Three-dimensional tongue surface shapes of English consonants and vowels

Maureen Stone
Division of Otolaryngology, The University of Maryland Medical School, 16 S. Eutaw St., Rm. 525, Baltimore, Maryland 21201

Andrew Lundberg
Department of Computer Science, The Johns Hopkins University, Baltimore, Maryland 21218

(Received 5 July 1995; accepted for publication 12 February 1996)

This paper presents three-dimensional tongue surfaces reconstructed from multiple coronal cross-sectional slices of the tongue. Surfaces were reconstructed for sustained vocalizations of the American English sounds /i, ɪ, e, ɛ, æ, ɑ, ɔ, o, ʊ, ʌ, ɝ, l, s, ʃ, *θ*, n, ŋ/. Electropalatography (EPG) data were also collected for the sounds to compare tongue surface shapes with tongue–palate contact patterns. The study was interested also in whether 3-D surface shapes of the tongue were different for consonants and vowels. Previous research and speculation had found that there were differences in production, acoustics, and linguistic usage between the two groups. The present study found that four classes of tongue shape were adequate to categorize all the sounds measured. These classes were front raising, complete groove, back raising, and two-point displacement. The first and third classes have been documented before in the midsagittal plane [cf. R. Harshman, P. Ladefoged, and L. Goldstein, J. Acoust. Soc. Am. **62**, 693–707 (1976)]. The first three classes contained both vowels and consonants, the last only consonants. Electropalatographic patterns of the sounds indicated three categories of tongue–palate contact: bilateral, cross-sectional, and combination of the two. Vowels used only the first pattern, consonants used all three. The EPG data provided an observable distinction in contact pattern between consonants and vowels. The ultrasound tongue surface data did not. The conclusion was that the tongue actually has a limited repertoire of shapes and positions them against the palate in different ways for consonants versus vowels to create narrow channels, divert airflow, and produce sound.

PACS numbers: 43.70.Aj, 47.70.Jt

INTRODUCTION

Three-dimensional reconstructions of the tongue's surface are difficult to acquire using current technology. Vocal tract shapes can be reconstructed from magnetic resonance imaging (MRI) images (cf. Baer *et al.*, 1987, 1991) and tongue surfaces can be extracted from them to some extent. Computed tomography (CT) can also be used to collect such slices but is not typically done because of the radiation exposure, and because only transverse and oblique slices are possible. X ray is unable to provide 3-D surfaces because it is a projection technique that results in an image from which the 3-D tongue surface is not recoverable. Ultrasound images are thin sections of tongue soft tissue that could theoretically be used for reconstructions, but with two drawbacks: multiple sections cannot be collected simultaneously, and the sections are not parallel to each other.

Three-dimensional reconstructions of the tongue surface are of interest because the tongue is a complex system for which we typically have incomplete information. The tongue is composed entirely of muscle and has a fixed volume. These two features mean that, unlike a rigid body, the tongue not only is moved by its muscles, but also shaped by them. These features also classify the tongue as a muscular hydrostat (Kier and Smith, 1985). Tongue shape is systematically related to tongue position, because tongue volume can be redistributed, but not increased or decreased. When the tongue contacts a hard surface like the palate, its volume distribution becomes more complicated and depends on, among other things, the location, force, and surface area of the contact.

Considering muscular activity alone, very complicated tongue surface shapes can be produced. The extrinsic muscles of the tongue insert either exclusively at midline (genioglossus) or at the extreme lateral margins of the tongue (hyoglossus, palatoglossus, styloglossus) (Abd-El Malek, 1939; Carpentier and Pajoni, 1989). Contraction of genioglossus will pull the midline tongue inward, producing a midline groove, and there is EMG evidence that genioglossus contraction can be very local, producing a groove, or dimple, in one place, but not another (Miyawaki *et al.*, 1975). Local contraction of genioglossus (GG) appears to be very important in producing midsagittal tongue grooves. In addition, the tongue has four intrinsic muscles, which can be thought of as connecting the four ''sides'' of the pseudo-rectangular tongue. Superior and inferior longitudinal muscles connect the anterior and posterior ends. The verticalis m. connects the superior and inferior surfaces. The transverse m. connects the left and right sides. When any of the intrinsic muscles contract they will bring the two attached sides closer together, causing shortening, lengthening, widening, narrowing, or combinations of these shapes. By combining various muscle contractions, complicated shapes can be made. It is

0001-4966/96/99(6)/3728/10/$6.00

not known just how complicated actual tongue shapes are, however.

A further complicating factor is the hard palate. When the tongue touches the palate, the muscles' activities and their resultant forces are combined with a boundary condition that offers resistance to the volume. In that case the surface shapes will represent an interaction between muscle forces and the boundary contact. Lingual consonants and many vowels touch the palate. In low vowels or the consonant /l/, however, contact can be quite minimal. For an English /l/ in /ɑ/ context, as in the present data, the contact may contain only minimal anterior lateral contact. This appears to be true in Catalan and German as well (Recasens *et al.*, 1995, Fig. 5). It is of interest to determine the relationship between tongue–palate contact pattern and 3-D surface shape to understand how the tongue uses the hard palate to shape the vocal tract.

The current wisdom regarding the shape of the tongue is derived largely from two-dimensional data such as x ray and fleshpoint tracking, and from acoustically based tube models of the vocal tract. Recently shape information has been added for the vocal tract and the tongue in multiple planes using MRI (Baer *et al.*, 1987, 1991; Moore, 1992) and ultrasound (Stone *et al.*, 1988, 1991, 1992; Stone, 1990). These techniques provide full 3-D tongue surfaces, though with limits of their own. The MRI cannot image the tongue–teeth interface, and ultrasound can lose entirely the tongue tip and lateral margins, due to the presence of air beneath these unattached portions of the tongue. Within these limits, however, 3-D tongue surface shapes can be reconstructed. The present paper is the first time 3-D ultrasound, a fairly new technique, has been used to capture lingual articulation.

Both consonants and vowels were examined in this study. A large number of readily observable features distinguish consonants from vowels acoustically and physiologically. For example, most vowels are produced with a relatively open vocal tract, consonants with a more obstructed one. Vowels tend to have large airflows and small intraoral pressure, while consonants have restricted airflows and large intraoral pressure. Vowel durations are longer and more sensitive to rhythm and rate changes than consonant durations. Vowels have low-frequency spectral energy and one sound source (phonation). Consonants have high-frequency spectral energy and up to three sound sources (phonation, frication, burst). Finally, vowels and consonants behave linguistically as different types of entities in that vowels are syllable nuclei whereas consonants typically occur as the syllable onset and coda. In fact some researchers have postulated two systems of production for consonants and vowels (Ohman, 1966; Fowler, 1977; Browman and Goldstein, 1990, 1992; Smith, 1993; Stone *et al.*, 1992; Stone and Lundberg, 1994).

Although there are many differences between consonants and vowels, it is not necessarily true that their production is controlled differently, at least with respect to the basic mechanisms that underlie their motions. The present study addresses the organization of commands subserving tongue shapes in speech through a comparison of the forms of 3-D tongue shapes and tongue–palate contact patterns. By comparing tongue shapes and EPG patterns for vowels and consonants it may be possible to assess the extent to which their basic organization is comparable. Different control parameters used in alternation, as consonants and vowels often are, would allow faster production speeds and reduce fatigue. Superficial differences, however, do not always indicate different strategies of production. The present study is unable in its design to shed much light on the control issue because it is simply a representation of 3-D tongue shapes and tongue–palate contact patterns. In addition, much tongue tip information is lost due to the ultrasound technique itself. The study will, however, compare tongue shapes and tongue–palate contact patterns from consonants and vowels to see if they differ.

I. METHODS

A. Data Collection

1. Ultrasound

An ultrasound image is a visual representation of density changes in a 2-D slice of tissue along the transducer's crystal array axis, using a 256-gradation gray scale. The surface of the tongue is a tissue–air interface and the largest density change in the scan. It is visible as the lower surface of a bright white line.

Ultrasound data for static speech sounds were collected at the Johns Hopkins University using a developmental 3-D ultrasound machine courtesy of Acoustic Imaging Inc. (Phoenix, AZ). The 3-D ultrasound transducer has a single curvilinear array of 128 ultrasound crystals that scans one 90° sector at a time. Each sector is scanned in 33 ms. Using a motorized pivot, the single array is moved, 1° at a time through a 60° arc, making a polar sweep of the 3-D space. In the present study, the transducer sweep collected 60 slices in the coronal plane, each 1° apart, in about 10 s. The tip of the tongue and lateral margins (especially anteriorly) often were not imaged due to air beneath, and the tongue root was sometimes obscured by the hyoid. As a result, the deformation of the edges of the tongue surface around the teeth are not seen. For the 18 sounds measured here, the number of slices needed to represent the entire tongue surface was between 42 and 55. The ultrasound images were stored as TIFF (tagged image file format) images, and computer image processing software was developed to reconstruct them into 3-D surfaces. Electropalatographic data were used to interpret and complement the tongue shape data.

2. Electropalatography (EPG)

The EPG data were collected using the Kay Elemetrics Palatometer 6300 system (Lincoln Park, NJ), at a separate session from the ultrasound data. The subject was custom fitted with a 0.5-mm-thick acrylic palate that covered the hard palate and the inner and outer surfaces of the teeth. Ninety-six electrodes were embedded along the surface of the palate and the inner edges of the teeth. The electrodes were sampled at 100 Hz. The schematic presentations of the EPG are a close representation of electrode locations in the mouth. The lateral-most row of electrodes was on the inner surface of the teeth, near the cutting edge of the molars and the gingival edge of the incisors, cuspids, and bicuspids. The second-most-lateral row was on the gingiva, at the dental edge. The posterior row was just anterior to the hard palate/soft palate junction.

3. Speech acoustics

The speech wave was recorded simultaneously with the EPG data and analyzed using Computer Speech Lab (Kay Elemetrics, Lincoln Park, NJ). Formants were extracted using LPC analysis of a steady central portion of each voiced sound.

4. Subject and speech materials

The speaker was a normal adult female speaker of English (age 26). She was a phonetically naive, native of Maryland and had a slight regional accent. The vowels were /i/, /ɪ/, /e/, /ɛ/, /æ/, /ɑ/, /ɔ/, /o/, /ʊ/, /u/, /ɝ/, and /ʌ/. The speaker had no difficulty sustaining the lax vowels. In order to prevent her tendency to diphthongize the tense vowels (especially /e/ and /o/), she was instructed to take a deep breath and sustain the vowel without diphthongization, until told to stop. She was told this would be about 15 s. The subject was also told to imagine the syllable would end with a /p/. She was not told to stop production until well after the ultrasound sweep was completed. In general the subject did not diphthongize. If diphthongization was heard during data collection, the subject was reinstructed and the sound rerecorded. For the ultrasound data collection, the vowels were produced in /pV(p)/ syllables with the vocalic portion of the syllable sustained for 10 s. For the EPG data collection the vowels were produced in /pVp/ syllables with a slight prolongation of the vowel.

The consonants were /s/, /ʃ/, /l/, /n/, /ŋ/, and /θ/. Nasals, rather than stops, were used to facilitate prolongation. For the ultrasound data collection, the subject said /ɑC/ and sustained the consonant for 10 s. The /ɑ/ context was used to stabilize production of the sustained consonants. For EPG data collection, the same consonants were spoken normally in /ɑCɑ/ context. The utterances were produced as spondees with special care taken that the first /ɑ/ was not reduced. Three glides were produced with their homorganic vowel: /iji/, /uwu/, and /ɝrɝ/, as well as in /ɑCɑ/ context. The sustainable consonants, all except /j/ and /w/, were additionally produced as sustained sounds to mimic the ultrasound data collection.

B. Data analysis

1. Data reduction and scaling

During data analysis, the EPG frame of maximal/minimal contact was chosen. For /e/ and /o/, a frame in the first third of the vowel was chosen to minimize the effects of diphthongization. For the ultrasound images, a custom edge detection program detected the surface profile of the tongue in each slice (Unser and Stone, 1992). These surface profiles were then stored as a series of points (*xy* coordinates). Unlike commercially available ultrasound machines, the 3-D transducer did not have a metric scale imposed on the video image. To obtain accurate scaling of the data, therefore, we collected 3-D volumetric ultrasound data of a phantom. An ultrasound phantom is a fluid-filled box containing objects of known size and position used to test and calibrate ultrasound machines. By measuring phantom objects, we were able to obtain scaling information.

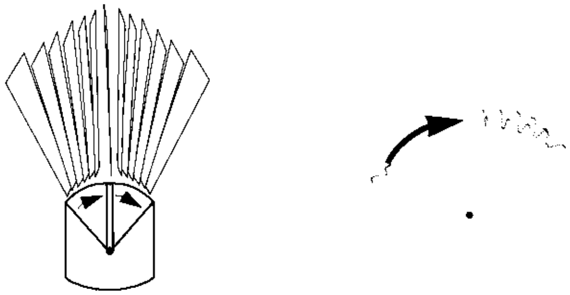

FIG. 1. A 3-D volumetric ultrasound transducer (schematic) that collects 60 ultrasound slices each one degree apart (left) and a 3-D reconstruction of a tongue surface from the individual slices (right). For this study the transducer was held to collect coronal slices with the first slice immediately posterior to the mandibular symphysis.

2. Reconstruction of 3-D tongue geometry

The *xy* coordinates for each slice were restored to their relative 3-D coordinate locations using the *virtual pivot point* inside the transducer (see Fig. 1). Since the geometry of the 3-D transducer was known to us, we were able to determine the distance from the pivot point to the surface of the tongue and the angle of each slice. Each tongue contour was rotated about the pivot point to restore it to its relative 3-D position. For a full set of 60 slices, the first contour would be rotated 60° anterior to the pivot point and each successive contour would be rotated backward 1°. This rotation would place the set of slices at angles 60–120, symmetrically about the 90° vertical. The actual slice sets contained 42–55 tongue surface contours, because the beginning and end slices were often anterior or posterior to the accessible part of the tongue. Therefore, the first image with a visible contour (and the subsequent images) was rotated backwards 1° to account for each missing anterior image. This assured that the positions of the slices were consistent and comparable across reconstructions.

To complete the tongue surface reconstruction, the surface points between the detected data points had to be filled in. Because the tongue is basically smooth, spline interpolation was used to do this. For each of the 2-D profiles we calculated an interpolating cubic *B*-spline that passed each of the data points with the smoothest possible curve. This was done for each profile, and also across the profiles to obtain a grid of *B*-splines. Each square on the grid was a ''patch.'' To define the shape of the local surface within each patch required additional control points. Four interior control points were picked and the bordering *B*-splines were converted to the equivalent cubic bezier segments. This produced a 4×4 grid of bezier control points which described a bicubic bezier patch and defined the patch's surface shape. The bezier patches described a continuous surface, and could be measured for statistical analysis or displayed visually for a more intuitive presentation (Farin, 1993). Once the patch was completed, a program was developed to render and view the

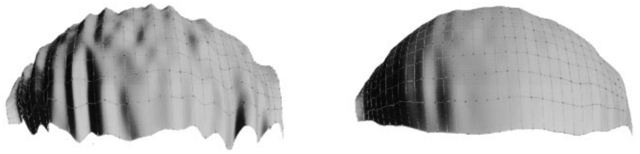

FIG. 2. Lengthwise smoothing of the upper surface of an egg-shaped phantom to reduce interslice measurement error.

reconstructed tongue surface, as well as perform analyses on the surface. The results are seen below in the 3-D surface figures.

C. Validation

1. Reconstruction validity

To validate the surface geometry of the reconstructions we used an ultrasound phantom developed specifically for calibrating 3-D volumetric transducers (CIRS, Norfolk, VA). The phantom object was a 3-D egg encased in a cube of polymer designed to mimic the acoustic properties of soft tissue. The egg itself was also made from tissue approximating polymer, but with different acoustic density. The egg phantom was scanned with the 3-D transducer, and its surface was reconstructed using the same procedure designed for the tongue surface reconstructions. Edge detection for the phantom data was more difficult than for the tongue surface data; in contrast to the strong density of the tongue/air interface, the egg surface was distinguishable only by a change in the scattering properties of the material (producing a low contrast edge). This required some input constraint in the edge detection and introduced some error. The egg's surface was reconstructed from the detected contours, and then compared to the known surface equations. We probed a large set of points regularly spaced over the reconstructed surface. We then measured the distance from a measured point to the nearest point on the idealized surface of the egg. The errors for the reconstructed phantom were as follows. The average error was 0.6273 mm, the standard deviation from the average error was 0.4523 mm, standard deviation from true was 0.7733 mm, and worst error was 2.028 mm. This was beyond the measurement error inherent in ultrasound (0.5 mm), but from the data appeared to be error introduced in the edge detection. We are continuing to improve the precision of the algorithm (Fig. 2).

2. Error reduction

Ultrasound data has the drawback of being fairly noisy. This results in difficulty in detecting the edges in the 2-D images. The edge detection program allowed operator intervention to help pick the best edges, but this did introduce some human error. The errors were not large, but tended to vary from slice to slice. To minimize these errors, we used a smoothing algorithm to realign the successive slices. Smoothing was performed on the control points of the B-spline grid. The squared distance between neighboring spline control points was minimized subject to a lambda term that balanced smoothing versus change in the control point set. In a single spline this minimized the equation

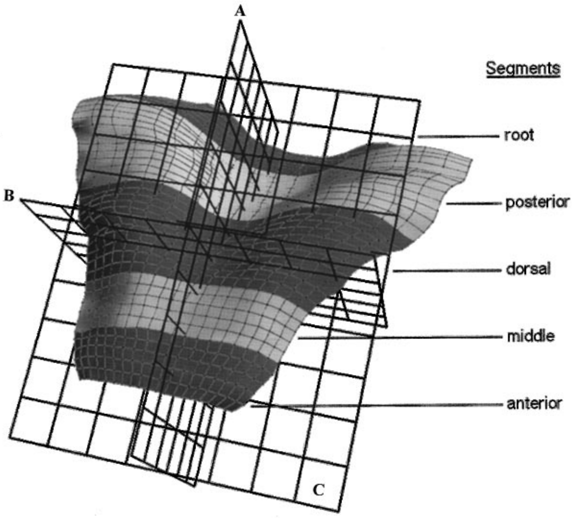

FIG. 3. Three-dimensional reconstruction of sustained /ʃ/ with lengthwise tongue segments labeled. Anterior is on the lower left.

$$\sum_{i=1}^{\substack{\text{spline}\\ \text{pts}-1}} \text{dist}(p^{i}_{\text{orig}},p^{i}_{\text{new}})^2+\lambda*\text{dist}(p^{i}_{\text{new}},p^{i+1}_{\text{new}})^2.$$

This was done simultaneously lengthwise and crosswise, with separate lambda terms to control separately the two directions of smoothing. Since the introduced error was primarily between untrasound slices, the data were optimized in the lengthwise direction, but not the crosswise one. This had the effect of smoothing out the differences from slice to slice, while not having much effect on the shape of each of the coronal slices (see Fig. 2).

II. RESULTS

A. Ultrasound data

Eighteen sounds of English were measured in this study. Recall that the images seen here are positioned relative to *jaw* position, not palate position, because the transducer was not displaced from the jaw when the measurements were made. The images and axes have been rotated to allow optimal viewing of the entire tongue surface and are drawn in perspective. The anterior tongue is on the lower left and the shapes are described according to the five lengthwise segments defined in Stone (1990) which from front to back are: anterior, middle, dorsal, posterior, root (see Fig. 3). The images were placed within the grids according to their actual position in the mouth, so that tongue surfaces that were farther back in the mouth appear farther back in the grid. In some cases, especially for high vowels, the tongue tip and lateral margins were not viewed due to air beneath the surface. The axes in these figures depict the (A) sagittal, (B) coronal and (C) transverse planes. Each square on the axes represents 0.5 cm.

1. Vowels

The 3-D tongue surfaces in Figs. 4 and 5 showed three basic tongue shape categories for vowels. In the first shape category, maximum tongue displacement (maxD) occurred

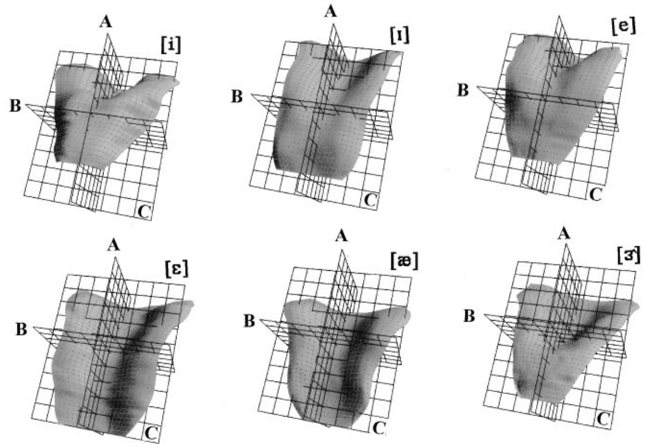

FIG. 4. Three-dimensional reconstructions of English front vowels and /ɝ/.

in the anterior portion of the tongue, and the midline approached the upper border of axis *A*. This shape was seen in the /i/, /ɪ/, /e/, and /ɝ/ (Fig. 4, top and lower right). For /ɪ/ and /ɝ/, the anterior and middle segments primarily were elevated; for /i/ and /e/ the dorsal segment was also. This shape was consistent with the first factor isolated in midsagittal tongue shapes by Harshman *et al.* (1976) and Jackson (1988): front raising. Unlike Harshman *et al.*, the term front raising (and later back raising) is not defined in the present paper using principal components, but rather by visual inspection. The present data set is consistent with a continuum of within-category shape changes, but is too small to strongly support that conclusion. Moreover, we do not believe these shapes should be treated as weighted sums of the elementary features front and back raising, but rather as unique categories whose most salient feature is captured by the category shape. At maxD, the tongue exhibited its maximum outward curvature (convexity) in the coronal plane.

Consistent with volume preservation, this convexity was accompanied by a complementary groove (concavity) in the posterior tongue, and the posterior groove depth was greater for phones with a more elevated anterior tongue. For this subject the tense /e/ had a higher anterior tongue and deeper posterior groove than the lax /ɪ/. The sustained /ɝ/ was treated as a vowel, because on the EPG data it was found to have less contact than the consonantal /r/ (see Fig. 9 below) and a virtually identical contact pattern to the vocalic /ɝ/ in /pɝp/ (Fig. 8 below). Acoustically, it fell between the consonantal and vocalic /r/ spectra (Table I below).

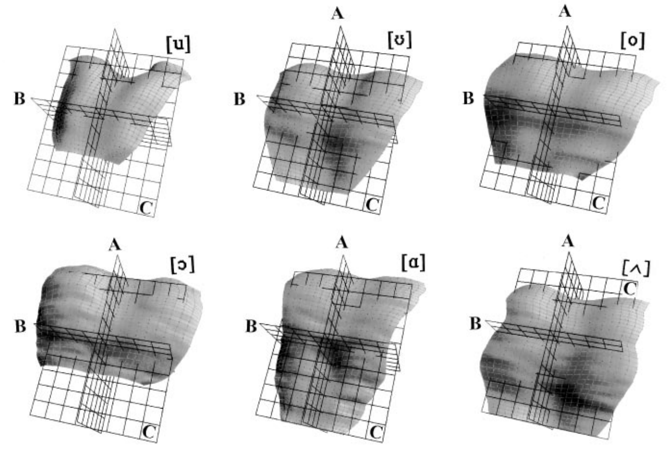

FIG. 5. Three-dimensional reconstructions of English back vowels and /ʌ/.

The second 3-D shape, complete groove, was found in the low vowels (/ɛ/, /æ/). For these vowels the tongue surface shape was channellike, the lateral margins of the tongue were elevated relative to the midsagittal plane, and maxD retained a substantial groove. The complete groove shape, with a shallower groove, was also seen in /ʌ/ (Fig. 5).

The vowels /ʊ/, /o/, /ɔ/ and /ɑ/ (Fig. 5) displayed a third tongue surface shape. The /u/ here appeared to have shape one (front raising). Previous x-ray data (cf. Harshman *et al.*, 1976) indicate that this sound often has shape three (back raising), in which case ultrasound may not have imaged the tongue tip. The formant patterns (Table I below), however, are consistent with an anterior tongue position for /u/ as shown here. Moreover, an anterior /u/ is not uncommon in the Maryland dialect. In shape three, maxD occurred more posteriorly than for the front vowels (cf. relative to axis *B*). MaxD also exhibited a diminished groove but did not become convex coronally, even for the /ʊ/. This third shape was consistent with the second midsagittal factor identified by Harshman *et al.* (1976) and Jackson (1988): back raising. Max *D* occurred in the dorsal segment for /o/, /ɔ/, and /ɑ/, in the middle segment for /ʊ/, and in the middle and anterior segments for /u/.

Two other surface features were specific to back vowels. The back vowels, except /u/ (Fig. 5), had a short midsagittal groove or ‘‘dimple’’ anterior to maxD. We have observed this dimple often in midsagittal scans of /ɑ/ as a natural variant of the sound. The back vowels also had a larger distance between the lateral margins than the front vowels. We believe this was due to lateral spreading of the tongue as a result of the front–back compression used to elevate the posterior tongue. The front vowels would not have this effect because they have posterior, but not anterior, compression, allowing the tongue to protrude forward. We cannot rule out, however, the possibility that the front vowels were equally wide, but had enough air under the lateral margins of the tongue to obscure the edges and create a narrower appearance. Finally, because of ultrasound's inability to image the tongue tip, it is not possible to determine exactly the anterior border of the tongue on these images.

2. Consonants

The consonants used the same three shape categories as the vowels plus one more. Shape one (front raising) was seen in the consonants /n/ and /ʃ/ (Fig. 6, left column). For the /n/ the anterior segment was level in the coronal plane; for /ʃ/ the anterior and middle segments were. Shape two (complete groove) was seen in /θ/ and /s/ (Fig. 6, upper right). Shape three (back raising) was seen in a very extreme form in /ŋ/. Not only was the tongue considerably more displaced for /ŋ/ than any of the back vowels (see axes *A* and *B*), but the surface shape was extremely convex (arched) in the coronal plane throughout the tongue's length, except for a short groove at the root and an anterior leveling. The fourth shape (two-point displacement) was seen exclusively in /l/. The /l/

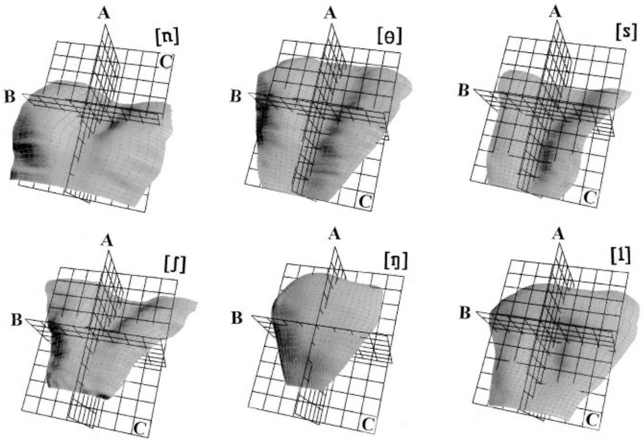

FIG. 6. Three-dimensional reconstructions of six consonants: /n/, /θ/, /s/, /ʃ/, /ŋ/, and /l/.

displayed a raised tip and a groove immediately posterior in the middle segment, which gradually gave way to a raised posterior segment.

To summarize the ultrasound data, four 3-D surface shapes were found for 18 American English sounds. Shape category I, front raising, exhibited an elevation of the anterior and middle segments (with or without an elevated dorsal segment), with a level or arched coronal shape, and a complementary groove in the posterior and root segments. All the sounds that used this shape contacted the palate bilaterally in the alveolar or palatal vault area and included /i, ɪ, e, ɝ, u, ʃ, n/. Shape category II, complete grooving, was found in /ɛ, æ, s, θ/. In the midsagittal plane, the /ɛ, æ, θ/ appeared to have a reduced version of shape one: front raising, while the /s/ was steeper and more linear (see Fig. 7). In the 3-D surface, however, one could see that they all had a continuous channel.

The third shape was consistent with the category back raising. In its less extreme form, found in vowels with minimal palatal contact /ʊ, o, ɑ, ʌ, ɔ/, the dorsal tongue was displaced upward with compression fore and aft. The tongue was always grooved at midline, though the groove diminished at maxD. In its more extreme form, seen in /ŋ/, the middle and dorsal segments were elevated considerably to produce an extreme arch at maxD. The fourth shape observed, two-point displacement, was found only in /l/, and consisted of two regions of displacement, an elevated anterior and posterior segment, with a short groove, almost a dimple, indicating compression of the middle segment.

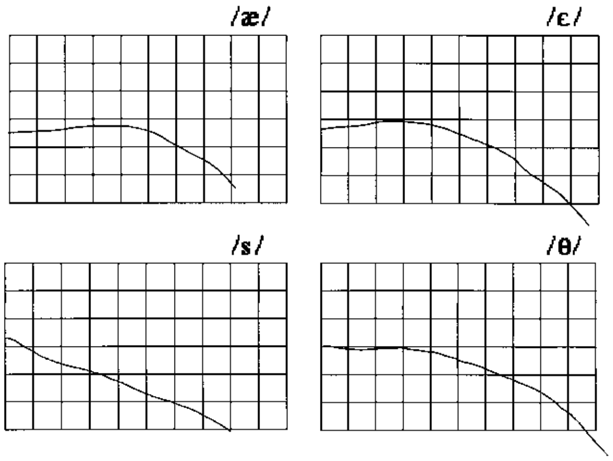

FIG. 7. Midsagittal contours extracted from the 3-D reconstructions for /s/, /θ/, /æ/, and /ɛ/.

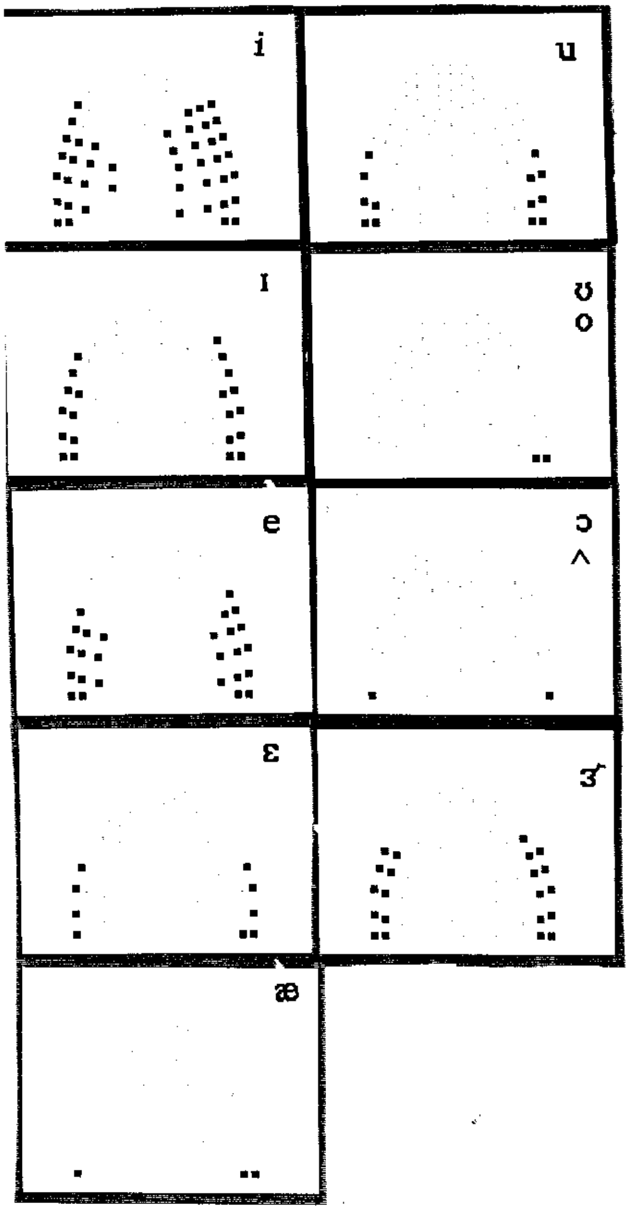

FIG. 8. The EPG patterns for English vowels. The palate contains 96 electrodes.

B. EPG data

1. Vowels

The EPG data for the vowels are seen in Fig. 8. The American English vowels studied here essentially used a single pattern of tongue–palate contact: bilateral contact. In this pattern the lateral margins of the tongue touched the inner molars and gingiva (outer two rows), and sometimes also the lateral margins of the palate. For the front vowels the EPG patterns formed a continuum in which anterior–posterior contact location varied with front–back tongue po-

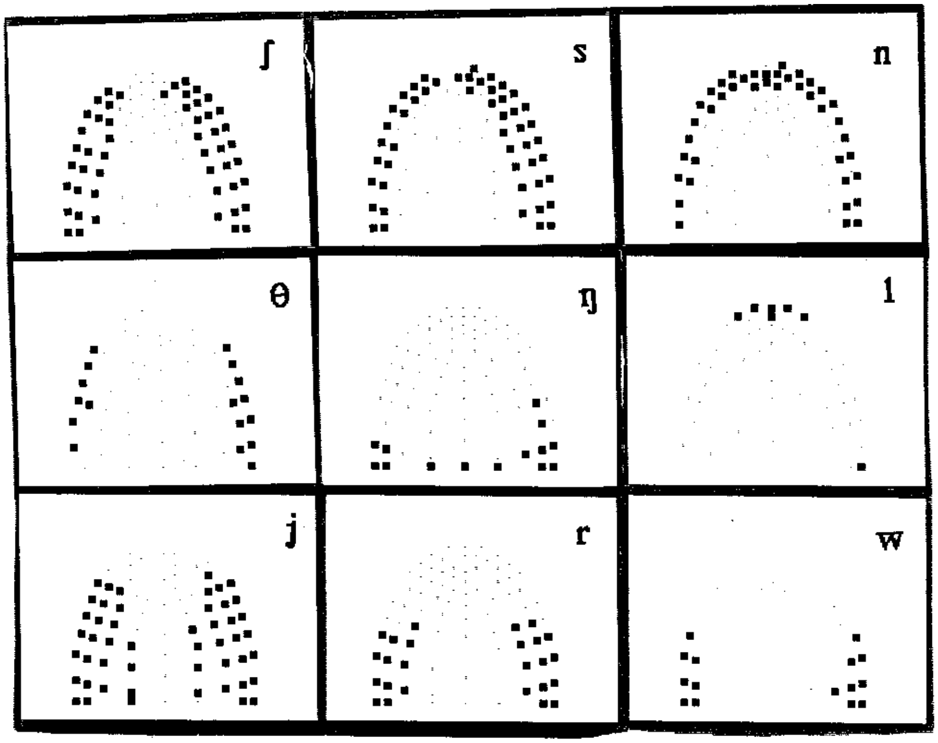

FIG. 9. The EPG patterns for English consonants. The palate contains 96 electrodes.

sition, and medial contact increased with presumed tongue height. The tense /e/ had more medial palatal contact than the lax /ɪ/, reflecting the higher medial tongue. The vocalic /ɝ/ pattern was between /e/ and /ɪ/. Its contacts were very lateral and posterior, apparently reflecting contact with the lateral margins of the dorsal and posterior tongue groove, not the anterior and middle bunching. The back vowels did not follow this straightforward change in pattern. Except for the /u/, the EPG data provided no usable information for the back vowels. This was presumably due to the large palatal vault space, which was not contacted even for the second highest vowels /ʊ/ and /o/.

2. Consonants

Figure 9 presents EPG data for nine consonants. The consonantal EPG patterns were much more varied than those found in vowels. The /ʃ/, /s/, and /n/ were somewhat similar in palatal contact pattern to each other. They had complete lateral occlusion, bilaterally, and more anterior contact than vowels. They were distinguished from each other by a decrease in medial contact posteriorly and an increase in medial contact anteriorly from /ʃ/ to /s/ to /n/. The /ŋ/ and /l/ patterns were even more dramatically different from vowels. For /ŋ/ complete crosswise occlusion occurred, with bilateral contact presumed to occur posterior to the crosswise contact. For /l/, tongue–palate contacts occurred almost exclusively in the front. The /θ/ showed bilateral contact on the teeth but not anteriorly where contact may have occurred on the cutting edge. The visible /θ/ contact pattern was not unlike that of the /ɪ/.

The EPG data for the three glides /r/, /j/, /w/ were compared to their nearest vowels to see if their patterns would reveal consonant-vowel differences. The glides had bilateral contact patterns similar to their homorganic vowels, but with narrower or longer channels. The /j/ had more anterior contact than the /i/ and a longer channel. The /r/ and /w/ had more medial contact than the vowels and a narrower channel; the /w/ in particular may have had more contact posterior to the pseudopalate as well, and may have elevated the tongue more at midline without touching the palatal vault. The consonants, thus, had three tongue–palate contact patterns: bilateral (/s/, /ʃ/, /r/, /j/, /w/, /θ/), crosswise (/l/), and a combination of the two (/n/ and /ŋ/), while the vowels had only one: bilateral.

C. Ultrasound/EPG comparison

A comparison of the tongue shapes and palatal contact patterns revealed several interesting relationships. For category I (front raising), the anterior tongue surface of the vowels was highly coupled to palatal shape, and midsagittal elevation caused more medial palatal contact. In the dorsal segments, where the palatal arch was higher, lateral palatal contact accompanied a midsagittal groove, whose depth was not reflected in the amount of medial contact. For the consonants, the EPG data indicate a more elevated tongue tip than that found in vowels. Although similar in tongue shape, therefore, the tip differences between consonants and vowels suggest possible production strategy differences. For category II (complete groove), consonants and vowels visibly behaved as different groups. The vowels in this group (/ɛ/ and /æ/) were low vowels, and tongue surface shape was made using minimal palatal contact. The consonants (/θ/ and /s/) were high and used considerable palatal contact to create essentially the same shapes. In category III (back raising), /ŋ/ was much more arched than any of the back vowels. For

TABLE I. Formant frequencies for the sounds spoken by the subjects. N.B. The voiceless consonants /s/, /ʃ/, and /θ/ did not have measurable formants.

	Formants		
	F1	F2	F3
Vowels in pVp context			
/i/	313	2784	3130
/ɪ/	563	1944	2980
/e/	554	2428	3161
/ɛ/	748	949	3085
/æ/	952	1862	3103
/ɑ/	957	1477	2802
/ʌ/	704	1485	2930
/ɔ/	706	1262	2676
/o/	585	1244	2838
/ʊ/	536	1180	2861
/u/	367	1461	2730
/ɝ/	488	1430	1934
Consonants in /bɑC/ context			
/n/	355	1782	3109
/ŋ/	375	1288	2389
Consonants in /ɑCɑ/ context			
/l/	529	1101	3175
/r/	489	1230	1688
Sustained sounds			
/ɝ/	485	1465	1783
/l/	455	1376	2928
Homorganic sounds:			
iji: /j/	356	2700	3112
/i/$_2$	395	2443	2879
uwu: /w/	238	1186	2390
/u/$_2$	497	1031	2702
ɝrɝ: /r/	452	1461	1461
/ɝ/$_2$	532	1461	1721

category IV (two-point displacement), the anterior palatal contact of /l/ reflected the anterior and lateral tongue raising, but the posterior raising, in the vault, was invisible.

D. Acoustic data

Formants were tracked for each of the voiced sounds from acoustic data collected simultaneously with the EPG data. The first three formants for each sound appears in Table I. The F2 data indicated that /u/ and /e/ were relatively anterior in tongue position, and the /e/ was high. The vowels showed generally higher F1s and lower F2s than the glides, indicating a more open vocal tract for the vowels.

III. DISCUSSION

In addition to reconstructing accurate 3-D tongue surface shapes, this study was interested in several features of tongue shape. First, how did the complex tongue musculature, with its volume-preserving constraints, affect the interdependence of tongue shape and position? Second, did consonants use palatal bracing to create qualitatively different tongue behaviors from vowels?

A. Complex muscular system: Muscular hydrostat.

Four categories of tongue surface shape emerged. For vowels, these categories supported expected phonetic dimensions, such as, position/shape correlations. Shape 1, front raising, occurred in higher front vowels. Shape 2, complete groove, occurred in lower vowels. Shape 3, back raising, occurred in back vowels. For consonants the first three shapes also occurred, but were not related as directly to tongue position. Shape 1 was found in /n/ and /ʃ/, an alveolar and palatal sound. Shape 2 was found in /s/ and /θ/, an alveolar and dental sound. Shape 3 was found only in the velar; shape 4, two-point displacement, only in the lateral. For consonants, the choice of shape seemed a function of aerodynamic needs and palatal morphology at the constriction site, not global tongue position.

A volume-preserving system must displace and compress local regions in a complementary fashion, because its internal volume can be shifted, but not reduced or increased. In the tongue, volume shifting was observed readily, as in the vowel /i/, where upward displacement at the front of the tongue was accompanied by inward compression at the back. This tradeoff between local compression and displacement, a basic volume-preserving principal, appeared to be the basic mechanism for positioning the tongue in vowels. The present data, however, are of tongue surfaces only, and full 3-D volume data are needed to confirm this hypothesis. Tongue movement could conceivably have been executed by maintaining a constant global configuration and using extrinsic muscles for positioning, as the tongue body does in some models (Mermelstein, 1973). In that case, however, tongue shape would not vary systematically with tongue position.

The three shape categories observed in the vowel data and their interdependence with tongue position suggested constraints on the tongue's ability to produce all shapes in all positions. For example, complete grooving (shape 2) was produced only in low front vowels; front raising, in mid-high front vowels. In such a tightly coupled relationship, both the shape and position features for the gesture need not be planned. One could simply be a consequence of the other. To allow these few shapes to be used in other positions, however, e.g., to form a channel using an elevated tongue position, as for /s/ and /θ/, or a complete occlusion, as for /n/ and /ŋ/, the boundary structures of the mouth offer an additional means of manipulating the tongue surface shape. In that case, shape and position both appear to be planned, because often slight differences in tongue–palate contact, for example in /s/ and /ʃ/, do not predict the large cross-category differences in tongue shape.

B. Palatal contact patterns

Examination of tongue surface shape in conjunction with the EPG pattern provided more understanding of consonant–vowel differences. In the vowel data a bilateral contact pattern was used exclusively. The bilateral contacts were located either on the middle or posterior palate, though contact may have continued posteriorly to the pseudopalate, especially for the back vowels. The EPG patterns for vowels roughly but systematically reflected tongue shape changes for front vowels, but not for back vowels, due to the steeply arched palatal vault. Vertical height from the gingiva to the highest point of the palatal vault was 1.4 cm.

The consonants explored in this study were /ʃ, s, n, θ, ŋ, l, r, j, w/. They represent almost the entire spectrum of linguopalatal consonants in English, because the tongue shapes for /s/, /θ/, and /ʃ/ are roughly comparable to /z/, /ð/ and /ʒ/, that of /n/ is comparable to those /t/ and /d/, and that of /ŋ/ is comparable to those of /k/ and /g/.

Many consonants used a bilateral contact pattern. These patterns differed from the vocalic ones, however, by being either longer (/s/, /ʃ/) or having a narrower channel (/r/, /w/). The /θ/, which was very vocalic looking, is predicted to approximate or contact the cutting edge of the anterior teeth, which was not visible in these data, also creating a much longer channel than seen for vowels. In addition to the bilateral-contact pattern, there were two other EPG contact patterns, both of which were unique to consonants. One was a crosswise-contact pattern, which produced a laterally directed airflow (seen in /l/). This pattern was never even approximated in vocalic productions and was far outside the continuum of vowel EPG patterns. The other purely consonantal pattern was a combination of the crosswise and bilateral patterns (seen in /n/ and inferred in /ŋ/). In the ''combination pattern,'' the bilateral contacts were posterior to the crosswise contact to create a complete vocal tract occlusion. By definition, vowels cannot have a complete occlusion of the vocal tract, precluding the combination pattern.

Tongue patterns alone suggest similar consonant and vowel production strategies. The EPG patterns taken alone provided strong support for different tongue control for consonants and vowels. Therefore the two data sets must be understood each in the context of the other.

C. Control of tongue positions involving contact forces: Muscular hydrostat with bracing

We have argued previously that consonants have a larger range of tongue shapes and EPG patterns than vowels (Stone, 1995; Stone and Vatikiotis-Bateson, 1995). Preliminary analysis of a few surfaces from the present data set (Stone and Lundberg, 1994), as well as assumptions based on midsagittal x-ray and midsagittal and single-slice coronal ultrasound images, led us to believe that shapes such as the midsagittal channel and the two-point displacement of /l/ would not be used by the vowels. This expectation was supported in the latter case, but not the former.

Since the tongue surface shapes for consonants and vowels had many common features, it was useful to consider the differences between the ultrasound and the EPG categories. These two data sets did not categorize the sounds comparably. First, three tongue shape categories were needed to describe the vowels (front raising, continuous channel, back raising); only one palatal contact category was needed: bilateral. Second, only one consonant had a nonvocalic tongue shape (/l/); virtually all had nonvocalic palatal patterns. These differences suggested that for consonants, vocalic tongue shapes were positioned in nonvocalic ways against the palate. This created global vocal tract shapes unique and appropriate to each consonant. For example, the /s/ and /θ/ tongue shapes used complete grooving, like the /ɛ/ and /æ/. The EPG data, however, indicated that the methods of producing the consonant and vowel tongue shapes were quite different. The tongue–palate channel created for the fricatives, which ensured a narrow, high-pressure, anteriorly directed air jet, needed more upward force, a complete lateral seal, and precise shaping of the groove against the palate. Thus, although no new tongue shapes were created, the complete groove was produced in a high tongue position through the use of palatal bracing.

In another example, /n/, /ʃ/, and /i/ all had front raised tongue shapes. The EPG contact patterns, however, were not comparable directly and suggested differences in tongue control for the consonant and vowel sounds. For the /n/ and /ʃ/, palatal contact extended more anteriorly and was occluded or funnel shaped. The /i/ contact pattern had a ''reverse'' funnel shape. Examination of the tongue surfaces revealed that the midsagittal elevation for /i/ included the anterior, middle, and dorsal sections of the tongue, for /n/ and /ʃ/ only the anterior and middle. The EPG also indicated that the unseen tongue tip was elevated for /n/ and /ʃ/, but not for /i/. The consonants' longer, more anterior, less arched shape, although still technically front raising, was different from that seen for front raised vowels, whose tips were lower than maxD. Thus, as with the complete groove, no new shape was created, but subtle control of local differences in the location and length of the raised front coupled with the flatter shape of the anterior palate created the fricative channel.

Let us now consider the remaining consonants, /ŋ/ and /l/. Their tongue shapes fell into two different categories. For /ŋ/, the tongue was back raised, and used an arched shape throughout its length, except in the tongue root and blade, which were compressed to help displace the dorsal tongue. The /ŋ/, although consistent with back raising, was unique in its extended midsagittal arching. The /l/ tongue shape was unique. The /l/ provided support for the palatal bracing theory of consonant production in that the contact pattern and tongue shape were unique in the data set and quite well related to each other. This shape and EPG contact pattern have been discussed extensively as an example of palatal bracing elsewhere (Stone, 1990; Stone *et al.*, 1992).

Two things were remarkable about these findings. The first was how similar the tongue shapes were and how few categories described the 3-D surfaces. The data suggested that the tongue, although flexible in its 3-D shape, nonetheless used a relatively small repertoire of shapes. The second and equally interesting finding was that the tongue used its limited shape repertoire to produce a large variety of vocal tract shapes. This variety appeared to be the result of positioning the tongue in various vocal tract locations, either entirely by muscular means or in conjunction with the palate. The complete groove shape (/θ/, /s/) created a tongue–palate channel when contacting the palate. The front raising shape created a tongue–palate channel (/ʃ/) or an occlusion (/n/), depending on where the tongue was placed and which segments were raised. The back raising shape created an occlusion (/ŋ/) and probably also a tongue–palate channel (/w/—no tongue data). Lowering of the lateral margins (/l/) diverted airflow laterally.

The /l/ was the only tongue shape to strongly support the notion of different, rather than more extreme, tongue shapes for consonants than vowels. The /ŋ/, /n/, and /ʃ/ provided weak support, because their extreme arching (/ŋ/) and elevated tip (/n/, /ʃ/) could be interpreted as outside the vocalic continuum. Better information about the tongue tip, and the addition of more sounds (e.g., retroflex /r/), might reveal more consonant shape variety. The MRI images of retroflex /r/ (Ong, 1995) indicate that it too uses two-point displacement, with an elevated tongue tip, a very displaced (and arched) posterior/root region, and compression between the two. The vocalic /ɝ/ in the present study leveled off posteriorly (Fig. 4), suggesting a slight bulge in the root, below the scan. A consonantal bunched /r/ might show a true two-point displacement.

It does seem, however, that most American English sounds employed a limited number of shapes. Consonants appeared to be distinguished from vowels by more extreme shapes, by subtle differences in shape, and by interaction with the palate to manipulate tract shape and create additional sound sources, rather than by large qualitative shape differences. The preponderance of bilateral tongue–palate contact patterns also suggests a ‘‘basic’’ nature to this pattern, consistent with nonspeech contact patterns, such as sucking and swallowing (Hamlet *et al.*, 1988), and this pattern is well adapted to the low lateral margins of palatal morphology.

IV. SUMMARY

This paper presented three-dimensional reconstructions of tongue surfaces from ultrasound data for 18 sounds of English. The tongue data indicated a limited repertoire of tongue surface shapes. These shapes were classified using four categories: (1) front raising, (2) complete groove, (3) back raising, and (4) two-point displacement. The fourth shape was used only for /l/. The other shape categories contained both consonant and vowel shapes.

The EPG data revealed three tongue–palate contact patterns. These were (1) bilateral contact, (2) crosswise contact, and (3) a combination of the two. Vowels used only the bilateral contact pattern. Consonants used all three.

Vowels and consonants were not distinguished categorically by tongue shape, but they were distinguished by EPG patterns. It appeared that the consonants used tongue shapes mostly similar to vowels in different vocal tract locations and with varied palatal contact patterns to create the variety of vocal tract shapes used in speech.

ACKNOWLEDGMENTS

This project was supported in part by Grant No. 7 R01 DC 01758 from the National Institute on Deafness and Other Communication Disorders, NIH, and the Center for Language and Speech Processing, Johns Hopkins University. The authors would like to thank Eric Vatikiotis-Bateson and David Ostry for their helpful comments on the manuscript.

Abd-El Malek, S. (**1939**). ‘‘Observations on the morphology of the human tongue,’’ J. Anat. **73**, 201–210.

Baer, T., Gore, J., Boyce, S., and Nye, P. (**1987**). ‘‘Application of MRI to the analysis of speech production,’’ Magn. Reson. Imag. **5**, 1–7.

Baer, T., Gore, J., Gracco, C., and Nye, P. (**1991**). ‘‘Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels,’’ J. Acoust. Soc. Am. **90**, 799–828.

Browman, C., and Goldstein, L. (**1990**). ‘‘Tiers in articulatory phonology, with some implications for casual speech,’’ in *Papers in Laboratory Phonology 1: Between the grammar and physics of speech*, edited by J. Kingston and M. Beckman (Cambridge U.P., Cambridge), pp. 341–376.

Browman, C., and Goldstein, L. (**1992**). ‘‘Articulatory phonology: An overview,’’ Phonetica, **49**, 155–180.

Carpentier, P., and Pajoni, D. (**1989**). ‘‘La Langue: un ensemble musculaire complexe. (The tongue: A complex muscular array),’’ Rev. Orthop. Dento. Faciale **23**, 19–28 (in French).

Farin, G. E. (**1993**). *Composite Surfaces and Spline Interpolation in Curves and Surfaces for Computer Aided Geometric Design: A Practical Guide*, edited by G. Farin (Academic, London), 3rd ed., pp. 294–301.

Fowler, C. (**1977**). ‘‘Timing Control in Speech Production,’’ Ph.D. dissertation, University of Connecticut.

Hamlet, S. L., Stone, M., and Shawker, T. (**1988**). ‘‘Posterior tongue grooving in deglutition and speech: Preliminary observations,’’ Dysphagia **3**, 65–68.

Harshman, R., Ladefoged, P., and Goldstein, L. (**1976**). ‘‘Factor analysis of tongue shapes,’’ J. Acoust. Soc. Am. **62**, 693–707.

Jackson, M. (**1988**). ‘‘Phonetic theory and cross-linguistic variation in vowel articulation,’’ in *UCLA Working Papers in Phonetics*, Vol. 71 (UCLA, Los Angeles).

Kier, W., and Smith, K. (**1985**). ‘‘Tongues, tentacles, and trunks: the biomechanics of movement in muscular-hydrostats,’’ Zool. J. Linnaean Soc. **83**, 307–324.

Mermelstein, P. (**1973**). ‘‘Articulatory model for the study of speech production,’’ J. Acoust. Soc. Am. **53**, 1070–1082.

Miyawaki, K., Hirose, H., Ushijima, T., and Sawashima, M. (**1975**). ‘‘A preliminary report on the electromyographic study of the activity of lingual muscles,’’ Ann. Bull. Res. Inst. Logoped. Phoniat. Tokyo University, **9**, 91–106.

Moore, C. (**1992**). ‘‘The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images,’’ J. Speech Hear. Res. **35**, 1009–1023.

Ohman, S. (**1966**). ‘‘Coarticulation in VCV utterances: Spectrographic measurements,’’ J. Acoust. Soc. Am. **39**, 151–168.

Ong, D. (**1995**). ‘‘Reconstruction of vocal tract shape from magnetic resonance images during production of [r] and [l],’’ unpublished Master’s thesis, Johns Hopkins University, Department of Biomedical Engineering, Baltimore, MD.

Recasens, D., Fontdevila, J., and Pallares, M. (**1995**). ‘‘Velarization degree and coarticulatory resistance for /l/ in Catalan and German,’’ J. Phonet. **23**, 37–52.

Smith, C. (**1993**). ‘‘Prosodic patterns in the coordination of vowel and consonant gestures,’’ Haskins Lab. Stat. Rep. Speech Res. **SR 115/116** (July–Dec. 1993), 45–55.

Stone, M., Shawker, T., Talbot, T., and Rich, A. (**1988**). ‘‘Cross-sectional tongue shape during the production of vowels,’’ J. Acoust. Soc. Am. **83**, 1586–1596.

Stone, M. (**1990**). ‘‘A three-dimensional model of tongue movement based on ultrasound and x-ray Microbeam data,’’ J. Acoust. Soc. Am. **87**, 2207–2217.

Stone, M., Faber, A., and Cordaro, M. (**1991**). ‘‘Cross-sectional tongue movement and tongue-palate contact patterns in /s/ and /ʃ/ syllables,’’ in *Proceedings of the XII International Congress of the Phonetic Sciences*, Aix-en-Provence, France, Vol. 2, pp. 354–357.

Stone, M., Faber, A., Raphael, L., and Shawker, T. (**1992**). ‘‘Cross-sectional tongue shape and lingua-palatal contact patterns in /s/, /ʃ/, and /l/,’’ J. Phonet. **20**, 253–270.

Stone, M., and Lundberg, A. (**1994**). ‘‘Tongue-palate interactions in consonants vs. vowels,’’ in *Proceedings of the Third International Conference on Spoken Language Processing*, Yokohama, Japan, pp. 49–52.

Stone, M. (**1995**). ‘‘How the tongue take advantage of the palate during speech,’’ in *Studies in Speech Production: A festschrift for Katherine Safford Harris* (American Institute of Physics, New York), Chap. 10.

Stone, M., and Vatikiotis-Bateson, E. (**1995**). ‘‘Trade-offs in tongue, jaw, and palate contributions to speech production,’’ J. Phonet. **23**, 81–100.

Unser, M., and Stone, M. (**1992**). ‘‘Automated detection of the tongue surface in sequences of ultrasound images,’’ J. Acoust. Soc. Am. **91**, 3001–3007.