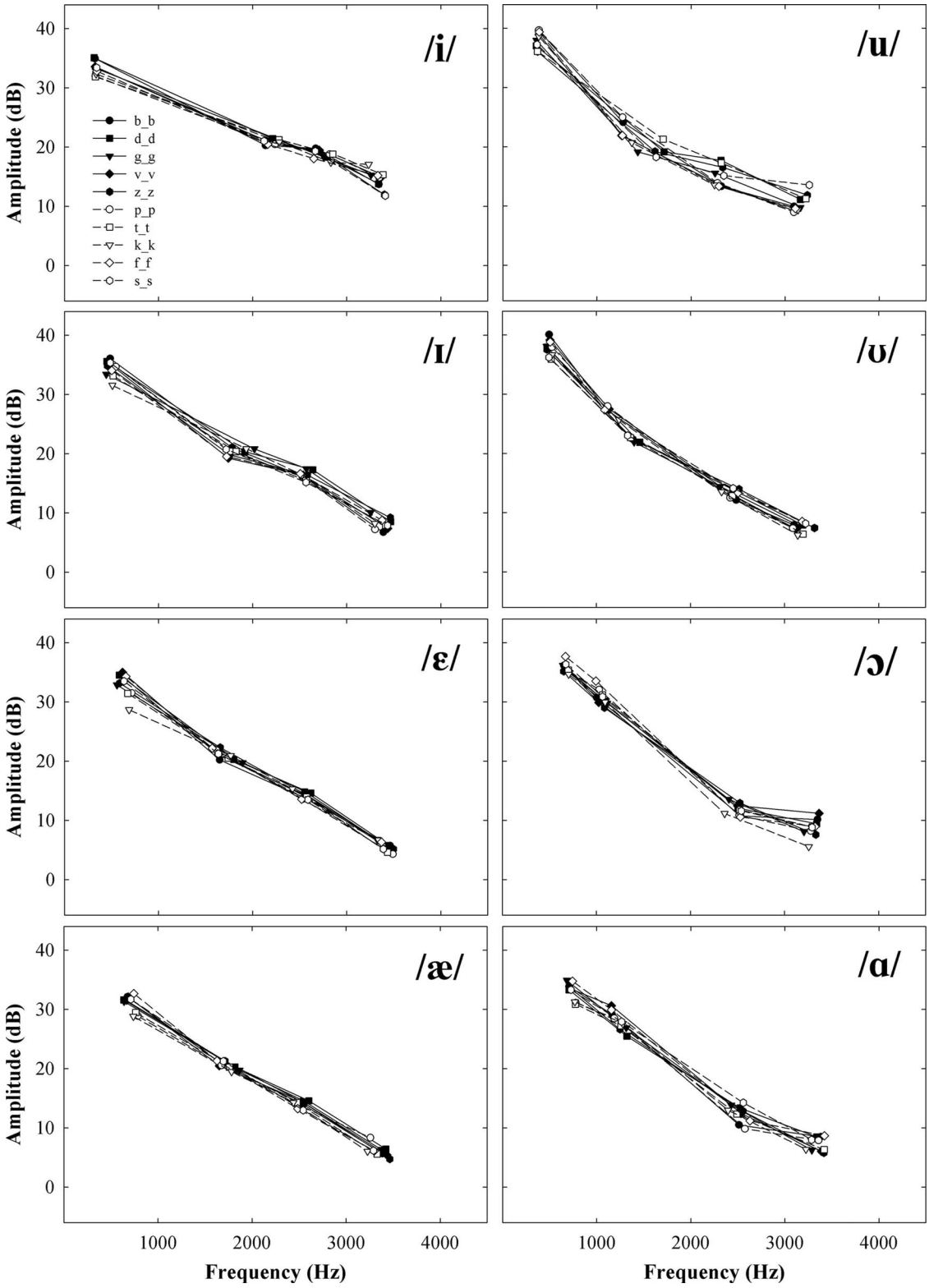

FIG. 5. Frequency and amplitude of formants 1–4 for individual vowels in ten consonantal contexts averaged for male speakers.

relative power cannot be advanced" (p. 112). The present results did not replicate this finding and did not provide any additional explanation for such atypically low amplitude values. However, in the current study, the amplitudes of these two vowels were not different from those for /ɪ/ and /ʊ/ within each front and back series, indicating that the greatest degree of opening may not automatically imply highest vowel amplitude.

Considering the consonantal context effects of voicing, manner, and place on amplitudes of individual vowels, the amplitude values did not differ greatly as a function of either voicing or manner. However, notable variation was found for consonant place: All vowels had higher amplitudes in the context of labials. The statistical assessments of the consonantal context effects are summarized in Table III. Table III lists the significant main effects only because significant in-

teractions were sparse and did not contribute greatly to the interpretation of the results. The main effect of voicing reached significance only for selected vowels (see Table III), and amplitude values were usually higher in the context of voiced stops. The main effect of manner was significant for the vowels /ɛ, æ, ɑ/ at rms peak. In each case, the peaks were higher in the context of fricatives. This tendency was consistent with HF results although decibel differences in the present data were much smaller, ranging from 0.5 to 0.7 dB, as compared to their relative power values. The effect of manner for the overall rms amplitude, however, produced greater amplitude in the context of stops. Overall, these results suggest that, generally, vowel amplitude does not change drastically as a function of either voicing or manner of articulation of the surrounding consonants.

To the contrary, the main effect of place was significant for all vowels except for /i/ for the stops-only analysis. Uniformly, the rms peaks were higher in the context of labial consonants as opposed to alveolar. The size of the place effects were greater too, with η^2 values ranging from 0.311 to 0.839. However, no consistent trend was found with respect to the alveolar and velar contexts (see Table III). The effects for the overall rms amplitude generally paralleled those at rms peak. The effect size of place was even stronger for the overall rms as evident from higher η^2 values listed in Table III. As a whole, the results for place do not match HF results, however, where vowels were greatest in relative power in the context of alveolars.

The main effect of gender was significant only for the vowel /i/ for both rms peak and overall rms, showing that male vowels had significantly higher amplitudes than female vowels. Given the relatively large degrees of freedom involved in the analyses, several interaction effects with gender as a factor were significant but had extremely low effect sizes ($\eta^2 < 0.10$), suggesting relatively unimportant (and, perhaps, even spurious) effects. They are not discussed here.

B. Relative amplitude of vowel formants

To get an overall picture of spectral variation for the present vowels in terms of both consonantal context and speaker gender, the mean frequency and amplitude values for the first four formants for each vowel in all ten consonantal contexts were plotted separately for males in Fig. 5 and for females in Fig. 6. As can be seen, consonantal context variously affected the amplitudes of individual formants and individual vowels. Frequency shifts are also detectable. Of particular note are the generally high F2 values for the vowel /u/ for both males and females which were additionally fronted in the context of alveolars. This is an outcome of regional variation in the vowel system of American English. The majority of the speakers for this study grew up in Ohio, where the variant of /u/ is significantly fronted. More acoustic evidence for the variation of F2 in the vowel /u/ as a function of speaker dialect can be found in Jacewicz *et al.* (2007a).

Figure 7 shows the mean differences between A2 and A1 (RelA2), A3 and A1 (RelA3), and A4 and A1 (RelA4) for each individual vowel and for each consonantal context. Because the first ANOVA showed that speaker gender was not significant for measure RelA2, RelA3, or RelA4, the display in Fig. 7 reflects amplitude values taken from Figs. 5 and 6 collapsed across speaker gender. As can be seen, mean RelA2 values remained rather steady for each front vowel (about −10 dB) whereas they were more variable for back vowels. In particular, they tended to increase with vowel openness, so that RelA2 for /u/ was about −18 dB and for /ɔ, ɑ/ was −5 dB. The variation as a function of consonantal context was quite extensive for most vowels although it was smaller for /u, ɑ/ and only minimal for /i, ɔ/.

Mean RelA3 values, although lower in general, tended to be higher for front vowels than for back (about −15 and −20 dB and lower, respectively). For back vowels, they were clearly the lowest for the vowel /ɔ/. The context-related variation was again evident for most vowels with the exception of /i, ɔ, ɑ/. The means for RelA4 again showed a sizeable variation as a function of consonantal context for some of the vowels.

The formant amplitude differences (RelA2, RelA3, and RelA4) were examined using separate ANOVAs which, for each vowel, assessed the effects of manner, voicing, and place separately for stops and fricatives sharing the same place of articulation (/b, d, p, t, v, z, f, s/) and for stops only differing in voicing and place (/b, d, g, p, t, k/) with gender as a between-subject factor. In the whole series of ANOVAs, the effect of speaker gender was significant only for the vowel /æ/ for RelA2 [$F(1,18)=6.65$, $p=0.019$, $\eta^2=0.270$] and RelA3 [$F(1,18)=7.76$, $p=0.012$, $\eta^2=0.301$] for the stops and fricatives analysis and for RelA2 [$F(1,18)=5.46$, $p=0.031$, $\eta^2=0.233$], RelA3 [$F(1,18)=8.97$, $p=0.008$, $\eta^2=0.333$], and RelA4 [$F(1,18)=4.77$, $p=0.042$, $\eta^2=0.210$] for the stops-only analysis, in each case indicating smaller decibel difference values for males. We therefore have chosen to exclude speaker gender as a factor for further consideration.

1. Variation of A1

In a separate analysis, we examined the variation of A1 as a function of consonantal context which is summarized in Fig. 8 for both male and female speakers.

The overall mean A1 values for male speakers were about 3 dB higher as compared to females (35 vs 32 dB). This relation was found in most vowels except for /i/ and /ʊ/ where the differences were greater (about 5 dB) and for /æ/, where mean male A1 was only 1 dB higher. As Fig. 8 shows, the variation as a function of consonantal context is particularly great for the vowels /ɪ, u/ for females and for /ɪ, ɛ/ for males.

We used separate ANOVAs with the within-subject factors manner, voicing, place, and gender as a between-subject factor to determine the significance of the A1 variations. A summary of significant effects is shown in Table IV.

The amplitude of A1, the strongest peak of a vowel, was affected by consonantal context to a relatively large extent. As can be seen, voicing had a significant effect on a number of vowels and was particularly strong for the short vowels /ɪ, ɛ, ʊ/. Unlike for the higher formants, all A1 peaks were higher in the context of voiced consonants and not voiceless. There were also significant effects of manner for three vow-

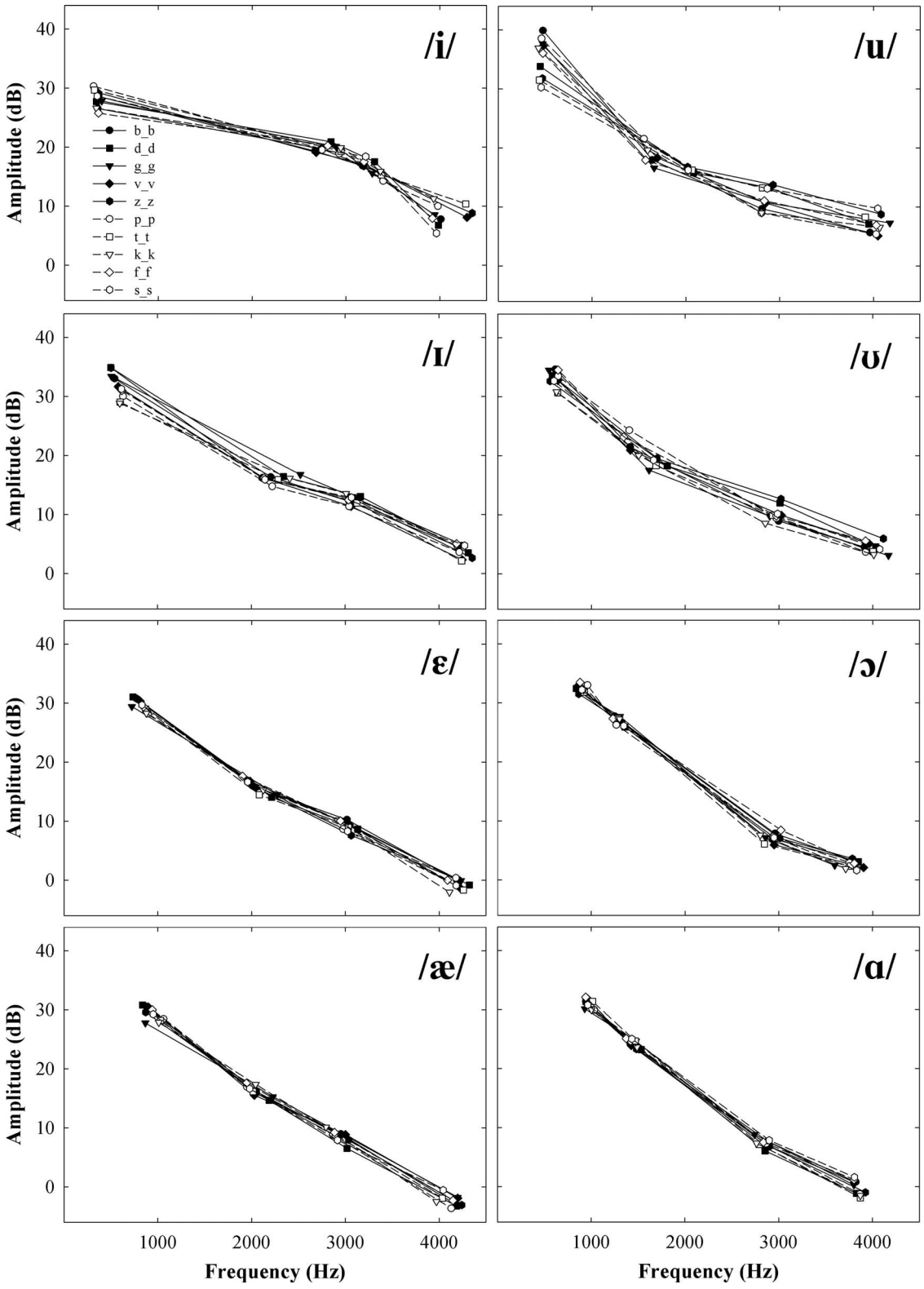

FIG. 6. Frequency and amplitude of formants 1–4 for individual vowels in ten consonantal contexts averaged for female speakers.

els /ɛ, æ, ɑ/, showing that A1 peaks were about 0.6–1.0 dB higher in the context of fricatives as opposed to stops. Place was significant for /ʊ, u/ for fricatives and stops, indicating that A1 in the context of labials was higher than in alveolars. For /u/, this effect was particularly strong (η^2=0.642) and A1 peaks were 4.4 dB higher. For the stops-only analysis, the results were variable. In general, the peaks in labial and alveolar contexts were higher as compared to velars. There were significant effects of gender for four vowels /i, ɛ, ɔ, ʊ/. In each case, A1 values for male speakers were higher than for female. One significant interaction deserves a mention here because it involved most of the vowels and showed a

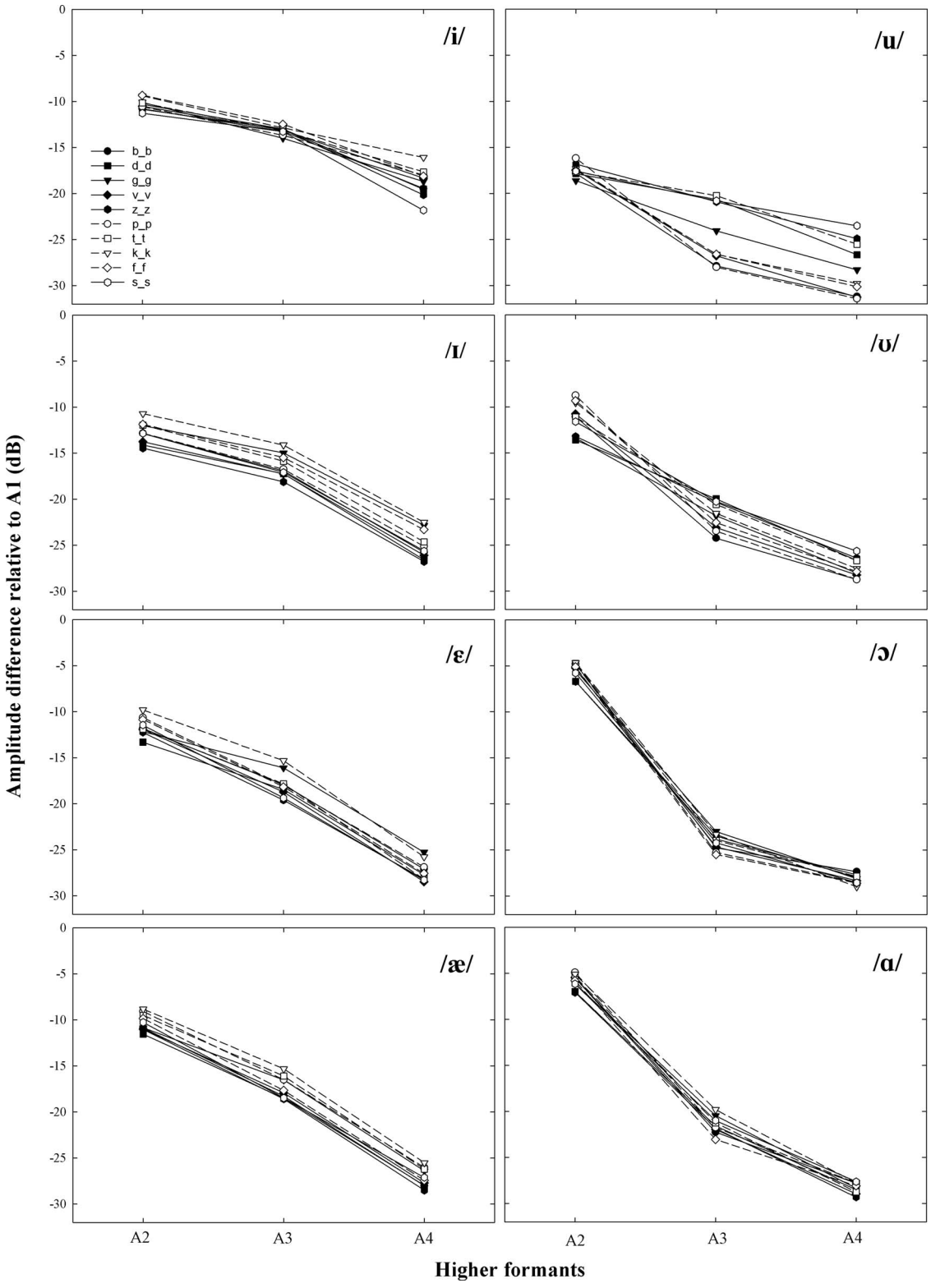

FIG. 7. Amplitude differences for each higher formant (A2, A3, A4) relative to A1 for individual vowels in ten consonantal contexts averaged across speaker gender.

consistent, relatively strong effect. The significant manner × voicing interaction arose from the fact that, for each affected vowel, the decibel difference between the higher A1 peak for voiced stops and the weaker A1 peak for voiceless stops was significantly greater (ranging from 2.18 to 2.88 dB) than the decibel difference between the A1 peaks for voiced and voiceless fricatives (ranging from −0.08 to 1.02 dB).

2. Results for RelA2, RelA3, and RelA4

Table V summarizes the statistical results for amplitude variations of the A2 and A3 relative to A1, i.e., RelA2 and RelA3. As can be seen, significant contextual effects were evident for all vowels except for /i/ and /ɔ/ for RelA2 and for /i, ɔ, ɑ/ for RelA3 analysis. In particular, the effect of voicing

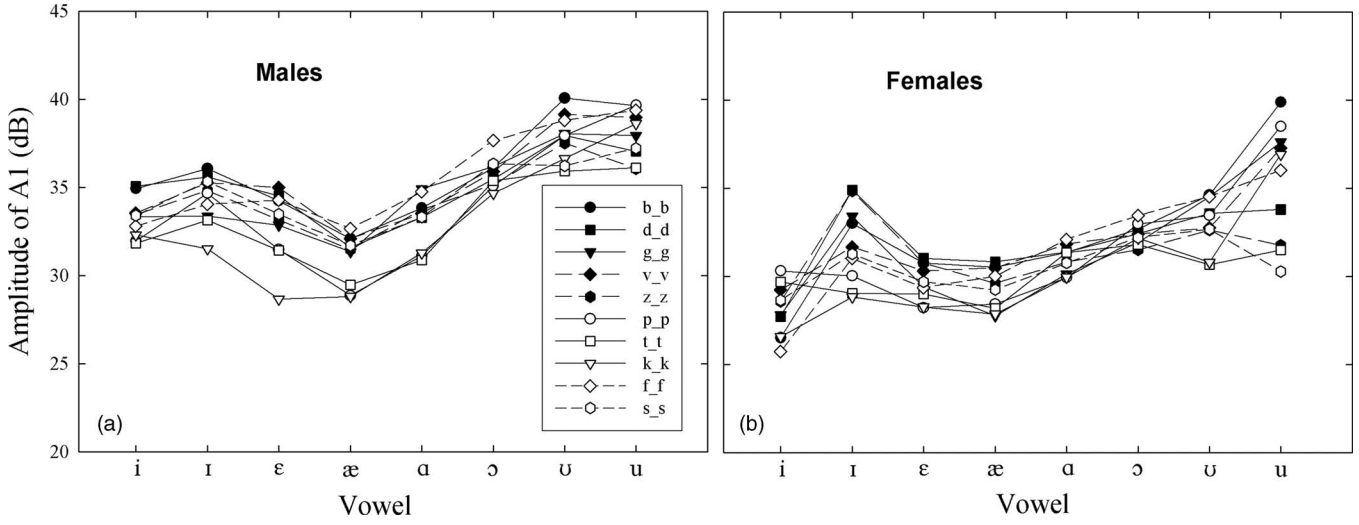

FIG. 8. Variation in the amplitude of the first formant (A1) for individual vowels in ten consonantal contexts averaged for male speakers (a) and for female speakers (b).

showed that the negative decibel differences were significantly smaller (about 2 to 3 dB) in the context of voiceless consonants as opposed to voiced, indicating that A2 was higher when surrounded by voiceless consonants. Consonant place introduced more variable results and affected primarily vowels in the context of stops. For front vowels, post-hoc analyses for stops revealed that decibel differences for both RelA2 and RelA3 were smallest in the context of velars, although this was not true for back vowels (compare Table V). The main effect of manner was not significant for any of the vowels. Significant interactions were sparse in the entire set and their small effects did not warrant a separate discussion.

While not necessarily related to consonantal context effects, visual inspection of Fig. 7 suggested that it was worthwhile to compare RelA3 values for front and back vowels. Across all contexts, these values tended to be smaller for front vowels than for back, indicating relatively stronger F3 peaks in front vowels. Results from a series of *t*-tests for RelA3 completed for all pairwise comparisons between vowels collapsed across all consonantal contexts confirmed these observations. There were no significant differences among the back vowels /u, ʊ, ɔ, ɑ/ except for /ɔ/ and /ɑ/ [$t(12) = -2.85$, $p = 0.015$]. For the front vowels, the differences among /ɪ, ɛ, æ/ were not significant. However, all pairwise comparisons with /i/ were significant, indicating that F3 for the vowel /i/ was significantly stronger than F3s for the remaining front vowels. No such consistency was found for all paired RelA2 and RelA4 comparisons, indicating that it is the relative amplitude of the third formant that most clearly marks the differences in spectral tilt between the front and back vowels.

TABLE IV. Summary of significant main effects and interactions from repeated measures ANOVAs for the amplitude of F1 (A1). Shown are partial eta squared values (η^2). (*) $p < 0.050$, (**) $p < 0.010$, (†) $p < 0.001$; (---) not significant; fr=fricative, st=stop, vd=voiced, vl=voiceless, l=labial, alv=alveolar, vel=velar, m=male, f=female.

	/i/	/ɪ/	/ɛ/	/æ/	/ɑ/	/ɔ/	/ʊ/	/u/
/b, d, p, t, v, z, f, s/								
Manner	---	---	0.206* fr>st	0.340** fr>st	0.241* fr>st	---	---	---
Voicing	---	0.441** vd>vl	0.546† vd>vl	0.326** vd>vl	---	---	0.246* vd>vl	---
Place	---	---	---	---	---	---	0.445** l>alv	0.642† l>alv
Manner × voicing	---	0.434**	0.227*	0.287*	---	0.301*	0.318*	---
Gender	0.312* m>f	---	0.364** m>f	---	---	0.361* m>f	0.371** m>f	---
/b, d, g, p, t, k/								
Voicing	---	0.544† vd>vl	0.432** vd>vl	0.357** vd>vl	0.418** vd>vl	---	0.506† vd>vl	---
Place	---	0.283** l>alv>vel	0.346† alv>l>vel	0.234** alv>l>vel	---	---	0.253** l>vel>alv	0.560† l>vel>alv
Gender	0.339** m>f	---	0.257* m>f	---	---	0.322* m>f	0.392** m>f	---

TABLE V. Summary of significant main effects from repeated measures ANOVAs for decibel difference A2-A1 (RelA2) and A3-A1 (RelA3). Shown are partial eta squared values (η^2). (*) $p<0.050$, (**) $p<0.010$, (†) $p<0.001$; (---) not significant; vd=voiced, vl=voiceless, l=labial, alv=alveolar, vel=velar.

	/i/	/ɪ/	/ɛ/	/æ/	/ɑ/	/ɔ/	/ʊ/	/u/
RelA2								
/b, d, p, t, v, z, f, s/								
Voicing	---	0.362** vl<vd	0.623† vl<vd	0.458** vl<vd	0.440** vl<vd	---	0.532† vl<vd	0.204* vl<vd
Place	---	---	0.252* l<alv	---	---	---	0.496** l<alv	---
/b, d, g, p, t, k/								
Voicing	---	0.433** vl<vd	0.632† vl<vd	0.667† vl<vd	0.427** vl<vd	---	0.575† vl<vd	0.339** vl<vd
Place	---	0.300** vel<alv<l	0.315** vel<l<alv	0.186* vel<l<alv	---	---	0.382† l<vel<alv	---
RelA3								
/b, d, p, t, v, z, f, s/								
Voicing	---	---	0.291* vl<vd	---	---	---	---	---
Place	---	---	---	---	---	---	0.466** alv<l	0.715† alv<l
/b, d, g, p, t, k/								
Voicing	---	0.253* vl<vd	0.289* vl<vd	0.287* vl<vd	---	---	0.233* vl<vd	---
Place	---	0.475† vel<alv<l	0.438† vel<l<alv	0.304** vel<l<alv	---	---	0.296** alv<vel<l	0.677† alv<vel<l

The context effects for RelA4 were comparatively smaller as shown in Table VI. The two high vowels /i, u/ were affected the most. Interestingly, place effects were different for each vowel. For /i/, the decibel difference was 1.3 dB smaller for the labial contexts as opposed to alveolar. The effect of place was particularly strong for the vowel /u/. For the stops and fricatives analysis, RelA4 was 7.5 dB higher in the context of alveolars as opposed to labials and the same tendency was also found for the stops-only analysis. Significant effects of place were also found for /ʊ/. For stops-only analyses, place was significant for /ɪ, ɛ, æ/ and the decibel differences were smallest in the context of velars, and for /u/, where the smallest decibel difference occurred in the context of alveolars. The effect of manner was significant for /u/ and /ɔ/, although showing opposite effects in that the decibel differences were smaller in the context of fricatives for /u/ (2.2 dB) and in the context of stops for /ɔ/ (1.4 dB). Voicing was significant for two vowels only, for /ɪ/ for fricatives and stops, and for /i/ for the stops-only analysis. In each case, the differences were about 2.5 dB smaller in the voiceless contexts as opposed to voiced.

Summarizing the results for formant amplitude, the consonantal context effects varied with vowel category and with a particular formant. For example, almost no significant effects were found for the vowel /ɔ/ across all three measures whereas numerous effects and interactions occurred for the short vowels /ɪ, ɛ, ʊ/. Similarly, there were no context effects for the vowel /i/ for RelA2 whereas they were abundant for

TABLE VI. Summary of significant main effects from repeated measures ANOVAs for decibel difference A4-A1 (RelA4). Shown are partial eta squared values (η^2). (*) $p<0.050$, (**) $p<0.010$, (†) $p<0.001$; (---) not significant; fr=fricaive, st=stop, vd=voiced, vl=voiceless, l=labial, alv=alveolar, vel=velar.

	/i/	/ɪ/	/ɛ/	/æ/	/ɑ/	/ɔ/	/ʊ/	/u/
/b, d, p, t, v, z, f, s/								
Manner	---	---	---	---	---	0.272* st<fr	---	0.288* fr<st
Voicing	---	0.261* vl<vd	---	---	---	---	---	---
Place	0.245* l<alv	---	---	---	---	---	0.235* alv<l	0.645† alv<l
/b, d, g, p, t, k/								
Voicing	0.216* vl<vd	---	---	---	---	---	---	---
Place	0.266** vel<alv<l	0.383† vel<l<alv	0.268** vel<l<alv	---	---	---	---	0.518† alv<vel<l

TABLE VII. Mean vowel durations (in milliseconds) (s.d.) in consonantal contexts grouped by voicing, manner, and place of articulation. All vowels are pooled. The data from House and Fairbanks (1953), shown as HF, are included for a comparison with the present results (JF).

Grouped consonant environments	JF females	JF males	HF males
Voicing			
Voiceless	200 (73)	172 (59)	174
Voiced	306 (68)	257 (67)	253
Manner of articulation			
Stop	235 (85)	200 (70)	203
Fricative	280 (85)	236 (80)	239
Place of articulation			
Labial	249 (85)	209 (73)	220
Alveolar	268 (91)	227 (80)	232
Velar	230 (82)	200 (71)	198

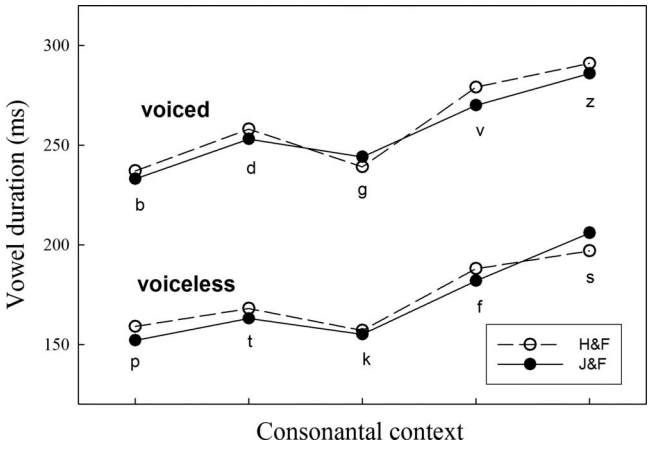

FIG. 9. Consonantal context effects on vowel duration pooled across all vowels. The present results, marked as JF, are compared with the data from House and Fairbanks (1953), marked as HF.

RelA4. Conversely, context affected RelA2 for /ɑ/ but not RelA3 or RelA4. Turning to the significant effects of consonantal features on formant amplitude variation, manner of articulation had almost no influence on amplitude changes except for a few interactions and main effects for RelA4. Both voicing and place affected formant amplitude variation to a much greater extent, as evident in a greater number of significant main effects. The effects of voicing were very consistent, showing uniformly that the formant peaks were higher in the voiceless contexts as opposed to voiced. Consonant voicing affected the greatest number of vowels and formants. The effects of place were more variable, although formant peaks tended to be higher in the context of labials as compared to alveolars. For stops only, the peaks were highest in the context of velars for front vowels whereas more variability was found in back vowels. Overall, there were more significant context-related main effects and interactions for front vowels (32) than for back (20).

C. Vowel duration

Although vowel duration was not of central interest to the study, this measure can be informative in light of the differences in amplitude results obtained between the HF study and the present report. Specifically, could differences in the articulation rates between the speakers in the two studies account for some of the discrepancies?

Table VII lists the duration values as a function of consonant environment grouped by voicing, manner, and place of articulation for all vowels pooled separately for males and females. With regard to the male data, the consistency with HF is remarkable given that only four vowels studied by HF were included in the present set.[2] The match is almost perfect despite the apparent differences in the way the material was recorded, i.e., as isolated disyllabic words in HF and as monosyllables in a short phrase in the present study. As might be expected, durations of female vowels were longer as compared to males (e.g., Adank *et al.*, 2007; Jacewicz *et al.*, 2007b).

Following HF, Fig. 9 displays mean durations for each individual consonantal context across all vowels pooled. Since the present focus was to compare the duration material with HF data,[3] only duration values from male speakers were included in Fig. 9. Again, the obtained pattern agrees remarkably with that found in HF. These results suggest that the articulation rates for the speakers in both the HF study and the present report were very similar.

As might be expected, the most robust duration differ-

TABLE VIII. Mean durations of individual vowels (in milliseconds) (s.d.) shown for voiced (Vd_Vd) and voiceless (Vl_Vl) environments and for all consonantal contexts pooled. Shown are also mean f_0 values (in hertz) (s.d.) for vowels for all consonantal contexts pooled.

Vowel	Duration Vd_Vd males	Duration Vl_Vl males	Duration Vd_Vd females	Duration Vl_Vl females	f_0 males	f_0 females
/ɪ/	204 (42)	125 (30)	250 (45)	142 (44)	138 (17)	217 (20)
/ʊ/	221 (43)	136 (35)	253 (42)	163 (47)	138 (16)	220 (23)
/ɛ/	225 (47)	153 (43)	277 (53)	172 (55)	131 (14)	213 (18)
/i/	257 (57)	173 (54)	305 (53)	205 (74)	145 (17)	237 (25)
/u/	267 (60)	166 (36)	309 (58)	209 (59)	144 (16)	235 (24)
/ɑ/	287 (70)	198 (56)	342 (58)	234 (68)	125 (14)	212 (19)
/ɔ/	300 (68)	211 (67)	374 (57)	246 (88)	129 (15)	207 (19)
/æ/	305 (66)	217 (62)	352 (63)	242 (69)	127 (15)	210 (18)

ences among the vowels came from the effects of consonant voicing. Table VIII lists durations of individual vowels (in ascending order) for male and female speakers, showing that vowels in the context of voiced consonants were longer than in the voiceless contexts. Also significantly longer vowels (18% longer) were produced by females [$F(1,10)=6.17$, $p=0.032$, $\eta^2=0.382$]. The main effect of consonant voicing was strong [$F(1,10)=134.7$, $p<0.001$, $\eta^2=0.931$] and the differences were very consistent within each gender group. On average, vowels in the context of voiced consonants produced by males were 36% longer than vowels in the voiceless contexts whereas the difference for female speakers was 37%. Comparing the present results with more recent published data, we found consistency with vowel duration material in Hillenbrand *et al.* (2001).

Considering the durations of individual vowels, the results for the voiced contexts were less variable than for the voiceless. In particular, paired two-tailed *t*-tests showed that there were no significant differences between /ɔ/ and /æ/ or between /æ/ and /ɑ/—the longest vowels or between the vowels /i/ and /u/—vowels of intermediate length. There were also no significant differences between /ɪ/ and /ʊ/ nor /ɛ/ and /ʊ/ —short vowels. All other differences were significant.

D. Variation in f_0

Although f_0 measure was also not of central interest to this study, the present f_0 values were compared with the results of HF for consistency. Following HF, f_0 values at the rms peak averaged across all vowels were plotted for each of the ten individual consonantal contexts (the plots are not included here). Although the present f_0 values for males were slightly higher than those in HF (on the order of 9%), the general trends were strikingly similar. Most important, f_0's of vowels in voiceless contexts were higher than those in voiced ones. As in HF, the difference in f_0 values as a function of voicing of the surrounding consonants tended to be greater for stops than for fricatives. These results match rather closely the f_0 material presented in HF. f_0 values for individual vowels averaged across all consonantal contexts and split by speaker gender are listed in Table VIII. The general tendency reflects the widely observed phenomenon that high vowels have higher intrinsic f_0 than low vowels (e.g., Whalen and Levitt, 1995).

IV. GENERAL DISCUSSION

The aim of this study was to characterize the nature, size, and range of acoustic amplitude variation in coarticulated vowels. Expecting large variation based on the findings in the House and Fairbanks (1953) study, we examined the changes to the amplitudes of the rms peak and the overall rms amplitude as a function of the consonantal features voicing, manner, and place. Contrary to the HF results, we found no systematic changes in vowel amplitude as a function of either consonant voicing or manner.

Both the present study and HF found significant effects of consonant place of articulation. However, although the small size of the amplitude variation was comparable, the directions of the effects for the place distinction differed. In particular, vowels in the context of alveolars were greater in relative power than labials in the HF data whereas the present results show the opposite: Both rms peak and overall rms amplitude were higher in the context of labials and not alveolars. In a similar fashion, vowels in the context of velars were lesser in relative power than in the context of alveolars in the HF study whereas no difference was found for these two contexts in the present study. On the whole, the present results show that consonantal context of a vowel does not introduce sizeable acoustic variation in vowel amplitude for either rms peak measure or overall rms amplitude.

A. Accounting for the discrepancy between the two sets of results

In attempting to explain the discrepancy between the central results of the HF study and the present results, we first considered the differences in the experimental protocol. Since the speech material in HF consisted of isolated words and the present stimuli were recorded in a short phrase, the possible differences in the articulation rate were a primary candidate for accounting for the discrepancy. As it turned out, the durations of vowels in the consonantal contexts in HF and in the present study were almost identical, indicating that the discrepancy was not due to a more formal or more casual speaking mode used in the recordings of the two sets of stimuli. We then turned to the equipment used in both studies and, although some slight incompatibilities might have been expected due to the analog and digital recordings, we considered each experimental setup to be adequate for the experimental purpose it served. The third possibility was in the way the acoustic measurements were made then and now. As will be shown in the following, the differences between the results of the two studies can probably be explained, at least to a great extent, by the different measurement techniques used.

Perhaps the greatest single discrepancy between the HF results and those presented here involves the voicing difference. In particular, HF's measurements indicated that vowels produced in the context of voiced consonants had twice the relative power of vowels produced in the context of voiceless consonants. One possible explanation for the difference lies in suggesting that the HF vowel measurements included some portion of the consonantal context, especially for voiceless consonants. In particular, we must assume that there was an integration window associated with their apparatus and associated measurement of intensity. If this window was large enough, then it might also have included a portion of the voiceless consonant context (syllable-initial, syllable-final, or both) since vowels produced in a voiceless context are significantly shorter than those in a voiced context. This account would explain many of the consonant context differences in relative power seen in HF (see their Fig. 3).

For example, we could expect that the voiced stops would show greater relative power than voiceless stops because the relative power measures for the vowels in a voiceless context would include portions of the voiceless stop itself. We would also expect vowels in the context of /s/ to

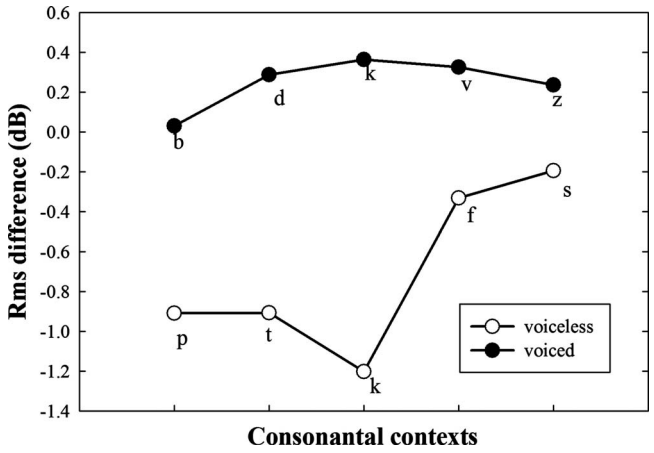

FIG. 10. Mean differences between overall rms (calculated from vowel onset to offset) and the rms from a 200 ms section of the utterance centered on the vowel across the ten consonantal contexts. Data are collapsed across two speakers and eight vowels.

show greater relative power than in the context of /f/ (since the relative power of the /s/ consonant, a portion of which is included in calculation of the vowel's power, is greater than that of the /f/). In turn, we would expect vowels in the context of /f/ to be slightly greater in relative power than when they occur in the context of /p, t, k/. This approach would thus also account for some of the differences between HF and the current study related to consonant manner. These patterns of consonantal feature differences were not found in the present study simply because our method only measured that portion of the speech signal that corresponded to the vowel.

As a test to determine whether this explanation is, in fact, plausible, the recordings from two speakers (one male, one female) were randomly selected and the rms of their vowels were analyzed in two separate ways. First, the rms of the target vowel, alone, was measured (this corresponded to the overall rms measure). Next, the rms of a 200-ms-long section of each utterance—centered on the midpoint of the target vowel—was measured. This 200 ms section of the utterance included portions of the initial and final consonants (which would lower the rms when the vowel was relatively short). Finally, the difference between the overall rms and the rms of the 200 ms section for each token was calculated. Shown in Fig. 10 are the mean rms differences collapsed across both speakers and all eight vowels. The pattern of the consonant context differences (between the voiced and voiceless contexts, and between the /p, t, k/ and /f, s/ context) shown in Fig. 10 bear a striking resemblance to HF (Fig. 3) suggesting the viability of this explanation for the obtained discrepancies.

B. Consonantal context effects on formant amplitude variation

Turning to the present results, we explored the possibility that significant context-related amplitude variations may affect specific spectral regions in a vowel (such as individual formants) rather than its overall energy. The changes in the amplitude of vowel formants were first examined by measuring amplitude of each higher formant relative to each vowel's strongest spectral peak, i.e., the amplitude of F1. This analysis revealed several systematic effects for each formant.

For F2, variation in A2 relative to A1 (RelA2) was significant for all vowels except for /i/ and /ɔ/. RelA2 was affected primarily by consonant voicing and place. For each vowel, the RelA2 values were uniformly smaller in the context of voiceless consonants as opposed to voiced, and they were variably affected by consonantal place. The variation ranged from 0.8 to 3.8 dB.

For F3, the RelA3 values did not vary greatly for the analysis involving labial and alveolar stops and fricatives. A large difference of 7.7 dB was found for the vowel /u/ as a function of place. However, more significant effects were found for the stops-only analysis. The RelA3 values were again smaller in the context of voiceless stops as opposed to voiced and the effects of place were variable. In general, the RelA3 variation was of the range 1.5–3.3 dB. The vowels /i, ɔ, ɑ/ remained unaffected by consonantal context. However, the place effects for the vowel /u/ were very strong. For this vowel, the decibel differences produced by context were large: 6.7 dB between alveolars and velars, and 8.7 dB between alveolars and labials. The consonant place effects for the vowel /u/ were thus the strongest in the entire set of vowels studied here. A3 peaks in the context of alveolars were higher than in the context of labials.

Finally, variation in the amplitude of F4, typically the weakest formant of a vowel, turned out significant for several vowels. Interestingly, the effects of place and voicing were significant for RelA4 for the vowel /i/ (1.3–2.6 dB) which otherwise was unaffected by contextual variation. The place effects for /u/ were again strong, reaching 7.5 dB difference for alveolars as opposed to labials for the contexts of stops and fricatives. For the stops-only analysis, A4 was 4.4 dB higher for alveolars as compared to velars and 7.3 dB higher in comparison with labials.

For the amplitude of F1 (A1), there was a significant effect of consonant voicing. A1 peaks had greater amplitude (ranging from 1.0 to 3.2 dB) in the voiced contexts than in the voiceless contexts. This was different from the patterns for the higher formants suggesting that A1 may be affected by other source characteristics. The results for place lead to the same conclusion. Although the amplitudes of higher formants were uniformly greater in the context of alveolars, A1 values were mostly greater in the context of labials. This was true for high vowels /u, ʊ, ɪ/ whose F1 frequency is low. For the vowels /ɛ, æ/ however, whose F1 frequencies are higher, A1 values were greater in the context of alveolars. This result was significant for the stops-only analysis. For the stop and fricatives analysis, the effect of place was not significant for these vowels.

C. Context-induced variation in relation to "basic rules of formant amplitude"

The observed amplitude variations as a function of consonantal context need to be situated within a well-established theoretical frame which lets us anticipate a substantial proportion of the effects found here. As Fant (1956) first noted,

formant frequencies and formant amplitudes are "intimately related" to the extent that two vowels differing in the frequency of F1 must also differ in the overall amplitude. This relation is carried over to the higher formants. Stevens 1998, (pp. 133–135) illustrates the three general rules of formant amplitude as stemming from the principles of linear resonances in series. All three rules are well represented in the present set of results, as shown in Figs. 4–6. In particular, the greater overall rms of back vowels as compared to front was due to the greater amplitude of F1 in back vowels. This can be accounted for by the first rule, which expresses the relation between the greater amplitude of F1 or higher frequency of F1 and the increased overall amplitude. Also, consistent with the first rule, the overall rms of the vowels increased as F1 rose in frequency. Although the overall rms of the vowels /æ, ɑ/ was slightly smaller than /ɛ, ɔ/ despite their higher F1 frequencies, this effect can be accounted for by the fact that the amplitude of F1 was on average about 2 dB greater for /ɛ, ɔ/ than for /æ, ɑ/.

A clear manifestation of the second rule, which states that increasing a particular formant frequency raises the amplitude of the spectrum at frequencies above that formant, can be found in front vowels. In the present set, the amplitude of F3 rose as a result of the increase in the frequency of F2 for the front series. This rule also explains the greater A3 for the vowel /u/ whose frequency of F2 rises in its variant spoken in Ohio. Finally, there is also an indication of the appearance of the third rule in the present set, which points to an increase in the amplitude spectrum in the vicinity of two or more formants coming close together. The increased amplitude of F2 in /ɔ, ɑ/ and of the F2–F3–F4 cluster in /i/ can be interpreted in light of the third rule.

The consonantal context effects on variation in formant amplitude observed in this study cannot be interpreted without reference to this pattern of expected changes to the amplitude spectrum. For example, consonant voicing did affect the amplitude of F1 but, apparently, the effect was not strong enough to influence the overall vowel amplitude and the first rule failed to apply. However, the context-induced variation in the frequency of a higher formant could in principle introduce changes to the amplitude spectrum at frequencies above that formant. The consonantal context effects on the vowel /u/ are the most illustrative example in the present set, where operation of the second and third rules cannot be excluded.

D. The role of formant amplitude variation in vowel perception

The size of variation in the amplitude of each higher formant observed here leads to the question of how relevant are these changes to the perception of vowel quality. In particular, one would not necessarily expect a vowel category change as a function of amplitude variation related to the place of articulation or voicing of consonants surrounding the vowel. In the typical speech perception paradigm involving isolated synthetic stimuli, the experiments demonstrated that adjustments made to formant amplitudes can change the perceived vowel quality (e.g., Miller, 1953; Lindqvist and Pauli, 1968; Carlson *et al.*, 1970; Ainsworth and Millar, 1972) and that a 8 dB change in the amplitude of A3 can be sufficient for the listener to perceive a different vowel category (Aaltonen, 1985). A much smaller difference limen was reported by Flanagan (1957) in a discrimination task, in which the 50%-crossover point for F2 of the vowel /æ/ was reached when the amplitude of this formant increased by 3 dB. Furthermore, listeners were shown to be equally sensitive to amplitude adjustments in F4 and in F2 for /ɪ/ and again, a 8 dB change caused a vowel category shift (Jacewicz, 2005).

These reports point to the potential role of variation in formant amplitudes in vowel perception. However, the variation ranging from 0.8 to 3.8 dB which was found in naturally produced vowels in this study is less likely to introduce a vowel quality change. After all, listeners do not depend on formant amplitude relations in making vowel quality distinctions to the extent that they rely on formant frequency pattern. Yet, the large 8.7 dB difference for the vowel /u/ produced by coarticulatory context as reported here has a great potential to affect the perception of vowel quality. In most cases, however, the sensitivity to formant amplitude variations may reflect listeners' experience with more subtle spectral changes which may contribute to the perceived vowel naturalness or timbre differences rather than vowel quality per se.

E. Consonantal context effects on amplitude spectrum at the auditory periphery

Before any expectations about the relevance of formant amplitude variation to vowel perception can be advanced, it is important to know whether such variation is still present in the auditory representation of the speech signal, at least at the auditory periphery. To estimate how the spectral amplitude variation may be encoded and represented at early stages of processing by the peripheral auditory system, we used the first level of the Perceptual Auditory Spectral Centroid model (Anantharaman, 1998) which simulates the peripheral processing through three psychoacoustically motivated stages corresponding to the middle-ear transduction, the filtering mechanism exhibited by basilar membrane, and hair cell activity transmitted in the auditory nerve which contributes to the sensation of loudness. Auditory spectra were calculated by passing the vowel spectra (as a positive time-frequency representation of the vowel signal) through the three stages of peripheral auditory processing: (1) equal-loudness preemphasis curve (provides a weighting along the frequency axis), (2) auditory filter bank (passes the output of the previous stage through 33 gammatone filters centered at each bin of the fft, center frequencies in ERBs), and (3) intensity-loudness compression (raises the output of the previous stage to the 0.3 power).

Figure 11 (bottom) shows the average auditory spectra of the vowel /ɪ/ in the contexts of labial, alveolar, and velar stops which were calculated from the productions of the present ten male speakers. For the same speakers, the corresponding measured mean frequencies of formants F1–F4 along with their amplitudes are included in the top panel. The formant peaks are situated within the four contiguous broad frequency bands B1–B4 [one band per octave, in kilohertz B1 (0–0.5), B2 (0.5–1.0), B3 (1.0–2.0), and B4 (2.0–

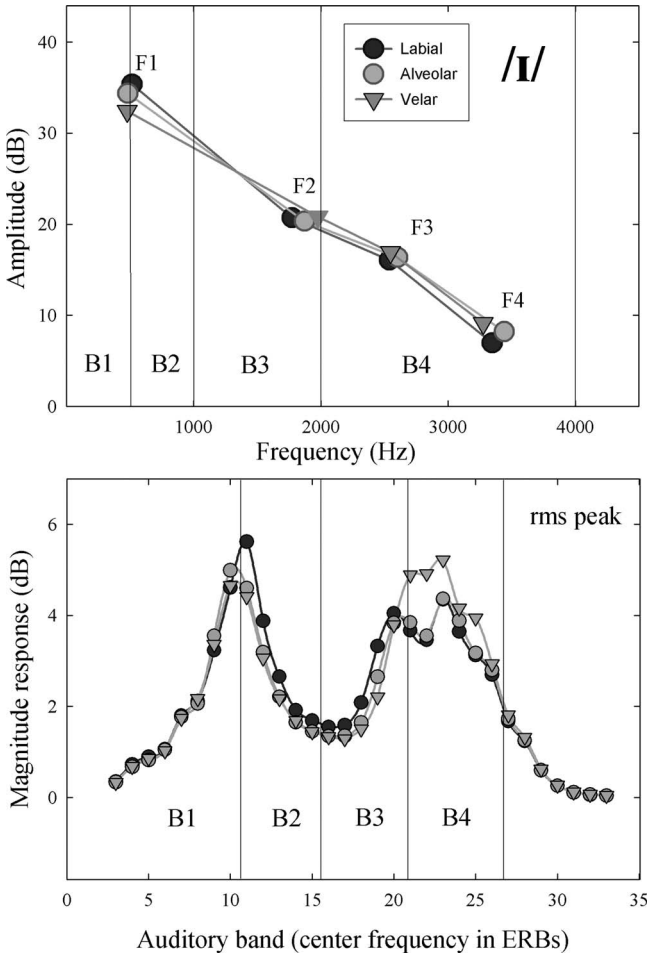

FIG. 11. Mean formant amplitude of the vowel /ɪ/ in the context of stops averaged across ten male speakers in this study (top) and the calculated auditory spectra based on the magnitude response of a gammatone filter bank (bottom).

4.0)]. These bands are also included in the auditory spectrum display to relate the two representations, i.e. acoustic (linear) and auditory (nonlinear), to one another. As evident, the acoustic amplitude variation is carried on through the early stages of auditory processing and the magnitude response in the auditory spectrum is remarkably consistent with the acoustic pattern. In particular, the specific significant consonantal context effects for the vowel /ɪ/ for the measures RelA2, RelA3, RelA4, and A1 summarized in Tables IV–VI can be clearly traced in the auditory spectrum. That is, the statistically strong place effects for RelA2, RelA3, RelA4 (see the results for the stops-only analysis for the vowel /ɪ/), indicating higher amplitude peaks in the context of velars are manifested in the auditory spectrum at the juncture of the bands B3 and B4 and throughout band B4. Clearly, the second peak in the auditory spectrum (which corresponds to the F2, F3, and F4 peaks) is highest in the context of velars. A close correspondence between the acoustic amplitude variations can also be found in the lower frequency region, representing F1 in the acoustic representation and the first peak in the auditory spectrum. The statistically strong effect of place for the A1 measure showing higher amplitude in the context of labials is also evident in the auditory spectrum.

The auditory spectra calculated from the measurements at rms peak show that two spectral regions—corresponding to the lower and higher parts of the measured acoustic spectrum—are emphasized. The question arises as to which spectral region contributes more to the overall vowel amplitude. In Table III, we observe that the significant effect of place for the vowel /ɪ/ for both the rms peak and the overall rms shows higher amplitude values in the context of labials. Relating this effect to the place effects within the spectral regions, we again find higher amplitude values in the context of labials for the lower spectral region (A1) but not for the higher (RelA2, RelA3, and RelA4), where the amplitude is greater in the context of velars. This may indicate that the lower spectral region dominates the vowel spectrum and the contribution of the higher region is comparatively smaller. However, despite the existing evidence (based upon psychophysical tuning) that in the region below 1000 Hz the ability to process spectral contrasts increases (e.g., Healy and Bacon, 2006), this interpretation needs to be tested in a series of further experiments.

At present, any generalization should be drawn with caution due to the large number of vowels and consonantal contexts used in this study. A more focused investigation should concentrate on selected vowels and contexts to gain more insight into the exact causes of differences in the size of amplitude variation found here. A detailed analysis deeply rooted in the acoustic theory of speech production is especially desirable in explaining some of the consonantal context effects on vowel amplitude variation. For example, spectral amplitude variations reflect a complex relation between the characteristics of the sound in the low-frequency region and acoustic attributes at frequencies above the first formant. These include the timing of acoustic events which depend on the length and the place of the consonantal constriction preceding the vowel. These events, such as rate of movement of F1 after the stop release and rate of movement of trajectories of higher formants following the release have an effect on amplitude spectrum (Stevens, 1998).

The significant effects of consonantal context on formant amplitude variations examined here were found at a time window relatively late in a vowel, i.e., at rms peak. This shows that significant variations do not occur only at vowel onsets and offsets (the vowel margins) whose formant frequencies are affected most by surrounding consonants but such effects can also be manifested at vowel nuclei. However, given the variable durations of the vowels and, understandably, variable locations of the rms peaks, it was not surprising to find different degrees of their susceptibility to the context effects. For example, the long vowels in the set such as /ɔ/ and /ɑ/ were relatively unaffected by contextual variation. This would indicate a greater temporal separation of their nuclei from syllable margins. However, this interpretation may not be true with respect to the vowel /æ/ which was the longest in the present set and exhibited more contextual effects than either /ɔ/ or /ɑ/. Also, the vowel /i/, being of intermediate length, was unaffected by contextual variation except for the variation in A4. These mixed results indicate that differences in vowel duration alone cannot entirely explain the varying degrees of context effects found at

vowel nucleus. Finally, there may also be a source variation in the vowel as a function of consonantal context—a study of contextual variations of the voice source can be found in Fant (1997). Fine-grained analyses of the voice source have shown that the influences on the vowel's mode of phonation depend on the manner of articulation of the consonant (Gobl and Ní Chasaide, 1999) and consonant voicing (Gobl and Ní Chasaide, 1993). The source variation in consonantal context may also have a profound effect on variation in the amplitudes of formants.

The type of variation in formant amplitude found in coarticulated vowels in the present study naturally co-occurs with speaker-intended changes to vocal intensity to convey emphasis, linguistic stress, emotions, or to communicate in noisy environments or over a longer distance. How listener deals with such complex interaction between several factors influencing vowel perception is a rich area for future research.

ACKNOWLEDGMENTS

This study was supported by the research Grant No. R03 DC005560 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health, to the first author. The authors would like to thank Chiung-Yun Chang for assistance at various stages of this research and Jeff Murray for editorial help. The comments and suggestions of the Associate Editor Anders Löfqvist and two anonymous reviewers on an earlier version of the paper are greatly appreciated.

[1]The location of the rms peak can be expressed in several ways. First, the peak rms position can be determined in terms of its absolute distance (in milliseconds) from vowel onset or CVC-word onset. Alternatively, the peak rms position can be expressed in terms of its relative position (with values ranging from 0 to 100%) with regard to the duration of the vowel or duration of the word. We experimented with each of these approaches to determine which measure best characterized the variation in the location of rms peak as a function of vowel category and consonantal context. Based on the statistical results, it became clear that the two relative positions (i.e., within the vowel and within the word) yielded the most comparable patterns. We therefore decided to present only the results for the most dependable measure.

[2]In the HF study, the voiced environments included the nasals /m, n/ which were absent in the present report. Also, in their Table 2, HF listed vowel durations separately for bilabial (/p, b, m/) and labiodental (/f, v/) consonants. In the present study, the HF's values for bilabials and labiodentals were averaged and listed in Table VII as labials. Also, postdental consonants in HF included /n/ which was absent in the present alveolar environments.

[3]HF data displayed in their Fig. 1 are taken from their Table 2.

Aaltonen, O. (**1985**). "The effects of relative amplitude levels of F2 and F3 on the categorization of synthetic vowels," J. Phonetics **13**, 1–9.

Adank, P., van Hout, R., and van de Velde, H. (**2007**). "An acoustic description of the vowels of northern and southern standard Dutch. II. Regional varieties," J. Acoust. Soc. Am. **121**, 1130–1141.

Ainsworth, W. A. (**1972**). "Duration as a cue in the recognition of synthetic vowels," J. Acoust. Soc. Am. **51**, 648–651.

Ainsworth, W. A. (**1981**). "Duration as a factor in the recognition of synthetic vowels," J. Phonetics **9**, 333–342.

Ainsworth, W. A., and Millar, J. (**1972**). "The effect of relative formant amplitude on the perceived identity of synthetic vowels," Lang Speech **15**, 328–341.

Anantharaman, J. N. (**1998**). "A Perceptual Auditory Spectral Centroid Model," Ph.D. dissertation, The Ohio State University, Columbus, OH.

Beckman, M. (**1986**). *Stress and Non-Stress Accent* (Foris, Dordrecht).

Carlson, R., Granström, B., and Fant, G. (**1970**). "Some studies concerning perception of isolated vowels," STL-QPSR 2-3/1970, pp. 19–35.

Fant, G. (**1956**). "On the predictability of formant levels and spectrum envelopes from formant frequencies," in *For Roman Jakobson: Essays on the Occasion of his Sixtieth Birthday*, edited by M. Halle, H. G. Lunt, H. McLean, and C. H. van Schooneveld (Mouton, The Hague), pp. 109–120.

Fant, G. (**1997**). "The voice source in connected speech," Speech Commun. **22**, 125–139.

Fant, G., Fintoft, K., Liljencrants, J., Lindblom, B., and Martony, J. (**1963**). "Formant-amplitude measurements," J. Acoust. Soc. Am. **35**, 1753–1761.

Fant, G., Kruckenberg, A., and Liljencrants, J. (**2000a**). "Acoustic-phonetic analysis of prominence in Swedish," in *Intonation: Analysis, Modeling and Technology*, edited by A. Botinis (Kluwer Academic, Dordrecht), pp. 55–86.

Fant, G., Kruckenberg, A., and Liljencrants, J. (**2000b**). "The source-filter frame of prominence," Phonetica **57**, 113–127.

Fant, G., and Lin, Q. (**1987**). "Glottal source–vocal tract acoustic interaction," STL-QPSR 1/1987, 13–27.

Flanagan, J. L. (**1957**). "Difference limen for formant amplitude," J. Speech Hear Disord. **22**, 205–212.

Fry, D. B. (**1955**). "Duration and intensity as physical correlates of linguistic stress," J. Acoust. Soc. Am. **27**, 765–768.

Gobl, C., and Ní Chasaide, A. (**1993**). "Contextual variation of the vowel voice source as a function of adjacent consonants," Lang Speech **36**, 303–330.

Gobl, C., and Ní Chasaide, A. (**1999**). "Techniques for analyzing the voice source," *Coarticulation: Theory, Data, and Techniques*, edited by W. J. Hardcastle and N. Hewlett (Cambridge University Press, Cambridge), pp. l00–321.

Gottfried, T. L., and Beddor, P. S. (**1988**). "Perception of temporal and spectral information in French vowels," Lang Speech **31**, 57–75.

Healy, E. W., and Bacon, S. P. (**2006**). "Measuring the critical band for speech," J. Acoust. Soc. Am. **119**, 1083–1091.

Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (**2001**). "Effects of consonantal environment on vowel formant patterns," J. Acoust. Soc. Am. **109**, 748–763.

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.

House, A., and Fairbanks, G. (**1953**). "The influence of consonant environment upon the secondary acoustical characteristics of vowels," J. Acoust. Soc. Am. **22**, 105–113.

Ito, M., Tsuchida, J., and Yano, M. (**2001**). "On the effectiveness of whole spectral shape for vowel perception," J. Acoust. Soc. Am. **110**, 1141–1149.

Jacewicz, E. (**2005**). "Listener sensitivity to variations in the relative amplitude of vowel formants," ARLO **6**, 118–124.

Jacewicz, E., Fox, R. A., and Salmons, J. (**2007a**). "Vowel space areas across dialects and gender," in *Proceedings of the XVIth International Congress of Phonetic Sciences*, edited by J. Trouvain and W. J. Barry, Saarbrücken, Germany, pp. 1465–1468.

Jacewicz, E., Fox, R. A., and Salmons, J. (**2007b**). "Vowel duration in three American English dialects," Am. Speech **82**, 367–385.

Katz, W. F., and Assmann, P. F. (**2001**). "Identification of children's and adults' vowels: Intrinsic fundamental frequency, fundamental frequency dynamics, and presence of voicing," J. Phonetics **29**, 23–51.

Kiefte, M., and Kluender, K. R. (**2005**). "The relative importance of spectral tilt in monophthongs and diphthongs," J. Acoust. Soc. Am. **117**, 1395–1404.

Lehiste, I., and Peterson, G. E. (**1959**). "Vowel amplitude and phonemic stress in American English," J. Acoust. Soc. Am. **31**, 428–435.

Lindqvist, J., and Pauli, S. (**1968**). "The role of relative spectrum levels in vowel perception," STL-QPSR 2-3/1970, pp. 12–15.

Mermelstein, P. (**1978**). "On the relationship between vowel and consonant identification when cued by the same acoustic information," Percept. Psychophys. **23**, 331–336.

Milenkovic, P. (**2003**). TF32 software program. University of Wisconsin, Madison, WI.

Miller, R. L. (**1953**). "Auditory tests with synthetic vowels," J. Acoust. Soc. Am. **25**, 114–121.

Nearey, T. M. (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.

Oppenheimer, A. V., and Willsky, A. S. (**1977**). *Signal and Systems* (Prentice Hall, New York).

Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.
Schwartz, J., and Escudier, P. (**1987**). "Does the human auditory system include large scale spectral integration?" in *The Psychophysics of Speech Perception*, edited by M. Schouten (Nijhoff, Dordrecht), pp. 284–292.
Sluijter, A. M. C., and Van Heuven, V. J. (**1996**). "Spectral balance as an acoustic correlate of linguistic stress," J. Acoust. Soc. Am. **100**, 2471–2485.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT, Cambridge, MA).
Titze, I. R. (**2000**). *Principles of Voice Production* (National Center for Voice and Speech, Iowa City, IA).
Whalen, D. H. (**1989**). "Vowel and consonant judgments are not independent when cued by the same information," Percept. Psychophys. **46**, 284–292.
Whalen, D. H., and Levitt, A. G. (**1995**). "The universality of intrinsic F0 of vowels," J. Phonetics **23**, 349–366.