Comparison of speech production in upright and supine position

M. Stone,[a] G. Stock, K. Bunin, K. Kumar, and M. Epstein
Department of Biomedical Sciences, University of Maryland, Baltimore, Maryland 21201

C. Kambhamettu and M. Li
Department of Computer Science, University of Delaware

V. Parthasarathy and J. Prince
Department of Electrical Engineering, Johns Hopkins University

(Received 7 July 2006; revised 16 February 2007; accepted 20 February 2007)

Speech is usually produced in an upright sitting or standing posture. Measurements and judgments of speech may be made in conditions requiring a supine position, however. These conditions include MRI recordings, and oral procedures, such as, adjustments to dental appliances, medical and surgical procedures. It is of interest, therefore, to see whether gravity has strong or systematic effects on tongue behavior. In the present study, 13 subjects repeated several words, which contained extreme consonant and vowel tongue positions, during upright and supine condition. Ultrasound imaging provided midsagittal tongue contours, in each condition, for comparison. A neck brace was used to stabilize transducer placement and the palate was used as a physiological reference to register the data sets. Results showed a significant subject effect. In supine position the tongue was more posterior than upright for seven subjects, more anterior for two subjects and varied by phoneme for four subjects. However, there was no significant phoneme effect. The direction of change and the amount of change were not directly related. Most subjects had small upright-supine differences. The largest differences, less than 3 mm on average, were in the posterior tongue.
 [DOI: 10.1121/1.2715659]

PACS number(s): 43.70.Aj, 43.70.Jt [BHS] Pages: 532–541

I. INTRODUCTION

Gravitational force is one of the major external loads experienced during speech (Shiller *et al.*, 1999), but it is not well known how the tongue compensates for this load or how this compensation differs in upright versus supine position. Under ordinary circumstances speech is produced in an upright position. However, in situations like sleeping, dental procedures, and MRI recordings, the subject is supine. In sleep, tongue position is of interest due to its relationship with sleep apnea. Rose *et al.* (2002) concluded that alteration of head posture and tongue position alone had a significant effect on the severity of obstructive sleep apnea. In dentistry, body and head position are also a concern. The measurements and placement of an upper removable appliance are often done in supine position, and errors can affect tongue position and disturb tongue function (Reinicke *et al.*, 1998). In MRI, vocal tract recordings are made in supine condition but are used in "upright" models of speech production. The present study is particularly focused on the effects of gravity on the tongue in studies using MRI.

Several methods have been used to study the effects of gravitational orientation on tongue behavior. EMG measurements of muscle activity showed greater genioglossus posterior (GGP) activity for some vowels than others, and for inspiration than expiration. A comparison of these behaviors in upright and supine position showed more GGP activity in supine position (Myamoto *et al.*, 1997; Niimi *et al.*, 1994; Otsuka *et al.*, 2000; Sauerland and Mitchell, 1975). It is likely that increased GGP activation results in a variety of tongue modifications, including maintenance of the upright pharyngeal position, a more anterior tongue position, or a more posterior position, as long as the airway is sufficiently open. These tongue modifications may be subject and/or task specific. Cephalograms have confirmed a variety of responses to gravity. Patients with sleep apnea maintained their upright tongue position when moved to supine condition, whereas nonapneic snorers had significant superior-posterior tongue movement in supine position (Myamoto *et al.*, 1997).

The introduction of MRI to the study of speech affords great advancement in the ability to visualize and quantify the behavior of the tongue (cf. Narayanan *et al.*, 1995; Stone *et al.*, 2001). Unfortunately, the supine data collection of MRI does not perfectly reflect upright speech and may require transformation to simulate upright positions if these data are to be compared to upright data sets. A better understanding of gravitational effects will allow better utilization of MRI data in speech analysis and modeling. Several MRI studies have examined sustained phonemes. Badin *et al.* (2002) found that sustained consonants and vowels had backward displacement of the tongue in MRI images when compared to cineradiographs. Engwall (2006) further found that gravity affected the posterior tongue to a greater extent than the an-

[a]Electronic mail: mstone@umaryland.edu

terior tongue. Kitamura *et al.* (2005) used Open MRI to compare upright and supine tongue positions for three subjects saying five sustained vowels. The results showed subject differences in the effects of gravity orientation. One subject had an almost identical position of the tongue body and posterior pharyngeal wall in both conditions. The other two subjects showed posterior displacement of both the tongue and the posterior pharyngeal wall. The authors proposed differences in head, rather than tongue position, as the major cause. However, the observed tongue retraction was smaller for front vowels, and these two subjects reduced jaw opening in supine position. The authors proposed stabilization of the anterior tongue against the hard palate for front vowels to account for this difference.

Engwall (2006) used MRI to compare supine and prone speech tasks. This work used three-dimensional reconstructions of supine and prone tongue surfaces for vowels and fricatives sustained for 5 seconds, and midsagittal contours of VCVs collected at nine frames per second. Overall, differences seen in supine versus prone condition were small and varied with phoneme. The effects of gravity were larger for sustained phonemes than speech articulations. In supine condition, the tongue was more posterior especially for front vowels reducing the tongue shape differences between /i/, and the back vowels /ɑ/, /u/.

Tiede *et al.* (2000) used x-ray microbeam measures of the tongue and jaw, to study two subjects saying vowels, /bV/ syllables, and phrases, in sitting and supine position. Results for the sustained vowels reflected subject differences. One subject had a slightly posterior jaw, and a slightly upward and anterior tongue when supine. The second subject had virtually no mandibular displacement, and a slightly posterior tongue when supine. During movements, however, the two subjects behaved similarly across gravity conditions. For jaw movements, both subjects showed reduced range of motion, and reduced opening/closing gestures, when supine. For tongue movements, both had minimal variability in the region of the constriction when comparing supine to upright. Differences in tongue position were somewhat dependent on the type of task. As with Engwall (2006) tongue position in static and /bV/ vowels was more variable and more affected by gravity than those in phrases. Although the jaw and tongue positions were variable between upright and supine posture, the output acoustics were relatively similar regardless of body orientation. The authors concluded that in supine position articulatory adaptation was idiosyncratic and designed to maintain acoustics, rather than preserve articulatory trajectories.

Shiller *et al.* (1999) modeled the effects of gravity on jaw displacement based on the hypothesis that no compensations to gravity are made. They predicted that in supine position the jaw would shift backward horizontally and rotate away from occlusion. These predictions were compared with empirical data from the nonhigh front vowels (/æ/ and /ɛ/). Five subjects produced CVCs with various consonants in upright, supine and prone positions to compare with the simulations. Measurements were made of jaw motion (Optotrak - Northern Digital, Canada) and vowel formant frequencies. Formant analysis indicated a significant upright-supine difference for /æ/ but not /ɛ/. Jaw measurements, averaged across all five subjects, indicated a significant difference in jaw rotation, but not horizontal jaw translation for five of the six syllables in supine condition. The lack of backward jaw translation indicated that differences in supine jaw motion are not due to gravity alone. The authors concluded that subjects do not completely compensate for gravitational load.

Taken together these studies indicate that the effects of supine position on tongue behavior are both local and global, are lesser for time-varying tokens than sustained phonemes, and vary among subjects. The studies primarily examined vowels. Lingual consonants are also of interest and may exhibit a smaller gravitational effect at their vocal tract constriction, than the vowels due to their tighter tongue palate contact (similar to Badin *et al.*, 2002).

The present study aims to replicate a cine-MRI session and determine potential effects in midline tongue contours due to gravity. In static MRI studies, where a single sound is sustained, the weak RF emissions from hydrogen protons are summed over time as the subject holds still. Cine-MRI requires multiple repetitions of a single utterance while the emissions of each time frame are summed in an ensemble fashion. The technique was developed and refined in a smaller study that examined the effect of gravitational orientation on tongue position during words and pauses for a small number of subjects (Stone *et al.*, 2002b). That study found backward rotation was the most salient feature in upright/supine comparisons for two subjects, and that the center of this rotation differed by phoneme, with /i/ being most anterior. There was no acoustic effect of the backward tongue rotation.

In the present study, tongue shapes are expected to differ locally and globally due to gravitational orientation and also to differ among subjects. To determine whether these expectations are true, ultrasound images presenting a variety of lingual tongue contours during consonants and vowels were taken from words spoken in upright and supine condition. Ultrasound allows tongue data to be collected in both supine and upright conditions, and provides good imaging of the pharyngeal region, which is of major interest in this study. In addition, midsagittal ultrasound contours of the tongue parallel those observed in midsagittal MRI, despite the loss of tissue visibility in the extreme anterior and posterior tongue (Stone *et al.*, 2003). Finally, the data were collected from continuous speech, which is more realistic than sustained utterances, and seems from previous work to have a lesser response to gravity. Ultrasound images are thus ideal for exploring the effects of gravity on tongue surface shape.

The goals of the present study were to measure the effects of gravity on tongue position in speech behaviors, and to determine whether task and subject effects interact with gravity orientation. Two responses to gravity are possible. First, the tongue might be positioned more posteriorly in supine orientation with little or no compensation to gravity, as modeled by Shiller *et al.* (1999). This would occur if the posterior displacement of the tongue did not significantly affect speech acoustics or airway patency. Second, the tongue might be positioned more anteriorly (this includes

TABLE I. Subject and task listing. Thirteen subjects said the words "bang," "dash," and "golly," except as noted in the table.

			Bang	Dash		Golly
Subject	Age	Sex	ŋ	dæ	ʃ	gɑli
1	16	f	×	×	×	×
2[a]	26	f	He rang	Dack	Highchair	×
3	25	f	He rang	Dack	Highchair	×
4	24	m	×	×	×	×
5	19	m	×	×	×	×
6	27	f	×	×	×	×
7	48	m		×	×	×
8	22	f	×			×
9[b]	24	m	×	×	×	×
10	29	f	×	×	×	×
11	16	m	Anger		Highchair	×
12	24	m	×	×	×	×
13	16	m	×	×	×	×

[a]Native speaker of Brazilian Portuguese.
[b]Native speaker of Russian.

maintaining the original posture) to protect the airway (Tiede *et al.*, 2000; Kitamura *et al.*, 2005). Strategies may vary with subject and phoneme.

II. METHOD

A. Subjects

Thirteen subjects participated in the study. They included seven males and six females, with ages ranging from 16 to 48 years old (mean age 24). All but two were native speakers of U.S. English; those two were native speakers of Portuguese and Russian and were fluent in English (see Table I).

B. Speech materials

The target speech tasks were designed to include a wide range of oral positions and to mimic tasks used in an MRI data session. Therefore the tasks were short words that could be repeated at one-second intervals and included the consonants /ŋ/, /d/, /ʃ/, /g/, /l/, and the vowels /i/, /ɑ/, /æ/. The words were "bang," "golly," and "dash." Due to slightly different protocols, the /æ/ and /ʃ/ were taken from "dack" (/dæk/ and "high chair" for some subjects. The /æ/ was always taken from the nonnasal context (dack or dash). Each word was repeated seven times. To eliminate possible "list-effects" on the first and last repetitions, the five medial repetitions were measured. Not all subjects said the same words and Table I lists the words for each subject. Of the 13 subjects, three words were missing due to errors in data collection. Thus the corpus consisted of 39 words, 37 vowels, and 61 consonants.

C. Data collection

An ultrasound machine (Acoustic Imaging, Inc., Phoenix, AZ, Model No. A15200S) was used to collect midsagittal images of the tongue from each subject during the speech tasks in two gravity conditions: upright (UP) and supine (SUP). A 2.0–4.0 MHz multifrequency convex-curved linear array transducer that produced 30 wedge-shaped scans per second was used. In order to create similar upright and supine recording conditions, the ultrasound transducer was fitted to a cervical collar in a midsagittal orientation. The collar was positioned around the subject's neck, in the upright position, so that the transducer was under the chin, and then adjusted until the shadows of the hyoid and jaw were as close as possible to being equidistant from the edges of the scan. This is meant to normalize transducer position across subjects. The transducer was immobile for both supine and upright data collection (Fig. 1, col. 1). This inhibited jaw motion severely, but allowed consistent transducer positioning across conditions. A microphone, attached to the dental chair, recorded speech.

The subjects were instructed to repeat the words seven times in each gravity condition to the beat of a metronome. The metronome was set at one beat per second to mimic an MRI data collection, where repetition time is one second. The entire recording procedure lasted 20–30 minutes. Both the ultrasound image sequences and the synchronized acoustic signal were digitized using the ADVC1394 (Canopus Inc.) I/O card. The ultrasound image sequences were saved as a series of jpeg files at 29.97 frames per second. The acoustic data were saved as a monaural wave file at a sampling rate of 44.1 kHz. All data were backed-up by simultaneous recording on videotape.

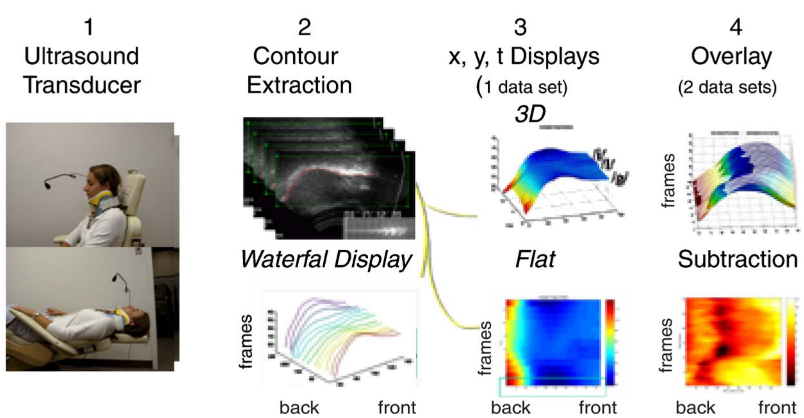

FIG. 1. (Color online) Data collection and analysis. (1) Transducer in neck brace; (2) tongue contour extraction and time-motion (waterfall) display; (3) time-motion displays as 3D surface and flat sequence; and (4) overlay of two contour sequences and their difference.

D. Palate alignment

The first four subjects were collected in upright position first. When put into supine position it was possible that some had moved their heads backwards. All subsequent subjects were measured in supine position first and instructed to keep the head immobile when moving into upright position. When supine was collected first the neck brace appeared to adequately inhibit head movement in the speakers. Validation was done on the supine-first method by measuring distances and angles between points on the head and neck brace on a single subject. These measurements indicated less than 3 mm error between the gravity conditions. However, it was believed that some subjects may have moved their head in the superior-inferior direction. Therefore, a more valid method was developed to align palates, described in detail in Epstein and Stone (2005). In brief, the supine palatal contour was overlaid and aligned to the upright contour using rigid body transformation (x-, y-translation and in-plane rotation). The palate transformation parameters for each subject were saved and used to transform the supine tongue contours into the upright orientation.

Palate shape is not normally visible in ultrasound. However, during a swallow the palate is visible because the sound reaches the palatine bone when the tongue touches the palate. When the ultrasound wave reaches the palatine bone it reflects back to the transducer. The reflected palate appears as a white line, which can be traced like the tongue. Palatal contours for each subject were extracted from spontaneous dry swallows in upright and supine condition, and the pair of palate contours was rigidly aligned. In addition, by measuring the palate during several frames of the swallow, the velum was captured in multiple positions as it moved upward into a maximally closed position. Overlaying the velar contours reveals the junction of the velum and the hard palate. The palatovelar junction is a tissue point that was used to normalize the location of tongue constrictions across subjects (for details see Epstein and Stone, 2005).

The quality of the UP-SUP palate alignment was determined for each subject by two calculations: maximum absolute error and root-mean-squared (RMS) error. Error measurements were made by comparing the nearest-neighbor points in the x, y directions. The RMS errors provide a global estimate of error, and the maximal absolute error indicates how the palatal differences compare to the ultrasound measurement error (max=0.7 mm, Unser and Stone, 1992). RMS errors were used rather than average distances, because they give increased weight to larger distances making differences more noticeable. After alignment seven subjects had maximum errors below measurement error (0.7 mm), for five others it was less than 1.3 mm. For the RMS errors, 10 subjects were below measurement error, two were at 0.8 mm. The remaining subject had a maximum error of 2.9 mm and an RMS error of 1.5 mm. This final subject (No. 11) had two entirely different palate shapes; one may have been the tongue. Since one swallow was collected per subject, a better tracing could not be obtained. However, because parts of the palatal contours were well aligned visually, his data were included. His results were among the larger differences, but not the most extreme.

E. Tongue data preprocessing and analysis

Frame numbers (i.e., temporal location) for each word, phoneme, and repetition were extracted from the speech wave. For the consonants and vowels, tongue surface contours were measured and analyzed in several steps using the Maryland Tongue Analysis Package (MTAP) (www.speech.umaryland.edu). First, tongue contours were extracted from the ultrasound images using EdgeTrak, which semiautomatically extracts and tracks tongue contours (and palate contours) for each subject and word (Li *et al.*, 2005a) (Fig. 1, col. 2). Second, the supine and upright palate contours were registered by rigidly transforming the supine palates to the upright orientation. These transformation parameters were then applied to all the supine tongue contours. All subsequent analyses were done using palatally aligned supine data. Third, before averaging the five repetitions of each word, temporal alignment was done using features in the ultrasound contours.

Spatial alignment (rigid body registration) also was done to reduce random variation between replicates introduced by subject imprecision, without modifying tongue shape (Li *et al.*, 2005b). It was assumed that inter-repetition variability was due to unintended speaker error (noise), and that the subject intended to produce the same token in each repetition. Inter-repetition variability varies by speaker and tongue location. The current wisdom is that variability is minimal at constriction locations and (Perkell and Nelson, 1982), because constrictions are acoustically important; acoustically unimportant regions have greater variability. Previous ultrasound studies have found average-RMS differences for replicate curves (data set: /a, i, u, e, o/) to be 1.36 mm or less (Morrish *et al.*, 1985). However, differences at the location of maximum variability (data set: /a, i, u/) were 1.4 cm or less (Stone *et al.*, 1983). Large maximum and small average variability are consistent with fairly small overall variability in multiple replicates. Our unpublished examinations of list-type replications support this as well. Rigid body alignment, based on least-squares minimization, would optimally reduce larger variations while minimally affecting smaller ones.

The fourth preprocessing step averaged the five spatiotemporally aligned repetitions for each gravity condition. The average dataset was considered to be a better exemplar of the word than any individual token. Moreover, MRI data are typically summed from many repetitions, and thus contain smoothing similar to the ultrasound average. All subsequent analyses used the average contours except range of motion.

In order to visualize and statistically analyze the data sets, they had to be of the same size. Therefore, for each subject, upright-supine word pairs were cut or extended to the same length, with the software Surfaces, which employs thin-plate splines (kriging) to extend a contour (Parthasarathy *et al.*, 2005). The resulting contour sequence for each word is represented as a 3D object, that is, a spatiotemporal

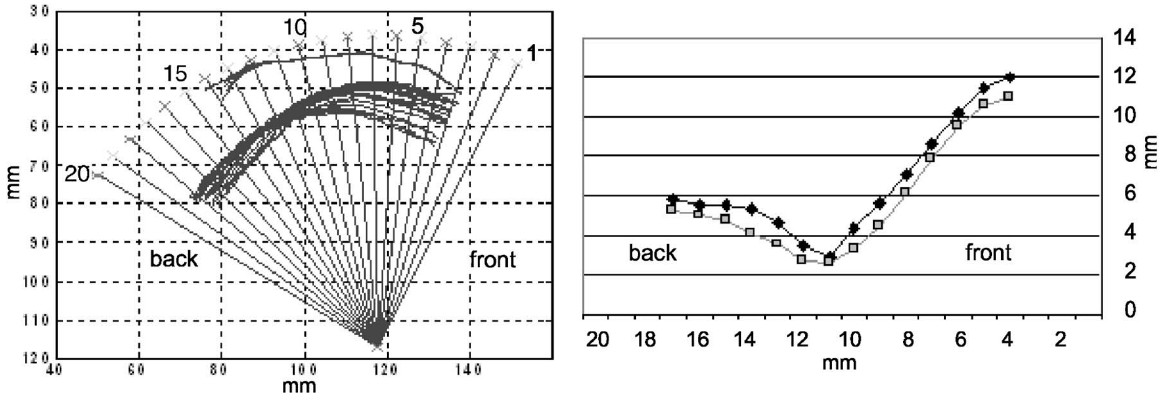

FIG. 2. (a) Tongue contours for one repetition of "dash" (upright), and the subject's palate trace (top), are intersected by up to twenty radii. The palate trace shows multiple velar positions. (b) Motion functions show the ROM at each radius for five repetitions dash in upright (black diamond) and supine (gray square) condition. Radius 1 is anterior; measured values are in millimeters.

"surface" consisting of x, y, and t values (see Fig. 1, col. 3). Height values are color-coded. From the upright and supine surfaces, the target phoneme contours were extracted and overlaid. UP-SUP contour differences were visualized by subtraction (Fig. 1, col. 4) and compared globally using RMS error. The tongue was then divided into six equal length regions and the average UP-SUP distance calculated for each. In ultrasound images the tongue tip and root are imaged less clearly, causing measurement error. This error is compounded by the use of kriging (spline extension) to equalize contour lengths for the comparisons. As the ends (regions 1 and 6) may have visibility and extrapolation errors, regions 2 and 5 were used as the extreme posterior and anterior tongue regions respectively. As transducer angle cannot be perfectly controlled in the neck brace, some subjects' data were collected at a more posterior angle than others. To compare constriction location across subjects the location of the velar junction, a tissue point on the hard palate, was used as a reference point across subjects.

For each subject, range of motion (ROM) was calculated in both upright and supine condition for five repetitions of each word. Differences in ROM between the two gravity conditions are of interest as they reflect tongue mobility effects. The need to maintain an open airway in supine position could reduce ROM. It is also possible that gravitational pull would increase backward movement while forward positioning was maintained for acoustic reasons, resulting in larger supine ROM. ROM was examined at multiple locations on the tongue surface by creating a polar grid template for each subject. Each template was based on the subject's entire dataset, that is, all the tongue contours for 3 (or 4) words $\times$ 5 repetitions $\times$ 2 gravity conditions. The tongue contours were overlaid and a vertex placed several mm below and midway between the two most extreme endpoints. Originating from this vertex, 20 equally spaced radii were positioned to intersect with the tongue surface; the first and last radii were aligned with the most extreme endpoints in the data. Figure 2(a) shows tongue contours for one repetition of dash drawn on such a template. To calculate ROM at each radius, the minimally displaced contour was subtracted from the maximally displaced contour (max-min difference). The max-min differences were averaged for each word, and the averages used to compare the effects of gravity, subject and word on the extent of tongue surface motion.

After all UP-SUP comparisons were made, the supine contours were "corrected " to match their upright counterparts using rigid body and rigid-plus-affine (hereafter, affine) alignment methods. The goal was to determine whether global transformations were sufficient to correct the effects of gravity on the tongue, and whether additional affine correction (homogeneous scale, stretch and shear) improved the match.

Acoustic analysis was performed on the upright and supine speech waves for nine subjects, and five repetitions, of the words dash (or dack) and golly using PRAAT. The first two formants were extracted from the midpoints of each vowel by visual inspection of the spectrogram, with LPC tracking for assistance. In addition, formants were extracted at vowel onsets or offsets to capture consonant information for /ŋ/, /d/, /g/, and at the onset and midpoint of /l/. For the acoustic analysis five repetitions of the sustained vowels /u/, /a/, and /i/ were also available. The formants were extracted at the midpoint of the vowel for each repetition. These sounds did not have visually observable formant variability and were used to determine whether sustained sounds were different from continuous speech. Because the ultrasound machine is fairly noisy, the acoustic wave was as well, and neither formant bandwidths nor fricative noise were measured.

F. Comparative statistics

Acoustic data were available for 198 upright-supine comparisons (nine subjects $\times$ 11 phonemes $\times$ 2 formants). Thirty of these comparisons had missing data, leaving 168 comparisons. Confidence intervals (at 95% confidence) were calculated for F1 and F2 in each gravity condition, and significant differences determined based on overlap.

The ultrasound contours were examined in three ways. (1) Paired t-tests compared the ROM of each word in the upright and supine condition. (2) Two one-way ANOVA's for repeated measures examined the UP-SUP differences in global contour (RMS error), and the pharyngeal zone, due to subject and phoneme. If a significant difference was found, a

Tukey HSD test was used to extract the significant comparisons. (3) Paired t-tests were used to compare the upright contours to the corrected supine contours on a point-to-point basis.

III. RESULTS

This study examined several global and local features of midsagittal tongue contours that might be affected by gravity. (1) Vowel formants were used to compare acoustic consequences of gravity on target consonants and vowels. (2) Range of motion was calculated to determine whether supine motions were more limited than upright. (3) Global contour differences were compared for each phoneme. (4) Pharyngeal contour differences were compared for each phoneme. (5) The effect of gravity on the location of the consonant constriction was measured. (6) The supine contours were corrected to upright using rigid and affine transformations. These global and local effects were considered with respect to subjects and tasks.

A. Acoustic effects due to gravity

The confidence intervals of the 168 comparisons showed 13 significant differences between upright and supine at the $p=0.05$ level. This number is consistent with chance, and indicates that the physiological effects of gravity were negligible acoustically. Interestingly, 9 of the 13 significant differences occurred in sustained vowels.

B. Range of motion in upright and supine condition

ROM was calculated for each of the 39 words in the upright and supine condition to examine tongue motion. The other measurements were applied to single frames extracted from the image-sequence to represent the maximum position of each phoneme.

The words differed in ROM based on phonemic content. Golly had the largest average ROM in both gravity conditions (mean=13.3 mm-UP, 13.6 mm-SUP) due to the large difference in tongue position for the high back /g/, low back /ɑ/ and high front /i/. A smaller ROM was seen for dash (8.3 mm, 8.5 mm), which had no back phonemes. The smallest was seen for bang (6.5 mm, 6.1 mm), with only two lingual phonemes. Figure 2(a) displays a set of overlaid tongue contours for one repetition of upright dash, and a palate contour showing several velar positions. The lowest tip position (radius 4) occurs in /ae/, the highest one in /d/, with the /ʃ/ in between. ROM is maximal anteriorly and minimal at a pivot region toward the back. The ROM calculated at each radius can be plotted with respect to the radius point (index) to generate a "motion function." Motion functions for the word dash are shown in Fig. 2(b) for the upright tongue contours presented previously in Fig. 2(a) (black diamonds) and for the corresponding supine contours (white squares). Motion functions were generated for each word in the UP-SUP conditions and indicated, across gravity conditions, a pattern similar to that in Fig. 2(b).

The motion functions shown in Fig. 2(b) reflect the large anterior ROM and a rotational movement during the word. At radius 11 is the motion minimum, which depicts the fulcrum of the tongue rotation. The difference in average (and SD) ROM for all the radii is very small [upright 9.64 mm (3.1), supine 9.47 mm (3.3)]. For each word, the average ROM values were compared using a matched pairs t-test and the differences were non-significant ($p=0.05$). Of the 39 subject/word comparisons, 27 had average ROMs that were larger in upright condition; 21 of these differences were less than 1 mm. The remaining 12 words had ROMs that were larger in supine condition; 8 of these were less than 1 mm.

C. Upright-supine differences in tongue contour

Subject and phoneme differences were compared globally by calculating RMS difference between each pair of UP-SUP contours. The distances ranged from 0.7 to 6.9 mm (Table II). A one-way ANOVA for repeated measures determined that subject had a significant effect on the RMS difference ($F=10.724, p\leqslant 0.0001$), but phoneme did not ($F=0.488, p\leqslant 0.841$). A Tukey HSD test was used to determine which subjects grouped together. In Table II, the bars indicate groups that are not significantly different. The single horizontal bar at the bottom indicates no phoneme differences. The vertical bars classify the majority of subjects into a single group with small UP-SUP RMS differences (mean $\leqslant 2.5$ mm), and other groups with larger UP-SUP differences. Two subjects (1,3) bad significantly greater UP-SUP differences than almost all the other subjects. These two subjects were among the four whose data were collected upright first. Repeat measurements and visual inspection of the raw images indicated a backwardly displaced tongue relative to the palate in supine condition. To artifactually create such a difference in tongue position relative to the palate reference, the transducer or head would need to rotate sufficiently that the angle of recording was vastly different. A largely different vertex location and radii angle could create such an error. Such a movement was not observed during data collection, however, it cannot be ruled out.

Subject and phoneme differences also were compared in the pharyngeal zone since gravity might have its greatest effect on this region. The pharyngeal zones varied in location by subject, and were determined by the velar junction to be 2, 3, or 4. The pharyngeal differences ranged from 0.2 to 8.9 mm. About a quarter of the differences (27 of 97) were less than 1 mm (Table III). A one-way ANOVA for repeated measures determined that subject had a significant effect on the pharyngeal contour difference ($F=12.728, p\leqslant 0.0001$), and phoneme did not ($F=0.600, p\leqslant 0.755$). The result of the Tukey HSD test is indicated by the black bars on the side and bottom of Table III. The single bar at the bottom indicates that phoneme differences were not significant. The bars on the right indicate that ten subjects were not significantly different from each other; they had from 1 to 3 mm differences in pharyngeal tongue position. Again subjects 1 and 3 had the largest posterior differences, over 6 mm.

Although the predominant response to the supine gravity condition was to move the pharyngeal tongue posteriorly, subject specific patterns showed that this was not universal, and that the size of the response did not reflect the direction of motion. The largest group had small, nonsignificant, pha-

TABLE II. Upright-supine tongue contour differences calculated using RMS error, in ascending order of subject mean. Means that are not within the same bar are significantly different. Values in millimeters.

Subject	ng	d	sh	g	l	a	ae	i	Mean (SD)
2	1.5	0.7	1.6	1.8	1.8	2	0.7	1.8	1.5 (0.5)
8	0.9			2.6	1.4	1.7		1.6	1.6 (0.6)
10	0.5	1.2	0.9	2.8	2.3	2.2	1.3	3.6	1.9 (1.1)
7		1.8	2.7	1.7	0.9	2.3	1.9	2	1.9 (0.6)
6	0.4	2.5	1.5	0.7	1.9	2.4	3.5	2.7	2.0 (1.0)
13	1	3.1	0.8	1.9	2.1	2.9	1.4	4.1	2.2 (1.1)
9	2.9	1.2	3.9	1.4	1.8	1.2	2.2	2.4	2.1 (0.9)
12	3.7	1.5	2.3	2.9	1.4	3.3	3.2	1.6	2.5 (0.9)
5	1	1.6	2.2	1.3	2.3	3.2	2.4	6.2	2.5 (1.6)
4	6.8	5.7	5.1	1.2	2.1	2.4	4.1	4	3.9 (1.9)
11	5.9		3.1	4.1	4.6	3.6		4.5	4.3 (1.0)
1	4.2	6	5.4	5.3	6.3	4.2	4.5	4.4	5.0 (0.8)
3	6.9	5	4.7	4.7	4.3	5.3	3.8	6.5	5.2 (1.1)
Mean	2.98	2.75	2.85	2.49	2.55	2.8	2.78	3.1	
SD	2.48	1.93	1.6	1.43	1.55	1.1	1.1	1.5	

ryngeal differences (subjects 2, 5, 6, 7, 8, 9, 10, 11, 12, and 13). Five of these moved backward (2, 5, 7, 9, and 11), one moved forward (7), and four varied by task (8, 10, 12, and 13). A second group (1 and 3) moved backward considerably in supine position (mean=6 mm). A single subject (4), had a moderate mean difference (4 mm), and moved forward in supine condition. Figure 3 displays typical contours in upright (black) and supine (gray) conditions. Backward and forward motion included translation (A,C) and rotation (B). Differences were largest posteriorly.

D. The effect of gravity on the consonant constriction

It is well documented that the effects of continuous speech, such as coarticulation, rate, multiple repetitions, have

TABLE III. Upright-supine contour differences in pharyngeal region of tongue, in ascending order of subject mean. Means that are not within the same bar are significantly different. Values in millimeters.

Subject	ng	d	sh	g	l	a	ae	i	Mean (SD)
6	0.2	1.6	1.6	0.4	0.2	0.8	0.7	2.4	0.9 (0.80)
8	0.9			1.1	1.3		0.4	2.6	1.3 (0.82)
7		2	2.3	0.7	0.2	1.9	1	0.8	1.3 (0.79)
2	0.4	0.4	1.6	1.4		0.6	2.5	2.5	1.3 (0.92)
13	0.2	1.4	0.3	1.5	2.9	1.7	3	0.7	1.5 (1.07)
10	0.6	0.8	0.8	2.3	2.4	1.1	2.2	1.8	1.5 (0.75)
9	0.7	0.7	6.9	0.2	1.5	1.3	0.3	3	1.8 (2.24)
12	3.3	0.7	1	1.1	2.2	3.7	0.8	2.8	1.9 (1.21)
5	0.4	0.6	3	3	2.6	3.1	3.4	3.7	2.5 (1.26)
11	1.9		1.7	3.1	4		2.7	5.7	2.8 (1.72)
4	3.7	7.4	6.7	1.5	2	5.6	1.7	5.8	4.3 (2.38)
1	3.9	8.9	7.7	5.4	7.8	6.4	5	5.9	6.4 (1.66)
3	9.3	5.2	5	6.4	6.1	3.8	6.5	8.8	6.4 (1.87)
Mean	2.13	2.7	3.22	2.16	2.77	2.73	2.32	3.58	2.61
SD	2.66	3.02	2.64	1.89	2.26	1.97	1.86	2.34	1.89

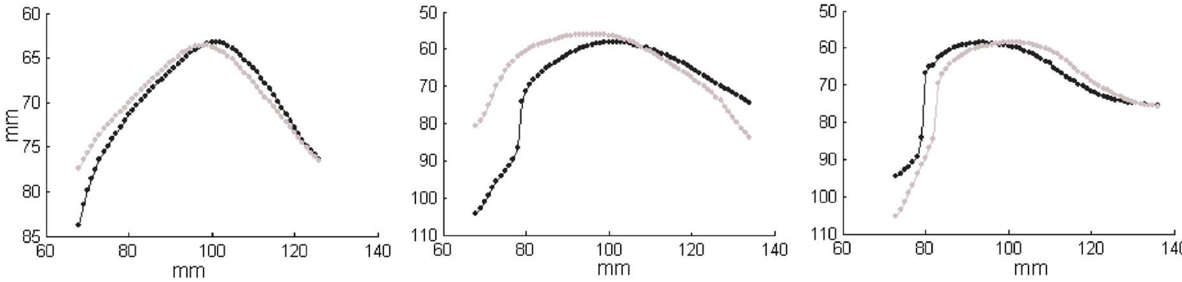

FIG. 3. (Color online) A comparison of upright (black) and supine (gray) contours for three subjects saying /g/ demonstrates backward translation, backward rotation and forward translation. (A) Subject 5 /æ/; (B) Subject 1 /ʃ/; and (C) Subject 8 /g/.

their smallest effect on vocal tract area at the location of the consonant constriction (cf. Perkell *et al.*, 1992). In the present study it was hypothesized that the UP-SUP difference would be minimal at the location of the consonant constriction to preserve this acoustically important feature. The velar junction and the minimal UP-SUP difference for each subject and phoneme were studied to determine whether the minimum difference was at the constriction location. Since the tongue tip is not visible in ultrasound images, the anterior constriction locations may not be well depicted.

Figure 4 depicts the number of times that the minimal upright/supine distance occurred at a particular zone for each consonant. The sounds are ordered from top-to-bottom in the order /l, d, ʃ, s, ŋ/. Some phonemes had uniform UP-SUP differences throughout the tongue, and often these differences were quite small. Distances of less than or equal to 1.0 mm throughout the tongue (zones 2–5) were considered to have no-change and were not plotted. The number of times each phoneme had no-change (out of 13 per phoneme) was 7 for /ŋ/, 6 for /d/, 6 for /ʃ/, 4 for /g/ and 2 for /l/. Thus, for some phonemes, such as /ŋ/, there were very few token differences to consider. For the remaining 36 comparisons, consonants with more anterior constriction locations were found to have minima at more anterior segments. Although the data were generally consistent with preserving size of the constriction, the pharyngeal zone differed across subjects and there were not enough data to do a statistical analysis, or provide unambiguous results.

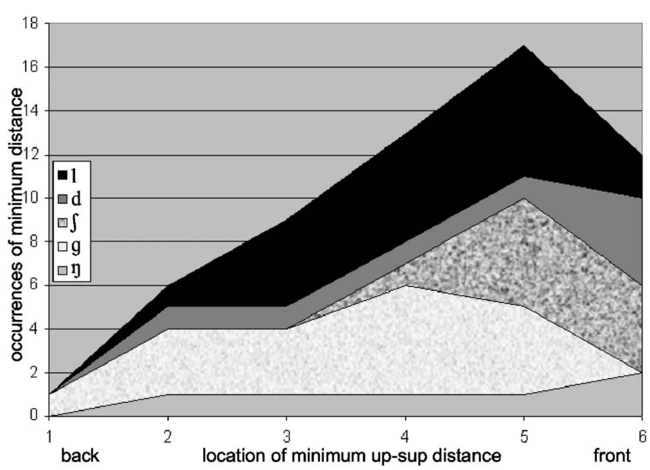

FIG. 4. Number of occurrences of minimal up-sup distance at each zone for each phoneme. Phonemes with anterior constrictions have more frequent occurrences of anterior minima. Order of phonemes from top to bottom is: /l, d, ʃ, s, and ŋ/.

E. Correcting the supine data for the effects of gravity

Once the UP-SUP differences were documented for each subject and phoneme, the next step was to correct the supine contours into upright ones to see how complex a transformation was needed. Correction of the supine contours into the upright ones was performed using two methods: rigid alignment (i.e., rotation and translation), and affine alignment (i.e., rigid plus homogeneous stretch, shear and scale). A one-way ANOVA found that there was no subject effect on the RMS distances for either the rigid or affine correction. That is, regardless of whether the correction was affine or rigid, the corrected supine contours were very similar to the original upright contours for all subjects. In only one subject did the RMS error show contour differences of more than 0.4 mm (see Table IV). A matched-pair t-test compared the effects of phoneme, within each subject, on transformation method. Although the affine transformation significantly improved the rigid alignment ($p=0.05$) for 5 subjects (3, 4, 6, 10, and 12), these improvements were on the order of 0.2 mm or less, and are probably not functionally meaningful.

TABLE IV. RMS error for rigid and affine transformation methods of aligning upright and supine contours (mm).

Subject	Rigid	Affine
1	0.8	0.7
2	0.5	0.5
3	0.6	0.4
4	0.7	0.5
5	0.4	0.4
6	0.6	0.4
7	0.5	0.6
8	0.7	0.5
9	0.7	0.7
10	0.4	0.3
11	0.8	0.7
12	1.1	0.7
13	0.7	0.8
Mean	0.66	0.54
SD	0.18	0.16

IV. DISCUSSION

Three factors appeared to contribute to the supine tongue displacements: preservation of speech quality, airway protection, and subject preference.

The first finding was that acoustic spectra (formant frequencies) were preserved despite the varying tongue responses to gravity. The number of significant differences was due to chance and occurred mostly on steady state sounds. This is consistent with the findings of others, who found tongue position was more sensitive to gravity in steady sounds than continuous speech (Tiede *et al.*, 2000; Engwall, 2006).

Airway protection appeared to be a consideration in tongue displacement as well. Only two subjects had large backward displacements in supine condition. The small backward translation in the other 11 is consistent with airway protection being a factor in tongue positioning. The fairly large number of subjects reveals this as a consistent behavior despite between-subject differences.

The presented data were midsagittal. It was not possible to position the transducer in the neck brace to collect comparable coronal or parasagittal slices across subjects. Our belief is that parasagittal tongue differences would be smaller than midsagittal. Lateral tissue is attached to bony structures or other muscles (e.g., pharyngeal constrictors) whose mass would resist anterior-posterior motion of the lateral tongue. In general we have found that tongue motions are smaller in the para- than the midsagittal planes during speech. In the present data, the consistent formants across gravity conditions is interesting, and suggests that either these fairly small motions are not occurring in acoustically sensitive regions, or that the front back trade-offs create cavity sizes that produce comparable formant patterns, or that some parasagittal changes are indeed compensating for the gravitational changes.

The supine tongue positions indicated the use of all three possible compensatory strategies: backward displacement (minimal or no compensation), maintenance of upright position (sufficient compensation), and forward positioning (strong compensation). Backward displacement in supine position was the most common response, seen in seven of thirteen subjects. It is possible that different strategies would have been observed if the jaw had been free to move. However, Shiller *et al.* (1999) found no anterior-posterior jaw displacement in supine position. A stable jaw position implies that tongue changes are not correlated to jaw changes. The range of backward displacements was consistent with two strategies: little or no opposition to gravity (e.g., Subjects 1 and 3), and "sufficient" opposition needed to maintain some feature, such as, airway opening or phoneme quality (e.g., Subjects 2, 5, 7, 9, and 11). Two additional subjects showed a greater opposition to gravity by positioning the tongue forward in supine position, clearly maintaining an open airway. These nine subjects used different strategies for supine speech (rotation, translation, local displacement), but each applied his/her strategy fairly consistently across phonemes. Four other subjects varied in a non-systematic manner across phonemes. These variable strategies suggest a variety of speaker dependent compensations to the change in gravity, which are motivated by constraints of airway protection, acoustic clarity and user idiosyncracies.

There was mild support for preservation of the constriction location consistent with Tiede *et al.* (2000) and Engwall (2006). However, the missing tongue tip and the large number of upright-supine comparisons in which the entire contour differed by less than 1 mm prevents strong interpretation of this finding.

The corrections used on supine contours suggest control strategies used to adjust the tongue to the changed gravitational orientation. Rigid body transformation implies a muscle synergy in which the tongue moves as a single entity to maintain shape in supine position. This is possible if orthogonal muscles, such as verticalis/GG versus transverse, contract simultaneously to form isometric stiffening. Orthogonal muscle contractions would reduce or prevent local shape changes.

Affine transformation implies a homogeneous, but less rigid response to the gravitational changes, so that the tongue can stretch or shear uniformly. This would occur if the additional GGP activity found in supine position (Sauerland and Mitchell, 1975) changed the shape of the posterior tongue. However, the present data showed that tongue shape is often constant; rigid body transformation accounts for the bulk of the difference between upright and supine tongue position. This suggests that, like the muscles that stabilize the middle ear bones, GGP contraction is usually exactly sufficient to overcome the airway pressures of supine position. In addition GGP's orthogonality to transverse would contribute further to the rigid body aspect of airway preservation. Muscle contraction patterns would be needed to confirm these hypotheses.

V. CONCLUSION

This study confiRMS several findings seen in previous studies including: subject specific strategies for tongue displacement in supine position; preservation of formant frequencies in continuous speech in supine position; possibly greater preservation of tongue position at the constriction location than at other regions of the vocal tract. The study also showed that ultrasound can determine a speaker's strategy for use in subsequent tongue correction. As MRI speech data sets tend to be short it is possible to collect and analyze the identical data set on MRI and ultrasound.

For some subjects there is enough variability between upright and supine contours that the datasets cannot be used interchangeably, for example, in some vocal tract models or when considering the effects of a supine procedure on speech or muscle tissue. However, in such cases rigid, affine, or local transformations may be useful to transform supine data to an upright position.

ACKNOWLEDGMENT

The research was supported in part by NIH Grant No. R01-DC01758 from the NIDCD.

Badin, P., Bailly, G., Reveret, L., Baciu, M., and Segebarth, C. (**2002**). “Three-dimensional linear articulatory modeling of tongue, lips and face, based on MR1 and video images,” J. Phonetics **30**, 533–553.

Engwall, O. (**2006**). “Assessing MRI measurements: Effects of sustenation, gravitation and coarticulation,” *Speech production: Models, Phonetic Processes and Techniques*, edited by J. Harrington and M. Tabain (Psychology Press, New York), pp. 301–314.

Epstein, M., and Stone, M. (**2002**). “Validation of upright - supine methods,” Working Papers and Technical Reports, No. 2002-01, Vocal Tract Visualization Laboratory, University of Maryland Dental School, Baltimore, MD, pp. 1–7.

Epstein, M., and Stone, M. (**2005**). “The tongue stops here: Ultrasound imaging of the palate,” J. Acoust. Soc. Am. **118**, 2128–2131.

Kitamura, T., Takemoto, H., Honda, K. Shimada, Y., Fujimoto, I., Syakudo, Y., Masaki, S., Kuroda, K., Oku-uchi, N., and Senda, M. (**2005**). “Difference in vocal tract shape between upright and supine postures: Observations by an open-type MRI scanner,” Acoust. Sci. & Tech. **26**, 465–468.

Li, Kambhamettu, and Stone. (**2005a**).

Li, *et al.* (**2005b**).

Morrish, K., Stone, M., Shawker, T., and Sonies, B. C. (**1985**). “Distinguishability of tongue shape during vowel production,” J. Phonetics **13**, 89–204.

Myamoto, K., Ozbek, M. M., Lowe, A. A., and Fleetham, J. A. (**1997**). “Effect of body position on tongue posture in awake patients with obstructive sleep apnea,” Thorax **52**, 255–259.

Narayanan, S., Alwan, A. A., and Haker, K. (**1995**). “An articulatory study of fricative consonants using magnetic resonance imaging,” J. Acoust. Soc. Am. **98**, 1325–1347.

Niimi, S., Kumada, M., and Niitsu, M. (**1994**). “Functions of tongue-related muscles during production of the five Japanese vowels,” Ann. Bull. RILP **28**, 33–39.

Otsuka, R., Ono, T., Ishiwata, Y., and Kuroda, T. (**2000**). “Respiratory related genioglossus electromyographic activity in response to head rotation and changes in body position,” Angle Orthod. **70**, 63–69.

Parthasarathy, V., Stone, M., and Prince, J. L. (**2005**). “Spatiotemporal visualization of the tongue surface using ultrasound and kriging (SURFACES),” Clin. Linguist. Phonetics **19**, 529–544.

Perkell, J., *et al.* (**1992**).

Perkell, J., and Nelson, W. (**1982**). “Articulatory targets and speech motor control: A study of vowel production,” *Speech Motor Control*, edited by S. Grillner, B. Lindblom, J. Lubker, and A. Persson (Pergamon, New York), pp. 187–204.

Perkell, J., and Zandipour, M. (**2002**). “Economy of effort in different speaking conditions. II. Kinematic performance spaces for cyclical and speech movements,” J. Acoust. Soc. Am. **112**, 1642–1651.

Reinicke, C., Obijou, N., and Trankmann, J. (**1998**). “The palatal shape of upper removable appliances. Influence on tongue position in swallowing,” J. Orofac. Orthop. **59**, 2002–2007.

Rose, E. C., Staats, R., Lehner, M., and Jonas, I. E. (**2002**). “Cephalometric analysis in patients with obstructive sleep apnea. Part I: diagnostic value,” J. Orofac. Orthop. **63**, 143–153.

Sauerland, E. K., and Mitchell, P. (**1975**). “Electromyographic activity of intrinsic and extrinsic muscles of the human tongue,” Tex Rep. Biol. Med. **33**, 445–455.

Shiller, D. M., Ostry, D. J., and Gribble, P. L. (**1999**). “Effects of gravitational load on jaw movements in speech,” J. Neurosci. **19** 9073–9080.

Stone, M., Sonies, B., Shawker, T., Weiss, G., and Nadel, L. (**1983**). “Analysis of real-time ultrasound images of tongue configuration using a grid-digitizing system,” J. Phonetics **11**, 207–218.

Stone, M., Davis, E., Douglas, A., Ness Aiver, M., Gullapalli, R., Levine, W., and Lundberg, A. (**2001**). “Modeling the motion of the internal tongue from tagged cine-MRI images,” J. Acoust. Soc. Am. **109**, 2974–2982.

Stone, M., Lundberg, A., and Grouse, U. (**2002a**). “Exploring muscle interactions using tagged cine-MRI,” presented at Speech Motor Control Conference, Williamsburg, VA, March 14.

Stone, M., Sutton, M. Parthasarathy, V., Prince, J., Li, M., and Kambhamettu, C. (**2002b**). “Effects of upright and supine orientation on tongue position during silence,” Joint Meeting of the Acoustical Society of America, Iberoamerican Congress of Acoustics and Mexican Congress on Acoustics. Cancun, Mexico. December 2–6.

Stone, M., Parthasarathy, V., Prince, J. L., and Iskarous, K. (**2003**). “Using ultrasound and tagged Cine-MRI to infer the functional relationship between tongue internal strains and surface shapes,” *Proceedings of the International Conference on Phonetic Sciences*, Barcelona, Spain, August 3–9.

Tiede, M. K., Masaki, S., and Vatikiotis-Bateson, E. (**2000**). “Contrasts in speech articulation observed in sitting and supine conditions,” *Proceedings of the 5th Seminar on Speech Production*. Bavaria, pp. 25–28.

Unser, M., and Stone, M. (**1992**). “Automated Detection of tongue surfaces in sequences of ultrasound images,” J. Acoust. Soc. Am. **91**, 3001-3007.