A geometric representation of spectral and temporal vowel features: Quantification of vowel overlap in three linguistic varieties

Alicia Beckford Wassink[a)]
Department of Linguistics, University of Washington, Seattle Washington 98195

(Received 14 March 2005; revised 9 December 2005; accepted 2 January 2006)

A geometrical method for computing overlap between vowel distributions, the spectral overlap assessment metric (SOAM), is applied to an investigation of spectral (F1, F2) and temporal (duration) relations in three different types of systems: one claimed to exhibit primary quality (American English), one primary quantity (Jamaican Creole), and one about which no claims have been made (Jamaican English). Shapes, orientations, and proximities of pairs of vowel distributions involved in phonological oppositions are modeled using best-fit ellipses (in F1 $\times$ F2 space) and ellipsoids (F1 $\times$ F2 $\times$ duration). Overlap fractions computed for each pair suggest that spectral and temporal features interact differently in the three varieties and oppositions. Under a two-dimensional analysis, two of three American English oppositions show no overlap; the third shows partial overlap. All Jamaican Creole oppositions exhibit complete overlap when F1 and F2 alone are modeled, but no or partial overlap with incorporation of a factor for duration. Jamaican English three-dimensional overlap fractions resemble two-dimensional results for American English. A multidimensional analysis tool such as SOAM appears to provide a more objective basis for simultaneously investigating spectral and temporal relations within vowel systems. Normalization methods and the SOAM method are described in an extended appendix. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2168414]

PACS number(s): 43.70.Jt [AL] Pages: 2334–2350

I. INTRODUCTION

This paper deals with one classification problem resulting from systematic sources of variability affecting acoustic vowel features—the classification of languages as length-versus quality-contrasting. A broad range of linguistic and nonlinguistic effects would seem to render impossible the identification of numerical cutoffs (durational ranges and concomitant formant averages) which might allow for direct categorization of vowel systems based upon whether temporal or spectral features play a primary or secondary role. In this paper, spectral/temporal relations are investigated in three linguistic varieties that have been claimed to differ with regard to the role of these features in phonological contrast. We consider how the *overlap fraction* yielded by the spectral overlap assessment metric, or SOAM (described below), may help to address the question of whether or how vowel quantity contributes to the distinction (in production) between phonologically tense and lax vowels in three linguistic varieties: Jamaican Creole (JC hereafter), Jamaican English (JE), and American English (Pacific Northwest variety; PNWEng hereafter).

Broad variation exists in the possible distribution and combination of acoustic features (e.g., vowel formants and segmental durations) resulting from language-specific coarticulatory effects (e.g., vowel-to-vowel coarticulation: Macken, 1980; Strange and Bohn, 1998; Perkell and Matthies, 1992), nonlinguistic influences such as physiological differences (e.g., vocal tract length differences between sexes, adults and children: Nordström and Lindblom, 1975), and sociolinguistic factors (e.g., dialectal variation and linguistic change: Labov, 1994; Hagiwara, 1997). This variation, though many of its causes yield systematic effects, has not yet been integrated into a unified theory of influences on acoustic vowel features. Researchers still lack acoustic characterizations of many vowel systems. Also lacking are models of vowel systems that enable the examination of the combined influences of sets of effects on vowel features. This paper particularly addresses the latter issue. We apply SOAM to a cross-linguistic investigation of three vowel systems. SOAM allows the simultaneous comparison of spectral and temporal vowel features. Statistical properties of normalized estimates of the first two acoustic vowel formants (F1, F2) and segmental duration provide the basis for ellipse and ellipsoidal models of vowels in phonological oppositions. The shapes of the two vowel distributions and their proximity in two- and three-dimensional space (i.e., F1 $\times$ F2 and F1 $\times$ F2 $\times$ duration) are calculated. Graphical representation then allows for the visual evaluation of algorithm results.

II. BACKGROUND: SPECTRAL/TEMPORAL INTERACTIONS

A good deal of research attention has been devoted to the typological classification of vowel systems. Increasingly, acoustic phonetic analyses are undertaken to support phonological classifications (i.e., into types of contrast) based upon spectral and temporal vowel features. *Primary quality* has been used to classify languages in which spectral contrasts,

[a)]Electronic mail: wassink@u.washington.edu

typically carried in the first two formants (F1 and F2), provide the basis for phonological contrast. The designations *primary* and *secondary quantity* are used to classify languages within which temporal features (typically segmental duration) play a crucial role. In the former, phonological contrast in a vowel system is based upon "robust" differences in the segmental durations of vowels in phonological oppositions. In the latter, phonological contrast is based upon robust differences in the quality of vowels in phonological oppositions while systematic differences in vowel length are also observed. One concern has been to understand what relations between vowel quality and quantity features must obtain in order to establish a basis for distinguishing between the two temporal designations. Certain inconsistencies occur in the literature regarding the definitions given to these three classifications. Namely, systematic differences in vowel length may result from several sources, some language-general and others language-specific. However, lengthening due to different sources is often not distinguished. For example, systematic differences in vowel length may result from (1) a language-specific phonetic tendency to "enhance" quality (i.e., phonemic) distinctions as in the tense/lax contrast in English (Lehiste, 1970) or (2) language-general (phonetic) processes, such as preconsonantal "lengthening," which supports the perception of consonant voicing. As an example of the former, the English tense/lax ratio is underlyingly 1.2:1—i.e., the average duration of short vowels is 83% of that of the long vowels. This represents a language-specific tendency related to phonological length. This ratio increases to 1.5:1 (67%) when the vowel precedes a voiced obstruent—a language-general tendency unrelated to phonological contrast (Lehiste, 1970). Kluender *et al.* (1988) suggest that true length contrasts (i.e., in primary quantity languages) may be so auditorily distinctive "that the contrast-enhancing effects of vowel-length variation are simply unnecessary." Thus, difficulty arises when classifications are used that do not distinguish between the phonetic outcomes of phonological length and processes of phonetic lengthening unrelated to phonological length. Additionally, classifications are often made without reference to spectral differences, e.g., whether a systematic difference in vowel duration is accompanied by a concomitant absence of quality overlap.[1] Crucially, as has been pointed out by Lehiste (1970, p. 35), a linguistic rationale (either auditory or acoustic) is lacking for determining how large a temporal contrast must be before it may be considered *primary*. Furthermore, it is even less clear what type of spectral distinction must obtain simultaneously with a given durational contrast. These two problems may be illustrated with examples related to perspectives on the role of *overlap in spectral features* and *overlap in temporal features*.

In the spectral domain, the lack of experimental research that uses a multidimensional approach in investigating the nature of spectral and temporal distributional patterns has impeded scholarly understanding of prototypical versus "ambiguous" vowel qualities. Hillenbrand *et al.* (1995) carry out a replication of the most widely cited study of the acoustics and perception of vowels (Peterson and Barney, 1952; P&B hereafter). They analyze spectral (f0, F1–F3) and temporal (segmental duration) information for speakers of a Midwestern dialect of American English and find differences between mean segmental durations for males' as compared to both children's and females' productions (males' durations being shortest). These results are statistically significant in *post hoc* comparisons, but the authors are unable to provide an explanation for this effect (Hillenbrand *et al.*, 1995, p. 3102). Their intention was to obtain, for General American English, a set of data comparable to that original study but more precisely controlled in terms of several parameters, including regular interval measures to trace the trajectory of the vowel, inclusion of measures for segmental duration, and controls for dialect of both speaker and listener dialect.

One key difficulty that reduces comparability between P&B and Hillenbrand *et al.* lies in the fact that the latter collected data from speakers of a Midwestern dialect. The authors were not unaware of dialect differences, which is clear in the fact that they attempted to screen for certain differences, excluding those who merged /a, ɔ/, or whose speech evidenced a "departure from general American English" (Hillenbrand *et al.*, 1995, p. 3100). However, although it is true that this dialect region is well known for not (yet) participating in the widespread merger of /a, ɔ/, it is not clear how meaningful it is to say that they also omitted speakers native to the area who were not speakers of "General American English."[2] There are differences between the productions of the final Midwestern sample and the P&B sample that reflect further spectral differences for which the authors cannot account. For example, sociophonetic descriptions of the vowel system for this region indicate marked raising of /æ/ that distinguishes the Midwest from other regions (Hagiwara, 1997; Labov, 1994), with women and children leading in this change. (Also, since the merger of /a, ɔ/ is widespread among dialects, the lack of merger already places this region outside the norm with respect to a "general American" feature). The Midwestern speakers' systems do indeed show raising of /æ/. For example, for the Midwestern women (the group with the *greatest* differences in mean F1, F2), Hillenbrand *et al.* report the following category means for /æ/ and /ɛ/: in F1—669 and 731 Hz (a difference of 62 Hz) and in F2—2349 and 2058 Hz (a difference of 291 Hz).[3] These differences in F1 and F2 are smaller than the mean differences reported for these formants in the earlier Peterson and Barney study (1952, p. 183), suggesting greater proximity between the /æ/ and /ɛ/ vowel categories of Midwestern speakers.[4]

This proximity appears to be relevant to the perception study results as well. Hillenbrand *et al.* report that listeners showed greatest confusion for /æ, ɛ, ɔ/. This is possibly the result of spectral overlap between some tokens of /æ/-/ɛ/ (which is perhaps most marked in males' tokens, as will be discussed for this complex situation, below). For /ɔ/, it seems likely that some speakers did lower or merge the quality of this vowel toward [a], so that while /a/ tokens remain unambiguous, lowered tokens of /ɔ/ were confusible with /a/. However, spectral overlap in the presence of systematic and auditorally salient temporal differences could still cue a distinction. In fact, there is evidence that temporal features aid in distinction.

First, Hillenbrand *et al.* report that a quadratic discriminant analysis was most successful when duration was included in the set of parameters used to classify tokens. Second, reexamination of production results suggests that temporal features indeed play a crucial role. Close examination of their Fig. 9, which shows plots of F3-F2 (mels) by F0-F1 (mels) of formant patterns sampled at 20% and 80% of the duration of the vowel, is quite telling. This figure shows longer curves for /æ/ than for /ɛ/ (for all groups) and for women than men.[5] Longer curves indicate greater spectral change over time, a temporal feature of vowel production. Specifically, women's /æ/ exhibited greatest change in the dimension of vowel height.

The duration difference between groups that Hillenbrand *et al.* found problematic may thus be understood as follows: females' longer durations are consistent with their greater advancement in the change by which /æ/ is produced as [iɛ]. This vowel is (for women and children) diphthongal, while still relatively phonetically static for males. Males' /æ/ and /ɛ/ are spectrally and temporally most extensively overlapping of the three Midwestern groups. Plots of midpoint values alone (as the authors acknowledge) mask crucial temporal distinctions. A representation of spectral and temporal features together is most helpful. Thus, knowledge of the phonetic extent of dialect differences is important in designing a replication and in interpreting study results. However, the more crucial point for the present study is that this difference suggests systematic dialectal differences in the relations between spectral and temporal cues to a phonological contrast.

Secondly, cross-linguistic comparisons of vowel length contrasts provide an illustration of how the lack of experimental research that uses a multidimensional approach has impeded scholarly understanding of the role of durational differences in phonological contrast. Japanese, Luganda, Icelandic, and Thai have been characterized as unambiguous examples of primary quantity languages on the basis of the typical durational differences between phonologically long and short vowels (conventionally expressed as a ratio of the duration of the phonologically long member of a pair of vowels to the phonologically short one). That is, vowel durations for contrasts in these languages typically fall above 1.6:1, while English and German are characterized as having secondary quantity (Lehiste, 1970). This length ratio is often accompanied by an impressionistic determination about spectral contrast (full overlap, partial overlap, or no overlap), which is unquantified in terms of phonetic characteristics of the vowels. Measures of F1 and F2 for the distributions are not typically presented. The conventional means of presentation is summarized in Table I.

Both of these problems demonstrate the need for a mathematical procedure that provides a means of simultaneously representing and evaluating the roles of spectral and temporal acoustic phonetic vowel features. Such a measure would have several advantages for between-language and within-language comparisons. Among these, it would facilitate categorization of primary and secondary quantity languages and provide a means of determining, within a language, whether a systematic difference in vowel duration is accompanied by a concomitant absence of quality overlap. An experiment was carried out in which spectral/temporal relations were tested for long-short vowel pairs in three language varieties: a Pacific Northwest variety of American English (referred to hereafter as PNWEng), Jamaican Creole (JC), and Jamaican English (JE). One longstanding question in the phonology of Jamaican varieties has been whether spectral contrasts or temporal ones are the basis for phonological vowel contrast (Wassink, 2001). Jamaican English (the term used in this paper to refer the Jamaican "acrolect") and Jamaican Creole (the Jamaican "basilect") represent distant ends of the post-Creole linguistic continuum in Jamaica. These two varieties are of interest because they differ systematically in the distribution of spectral (F1 and F2) and temporal (duration) features in their vowel inventories (Wassink, 1999, 2001), such that Jamaican Creole has been characterized as retaining more of the phonetic features of its West African adstrates (which utilized phonemic vowel length), and Jamaican English of its English adstrates (several of which utilized secondary quantity). Thus, researchers are interested in understanding the relative contributions of the phonologically different input languages to the Jamaican continuum. The experiment that follows reports the results of the within-language characterizations, followed by a between language-comparison which allows some observations to be made about the application of the terms primary quantity, primary quality, and secondary quantity to these varieties.

TABLE I. Languages reported to utilize vowel length (from Crothers, 1978).

Language	Long: short ratio	Spectral overlap	Primary/secondary quantity
Japanese	2.5:1	Full	Primary
Thai	1.9:1	Full	Primary
Icelandic	2.0:1	Partial	Primary
Luganda	2.5:1	Partial	Primary
German	1.5:1	Partial	Secondary
English	1.2:1	Partial	Secondary
(arbitrary cutoff)	1.6:1		

III. EXPERIMENT

A. Speakers

The Pacific Northwest sample included six female speakers born and raised in Washington state. The Jamaican English sample included nine speakers (five male and four female) native to the Kingston Corporate Area (St. Andrew Parish), and the Jamaican Creole sample included ten speakers (five male and five female) native to the rural parish of St. Thomas. None reported a knowledge of hearing difficulties.

B. Materials

Target data for this study include six vowels representing three phonological oppositions, namely /iː/ *beat* ~ /i/ *bit*, /aː/ *bought*~ /a/ *bat*, /uː/ *boot*~ /u/ *book*. Here, the symbols /iː,i,aː,a,uː,u/ are used to provide a consistent, convenient means of representing the word classes indicated and to signal the phonological specification of the lengths of the vowels in each and not the phonetic quality of vowels in any of the language varieties under study. The phonetic qualities differ, as will be discussed below, and phonetic qualities will be represented using square brackets "[ ]". Vowels from these six classes were elicited for all speakers of the three linguistic varieties. Vowels representing all other monophthongal qualities were also elicited, but were used solely for purposes of normalization (see below). Tokens were collected in word-list elicitation tasks in language-appropriate carrier frames: "Write ___ please" for Jamaican English, "Unu rait ___ pon i" for Jamaican Creole, and "Say ___ again" for PNWEng. The carrier frames and token numbers vary for PNWEng and the two Jamaican varieties because data were originally elicited for two different investigations. For all language varieties, vowels were elicited in real-word monosyllables of the shape /b,d,k_p,b,t,d,k,g/, as well as /h_d/ for PNWEng.[6] Three hundred ninety-one tokens were analyzed for PNWEng (6 speakers × 6 vowels × 5 repetitions × a subset of the 10 contexts), 2362 tokens for Jamaican English (9 speakers × 6 vowels × ≤5 repetitions × 10 contexts), and 2437 tokens for Jamaican Creole (10 speakers × 6 vowels × ≤5 repetitions × 10 contexts).

All Jamaican English and Jamaican Creole tokens were collected in a field research setting, which facilitated the collection of data for two linguistic varieties not previously subjected to experimental phonetic study. However, environmental conditions had to be carefully controlled to ensure a high-fidelity signal for acoustic analysis. To this end, recordings were made with a close-talking AIWA lapelle microphone and Sony TCD-D8 digital audio tape recorder. Recordings of the PNWEng speakers were made in the Linguistic Phonetics Laboratory at the University of Washington, using a TASCAM DAP-11 and Nakamichi (CM-100) cardioid microphone. The speech signal was recorded at a 44.1 kHz sampling rate and digitally transferred for analysis in Praat software. Tokens were downsampled to 11.025 kHz for analysis.

The first three vowel formants (F1, F2, F3) were measured at the temporal midpoint for each vowel token. Vowel onset was determined conservatively from the waveform as the beginning of the first clear pitch pulse associated with modal voicing following the release of the preceding consonant. Vowel offset was defined as the end of the final clear pitch pulse before the closure of the following consonant. Vowel duration was automatically calculated by the software (Praat) as the difference between vowel onset and offset. F1, F2, and F3 were obtained using superimposed FFT and LPC spectra from a 25.6-ms window. FFT analyses used a 59-Hz filter; LPC spectra used 14 predictor coefficients. In an initial pass, formant measures were taken using an automatic peak-picking routine based on the LPC spectra. Then, all automatically calculated frequencies were checked in the combined FFT/LPC display. Questionable values were corrected or confirmed using a narrow-band spectrogram. To check the reliability of the measures, approximately 20% of the dataset was independently analyzed by a second phonetician. F1, F2, and F3 were obtained using Praat's autocorrelation-based procedure and the values generated by this procedure were manually inspected using Praat's formant trace function. Intermeasurer agreement was assessed at 92%.

Individual F1 and F2 values for all monophthongs, including (/iː, ɪ, ɛ, æ, (ɔ,) aː, uː, ʊ/[7] were normalized using an adaptation of Nearey's (1977) logarithmic uniform scaling technique (Wassink, 1999). This procedure is detailed in Appendix A. This method was preferred because it suppresses variation due to sex-related physiological differences, but appears to retain variation such as may be introduced by differences in linguistic quality (Adank *et al.*, 2004; Hindle, 1978). This vowel-extrinsic procedure achieves within-speaker normalization using information distributed over the formant frequencies for all monophthongs in that speaker's system.

C. Data analysis procedures: The spectral overlap assessment metric (SOAM)

Mathematical determination of the relationships between two vowel distributions is accomplished using one of two algorithms: for spectral features alone (SOAM 2-D, two-dimensional model) or between spectral and temporal vowel features (SOAM 3-D, three-dimensional model). Both algorithms take as input normalized values for individual vowel tokens. The two- and three-dimensional algorithms are summarized in Appendix B.

IV. RESULTS

In this section, the results of the cross-linguistic investigation of JE, JC, and PNWEng vowel features are presented. First, we examine the results yielded by a conventional F1 × F2 representation, using a Plot Formants-style representation (Ladefoged, 1996). Next, the results of overlap calculations generated in SOAM 2-D (F1 × F2) and SOAM 3-D (F1 × F2 × duration) are presented.

Table II presents the mean and standard deviation information for F1 and F2 (in Hz) for each vowel for each variety. Figures 1(a)–1(c) show conventional inverted F1 × F2 scatterplots of the normalized formant values calculated using these means. In Fig. 1(d), the raw data (in Hz) for PNWEng are presented, encircled by best-fit ellipses generated in PlotFormants (Ladefoged, 1996). The main benefit of the conventional method, i.e., of reporting mean and deviation information along with a supplemental inverted F1 × F2 plot of vowel data, is that it enables one to generally locate vowel qualities in particular regions of acoustic vowel space. For example, comparison of the scatter plots reveals a different overall shape for the Jamaican vowel spaces relative to that of PNWEng. While the high front pair /iː, i/ lie in roughly the same locations, the low vowels /aː, a/ are central in Jamaican varieties (the apex of a V-shaped vowel space), while the PNWEng low vowels occupy different regions entirely in that space: /a/ *bat* is phonetically [æ], a low front vowel, while /aː/ *bought* has a low back place of articulation.[8] Similarly, the high back pair /uː, u/ occupy different regions. This, however, is most easily determined from inspection of the mean F2 data in Table II. Mean F2 values for PNWEng /uː, u/ (1653.96 and 1480.17 Hz, respectively) are higher than

TABLE II. Mean and standard deviations for F1, F2, and duration for Jamaican English, Jamaican Creole, and PNWEng /i:, i, a:, a, u:, u/ (n =number of tokens).

	n=2362	F1 (Hz)	Standard deviation	F2 (Hz)	Standard deviation	Duration (s)	Standard deviation
				Jamaican English			
i:	493	305.04	*53.36*	2440.43	*237.29*	0.1271	*0.0308*
i	525	372.52	*62.05*	2219.33	*203.78*	0.0751	*0.0291*
a:	130	691.78	*87.24*	1179.93	*135.61*	0.1803	*0.0232*
a	527	755.97	*87.07*	1525.81	*197.90*	0.1200	*0.0297*
u:	588	338.31	58.46	986.00	*396.64*	0.1303	*0.0365*
u	99	365.99	*81.32*	858.42	*190.48*	0.0867	*0.0542*
				Jamaican Creole			
	n=2437	F1 (Hz)	Standard deviation	F2 (Hz)	Standard deviation	Duration (s)	Standard deviation
i:	511	309.77	*77.40*	2463.66	*251.38*	0.1386	*0.0317*
i	551	379.36	*63.08*	2277.18	*206.47*	0.0748	*0.0228*
a:	136	688.37	*95.52*	1231.58	*162.71*	0.1868	*0.0355*
a	565	728.56	*110.32*	1493.20	*198.78*	0.1174	*0.0278*
u:	567	365.94	*58.96*	834.33	*200.98*	0.2294	*1.9129*
u	107	421.12	*56.63*	892.17	*186.15*	0.0624	*0.0190*
				American English (PNW variety)			
	n=391	F1 (Hz)	Standard deviation	F2 (Hz)	Standard deviation	Duration (s)	Standard deviation
i:	56	382.75	*53.50*	2614.63	*365.08*	0.0843	*0.0231*
i	64	465.07	*50.76*	2177.36	*169.13*	0.1003	*0.0401*
a:	101	818.59	*128.91*	1366.01	*117.99*	0.1047	*0.0397*
a	66	837.61	*146.76*	1893.62	*168.13*	0.1015	*0.0343*
u:	68	400.30	*48.74*	1653.96	*345.66*	0.1399	*0.0425*
u	36	531.72	*62.54*	1480.17	*285.53*	0.1629	*0.0481*

the corresponding Jamaican English values (986.00 and 858.42 Hz, respectively) and the corresponding Jamaican Creole values (834.33 and 892.17 Hz, respectively). Thus, PNWEng high back vowels lie further to the front of the acoustic vowel space, consistent with reports by Hillenbrand *et al.* (1995) for the American Midwest. While it might seem reasonable to impute the difference in the overall dispersion of acoustic vowel space to the presence in the Jamaican samples of male speakers (whose high back vowels tend to have lower F1 and F2 frequencies than those of adult females), differences of the magnitude observed (for JC versus PNWEng /u:/=819.63 Hz; for /u/=588.00 Hz) are far greater than would be expected due to sex-related differences (Hindle, 1978). However, it is important to test this by examining the normalized values. Table III presents mean and standard deviation information for the log-mean normalized F1 and F2 values of all three language varieties. Log-mean normalized values for /u:/ ~ /u/ (averaged across all speakers within the language group) for Jamaican Creole are −0.116(log F1) and −0.262 (log F2); for PNWEng −0.289(log F1) and −0.157(log F2). These values are more difficult to interpret than values in raw Hz. They represent mean token deviations from the center of each language system's vowel space. Therefore, if both spaces are centered at (0,0), we see that the PNWEng /u:/ data tend to lie further front along the front-back dimension in that system: normalized F2 values tend to be smaller than for the Jamaican Creole /u:/. Normalized data provide the basis for the comparisons that follow, which examine, first, the extent of overlap of each vowel onto its acoustically adjacent partner in two and then in three dimensions.

Standard deviation information is of some utility in determining the proximity of vowel distributions. We may observe, for example, that the standard deviations calculated for Jamaican English /a:,a/ are sufficient to close the distance between the F1 means for these vowels, suggesting that these distributions are quite proximal, and possibly overlapping, at least in this dimension. We may independently observe that the standard deviations for this same pair of vowels indicate that some scatter lying near the edges (=2σ of the mean) of each distribution approximate data in the other distribution by up to ~135 Hz. Such comparisons are limited, however, because it is possible that two distributions that are proximal in one spectral dimension may yet be spectrally distinct, because they may show significant separation in the other dimension. The results of the spectral overlap assessment will therefore be enlightening.

Overlap in two dimensions. The leftmost columns listed for each language in Table IV provide the *overlap fractions* yielded by SOAM 2-D for each phonologically contrasting vowel pair. An overlap average, calculated across all pairs within a language variety, is provided in the final row, for convenience in comparing the three linguistic varieties. This value must be considered with some caution, however, because it gives only a global sense of systems that show internal variability. As the discussion above demonstrated for Midwestern American English *bag*, intrinsic and extrinsic factors (such as those operating in linguistic change) can

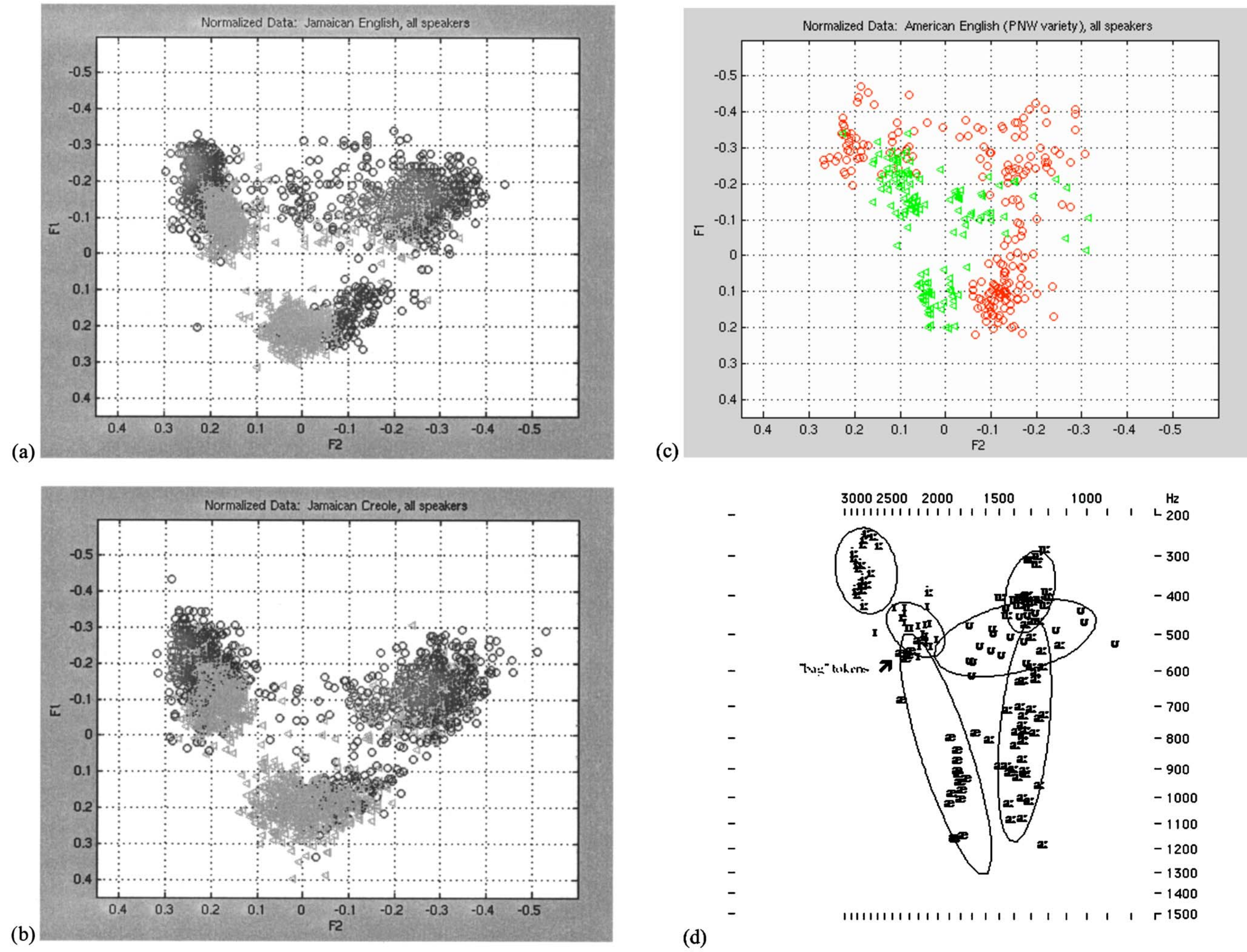

FIG. 1. (Color online) Scatterplots illustrating the log-mean normalized values for vowels in three tense-lax oppositions for three language varieties: (a) Jamaican English, (b) Jamaican Creole, and (c) PNWEng. Tense vowels /iː, aː, uː/ are represented with "∘," lax vowels /i, a, u/ with "◁". (d) is a second representation of the PNWEng data in (c), generated from raw F1, F2 values (in Hz) by Plot Formants software, "stats" function. Note vowel raising of several tokens of /a/ (phonetically [as]: specifically, those in the word "bag").

result in the displacement of one member of a pair. While it might be convenient to be able to accord classifications such as primary quantity or primary quality on the basis of the overall system, it cannot be expected that all pairs exist in the same phonetic relationships. For this reason, the discussion will treat each pair separately. For PNWEng, the vowel pair /iː∼i/ shows the greatest overlap, 46%. This result for the Pacific Northwest speakers is consistent with a visual inspection of Hillenbrand *et al.*'s findings for a group of Midwest American English speakers. The pair /uː∼u/ shows slightly less overlap, 34%, and /aː∼a/ the least, 15%. (Recall that /a/ represents the *cat* word class.) This vowel shows some raising in the Pacific Northwest vowel space as in the Midwestern one, however, raising is apparent for just one lexical item, namely "*bag*," which is widely realized by PNW speakers as [beg] or [beɪg]. When *bag* tokens are removed, the distribution mean for this class is still close enough to [æ] that the pattern observed above remains: that is, PNWEng /aː∼a/ show the least distributional crowding across all three varieties in a two-dimensional representation (reflecting linguistic changes that have taken place in varieties of American, but not Jamaican, English).

Overlap fraction values for Jamaican English appear to be similar to those for PNWEng. Percent overlap for the pair /iː∼i/ is 36% (10 percentage points less than PNWEng) and 23% for /aː∼a/ (3 points less). It is in the /uː∼u/ pair that Jamaican English values are most different from those of PNWEng: the overlap fraction for the former language variety is 75% (against 34% for PNWEng, a difference of 41 percentage points). Thus, whereas the high front and low pairs appear to be less coextensive in F1×F2 acoustic space, the high back pair shows greater overlap.

The Jamaican Creole oppositions received substantially higher overlap fraction values than either Jamaican English or PNWEng. All vowel pairs overlap in F1×F2 space by more than 55% of their total area: /iː∼i/—75%, /aː∼a/—55%, /uː∼u/—86%.

Overlap in three dimensions. As described below in Appendix B, SOAM 3-D includes a second least-squares fit calculation to transform the data in a manner that reflects their orientation in $x^{*}\times$ duration space. By thus incorporating duration into the model of our best-fit ellipsoids, we may recalculate overlap fractions and observe how segmental duration information influences the extent of protrusion of ad-

TABLE III. Log-mean normalized F1 and F2 values for three language varieties. (Tokens collected for the word bag have been omitted for the PNWEng speakers.)

Vowel	Jamaican English		Jamaican Creole		PNW English	
i:	$n=533$		$n=579$		$n=56$	
log F1 (*std dev*)	−0.198	*0.057*	−0.191	*0.076*	−0.311	*0.063*
log F2 (*std dev*)	0.214	*0.029*	0.218	*0.039*	0.176	*0.061*
i	$n=586$		$n=619$		$n=62$	
log F1 (*std dev*)	−0.096	*0.048*	−0.065	*0.106*	−0.209	*0.060*
log F2 (*std dev*)	0.169	*0.036*	0.150	*0.103*	0.102	*0.034*
a:	$n=132$		$n=136$		$n=97$	
log F1 (*stddev*)	0.163	*0.047*	0.166	*0.051*	0.085	*0.080*
log F2 (*std dev*)	−0.101	*0.048*	−0.083	*0.061*	−0.128	*0.036*
a	$n=559$		$n=601$		$n=47$	
log F1 (*std dev*)	0.205	*0.035*	0.191	*0.052*	0.117	*0.043*
log F2 (*std dev*)	0.009	*0.037*	−0.004	*0.058*	0.021	*0.029*
u:	$n=618$		$n=590$		$n=58$	
log F1 (*std dev*)	−0.148	*0.066*	−0.116	*0.066*	−0.289	*0.070*
log F2 (*std dev*)	−0.221	*0.109*	−0.262	*0.091*	−0.157	*0.084*
u	$n=99$		$n=107$		$n=36$	
log F1 (*std dev*)	−0.065	*0.050*	−0.073	*0.061*	−0.138	*0.058*
log F2 (*std dev*)	−0.209	*0.070*	−0.249	*0.076*	−0.102	*0.093*

jacent vowels. In general, we expect to find an inverse relationship between durational separation and overlap fraction. That is, as mean duration differences for a given vowel pair increase, the overlap fraction will decrease. The rightmost column presented below each language variety in Table IV provides the results for the three-dimensional overlap calculations (SOAM 3-D).

We may first observe that overall, overlap for all three language varieties decreases in the three-dimensional model. The overlap fraction decreases the least for PNWEng: by 19 percentage points for /i:∼i/, 26 for /a:∼a/, and 19 for /u:∼u/. The decrease, however, does indicate that incorporation of duration into the model pulls the members of each pair further apart. All oppositions now overlap by 27% or less (14% on average).

Overlap fractions for Jamaican English under the three-dimensional representation show a greater change than do those of PNWEng. Amount of overlap decreases are 27 percentage points for /i:∼i/, 21 for /a:∼a/, and 57 for the pair that showed greatest overlap in the two-dimensional model, /u:∼u/. These changes bring the Jamaican English oppositions into relationships where each pair overlaps, on average by only 10%.

Jamaican Creole overlap fractions show the greatest change from the two- to the three-dimensional model. Degree of overlap decreases are 58 percentage points for /i:∼i/, 32 for /a:∼a/, and 39 for /u:∼u/. These changes bring the Jamaican Creole oppositions into relationships where each pair protrudes, on average, by 29%.

Because the transformation of a two-dimensional to a three-dimensional overlap outcome is not linear, it is inappropriate to subtract the three-dimensional from the two-dimensional overlap fractions to determine the contribution of duration. At present, we must accomplish this impressionistically by manually rotating the three-dimensional graph and visually observing the extent of overlap from several angles. This is shown in Fig. 2, which provides three views of each /a:∼a/ model ellipsoid pair generated for each of the test language varieties. Unfilled circles represent datapoints contained within each model ellipsoid. For each language variety, proximity between pairs of model ellipsoids may be visually inspected from any number of angles. The top panel in each set of figures displays the default three-dimensional output of VOIS3D software. The second and third panels show rotations of the data as projected onto the F1×F2 and the F2×duration planes, respectively. It may be observed from the second panel in each set [Figs. 2(a), 2(b), 2(c)] that coordinates for /a:/ and /a/ coextend along a similar range of values in F1 for all three varieties: distributions for /a:/ are typically larger, while /a/ typically has a more compact distribution in F1. The PNWEng ellipsoids differ from the Jamaican ones in two respects. First, as may be observed in Fig. 2(c) model distributions do not overlap along the F2 dimension in PNWEng, consistent with the overlap percentages reported above. Second, it may be observed from Fig. 2(c) that duration values exhibit a similar range for both the /a:/ and /a/ word classes.

Comparison of the two- and three-dimensional overlap fractions supplied by SOAM 2-D and SOAM 3-D leads us to conclude that for the language varieties and dimensions investigated, duration is most critical for production contrasts in Jamaican Creole. That is, duration has a more notable reducing effect on overlap in Jamaican Creole than in the other varieties, and within this variety an effect of similar size obtains in all oppositions. Interestingly, although Jamaican English vowels already show little overlap in F1×F2 for two contrasts, overlap fractions decrease most substantially

TABLE IV. Overlap fractions calculated in two (by SOAM 2-D) and three (by SOAM 3-D) dimensions, F1×F2 and F1×F2×duration, respectively, for three language varieties.

	Jamaican English		Jamaican Creole		PNW English	
	F1×F2	F1×F2×Dur	F1×F2	F1×F2×Dur	F1×F2	F1×F2×Dur
i:∼i	36%	9%	75%	17%	46%	27%
a:∼a	23%	2%	55%	23%	15%	0%
u:∼u	75%	18%	86%	47%	34%	15%
Average	45%	10%	72%	29%	35%	14%

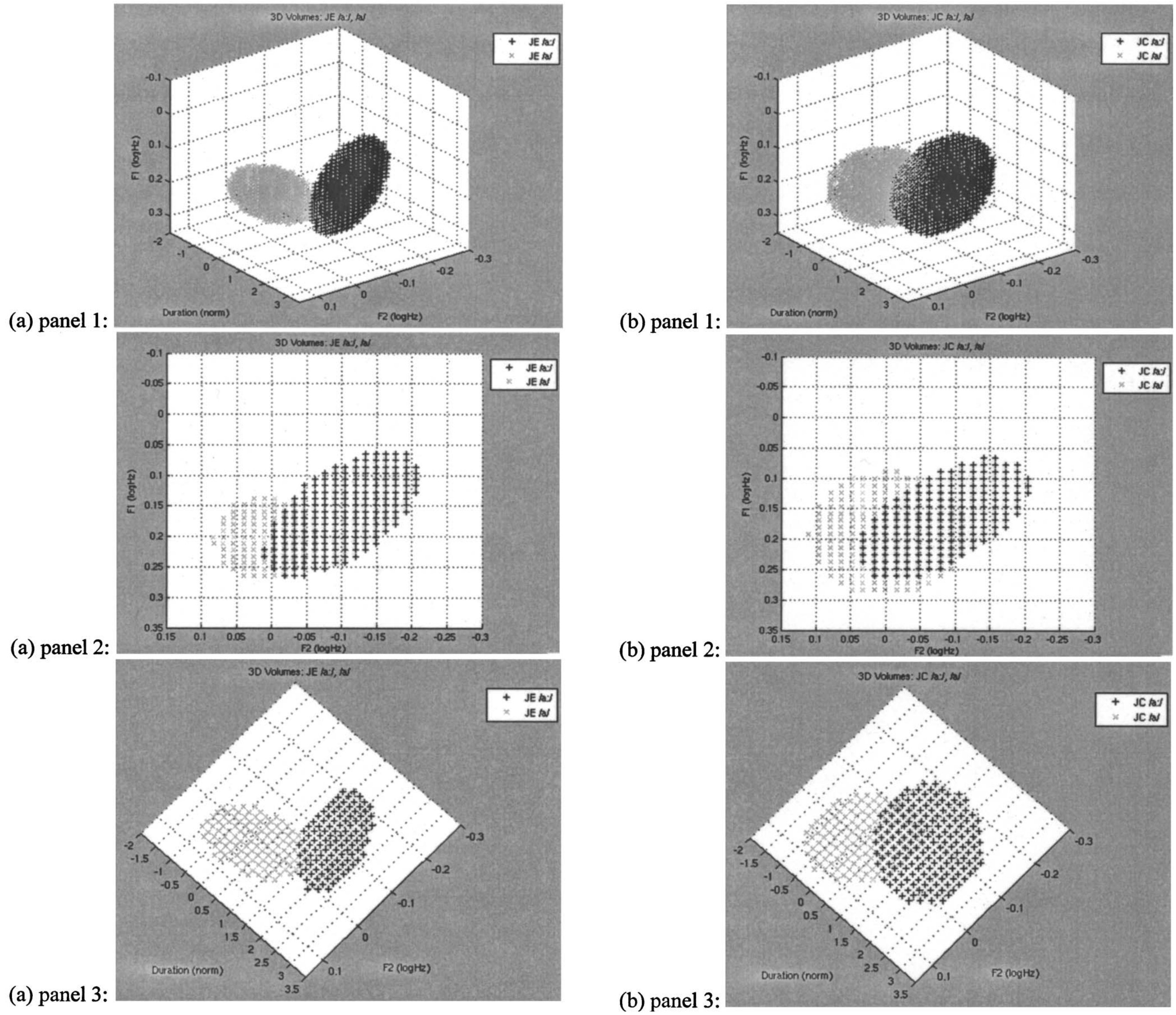

FIG. 2. Three-dimensional representations of /a: ~ a/ for (a) Jamaican English, (b) Jamaican Creole and (c) PNWEng data, generated using VOIS3D software. Panels 2 and 3 represent manual rotations of the 3-D volume graphs in panel 1 oriented in the F1 × F2 (panel 2) and F2 × duration (panel 3) planes.

for the high back pair (the magnitude of change in the overlap fraction from a two- to a three-dimensional model being most similar to Jamaican Creole). For PNWEng the overlap between volumes decreases, but less substantially—volumes that already shared only 26–47% of their volumes show an additional decrease of 14 percentage points. Thus, the most complex relation obtains for Jamaican English. It appears that duration contributes to a distinction already primarily carried along the spectral dimension (for high front and low central contrasts). In the high back pair, this enhancement is quite large. Listener identification might be tested to determine whether independent evidence may be found, in the perceptual domain, to suggest whether spectral or temporal features, or a complex relation between the two, serves as the primary basis for distinction. It appears to be important to investigate whether one parameter can be more crucial to contrast for one particular opposition than for others.

V. DISCUSSION

The findings presented above are illuminating in relation to the typological question of whether we might classify the three varieties investigated as showing primary quantity, secondary quantity, or primary quality. Ideally, overlap patterns will be investigated for a wide range of languages, and these findings are used to see how overlap patterns cluster cross-linguistically. This is preferred over the designation of arbitrary cutoffs. However, it may be instructive to assign a more principled (though still somewhat arbitrary) cutoff for purposes of informal comparison, as such data are not presently available. We may employ terms currently used in the literature for describing impressionistic relations in the quality dimension only, applying these instead to our multidimensional representations: complete overlap, partial overlap, and no overlap. We may provisionally define a fairly conservative cutoff for partial overlap as protrusion in less than half

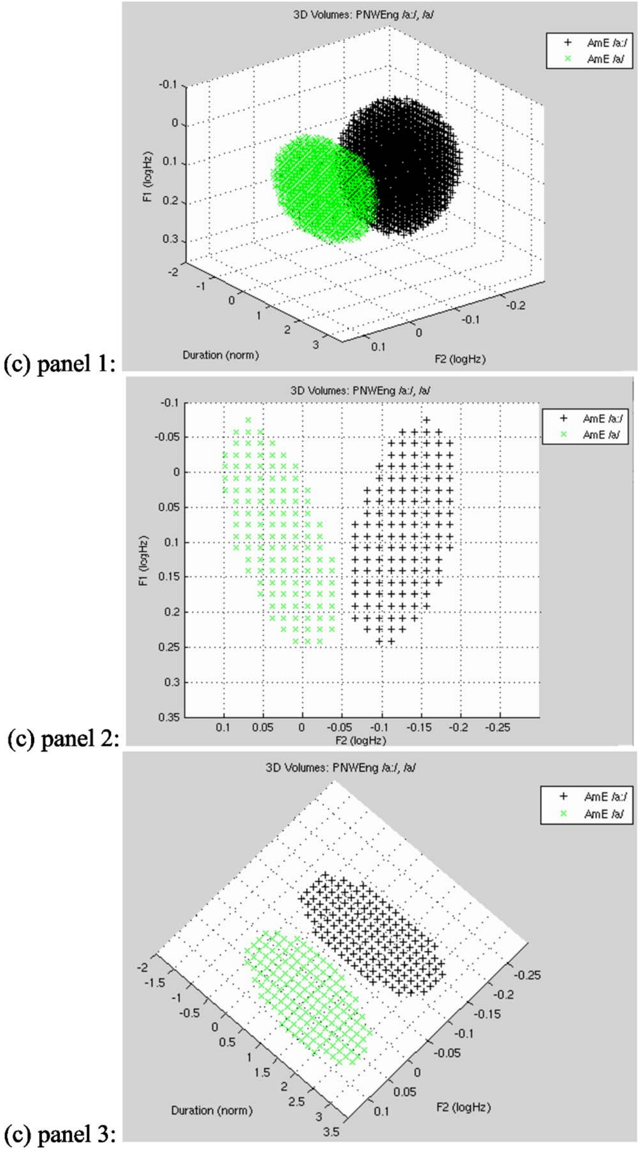

FIG. 2. (*Continued*).

of the volume of the shortest vowel in an acoustically adjacent pair (e,g., an overlap fraction of 20%–40%), one for no overlap as 0%–20%, and a third for complete overlap as ⩾40%. We may observe that once a factor is included for duration, all oppositions for all three language varieties must be classed as showing partial or no overlap, with the exception of Jamaican Creole /u: ∼ u/, which is only slightly above the cutoff for partial overlap (47%). The Jamaican English oppositions and two of three PNWEng ones must be classed as showing no spectral overlap, while the third PNWEng opposition (/i: ∼ i/) falls just above the cutoff for partial overlap. The Jamaican Creole oppositions would be classed as no overlap (/i: ∼ i/), partial overlap (/a: ∼ a/), and complete overlap (/u: ∼ u/), respectively.

When we consider the magnitude of change within each language variety from a two- to a three-dimensional best-fit model, we find that in the initial two-dimensional analysis, two of the three PNWEng oppositions already show no overlap (low central and high back), and the third may be classified as just above the cutoff for partial overlap. All Jamaican Creole oppositions must be classed as exhibiting complete overlap when the model includes F1 and F2 alone. As mentioned, these relations are reduced to no or partial overlap with the incorporation of a factor for duration in /i: ∼ i/ and /a: ∼ a/, while /u: ∼ u/ falls just above the cutoff for partial overlap. Jamaican English /i: ∼ i/ and /a: ∼ a/ are reduced from partial to no overlap when modeled in three dimensions, and /u: ∼ u/ from complete to no overlap. Thus, once duration is accounted for, JE overlap fractions come to most closely resemble those of the two-dimensional results for PNWEng (10%, 14%; no overlap). The overall average for JC decreases from complete to partial overlap (29%).

VI. SUMMARY

We have considered an approach to vowel characterization that uses linear algebraic formulas and geometrical models to represent vowels as ellipses (with area) and ellipsoids (with volume). Two algorithms are used to calculate the area of overlap (for ellipses) or the shared regions within overall vowel volumes (for ellipsoids). The output of each algorithm is an overlap fraction. These models are statistical in that they are not directly based upon calculations of acoustic vowel features, but rather take as their input summary statistical information (formant and duration means and standard deviations) that is used to locate the center and define the spread of each vowel in acoustic vowel space. Two least-squares lines are used to orient each vowel relative to F1, F2, and duration axes. Best-fit ellipses (or ellipsoids) are then used for area (or volume) overlap calculations. This method provides a more objective basis for simultaneously investigating spectral and temporal relations within vowel systems.

These methods have been applied to an investigation of spectral/temporal relations in three linguistically different varieties. American English has traditionally been classified, on the basis of impressionistic evidence, as a secondary quantity language. This investigation has demonstrated that for a sample of speakers of a Pacific Northwestern variety of American English, the domains for three tense/lax vowel pairs may be considered to be partially overlapping when modeled in F1 and F2, and that incorporation of durational information in the model reveals systematic but small additional reductions in overlap. Previous research into Jamaican phonology has reported that Jamaican Creole distinguishes vowels only by duration (Akers, 1981). What has not been previously reported is the nature of vowel quality differences in Jamaican Creole in light of concomitant differences in vowel length. Differences between Jamaican Creole and Jamaican English have also gone undescribed. The present study suggests that spectral separation of vowels does occur for Jamaican English speakers (as revealed in the overlap fractions in the two-dimensional model), but these differences are not equivalent for all oppositions. For Jamaican Creole, measures of overlap in two parameters support the interpretation that complete spectral overlap obtains for all oppositions. For Jamaican English, it appears that duration plays an enhancing role (for high front and low central contrasts) and serves as the primary basis for acoustic distinction in the high back pair.

The finding that vowel length may play a different role for different pairs of vowels within the same linguistic system suggests a more complex view of vowel system classifications than is typically taken in the linguistics literature. While it may be necessary from a phonological point of view to theorize that a particular abstract phonological feature such as vowel length operates at the level of the system (or subsystem) so that each contrast in the system bears a specification for that feature, it has been informative to see the differences between pairs of oppositions, revealed by the analysis above. While theories of phonology typically acknowledge that robust phonetic (defined broadly) differences may provide the basis of phonological contrast, it is well understood that phonetic surface features are complex and difficult to predict. Even when normalization algorithms have suppressed the effects of nonlinguistic factors (e.g., physiological differences between speakers), there are additional, systematic sources of variability that remain and must be explained. Because vowel-extrinsic (see Appendix A) normalization methods have been applied in the present analysis, the variability we observe in overlap fractions appears to be linguistic or sociolinguistic in nature. We have considered all oppositions separately (e.g., before duration is modeled, Jamaican Creole oppositions overlap by 75%, 55%, and 86%). We have also noted whether or not all three oppositions may be given the same classification (e.g., before duration is modeled, all Jamaican Creole oppositions fall above the arbitrary cutoff for complete spectral overlap). The latter perspective allows us to see that duration differences of a certain magnitude are maintained in all contrasts, which is useful in understanding forces possibly operating within an overall system. However, it should not be surprising that different vowel pairs stand in different relations (i.e., that different overlap fractions obtain for different pairs in the same system; e.g., compare Jamaican English /aː∼a/, 23%, to /uː∼u/, 75%). This finding is interpretable from the perspective of several phenomena well known to linguists, such as sound change. Historical phonological research has amply demonstrated that vowels within a system may begin a course of change one at a time (e.g., lowering of the nucleus of Middle English long-u, the vowel in *mouse*, “breaking” [uː] into diphthong [aʊ] in Early Modern English which was subsequently followed by raising of Middle English long-o). Such changes occur within the system in an orderly fashion, as so-called “chain shifts.” These changes often disturb the symmetry of a vowel system, bringing one vowel (the one undergoing change in this case) into a closer spectral or temporal relation with a new neighbor but further from vowels in other parts of the system. Thus, surface phonetic features must be understood to reflect both diachronic (historical) and synchronic (sociolinguistic) variation, from one community to another and from one point in time to another.

There is a continued need within experimental phonetic research to critically consider the use of the ellipse and the ellipsoid to represent vowel distributions in acoustic space. This convention appears to exist out of convenience while having no basis in auditory or acoustic reality (Ladefoged, p.c.). We find that best-fit ellipses (using two standard deviations of the mean for calculation of ellipse vertices) superimposed upon the original values for vowel data typically fit the data closely, without appearing to leave large “gaps” within a distribution, however this has by no means been tested mathematically. It would, of course, be desirable to find optimal geometrical representations on a more principled, by-vowel basis. Such an approach might be illuminating with respect to different patterns in acoustic clustering between vowels, as well as possible auditory or acoustic motivations for this. Finally, further research would benefit from dynamic representations of time-varying vowel features. The present method, while enabling examination of spectral and temporal features in a multidimensional geometrical representation, is limited to use of spectral vowel measures from one temporal point. The next phase of this project involves seeking means of representing spectral features across the duration of the vowel.

Widespread continued use of the results of descriptive studies (such as those of Peterson and Barney, 1952; Hagiwara, 1997, etc.) underscores the ongoing need for research providing baselines for the features of vowel systems. However, as argued by Hagiwara (1997) and others, such studies have not tended to adequately account for dialectal variation, to say nothing of the effects of ongoing phonological change embedded within the speech community. Further studies are needed to provide current baselines, drawn on carefully designed samples, that reflect what is currently known about phonetic variation. We have argued here that descriptive applications such as the one undertaken in this article further require the transformation of data using a normalization process that minimally suppresses possible linguistic effects. Nearey’s log-mean uniform scaling technique was used to achieve within-speaker normalization. The normalized values for each speaker constituted the input values (i.e., scatter) for between-vowel comparisons. Other sociolinguists have similarly found this method to be suitable for phonetic analysis of data in studies of the speech community (e.g., Labov, 1994; Adank *et al.*, 2004). For both sociolinguistic and phonetic research, integration of the methods, and increased attention to the results, of the other holds clear promise for the advancement of both disciplines.

ACKNOWLEDGMENTS

The author wishes to express gratitude to Jeremy Waltmunson for extension of the SOAM graphical user interface into the VOIS3D (Vowel Overlap Indication Software, in 3 Dimensions) software package, that enables visualization of SOAM output, and to Setsuko Shirai for assistance in researching normalization techniques. Development of the VOIS3D graphical user interface was supported by University of Washington Royalty Research Grant No. 65-2597.

APPENDIX A: NORMALIZATION METHODS USEFUL FOR OVERLAP CALCULATIONS

In the last 20 years, a good deal of discussion around the topic of normalization of the acoustic signal has been carried in this journal (e.g., Lobanov, 1971; Nearey, 1977, 1989; Disner, 1980; Syrdal and Gopal, 1986; Miller, 1989). Other studies (e.g., Hindle, 1978; Syrdal, 1984; Deterding, 1990) have also demonstrated the need for normalization algorithms that allowed for the suppression of systematic, predictable effects of various types, such as the influence of sex-based differences on the fundamental frequency domain, due to physiological differences such as vocal tract length and circumference. Adank *et al.* (2004) evaluate the effectiveness of 11 vowel normalization procedures for application in language variation research. These procedures are classed into one of two types depending on the information they employ. Vowel-intrinsic normalization procedures make use of formant data for a single vowel token and yield normalized vowels for that token. Common methods include nonlinear transformations of the frequency scale (log, mel, bark). Vowel-extrinsic normalization procedures make use of (spectral) information about more than one vowel or about the speaker's vowel system to normalize the data for each vowel. For example, some vowel-extrinsic methods make use of the formant frequencies of the point vowels, while others make use of formant frequencies for all vowels in the system. Such methods often calculate a grand mean, representing the center of a speaker's vowel space.

The present study constitutes a language variation study of a different type than that presented in Adank *et al.*, however, we also find that vowel-extrinsic procedures are most appropriate. We are concerned here with modeling vowel distributions and assessing the relative contributions of spectral and temporal features to vowel distribution volume. This requires an approach to normalization and quantification that both suppresses the effects of unwanted sources of variability (physiological effects of speaker sex) and preserves extralinguistic dimensions of contrast (e.g., regional or cross-linguistic differences) as well as phonemic differences (i.e., vowel identity). This need may be reduced to two problems: a normalization problem and a multidimensional modeling problem. SOAM uses one of several "vowel-extrinsic" normalization methods to accomplish within-speaker normalization to address the first need. This process yields normalized formant and duration values for each speaker that may be used for between-speaker comparisons. Then, the multidimensional modeling problem is solved by using a geometric algorithm that represents normalized scatter for two vowel distributions as best-fit ellipsoids oriented at angles with respect to the F1, F2, and duration axes. The output of SOAM is an overlap fraction, a value that represents the volume of the region of overlap (the coordinates of normalized vowel space shared by both best-fit ellipsoids). Three vowel-extrinsic methods for spectral normalization and two methods for temporal normalization have to date been tested for use with SOAM. Since the primary purpose of the present study was to develop a method for modeling vowels in three dimensions and not to evaluate vowel normalization techniques, the discussion of normalization below takes as its point of departure the evaluations of normalization methods described in the literature. However, it will be necessary to describe limitations that were discovered while researching particular procedures, and the steps taken to address these. Formant-based methods tested for use in SOAM include log-mean normalization (Nearey, 1989), known-extremes normalization (Shirai, 2004), and a modified *Z*-score normalization. Duration-based methods include segmental duration *Z*-score and phrase duration *Z*-score. Methods are classed as "vowel-extrinsic" or "vowel-intrinsic," following Adank *et al.* (2004). The overlap method is described in Appendix B.

1. Methods for spectral normalization

I. *Log-mean normalization* (vowel-extrinsic).

This procedure, adapted by Wassink [1999; based on Nearey's (1977) Uniform Scaling technique; see also Disner, 1980], is appropriate for normalizing data for two vowels produced by a single speaker when the spectral locations (in an F1×F2 acoustic space) of all vowels in the speaker's linguistic system are known. Nearey's method has become perhaps the most widely used method of normalization used in sociolinguistic studies of language variation because it minimally reduces linguistic and sociolinguistic variation while suppressing certain effects of physiological factors (Hindle, 1978; Labov, 2001).

In Nearey's normalization procedure Hertz frequency values for vowel formants are first log-transformed to better reflect the scale of sensitivity of the human ear to changes in frequency (Moore, 1989; Nearey, 1989). A second normalization step suppresses the effects of interspeaker variation. For example, males tend on average to have lower fundamental and lower formant frequencies than females, and their vowel spaces tend to be more compact overall than those of females. The log-transformed frequency G_{hijk} of a particular formant h (F1 or F2) occurring for a token i of a given vowel j for a particular speaker k is converted to a difference score F_{hijk}: the difference of that value from the mean of all values for that speaker for that formant, $\bar{G}_{hk}$ [Eq. (A1)], is

$$F_{hijk} = G_{hijk} - \bar{G}_{hk}. \tag{A1}$$

If the number of tokens is the same for all vowel categories in the speaker's system, the speaker grand mean $\bar{G}_{hk}$ (i.e., the F1 or F2 grand mean) is calculated over all realizations of that formant for that speaker, using Eq. (A2):

$$\bar{G}_{hk} = \frac{\left(\sum_{j=1}^{n} \sum_{i=1}^{n_j} G_{hijk}\right)}{nn_j}, \tag{A2}$$

where n is the number of vowels in the speaker's system and n_j is the number of data tokens for each vowel. This method, however, results in skewing of the grand mean for a given dimension away from the true center of speaker's vowel space when the number of vowels in any category is disproportionately represented in the sample. Therefore, it was necessary to adapt the method. If the number of tokens is different for different vowel categories, then $\bar{G}_{hk}$ is calculated as the mean of the category means for each formant for each vowel, as in Eq. (A3):

$$\bar{G}_{hk} = \frac{\sum_{j=1}^{n}\left(\sum_{i=1}^{n_j} G_{hijk}/n_j\right)}{n}, \tag{A3}$$

where n is the number of vowels in the speaker's system and n_j is the number of data tokens for vowel j.

II. *Known extremes normalization* (vowel extrinsic). This method was proposed by Shirai (2004) for normalization of spectral data for languages that have asymmetrical acoustic distributions, such as Japanese, where the so-called "point vowels" /i, a, u/ do not demarcate the edges of vowel space in F1 and F2. The "known extremes" method is also appropriate when locations are only known for a subset of the vowels in a speaker's system where these vowels demarcate the edges of the system. This method allows for the calculation of a grand mean for the system and, for this reason, may be used for cross-speaker comparisons (within one language). Unlike methods I and III, this method includes a formula for the determination of the vowel categories that represent the maxima and minima in F1 and F2. This method also differs from I and III in the method for calculating a grand mean for each speaker [Eq. (A4)]:

$$\bar{G}_{hk} = \left(\sum_{j=1}^{2}\left(\frac{\sum_{i=1}^{n_j} G_{hijk}}{n_j}\right)\right)/2, \tag{A4}$$

where $j=1$ is vowel at the lower extreme of the formant h for the speaker's vowel system, and $j=2$ is a vowel at the upper extreme of the formant h for the speaker's vowel system. Thus only two vowels are used in the calculation of the known extremes grand mean for a particular formant. Once the grand mean has been calculated in this way, the procedure for normalizing individual formant values follows as in Eq. (A1).

III. *Modified Z-score normalization* (vowel extrinsic). This method, introduced by Lobanov (1971), is appropriate when only a subset of the vowel locations is known (e.g., the language has five vowels, but we have data only for the two vowels whose overlap we desire to compare). Crucially, this method is only used for overlap calculations within a single speaker. If the number of tokens is the same for each vowel for which data are available, a formant grand mean $\bar{G}_{hk}$ (i.e., an F1 or F2 grand mean) for the subject is calculated using Eq. (A2) above, only that n is the number of vowels for which data are available. As with log mean normalization, because numbers of datapoints for each vowel category may be unequal, a vowel category mean is used to represent each vowel category rather than individual observed datapoints. In this case Eq. (A3) above may be used, again with n referring to the number of vowels for which data are available, rather than the total number of vowels in the speaker's system. Once again, the normalized formant values are calculated using Eq. (A1).

2. Methods for temporal normalization

IV. *Segmental duration Z-score normalization*. This procedure is similar to the one used for calculating Z scores for spectral data, except that the data are not log transformed. Here, segmental duration values are used. Calculate the mean duration for each vowel category (e.g., unique category means are calculated for /i/, /a/, /u/, etc.) A grand mean of duration, $\bar{D}_{o,k}$, for speaker k is calculated across all category durations in Eq. (A5):

$$\bar{D}_{o,k} = \frac{\sum_{j=1}^{n}\left(\sum_{i=1}^{n} D_{ijk}/n_j\right)}{n}, \tag{A5}$$

where D_{ijk} is the observed segment duration for token i of vowel j for speaker k. The normalized value of duration, δ_{ijk}, is equal to the observed value minus the category grand mean [Eq. (A6)]:

$$\delta_{ijk} = D_{ijk} - \bar{D}_{ok}. \tag{A6}$$

V. *Phrase-duration normalization*. Quite frequently, it is necessary to use phrase durations to normalize segmental durational values. For example, in experimental methods that rely upon use of reading lists, or other methods where rate of speech is not readily controlled, within- or between-speaker differences may arise in the durations due to fluctuation in speech or reading rate, rather than intrinsic differences in vowel durations. Ericsdotter and Ericsson (2001) and Simpson and Ericsdotter (2003) normalize duration of a vowel segment relative to the word or phonological phrase in which it was uttered. One adaptation of this idea is given in Eq. (A7),

$$D_{\text{norm}} = D_{\text{segment}}/D_{\text{phrase}}, \tag{A7}$$

where D_{norm} is the phrase-duration normalized segment duration, D_{segment} is the measured segment duration of a vowel token, and D_{phrase} is the duration of the carrier word or phrase for that token. This preprocessing constitutes an initial normalization step and is done before normalizing these durations to the system grand mean.

In summary, vowel-extrinsic procedures are always used for spectral and temporal normalization in vowel overlap calculations. This is because the majority of types of overlap comparisons that we might desire to compute appear to require scaling to a (system or language) grand mean. Furthermore, several vowel-extrinsic procedures incorporate nonlinear transformations of the frequency scale, for representing auditory sensitivity. For example, grand means (or uniform scaling) are desirable: for comparisons of overlap in two acoustically adjacent vowel categories in a single language where data are pooled across more than one speaker; for cross-linguistic comparisons, such as those reported above, in pairs of vowels present in the systems of both languages; for comparisons of the extent of overlap in vowel pairs for speakers of different (developmental) ages in studies of the development of phonemic contrast; and in investigations of the extent of loss of contrast (decrease in overlap) between two vowels under different speaking rates or registers. Even where we might wish to compute overlap between two vowels produced by a single speaker, we typically want to know how that speaker's overlap compares to that of other speakers (or to other vowels produced by the same speaker in different environmental conditions). Because calculation of the size of a vowel distribution (as compared to the spectral

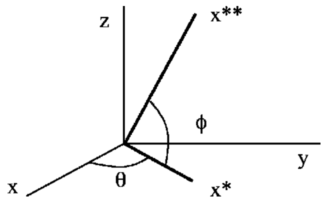

FIG. 3. Graphic of 3-D coordinate system, showing x,y,z,x^{*}, and x^{**} axes, angles θ, and ϕ.

location of a single formant) relies on distributional information (as described below), we need a method that will reflect distances between elements of the system.

APPENDIX B: AN ANALYTIC GEOMETRIC METHOD FOR QUANTIFYING VOWEL OVERLAP

1. Introduction

The spectral overlap assessment metric (SOAM) uses two algorithms for the calculation of the extent of overlap between pairs of vowels. Extent of overlap between two normalized vowel distributions is determined either by considering spectral features alone (two-dimensional overlap, SOAM 2-D) or by considering spectral and temporal features together (three-dimensional overlap, and SOAM 3-D). Each member of a pair of vowel distributions is modeled geometrically as an ellipse in F1×F2 space or an ellipsoid in F1×F2×duration space. The ellipse or ellipsoid is best fit to the data scatter for the vowel using least-squares fitting to determine principal axes. Overlap between a pair of vowel distributions is calculated based on a fraction of uniformly distributed test points in the region of overlap relative to the number of test points in each vowel distribution.

Vowel distributions have been modeled in the past as ellipses in F1×F2 space (Wakita, 1976; Ladefoged, 1996). In this study, a vowel distribution is modeled as a best-fit ellipse oriented at an angle with respect to the F1×F2 axes or as an ellipsoid in F1×F2×duration space. Wassink (1999) described a method for assessing overlap between two vowel distributions modeled as ellipses based on the extent of protrusion of one ellipse into the other. Here, instead of a protrusion metric, area of overlap is calculated.

2. Calculation of vowel distribution overlap in three dimensions

a. Modeling two vowel distributions in three dimensions

In this case, spectral information is supplied for both vowels in each pairwise comparison for two dimensions (where F2 is the x axis and F1 is the y axis) while temporal information is supplied for the third (z axis). Prior to application of this method, data are normalized using one of the procedures described in Appendix A.

For convenience, plotting of data in three dimensions is accomplished using a polar coordinate system, as illustrated in Fig. 3. x,y, and z represent the axes associated with F2, F1, and duration dimensions, respectively.

Each of the vowel distributions may be modeled as an ellipsoid in F2×F1×duration space. Two vowel distributions may be defined as $\boldsymbol{v}_1=(\boldsymbol{x}_1,\boldsymbol{y}_1,\boldsymbol{z}_1)$ and $\boldsymbol{v}_2=(\boldsymbol{x}_2,\boldsymbol{y}_2,\boldsymbol{z}_2)$ where $\boldsymbol{x}_j,\boldsymbol{y}_j$, and $\boldsymbol{z}_j$ are column vectors of F2, F1, and duration data, respectively, for vowel j, and (x_{ij},y_{ij},z_{ij}) are F2×F1×duration data triplets for a particular vowel token.

We begin by subtracting the mean values of F2, F1, and duration for vowel j, $\boldsymbol{x}_{o,j}=(x_{o,j},y_{o,j},z_{o,j})$, from each of the tokens in that distribution. Next, the orientation of the principal axis projected into the F2×F1 plane is calculated through a least squares fit of a line to the F2×F1 data. The angle, θ_j, between the $\boldsymbol{x}$ axis and the $\boldsymbol{x}_j^{*}$ axis is calculated from the slope of the least squares line [Eq. (B1)]

$$\theta_j=\tan^{-1}(k_{xy,j}). \tag{B1}$$

A rotation matrix $\boldsymbol{R}_{\theta j}$ is developed [Eq. (B2)] to rotate the data into a coordinate system aligned with $\boldsymbol{x}_j^{*}$:

$$\boldsymbol{R}_{\theta j}=\begin{bmatrix}\cos(-\theta_j) & -\sin(-\theta_j) & 0\\ \sin(-\theta_j) & -\cos(-\theta_j) & 0\\ 0 & 0 & 1\end{bmatrix}. \tag{B2}$$

The matrix containing translated values for each data point is multiplied by the above rotation matrix producing data in $(\boldsymbol{x}_j^{*},\boldsymbol{y}_j^{*},\boldsymbol{z}_j^{*})$ coordinates [Eq. (B3)]:

$$\boldsymbol{v}_j^{*}=[\boldsymbol{x}_j-\boldsymbol{x}_{o,j}]\boldsymbol{R}_{\theta,j}. \tag{B3}$$

Once the rotation within the F2×F1 plane has been carried out, a least squares line is fit to the $(\boldsymbol{x}_j^{*},\boldsymbol{z}_j^{*})$ data to get a second principal axis for the ellipsoid, oriented at an angle ϕ_j relative to the x_j^{*} axis [Eq. (B4)]. Finally, the $\boldsymbol{v}_j^{*}$ data are transformed to a coordinate system $\boldsymbol{v}_j^{**}=(\boldsymbol{x}_j^{**},\boldsymbol{y}_j^{**},\boldsymbol{z}_j^{**})$ aligned with this angle, using a second rotation matrix $\boldsymbol{R}_{\phi j}$ [Eqs. (B5) and (B6)]:

$$\phi_j=\tan^{-1}(k_{x^{*}z^{*},j}), \tag{B4}$$

$$\boldsymbol{R}_{\phi j}=\begin{bmatrix}\cos(-\phi_j) & 0 & -\sin(-\phi_j)\\ 0 & 1 & 0\\ \sin(-\phi_j) & 0 & \cos(-\phi_j)\end{bmatrix}, \tag{B5}$$

$$\boldsymbol{v}_j^{**}=\boldsymbol{v}_j^{*}\boldsymbol{R}_{\phi,j}, \tag{B6}$$

Figures 4(a) and 4(b) show the first vowel distribution prior to and following translation and rotation, while Figs. 5(a) and 5(b) show the second vowel distribution prior to and following translation and rotation.

Once the vowel distribution data are converted to a coordinate system aligned with the principal axes, it is simple to calculate the principal radii of the ellipsoid. The standard deviation of data along each axis (that is, in each of the coordinates) is calculated, and the principal radius along that axis is defined as m times the standard deviation in that direction. A value of $m=2.0$ is used. The formula for each ellipsoid is solved separately, using Eq. (B7):

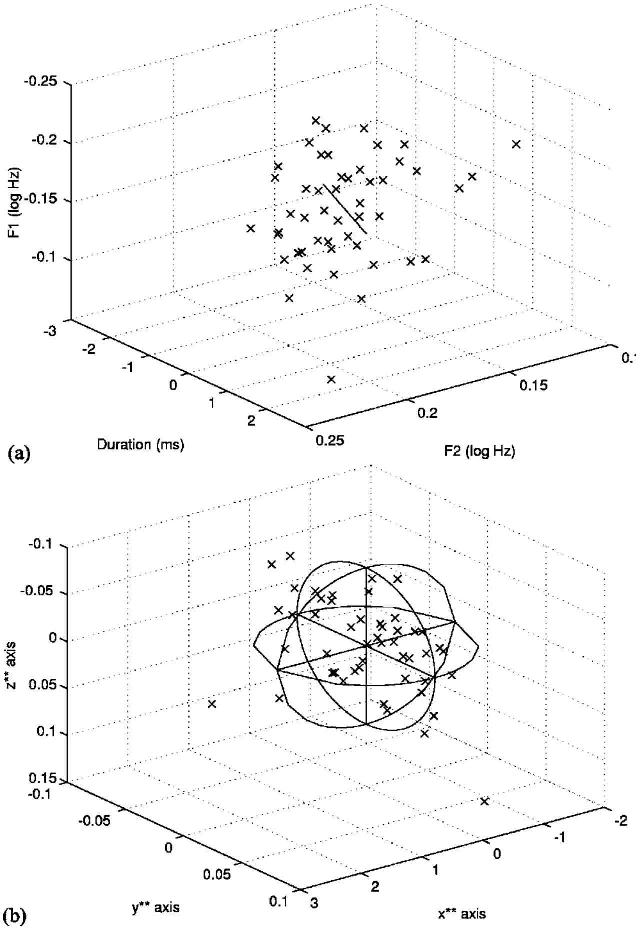

FIG. 4. Distribution of scatter for /i/ for speaker TX (a) before data translation and rotation and (b) after. The solid line in (a) represents the major axis defined by least-squares line.

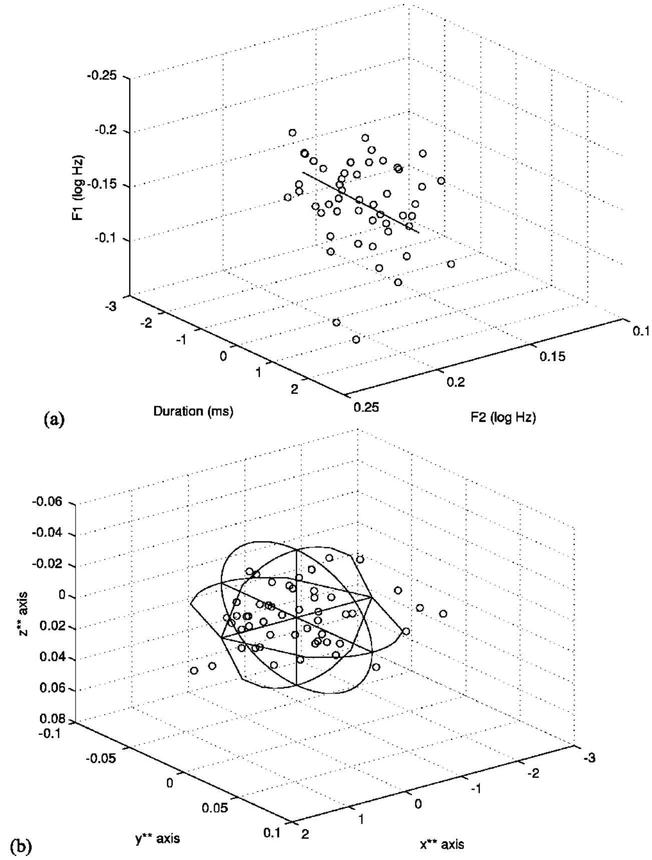

FIG. 5. Distribution of scatter for /iː/ for speaker TX (a) before data translation and rotation and (b) after. The solid line in (a) represents the location of the major axis.

$$\frac{x^2}{a_j^2}+\frac{y^2}{b_j^2}+\frac{z^2}{c_j^2}=1, \tag{B7}$$

where a_j, b_j, and c_j are the principal radii for ellipsoid j. The ellipsoid equation for vowel j may be converted to polar coordinates, using $x=r\cos\theta\cos\phi$; $y=r\sin\theta\cos\phi$; $z=r\sin\phi$ [Eq. (B8)]:

$$\frac{r^2\cos^2\theta\cos^2\phi}{a_j^2}+\frac{r^2\sin^2\theta\cos^2\phi}{b_j^2}+\frac{r^2\sin^2\phi}{c_j^2}=1. \tag{B8}$$

Equation (B8) can be solved for r and written as a function of $r_j(\theta,\phi)$ [Eq. (B9)]:

$$r_i(\theta,\phi)=\left[\sqrt{\left(\frac{\cos^2\theta\cos^2\phi}{a_j^2}+\frac{\sin^2\theta\cos^2\phi}{b_j^2}+\frac{\sin^2\phi}{c_j^2}\right)}\right]^{-1}, \tag{B9}$$

where j is the vowel index ($j=1$ for vowel 1, $j=2$ for vowel 2); a_j, b_j, and c_j are calculated from the standard deviations in data in the x_j^{**}, y_j^{**}, and z_j^{**} directions, respectively [Eqs. (B10)–(B12)]:

$$a_j=m\sigma_{x^{**},j}, \tag{B10}$$

$$b_j=m\sigma_{y^{**},j}, \tag{B11}$$

$$c_j=m\sigma_{z^{**},j}. \tag{B12}$$

Once followed for both vowels, the steps above yield two ellipsoids, each representing a vowel distribution. The representative ellipsoid for vowel 1 (referred to as ellipsoid 1) is shown along with original data for vowel 1 in Fig. 4(b). The representative ellipsoid for vowel 2 (referred to as ellipsoid 2) is shown along with original data for vowel 2 in Fig. 5(b).

The overlap calculations are based upon the volume of overlap of the two ellipsoids in (x,y,z) space. A graphical representation of ellipsoid j can be created in (x,y,z) space by generating parametric data E_j^{**} using Eq. (B9) for varying values of (θ,ϕ) in $(x_j^{**},y_j^{**},z_j^{**})$ space, and then rotating and translating this data from $(x_j^{**},y_j^{**},z_j^{**})$ space to (x,y,z) space. Assuming that we have E_j^{**}, the transformation is given by Eqs. (B13) and (B14):

$$\mathbf{E}_j^{*}=\mathbf{E}_i^{**}\begin{bmatrix}\cos\phi_j & 0 & -\sin\phi_j\\ 0 & 1 & 0\\ \sin\phi_j & 0 & \cos\phi_j\end{bmatrix}, \tag{B13}$$

$$\mathbf{E}_j=\mathbf{E}_j^{*}\begin{bmatrix}\cos\theta_j & -\sin\theta_j & 0\\ \sin\theta_j & \cos\theta_j & 0\\ 0 & 0 & 1\end{bmatrix}+E_{o,j}, \tag{B14}$$

where the mean values of F2, F1, and duration for the vowel distribution, $E_{o,j}$, are added to the coordinates for the ellipsoid representing that vowel (Fig. 6). In summary, the para-

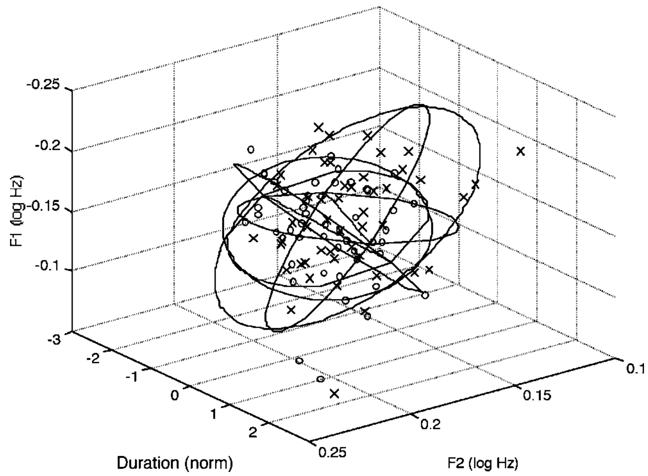

FIG. 6. Scatter data and representative ellipsoids for /i/ (x-marks) and /i:/ (o-marks) of speaker TX. For each ellipsoid, the line represents the major axis.

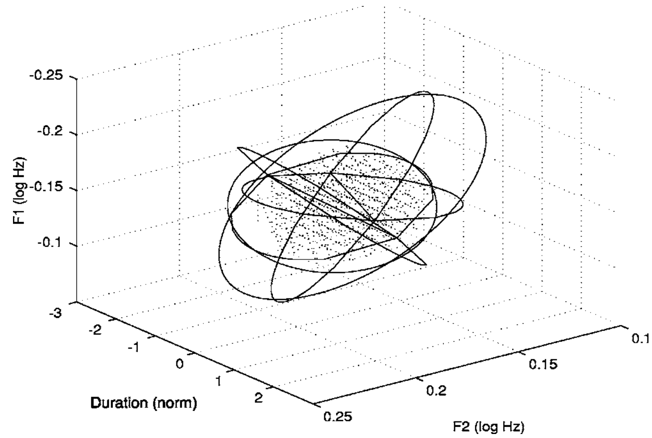

FIG. 7. Representative ellipsoids for /i/ and /i:/ of speaker TX, male, Jamaican Creole. The overlap region between the two vowels is shaded. Calculated 3-D overlap is $\Omega = 23\%$.

metrically generated ellipsoid data go through the reverse of the process used to align the original scatter data to the principal axes.

b. Calculating vowel overlap

To calculate the extent of overlap between the two ellipsoids, a somewhat "brute-force" method is employed. A space slightly larger than the space occupied by both ellipsoids is divided into a grid of evenly spaced test points. For most test cases, a $20\times 20\times 20$ test grid was found to give a good compromise between precision and computation time. For each test point, l, it is calculated whether or not the test point falls within ellipsoid j. The test point (x_l, y_l, z_l) is transformed into $(x_{lj}^{**}, y_{lj}^{**}, z_{lj}^{**})$ space, and the coordinate system aligned with ellipsoid j, by subtracting mean values and rotating the coordinate system [Eqs. (B15) and (B16)]:

$$[x_{lj}^{*} \quad y_{lj}^{*} \quad z_{lj}^{*}] = [(x_{lj} - x_{o,j}) \quad (y_{lj} - y_{o,j}) \quad (z_{lj} - z_{o,j})]\mathbf{R}_{\theta j}, \tag{B15}$$

$$[x_{lj}^{**} \quad y_{lj}^{**} \quad z_{ij}^{**}] = [x_{lj}^{*} \quad y_{ij}^{*} \quad z_{lj}^{*}]\mathbf{R}_{\phi j}. \tag{B16}$$

Next, the distance from the test point to the center of the ellipsoid is calculated [Eq. (B17)]:

$$d_{lj} = \sqrt{(x_{lj}^{**})^2 + (y_{ij}^{**})^2 + (z_{lj}^{**})^2}. \tag{B17}$$

Second, a radius is determined for ellipsoid j at the angle from the $\mathbf{x}^{**}$ axis to the line from the ellipsoid center to test point l. The angles are calculated from Eqs. (B18) and (B19):

$$\theta_{lj} = \tan^{-1}(y_{lj}^{**}/x_{lj}^{**}), \tag{B18}$$

$$\phi_{ij} = \tan^{-1}\left(\frac{z_{lj}^{**}}{\sqrt{(x_{lj}^{**})^2 + (y_{lj}^{**})^2}}\right), \tag{B19}$$

which can then be entered into the ellipse formula [Eq. (B19)] for $r_{lj}(\theta, \phi)$.

If the local radius is greater than or equal to the test point distance $[r_{lj}(\theta,\phi) >= d_{lj}]$, the test point is taken to be within the ellipsoid. This process is repeated for the test point relative to each of the ellipsoids $(j=1,2)$. The entire procedure is repeated for each test point l. The results are tallied for each of the test points: Is it in ellipsoid 1? Is it in ellipsoid 2? Is it in both ellipsoids?

The number and coordinates of all test points within the overall grid that are contained within each vowel are tracked. (The number of test points in an ellipsoid is approximately proportional to the "volume" of that ellipsoid.) The set of points that are contained both within vowel 1 and vowel 2 define the overlap region. Initially the overlap region is calculated relative to each vowel. The overlap fraction of vowel 1, Ω_1, is calculated [Eq. (B20)] as the total of test points in the overlap region between the two ellipsoids, N_{both}, divided by the total test points in ellipsoid 1, N_1,

$$\Omega_1 = N_{\text{both}}/N_1. \tag{B20}$$

Similarly, the overlap fraction of vowel 2, Ω_2, is calculated [Eq. (B21)] as the total of test points in the overlap region, N_{both}, divided by the total test points in ellipsoid 2, N_2.

$$\Omega_2 = N_{\text{both}}/N_2. \tag{B21}$$

The overlap fraction of the vowel pair, Ω, is defined [Eq. (B22)] as the larger of these two values:

$$\Omega = \max[\Omega_1, \Omega_2]. \tag{B22}$$

Since the two ellipsoids may be of different volumes, or because one ellipsoid may be contained within the other, the larger of the resulting calculations is taken as the overlap percentage. That is, it is desirable to know when one of the volumes protrudes substantially onto the other or is enveloped by the other. Note that this method does not tell us the region in three-dimensional space where the overlap is most extensive; the overlap percentage reflects only the extent of overlap. In Fig. 7, the representative ellipsoids are shown along with a shaded area representing the overlap region.

c. Calculation of vowel distribution overlap in two dimensions

The method for calculation of vowel distribution overlap in two dimensions is analogous to the three-dimensional case. Vowels are modeled as ellipses oriented at an angle to

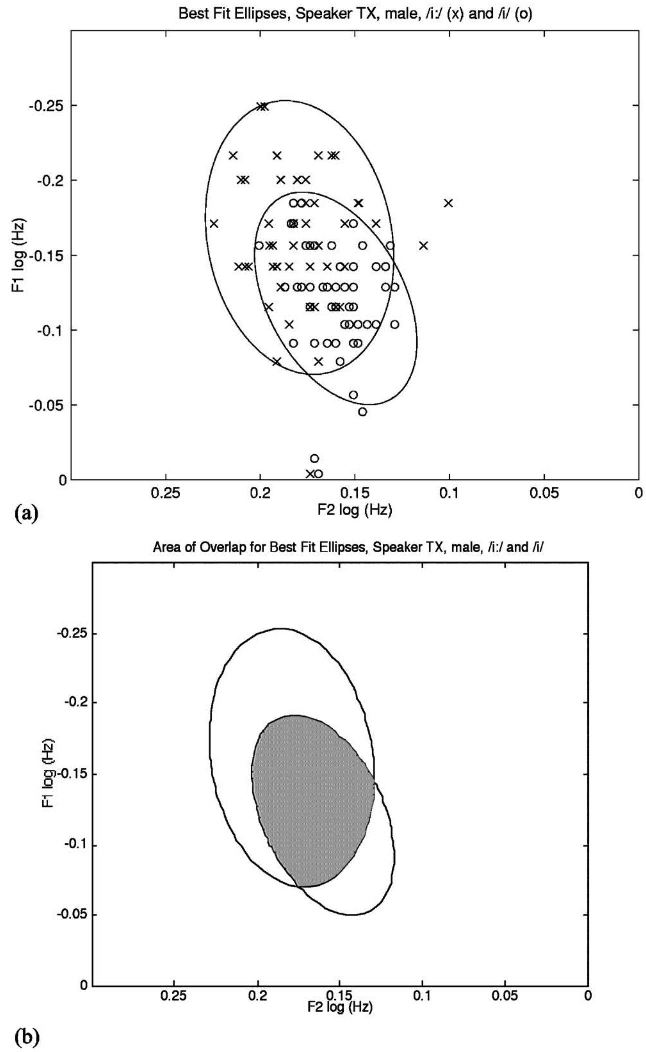

FIG. 8. (a) Best-fit ellipses fitted to scatter for vowel /i:/ (represented by "x") and /i/ ("o") in F1×F2 space. (b) Best-fit ellipses, in F1×F2 space, following calculation of overlap area (shaded). Calculated overlap fraction $\Omega = 76\%$. For Jamaican Creole speaker TX, male, vowels /i:/ and /i/.

F1×F2 axes, using least squares fitting. Overlap is determined using a uniform grid of test points. Results of the two-dimensional method (for vowels /i/ and /i:/ for one male Jamaican Creole speaker) are shown in Fig. 8.

For the example we have followed throughout this Appendix, note that the calculated overlap Ω in the three-dimensional case (23%) is significantly lower than that (76%) calculated in the two-dimensional case where duration is not considered. This result suggests that duration contributes significantly to the contrast between /i/ and /i:/ for this speaker.

[1]That is, a systematic difference between the members of a phonological opposition. The convention of classifying vowel systems assumes similar spectral/temporal relations between all contrasting pairs in the system. This is an oversimplification, accomplished for theoretical convenience. In fact, empirical studies must take specific pairwise contrasts as a starting point, as will be accomplished in the study reported here, for reasons to be discussed in Sec. IV, below.

[2]In addition, the passage of time between the two studies raises the possibility that phonological change may also complicate comparability.

[3]Standard deviations are not provided. Furthermore, looking only at mean differences in F1 and F2 is only partly revealing. Women's and childrens' mean F1 for /a/-class words (*"bag,"* phonetically [æ]) is actually *lower* (i.e., this vowel is situated higher in inverted F1×F2 space) than /ε/ in the same system. Thus, they show greater raising than males in the same group.

[4]Dialect may have been a confounding factor in the Peterson and Barney study. Their 28 females and 15 children were primarily speakers of a Middle Atlantic dialect, while the 33 males represented "a much broader regional sampling." Thus, in addition to representing a different dialect than that of Hillenbrand *et al.*, the Peterson and Barney sample itself may have been drawn using speakers from different dialect regions (males versus females and children).

[5]Hillenbrand *et al.* do not give actual values for the curves presented in this graph, but hand-measuring yields the following lengths for each curve:

	Males	Females	Children
/æ/ (mm)	7	18	16
/ε/	6	6	7

[6]/h_d/ data were not able to be collected for the Jamaican speakers due to the deletion of /h/ in stressed, word-initial positions in all varieties of this language (Akers, 1981).

[7]The symbol /ɔ/ represents the fact that this word class was elicited for all speakers, all varieties. In the PNWEng sample, however, this phonetic quality did not exist. This word class was merged with /a:/.

[8]*Cot* and *caught* word classes are merged in this variety of American English. Thus, the /a:/ category in this study contains pooled data from both classes.

Adank, P., Smits, R., and van Hout, R. (**2004**). "A comparison of vowel normalization procedures for language variation research," J. Acoust. Soc. Am. **116**, 3099–3107.

Akers, G. (**1981**). *Phonological Variation in the Jamaican Continuum* (Karoma, Ann Arbor, MI).

Crothers, J. (**1978**). "Typology and universals in vowel systems," in *Universals of Human Language*, edited by J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (Stanford U.P., Stanford), pp. 93–152.

Deterding, D. (**1990**). "Speaker normalization for automatic speech recognition," Ph.D. thesis, University of Cambridge.

Disner, S. F. (**1980**). "Evaluation of vowel normalization procedures," J. Acoust. Soc. Am. **67**, 253–261.

Ericsdotter, C., and Ericsson, A. M. (**2001**). "Gender differences in vowel duration in read Swedish: preliminary results," in *Proc. of Phonetic 2001, the XIVth Swedish Phonetics Conferences, Örenaäs*, 30 May 30–1 June 2001. Reprinted in *Working Papers*, *Department of Linguistics*, *Lund University*, Vol. **49**, pp. 34–37.

Hagiwara, R. (**1997**). "Dialect variation and formant frequency: the American English vowels revisited," J. Acoust. Soc. Am. **102**, 655–658.

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

Hindle, D. (**1978**). "Approaches to normalization in the study of natural speech," in *Linguistic Variation, Models and Methods*, edited by D. Sankoff (Academic, New York), pp. 161–171.

Kluender, K., Diehl, R., and Wright, B. (**1988**). "Vowel-length differences before voiced and voiceless consonants: an auditory explanation," J. Phonetics **16**, 153–169.

Labov, W. (**1994**). *Principles of Linguistic Change: Internal Factors* (Blackwell, Oxford).

Labov, W. (**2001**). *Principles of Linguistic Change: Volume 2, Social Factors* (Blackwell, Oxford).

Ladefoged, P. (**1996**). Plot Formants software, UCLA Phonetics Laboratory, http://www.linguistics.UCLA.edu/faciliti/facilities/acoustics/acoustics.html

Lehiste, I. (**1970**). *Suprasegmentals* (MIT, Cambridge, MA), pp. 18–33.

Lobanov, B. M. (**1971**). "Classification of Russian vowels spoken by different speakers," J. Acoust. Soc. Am. **49**, 606–608.

Macken, M. L. (**1980**). "Phonological universals and variation: on pattern and diversity in acquisition," unpublished Ph.D. dissertation, Stanford University.

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

Moore, B. C. J. (**1989**). *An Introduction to the Psychology of Hearing*, 3rd ed. (Academic, London).

Nearey, T. M. (**1977**). "Applications of generalized linear modeling to vowel data," Proc. ICLSP **92**, 583–586.
Nearey, T. M. (**1989**). "Static, dynamic, and relational properties in speech perception," J. Acoust. Soc. Am. **85**, 2088–2113.
Nordström, P. -E., and Lindblom, B. (**1975**). "A normalization procedure for vowel formant data," Proc. ICPhS **8**, 212.
Perkell, J. S., and Matthies, M. L. (**1992**). "Temporal measures of anticipatory labial coarticulation for the vowel /u/: within- and cross-subject variability," J. Acoust. Soc. Am. **91**, 2911–2925.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.
Shirai, S. (**2004**). "Lexical effects in Japanese vowel reduction," unpublished Ph.D. thesis, University of Washington.
Simpson, A., and Ericsdotter, C. (**2003**). "Sex-specific durational differences in English and Swedish," *Proc. of the XVth ICPhS*, Barcelona, 3–9 August **2003**, pp. 1113–1116.
Strange, W., and Bohn, O.-S. (**1998**). "Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies," J. Acoust. Soc. Am. **104**, 488–504.
Syrdal, A. K. (**1984**). "Aspects of a model of the auditory representation of American English vowels," Speech Commun. **4**, 121–135.
Syrdal, A. K., and Gopal, H. S. (**1986**). "A perceptual model of vowel recognition based on the auditory representation of American English vowels," J. Acoust. Soc. Am. **79**, 1086–1100.
Wakita, H. (**1976**). "Normalization of vowels by vocal-tract length and its application to vowel identification," IEEE Trans. Acoust., Speech, Signal Process. **25**(2), 183–192.
Wassink, A. B. (**1999**). "A sociophonetic analysis of Jamaican vowels," unpublished Ph.D. dissertation, University of Michigan, Ann Arbor.
Wassink, A. B. (**2001**). "Theme and variation in Jamaican vowels," Language Variation Change, **13**(2), 135–159.