Tongue surface displacement during bilabial stops

Mario A. Svirsky,[a] Kenneth N. Stevens, Melanie L. Matthies, Joyce Manzella, Joseph S. Perkell, and Reiner Wilhelms-Tricarico
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 21 November 1995; accepted for publication 20 March 1997)

The goals of this study were to characterize tongue surface displacement during production of bilabial stops and to refine current estimates of vocal-tract wall impedance using direct measurements of displacement in the vocal tract during closure. In addition, evidence was obtained to test the competing claims of passive and active enlargement of the vocal tract during voicing. Tongue displacement was measured and tongue compliance was estimated in four subjects during production of /aba/ and /apa/. All subjects showed more tongue displacement during /aba/ than during /apa/, even though peak intraoral pressure is lower for /aba/. In consequence, compliance estimates were much higher for /aba/, ranging from 5.1 to 8.5×10^{-5} cm^3/dyn. Compliance values for /apa/ ranged from 0.8 to 2.3×10^{-5} cm^3/dyn for the tongue body, and 0.52×10^{-5} for the single tongue tip point that was measured. From combined analyses of tongue displacement and intraoral pressure waveforms for one subject, it was concluded that smaller tongue displacements for /p/ than for /b/ may be due to active stiffening of the tongue during /p/, or to intentional relaxation of tongue muscles during /b/ (in conjunction with active tongue displacement during /b/). © *1997 Acoustical Society of America.* [S0001-4966(97)03407-3]

PACS numbers: 43.70.Aj, 43.70.Bk [AL]

INTRODUCTION

What happens inside the vocal tract when an obstruent stop consonant is produced? Initially there is an increase in intraoral pressure, which can result in outward movement of the vocal-tract walls. All stop sounds require this pressure buildup to achieve a burst when they are released, but voiced and unvoiced stops require different articulatory adjustments. In particular, voiced stops require a transglottal pressure difference to sustain glottal vibration (although it may not always be sustained throughout closure). Two main mechanisms have been proposed that may be used by speakers to maintain this pressure difference: it may be done with *active* enlargement of the supraglottal vocal tract (Rothenberg, 1968; Kent and Moll, 1969; McGowan *et al.*, 1995), and/or with a relaxed state of supraglottal muscles (which results in *passive* deformation of the vocal tract; cf. Perkell, 1969). Bell-Berti (1974) proposed that pharyngeal cavity size may be controlled with predominantly passive or active mechanisms by different speakers. Westbury (1983) pointed out that while there was good reason to believe in an active component to facilitate voicing during a stop, the issue of active and passive contributions to vocal-tract enlargement would be difficult to settle on the basis of kinematic data alone.

At the end of a vowel preceding an unvoiced stop, glottal vibration may be stopped quickly by spreading the glottis, perhaps in combination with stiffening the vocal-tract walls. Following release of the consonant, the intraoral pressure decreases and the walls are hypothesized to move inward with a time constant that depends on their physical properties. Vocal-tract wall deformation has traditionally been analyzed using a lumped parameter model whose low-frequency equivalent is shown in Fig. 1. The model parameters are subglottal pressure P_s, transglottal impedance Z_g, and vocal-tract wall resistance R_w and compliance C_w. In this low-frequency approximation, valid below about 20 Hz, the mass of the vocal-tract walls is neglected. The parameters in this model are not fixed. For example, compliance values change greatly depending on whether articulatory muscles adopt a tense or lax state.

Accurate measurements of the parameters listed above are necessary to improve models for how long glottal vibration lasts during closure, and also to improve our models of the release of stop sounds. In particular, accurate estimates of vocal-tract compliance (i.e., the extent to which the vocal tract deforms in response to pressure) are important for designing articulatory models, because assuming rigid walls would lead to major errors in any model. Some existing estimates were obtained with indirect methods (Rothenberg, 1968) or using measurements made on surfaces outside the vocal tract (Ishizaka *et al.*, 1975; Wodicka *et al.*, 1993). The reason for such indirect approaches is that it is difficult to measure movement parameters inside the vocal tract, especially when the mouth and velum are closed. The goal of this study was to refine current estimates of vocal-tract impedance using direct measurements of tongue surface displacement during production of labial stops. In particular, we wanted to obtain evidence to test the competing claims of passive and active enlargement of the vocal tract during voicing. Our approach was made possible by the use of an electromagnetic midsagittal articulometer (EMMA; see Perkell *et al.*, 1992), which allowed us to gather sufficient quan-

[a]Currently at the Department of Otolaryngology—Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, IN 46202-5200.

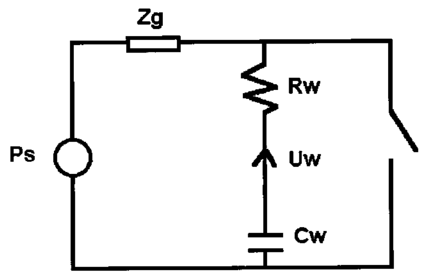

FIG. 1. Lumped parameter circuit model of the supraglottal cavity during a stop consonant. Z_g represents transglottal resistance; P_s is subglottal pressure; R_w and C_w are vocal-tract wall resistance and compliance; and U_w represents airflow. Vocal-tract wall mass is neglected at very low frequencies. Release of the stop is represented by the closure of the switch.

tities of kinematic data from vocal-tract structures during bilabial closure.

I. METHODS

The subjects were four young adult males. Two of them (S1 and S2) were native speakers of American English and the other two (S3 and S4) were native speakers of Ewe, an African language. Both English and Ewe voiced stops are typically produced with glottal vibration, at least at the beginning of the segment, when they are in intervocalic position. The productions from these four subjects were no exception, and the /b/ segments frequently showed glottal vibration throughout the entire closure. Subjects read at least ten repetitions each of three utterances: /apa/ and /aba/ (in which we were primarily interested), and /ama/ (included as a control condition). We expected to find minimal pressure-induced tongue surface displacement in the /ama/ control condition, because during the production of /m/ the velopharyngeal port is open so intraoral pressure remains low, precluding any passive tongue deformation. We used an electromagnetic midsagittal articulometer (EMMA: cf. Perkell *et al.*, 1992) to measure displacement of the tongue dorsum. A small transducer coil (4 mm by 4 mm base, 2.5 mm height) was fixed to the tongue surface (between 4 and 6 cm from the tongue tip) with a biocompatible cement. Alternating magnetic fields generated by a three-transmitter system induce an alternating voltage in the transducer coil. The transduced voltages from the tongue coil, as well as from two other coils placed on the upper incisors and the bridge of the nose for fixed reference points, were low-pass filtered at 100 Hz and digitized at 312.5 samples per s, (sps). Subject S4 had an additional transducer attached to the tongue blade, approximately 1 cm from the tip.

With subject S2, a Glottal Enterprises differential pressure transducer was used to simultaneously measure intraoral pressure (*re*: atmospheric pressure). The transducer was coupled to the oral cavity via a plastic (polyethylene) probe tube approximately 10 cm in length and with an inner diameter of 0.2 cm. The pressure measurements were calibrated using a U-tube manometer and were found to be linear over the pressure range of interest (0–24 cm H_2O).

The acoustic signal was recorded through a directional microphone, low-pass filtered at 4.8 kHz and sampled at 10 kHz. After digitizing the acoustic, movement, and (in S2's case) pressure data, the signals were demultiplexed into separate, time-aligned signal streams. The digitized movement data were then low-pass filtered with 71-tap FIR filters (cutoff frequency about 150 Hz), and converted to x,y coordinates with reference to an x axis which lies in the midsagittal plane, and is parallel to the subject's occlusal plane.

Figure 2 shows a display generated by the software used to extract movement data. The acoustic signal during an /apa/ utterance is shown in window 1, where T_b, T_r, and T_a indicate closure, release, and beginning of the second /a/, respectively. Window 2 shows the expanded acoustic signal around the time of release, T_r. Time-aligned movement data (the positions of a tongue transducer) are shown in window 3: The three traces are the x and y coordinates of the tongue transducer over time, and a measure of transducer misalignment (see Perkell *et al.*, 1992). Window 4 shows an x-y plot of the displacement data (the front of the vocal tract is to the right) and a reference circle of 1 mm radius. The trajectory of the tongue transducer is indicated by the dotted line, with dots representing samples taken at 3.2-ms intervals. The acoustic events were marked interactively for each utterance: the time of closure T_b, defined as the point in the acoustic time waveform just after the last full glottal cycle of the first /a/; the time of release, T_r, defined as the point in the acoustic waveform where there is a sudden increase in high-frequency noise amplitude at the end of the closure period; and T_a, defined as the point in the acoustic waveform just before the first full glottal cycle of the second /a/. A cursor, time-aligned across all the displays, was used to mark the events T_b, T_r, and T_a. The zoomed speech waveform with the cursor located at release time T_r, shown in window 2, was used to locate T_r more precisely. As the x-y data in window 4 show, beginning at the time of closure there is a downward displacement of the tongue, which is presumably due to increased intraoral pressure. At the time of release intraoral pressure is sharply reduced and the tongue moves back up.

Because the x-y data were rotated to make the occlusal plane horizontal, the y axis is the direction perpendicular to the occlusal plane. Mechanical compliance per unit area was calculated as the displacement perpendicular to the occlusal plane (i.e., along the y coordinate) that occurred between T_b and T_r, divided by the estimated peak intraoral pressure. This way of calculating compliance assumes that the tongue displacement perpendicular to the occlusal plane between T_b and T_r is due to increased intraoral pressure. Therefore, systematic movements perpendicular to the occlusal plane and unrelated to pressure would result in over- or underestimates in our compliance measurements. Peak intraoral pressure estimates were obtained from a study of pressure during VCV syllables in males, females, and children by Subtelny *et al.* (1966). We used the average male values from that study: 6.43 cm of H_2O for /p/ and 4.37 cm of H_2O for /b/. While this methodology is appropriate to obtain reasonable estimates of compliance, it may be refined in future studies by measuring intraoral pressure for all subjects. Another possible refinement would involve the simultaneous analysis of the locations of several pellets to separate the local reaction

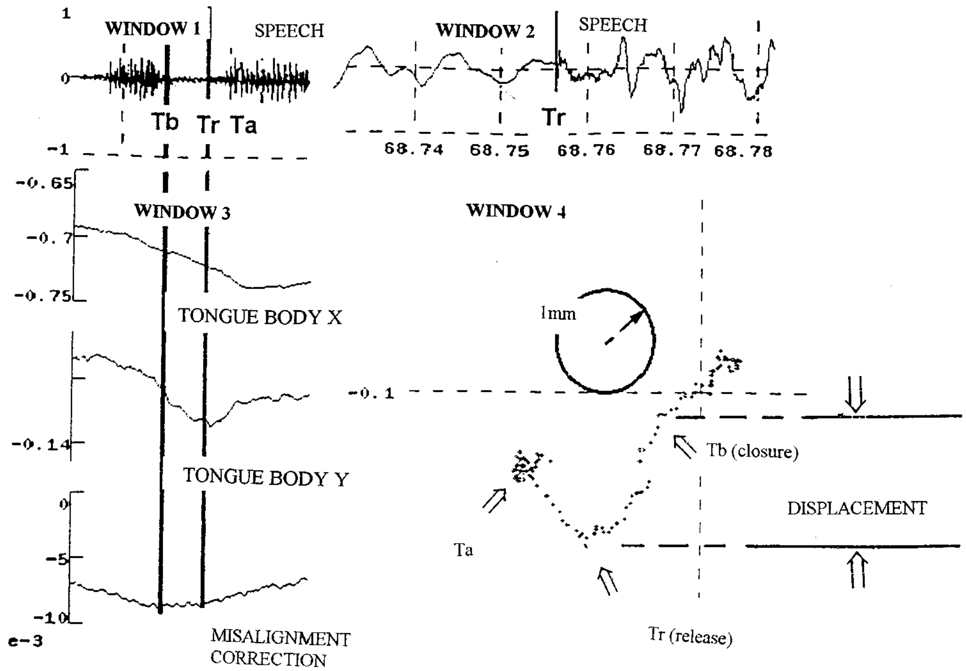

FIG. 2. Display used to extract movement data. Window 1 shows the speech signal for one /aba/ token. Three points are marked: T_b, the beginning of closure, defined as the point in the acoustic time waveform just after the last full glottal cycle of the first /a/; the time of release, T_r, defined as the point in the acoustic waveform where there is a sudden increase in high-frequency noise amplitude at the end of the closure period; and T_a defined as the point in the acoustic waveform just before the first full glottal cycle of the second /a/. Window 2 shows the expanded acoustic signal around T_r. Time axis values are in s. Note the weak periodicity before T_r and the release noise immediately following T_r. Time-aligned movement data are shown in window 3: x and y coordinates of the tongue transducer (in dm), and a measure of transducer misalignment correction (see Perkell *et al.*, 1992). Window 4 shows an x-y plot of the same data, with the occlusal plane being parallel to the x axis and the front of the vocal tract placed to the right. The tongue surface starts a more pronounced downward displacement at closure time T_b, it reverses its movement direction at release time T_r, and remains relatively steady after the second vowel starts at T_a. A 1-mm circle is displayed as a reference, and the two horizontal lines to the right of the x-y data indicate measured displacement for this token.

to pressure on the tongue surface from the global component of tongue deformation.

The pressure and movement signals obtained from subject S2 were used to assess the validity of a passive tongue displacement model. If tongue displacements during closure were mostly passive and driven by intraoral pressure, the displacement trace should lag pressure by a small time interval (determined by the impedance of the vocal-tract walls). On the other hand, if vocal-tract expansion was actively driven, displacement and pressure signals would not necessarily be time synchronized. We assessed the synchrony between pressure and displacement waveforms in two ways. First, we found the time difference corresponding to the maximum of the correlation function between each pair of pressure and displacement signals (holding one signal fixed and applying variable time shifts to the other signal). Second, we determined the relative positions of the most prominent peak in the (smoothed) first derivatives of each pressure and displacement signal. These peaks indicate the instants when each signal is changing at a maximum rate.

We also assessed the pressure-displacement relation by applying two kinds of mathematical models to the pressure-displacement data. First we used a second-order model of the form

$$\text{Pressure} = m\,\frac{d^2y}{dt^2} + b\,\frac{dy}{dt} + ky, \tag{1}$$

where y is vocal-tract displacement perpendicular to the occlusal plane and m, b, and k are constants representing mass, damping, and stiffness per unit area, respectively. Equation (1) can be discretized and rewritten as:

$$\text{Pressure}(n) = a_0 y(n) + a_1 y(n-1) + a_2 y(n-2) \tag{2}$$

for the purpose of fitting discrete-time signals. For each of the ten /aba/ and ten /apa/ tokens, a least-squares procedure was applied to obtain values of a_1, that minimized the difference between predicted and actual displacement. To the extent that tongue displacement is passive and the second-order model in Eq. (2) is accurate, this model should give a good prediction of displacement as a function of the driving pressure. The other model that we used incorporated a linear trend, in addition to the parameters of the first model:

$$\text{Pressure}(n) + a_3 n + a_4 = a_0 y(n) + a_1 y(n-1) + a_2 y(n-2). \tag{3}$$

The left side of the equation now includes, in addition to the measured intraoral pressure, a linear term that represents active expansion of the vocal tract (i.e., downward tongue displacement). If vocal-tract expansion was passive, Eq. (2) would give good predictions of tongue displacement, and Eq. (3) would not yield substantially better fits than Eq. (2). On the other hand, if vocal-tract expansion was at least partly active, Eq. (3) should provide better fits to the data than Eq. (2) and the estimated values of a_3 should be positive, indicating that downward (rather than upward) displacement of

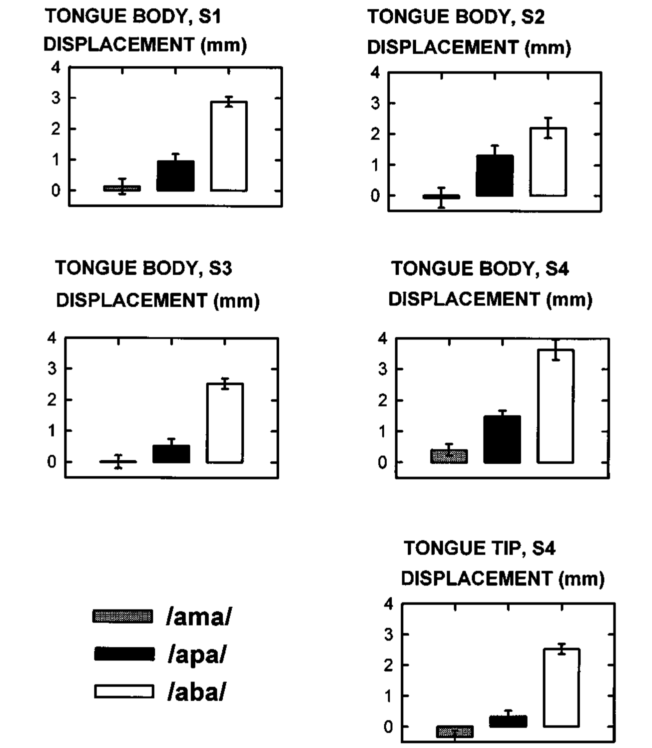

FIG. 3. Mean tongue displacement and standard errors for each category, subject, and (in S4's case) each tongue transducer. Transducers are in the tongue body for all subjects and, in S4's case, there is an additional transducer near the tongue tip.

the vocal tract is taking place and that this displacement is unrelated to intraoral pressure.

II. RESULTS

A. Tongue displacement and estimated compliance

Figure 3 shows plots of mean tongue displacement and standard error values for each token, subject, and (in S4's case) each tongue transducer. A series of t-tests revealed that in all cases, the magnitude of peak displacement for /aba/ (in the direction perpendicular to the occlusal plane) was significantly higher than for /apa/, and significantly higher for /apa/ than for /ama/. Displacements during /ama/ were not significantly different from zero. Table I shows average peak displacements and the compliances associated with them.

TABLE I. Average displacement perpendicular to the occlusal plane for each subject, token type and (for subject S4) transducer location. All numbers are the mean from ten measurements. The two bottom rows list compliance values calculated based on the displacements that appear in the first three rows.

		S1	S2	S3	S4 (tongue body)	S4 (tongue tip)
displacement (mm)	ama	0.13	-0.07	0.01	0.39	-0.32
	apa	0.95	1.30	0.52	1.48	0.33
	aba	2.89	2.20	2.53	3.64	2.53
Compliance (cm^3/dyn) (all numbers are $\times 10^{-5}$)	apa	1.51	2.06	0.82	2.35	0.52
	aba	6.74	5.13	5.90	8.49	5.90

B. Synchrony between intraoral pressure and tongue displacement

Figures 4 and 5 show intraoral pressure and displacement over time for S2, for /apa/ and /aba/ tokens, respectively. Intraoral pressure rises rapidly during /apa/, which is consistent with oral closure and the open glottal configuration necessary to stop voicing. Pressure during /aba/ rises more slowly, with lower values at the time of release. This result is consistent with the need to maintain a transglottal pressure difference that is compatible with continued glottal vibration (the pressure traces do indeed show fluctuations due to glottal vibration almost through the end of most /aba/ tokens). It is interesting to observe that the relatively sharp, fast downward tongue dorsum displacements during /apa/ or /aba/ were generally close to the rise in intraoral pressure. Consistent with this observation, maxima of the cross-correlation functions between pressure and displacement for all /apa/ tokens were obtained, with displacement lagging pressure by a mean of 1.3 ms and a standard error of 1.0 ms. In other words, the cross-correlation function is maximized when pressure is delayed by about 1.3 ms. Results for /aba/ were qualitatively similar, with a mean lag of 5.4 ms and a standard error of 1.9 ms.

We also compared the timing of the peaks of the first derivative for the pressure and displacement traces. This comparison indicates the timing between rapid changes in the two signals. The comparison was made both for the timing between pressure increase and downward tongue displacement, and for the timing between pressure release and the concomitant upward displacement of the tongue. For /apa/, the maximum of the first derivative of the pressure rise preceded the one for displacement by 10.0 ms (standard error: 1.2 ms) and the maximum for the pressure decrease preceded the one for displacement by 8.2 ms (standard error: 0.9 ms). First derivative maxima for displacement traces showed longer lags for /aba/: 18.2 ms for the pressure rise and 12.1 ms for the release (standard errors of 1.4 and 2.7 ms, respectively). Longer lags for /aba/ than for /apa/ are consistent with the greater compliance measured during the voiced stop, since (according to the model depicted in Fig. 1) the displacement time constants are the product of vocal-tract wall resistance and compliance.

C. Modeling tongue displacement

Figure 6 shows vertical tongue displacement for two representative /apa/ tokens, as well as the two model predictions for each token. Connected dots show actual data; dotted lines are the predictions made by the second-order model with no trend, and the solid lines are predictions made by the model that includes a linear trend. Visual examination of the top panel reveals that the model with a linear trend predicts displacement much better than the model with no trend, while both models provide similarly good predictions for the token in the bottom panel. One way to quantify this observation is to calculate the rms of the residuals (i.e., the rms of the difference between model predictions and actual displacement) for each token and each model. Residuals for the token in the top panel were 0.78 mm (model with no trend)

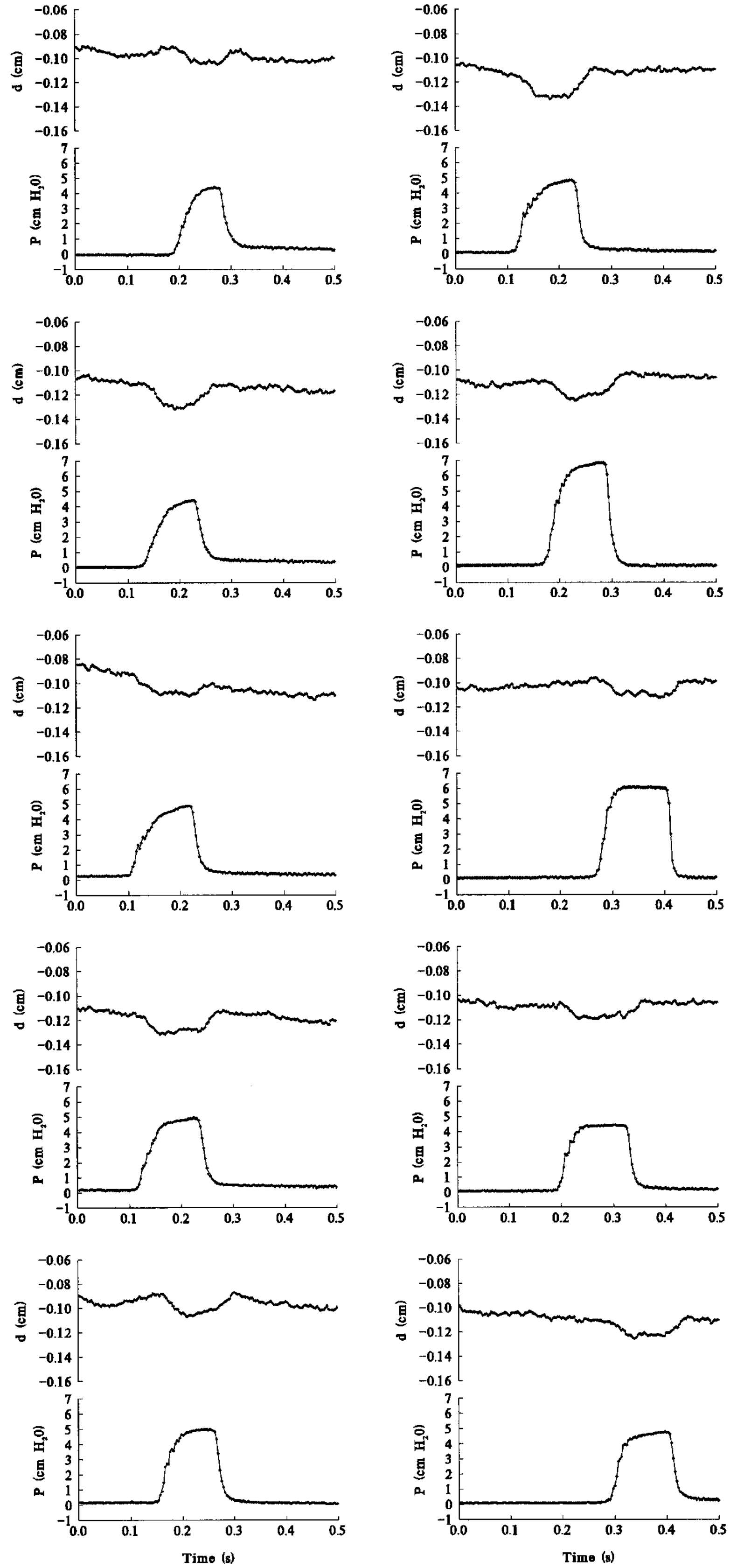

FIG. 4. Tongue displacement (perpendicular to the occlusal plane) and pressure waveforms for /apa/ utterances from subject S2. The top trace in each panel shows tongue displacement in the direction perpendicular to the occlusal plane (in dm), and the bottom trace shows pressure (in cm H_2O). The horizontal axis is time (s), and the plots span 500 ms.

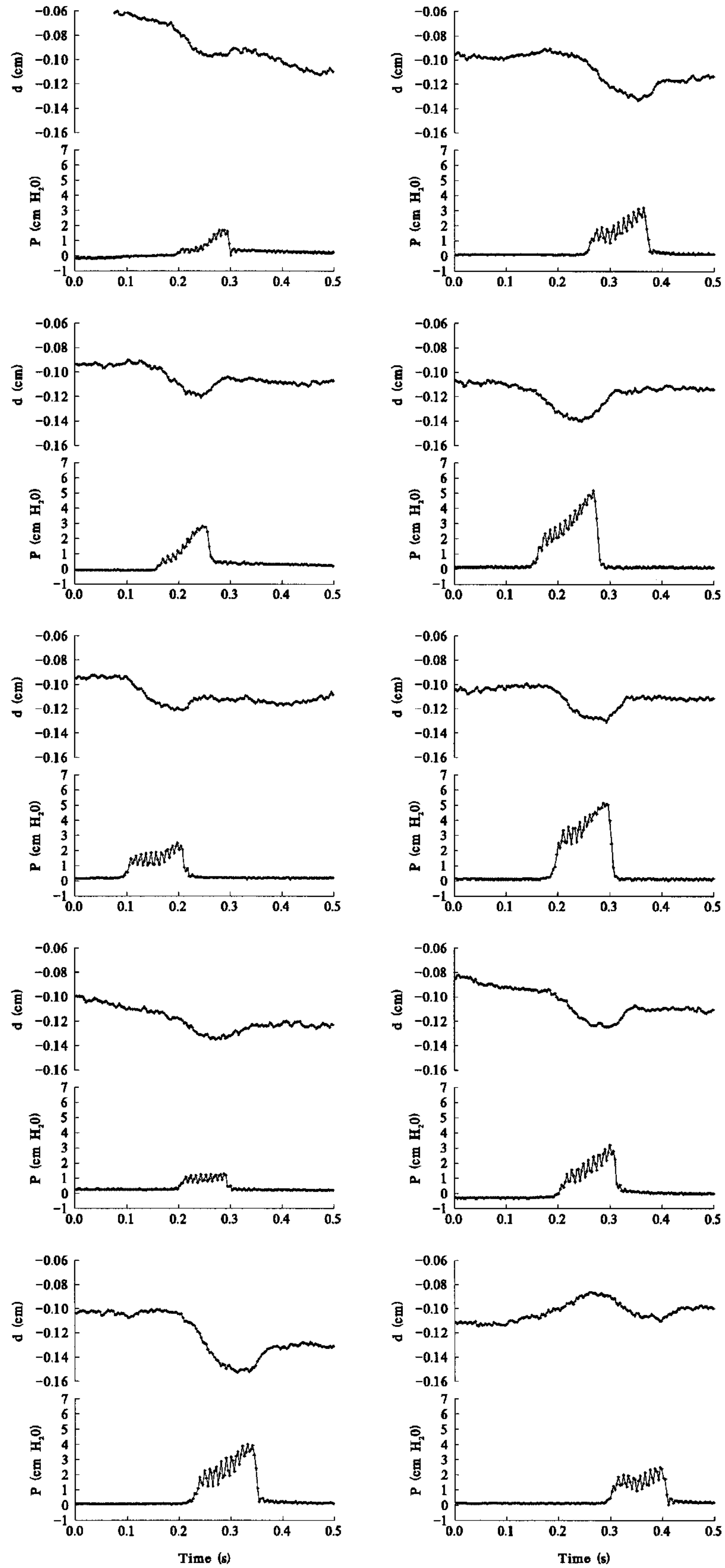

FIG. 5. Tongue displacement and pressure waveforms for /aba/ utterances, subject S2. The top trace in each panel shows tongue displacement in the direction perpendicular to the occlusal plane (in dm), and the bottom trace shows pressure (in cm H_2O). The horizontal axis is time (s), and the plots span 500 ms.

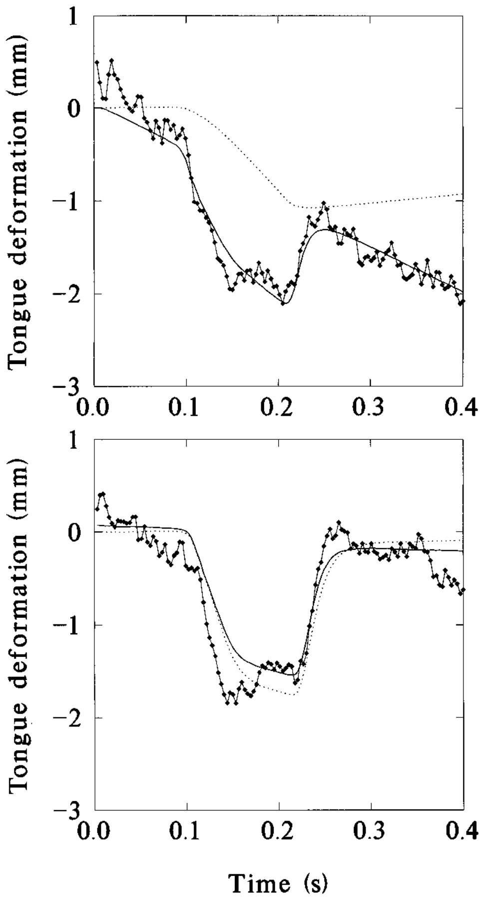

FIG. 6. Black dots show tongue deformation in the direction perpendicular to the occlusal plane (versus time) for two representative /apa/ tokens. The superimposed lines show model predictions for each token. Dotted lines correspond to a second-order model with no trend, which gave a poor fit for the token shown in the top panel (residual of 0.78 mm) and a good fit for the token in the bottom panel (residual was 0.28 mm). Predictions of the model with a linear trend are represented with solid lines. This model yielded good fits for both tokens, with residuals of 0.17 and 0.23 mm for the top and bottom panels, respectively (note how the solid lines seem to match the data relatively well).

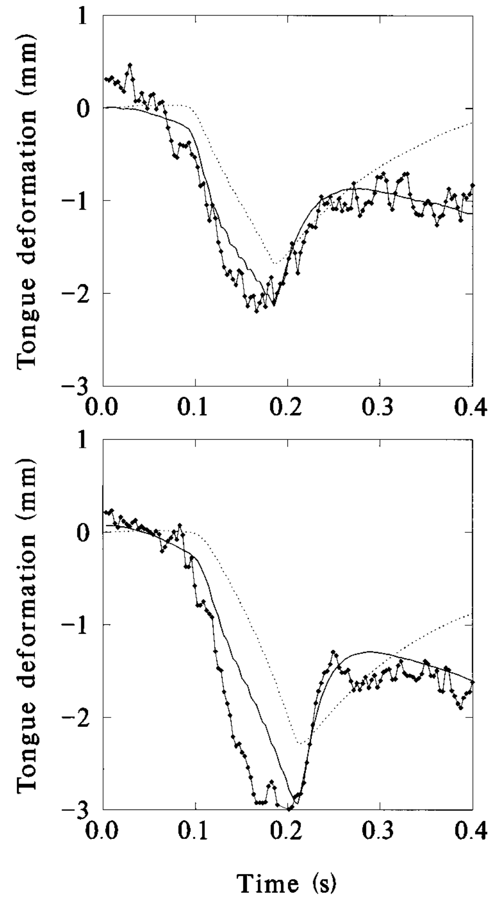

FIG. 7. Black dots show tongue deformation in the direction perpendicular to the occlusal plane for two representative /aba/ tokens, and the superimposed lines show model predictions for both tokens. Residuals were poor for the model with no trend (residuals of 0.52 and 0.75 mm, respectively) but were good for the model that included a linear trend (0.21 and 0.29 mm).

and 0.17 mm (model with linear trend); whereas for the token in the bottom panel residuals were 0.28 mm (model with no trend) and 0.23 mm (model with linear trend). Figure 7 shows two representative /aba/ tokens and model predictions. Residuals were poor for the model with no trend (residuals of 0.52 and 0.75 mm, respectively) but were good for the model that included a linear trend (0.21 and 0.29 mm).

Table II lists the rms residuals obtained for each /aba/ and /apa/ token using each model. To summarize the data, we may classify residuals in three categories: residuals smaller than 0.3 mm indicate a very good fit since measurement noise for the EMMA system we used is in the order of 0.2 mm; residuals greater than 0.5 mm are classified as ''poor'' and residuals between 0.3 and 0.5 mm are classified as ''intermediate.'' Although any classification scheme such as this one has to be somewhat arbitrary, other definitions of good, bad, and intermediate tokens would not change the conclusions we draw from the data.

The first column in Table II shows that the fit was quite poor for /aba/ tokens using the model with no trend: Eight tokens gave a poor fit, two were intermediate, and none were well fit by the model. The average residual (0.93 mm) was quite large. Results using the model with no trend were much better for /apa/ than for /aba/: Five tokens gave a good fit, only two gave a poor fit and three were intermediate. The

TABLE II. rms values of the difference between modeling results and actual tongue displacement (i.e., residuals) for each /aba/ and /apa/ token, for second-order models with and without a superimposed linear displacement trend. Estimates for the trends superimposed on each individual token are also shown.

	Model with no trend		Model with linear trend			
	aba	apa	aba		apa	
	residuals (mm)	residuals (mm)	resid (mm)	trend (mm/s)	resid (mm)	trend (mm/s)
1	2.53	0.41	0.47	11.0	0.29	1.5
2	0.74	0.34	0.27	4.1	0.19	1.3
3	0.53	0.78	0.21	5.4	0.17	4.9
4	0.52	0.28	0.21	2.8	0.23	0.7
5	1.54	0.42	0.67	7.1	0.40	0.5
6	1.05	0.27	0.36	6.1	0.27	0.0
7	0.68	0.69	0.63	0.5	0.23	−1.9
8	0.50	0.21	0.19	2.8	0.16	−0.7
9	0.75	0.27	0.29	4.4	0.14	−1.0
10	0.46	0.22	0.33	2.5	0.17	−0.5
Mean	0.93	0.39	0.36	4.7	0.22	0.48

average residual was 0.39 mm. Residuals for /aba/ were substantially better with the model that incorporated a linear trend than with the model with no trend: Five tokens were well fit by the model, only two were poorly fit, and three were intermediate, with a mean residual of 0.36 mm over all tokens. Addition of a linear trend to the model improved the fit to the /apa/ tokens as well (mean residual was 0.22 mm) but this improvement was less dramatic than for /aba/, mostly because the model without a linear trend was quite good for /apa/ tokens to begin with.

Table II also shows that the linear displacement trends obtained with the model were quite different for voiced and unvoiced tokens: /aba/ tokens required sizeable trends that were always positive, indicating a downward trend in tongue movement (4.8 mm/s on the average), while /apa/ tokens were best fit with rather small trends (mean of 0.48 mm/s) that were sometimes positive and sometimes negative. This result suggests that a significant downward tongue movement, unrelated to intraoral pressure, occurred during production of /aba/ but not during /apa/. However, this was not enough to influence initial compliance estimates significantly. Finally, the second-order coefficients were quite small, suggesting that the second-order model does not provide a substantially better description than the first-order model depicted in Fig. 1.

III. DISCUSSION

A. Values of tongue compliance

All subjects showed more tongue displacement during /aba/ than during /apa/ ($p<0.05$, two-tailed t-tests), even though peak intraoral pressure was lower for /aba/. In consequence, compliance estimates were much higher for /aba/, ranging from 5.1 to 8.5×10^{-5} cm^3/dyn. Compliance values for /apa/ ranged from 0.82 to 2.35×10^{-5} cm^3/dyn for the tongue body and 0.52×10^{-5} for the single tongue tip point that was measured. Tongue displacement values for /ama/ were not significantly different from zero, indicating that the tongue movement observed during the plosives was not simply a coarticulatory effect of bilabial closure.

To the extent that they can be compared, these values are consistent with the literature. Ishizaka *et al.* (1975) measured the compliance of the cheek and neck outer surfaces, finding cheek compliance values of 0.3×10^{-5} cm^3/dyn for a tense posture and 1.18×10^{-5} cm^3/dyn for a lax posture of the cheek, roughly equal to the values found by Wodicka *et al.* (1993) for the maternal abdomen. These values are 2–8 times smaller than the ones we found for the tongue, possibly reflecting an actual difference in compliance for tongue versus cheek surfaces. In another study, Rothenberg (1968) measured the average compliance of the vocal tract by measuring the change of pressure resulting from the introduction or removal of known quantities of air from the oral cavity. With the articulators positioned for a bilabial stop and assuming a lax posture, the measured value was 4×10^{-5} cm^3/dyn; when the cheeks and lips were tensed, the measured value was 0.68×10^{-5} cm^3/dyn. Assuming an alveolar closure position with lax vocal-tract walls, the compliance was 0.53×10^{-5} cm^3/dyn and with tense walls it was 0.38×10^{-5} cm^3/dyn. These values are somewhat lower than the ones we measured. This discrepancy may be due to the different nature of the measurements: Rothenberg measured average vocal-tract compliance while we measured compliance of the tongue dorsum and tip, which should be higher than average vocal-tract compliance (as discussed below).

It is important to note that the method we used for estimating compliance is correct only for pressure-driven movements (as tongue displacement during /p/ may have been during these experiments) but not for movements that are at least partly active. Assuming that the 4.7-mm/s linear displacement trend estimated for /b/ was the active part of the movement, and that the average time from closure to release was 100 ms, the average deformation due to active movement would be about 0.47 mm. If this estimate applied to all subjects in the study, the compliance estimates for /b/ would be 13%–21% too high, because the displacement due to pressure alone would be 0.47 mm less than the values listed in Table I. In addition to this systematic error, there is a ''random'' source of error in our measurements, arising from using values from the literature for intraoral pressure. Intersubject variability in intraoral pressure, as measured by the standard deviations in Subtelny and Worth's study, was 1.07-cm H_2O for /b/ and 1.42-cm H_2O for /p/. These standard deviations amount to 17% of the mean values for /p/ and 32% of the mean values for /b/. Since the relative accuracy of distance measurements obtained with EMMA is much lower than this 17%–32% range (Perkell *et al.*, 1992), the accuracy of our compliance estimates is limited by the accuracy of the intraoral pressure values that we used. Future studies should obtain simultaneous pressure and kinematic measures in a large number of subjects, and attempt to factor out the effect of active vocal tract movements.

B. Tongue compliance versus average vocal-tract compliance

In order to make a more informed comparison between our compliance measurements and those obtained by Rothenberg, we need to assess the possible range of average vocal-tract compliance. An indirect way of estimating average vocal-tract compliance during /aba/ is to estimate the volume of air passing into the vocal tract (during the closure for the voiced stop) and the average displacement of the subject's vocal-tract surface. The ratio of these two parameters is the average displacement of the vocal tract, and it can be used to estimate compliance. Since all our subjects are male, we used a vocal-tract surface area of 100 cm^2 for these calculations (the approximate area of a cylinder that has an average cross section of 3 cm^2 and a length of about 17 cm). Holmberg *et al.* (1988) measured average flow during production of a vowel in 25 male subjects, finding an average of 0.19 l/s and a standard deviation of 0.07 l/s. Average glottal flow during /b/ is less than during a vowel because transglottal pressure decreases during closure, as intraoral pressure rises. As a rough approximation, we estimate average flow during /b/ as 50% of the vowel flow. Average closure duration during /b/ was approximately 100 ms for our subjects. In consequence, the amount of air entering the vocal tract during /b/ is estimated as

$$\text{volume}=0.19\ \text{l/s}\times0.5\times0.1\ \text{s}=0.0095\ \text{l}=9.5\ \text{cm}^3.$$

In a vocal tract with 100-cm^2 surface, the average deformation would be 9.5 cm^3/100 cm^2=0.095 cm=0.95 mm.

Assuming a peak pressure of 5-cm H_2O for /b/, the resulting average compliance during /b/ is 1.94×10^{-5} cm^3/dyn, consistent with Rothenberg's values and between 2.6 and 4.4 times lower than our tongue compliance values. Even if we use very conservative values (0.3 l/s, the highest flow measured by Holmberg *et al.* among 25 subjects; /b/ duration of 150 instead of 100 ms), the resulting estimate of average vocal-tract compliance would be 4.6×10^{-5} cm^3/dyn, lower than any of our measured values for the tongue. We conclude that compliance of the tongue dorsum (at least during production of /b/) is quite probably higher than for the rest of the vocal tract.

C. Passive displacement or active expansion?

Cineradiographic studies have found that voiced stops are produced with a larger supraglottal volume than their voiceless cognates (Kent and Moll, 1969; Perkell, 1969). Perkell proposed that this difference may be due to passive expansion permitted by more lax vocal-tract walls during production of the voiced consonant (at least for alveolars). Kent and Moll preferred the explanation that the larger supraglottal volume during voiced stops was due to active expansion of the vocal tract. In a study that modeled the aerodynamic data obtained by Löfqvist *et al.* (1995), McGowan *et al.* (1995) found support for the hypothesis of active upper vocal-tract volume control. Rothenberg (1968) estimated that ''in bilabial and retroflexed closures the [supraglottal] cavity can be used to absorb the glottal air flow to maintain voicing during a reasonably long articulatory closure,'' but he thought this explanation was less plausible for some occurrences of voiced alveolar stops. Bell-Berti (1974) found that speakers use different mechanisms to allow pharyngeal expansion during utterance-medial voiced stops. It is indeed difficult to select one of the two explanations without access to appropriate experimental data. Westbury (1993) correctly indicated that ''it is at least difficult, if not impossible [...] to differentiate changes in vocal-tract dimensions resulting from vocal-tract expansion and cavity enlargement'' (i.e., passive or active expansion) using only kinematic data.

However, we believe our combined kinematic and aerodynamic data from subject S2 may help shed light on this issue, suggesting that tongue deformation may be passive and pressure driven for /apa/ and that active vocal-tract expansion is superimposed on the passive, pressure driven deformation for /aba/. The present data confirm that there is more vocal-tract expansion during /b/ than during /p/, but do not support the active expansion hypothesis as the *sole* explanation. Active expansion could conceivably start anytime around the beginning of the voiced stop. However, study of the synchrony between pressure and displacement data in subject S2 (see Figs. 4 and 5) shows that pressure increases are always followed a few milliseconds later by downward tongue displacements. This close synchrony is consistent with passive, pressure driven increase of supraglottal size but not with solely active expansion, unless we are willing to postulate that active tongue movements and pressure increases are synchronized so well that the lag between them has a standard deviation of 1–2 ms, a time interval about the same length as a single action potential. Thus the close synchrony favors a passive component. An active component is indicated by the modeling of tongue deformation for /aba/ tokens, which shows reasonably good results only when a linear trend is included in the model. The fitted slopes are positive (i.e., downward tongue displacement) for all ten /aba/ tokens and are quite substantial, averaging 4.7 mm/s. On the other hand, modeling /apa/ tokens shows that reasonably good fits can be obtained without a linear trend. When a linear trend is included in the voiceless data, the slopes are sometimes positive, sometimes negative, and they average only 0.48 mm/s. In summary, a presumably active, substantial, consistent linear displacement is superimposed on pressure driven tongue displacement during /aba/ but a similar case may not be made for /apa/, where tongue displacement can be reasonably explained by passive, pressure driven deformation only. However, these conclusions cannot be overgeneralized because they are based on few data: a single fleshpoint of a single talker. The relative contribution of pressure driven deformation and active movement during bilabial stops remains an open question that should be explored in future studies, to refine the estimates of tongue compliance.

IV. SUMMARY

We conclude that our kinematic data are consistent with the physical description provided in the Introduction, based on a simple lumped parameter model that includes vocal-tract resistance and compliance. Smaller tongue displacements for /p/ than for /b/ may be due to active stiffening of the tongue during /p/, and/or to intentional relaxation of tongue muscles during /b/ (in conjunction with active tongue displacement during /b/), in order to accommodate airflow into the oral cavity while maintaining a transglottal pressure differential that will allow vocal fold vibration. Finally, these data allow us to refine estimates of vocal-tract wall compliance obtained with indirect methods.

ACKNOWLEDGMENTS

This research was supported by NIDCD Grant Nos. DC00075, DC01925, and DC00423.

Bell-Berti, F. (**1975**). ''Control of pharyngeal cavity size for English voiced and voiceless stops,'' J. Acoust. Soc. Am. **57,** 456–461.

Holmberg, E., Hillman, R., and Perkell, J. (**1988**). ''Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice,'' J. Acoust. Soc. Am. **84,** 511–529.

Ishizaka, K., French, J. C., and Flanagan, J. L. (**1975**). ''Direct determination of vocal tract wall impedance,'' IEEE Trans. Acoust. Speech Signal Process. **4,** 370–373.

Kent, R. D., and Moll, K. L. (**1969**). ''Vocal tract characteristics of the stop cognates,'' J. Acoust. Soc. Am. **46,** 1549–1555.

Löfqvist, A., Koenig, L. L., and McGowan, R. S. (**1995**). ''Vocal tract aerodynamics in /aCa/ utterances: Measurements,'' Speech Commun. **16,** 49–66.

McGowan, R. S., Koenig, L. L., and Löfqvist, A. (**1995**). ''Vocal tract aerodynamics in /aCa/ utterances: Simulations,'' Speech Commun. **16,** 67–88.

Perkell, J. S. (**1969**). ''Physiology of speech production: Results and implications of a quantitative cineradiographic study,'' Research Monograph No. 53 (MIT, Cambridge, MA).

Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta, I., and Jackson, M. T. T. (**1992**). ''Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements,'' J. Acoust. Soc. Am. **92,** 3078–3096.

Rothenberg, M. (**1968**). ''The breath-stream dynamics of simple-released-plosive production,'' *Bibliotheca Phonetica*, No. 6 (Karger, Basel).

Subtelny, J. D., Worth, J. H., and Sakuda, M. (**1966**). ''Intraoral pressure and rate of flow during speech,'' J. Speech Hear. Res. **9,** 498–518.

Westbury, J. (**1983**). ''Enlargement of the supraglottal cavity and its relation to stop consonant voicing,'' J. Acoust. Soc. Am. **73,** 1322–1336.

Wodicka, G. R., Lam, A. M., Bhargava, V., and Sunkel, D. (**1993**). ''Acoustic impedance of the maternal abdomen,'' J. Acoust. Soc. Am. **94,** 13–18.