Synergistic modes of vocal tract articulation for American English vowels[a)]

Brad H. Story[b)]
Speech Acoustics Laboratory, Department of Speech and Hearing Sciences, University of Arizona, Tucson, Arizona 85721

(Received 6 June 2005; revised 19 September 2005; accepted 20 September 2005)

The purpose of this study was to investigate the spatial similarity of vocal tract shaping patterns across speakers and the similarity of their acoustic effects. Vocal tract area functions for 11 American English vowels were obtained from six speakers, three female and three male, using magnetic resonance imaging (MRI). Each speaker's set of area functions was then decomposed into mean area vectors and representative modes (eigenvectors) using principal components analysis (PCA). Three modes accounted for more than 90% of the variance in the original data sets for each speaker. The general shapes of the first two modes were found to be highly correlated across all six speakers. To demonstrate the acoustic effects of each mode, both in isolation and combined, a mapping between the mode scaling coefficients and [F1, F2] pairs was generated for each speaker. The mappings were unique for all six speakers in terms of the exact shape of the [F1, F2] vowel space, but the general effect of the modes was the same in each case. The results support the idea that the modes provide a common system for perturbing a unique underlying neutral vocal tract shape. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2118367]

PACS number(s): 43.70.−h, 43.70.Bk, 43.70.Gr [AL] Pages: 3834–3859

I. INTRODUCTION

The upper airway in humans, extending from the larynx to the lips, is utilized as a conduit for the physiologic processes of respiration, deglutition, and sound production. Each physiologic function requires specific types of coordinated articulatory movement involving the larynx, tongue soft palate, lips, and jaw to accomplish a particular goal. For respiration, the vocal folds must be abducted and, depending on the amount of oxygen needed, the jaw may be lowered and the tongue moved anteriorly and inferiorly to allow a high volume of air to flow into the lungs. A typical swallow proceeds as a sequence of distinct "phases" in which movements of articulatory structures, and their generated forces are precisely orchestrated to propel a bolus of food or liquid through the oral cavity, the pharynx, and eventually into the esophagus and stomach (e.g., Logeman, 1983). For both respiration and deglutition, the goals of the process are to achieve the efficient transport of fluids and solids to a destination (i.e., lungs and stomach). In contrast, the goals of speech production require that the articulatory system not only *transport* sound waves from the larynx to the lips, but also *transform* them into highly structured, linguistically relevant sounds.

The transformation of sounds, produced by vocal fold vibration or other sources, into speech is accomplished through specific shapes and shape changes of the tubelike structure formed collectively by the epilarynx, pharynx, and oral cavity (henceforth referred to as the "vocal tract"). This vocal tract tube creates an acoustic filter whose resonances generate enhanced regions of amplitude (i.e., formant frequencies) that are observable in the spectrum of a speech signal. The relation between vocal tract shape and the acoustic characteristics it generates, especially formant frequencies, is known to be nonlinear (e.g., Fant, 1960; Stevens and House, 1955) and, theoretically, many-to-one (Schroeder, 1967; Mermelstein, 1967). But even though many possible vocal tract shapes can support the same pattern of formant frequencies, speakers easily coordinate the actions of the articulators, efficiently modifying the vocal tract shape over time, to produce a coarticulated stream of intelligible speech sounds. Thus, part of understanding speech production is to know how the vocal tract shape, represented as a tubular entity, may be systematically controlled to specify and simplify its relation to the formant frequencies.

While all of the articulators potentially contribute to any particular vocal tract shape, it is largely dominated by the configuration of the tongue. This is evidenced by common use of the high/low and front/back descriptors for tongue position in vowel production and implies a simple relation to formant frequencies. Quantitative studies tend to support a view that the tongue shape can be described by underlying patterns of spatial deformation. Harshman *et al.* (1977) statistically decomposed tongue profiles of ten English vowels taken from midsagittal x-ray pictures into basic displacement patterns called "factors." They found that only two factors were needed to reconstruct the profiles within a small error of the originals. Furthermore, the factors had an apparent articulatory phonetic interpretation. The first indicated a forward movement of the tongue root along with an elevation of the front of the tongue; this was suggested to correspond to the action of the genioglossus muscle. The second factor produced an upward and backward movement of the tongue

[a)]Portions of this paper were presented at the 144th Meeting of the Acoustical Society of America.
[b)]Electronic mail: bstory@u.arizona.edu

which was compared to the action of the styloglossus muscle. Nearly parallel to this work was research reported by Shirai and Honda (1977), who also represented midsagittal tongue shapes with two empirically determined displacement patterns. The shapes were generally similar to those of Harshman *et al.* (1977). Using tongue shapes obtained from cineradiographic tracings for Icelandic speakers, Jackson (1988) reported that *three* factors were necessary to describe Icelandic vowels, where the second factor was of significantly different shape than that reported by Harshman *et al.* (1977). A reanalysis of these data by Nix *et al.* (1996), however, showed that two factors could, in fact, describe the Icelandic tongue shapes too. More recently, Zheng *et al.* (2003) have applied a factor analysis to tongue shapes determined from 3-D image sets based on MRI. While their results were not identical to previous studies, they suggest that a large percentage of the variance in tongue shape for vowels can be described with two shaping factors.

Although in a strict sense "factors" are only a statistical description of tongue shape, a physiologic basis for them was reported by Maeda and Honda (1994). During vowel production, the electromyographic activity of two antagonistic muscle pairs, hyoglossus (HG)-genioglossus posterior (GGp) and styloglossus (SG)-genioglossus anterior (GGa), was observed to coincide with the spatial tongue shaping characteristics of two factors that had been derived from midsagittal images of the vocal tract. The vowel-dependent differential activity of each antagonistic pair was also shown to form a vowel triangle much like that of the first two formant frequencies (Kusakawa *et al.*, 1993). To further investigate the relation between muscle activity patterns, vocal tract shape, and formant frequencies, Maeda and Honda (1994) used the analyzed EMG activities (of the HG-GGp and SG-GGa pairs) as the driving input to an articulatory model based primarily on two tongue shaping patterns [similar to the factors derived by Harshman *et al.* (1977)]. The results were F1-F2 patterns that corresponded closely to measured formant values and led them to state: "We speculate the brain optimally exploits the morphology of the vocal tract and the kinematic functions of the tongue muscles so that the mappings from the muscle activities (production) to the acoustic patterns (perception) are simple and robust."

The results of these experiments suggest that tongue shaping patterns obtained from statistical analyses may capture the spatial, and perhaps kinematic, representation of a coordinated pattern of individual articulator positions or movements that is directly related to the acoustic characteristics of speech. This is similar to the concept of "coordinative structures" or "synergies" that has been proposed as the means by which a potentially large number of articulatory degrees of freedom can be significantly reduced to enable the efficient production of phonetically relevant gestures (e.g., Kelso *et al.*, 1986; Fowler and Saltzman, 1993). In a review of movement control and speech production literature, Löfqvist (1997) defines coordinative structures as "…linkages between muscles that are set up for the execution of specific tasks," and further that they are "…a set of constraints between muscles that are set up to make the set of muscles behave as a unit." Thus, the observed relation between tongue shaping patterns (e.g., factors) and muscle activations appears to represent some underlying synergy of muscles controlled as a functional unit for the production of speech. Whether or not the nature of such functional units is specific to speech is still an open question, however. Perrier *et al.* (2000) proposed that the two degrees of freedom expressed by factor analyses of the tongue have an anatomical and biomechanical basis. Using a model of the tongue that included a representation of the lingual musculature, they showed that much of the variation in tongue shape (84%) allowed by the model could be described by two factors with shapes similar to those reported by Nix *et al.* (1996). Thus, biomechanical constraints, as well as task-specific constraints, may play a role in reducing the degrees of articulatory freedom.

While tongue shape for vowel production apparently has a consistent, systematic description, similar analyses of area functions (i.e., cross-sectional area of the upper airway as a function of distance from the glottis) have indicated that control of the vocal tract shape may also be characterized by a small set of spatial deformation patterns that extend over the entire length of the vocal tract. Story and Titze (1998) applied a principal components analysis (PCA) to a set of MRI-based area functions for ten vowels from an adult male speaker. The two most significant orthogonal components accounted for nearly 88% of the total variance in the set of area functions. When the first component was superimposed on the mean area function in isolation and weighted with its minimum (negative) coefficient, the area function approximated an [i] vowel and produced widely spaced F1 and F2 formant frequencies. Conversely, when weighted with its maximum coefficient, an [ɑ]-like shape and closely spaced F1 and F2 frequencies were produced. Similar isolated weightings of the second component showed that it specified area functions and formants approximately representative of [æ] and [o]. Combinations of incremental coefficient weights for both components led to development of an area function model that can generate a wide variety of vocal tract shapes that were not included in the original set of area functions, and produce a nearly one-to-one mapping between component coefficients (weights) and the F1-F2 vowel space (Story and Titze, 1998; Story, 2005). Other studies based on principal components analysis of area functions derived from sagittal x-ray images (Meyer *et al.*, 1989; Yehia *et al.*, 1996) have also shown that two to five components account for the majority of the variance in their respective data sets.

It is noteworthy that a similar system of spatial shaping patterns has been determined for hand postures. Santello *et al.* (1998) recorded spatial coordinates from sensors on subjects' hands while they grasped 57 different objects with wide variations in shape (analogous to speakers producing a variety of vowel sounds in an MRI scanner). A PCA performed on the coordinates indicated that, for each subject, two principal components could account for greater than 80% of the variance in hand posture. A model for controlling the general shape of the hand was developed based on superimposing coefficient weighted components, either in isolation or combination, on the mean hand shape. In this manner, many hand shapes not in the original set could be produced.

Santello *et al.* note that if control of the general shape of the hand were based on the principal components, the number of degrees of freedom would be significantly reduced. Drawing on the concepts of Macpherson (1991), they further suggested that each component represents a postural "synergy" between individual muscles of the hand that can be utilized alone or in combination with other synergies (i.e., another principal component).

An area function representation of the vocal tract is, strictly, only a description of the nonuniform shape of a tubular structure and cannot necessarily be considered an "articulatory" representation. Based on definitions of coordinative structures (e.g., Löfqvist, 1997), and also by analogy to Santello *et al.* (1998), however, it is conceivable that the principal components derived from a set of vowel area functions (e.g., Story and Titze, 1998) could represent some form of synergy of muscles associated with individual articulators, that facilitates the production of predictable patterns of formant frequencies (cf. Story, 2004). While speculative at this point, the idea is consistent with Gracco's (1992) statement that speech motor control is apparently "organized at a functional level according to sound-producing vocal tract actions." Story and Titze (1998) referred to the principal components in their model as "modes" to emphasize a similarity to a modal decomposition of a dynamical system into natural modes. Hence, to the degree that they capture some aspect of possible muscle synergies, the modes may be conceptualized as "synergistic modes."

To date, modes or principal components reported for area functions measured with 3-D imaging techniques (e.g., MRI) have been based on only one adult male speaker's area function set (Story *et al.*, 1996). While they provide an efficient parametric representation of the vocal tract area function under both normal and perturbed conditions (Story, 2004, 2005), it is necessary to obtain mode shapes for additional speakers to determine if the concept of area function modes can be generalized. This study was motivated by a simple question: Are synergistic modes, based on area functions, similar across speakers? The specific aims of the project were to (1) acquire area functions for vocal tract shapes of vowels from six speakers (three adult males and three adult females) using magnetic resonance imaging (MRI), (2) compare formant frequencies extracted from acoustic recordings to those calculated for each area function, (3) determine mode shapes and coefficient weights for each speaker's area function set and compare across speakers, and (4) calculate the effects of the modes on the first two formant frequencies (F1 and F2) for each speaker.

II. ACQUISITION AND ANALYSIS OF IMAGES AND AUDIO SAMPLES

A. Image collection

Magnetic resonance imaging (MRI) was used to obtain volumetric image sets for a variety of vocal tract shapes from three male and three female speakers. The vocal tract shapes corresponded to each speaker's production of the American English vowels [i, ɪ, e, ɛ, æ, ʌ, ɑ, ɔ, o, ʊ, u]. Image sets for [l], [ɝ], the nasal tract, and trachea were also obtained but will not be reported in this article.

TABLE I. The age of each speaker at the time of MR scanning and the geographic area they reported as being most representative of their "growing up" years. The female speakers are denoted by "SF" and the male speakers by "SM."

Speaker	Age (years)	Region
SF1	29	N. Carolina, USA
SF2	25	Oklahoma, USA
SF3	23	Manitoba, Canada
SM1	33	N. Carolina, USA
SM2	41	New York (Long Island), USA
SM3	30	Los Angeles, CA, USA

The six speakers were recruited from the student, faculty, and staff population at the University of Arizona. They will be identified in this article as SF1, SF2, SF3, SM1, SM2, and SM3, where the "F" denotes *female* and "M" *male*. None of the speakers had any specialized voice or speech training (e.g., singing, acting, etc.), although it may be noted that all three female speakers were students in the Department of Speech, Language, and Hearing Sciences. Five of the speakers had no history of speech, language, or hearing disorders. SM2 reported a cleft of his soft palate that had been surgically repaired at approximately 2.5 years of age. At the time of image collection his speech was judged to be normal. The age of each speaker is shown in Table I, along with the geographic area they reported as being most representative of their "growing up" years.

Prior to image collection, each speaker participated in three practice/training sessions. The purpose was to present the speakers with some of the conditions that are experienced in the MR scanner and to provide them with ample time to practice producing the set of speech sounds under these conditions. They were first given earplugs to partially simulate the limited auditory feedback conditions in the MR scanner. Then while lying supine on a cushioned table in a sound treated room, they practiced sustaining each speech sound. Throughout each session an emphasis was placed on the concentration required to maintain a steady vocal tract shape. A high-quality digital audio tape recording of the third session was made and later used for formant frequency analysis.

The MR images were acquired with a General Electric Signa 1.5-T scanner at the University of Arizona Medical Center. The data acquisition mode was fast spin echo and the scanning parameters were set to TE=13 ms, TR=4000 ms, ETL=16 ms, and NEX=2. During each speaker's production of a particular vocal tract shape, a 24–30 slice series was collected with an interleaved acquisition sequence. Each image set consisted of contiguous, parallel, axial sections (slices) extending from a location just superior of the hard palate to an inferior location near the first tracheal ring. The field of view (slice dimensions) for each slice was 24 cm $\times$ 24 cm which, with a pixel matrix of 256$\times$256, provided an in-plane spatial resolution of 0.938 mm/pixel. Image collection for the female speakers was performed in late 2001 and early 2002, and at that time only a flexible anterior neck

coil was available. By the time that the male speakers were imaged in the summer of 2003, a new rigid, anterior/posterior neck coil had been obtained and was used for the acquisition of their image data. The newer coil produced slightly better image quality and allowed thinner slices to be obtained. Thus, the image slice thickness was 5 mm for all of the female data and 4 mm for the male data.

After the subject had been positioned in the MR scanner [see Story *et al.* (1996, 1998) for more details], the image acquisition protocol proceeded as follows. For a specific vocal tract shape, a corresponding example word (hVd context for each vowel, "luck" and "earth" for [l] and [ɜ˞], respectively) was spoken to the subject over the intercom. The subject was then asked to produce a sustained version of the particular speech sound that the investigator verified before beginning the actual image acquisition. When the subject was ready he/she began phonation and the MR technologist followed by initiating the scan. After 8 s the scan was paused to allow time for the subject to breathe. The scanning was continued when the subject resumed phonation. The scanning time required for each image set (i.e., for one complete tract shape) was 4 min and 16 s which required approximately 30 repetitions. With pauses for respiration between repetitions, each image set was completed in about 10–15 min. Because of the potential for vocal and general physical fatigue, the image collection for each speaker was separated into at least two sessions which occurred different days.

B. Image analysis

The first step in the image analysis procedure was the application of an airway segmentation technique and was followed by shape-based interpolation to generate a 3-D reconstruction of each vocal tract shape [identical to methods reported in Story *et al.* (1996, 1998)]. Based on raw images and 3-D reconstructions, the location of the glottis was determined, and a point just above it was identified as the inlet to the vocal tract. Cross-sectional areas between this point and the lip termination were obtained by first finding the centerline through the 3-D reconstruction with an iterative bisection algorithm (see Story *et al.*, 1996, p. 542). Areas were then measured from oblique sections calculated to be locally perpendicular to the centerline. The collection of these areas extending from just above the glottis to the lips, along with the distance of each cross section from the glottis, comprises the area function. Each area function was subsequently resampled with a cubic spline from which 44 area sections[1] were obtained at equal length increments. A smoothing filter was also applied to remove small discontinuities assumed to be imaging artifacts (see Story *et al.*, 2001, p. 1653).

The piriform sinuses were segmented in each image set and included in the shape-based interpolation. Their shapes and sizes, however, were not generally consistent across the vowels of a given speaker and were, in many cases, discontinuous with the main vocal tract. While the vowel dependency of the piriform sinuses may be an interesting study in itself, including this component was considered to be beyond the scope of the present study. Hence, their cross-sectional areas are not reported here.

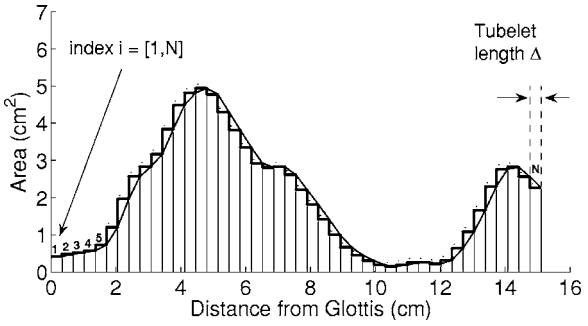

FIG. 1. Example of an area function ([i] vowel produced by SF1) shown as a succession of tubelets, denoted by the index i, extending from just above the glottis to the lips. The tubelet length is Δ and is shown at the rightmost section.

An area *function* can be considered to be comprised of two components, an area vector and a length vector. The area *vector* $A(i)$ contains the $N=44$ cross-sectional areas, assumed to represent a concatenation of "tubelets" ordered consecutively from glottis to lips. The index i denotes this ordering and extends from 1 to N. Similarly, a length vector $L(i)$ contains $N=44$ elements representing the length Δ of each tubelet (i.e., distance increment between consecutive x-sect. areas).[2] A cumulative length vector $\mathcal{X}(i)$ representing the actual distance from the glottis can be derived from L as

$$\chi(i) = \sum_{z=1}^{i} L(z), \quad i = [1,N]. \tag{1}$$

An area function can be shown graphically by plotting $A(i)$ versus $\mathcal{X}(i)$. An example is given in Fig. 1 for speaker SF1's [i] vowel where the "stair-step" plot indicates the discretization of the vocal tract shape into consecutive tubelets, each of length Δ. For graphical clarity in subsequent plots, the area functions will be shown with a smooth curve rather than in stair-step fashion (e.g., see the thin, smooth line in Fig. 1).

C. Audio recording, analysis, and theoretical calculation of formants

As stated in Sec. II A, the third practice session for each speaker was recorded onto digital audio tape and served as the high-quality recording used for formant frequency analysis. The audio signal was transduced with an AKG CK92 microphone that was positioned 30 cm from the speaker and off-axis at 45°. An electroglottographic (Glottal Enterprises, EG-2) signal was simultaneously recorded on the second channel of the digital tape. This signal was intended to be used in later studies for the accurate extraction of fundamental frequency and detection of voicing.

The recordings were transferred to digital files and then downsampled from 44.1 to 22.05 kHz. Formant frequencies were obtained with a LPC algorithm (autocorrelation method) written in Matlab that first estimated the frequency response over a 25-ms window and then found the formants with a peak-picking technique. For the vowels of the female speakers, 16–20 LPC coefficients were used, whereas, for

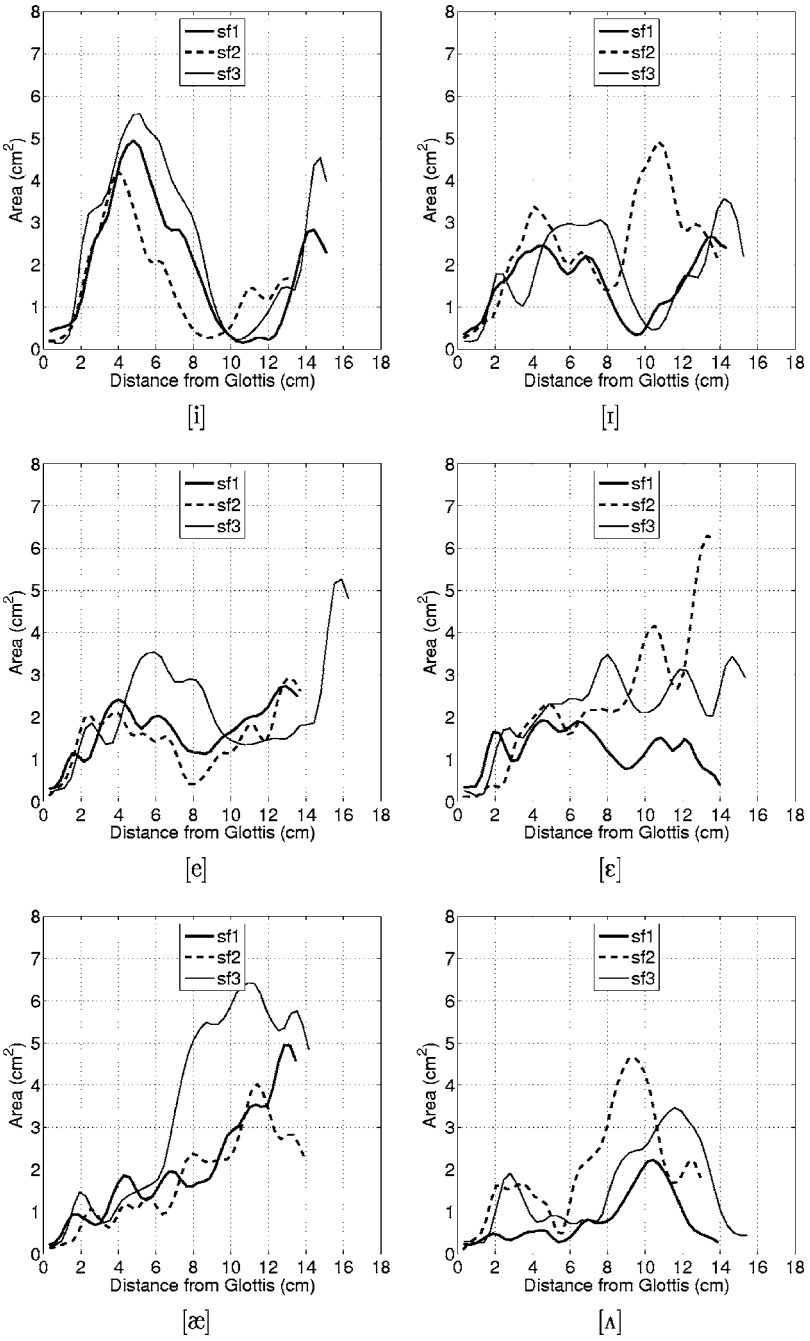

FIG. 2. Area functions for the six vowels, [i, ɪ, e, ɛ, æ, ʌ], produced by three female speakers. Each plot shows one of the vowels for all three speakers and is identified with the corresponding IPA symbol. The thick solid line represents SF1, the dashed line represents SF2, and the thin solid line is for SF3.

the male vowels 26 coefficients were used. This technique was applied to consecutive 25-ms windows (with 12.5-ms overlap) over the time course of each vowel sample which ranged from 4 to 8 s. Thus, with more than 150 sets of formant values obtained for each vowel, the standard deviations reported in Sec. III refer to the variability of a given formant over the time course of a sustained vowel. For each vowel, the mean fundamental frequency was also estimated.

Frequency response functions were calculated for each area function with a frequency-domain technique based on cascaded "ABCD" matrices (Sondhi and Schroeter, 1987; but specifically as presented in Story *et al.*, 2000). This calculation included energy losses due to yielding walls, viscosity, heat conduction, and radiation. Formant frequencies were determined by finding the peaks in the frequency response functions.

III. AREA FUNCTIONS AND FORMANT FREQUENCIES

Area functions for vowels produced by the six speakers are presented in Figs. 2–5. Within each figure panel, three area functions are plotted that correspond to a particular vowel for either the three female speakers (Figs. 2 and 3) or the three male speakers (Figs. 4 and 5). These data are tabulated numerically in vector form in Appendix A.

A. Female speakers

A general observation that can be made from Figs. 2 and 3 is that the vocal tract length for SF2 was consistently shorter than the other two speakers. She, in fact, had a mean tract length of 13.8 cm, whereas for SF1 and SF3 the mean lengths were 14.4 and 15.4 cm, respectively. Taking the length differences into account, the gross shape of each

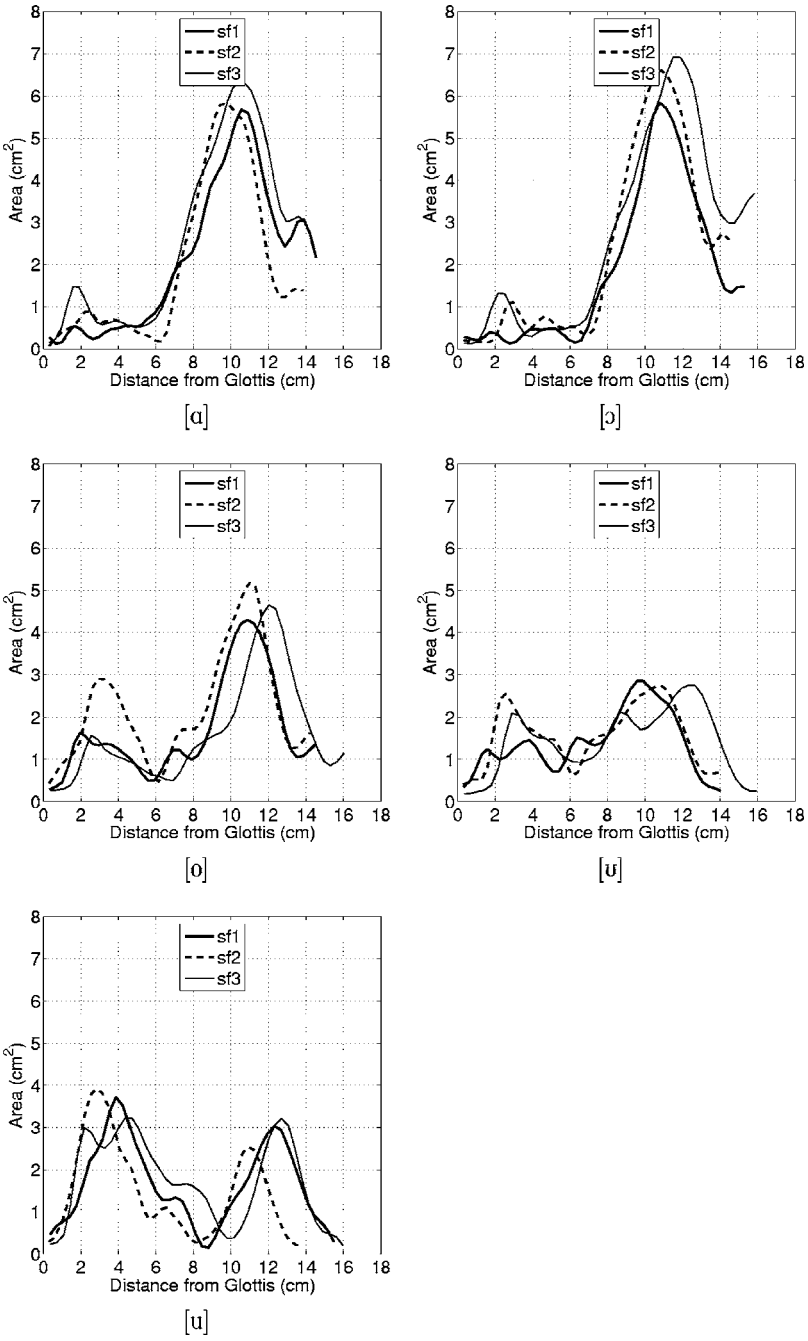

FIG. 3. Area functions for the five vowels, [ɑ, ɔ, o, ʊ, u], produced by three female speakers. Each plot shows one of the vowels for all three speakers and is identified with the corresponding IPA symbol. The line-type assignments are the same as in Fig. 2.

vowel, in terms of the location of major constrictions and expansions, is fairly similar across the three speakers. Some notable exceptions are the wide opening at the lips for SF3's [e] and SF2's [ɛ], the large oral cavity areas for SF3's [æ], and the wide lower pharyngeal space of SF2's [o]. The shape of area functions for the corner vowels, [i, æ, ɑ, u], all reasonably fit expected, prototypical vocal tract configurations of high/low and front/back vowels. The area functions for vowels that exist between the "corners" tend to be somewhat more unique to the speaker.

All three female speakers produced relatively small areas just superior to the glottis. Often referred to as the epilaryngeal space or tube, cross-sectional areas in this region averaged about 0.25 cm^2. The artificial speaking conditions imposed by the MR scanner (e.g., supine position, highly restricted movement, interfering sounds) likely affect the configuration in this part of the vocal tract because they force the speaker to produce "loud" speech (Story *et al.*, 1998). But a constricted epilarynx has also been theoretically implicated as a means by which phonation threshold pressure may be reduced (Titze and Story, 1997). Thus, it is possible that speakers may constrict this part of the vocal tract to a similar degree during normal speech production.

Comparisons of measured and calculated formant frequencies for the female speakers are given in Tables II–IV. The first row in each table contains the mean fundamental frequency produced during the sustained production of a particular vowel. Measurements of the first three formant frequencies and their respective standard deviations are shown in the next six rows. Each formant is labeled with the superscript *N* to denote "natural" speech. In the next three rows are the calculated formant frequencies based on the area

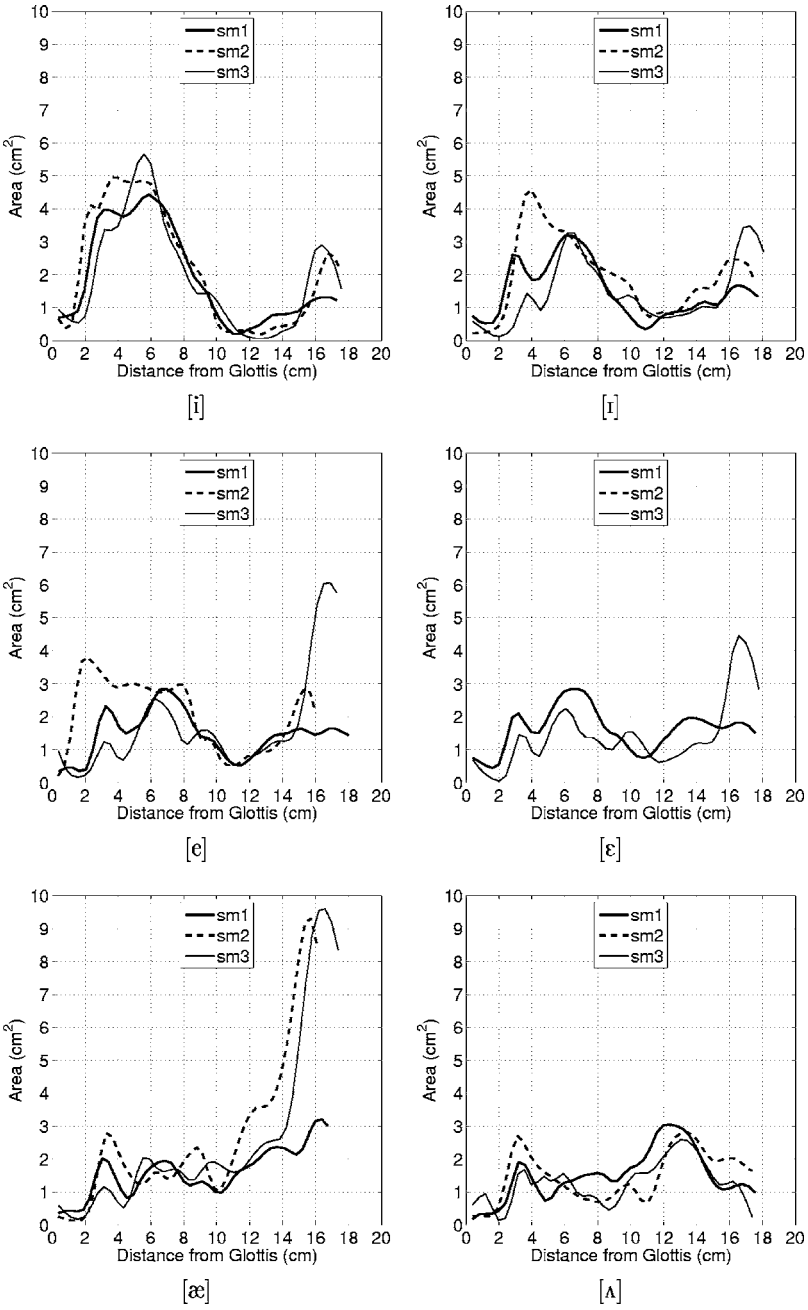

FIG. 4. Area functions for the six vowels, [i, ɪ, e, ɛ, æ, ʌ], produced by three male speakers. Each plot shows one of the vowels for all three speakers and is identified with the corresponding IPA symbol. The thick solid line represents SM1, the dashed line represents SM2, and the thin solid line is for SM3. For reasons explained in the text, an area function for SM2's [ɛ] is not shown.

functions; a superscript *C* is used to denote that they are "calculated." Finally, the percentage errors between measured and calculated formants are given in the bottom three rows. Positive errors indicate that a calculated formant was an overestimation of its measured counterpart, whereas negative errors suggest the opposite.

Across the three speakers, the tabulated errors range from an absolute minimum of −0.1% for the second formant of SF2's [ɛ] vowel to a maximum of 31.4% for the third formant of SF3's [i]. While a large error such as 30% or more is undesirable, it is noted that over half of the errors shown in the tables are less than 10%, and more than a third are less than 5%. It is also noted that the smallest combined error occurs for each speaker's [æ] vowel. This may be coincidental but perhaps there is some characteristic of [æ], either acoustic or articulatory, that facilitates a more consistent production than other vowels.

B. Male speakers

The area functions for the three male speakers are shown in Figs. 4 and 5. Consistent with the plots for the females, each figure panel contains three area functions corresponding to a particular vowel, except for the [ɛ] in Fig. 4, where area functions are shown only for speakers SM1 and SM3. Because of movement artifact, the 3-D reconstruction of SM2's [ɛ] vowel produced a discontinuous air space, and hence an area function could not be determined.

The [i] vowel was produced by all three male speakers with similar vocal tract configurations, both in terms of cross-sectional area distribution and vocal tract length. The expanded pharynx and constricted oral cavity are prototypical for an [i] vowel. There were, however, slight differences in the location and extent of the oral cavity constrictions which, because of their small cross-sectional areas, could

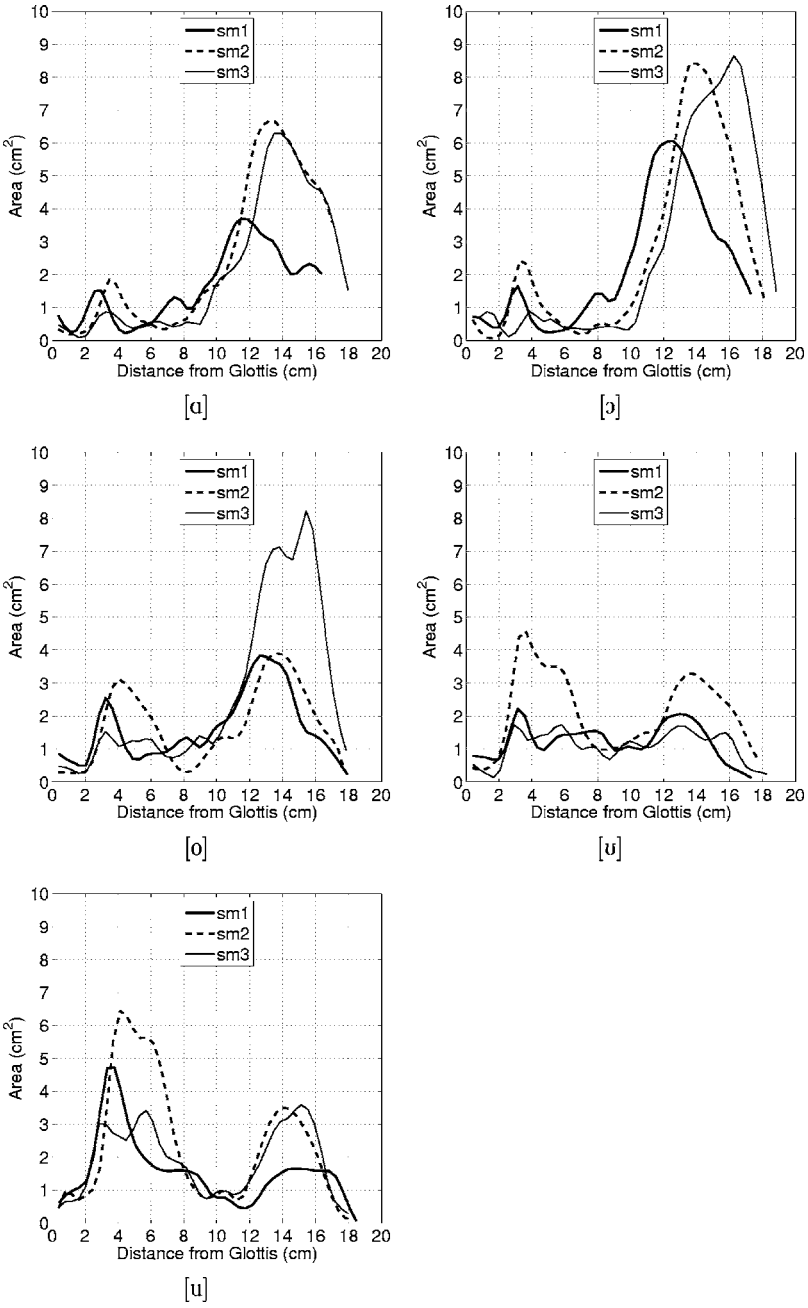

FIG. 5. Area functions for the five vowels, [ɑ, ɔ, o, ʊ, u], produced by three male speakers. Each plot shows one of the vowels for all three speakers and is identified with the corresponding IPA symbol. The line-type assignments are the same as in Fig. 4.

have significant effects on the acoustic characteristics. The gross shape of the area functions for the other corner vowels, [æ, ɑ, u], were also fairly prototypical, but were produced quite differently by each speaker. In particular, the cross-sectional area within the oral cavity for SM1's version of these vowels, as well as for [ɔ], is considerably smaller than in those of the other two speakers. SM1 also exhibited a comparatively shorter tract length for both the [ɑ] and [o] vowels. The vowels [i], [e], [ɛ], and [ʌ] were similarly shaped for each speaker, although SM2 does maintain slightly larger areas in the lower pharyngeal portion for these vowels. For the [ʊ] vowel, SM2 produced a tract shape with an expansion in the pharynx and oral cavity somewhat like an [u], whereas SM1 and SM3 kept the cross-sectional area nearly constant except at the glottal and lip ends.

Similar to the females, the cross-sectional areas representing the epilaryngeal portion of the vocal tract were less than 1 cm^2 for all of the vowels produced by each speaker. The mean area in this region, however, was roughly 0.5 cm^2, which is about twice that measured for the female speakers. Also in comparison to the females, the epilarynx for each male was more distinctly shaped like a tube. This was true for most of vowels where the area remained on the order of 0.5 cm^2 over a distance of about 1.5–2 cm from the glottis. Beyond this point, the cross-sectional area tended to increase rapidly.

Comparisons of measured and calculated formant frequencies are shown in Tables V–VII. The data are arranged in exactly the same format as they were for the female speakers. The absolute minimum error is −0.1% for the first formant in SM2's [e] and [ɔ] vowels. The maximum error of 48.5% occurs in the second formant for SM3's [u]; the next largest error is −28.8% for the third formant of SM3's [e]. Notwithstanding these largest errors, the match between

TABLE II. Fundamental frequencies F_0 and measured and calculated formants for the 11 vowels of speaker SF1. Each measured formant (denoted by superscript "*N*") is the mean across several seconds of recording and *sd* is the standard deviation. The calculated formant values are denoted by "*C*." The Δ's represent the percent error of the computed formants relative to the mean value of the natural speech formants (e.g., $\Delta 1 = 100(F1^C - F1^N)/F1^N$).

	i	ɪ	e	ɛ	æ	ʌ	ɑ	ɔ	o	ʊ	u
F_0	280	276	278	275	267	264	245	251	266	268	286
$F1^N$	391	654	704	770	903	843	917	811	773	721	391
sd	±16	±11	±8	±8	±9	±7	±9	±11	±9	±24	±16
$F2^N$	2720	2269	2404	2240	2087	1574	1484	1242	1326	1602	1163
sd	±13	±44	±46	±26	±25	±10	±8	±9	±6	±12	±23
$F3^N$	3332	3264	3383	3152	3157	3129	3366	2985	2977	3549	2855
sd	±17	±23	±107	±59	±102	±33	±57	±26	±38	±97	±79
$F1^C$	340	537	706	555	901	655	842	737	663	511	417
$F2^C$	2757	2311	2156	1838	2044	1461	1679	1247	1296	1452	1081
$F3^C$	4235	2849	3247	2852	3114	3228	3077	2932	3290	3221	3167
Δ1	−13.0	−18.0	0.2	−28.0	−0.2	−22.3	−8.1	−9.1	−14.2	−29.1	6.5
Δ2	1.3	1.8	−10.3	−18.0	−2.0	−7.2	13.1	0.4	−2.3	−9.3	−7.1
Δ3	27.1	−12.7	−4.0	−9.5	−1.4	3.2	−8.6	−1.8	10.5	−9.2	10.9

TABLE III. Fundamental frequencies F_0 and measured and calculated formants for the 11 vowels of speaker SF2.

	i	ɪ	e	ɛ	æ	ʌ	ɑ	ɔ	o	ʊ	u
F_0	249	249	249	247	248	246	248	249	247	247	250
$F1^N$	446	599	639	718	892	741	862	842	643	666	460
sd	±6	±4	±11	±13	±7	±17	±7	±10	±9	±20	±10
$F2^N$	2875	2261	2445	2156	2004	1616	1337	1184	1240	1483	1319
sd	±60	±13	±13	±36	±16	±32	±17	±15	±11	±5	±34
$F3^N$	4452	3209	3193	3159	3080	3437	3266	3230	3082	3252	3012
sd	±31	±10	±25	±29	±31	±59	±28	±13	±32	±34	±36
$F1^C$	457	710	622	910	878	773	711	839	619	569	412
$F2^C$	2512	1721	2152	2155	1828	1588	1166	1324	1295	1442	1235
$F3^C$	3591	3549	3255	3703	3137	3926	4000	3650	3570	3472	3310
Δ1	2.6	18.6	−2.6	26.8	−1.6	4.3	−17.6	−0.4	−3.7	−14.5	−10.5
Δ2	−12.6	−23.9	−12.0	−0.1	−8.8	−1.7	−12.8	11.8	4.4	−2.8	−6.3
Δ3	−19.3	10.6	2.0	17.2	1.8	14.2	22.5	13.0	15.8	6.8	9.9

TABLE IV. Fundamental frequencies F_0 and measured and calculated formants for the 11 vowels of speaker SF3.

	i	ɪ	e	ɛ	æ	ʌ	ɑ	ɔ	o	ʊ	u
F_0	280	272	280	275	277	279	273	273	270	274	267
$F1^N$	368	557	694	771	820	617	708	683	603	631	433
sd	±5	±14	±14	±19	±44	±25	±37	±37	±5	±15	±16
$F2^N$	2828	2180	2277	2024	1712	1432	1329	1231	1017	1353	1144
sd	±47	±45	±22	±38	±55	±25	±36	±29	±9	±18	±46
$F3^N$	3552	3270	3213	3221	3253	3117	3299	3299	3086	3162	2881
sd	±46	±45	±23	±60	±62	±41	±21	±34	±43	±62	±25
$F1^C$	371	536	563	672	877	773	848	775	614	458	391
$F2^C$	2914	2152	2094	1914	1747	1588	1366	1216	1190	1331	1153
$F3^C$	4666	2525	2801	2906	3049	3926	3365	3015	3036	2845	2976
Δ1	0.7	−3.8	−18.9	−12.8	6.9	25.3	19.8	13.5	1.9	−27.4	−9.5
Δ2	3.0	−1.3	−8.0	−5.4	2.1	10.9	2.8	−1.2	17.1	−1.7	0.7
Δ3	31.4	−22.8	−12.8	−9.8	−6.3	25.9	2.0	−8.6	−1.6	−10.0	−7.1

TABLE V. Fundamental frequencies F_0 and measured and calculated formants for the 11 vowels of speaker SM1.

	i	ɪ	e	ɛ	æ	ʌ	ɑ	ɔ	o	ʊ	u
F_0	122	126	125	126	124	125	125	123	124	123	126
$F1^N$	311	449	495	496	566	521	640	631	493	463	326
sd	±5	±5	±4	±4	±4	±4	±6	±6	±3	±5	±3
$F2^N$	2086	1765	1729	1693	1598	1236	1105	925	912	1137	928
sd	±20	±19	±18	±38	±7	±8	±26	±12	±20	±24	±13
$F3^N$	2570	2299	2278	2272	2172	2246	2465	2513	2356	2277	2282
sd	±23	±29	±30	±33	±12	±11	±29	±20	±16	±32	±13
$F1^C$	327	414	470	503	638	552	701	636	442	390	314
$F2^C$	2032	1771	1650	1599	1671	1361	1184	984	1047	1211	954
$F3^C$	2514	2418	2356	2400	2497	2605	2596	2596	2488	2366	2235
Δ1	5.2	−7.7	−5.0	1.4	12.7	5.9	9.5	0.9	−10.3	−15.8	−3.8
Δ2	−2.6	0.3	−4.6	−5.5	4.6	10.1	7.1	6.4	14.8	6.5	2.8
Δ3	−2.2	5.2	3.4	5.7	15.0	16.0	5.3	3.3	5.6	3.9	−2.1

TABLE VI. Fundamental frequencies F_0 and measured and calculated formants for the 11 vowels of speaker SM2.

	i	ɪ	e	ɛ	æ	ʌ	ɑ	ɔ	o	ʊ	u
F_0	144	145	144	143	141	142	139	138	135	137	135
$F1^N$	289	428	425	547	606	568	659	569	406	515	352
sd	±1	±2	±4	±3	±4	±3	±4	±2	±11	±10	±11
$F2^N$	2215	1866	1983	1740	1719	1329	1180	875	799	974	818
sd	±17	±6	±13	±12	±10	±19	±11	±9	±21	±19	±6
$F3^N$	2729	2429	2454	2664	2606	2675	2678	2924	2867	2892	2424
sd	±45	±7	±10	±14	±12	±22	±24	±22	±14	±45	±24
$F1^C$	294	446	425		748	559	742	568	462	457	286
$F2^C$	2317	1845	2013		1816	1340	1170	837	942	1185	864
$F3^C$	2651	2577	2624		2648	2691	2928	2338	2979	2870	2934
Δ1	1.6	4.2	−0.1		23.5	−1.6	12.6	−0.1	13.7	−11.3	−18.8
Δ2	4.6	−1.1	1.5		5.6	0.8	−0.9	−4.4	17.9	21.7	5.6
Δ3	−2.9	6.1	6.9		1.6	0.6	9.4	−20.0	3.9	−0.8	21.1

TABLE VII. Fundamental frequencies F_0 and measured and calculated formants for the 11 vowels of speaker SM3.

	i	ɪ	e	ɛ	æ	ʌ	ɑ	ɔ	o	ʊ	u
F_0	134	129	129	131	126	122	122	122	123	124	125
$F1^N$	307	499	519	538	652	619	678	627	498	514	355
sd	±7	±3	±1	±4	±7	±3	±4	±3	±3	±9	±5
$F2^N$	2171	1773	1720	1723	1595	1244	1041	866	845	1044	696
sd	±8	±17	±12	±28	±6	±7	±23	±4	±12	±27	±14
$F3^N$	3224	2322	3311	2367	2345	2142	2134	2276	2413	2319	2338
sd	±45	±18	±11	±24	±35	±16	±15	±14	±29	±69	±32
$F1^C$	265	503	567	563	758	495	722	637	611	429	404
$F2^C$	2428	1897	1903	1662	1753	1134	1089	875	1068	1207	1034
$F3^C$	3328	2187	2356	2020	2304	2122	2489	2044	2591	2503	2618
Δ1	−13.6	0.7	9.2	4.6	16.3	−20.0	6.5	1.6	22.7	−16.6	13.9
Δ2	11.9	7.0	10.7	−3.6	9.9	−8.8	4.6	1.1	26.4	15.6	48.5
Δ3	3.2	−5.8	−28.8	−14.6	−1.7	−0.9	16.6	−10.2	7.4	8.0	12.0

measured and calculated formants for all of the vowels produced by each of the three speakers is reasonably accurate. Nearly two-thirds of the errors were calculated to be less than 10% and more than a third were less than 5%.

IV. "SYNERGISTIC" MODES

Based on visual inspection, and comparison of calculated and measured formant frequencies, it was assumed that the six sets of area functions provide a reasonable "sampling" of possible vocal tract configurations used by each speaker to produce these vowels. From this "sampling," a set of *modes*, as defined in the Introduction, are reported for *each* speaker. The similarity of the modes across speakers was assessed visually, with correlation analysis and with an acoustic mapping technique.

A. Principal components analysis

The collection of area vectors for a given speaker (i.e., Tables XI–XVI) can be represented in matrix form as $A(i,j)$, where i is the area index as defined in Sec. II B, and j denotes the particular vowel in the left-to-right order used in the tables in Appendix A (i.e., $j=1$ for [i], $j=2$ for [ɪ]…, $j=11$ for [u]). A principal components analysis (PCA) was used to derive, for each of the six $A(i,j)$'s, a set of orthogonal eigenvectors representing the prominent vocal tract shaping features utilized by each speaker. Following Story and Titze (1998), the PCA was performed on the equivalent diameters of the cross-sectional areas rather than on the areas themselves.[3] Thus, an area matrix $A(i,j)$ was first converted to a diameter matrix $D(i,j)$ by

$$D(i,j) = \sqrt{\frac{4}{\pi}A(i,j)}. \tag{2}$$

The square root operation has the effect of compressing and expanding the portions of a vocal tract shape with the largest and smallest areas, respectively, and has been shown to produce somewhat more accurate reconstructions of vowel configurations than a PCA performed on areas alone (Story and Titze, 1998).

The next step was to assume that a speaker's $D(i,j)$ can be represented by a mean and variable part,

$$D(i,j) = \Omega(i) + \alpha(i,j), \tag{3}$$

where $\Omega(i)$ is the mean diameter vector across $D(i,j)$, and $\alpha(i,j)$ is the variation superimposed on $\Omega(i)$ to produce a specific diameter vector. The PCA was then carried out by calculating the eigenvectors of a covariance matrix formed with $\alpha(i,j)$. The specific implementation was essentially the same method as reported in Story and Titze (1998), and results in the following parametric representation of the original *area* matrix,

$$A(i,j) = \frac{\pi}{4}\left(\Omega(i) + \sum_{i=1}^{N} q_k(j)\phi_k(i)\right)^2,$$
$$i=[1,N] \quad j=[1,11], \quad k=[1,N] \tag{4}$$

where the $\phi_k(i)$'s are 44-element eigenvectors that, when multiplied by the appropriate scaling coefficients $q_k(j)$, will reconstruct each area vector in $A(i,j)$. As explained in the Introduction, the eigenvectors or principal components obtained from the area vector matrices have been called *modes* and will henceforth be referred to as such.

This particular implementation of the PCA effectively normalizes the vocal tract lengths for each speaker's vowels by excluding the length data in the analysis. That is, for each speaker's $A(i,j)$, and subsequent $D(i,j)$, the ith section was assumed to correspond to the same location within the vocal tract. This is not strictly true because each area function in Tables XI–XV has its own unique length increment Δ, thus the distance of the ith tubelet from the glottis is vowel dependent. Comparisons of the formant frequencies calculated with the actual lengths of each area function to those based on the normalized length showed approximately a 5% shift upward for the longest area function in each speaker's set and a 5% downward shift for the shortest. Whereas this approach somewhat limits the interpretation of the results, especially with regard to rounded vowels which typically have long vocal tract lengths, it simplified the execution of the PCA and facilitated the comparison of modes across speakers.

To further aid interspeaker comparison, the three modes that accounted for most of the variance in each speaker's $D(i,j)$ [referred to as $\phi_1(i), \phi_2(i), \phi_3(i)$] have been smoothed by fitting them with eighth-order polynomials. This simplified the visual comparison of the modes but maintained their gross characteristics. The amount of variance in each speaker's diameter matrix that is accounted for by each mode was calculated based on these smoothed modes.

TABLE VIII. Percentage of the total variance in each speaker's diameter matrix $D(i,j)$ accounted for by the most significant modes (in their smoothed form).

Mode	SF1	SF2	SF3	SM1	SM2	SM3
ϕ_1	63.6	61.1	63.0	76.2	67.4	63.7
ϕ_2	21.5	22.0	28.7	12.3	20.5	25.2
ϕ_3	7.7	7.4	3.0	4.5	4.7	8.4
Total	92.8	90.5	94.7	93.0	92.6	97.3

B. Modes and mean area functions for six speakers

Percentages of the total variance in each speaker's diameter matrix $D(i,j)$ that are accounted for by the three most significant modes are shown in Table VIII. Together, the three modes accounted for more than 90% of the variance for each speaker. The first mode, ϕ_1, accounted for variance of over 60%, while ϕ_2 typically accounted for 20% or more. The exception was SM2 whose second mode accounted for only 12.3% of the variance in his diameter matrix, but was balanced by the high 76.2% that was attributed to his first mode. The first two modes combined accounted for a minimum of 83% of the variance for SF2 and a maximum of 91.7% for SF3. The amount of variance accounted for by ϕ_3

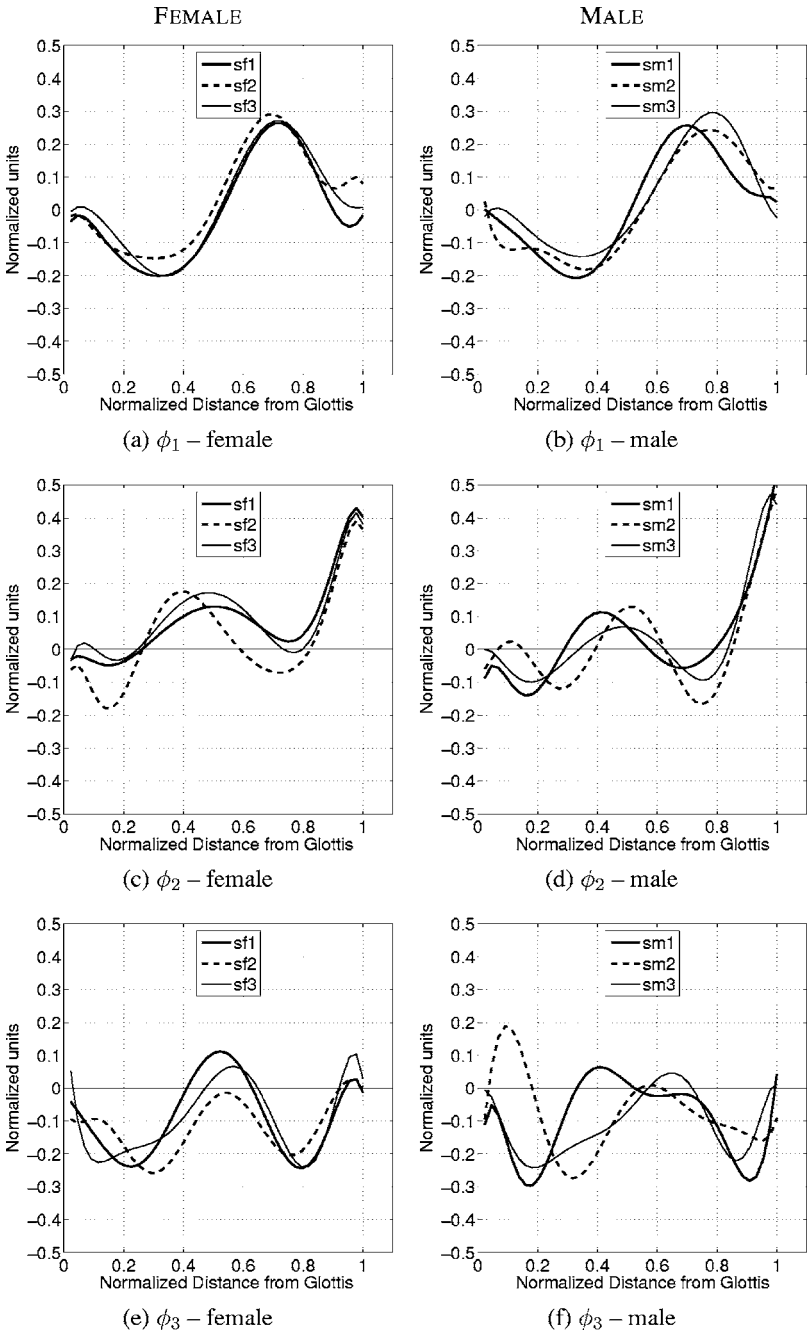

FIG. 6. Three modes obtained from the area function data of the three female and three male speakers. The left column shows plots corresponding to the females (a, c, and e) and the right corresponds to the males (b, d, and f). Each row of the plot is associated with mode ϕ_1, ϕ_2, or ϕ_3.

was no greater than 8.4% for any of the speakers. Thus, the third mode is much less significant than either ϕ_1 or ϕ_2.

The three modes are shown graphically in Fig. 6 and, for reference, are numerically tabulated in Appendix B. This figure is arranged so that the left and right columns of plots correspond to the three female and three male speakers, respectively. Each row corresponds to one of the three modes ϕ_1, ϕ_2, or ϕ_3. Note that the x axis of each plot ranges from 0 to 1.0 to indicate a normalized distance from the glottis. An approximate actual distance could be obtained for each speaker by scaling the x axis with their respective mean vocal tract lengths. Associated with the plots of the three modes are their respective scaling coefficients that correspond to each speaker's vowels. These are given in Table IX, where the minimum and maximum values along each row are shown in boldface.

The first mode, shown in Figs. 6(a) and 6(b), presents similar characteristics for all of the speakers. The amplitude is near zero at both the glottal and lip ends, a zero crossing occurs at approximately the mid-point of the vocal tract, and negative and positive maxima occur at normalized distances of about 0.35 and 0.75, respectively. When scaled with a positive coefficient and superimposed on a mean vocal tract shape, as specified by Eq. (4), this mode would have the effect of constricting the pharyngeal portion of the tract while expanding the oral cavity; the glottal and lip ends would be minimally affected. Such a shape would be characteristic of a low-back vowel. A negative scaling coefficient would impose the opposite changes in vocal tract shape and would be suggestive of a high-front vowel. It is not surprising then that the scaling coefficient for the first mode, q_1, shown in Table IX, is maximally positive for the [ɔ] vowel

TABLE IX. Scaling coefficients of the three modes that will reconstruct each of the 11 vowels. Coefficients are given for all six speakers. The minimum and maximum values along each row are shown in boldface type.

Subject	Coefficient	i	ɪ	e	ɛ	æ	ʌ	ɑ	ɔ	o	ʊ	u
SF1	q_1	**−4.667**	−2.456	−1.206	−1.275	0.359	1.168	3.170	**3.790**	1.572	0.756	−1.211
	q_2	0.770	0.610	0.894	−1.119	**2.455**	**−2.108**	1.619	0.342	−0.485	−1.500	−1.478
	q_3	0.496	0.330	−0.135	0.690	−0.222	**1.406**	0.197	0.250	−1.044	−0.018	**−1.949**
SF2	q_1	**−4.025**	0.070	−2.213	0.930	0.821	1.263	2.969	**3.457**	0.563	−0.692	−3.142
	q_2	0.859	**1.059**	0.761	2.833	0.789	−0.355	−1.311	−0.123	−1.075	−1.485	**−1.952**
	q_3	0.357	**−1.728**	0.852	−0.150	**1.067**	−0.244	0.734	0.332	−1.270	0.230	−0.180
SF3	q_1	**−4.927**	−2.506	−2.228	−0.591	2.999	0.961	3.460	**3.761**	1.378	−0.179	−2.128
	q_2	1.503	0.825	1.618	1.000	**2.842**	−2.290	0.546	0.715	−1.942	**−2.467**	−2.351
	q_3	−0.249	**0.702**	0.384	0.233	−0.608	0.549	−0.347	0.178	0.168	0.499	**−1.508**
SM1	q_1	**−3.860**	−2.192	−1.334	−0.956	0.295	1.169	2.659	**4.260**	1.761	−0.120	−1.683
	q_2	0.131	0.510	0.636	0.717	**1.437**	−0.251	0.542	0.294	−1.206	**−1.448**	−1.360
	q_3	−0.410	0.242	0.400	0.138	−0.184	0.658	0.038	−0.687	−0.001	**0.895**	**−1.090**
SM2	q_1	**−4.777**	−2.109	−3.115	⋯	1.864	0.396	4.175	**4.816**	0.915	−0.449	−1.715
	q_2	0.879	0.529	1.276	⋯	**3.209**	−0.142	0.694	−0.727	−1.621	−1.242	**−2.854**
	q_3	0.298	−0.670	1.039	⋯	**−1.300**	**1.254**	0.084	0.461	0.332	−0.214	−1.285
SM3	q_1	**−5.047**	−2.349	−1.740	−1.682	0.916	−0.235	3.409	**5.066**	3.725	−1.350	−0.714
	q_2	−0.079	0.851	2.176	1.184	**3.531**	−1.979	−0.213	0.298	−0.656	**−2.657**	−2.457
	q_3	−1.551	0.812	0.584	0.901	−0.801	**1.453**	0.709	−0.012	−1.246	0.999	**−1.849**

across all six speakers. Likewise, for every speaker, q_1 has the largest negative value for the vowel [i]. It is also noted that q_1 generally increases stepwise from this most negative value toward the most positive as the vowel changes in the order given in the table. This ordering of vowels follows a counter-clockwise rotation through a articulation-based vowel quadrilateral (e.g., Shriberg and Kent, 2003; p. 28), or clockwise through a typical F1-F2 vowel space plot.

Collections of the second mode, ϕ_2, for the six speakers are shown in Figs. 6(c) and 6(d). Though the similarity across speakers is perhaps not as visually apparent as for the first mode, the ϕ_2's do possess some common features. The lip end of these modes is uniformly of high amplitude, suggesting that it would expand or constrict the mouth opening with positive and negative scaling coefficients, respectively. Moving posteriorly from the lips, each ϕ_2 drops sharply in amplitude and, for five speakers, crosses zero at a normalized distance from the glottis of about 0.8. SF1's second mode does not have a zero crossing at this point but does maintain a shape that is similar to the other speakers. The amplitude then drops further until a local minima is reached at a location ranging from 0.68 to 0.78 across the speakers. Continuing toward the glottal end, a local maxima is encountered for each speaker at a normalized distance between 0.4 and 0.5. As the amplitude drops again, the second mode for each of the females has a zero crossing at a distance of about 0.25. The ϕ_2 amplitudes for the males also drop but their subsequent zero crossings are not as closely aligned as they are for the females. Finally, for all speakers, the ϕ_2 amplitudes dip down to another local minima before returning to nearly zero amplitude at the glottal end. In general, the primary effect of the second mode on vocal tract shape is to expand or constrict both the lip end and the middle portion, depending on whether the scaling coefficient q_2 is positive or negative. From Table IX, it is apparent that the largest positive value of q_2 occurred for the [æ] vowel of every speaker except SF2. Also, with the exception of SF2, the most negative value of q_2 is associated with either the vowel [ʊ] or [u]. Notwithstanding the results for SF2, this suggests that the second mode roughly corresponds to a continuum from low-front to high-back vowels.

The primary features contributed to the vocal tract shape by the third mode, shown in Figs. 6(e) and 6(f) are complementary expansive or constrictive effects in both the pharyngeal and oral cavities. This is most apparent for the female speakers, but approximately true for the male speakers as well. Although only a small percentage of the variance is accounted for by this mode (see Table VIII), the simultaneous expansive effects in the pharyngeal and oral cavities, obtained with a negative scaling coefficient, would seem to be useful for enhancing vocal tract shapes with a mid-tract constriction. This is supported by Table IX where, for four of the six speakers, the largest negative value of q_3 occurred for the [u] vowel.

To provide a numerical assessment of the similarity of the mode shapes, a correlation coefficient was calculated for each speaker's modes relative to those of the other five speakers. The calculation was performed by dividing the covariance of a given pair of modes (e.g., ϕ_1 from SF1 and SF2) by the product of their standard deviations (e.g., Taylor, 1982). Note that because normalized vocal tract length is assumed for all correlation calculations, female modes can be compared to male modes. The correlation coefficients are presented in Table X and are arranged so that female-female, female-male, and male-male comparisons within each mode can be readily observed. The upper part of the table indicates that the shape of the first mode is highly correlated among all of the speakers, which supports the visual comparisons dis-

TABLE X. Matrix of correlation coefficients quantifying the similarity of ϕ_1, ϕ_2, ϕ_3, and $(\pi/4)\Omega^2$ across the six speakers. The table is divided into sections showing female-female, male-male, and female-male comparisons.

		Female			Male		
Quantity	Subject	SF1	SF2	SF3	SM1	SM2	SM3
ϕ_1	SF1	1.00	0.97	0.99	0.98	0.90	0.91
$\bar{R}_1=0.94$	SF2		1.00	0.96	0.99	0.91	0.87
	SF3			1.00	0.99	0.92	0.93
	SM1				1.00	0.93	0.90
	SM2					1.00	0.97
	SM3						1.00
ϕ_2	SF1	1.00	0.90	0.97	0.94	0.86	0.97
$\bar{R}_2=0.91$	SF2		1.00	0.93	0.95	0.74	0.89
	SF3			1.00	0.93	0.90	0.96
	SM1				1.00	0.82	0.94
	SM2					1.00	0.92
	SM3						1.00
ϕ_3	SF1	1.00	0.74	0.90	0.53	0.11	0.59
$\bar{R}_3=0.44$	SF2		1.00	0.74	−0.06	0.41	0.43
	SF3			1.00	0.39	−0.04	0.69
	SM1				1.00	−0.18	0.60
	SM2					1.00	0.25
	SM3						1.00
	SF1	1.00	0.93	0.92	0.64	0.66	0.70
$\frac{\pi}{4}\Omega^2$	SF2		1.00	0.89	0.69	0.74	0.68
	SF3			1.00	0.64	0.57	0.67
$\bar{R}=0.71$	SM1				1.00	0.72	0.52
	SM2					1.00	0.73
	SM3						1.00

cussed previously. The highest correlation coefficients were $R=0.99$ for the comparisons of SF1-SF2, SF2-SM1, and SF3-SM1. The lowest correlation for the first mode was $R=0.87$ resulting from the comparison of SF2-SM3; this was the only R below 0.90. The mean correlation coefficient across all comparisons of the first mode was $\bar{R}_1=0.94$ and is shown in the far left-hand column of the table. The correlation across speakers for the second mode was also fairly high as indicated by the mean coefficient of $\bar{R}_2=0.91$. In addition, 11 of the 15 coefficients were greater than 0.90, while the low was 0.74. As expected from visual inspection, the shape of the third mode is less well correlated across speakers with $\bar{R}_3=0.44$. The female-female comparisons did have R values that were 0.74 or greater, but any of the correlations involving male speakers were typically much lower than 0.70.

Correlation coefficients are also presented in Table X for the mean area vectors of the six speakers. These vectors were generated from Eq. (4), but with the mode coefficients set to zero [i.e., $\pi/4\Omega^2(i)$]. For female-female comparisons, the correlation coefficients ranged from 0.89 to 0.93. These are fairly high, but somewhat less so than their correlations for either ϕ_1 or ϕ_2. The correlations for male-male mean area vectors are comparatively low, and the female-male correlations are lower yet. The mean correlation coefficient across all of the comparisons is $R=0.71$.

The low correlation for the mean area vectors across speakers suggests that they are somewhat more speaker-specific than either of the ϕ_1 or ϕ_2 modes. They are plotted for each female speaker in Fig. 7(a) and for the male speakers in Fig. 7(b). Visually, the three female area vectors are similar to each other, but they also possess some idiosyncratic features such as differences in maximum cross-sectional area and shape of the mid-portion of the vocal tract. For the male speakers, the mean area vectors appear more idiosyncratic than those of the females, as expected from the correlation analysis. For example, the cross-sectional areas within the pharyngeal portion of the vocal tract (i.e., between about 0.2 and 0.4) are similar for SM1 and SM3, but they differ in the oral cavity. In contrast, the oral cavity areas are similar for SM2 and SM3, but their pharyngeal areas differ by about 1.5–2.0 cm^2.

Frequency response functions for the six mean area *functions* were calculated and are plotted in Figs. 7(c) and 7(d). This calculation was performed with the methods described in Sec. II C and with the tubelet length set to the mean of a given speaker's set of Δ's as given in the next to last row of Tables XI–XVI. As indicated by Fig. 7(c), the first two formant frequencies were closely aligned for the female speakers, where F1 ranged from 684 to 744 Hz and F2 from 1675 to 1741 Hz. F1 and F2 were also similar for the male speakers [Fig. 7(d)]. Their F1 values ranged from 536 to 626 Hz, while F2 ranged from 1430 to 1553 Hz. The

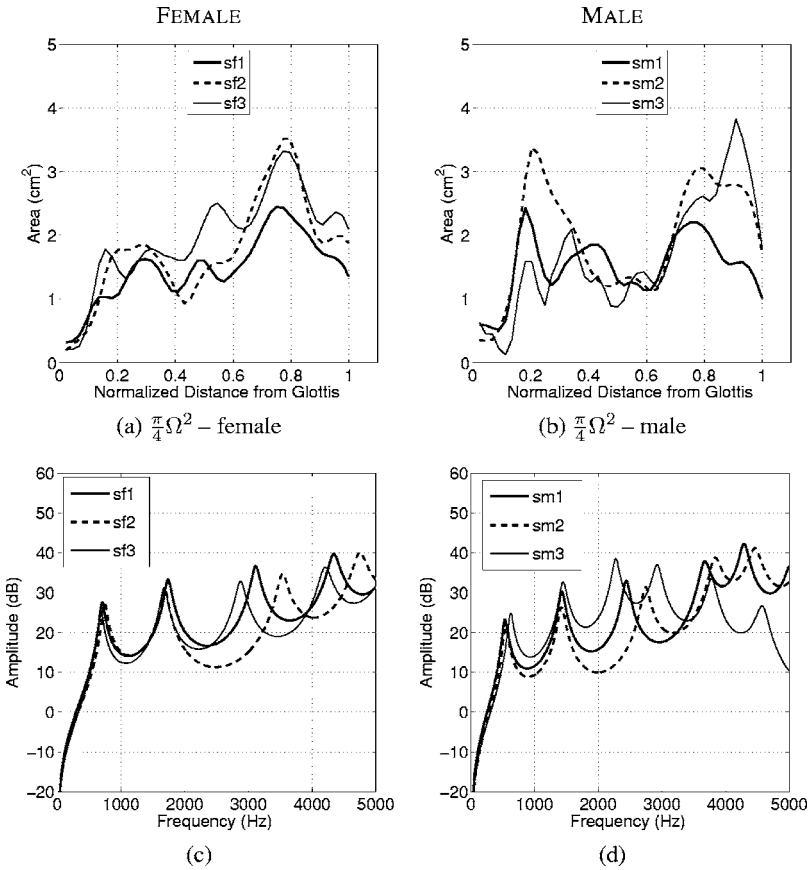

FIG. 7. Mean area vectors and calculated frequency response functions for six speakers. (a) Area vectors for SF1 (thick), SF2 (dashed), and SF3 (thin); (b) area vectors for SM1 (thick), SM2 (dashed), and SM3 (thin); (c) frequency response functions corresponding the female area vectors in (a); and (d) frequency response functions corresponding the male area vectors in (b).

values of F1 and F2 for both female and male speakers are approximately representative of a "neutral" vowel. That is, they are similar to those expected from a uniform tube with length appropriate for each speaker. There were, however, large differences in the locations of the upper formants, a frequency range where the idiosyncratic differences of the area vectors seem to exert their largest effect on the acoustic characteristics. Thus, each speaker's mean area vector could be considered to be a phonetically neutral vocal tract shape with regard to F1 and F2, but speaker-specific for higher frequency formants.

C. Mode coefficient-to-formant mapping

In the previous section, the spatial similarity of the mode shapes across speakers was assessed by visual inspection and with correlation analysis. To a large degree, the vocal tract shape changes imposed by the modes were found to be similar regardless of the speaker. It follows that the modes should also similarly affect the formant frequencies supported by area functions generated with Eq. (4), but perhaps with speaker-specific characteristics due to the underlying neutral shape on which the modes are superimposed.

To demonstrate the acoustic effects of the modes, a mapping was generated for each speaker that linked the three scaling coefficients (q_1, q_2, and q_3) to the first two formant frequencies (F1 and F2). This was motivated by a similar approach used by Story and Titze (1998) for two modes and is capable of producing thousands of area function configurations that did not exist in a given speaker's original area function set. The first step was to generate an equal increment continuum for each mode coefficient that ranged from their respective minimum to maximum values (see Table IX). The increments were specified as

$$\Delta q_1 = \frac{q_1^{\max} - q_1^{\min}}{M-1}, \tag{5a}$$

$$\Delta q_2 = \frac{q_2^{\max} - q_2^{\min}}{N-1}, \tag{5b}$$

$$\Delta q_3 = \frac{q_3^{\max} - q_3^{\min}}{K-1}, \tag{5c}$$

where M, N, and K were the number of increments along each coefficient dimension. Since the first two modes accounted for most of the variance in the original data sets (i.e., Table VIII), it was assumed that they would have the largest acoustic effects. Hence, the sampling along the q_1 and q_2 continua was chosen to be more dense than for q_3, such that M, N, and K were set to 60, 60, and 5, respectively. The coefficient continua were then generated by

$$q_{1m} = q_1^{\min} + m\Delta q_1, \quad m = 0, \ldots, M-1, \tag{6a}$$

$$q_{2n} = q_2^{\min} + n\Delta q_2, \quad n = 0, \ldots, N-1, \tag{6b}$$

$$q_{3k} = q_3^{\min} + k\Delta q_3, \quad k = 0, \ldots, K-1, \tag{6c}$$

with m, n, and k serving as indexes along each continuum.

By modifying Eq. (4) to contain only three modes and to eliminate the dependency on a specific vowel (i.e., jth vowel), an area vector can be generated with

TABLE XI. Area vectors for the 11 vowels of speaker SF1. Each original area function has been segmented to consist of 44 area sections; the length of each section is given by Δ. The glottal end of each area vector is at section 1 and the lip end at section 44. The total vocal tract length (VTL) is computed as 44Δ.

Section i	i	ɪ	e	ɛ	æ	ʌ	ɑ	ɔ	o	ʊ	u
1	0.42	0.34	0.29	0.34	0.22	0.24	0.27	0.28	0.28	0.41	0.46
2	0.48	0.45	0.33	0.35	0.26	0.23	0.12	0.24	0.35	0.47	0.67
3	0.53	0.53	0.51	0.35	0.44	0.25	0.16	0.20	0.46	0.77	0.77
4	0.58	0.66	0.89	0.63	0.72	0.34	0.39	0.27	0.83	1.11	0.88
5	0.74	1.01	1.14	1.21	0.94	0.42	0.55	0.40	1.39	1.24	1.14
6	1.21	1.39	1.08	1.65	0.94	0.49	0.47	0.36	1.63	1.11	1.67
7	1.97	1.57	0.94	1.61	0.85	0.45	0.31	0.21	1.51	0.99	2.21
8	2.58	1.64	1.03	1.26	0.76	0.36	0.23	0.13	1.36	1.05	2.41
9	2.84	1.83	1.39	0.96	0.69	0.34	0.28	0.17	1.33	1.19	2.73
10	3.16	2.15	1.82	0.98	0.73	0.40	0.40	0.33	1.37	1.31	3.38
11	3.84	2.29	2.16	1.29	0.99	0.48	0.47	0.45	1.34	1.41	3.72
12	4.49	2.32	2.34	1.59	1.37	0.53	0.49	0.46	1.26	1.46	3.52
13	4.82	2.43	2.41	1.80	1.68	0.54	0.53	0.45	1.14	1.32	3.06
14	4.94	2.46	2.34	1.92	1.86	0.56	0.56	0.48	1.02	1.08	2.60
15	4.78	2.36	2.14	1.90	1.82	0.54	0.52	0.48	0.85	0.85	2.23
16	4.30	2.17	1.86	1.76	1.56	0.40	0.59	0.36	0.64	0.70	1.85
17	3.82	1.91	1.72	1.65	1.34	0.28	0.73	0.22	0.50	0.71	1.49
18	3.35	1.77	1.84	1.68	1.28	0.31	0.85	0.15	0.51	0.95	1.28
19	2.92	1.87	2.00	1.79	1.36	0.41	1.06	0.20	0.71	1.31	1.28
20	2.82	2.07	2.03	1.90	1.58	0.55	1.39	0.53	1.01	1.51	1.34
21	2.84	2.20	1.96	1.87	1.82	0.73	1.78	1.07	1.22	1.47	1.25
22	2.62	2.14	1.81	1.73	1.96	0.82	2.05	1.47	1.22	1.37	0.93
23	2.22	1.90	1.59	1.59	1.94	0.77	2.15	1.67	1.08	1.33	0.50
24	1.83	1.56	1.35	1.41	1.76	0.74	2.29	1.91	0.99	1.38	0.18
25	1.42	1.20	1.20	1.21	1.60	0.76	2.64	2.26	1.10	1.53	0.15
26	1.00	0.88	1.17	1.05	1.60	0.85	3.29	2.75	1.36	1.78	0.40
27	0.67	0.63	1.16	0.90	1.69	1.03	3.85	3.31	1.75	2.09	0.75
28	0.46	0.45	1.12	0.77	1.72	1.25	4.13	3.91	2.32	2.39	1.07
29	0.31	0.34	1.19	0.79	1.79	1.49	4.41	4.72	2.95	2.67	1.33
30	0.19	0.35	1.37	0.92	2.05	1.74	4.86	5.53	3.48	2.84	1.54
31	0.15	0.54	1.51	1.07	2.44	1.99	5.37	5.84	3.93	2.84	1.81
32	0.19	0.84	1.61	1.27	2.79	2.19	5.67	5.70	4.22	2.72	2.20
33	0.26	1.05	1.70	1.45	2.93	2.23	5.60	5.42	4.30	2.58	2.59
34	0.27	1.11	1.82	1.51	3.03	2.12	5.15	4.97	4.21	2.44	2.87
35	0.22	1.16	1.95	1.37	3.26	1.93	4.44	4.36	3.98	2.33	3.03
36	0.32	1.35	2.01	1.20	3.48	1.65	3.72	3.79	3.62	2.17	2.94
37	0.64	1.61	2.03	1.30	3.54	1.33	3.17	3.33	3.10	1.83	2.56
38	1.09	1.84	2.10	1.48	3.49	1.01	2.69	2.89	2.42	1.44	2.10
39	1.67	2.10	2.26	1.34	3.52	0.72	2.43	2.37	1.71	1.06	1.64
40	2.31	2.42	2.50	1.02	3.92	0.57	2.64	1.82	1.23	0.70	1.20
41	2.77	2.66	2.70	0.80	4.55	0.48	3.00	1.41	1.05	0.46	0.91
42	2.83	2.64	2.73	0.71	4.95	0.43	3.07	1.33	1.08	0.36	0.75
43	2.57	2.48	2.63	0.62	4.94	0.37	2.75	1.46	1.21	0.31	0.58
44	2.28	2.40	2.49	0.38	4.57	0.27	2.15	1.48	1.35	0.26	0.29
Δ	0.343	0.326	0.308	0.318	0.306	0.316	0.330	0.348	0.330	0.319	0.352
VTL	15.11	14.32	13.55	14.01	13.46	13.90	14.53	15.29	14.51	14.02	15.51

$$A_{mnk}(i) = \frac{\pi}{4}[\Omega(i) + q_{1m}\phi_1(i) + q_{2n}\phi_2(i) + q_{3k}\phi_3(i)]^2, \quad i = [1,44]. \tag{7}$$

This suggests that the three continua [Eq. (6)] form a parametric "articulation" space in the sense that any combination of the coefficients produces a vocal tract shape. With q_3 held constant at zero, the continua for the first two modes can be viewed as a 60×60 two-dimensional space consisting of all possible combinations of q_{1m} and q_{2n}. An example, based on SF1's coefficient set, is shown in Fig. 8(a), where every intersection point in the grid represents a $[q_1, q_2]$ pair. The thick solid and dashed lines indicate the continua for the first and second modes, respectively, when the other coefficient is zero. These represent the isolated spatial effect of each mode on the vocal tract shape, while all other lines (vertical or horizontal) in the grid are comprised of a nonzero contribu-

TABLE XII. Area vectors for SF2.

Section i	i	ɪ	e	ɛ	æ	ʌ	ɑ	ɔ	o	ʊ	u
1	0.19	0.27	0.13	0.12	0.15	0.10	0.07	0.20	0.44	0.34	0.29
2	0.19	0.33	0.27	0.11	0.16	0.26	0.24	0.17	0.66	0.51	0.42
3	0.25	0.44	0.39	0.11	0.22	0.36	0.38	0.17	0.88	0.52	0.70
4	0.34	0.59	0.59	0.17	0.27	0.45	0.46	0.17	1.04	0.50	1.15
5	0.53	0.68	0.97	0.33	0.33	0.71	0.53	0.16	1.15	0.82	1.72
6	0.94	0.77	1.52	0.41	0.55	1.23	0.68	0.21	1.40	1.66	2.49
7	1.58	1.10	1.98	0.34	0.90	1.63	0.88	0.59	2.03	2.42	3.28
8	2.19	1.67	2.05	0.42	1.07	1.62	0.88	1.07	2.65	2.55	3.77
9	2.55	2.13	1.87	0.87	0.96	1.53	0.70	1.11	2.90	2.34	3.91
10	2.84	2.31	1.82	1.38	0.74	1.57	0.63	0.84	2.89	2.08	3.83
11	3.38	2.58	1.98	1.67	0.63	1.65	0.67	0.58	2.77	1.83	3.52
12	3.95	3.08	2.11	1.84	0.78	1.67	0.69	0.50	2.61	1.72	2.99
13	4.21	3.38	2.08	1.98	1.06	1.54	0.66	0.66	2.29	1.65	2.49
14	4.09	3.26	1.84	2.13	1.18	1.35	0.57	0.77	1.91	1.53	2.23
15	3.72	3.06	1.58	2.28	1.09	1.28	0.47	0.69	1.66	1.49	2.02
16	3.28	2.83	1.55	2.31	1.10	1.19	0.39	0.55	1.39	1.47	1.58
17	2.75	2.48	1.61	2.18	1.29	0.89	0.34	0.50	0.98	1.28	1.08
18	2.25	2.11	1.54	1.88	1.34	0.52	0.29	0.53	0.59	0.91	0.84
19	2.06	2.00	1.42	1.59	1.13	0.49	0.19	0.51	0.48	0.61	0.89
20	2.11	2.19	1.48	1.64	0.94	0.97	0.17	0.40	0.72	0.69	1.05
21	2.04	2.30	1.55	1.94	1.01	1.62	0.43	0.32	1.16	1.09	1.10
22	1.70	2.14	1.39	2.13	1.31	2.01	1.12	0.46	1.55	1.41	0.95
23	1.32	1.88	1.03	2.17	1.78	2.15	1.92	0.95	1.71	1.52	0.78
24	1.02	1.59	0.64	2.18	2.20	2.26	2.48	1.75	1.71	1.56	0.59
25	0.76	1.40	0.42	2.19	2.39	2.39	2.98	2.53	1.72	1.61	0.35
26	0.53	1.41	0.43	2.15	2.33	2.56	3.55	3.20	1.86	1.71	0.27
27	0.38	1.55	0.54	2.14	2.20	2.96	4.19	3.89	2.20	1.87	0.33
28	0.29	1.99	0.69	2.19	2.16	3.57	4.90	4.59	2.75	2.09	0.41
29	0.26	2.83	0.89	2.35	2.21	4.04	5.48	5.19	3.35	2.33	0.53
30	0.28	3.69	1.09	2.66	2.24	4.37	5.78	5.69	3.81	2.48	0.72
31	0.34	4.14	1.15	3.17	2.24	4.62	5.82	6.14	4.15	2.57	0.99
32	0.40	4.34	1.13	3.72	2.41	4.67	5.70	6.48	4.53	2.65	1.41
33	0.53	4.66	1.27	4.07	2.81	4.53	5.57	6.61	4.93	2.72	1.97
34	0.78	4.91	1.55	4.16	3.34	4.26	5.35	6.49	5.18	2.72	2.40
35	1.15	4.79	1.78	3.91	3.86	3.80	4.80	6.18	5.07	2.62	2.54
36	1.43	4.22	1.82	3.32	4.03	3.08	3.98	5.74	4.44	2.33	2.46
37	1.44	3.40	1.56	2.79	3.83	2.36	3.04	5.12	3.49	1.95	2.14
38	1.29	2.85	1.39	2.67	3.39	1.91	2.17	4.19	2.68	1.58	1.70
39	1.17	2.81	1.72	3.05	2.90	1.68	1.56	3.22	2.02	1.19	1.28
40	1.20	2.96	2.32	3.95	2.71	1.71	1.24	2.58	1.50	0.85	0.88
41	1.43	2.93	2.77	5.10	2.82	1.98	1.23	2.37	1.26	0.68	0.54
42	1.64	2.73	2.94	5.96	2.83	2.21	1.37	2.57	1.28	0.65	0.32
43	1.67	2.48	2.85	6.30	2.61	2.16	1.44	2.74	1.46	0.68	0.22
44	1.63	2.17	2.61	6.23	2.27	1.82	1.38	2.58	1.63	0.69	0.20
Δ	0.303	0.315	0.311	0.309	0.317	0.294	0.315	0.329	0.323	0.321	0.313
VTL	13.32	13.84	13.71	13.60	13.93	12.94	13.85	14.47	14.22	14.11	13.76

tion from both coefficients. The inset plots within Fig. 8(a)demonstrate the area vector shapes generated at the end points of the ϕ_1 and ϕ_2 lines.

When q_3 is nonzero a third dimension is added to the articulation space. The q_2 vs q_1 grid will be the same as in Fig. 8(a), but can be considered to be shifted along the q_3 dimension. This is demonstrated in Fig. 8(b), again based on SF1's coefficients, where the dark grid in the middle is the case when $q_3=0$, and the other two lighter grids are positioned at the minimum and maximum values of q_3, respectively. Although $K=5$, only three grids are shown along the q_3 dimension to preserve the visual clarity of the figure.

To complete the mapping for a given speaker (e.g., SF1), Eq. (7) was used to generate an area vector for every coefficient combination within that speaker's three-dimensional coefficient space. A length vector was also generated in which the tubelet length (and consequently vocal tract length) was held constant for each speaker at the same mean values used to calculate frequency responses of the mean area functions in Sec. IV B. Finally, for each newly created area function, a set of formant frequencies was calculated.

Results for each of the six speakers are shown in the vowel space plots of Fig. 9. Within each figure panel is a dark, solid-lined grid of 3600 [F1, F2] pairs that corresponds

TABLE XIII. Area vectors for SF3.

Section i	i	ɪ	e	ɛ	æ	ʌ	ɑ	ɔ	o	ʊ	u
1	0.21	0.20	0.21	0.25	0.22	0.30	0.15	0.14	0.26	0.18	0.23
2	0.13	0.18	0.27	0.20	0.22	0.29	0.23	0.12	0.27	0.19	0.28
3	0.15	0.22	0.31	0.12	0.31	0.26	0.49	0.18	0.30	0.23	0.46
4	0.33	0.49	0.56	0.19	0.63	0.28	1.04	0.47	0.39	0.28	1.18
5	1.17	1.17	1.16	0.62	1.13	0.54	1.47	0.96	0.67	0.39	2.34
6	2.47	1.79	1.72	1.27	1.47	1.15	1.46	1.33	1.19	0.79	2.99
7	3.20	1.78	1.85	1.70	1.37	1.76	1.17	1.30	1.57	1.54	2.88
8	3.33	1.44	1.62	1.73	1.07	1.92	0.84	0.98	1.47	2.09	2.60
9	3.42	1.13	1.35	1.54	0.86	1.66	0.63	0.58	1.23	2.05	2.51
10	3.72	1.02	1.39	1.51	0.74	1.29	0.58	0.32	1.12	1.78	2.67
11	4.34	1.29	1.86	1.73	0.79	0.95	0.62	0.29	1.06	1.59	2.98
12	4.95	1.86	2.51	1.92	1.04	0.76	0.66	0.40	1.00	1.51	3.22
13	5.29	2.40	3.04	2.12	1.26	0.78	0.65	0.49	0.92	1.49	3.22
14	5.56	2.72	3.38	2.31	1.38	0.90	0.58	0.48	0.82	1.37	2.97
15	5.57	2.86	3.51	2.31	1.45	0.92	0.53	0.46	0.71	1.16	2.58
16	5.25	2.93	3.54	2.31	1.51	0.84	0.52	0.48	0.63	1.01	2.27
17	5.10	2.97	3.42	2.43	1.59	0.73	0.55	0.53	0.60	0.94	2.03
18	4.93	2.95	3.08	2.44	1.66	0.72	0.62	0.57	0.52	0.93	1.78
19	4.45	2.93	2.82	2.38	1.77	0.81	0.74	0.70	0.50	0.98	1.64
20	3.99	2.94	2.83	2.53	2.10	0.79	0.98	1.11	0.70	1.05	1.64
21	3.69	3.01	2.90	2.94	2.73	0.71	1.32	1.66	1.02	1.21	1.67
22	3.46	3.06	2.88	3.35	3.49	0.94	1.74	2.24	1.25	1.52	1.61
23	3.19	2.94	2.65	3.47	4.19	1.48	2.29	2.80	1.37	1.90	1.49
24	2.74	2.55	2.22	3.26	4.75	1.95	2.88	3.19	1.47	2.12	1.29
25	2.05	2.02	1.82	2.91	5.10	2.22	3.45	3.51	1.57	2.06	0.93
26	1.36	1.52	1.57	2.53	5.35	2.39	3.92	3.96	1.64	1.84	0.56
27	0.85	1.10	1.45	2.24	5.49	2.45	4.24	4.58	1.80	1.70	0.37
28	0.51	0.78	1.38	2.10	5.43	2.48	4.53	5.17	2.11	1.75	0.39
29	0.30	0.56	1.34	2.11	5.44	2.63	4.90	5.54	2.66	1.91	0.60
30	0.22	0.45	1.35	2.18	5.60	2.92	5.35	5.98	3.36	2.05	0.98
31	0.25	0.49	1.39	2.33	5.87	3.17	5.86	6.54	3.95	2.23	1.49
32	0.35	0.72	1.44	2.60	6.18	3.35	6.21	6.92	4.40	2.50	2.11
33	0.49	1.09	1.49	2.92	6.37	3.47	6.29	6.91	4.65	2.68	2.69
34	0.68	1.51	1.48	3.13	6.43	3.38	6.29	6.65	4.56	2.75	3.06
35	0.90	1.76	1.48	3.10	6.41	3.17	6.13	6.20	4.07	2.74	3.22
36	1.18	1.73	1.60	2.80	6.16	2.97	5.76	5.34	3.33	2.53	3.04
37	1.44	1.70	1.80	2.39	5.77	2.67	5.29	4.28	2.67	2.16	2.46
38	1.46	2.02	1.82	2.04	5.49	2.15	4.67	3.55	2.19	1.74	1.77
39	1.38	2.69	1.86	2.02	5.29	1.55	3.91	3.19	1.74	1.29	1.18
40	1.88	3.32	2.56	2.55	5.36	1.03	3.28	3.00	1.29	0.87	0.75
41	3.17	3.56	4.02	3.22	5.69	0.67	3.01	2.97	0.97	0.54	0.54
42	4.36	3.47	5.17	3.43	5.76	0.50	3.06	3.21	0.85	0.34	0.48
43	4.53	3.06	5.27	3.23	5.42	0.45	3.14	3.52	0.95	0.25	0.41
44	3.97	2.20	4.81	2.93	4.84	0.45	3.00	3.69	1.14	0.25	0.20
Δ	0.343	0.346	0.370	0.349	0.322	0.351	0.316	0.360	0.364	0.362	0.363
VTL	15.10	15.24	16.27	15.34	14.15	15.44	13.92	15.83	16.03	15.92	15.97

to a particular speaker's $[q_1, q_2]$ grid when $q_3=0$. The thick white lines (solid and dashed) indicate the formant frequency pairs that are produced in this same grid when either q_1 or q_2 are zero, thus demonstrating the acoustic effect of ϕ_1 or ϕ_2 by itself. The intersection point of these two lines gives the values of F1 and F2 when all three scaling coefficients are zero (i.e., the "neutral" location). There are five lighter, dashed grids "underneath" the dark one that are [F1, F2] pairs associated with the $[q_1, q_2]$ grids as they are shifted along the q_3 dimension. It is not intended that their separation be visible, but rather it is the extension of the vowel space around the dark grid provided by the third mode that is of interest. The four white dots (outlined in black) in each panel, plotted in clockwise order starting at the upper left corner, are the [F1, F2] pairs *measured* from the audio recording of each speaker's [i], [æ], [ɑ], and [u] vowels (see Tables II–VII).

Taking SF1 as an example, the solid white line in Fig. 9(a) represents the acoustic mapping of the q_1 coefficient continuum shown previously as the black, horizontal line in Fig. 8(a). This demonstrates that the primary acoustic effect of ϕ_1 is to generate a low F1 and high F2 when its scaling

TABLE XIV. Area vectors for SM1.

Section i	i	ɪ	e	ɛ	æ	ʌ	ɑ	ɔ	o	ʊ	u
1	0.70	0.75	0.34	0.75	0.36	0.18	0.77	0.73	0.86	0.79	0.62
2	0.73	0.58	0.45	0.64	0.43	0.33	0.43	0.70	0.71	0.77	0.85
3	0.78	0.52	0.46	0.52	0.44	0.34	0.25	0.56	0.60	0.75	0.99
4	0.90	0.53	0.36	0.45	0.41	0.35	0.30	0.41	0.50	0.69	1.08
5	1.54	0.83	0.40	0.55	0.46	0.44	0.62	0.39	0.49	0.68	1.29
6	2.77	1.75	0.96	1.18	0.83	0.66	1.10	0.72	1.05	0.99	2.02
7	3.73	2.60	1.87	1.97	1.53	1.31	1.50	1.37	2.05	1.69	3.49
8	3.98	2.56	2.33	2.10	2.02	1.91	1.52	1.65	2.55	2.22	4.70
9	3.95	2.09	2.09	1.79	1.92	1.81	1.12	1.21	2.23	2.03	4.74
10	3.86	1.84	1.66	1.53	1.49	1.43	0.62	0.67	1.58	1.45	3.98
11	3.77	1.88	1.50	1.50	1.07	1.05	0.32	0.39	1.03	1.04	3.05
12	3.86	2.14	1.61	1.78	0.82	0.74	0.23	0.27	0.71	0.96	2.38
13	4.08	2.55	1.76	2.20	0.93	0.83	0.28	0.24	0.69	1.16	2.06
14	4.33	2.97	2.04	2.56	1.32	1.14	0.39	0.28	0.82	1.36	1.83
15	4.44	3.20	2.52	2.78	1.61	1.29	0.43	0.31	0.87	1.43	1.67
16	4.29	3.18	2.83	2.85	1.79	1.33	0.49	0.40	0.87	1.43	1.60
17	4.11	3.06	2.85	2.85	1.91	1.43	0.67	0.60	0.91	1.45	1.57
18	3.82	2.88	2.70	2.78	1.95	1.51	0.91	0.85	1.05	1.50	1.60
19	3.36	2.57	2.50	2.52	1.89	1.54	1.15	1.16	1.25	1.54	1.59
20	2.88	2.14	2.19	2.09	1.63	1.58	1.32	1.43	1.35	1.56	1.55
21	2.37	1.66	1.74	1.69	1.33	1.52	1.23	1.42	1.22	1.46	1.47
22	1.93	1.35	1.42	1.52	1.21	1.35	0.98	1.20	1.05	1.17	1.19
23	1.70	1.15	1.36	1.45	1.27	1.34	0.99	1.25	1.22	0.95	0.86
24	1.43	0.91	1.26	1.27	1.32	1.56	1.31	1.73	1.56	1.00	0.77
25	0.99	0.66	0.97	0.99	1.21	1.74	1.63	2.30	1.76	1.09	0.77
26	0.63	0.45	0.70	0.81	1.02	1.82	1.82	2.86	1.87	1.02	0.63
27	0.38	0.34	0.55	0.75	0.97	1.97	2.11	3.72	2.13	0.99	0.48
28	0.20	0.44	0.52	0.81	1.17	2.35	2.63	4.84	2.58	1.11	0.44
29	0.19	0.65	0.65	1.03	1.46	2.80	3.18	5.63	3.09	1.39	0.55
30	0.30	0.79	0.86	1.30	1.65	3.03	3.56	5.91	3.57	1.74	0.83
31	0.39	0.85	1.07	1.50	1.75	3.05	3.71	6.04	3.83	1.95	1.16
32	0.46	0.89	1.29	1.72	1.85	3.01	3.67	6.06	3.78	2.03	1.37
33	0.62	0.91	1.44	1.91	2.02	2.92	3.51	5.87	3.65	2.07	1.50
34	0.77	0.97	1.47	1.97	2.22	2.71	3.27	5.50	3.56	2.04	1.63
35	0.80	1.11	1.51	1.95	2.34	2.34	3.12	5.01	3.24	1.92	1.66
36	0.80	1.18	1.61	1.90	2.37	1.92	3.03	4.46	2.56	1.73	1.64
37	0.83	1.11	1.64	1.80	2.34	1.50	2.75	3.89	1.84	1.45	1.63
38	0.89	1.09	1.54	1.70	2.23	1.19	2.31	3.36	1.50	1.11	1.58
39	1.04	1.32	1.46	1.65	2.14	1.07	2.01	3.09	1.43	0.78	1.58
40	1.20	1.60	1.52	1.72	2.31	1.10	2.04	2.99	1.30	0.54	1.58
41	1.28	1.69	1.63	1.82	2.76	1.19	2.23	2.71	1.09	0.42	1.42
42	1.33	1.63	1.64	1.83	3.15	1.25	2.33	2.26	0.83	0.33	0.99
43	1.32	1.50	1.55	1.72	3.22	1.16	2.25	1.85	0.53	0.23	0.47
44	1.22	1.34	1.44	1.50	3.01	0.98	2.03	1.42	0.22	0.12	0.05
Δ	0.393	0.403	0.409	0.399	0.381	0.400	0.372	0.393	0.408	0.394	0.420
VTL	17.28	17.73	18.02	17.56	16.75	17.58	16.35	17.31	17.94	17.34	18.48

coefficient q_1 is large and negative, and produces a fairly high F1 and low F2 when q_1 is large and positive. Such a result is not unexpected based on the mode coefficient table (Table IX) which suggests that ϕ_1 roughly represents vowels along the high-front to low-back continuum. The acoustic effect of ϕ_2, shown by the dashed white line, is to increase both F1 and F2 as the q_2 coefficient is increased from its most negative to most positive values. This result is also suggested by the mode coefficients in Table IX, where the largest negative and positive values of q_2 were typically associated with high-back and low-front vowels, respectively.

The dark grid in Fig. 9(a) shows how the entire rectangular $[q_1, q_2]$ coefficient space of Fig. 8(a) becomes deformed when mapped to the acoustic space. For the most part the characteristic shape of the ϕ_1 and ϕ_2 mappings (white lines) are retained throughout the grid and appear to influence its overall shape. This grid also demonstrates that the mapping between formant frequencies and mode coefficients is essentially one-to-one. That is, one [F1, F2] pair corresponds to one $[q_1, q_2]$ pair in Fig. 8(a). A similar one-to-one mapping of this type was reported by Story and Titze (1998).

TABLE XV. Area vectors for SM2. An area function for [ɛ] could not be obtained from this speaker.

Section i	i	ɪ	e	ɛ	æ	ʌ	ɑ	ɔ	o	ʊ	u
1	0.66	0.21	0.21	⋯	0.25	0.27	0.34	0.64	0.29	0.42	0.45
2	0.37	0.24	0.48	⋯	0.19	0.29	0.23	0.27	0.28	0.37	0.99
3	0.49	0.25	1.34	⋯	0.14	0.26	0.18	0.09	0.28	0.40	0.85
4	1.79	0.30	2.74	⋯	0.12	0.28	0.20	0.07	0.26	0.50	0.72
5	3.63	0.44	3.69	⋯	0.16	0.54	0.26	0.17	0.32	0.77	0.84
6	4.11	0.86	3.81	⋯	0.37	1.24	0.40	0.69	0.58	1.51	1.02
7	3.96	1.93	3.60	⋯	1.05	2.22	0.78	1.70	1.18	3.01	1.65
8	4.48	3.42	3.37	⋯	2.11	2.71	1.41	2.43	2.08	4.38	3.40
9	4.94	4.42	3.15	⋯	2.79	2.45	1.85	2.31	2.84	4.57	5.54
10	4.95	4.58	2.96	⋯	2.69	2.07	1.72	1.66	3.12	4.09	6.45
11	4.84	4.20	2.88	⋯	2.21	1.80	1.29	1.10	3.00	3.68	6.27
12	4.80	3.77	2.95	⋯	1.80	1.61	0.96	0.86	2.76	3.49	5.87
13	4.82	3.49	3.01	⋯	1.51	1.48	0.75	0.70	2.49	3.50	5.61
14	4.87	3.36	2.99	⋯	1.29	1.33	0.60	0.49	2.21	3.50	5.63
15	4.77	3.33	2.94	⋯	1.23	1.19	0.53	0.35	1.94	3.25	5.43
16	4.39	3.15	2.86	⋯	1.38	1.05	0.44	0.26	1.56	2.66	4.71
17	3.90	2.85	2.79	⋯	1.59	0.91	0.33	0.18	1.08	1.96	3.66
18	3.39	2.60	2.75	⋯	1.59	0.81	0.36	0.23	0.66	1.45	2.58
19	2.95	2.40	2.76	⋯	1.41	0.76	0.48	0.44	0.40	1.13	1.82
20	2.66	2.25	2.84	⋯	1.43	0.69	0.56	0.55	0.30	0.98	1.34
21	2.42	2.16	2.97	⋯	1.70	0.70	0.64	0.47	0.34	0.98	1.00
22	2.19	2.07	2.99	⋯	1.97	0.80	0.87	0.49	0.52	1.03	0.83
23	1.90	1.95	2.50	⋯	2.20	0.96	1.19	0.68	0.77	1.06	0.77
24	1.29	1.86	1.62	⋯	2.36	1.17	1.47	0.92	1.04	1.15	0.84
25	0.60	1.70	1.26	⋯	2.15	1.24	1.62	1.28	1.29	1.28	0.97
26	0.28	1.30	1.37	⋯	1.55	1.05	1.68	1.86	1.40	1.39	0.90
27	0.26	0.86	1.14	⋯	1.08	0.74	1.95	2.53	1.33	1.49	0.69
28	0.30	0.72	0.73	⋯	1.09	0.68	2.46	3.06	1.37	1.52	0.80
29	0.31	0.81	0.55	⋯	1.45	1.06	3.12	3.74	1.74	1.58	1.29
30	0.28	0.86	0.51	⋯	1.96	1.71	4.11	4.82	2.35	2.01	1.90
31	0.20	0.81	0.57	⋯	2.55	2.31	5.29	6.22	3.05	2.60	2.53
32	0.19	0.83	0.73	⋯	3.06	2.64	6.14	7.67	3.59	3.00	3.06
33	0.30	1.02	0.85	⋯	3.40	2.79	6.52	8.40	3.83	3.21	3.37
34	0.41	1.31	0.90	⋯	3.57	2.82	6.67	8.41	3.89	3.29	3.50
35	0.44	1.54	0.93	⋯	3.59	2.72	6.65	8.28	3.84	3.23	3.45
36	0.47	1.62	0.95	⋯	3.65	2.46	6.42	7.93	3.59	3.04	3.22
37	0.53	1.58	1.03	⋯	3.96	2.14	6.10	7.22	3.17	2.84	2.89
38	0.72	1.71	1.24	⋯	4.57	1.95	5.75	6.52	2.73	2.66	2.51
39	1.03	2.13	1.55	⋯	5.50	1.95	5.38	5.87	2.29	2.45	2.01
40	1.49	2.45	2.00	⋯	6.77	2.03	5.06	4.88	1.86	2.26	1.42
41	2.18	2.46	2.52	⋯	8.15	2.04	4.84	3.78	1.56	1.96	0.84
42	2.66	2.45	2.85	⋯	9.20	1.95	4.61	2.86	1.35	1.49	0.40
43	2.54	2.32	2.74	⋯	9.30	1.78	4.22	2.06	1.00	1.06	0.16
44	2.11	1.84	2.11	⋯	8.49	1.63	3.59	1.26	0.35	0.69	0.06
Δ	0.398	0.398	0.364	⋯	0.366	0.395	0.387	0.411	0.404	0.403	0.414
VTL	17.53	17.50	16.03	⋯	16.10	17.37	17.02	18.10	17.76	17.71	18.20

Also observable in Fig. 9(a) are the effects of the third mode on the acoustic mapping. Denoted by the light grids are the [F1, F2] pairs corresponding to the $[q_1, q_2]$ grids in Fig. 8(b) that have been shifted along the q_3 dimension. The most negative values of q_3 tend to lower both F1 and F2, extending the formant space to a region more representative of an [u] vowel. Large positive values of q_3 generally have the opposite effect of increasing F1 and F2. It is noted that all of the formant pairs for the measured vowels lie outside the dark grid, although [i] and [ɑ] are located just at the edge. This suggests that, for this speaker, the third mode is necessary to adequately represent her vowel space, especially for [u] and [æ].

The characteristics of the acoustic mappings for the other speakers shown in Fig. 9 are similar to those of SF1. It is evident that the independent effect of ϕ_1 on the vowel space (thick white lines) is to generate low F1 and high F2 frequencies at one end of the q_1 continuum and high F1 and low F2 frequencies at the other. Likewise, for all speakers, ϕ_2's effect (dashed white lines) is to produce F1s and F2s

TABLE XVI. Area vectors for SM3.

Section i	i	ɪ	e	ɛ	æ	ʌ	ɑ	ɔ	o	ʊ	u
1	0.95	0.58	0.96	0.69	0.60	0.61	0.47	0.69	0.47	0.53	0.50
2	0.73	0.44	0.45	0.44	0.38	0.84	0.34	0.73	0.44	0.37	0.65
3	0.58	0.29	0.22	0.25	0.23	0.95	0.20	0.87	0.36	0.25	0.66
4	0.54	0.15	0.17	0.11	0.17	0.53	0.10	0.78	0.26	0.14	0.75
5	0.73	0.13	0.21	0.05	0.25	0.15	0.14	0.35	0.28	0.35	1.11
6	1.48	0.21	0.41	0.21	0.55	0.19	0.42	0.12	0.64	1.11	2.05
7	2.70	0.41	0.83	0.83	0.96	0.73	0.75	0.23	1.24	1.73	3.00
8	3.37	0.96	1.24	1.46	1.16	1.55	0.87	0.57	1.53	1.58	3.01
9	3.34	1.43	1.17	1.38	1.03	1.68	0.82	0.89	1.32	1.26	2.75
10	3.48	1.23	0.80	0.92	0.72	1.24	0.66	0.74	1.08	1.33	2.63
11	3.98	0.92	0.68	0.81	0.52	1.31	0.46	0.57	1.16	1.41	2.51
12	4.71	1.22	0.96	1.21	0.82	1.48	0.37	0.65	1.25	1.45	2.82
13	5.38	1.94	1.41	1.76	1.55	1.33	0.44	0.52	1.23	1.66	3.26
14	5.67	2.69	1.93	2.11	2.04	1.43	0.52	0.37	1.31	1.74	3.42
15	5.37	3.26	2.37	2.25	2.01	1.57	0.57	0.38	1.27	1.47	3.07
16	4.50	3.28	2.55	2.03	1.75	1.33	0.58	0.36	0.98	1.13	2.40
17	3.59	2.80	2.42	1.57	1.61	1.00	0.48	0.33	0.77	0.98	2.02
18	3.04	2.35	2.20	1.38	1.66	0.88	0.41	0.35	0.74	1.06	1.91
19	2.68	2.16	1.84	1.38	1.72	0.91	0.45	0.40	0.79	1.05	1.82
20	2.24	1.86	1.30	1.24	1.58	0.83	0.56	0.43	0.92	0.81	1.67
21	1.74	1.41	1.16	1.04	1.36	0.59	0.54	0.42	1.15	0.68	1.27
22	1.44	1.22	1.41	1.01	1.39	0.46	0.49	0.36	1.39	0.88	0.83
23	1.42	1.30	1.58	1.23	1.66	0.59	0.89	0.31	1.32	1.15	0.74
24	1.45	1.39	1.60	1.52	1.89	0.93	1.54	0.54	1.16	1.25	0.87
25	1.24	1.27	1.45	1.55	1.92	1.34	1.89	1.25	1.39	1.16	0.99
26	0.91	1.01	1.15	1.33	1.80	1.56	2.03	2.01	1.85	1.06	0.95
27	0.60	0.85	0.79	1.03	1.66	1.58	2.20	2.40	2.27	1.06	0.86
28	0.33	0.75	0.54	0.74	1.59	1.58	2.47	2.83	2.69	1.17	0.92
29	0.17	0.69	0.51	0.62	1.66	1.73	2.92	3.79	3.30	1.37	1.16
30	0.10	0.70	0.65	0.66	1.87	1.98	3.73	5.08	4.37	1.58	1.49
31	0.07	0.74	0.79	0.74	2.15	2.23	4.86	6.19	5.68	1.70	1.88
32	0.06	0.78	0.91	0.84	2.38	2.44	5.84	6.82	6.62	1.69	2.34
33	0.08	0.83	1.03	0.98	2.51	2.58	6.29	7.14	7.05	1.55	2.82
34	0.15	0.94	1.19	1.14	2.56	2.56	6.30	7.38	7.12	1.38	3.09
35	0.28	1.04	1.27	1.21	2.60	2.39	6.10	7.59	6.85	1.27	3.18
36	0.35	1.01	1.25	1.18	2.95	2.12	5.75	7.83	6.74	1.30	3.40
37	0.45	1.00	1.34	1.23	3.92	1.78	5.27	8.26	7.45	1.44	3.60
38	0.96	1.21	1.72	1.58	5.52	1.44	4.82	8.65	8.22	1.50	3.46
39	1.95	1.82	2.65	2.58	7.35	1.23	4.64	8.35	7.67	1.26	3.01
40	2.75	2.78	4.11	3.92	8.82	1.25	4.54	7.26	6.07	0.78	2.20
41	2.91	3.43	5.41	4.45	9.54	1.33	4.13	5.99	4.34	0.44	1.25
42	2.71	3.48	6.03	4.25	9.61	1.17	3.39	4.65	2.78	0.34	0.66
43	2.28	3.23	6.06	3.74	9.18	0.75	2.43	3.07	1.67	0.29	0.44
44	1.57	2.70	5.76	2.84	8.34	0.25	1.52	1.49	0.96	0.24	0.29
Δ	0.399	0.411	0.393	0.404	0.395	0.395	0.409	0.428	0.406	0.415	0.409
VTL	17.57	18.07	17.28	17.79	17.38	17.38	17.98	18.84	17.87	18.28	17.98

that are both low in frequency at the negative end of the q_2 continuum and both high in frequency at the positive end. There are, however, speaker-specific differences. For example, the ϕ_1 line for SF2 has a relatively shallow slope and lacks the distinct curvature near the endpoints exhibited by the other speakers. This contrasts most apparently with the steepness of SM3's ϕ_1 line as well as with the curved portions at both the upper left and lower right ends of the line. For each speaker, the characteristic shape of their ϕ_1 and ϕ_2 lines is essentially retained throughout the entire vowel space mapping and influences its overall shape. Furthermore, the dark grids, representing the cases when $q_3=0$, all represent nearly one-to-one mappings between the $[q_1,q_2]$ coefficient space and the [F1, F2] acoustic space. The exception is for SM3 [Fig. 9(f)] where the upper left portion of his [F1, F2] grid indicates that multiple coefficient pairs correspond to the same [F1, F2] pair. This seems to be due in part to the steepness of ϕ_1 line.

When the q_3 dimension is included, the formant space for all speakers is slightly expanded. A consequence of this expansion is that, with the exception of the [u] vowel for SM1, all of the measured formant pairs for the four vowels

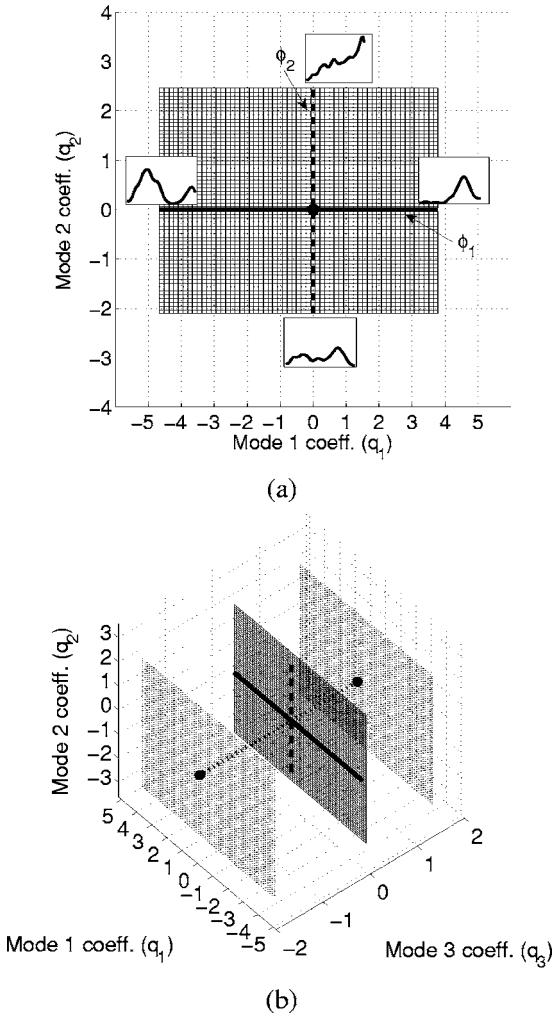

FIG. 8. (a) Grid of q_1 and q_2 scaling coefficients based on speaker SF1. The thick solid and dashed lines represent the continua for q_1 and q_2, respectively, when the other coefficient is equal to zero and are labeled with ϕ_1 and ϕ_2. The inset plots demonstrate the area vector shapes generated at the end points of the ϕ_1 and ϕ_2 lines. (b) Extension of the coefficient space to account for three modes. The (q_1,q_2) grid located at $q_3=0$ is identical to the grid shown in (a). The other two grids include the same collection of (q_1,q_2) values but are shifted along the q_3 dimension.

are included within each speaker's producible formant space. The mapping between the three-dimensional coefficient space and [F1, F2] formant space, however, is not one-to-one because many $[q_1,q_2,q_3]$ triplets may correspond to the same [F1, F2] pair. It is apparent though, from all of the speakers, that the primary effect of ϕ_3 is to slightly extend the edges of the vowel space relative to the $q_3=0$ condition. Thus, based on these figures, as well as on the calculated variances for each mode (see Table VIII), it seems reasonable to suggest that the nearly one-to-one mappings based on ϕ_1 and ϕ_2 capture most of a speaker's vowel production, while ϕ_3 may be "activated" to tune the vocal tract for the extreme vowels.

V. DISCUSSION

The main goal of this study was to determine whether vocal tract shaping patterns, obtained from sets of area functions, were similar across speakers. Toward this goal, area functions for 11 vowels were obtained from six speakers, three female and three male, using MRI. From each speaker's set of area functions, eigenvectors or *modes* were determined with principal components analysis (PCA).

The spatial similarity of each mode was assessed by visual comparison across all of the speakers as well as with correlation analysis. In both cases, the spatial features present in the first and second modes, ϕ_1 and ϕ_2, were highly correlated within the female and male groups, and across sex. The average correlation coefficient across all six speakers was 0.94 for the first mode and 0.91 for the second mode. The shape of third mode, ϕ_3, was fairly similar across the female speakers, but less so for the males. It is also noted that the mean area vectors, upon which the modes are superimposed to reconstruct specific vowels, contained more idiosyncratic features than did the modes. Acoustically, these idiosyncratic features primarily influenced the frequencies of the upper formants, while leaving F1 and F2 at locations representative of a neutral vowel. Thus, at this stage, the modes appeared to provide roughly a *common* system for perturbing a *unique* underlying neutral vocal tract shape.

This idea was further supported by the mappings generated between the scaling coefficients of the modes $[q_1,q_2,q_3]$ and the [F1, F2] frequencies of the resulting area functions. The mappings were unique for all six speakers in terms of the exact shape of the [F1, F2] vowel space, but the general effect of the modes was the same in each case. Vocal tract configurations produced by the first modes gave rise to a continuum of formant pairs over which F1 and F2 monotonically increased and decreased, respectively. In contrast, the second mode produced a continuum of formants over which both F1 and F2 monotonically increased. In addition, these mappings were essentially one-to-one when just the first two modes were considered. This means that an [F1, F2] pair in a given speaker's vowel space could be associated with one pair of $[q_1,q_2]$ coefficients.

A similar one-to-one mapping, based on the calculated modes of a male speaker, was reported by Story and Titze (1998). They later used the mapping in the "reverse" direction to transform time-varying formant frequencies obtained from recorded speech into time-varying mode coefficients. These were then used to generate area functions for synthesis of the original speech. Presumably the mappings presented in Fig. 9 could be similarly used to map formants to coefficients and ultimately to area functions. With regard to understanding how the vocal tract airspace is utilized for speech production, however, perhaps it is of more interest to note that the magnitudes of the mode scaling coefficients for the speakers in the present study (see Table IX) as well as in Story and Titze (1998) are of nearly the same range. This means that a transformation of formant frequencies of some utterance for a given speaker may yield scaling coefficients that could potentially generalize across all of the speakers. Likewise, "forward" specification of a sequence of coefficients that vary in time would, for all speakers, generate the same vowel-to-vowel transition but with different absolute formant frequencies.

Whereas the modes seem to capture some type of common shaping patterns of the vocal tract airspace, they cannot

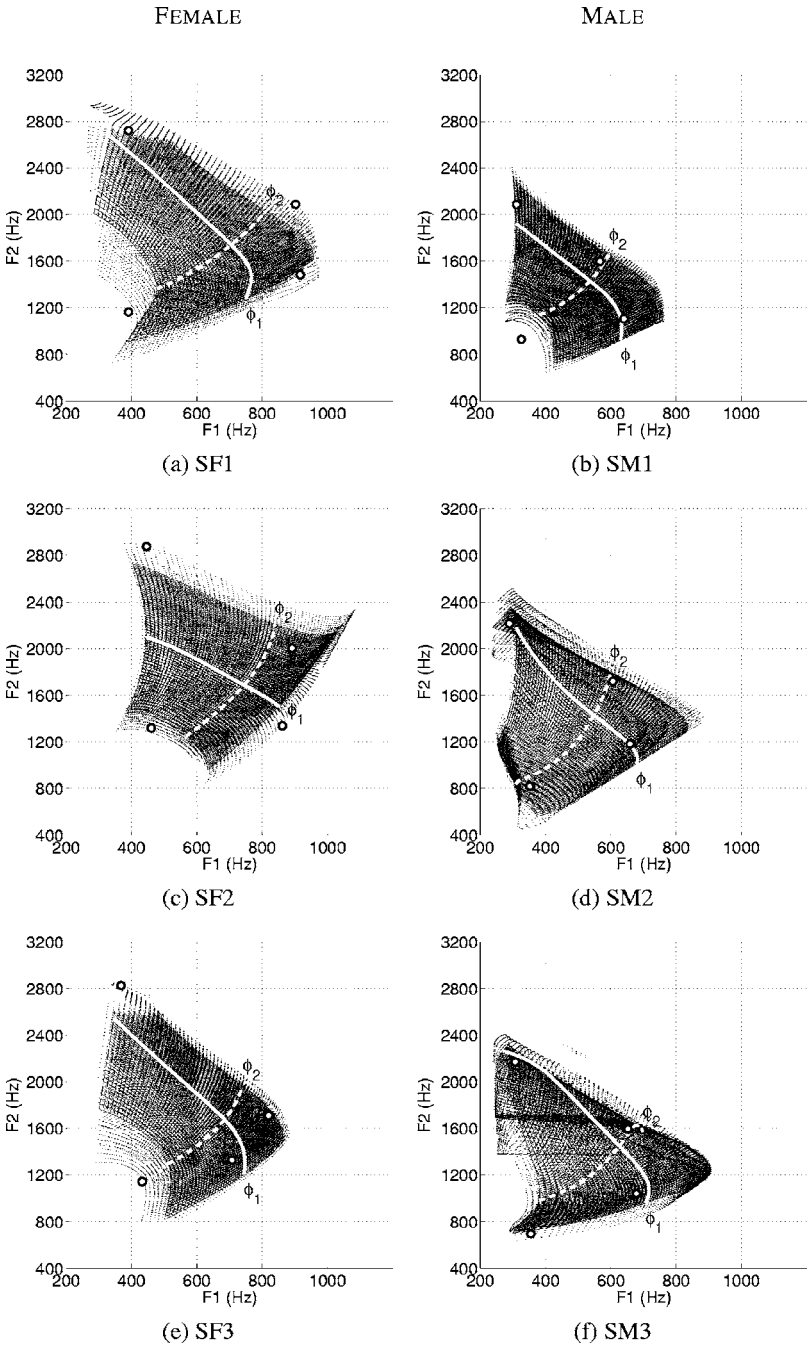

FIG. 9. Coefficient-to-formant mappings for each of the six speakers based the area functions generated with Eq. (7). The left column of plots corresponds to the female speakers and the right column to the males. Within each figure panel, the dark grid is the mapping obtained when $q_3=0$, the white lines (solid and dashed) represent the effect of each mode in isolation, the four white dots are the [F1, F2] pairs *measured* from each speaker's audio recording of the [i, æ, ɑ, u] vowels, and the light grids indicate the effect of ϕ_3.

be directly related to specific articulators. This may be regarded as a limitation of the study because little can be said about the actual articulatory coordination used for speech production. The data and subsequent analyses presented here, however, imply the existence of a common system for deforming the vocal tract shape that facilitates production of predictable patterns of formant frequencies, an essential component of efficient speech production. In this sense, perhaps the modes or combinations thereof prescribe a "goal" that some collective coordination of the articulators must achieve. Thus, even though variable positions and movement of articulators may be utilized during speech production, the goal would be achieved if the common deformation patterns, of the type predicted by the modes, are indeed generated by their collective effect. Although no information concerning muscle activity was presented in this study, the mode shapes apparently must represent some level of muscle orchestration that influences the shape of the vocal tract. To a degree consistent with the concept of coordinative structures or synergies in which muscles are organized into functional units, each of the modes may be thought to represent some abstract synergy of articulatory muscle activations that produce a desired acoustic effect.

Further investigation of the mode-based approach to vowel articulation needs to include vocal tract information based on languages other than American English. The results presented here may be inadvertently biased toward vocal tract configurations with expanded palatal regions (i.e., only 4 of the 11 vowels have significant constrictions in this region). It is possible that speakers of a language that includes the same American English vowels used in this study, as well as additional vowels with constricted front cavities (e.g.,

TABLE XVII. Mode and mean diameter vectors for the three female speakers. The glottal end of each area vector is at section 1 and the lip end at section 44. Each speaker's data are grouped into four consecutive columns within the table.

	SF1				SF2				SF3			
Section i	$\Omega(i)$	$\phi_1(i)$	$\phi_2(i)$	$\phi_3(i)$	$\Omega(i)$	$\phi_1(i)$	$\phi_2(i)$	$\phi_3(i)$	$\Omega(i)$	$\phi_1(i)$	$\phi_2(i)$	$\phi_3(i)$
1	0.638	−0.035	−0.030	−0.040	0.499	−0.018	−0.062	−0.093	0.518	−0.005	−0.035	0.053
2	0.661	−0.016	−0.021	−0.070	0.600	−0.012	−0.046	−0.107	0.520	0.008	0.011	−0.096
3	0.739	−0.022	−0.025	−0.099	0.688	−0.027	−0.075	−0.101	0.580	0.008	0.019	−0.177
4	0.901	−0.041	−0.034	−0.128	0.781	−0.051	−0.119	−0.093	0.792	−0.003	0.009	−0.215
5	1.065	−0.066	−0.042	−0.157	0.918	−0.076	−0.156	−0.093	1.128	−0.020	−0.008	−0.226
6	1.147	−0.092	−0.047	−0.184	1.120	−0.097	−0.177	−0.103	1.406	−0.041	−0.023	−0.222
7	1.151	−0.117	−0.048	−0.207	1.335	−0.113	−0.178	−0.123	1.508	−0.065	−0.032	−0.213
8	1.136	−0.139	−0.044	−0.226	1.467	−0.125	−0.159	−0.151	1.457	−0.090	−0.034	−0.202
9	1.174	−0.158	−0.035	−0.237	1.506	−0.133	−0.125	−0.182	1.355	−0.114	−0.028	−0.193
10	1.274	−0.173	−0.023	−0.239	1.503	−0.139	−0.079	−0.211	1.302	−0.138	−0.015	−0.186
11	1.374	−0.185	−0.008	−0.233	1.508	−0.142	−0.027	−0.236	1.344	−0.159	0.003	−0.179
12	1.428	−0.194	0.009	−0.218	1.531	−0.145	0.025	−0.252	1.427	−0.176	0.025	−0.173
13	1.440	−0.200	0.027	−0.194	1.539	−0.146	0.073	−0.259	1.491	−0.190	0.049	−0.166
14	1.428	−0.202	0.045	−0.163	1.502	−0.146	0.114	−0.254	1.512	−0.198	0.073	−0.157
15	1.375	−0.200	0.062	−0.127	1.443	−0.144	0.145	−0.239	1.491	−0.201	0.096	−0.143
16	1.282	−0.194	0.078	−0.086	1.375	−0.139	0.166	−0.216	1.464	−0.197	0.117	−0.126
17	1.202	−0.183	0.093	−0.044	1.283	−0.130	0.176	−0.185	1.450	−0.187	0.135	−0.104
18	1.193	−0.167	0.105	−0.003	1.163	−0.115	0.175	−0.150	1.431	−0.169	0.150	−0.079
19	1.255	−0.146	0.116	0.035	1.091	−0.094	0.165	−0.115	1.430	−0.145	0.162	−0.051
20	1.356	−0.119	0.123	0.067	1.142	−0.067	0.149	−0.081	1.488	−0.115	0.169	−0.022
21	1.430	−0.087	0.128	0.091	1.256	−0.034	0.126	−0.052	1.580	−0.079	0.172	0.005
22	1.428	−0.050	0.131	0.107	1.345	0.004	0.101	−0.030	1.684	−0.038	0.171	0.030
23	1.366	−0.010	0.130	0.112	1.394	0.047	0.074	−0.017	1.766	0.006	0.166	0.050
24	1.296	0.034	0.126	0.106	1.410	0.092	0.047	−0.014	1.786	0.051	0.157	0.062
25	1.275	0.078	0.119	0.090	1.411	0.138	0.021	−0.020	1.748	0.096	0.145	0.066
26	1.316	0.122	0.109	0.063	1.439	0.182	−0.003	−0.036	1.688	0.140	0.129	0.060
27	1.371	0.163	0.096	0.028	1.509	0.222	−0.023	−0.059	1.645	0.180	0.110	0.044
28	1.419	0.200	0.082	−0.013	1.612	0.255	−0.041	−0.087	1.634	0.214	0.088	0.018
29	1.482	0.230	0.066	−0.059	1.731	0.278	−0.054	−0.117	1.664	0.242	0.065	−0.017
30	1.565	0.252	0.051	−0.105	1.834	0.290	−0.065	−0.147	1.731	0.261	0.043	−0.059
31	1.653	0.264	0.038	−0.150	1.913	0.291	−0.071	−0.174	1.825	0.271	0.022	−0.105
32	1.729	0.264	0.028	−0.189	1.981	0.279	−0.072	−0.193	1.931	0.271	0.004	−0.150
33	1.768	0.254	0.023	−0.220	2.055	0.257	−0.068	−0.203	2.014	0.261	−0.007	−0.191
34	1.760	0.232	0.027	−0.239	2.114	0.226	−0.057	−0.203	2.056	0.243	−0.010	−0.222
35	1.721	0.200	0.040	−0.244	2.116	0.189	−0.038	−0.191	2.051	0.216	−0.002	−0.240
36	1.678	0.160	0.064	−0.234	2.029	0.150	−0.008	−0.168	1.992	0.184	0.020	−0.239
37	1.640	0.114	0.101	−0.209	1.866	0.114	0.032	−0.135	1.894	0.148	0.056	−0.218
38	1.588	0.067	0.151	−0.170	1.701	0.085	0.085	−0.096	1.784	0.111	0.108	−0.175
39	1.523	0.023	0.212	−0.120	1.590	0.069	0.150	−0.056	1.687	0.076	0.173	−0.113
40	1.478	−0.015	0.280	−0.066	1.548	0.065	0.222	−0.018	1.661	0.047	0.248	−0.038
41	1.460	−0.041	0.348	−0.015	1.564	0.074	0.297	0.011	1.701	0.024	0.324	0.037
42	1.446	−0.051	0.404	0.021	1.594	0.090	0.360	0.026	1.736	0.011	0.386	0.094
43	1.405	−0.043	0.430	0.028	1.591	0.100	0.392	0.025	1.716	0.006	0.415	0.104
44	1.313	−0.016	0.403	−0.014	1.545	0.081	0.362	0.005	1.633	0.008	0.380	0.028

Swedish) may produce somewhat different mode shapes than those reported here. In addition, the vocal tract length differences across vowels should be more adequately accounted for in the PCA, as well as in subsequent models for generating area functions.

ACKNOWLEDGMENTS

The author would like to thank Ted Trouard for consulting on image acquisition, Jennifer Johnson for operating the MR scanner, Kristen Bencala and Kang Li for assisting in the image analysis, and Wolfgang Golser for assistance with acoustic analysis. This work was supported by NIH Grant No. R01-DC04789.

APPENDIX A—Area Vectors and Vocal Tract Lengths

Area vectors for the three female and three male speakers are presented numerically in Tables XI–XVI. The first column in each table is the index i, which denotes successive sections or "tubelets" along the length of the vocal tract. Tubelet I is located just above the glottis and tubelet 44 at the

TABLE XVIII. Mode and mean diameter vectors for the three male speakers. The glottal end of each area vector is at section 1 and the lip end at section 44. Each speaker's data are grouped into four consecutive columns within the table.

	SM1				SM2				SM3			
Section i	$\Omega(i)$	$\phi_1(i)$	$\phi_2(i)$	$\phi_3(i)$	$\Omega(i)$	$\phi_1(i)$	$\phi_2(i)$	$\phi_3(i)$	$\Omega(i)$	$\phi_1(i)$	$\phi_2(i)$	$\phi_3(i)$
1	0.873	0.002	−0.089	−0.112	0.676	0.025	−0.059	0.094	0.897	−0.019	0.000	−0.005
2	0.867	−0.014	−0.050	−0.052	0.666	−0.060	−0.029	−0.063	0.762	0.002	−0.006	−0.027
3	0.835	−0.031	−0.056	−0.081	0.683	−0.102	0.001	−0.154	0.753	0.006	−0.027	−0.076
4	0.815	−0.047	−0.080	−0.146	0.811	−0.118	0.020	−0.189	0.529	−0.002	−0.051	−0.131
5	0.917	−0.065	−0.108	−0.215	1.004	−0.121	0.023	−0.179	0.405	−0.016	−0.073	−0.179
6	1.242	−0.083	−0.130	−0.267	1.251	−0.119	0.013	−0.136	0.698	−0.034	−0.089	−0.214
7	1.610	−0.102	−0.139	−0.295	1.589	−0.117	−0.010	−0.073	1.195	−0.052	−0.098	−0.235
8	1.759	−0.121	−0.136	−0.296	1.921	−0.117	−0.038	0.001	1.425	−0.071	−0.100	−0.243
9	1.662	−0.141	−0.120	−0.274	2.075	−0.122	−0.068	0.075	1.420	−0.088	−0.095	−0.240
10	1.470	−0.159	−0.094	−0.233	2.041	−0.130	−0.093	0.143	1.206	−0.103	−0.085	−0.231
11	1.312	−0.176	−0.061	−0.181	1.932	−0.140	−0.112	0.200	1.072	−0.116	−0.070	−0.217
12	1.247	−0.191	−0.025	−0.125	1.841	−0.152	−0.120	0.241	1.322	−0.127	−0.053	−0.203
13	1.299	−0.201	0.011	−0.071	1.778	−0.164	−0.117	0.266	1.437	−0.135	−0.034	−0.189
14	1.398	−0.207	0.045	−0.022	1.722	−0.173	−0.102	0.274	1.575	−0.140	−0.016	−0.176
15	1.456	−0.207	0.073	0.016	1.666	−0.180	−0.078	0.267	1.632	−0.142	0.003	−0.165
16	1.482	−0.200	0.094	0.044	1.578	−0.182	−0.046	0.248	1.471	−0.141	0.020	−0.156
17	1.512	−0.186	0.108	0.059	1.455	−0.179	−0.009	0.218	1.312	−0.137	0.035	−0.146
18	1.536	−0.165	0.113	0.064	1.351	−0.170	0.029	0.181	1.268	−0.129	0.047	−0.136
19	1.538	−0.138	0.111	0.060	1.286	−0.155	0.065	0.142	1.317	−0.118	0.057	−0.124
20	1.499	−0.104	0.102	0.048	1.244	−0.135	0.095	0.103	1.219	−0.103	0.064	−0.110
21	1.400	−0.064	0.087	0.033	1.235	−0.109	0.117	0.067	1.066	−0.084	0.067	−0.092
22	1.284	−0.021	0.068	0.017	1.266	−0.079	0.129	0.037	1.058	−0.061	0.067	−0.072
23	1.246	0.025	0.047	0.001	1.300	−0.046	0.130	0.013	1.123	−0.034	0.063	−0.049
24	1.270	0.071	0.025	−0.012	1.306	−0.010	0.118	−0.002	1.276	−0.004	0.055	−0.025
25	1.256	0.115	0.003	−0.020	1.291	0.027	0.095	−0.009	1.338	0.029	0.043	−0.001
26	1.207	0.157	−0.017	−0.024	1.250	0.064	0.062	−0.008	1.347	0.065	0.027	0.020
27	1.204	0.193	−0.033	−0.025	1.197	0.100	0.023	0.000	1.289	0.102	0.009	0.036
28	1.282	0.222	−0.046	−0.022	1.212	0.134	−0.021	0.012	1.229	0.139	−0.012	0.045
29	1.415	0.243	−0.054	−0.019	1.337	0.165	−0.065	0.028	1.311	0.176	−0.034	0.046
30	1.536	0.255	−0.057	−0.018	1.516	0.192	−0.105	0.045	1.451	0.210	−0.055	0.036
31	1.609	0.257	−0.054	−0.020	1.689	0.214	−0.138	0.062	1.616	0.241	−0.074	0.016
32	1.650	0.250	−0.047	−0.029	1.828	0.230	−0.159	0.078	1.709	0.266	−0.088	−0.015
33	1.675	0.234	−0.035	−0.046	1.922	0.241	−0.166	0.090	1.763	0.285	−0.094	−0.053
34	1.678	0.210	−0.018	−0.072	1.968	0.244	−0.157	0.100	1.806	0.295	−0.090	−0.097
35	1.653	0.182	0.004	−0.107	1.973	0.241	−0.131	0.107	1.824	0.296	−0.072	−0.141
36	1.600	0.151	0.031	−0.149	1.940	0.232	−0.088	0.112	1.798	0.287	−0.038	−0.180
37	1.519	0.120	0.063	−0.193	1.894	0.216	−0.030	0.118	1.836	0.266	0.013	−0.209
38	1.437	0.092	0.102	−0.236	1.875	0.194	0.039	0.125	1.965	0.235	0.081	−0.221
39	1.402	0.069	0.148	−0.268	1.885	0.168	0.116	0.134	2.076	0.195	0.165	−0.212
40	1.411	0.053	0.203	−0.283	1.889	0.139	0.197	0.145	2.207	0.147	0.258	−0.180
41	1.419	0.045	0.268	−0.268	1.875	0.110	0.276	0.155	2.121	0.096	0.352	−0.127
42	1.381	0.041	0.346	−0.215	1.824	0.084	0.351	0.159	1.983	0.045	0.431	−0.062
43	1.288	0.038	0.436	−0.114	1.701	0.067	0.421	0.144	1.801	0.003	0.471	−0.005
44	1.136	0.025	0.541	0.043	1.483	0.065	0.490	0.091	1.521	−0.023	0.441	0.012

lips. The other columns are the area data for the vowels in the order: [i, ɪ, e, ɛ,æ, ʌ, ɑ, ɔ, o, ʊ, u]. For reasons noted in the main text, the [ɛ] vowel for SM2 cannot be made available. The bottom two rows in each table indicate the length Δ of each successive section and the total vocal tract length (VTL=44Δ).

APPENDIX B—Mean Diameter and Mode Vectors

Numerical versions of the mode vectors, $\phi_1(i)$, $\phi_2(i)$, and $\phi_3(i)$, along with the mean diameter vectors $\Omega(i)$ are presented in Table XVII for the female speakers and in Table XVIII for the males. As with the area vectors in Appendix A, the first column of each table shows the index i which denotes successive sections along the length of the vocal tract.

[1]The use of 44 sections derives from the approximate spatial resolution obtained in MRI-based reconstructions of vocal tract shape. It is also convenient to use 44 elements for simulating male speech with acoustic waveguide models because it allows for a sampling frequency of 44.1 kHz when the tract length is approximately 17.5 cm (typical adult male). The female

area functions have also been segmented into 44 sections to simplify the management of the data.

[2]For this study, the length of each tubelet in a given area function is equal to the Δ shown in the area functions tables in Appendix A. Hence, $L(i)=\Delta$ for every tubelet within an area function. The form of the length vector used here, however, would generalize to cases where the tubelet length may be unequal (e.g., see Story, 2005).

[3]Story and Titze (1998) performed the PCA on the square root of the areas. In this study, the scaling factor of $4/\pi$ within the square root operation generates equivalent diameters which are a more intuitively appealing quantity and are more convenient to use for explanation and discussion.

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

Fowler, C. A., and Saltzman, E. L. (**1993**). "Coordination and coarticulation in speech production," Lang Speech **36**(2,3), 171–195.

Gracco, V. L. (**1992**). "Some organizational characteristics of speech movement control," J. Speech Hear. Res. **37**, 4–27.

Harshman, R., Ladefoged, P., and Goldstein, L. (**1977**). "Factor analysis of tongue shapes," J. Acoust. Soc. Am. **62**, 693–707.

Jackson, M. T. T. (**1988**). "Analysis of tongue positions: Language-specific and cross-linguistic models," J. Acoust. Soc. Am. **84**, 124–143.

Kelso, J. A. S., Saltzman, E. L., and Tuller, B. (**1986**). "The dynamical perspective on speech production: Data and theory," J. Phonetics **14**, 29–59.

Kusakawa, N., Honda, K., and Kakita, Y. (1993). "Construction of articulatory trajectories in the space of tongue muscle contraction force," ATR Technical Report, TR-A-0717 (in Japanese).

Löfqvist, A. (**1997**). "Theories and models of speech production," in *The Handbook of Phonetic Sciences*, edited by W. J. Hardcastle and J. Laver (Blackwell, Oxford), pp. 405–426.

Logeman, J. (**1983**). *Evaluation and Treatment of Swallowing Disorders* (College-Hill, San Diego, CA).

Macpherson, J. M. (**1991**). "How flexible are muscle synergies?" in *Motor Control: Concepts and Issues*, edited by D. R. Humphrey and H.-J. Freund (Wiley, Chichester), pp. 33–47.

Maeda, S., and Honda, K. (**1994**). "From EMG to formant patterns of vowels: the implication of vowel spaces," Phonetica **51**, 17–29.

Mermelstein, P. (**1967**). "Determination of the vocal-tract shape from measured formant frequencies," J. Acoust. Soc. Am. **41**, 1283–1294.

Meyer, P., Wilhelms, R., and Strube, H. W. (**1989**). "A quasiarticulatory speech synthesizer for German language running in real time," J. Acoust. Soc. Am. **86**, 523–539.

Nix, D. A., Papcun, G., Hogden, J., and Zlokarnik, I. (**1996**). "Two cross-linguistic factors underlying tongue shapes for vowels," J. Acoust. Soc. Am. **99**, 3707–3717.

Perrier, P., Perkell, J., Payan, Y., Zandipour, M., Guenther, F., and Khalighi, A. (**2000**). "Degrees of freedom of tongue movements in speech may be constrained by biomechanics," in Proc. of the Sixth Intl. Conf. on Spoken Lang. Proc., ICSLP-2000, Vol. **2**, pp. 162–165.

Santello, M., Flanders, M., and Soechting, J. F. (**1998**). "Postural hand synergies for tool use," J. Neurosci. **18**(23), 10105–10115.

Schroeder, M. R. (**1967**). "Determination of the geometry of the human vocal tract by acoustic measurements," J. Acoust. Soc. Am. **41**, 1002–1010.

Shirai, K., and Honda, M. (**1977**). "Estimation of articulatory motion," in *Dynamic Aspects of Speech Production*, edited by M. Sawashima and F. Cooper (Univ. of Tokyo, Tokyo), pp. 279–302.

Shriberg, L. D., and Kent, R. D. (**2003**). *Clinical Phonetics*, 3rd ed. (Allyn and Bacon, Boston).

Sondhi, M. M., and Schroeter, J. (**1987**). "A hybrid time-frequency domain articulatory speech synthesizer," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**(7), 955–967.

Stevens, K. N., and House, A. S. (**1955**). "Development of a quantitative description of vowel articulation," J. Acoust. Soc. Am. **27**, 484–493.

Story, B. H. (**2004**). "On the ability of a physiologically-constrained area function model of the vocal tract to produce normal formant patterns under perturbed conditions," J. Acoust. Soc. Am. **115**, 1760–1770.

Story, B. H. (**2005**). "A parametric model of the vocal tract area function for vowel and consonant simulation," J. Acoust. Soc. Am. **117**, 3231–3254.

Story, B. H., and Titze, I. R. (**1998**). "Parameterization of vocal tract area functions by empirical orthogonal modes," J. Phonetics **26**(3), 223–260.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am. **100**, 537–554.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1998**). "Vocal tract area functions for an adult female speaker based on volumetric imaging," J. Acoust. Soc. Am. **104**, 471–487.

Story, B.H., Laukkanen, A.M., and Titze, I.R. (**2000**). Acoustic impedance of an artificially lengthened and constricted vocal tract, J. Voice **14**(4), 455–469.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**2001**). "The relationship of vocal tract shape to three voice qualities," J. Acoust. Soc. Am. **109**, 1651–1667.

Taylor, J. R. (**1982**). *An Introduction to Error Analysis* (University Science Books, Mill Valley, CA).

Titze, I. R., and Story, B. H. (**1997**). "Acoustic interactions of the voice source with the lower vocal tract," J. Acoust. Soc. Am. **101**, 2234–2243.

Yehia, H. C., Takeda, K., and Itakura, F. (**1996**). "An acoustically oriented vocal-tract model," IEICE Trans. Inf. Syst. **E79-D**(8), 1198–1208.

Zheng, Y., Hasegawa-Johnson, M., and Pizza, S. (**2003**). "Analysis of the three-dimensional tongue shape using a three-factor analysis model," J. Acoust. Soc. Am. **113**, 478–486.