THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 56, No. 6, July–August 1977
Printed in U.S.A.

Acoustic Properties of Longitudinal Displacement in Vocal Cord Vibration

By K. ISHIZAKA and J. L. FLANAGAN

(Manuscript received January 12, 1977)

We examine the acoustic significance of longitudinal displacement in the self-oscillatory behavior of the vocal cords, and inquire into the need for representing this detail in speech synthesis. We use computer techniques and a previously derived model of the vocal cords to study the contribution of longitudinal displacement to the total acoustic volume velocity generated at the vocal cords. This volume velocity is the effective sound source for production of voiced speech. From computational results, and from speech sounds synthesized by the programmed model, we find that the contribution of longitudinal displacement is not significant perceptually, and is not essential for modeling the dominant acoustic properties of voiced speech.

I. VOCAL-CORD MODEL

In earlier work[1,2] we derived an analytical model for the self-oscillatory motion of the human vocal cords. We consider the displacing tissue of each cord to be approximated by two stiffness-coupled masses (see Fig. 1). For normal (nonpathological) conditions of phonation, the oscillator is bilaterally symmetric, and the mechanical constants of the opposing cords are identical. The left-hand mass pair (denoted m_1, m_1') constitutes the bulk of the firm cord tissue, while the smaller right-hand mass pair (m_2, m_2') represents the more flaccid mucous membrane covering of the firmer tissue. Each mass has associated with it a restoring stiffness and a resistive loss. All the stiffnesses and resistances are substantially nonlinear,[1] and in the original work, these elements act to oppose lateral motion (x-direction) only. The restriction to lateral motion still permits, of course, phase differences in the motion of the coupled masses. Lateral displacement of each mass pair determines the cross-sectional area of opening at each position. If the length of the cords, or glottal opening, is taken as ℓ_g, then the cross-sectional glottal areas are taken as rectangular shapes whose areas are $A_{gi} = 2\ell_g x_i$, $i = 1, 2$, where the factor

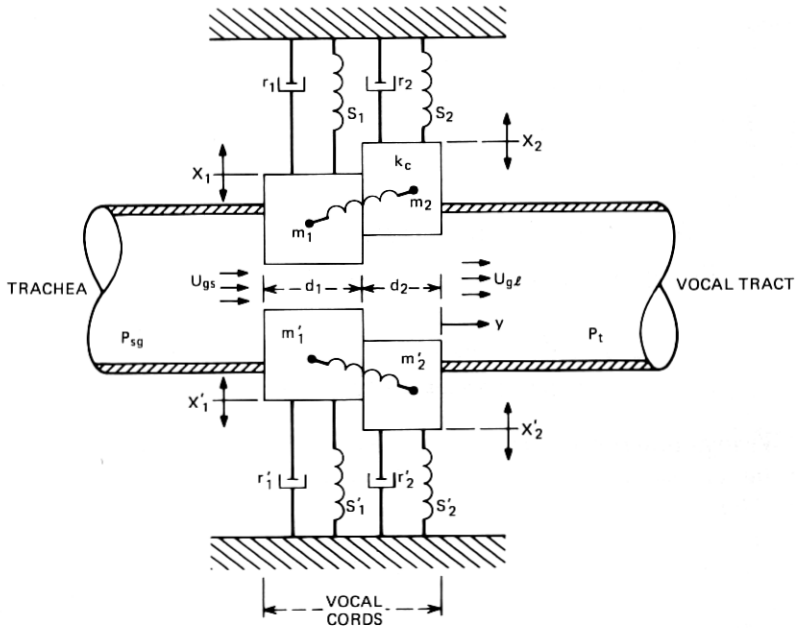

Fig. 1—Two-mass model of the vocal cords. Translational displacement is permitted in lateral (x) and longitudinal (y) directions.

2 arises from the bilaterally symmetric cord configuration. These cross-sectional areas determine the acoustic properties of the glottal volume current U_{gs}, which enters the cord orifice (from the subglottal system), and that which leaves it $U_{g\ell}$ (to pass into the larynx tube). The latter volume velocity is the effective sound source for all voiced speech sounds. The air pressure just to the left of (beneath) the vocal cords is the subglottal pressure P_{sg}, and the pressure just to the right of (above) the cords, at the entrance to the vocal tract, is P_t. The differential pressure $(P_{sg} - P_t)$ is the potential that creates the glottal volume currents.

The resulting volume currents depend upon serial acoustic impedances dictated by A_{g1} and A_{g2} and, hence, upon the cord motion, which, in turn, is conditioned by the intraglottal pressure distribution in the orifice and by the transglottal pressure $(P_{sg} - P_t)$. These serial acoustic impedances also are nonlinear (and flow dependent), and represent the mass (inertance) of air contained within the glottal orifice and the associated resistive flow losses.[1]

Additionally, there is another potential influence upon the glottal flow, namely, the volume of air displaced by the vibrating mass pairs. In general, this volume displacement can have components associated with lateral and longitudinal motion. In the original work, components of

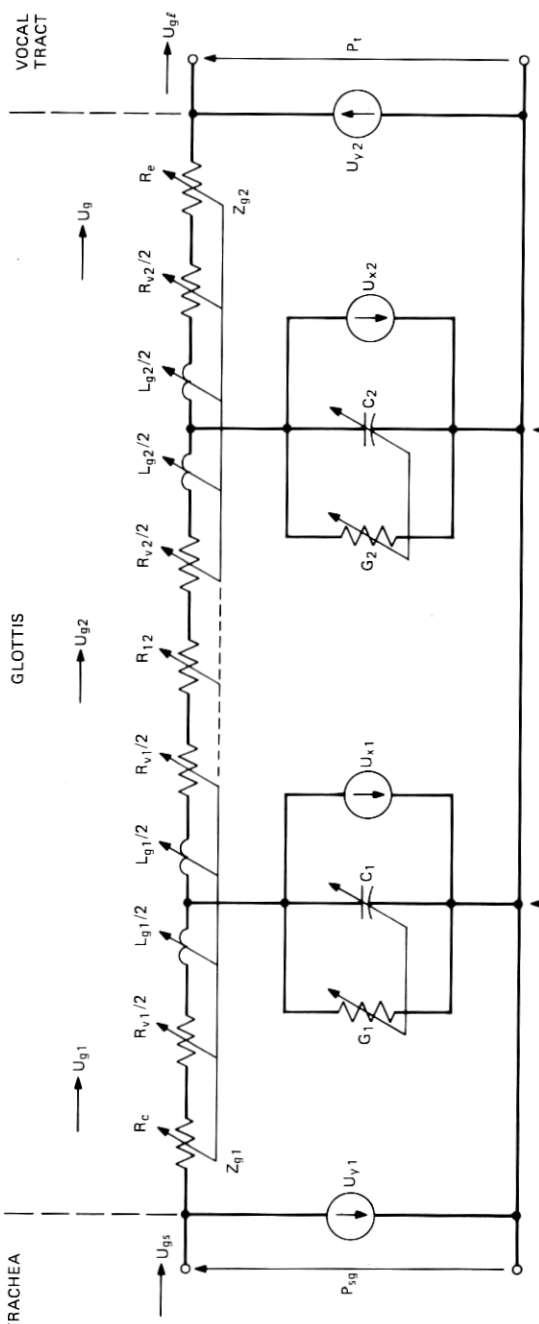

Fig. 2—Equivalent acoustic circuit including lateral and longitudinal displacement-volume velocities.

Table I — Values of impedance components of Fig. 2

Serial Impedances	$R_c = 1.37\frac{\rho}{2}\frac{\lvert U_{g1}\rvert}{A_{g1}^2}$,	$R_{12} = \frac{\rho}{2}\left(\frac{1}{A_{g2}^2} - \frac{1}{A_{g1}^2}\right)\lvert U_{g2}\rvert$
		$R_e = -\frac{\rho}{2}\frac{2}{A_{g2}A_1}\left(1 - \frac{A_{g2}}{A_1}\right)\lvert U_g\rvert$
	$R_{v1} = 12\mu\ell_g^2 d_1/A_{g1}^3$, $L_{g1} = \rho d_1/A_{g1}$,	$R_{v2} = 12\mu\ell_g^2 d_2/A_{g2}^3$ $L_{g2} = \rho d_2/A_{g2}$
Longitudinal Components	$U_{y1} = 2\ell_g(d_1 + d_2)\frac{dy}{dt}$,	$U_{y2} = 2\ell_g(d_1 + d_2)\frac{dy}{dt}$
Lateral Components	$U_{x1} = d_1\frac{d}{dt}(A_{g1})$,	$U_{x2} = d_2\frac{d}{dt}(A_{g2})$
	$= 2\ell_g d_1\, d_{x_1}/dt$,	$= 2\ell_g d_2\, d_{x2}/dt$
	$C_{g1} = A_{g1}d_1/\rho c^2$	$C_{g2} = A_{g2}d_2/\rho c^2$
	$G_1 = S_1\frac{\eta - 1}{\rho c^2}\sqrt{\frac{\lambda\omega_0}{2c_p\rho}}$,	$G_2 = S_2\frac{\eta - 1}{\rho c^2}\sqrt{\frac{\lambda\omega_0}{2c_p\rho}}$
	$S_1 = 2(\ell_g + 2x_1)d_1$	$S_2 = 2(\ell_g + 2x_2)d_2$
Constants (for vocal system, moist air at body temperature)*	$\rho = 1.14 \times 10^{-3}$gm/cm^3, air density	
	$\mu = 1.86 \times 10^{-4}$dyne-sec/cm^2, kinematic-coefficient of viscosity.	
	$c = 3.5 \times 10^4$cm/sec, sound velocity	
	$\eta = 1.4$, adiabatic constant	
	$\lambda = 0.055 \times 10^{-3}$cal/cm-sec-deg, coefficient of heat conduction	
	$\omega_0 = 2\pi(1000)$, mid audio range radian frequency	
	$c_p = 0.24$ cal/gm-degree, specific heat at constant pressure.	

* From J. L. Flanagan, *Speech Analysis, Synthesis and Perception*, second edition, New York: Springer Verlag, 1972.

glottal current corresponding to rate of volume displacement (both lateral and longitudinal) were neglected.

II. ACOUSTIC CIRCUIT

Recognizing that the cord dimensions are very small compared to sound wavelengths at the frequencies of interest, and that all mechanical velocities are small compared to the sound velocity, we derive a one-dimensional equivalent circuit for the acoustic quantities involved. Its complete form is shown in Fig. 2. The values of all impedance elements are given in Table I.

The serial elements (top branch in Fig. 2) are identical to those of our original work[1], and relate to time-variation of the acoustic impedance of the glottal opening.

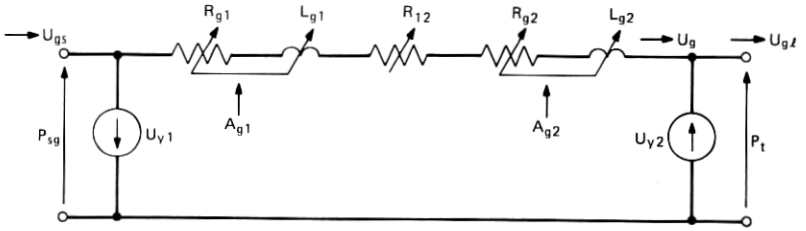

Fig.3—Simplified equivalent acoustic circuit, including longitudinal displacement currents.

All shunt elements relate to rate of displacement of air volume by the moving cord masses. The time variation of all shunt quantities is also determined by the motion of the cord masses.

Lateral motion of the cord masses (normal to the direction of glottal flow) displaces air volume at the rate of

$$U_{xi} = 2\ell_g\, d_i \frac{dx_i}{dt}\ \text{cm}^3/\text{s},\ i = 1, 2,$$

where x_i is the lateral displacement and d_i is the depth (thickness) of the cord element (mass). Again, the factor 2 arises from the two bilaterally opposing cords. The acoustic compliances, C_1 and C_2, represent the compressibility of the small air volumes contained between the opposing cords and the conductances G_1 and G_2 represent the heat-conduction loss at the tissue surfaces of the cords.

Longitudinal motion of the cord masses is assumed to occur cophasically and to be translational only. In this regard, consider the y-motion of the locked masses to be opposed by a nonlinear spring and loss similar to that of k_1 and r_1. The effective surface area exposed to the transglottal pressure difference is taken to be the product of cord length and total cord thickness, $\ell_g(d_1 + d_2)$. No cavity compliances or losses are associated with the longitudinal motion, and the longitudinal contribution to the total volume velocity is

$$U_{yi} = 2\ell_g(d_1 + d_2)\frac{dy}{dt},\ i = 1, 2.$$

In other words, U_{y1} and U_{y2} are equal and oppositely poled.

Notice that in the earlier formulation,[1,2] the absence of the shunt elements imposes the constraint $U_{g\ell} = U_{gs} = U_g$. The presence of the shunt elements (all time-varying with displacements that are determined by the equations of motion for the mechanical system which, in turn, is forced by the intraglottal and transglottal pressures to close the feedback loop of the oscillator) makes the input flow U_{gs} and the output $U_{g\ell}$ typically different.

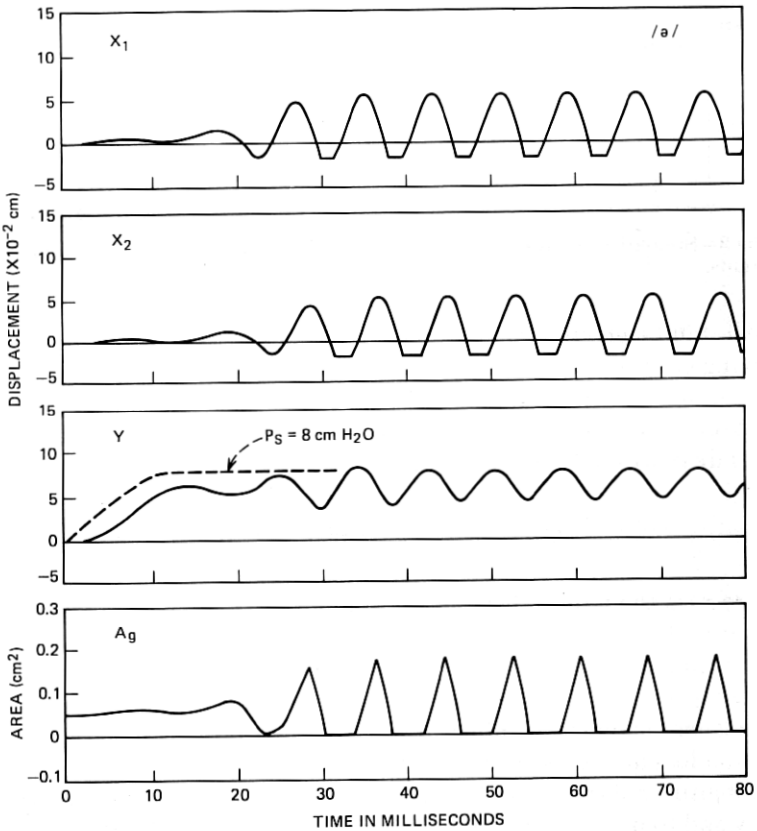

Fig. 4—Computed mechanical behavior of the vocal-cord/vocal-tract model. The vowel configuration is /ə/.

A recent related study[3] examined the influence of the U_{xi} upon $U_{g\ell}$. The present study considers separately the effects of the U_{yi}. For this purpose, the circuit of Fig. 3 is a simplification of Fig. 2.

III. PURPOSE OF PRESENT STUDY

In the original work,[1,2] we made estimates of the volume displacement currents, based upon long-wave assumptions and one-dimensional sound propagation, together with what we believed to be reasonable physiological estimates of cord velocities (compared with volume velocities responsive to transglottal pressure). We concluded that displacement currents are of second order, and in the original work we chose to neglect them in favor of elucidating dominant principles. The original formulation, therefore, treated only lateral displacement as it affects the serial glottal impedances. As a matter of completeness, we more recently have

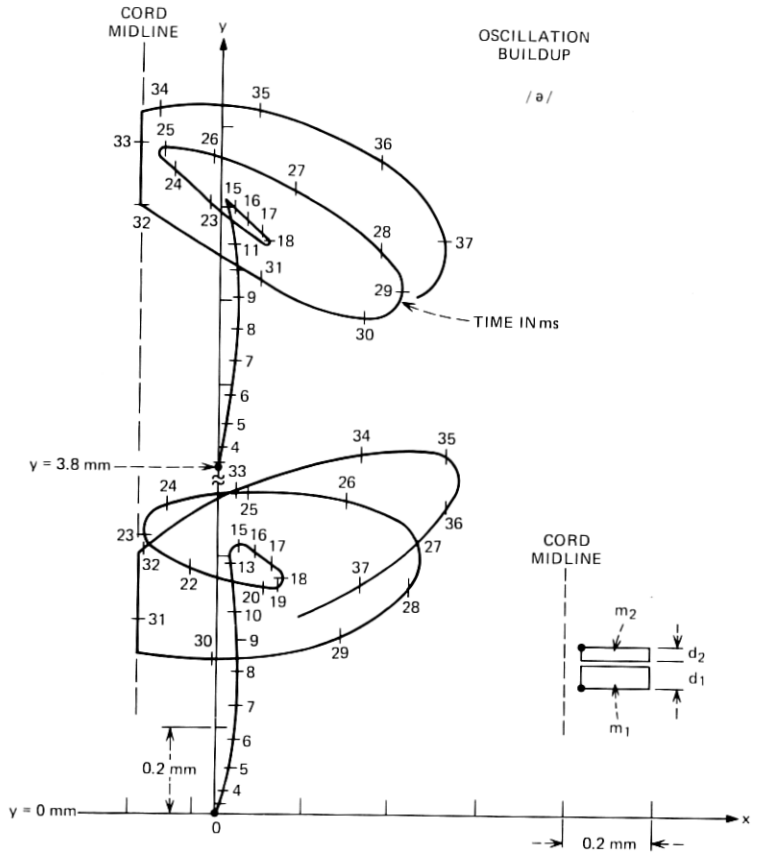

Fig. 5—X-Y trajectories for initiation of oscillation. Trajectories are for pellet positions shown in insert.

returned to a quantitative examination of these assumptions. A first study,[3] now completed, considered the importance of the shunt branches that represent the lateral components of volume velocity generated by the displacing masses—that is, from the volume current sources U_{x1} and U_{x2}. The results of the study support the original assumptions, and show the lateral components to be second order by comparison to the currents actuated by the pressure difference acting across the glottal opening.

The present study examines the contributions of the longitudinal displacement to the total glottal volume velocity (specifically, the contributions of U_{y1} and U_{y2}) and the importance of longitudinal displacement to the self-oscillatory dynamics of the cord model and to sound perception.

We take the longitudinal restoring stiffness k_y typically to be the same as the lateral restoring stiffness k_1, namely 80 kdynes/cm. The longitu-

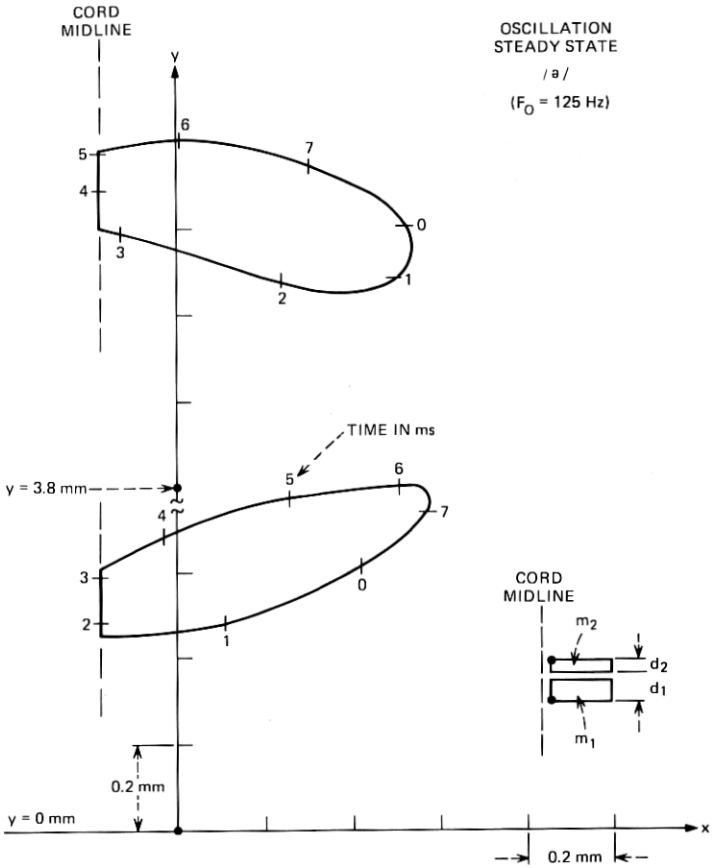

Fig. 6—X-Y trajectories for steady-state oscillation of the cord model.

dinal loss (or damping ratio) is also taken to be similar to the lateral one, namely $\zeta_y = \zeta_1 = 0.2$ These values are based upon clinical observations.[9] We examine these choices subsequently. Further, since the longitudinal and lateral motions of the cord masses are considered to be translational only, no rotational behavior is included. Still further, while the lateral translations of the coupled masses m_1 and m_2 can have large (and physiologically natural) phase differences, their longitudinal translations are considered to be cophasic, and the internal coupling stiffness is assumed to act only for lateral motion. The lateral and longitudinal motions are, therefore, coupled only through the acoustic variables that determine the oscillator forcing functions. In the course of our discussion, we will indicate comparisons to actual physiological data to assess the realism of these assumptions.

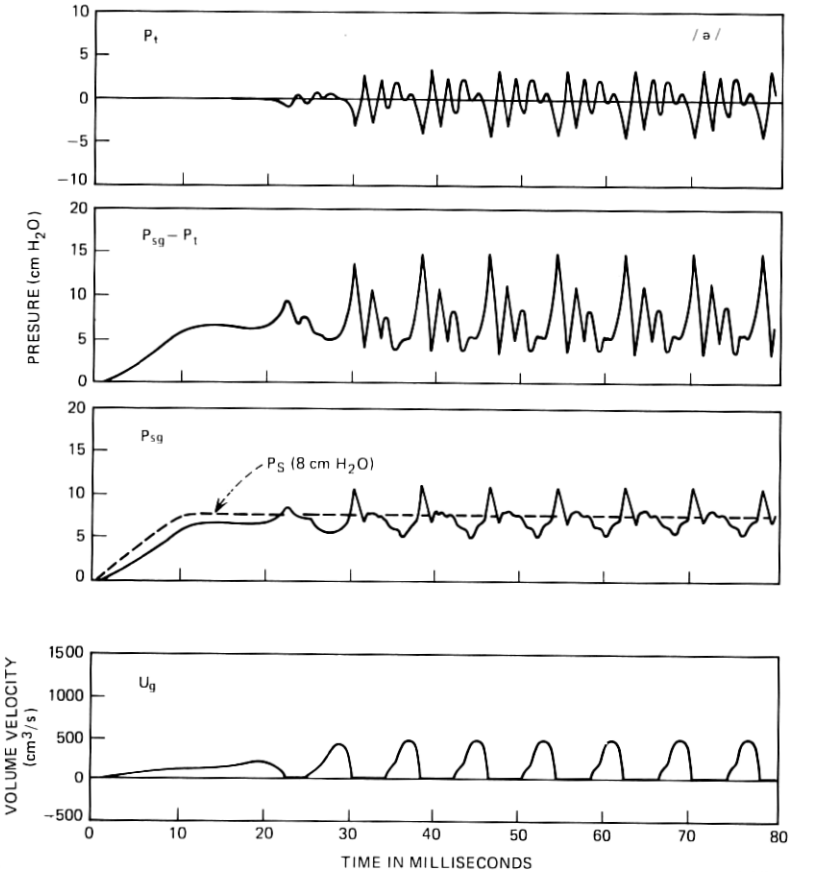

Fig. 7—Computed acoustic qualities for the vocal-cord/vocal-tract model. The vowel is /ə/.

IV. RESULTS OF COMPUTER SIMULATIONS

The vocal-cord model, as represented by Fig. 3, was combined with a transmission-line formulation of the vocal tract that we have used previously in speech synthesis studies.[4] The programmed vocal tract contains 20 sections which, in addition to the classical acoustic elements, represents the yielding soft walls of the tract and sound radiation from the yielding walls. This formulation is based upon measurements of tissue impedances that we reported earlier.[5] Also included for the present study is a transmission-line representation of the subglottal system. Six sections of line represent the trachea, bronchi and lungs, as previously described.[6] We implemented the entire system in terms of difference equations programmed on a laboratory computer by techniques we have described in detail previously.[1,2]

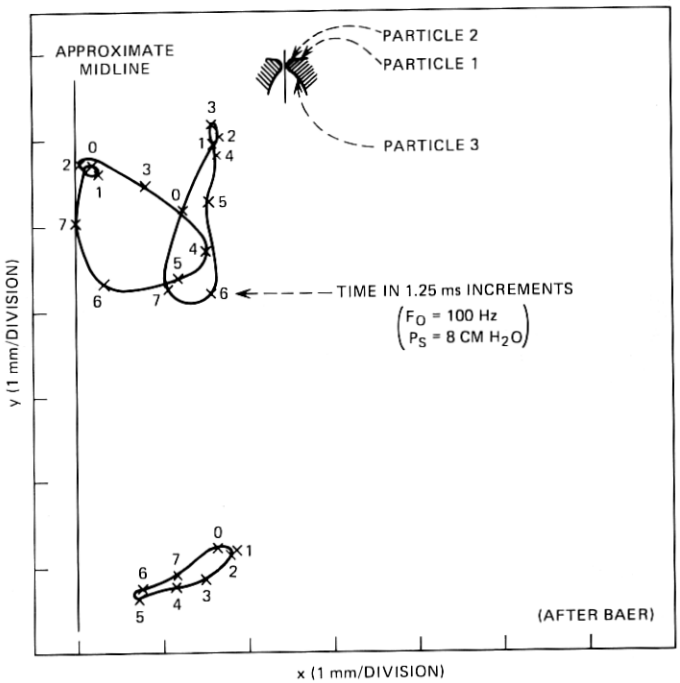

Fig. 8—X-Y trajectories observed from excised larynx of dog (after Baer).

Most of the data reported here are for the vocal tract configured in the shape for the neutral (schwa) vowel /ə/. Some data are also included for the vowels /i/ and /a/.

A first step is to ascertain if the cord oscillator, so arranged for longitudinal motion, performs realistically when compared with observations on the human larynx. A second step, then, is to determine the acoustic significance of the volume displacement current arising from longitudinal motion.

Throughout these calculations, the laryngeal parameters are set to the "standard" values used earlier for phonation by a man's voice in the chest register[1] (i.e., neutral glottal area $A_{go} = 0.05$ cm^2, cord tension parameter $Q = 0.78$, $d_1 = 0.25/Q$ cm, $d_2 = 0.05/Q$ cm). Recall that the Q parameter scales the values of mass and stiffness and, hence, also the values of the d_i. Phonation is initiated by raising the lung pressure P_s smoothly from zero to the standard value of 8 cm H_2O. The pressure is elevated in a 10-ms interval.

4.1 Mechanical behavior

As the lung pressure is elevated, the model commences a buildup of oscillation. After four or five transient swings, the oscillation settles into

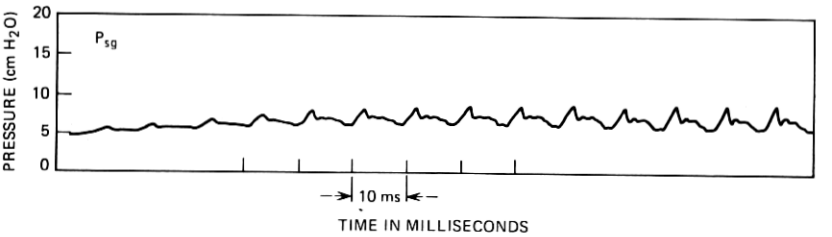

Fig. 9—Subglottal pressure variation measured on a human subject (after Sawashima).

a steady state behavior with a fundamental frequency (pitch) determined by the model parameters (the tension parameter Q has the dominant effect on pitch[1]). The initial 80 ms of this synthetic phonation is illustrated in Fig. 4 for the mechanical variables.

The top two curves show the displacements x_1 and x_2 of mass pair m_1 and mass pair m_2, respectively. The first collision of each mass pair is indicated by the first flat, negative-going portion of the displacement waveforms. For the $A_{go} = 0.05$ cm^2 value, this occurs for $x_i = -0.0178$ cm. Note, too, that x_1 leads x_2 in phase by the order of 60°, which is

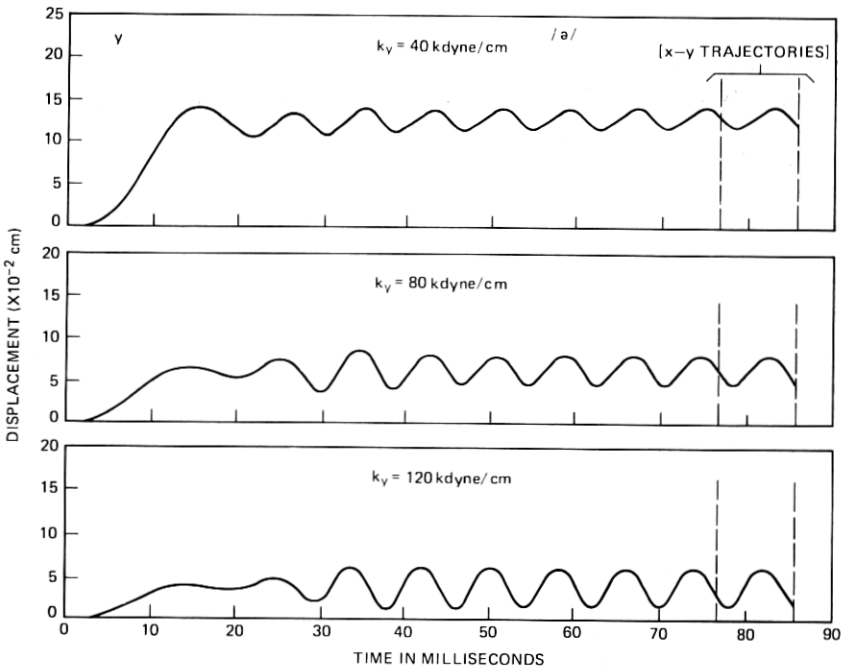

Fig. 10—Effect of longitudinal restoring stiffness k_y upon the longitudinal displacement, y. Data show oscillation buildup for a lung pressure P_s that is raised smoothly from zero to 8 cm H_2O.

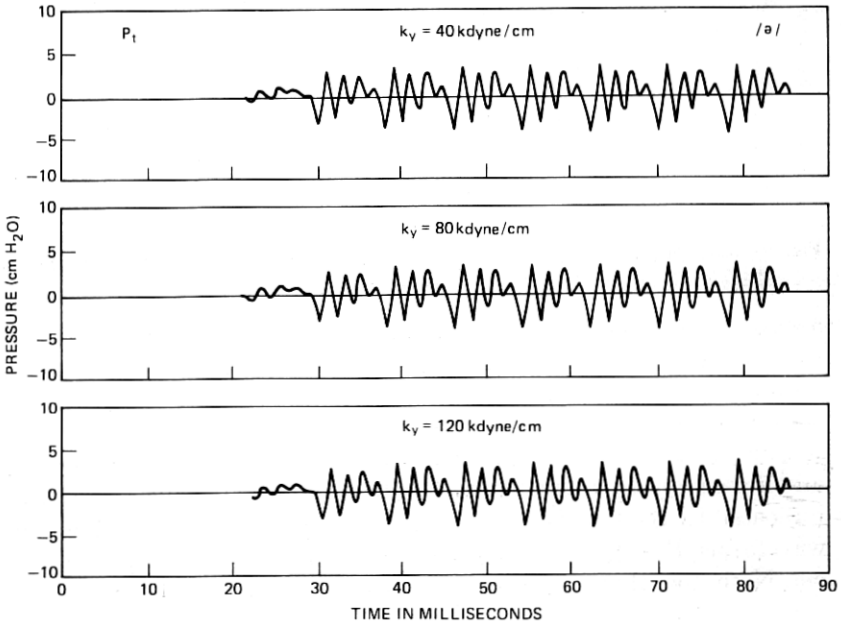

Fig. 11—Effect of k_y upon P_t.

consistent with observations from high-speed motion pictures of the human vocal cords. The third trace shows the longitudinal displacement, y, which bulges upward as P_s is raised. The y motion is roughly sinusoidal. The lower trace shows the net area of glottal opening A_g (namely the minimum of A_{g1} and A_{g2}). The y-displacement is seen to lead in phase the A_g wave, again consistent with the upward, rolling motion seen in high-speed photography of the real cords.

An x-y plot of the buildup transient portrays the behavior perhaps more graphically. Figure 5 shows the x_1 vs y and the x_2 vs y values with time as the parameter. Imagine pellets fixed to the lower and upper inner edges of one simulated cord, shown by the inserted anterior-posterior view of Fig. 5. The trajectories of the two pellets are plotted for the oscillation buildup. The y-axis is broken and re-originated at $y = (d_1 + d_2) = [(2.5 + 0.5\text{mm})/Q] = 3.8$ mm. The flat portion of the tracks, along the vertical midline, reflect collision with the opposing vocal-cord mass.

After several initial swings, the oscillator settles into a steady-state behavior. One cycle of this trajectory is shown in Fig. 6. The steady-state pitch frequency in this case is $Fo = 125$ Hz, or a period of $T = 8$ ms.

4.2 Acoustical behavior

The corresponding acoustical parameters, calculated for the same buildup period, are shown in Fig. 7. The acoustic pressure at the input

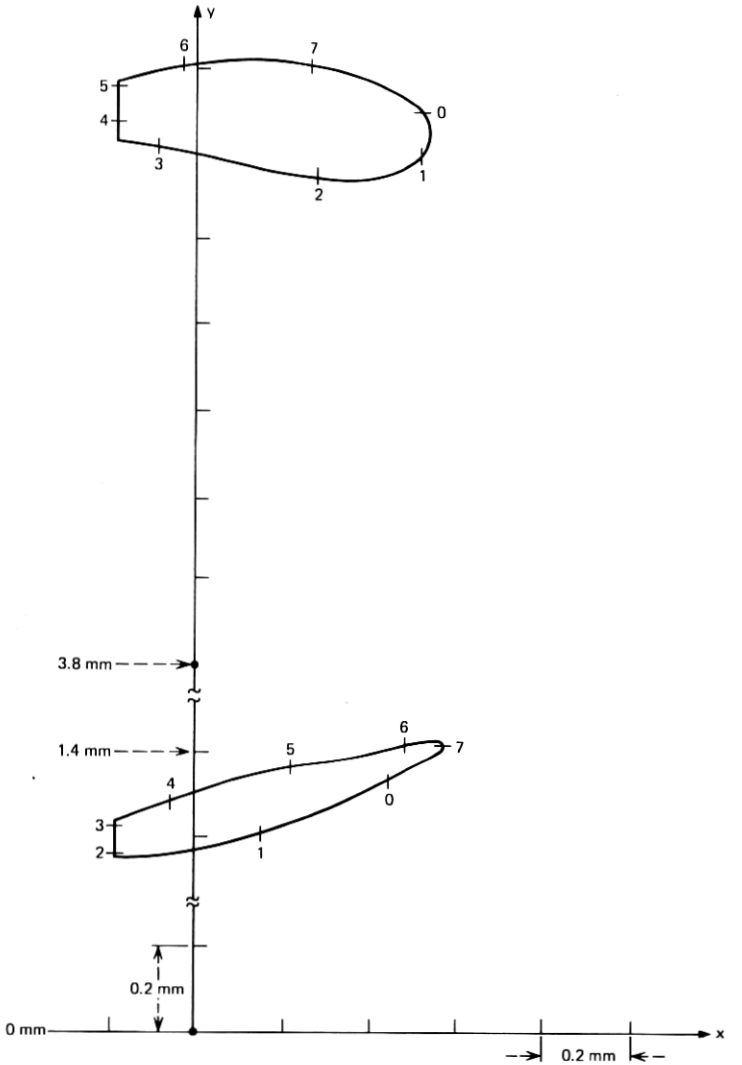

Fig. 12(a)—Steady-state X-Y trajectory for k_y = 40 kdyne/cm.

to the vocal tract P_t is shown in the top trace. It reflects strongly the eigenfrequency structure of the tract, in this case configured for /ə/ and having formant frequencies of approximately 500 Hz, 1500 Hz, 2500 Hz . . . The transglottal pressure ($P_{sg} - P_t$), which is the forcing function for the y-motion and the pressure potential for the volume flow through the glottal opening, exhibits a pronounced pitch-synchronous variation. Its peak values, in fact, approach twice the lung pressure value of P_s = 8 cm H_2O. Recall that P_s is the lung pressure input to the simulated

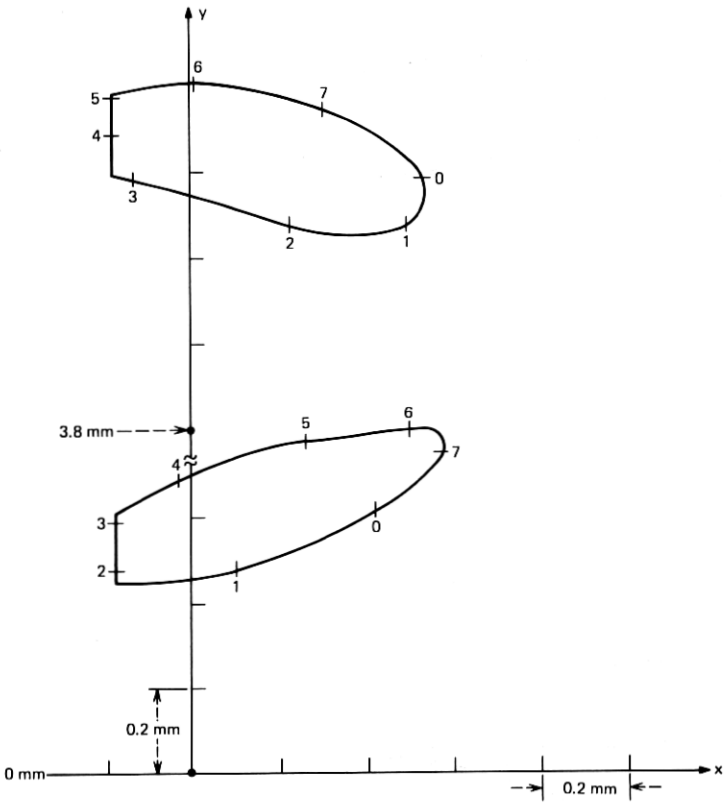

Fig. 12(b)—Steady-state X-Y trajectory for k_y = 80 kdyne/cm.

subglottal system, representing trachea and bronchi. But notice that the mechanical y-displacement (Fig. 4) does not respond with this detail. (Neither do the x_1 and x_2 displacements respond to high-frequency detail in their forcing functions—that is, the mechanical system, being mass-controlled, filters out this detail.)

The subglottal pressure P_{sg} (the pressure just beneath the vocal cords) also exhibits a pitch-synchronous fluctuation, but of somewhat less amplitude, namely about ±20 to ±30 percent of the mean subglottal pressure. Its positive peaks correspond to the closing epochs of the glottal port. The calculated volume velocity passing the glottal opening U_g (bottom trace) appears as a traditionally shaped, pulsive waveform. This wave is similar to that calculated in previous work (without longitudinal motion) but differs in that its values are modified by the effects that U_{y1} and U_{y2} couple into the pressure variables. That is, U_{y1} and U_{y2} can influence P_{sg} and P_t and, hence, $U_{g\ell}$. The latter three variables, in turn, close the oscillator feedback loop by constituting the forcing functions

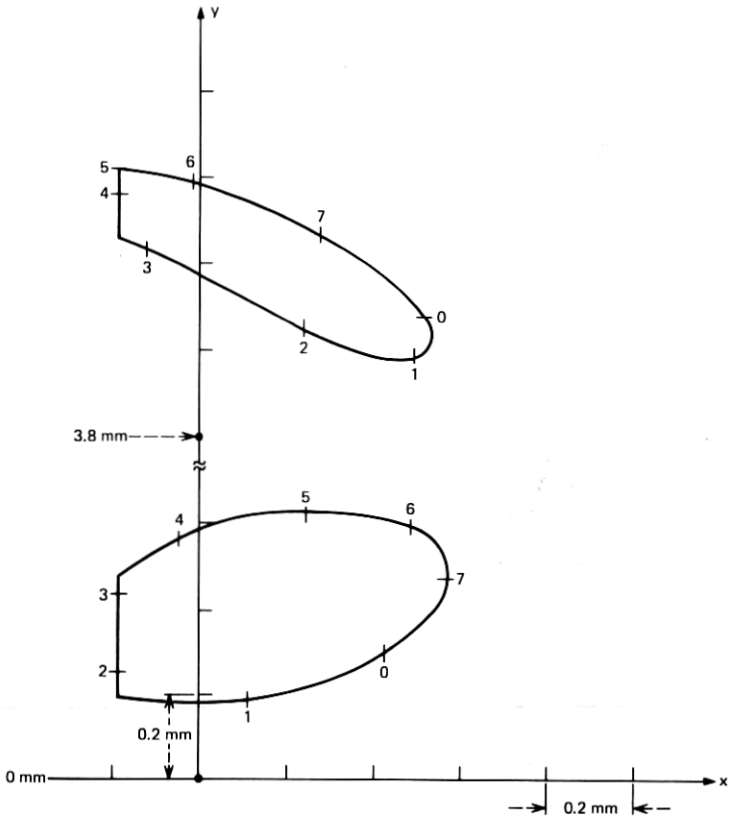

Fig. 12(c)—Steady-state X-Y trajectory for k_y = 120 kdyne/cm.

for the lateral displacement. As was the case in the mechanical variables, the U_g flow does not reflect a temporal fine structure comparable, say, to the $(P_{sg} - P_t)$ waveform. The resistive and inertive components of the glottal impedance (i.e., the serial components in Fig. 3) act effectively as a low-pass filter. It is not unusual, however, to see pronounced temporal structure that corresponds to the lowest eigenfrequency of the vocal tract, especially for low, back vowels (such as /a/), or for tightly articulated sounds.

A next question, then, is how do these mechanical and acoustical quantities, resulting from the model with longitudinal displacement, compare with physiological data.

4.3 Comparisons to physiological observations

One qualitative comparison can be made for the mechanical displacement behavior. Baer[7] performed studies on the excised larynx of

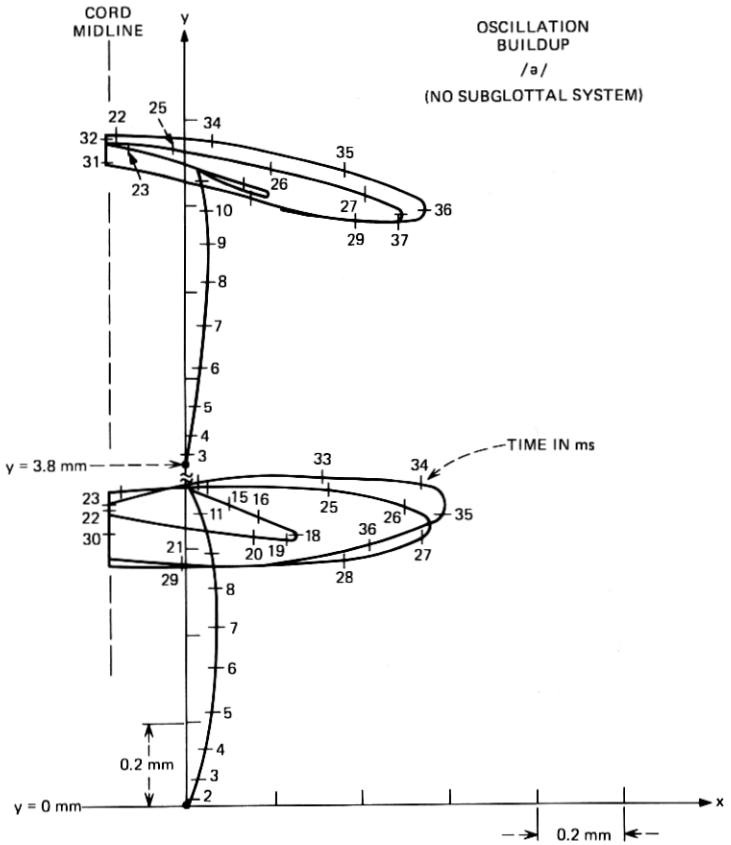

Fig. 13—Oscillation buildup without subglottal system.

a dog in which he fixed pellets to the displacing tissue and made optical observations under stroboscopic illumination. While his pellet positions do not correspond exactly to our mass-pair corners, we can roughly compare his observations with our data. Figure 8 shows x–y trajectories for one set of conditions for the dog larynx that approximates values used in human phonation (namely P_s = 8 cm H_2O, U_g = 275 cm^2/s, and Fo = 100 Hz). Particles (pellets) 2 and 3 are of interest. While the vibratory excursions of the excised dog larynx are larger than those we calculate with the model, the qualitative motions are gratifyingly similar. One question that arises is how much does the longitudinal (vertical) displacement depend upon the choice of longitudinal stiffness constant. We shall examine this question in more detail subsequently.

Another comparison can be made in the acoustic domain—namely, to the subglottal pressure variation P_{sg} shown previously in Fig. 7. Sawashima[8] has measured the subglottal pressure during phonation in a

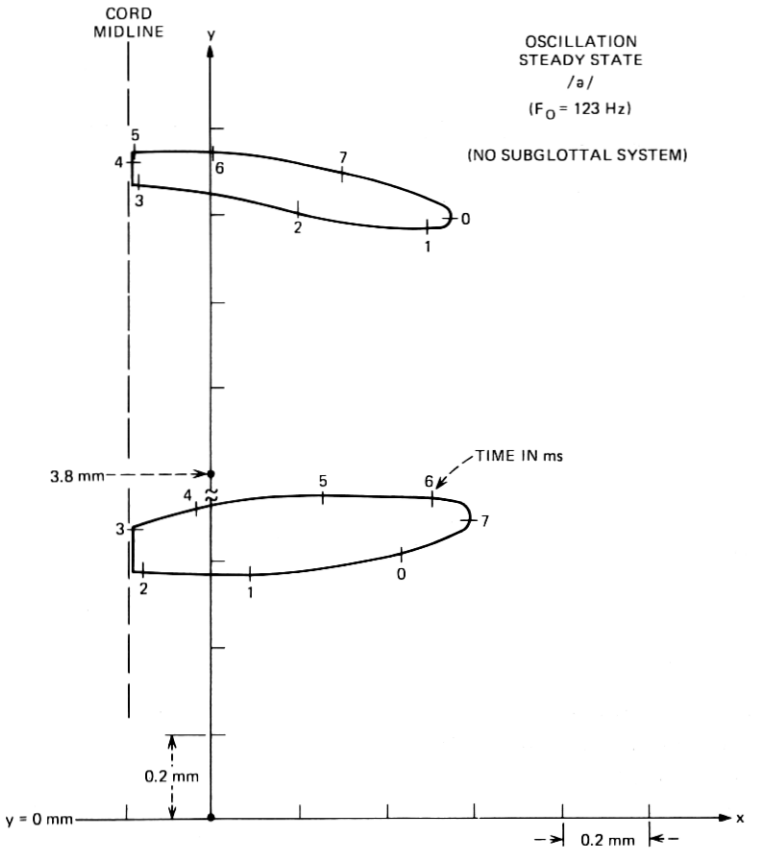

Fig. 14—Steady-state oscillation without subglottal system.

human subject. One of his results is shown in Fig. 9. The qualitative correspondence to the model calculation appears relatively good, and the acoustic interaction among the simulated cords, vocal tract, and subglottal system is realistic.

4.4 Effect of longitudinal stiffness constant

In view of uncertainties in the measurement of the stiffness constants in physiological preparations, it is important to examine how critical the value of k_y (the longitudinal restoring stiffness) is to the oscillatory behavior of the model.

For the bulk of our studies, we have taken k_y equal to our typical "standard" value of k_1, namely 80 kdynes/cm.[1] We have also used the standard value for the damping ratio, $\zeta_y = \zeta_1 = 0.2$. This choice is based upon the physiological measurements on cord tissue conducted by Ka-

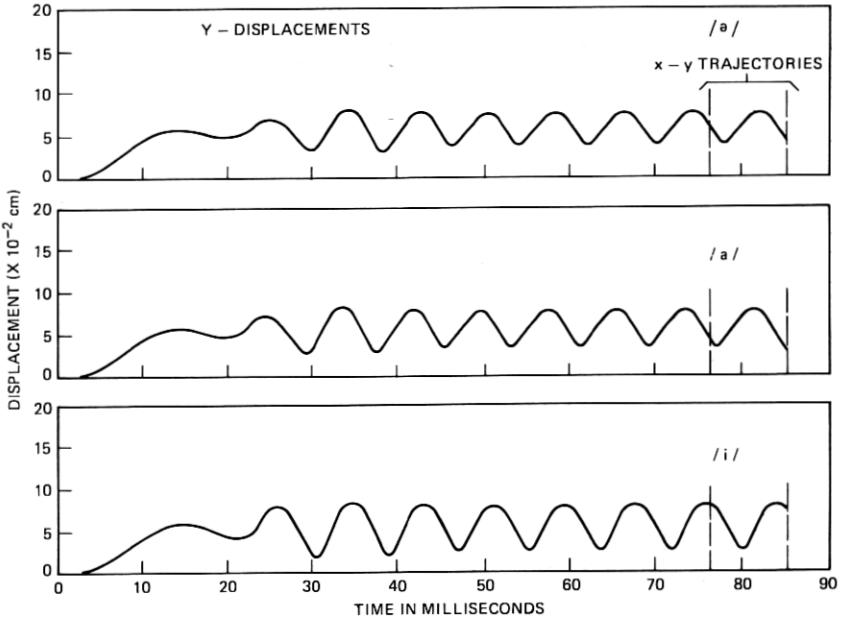

Fig. 15—Effect of vowel configuration upon longitudinal displacement.

neko,[9] who found the stiffness constants for lateral and for longitudinal displacement to be similar. Both stiffnesses are also taken to have the same nonlinearity, namely, a cubic nonlinearity in the restoring force (see Ref. 1).

To assess the model's sensitivity to large variations in k_y, we calculated the buildup of synthetic phonation for $k_y = 40$, 80, and 120 kdynes/cm. The resulting y-displacements for these values is shown in Fig. 10, and the sound pressure at the entrance to the vocal tract is shown in Fig. 11. Also, the x-y trajectories are shown in Fig. 12a, b, and c.

As would be expected, the greatest influence in this variation is reflected in the y-displacements. The "softer" k_y gives larger dc displacement and smaller peak-to-peak vibratory excursions. The P_t data indicate that the variations in acoustic behavior and glottal excitation are very small. The fundamental pitch is sensibly the same for all cases, namely 125 Hz. This factor is almost completely dominated by the lateral motion. In auditory assessment of the output synthetic sound, the differences are virtually imperceptible, suggesting that the longitudinal displacement current is insignificant for speech synthesis.

4.5 Effect of subglottal system

A side issue, of some interest in passing, is the effect of acoustic interaction between the subglottal system and the cord oscillator. If the

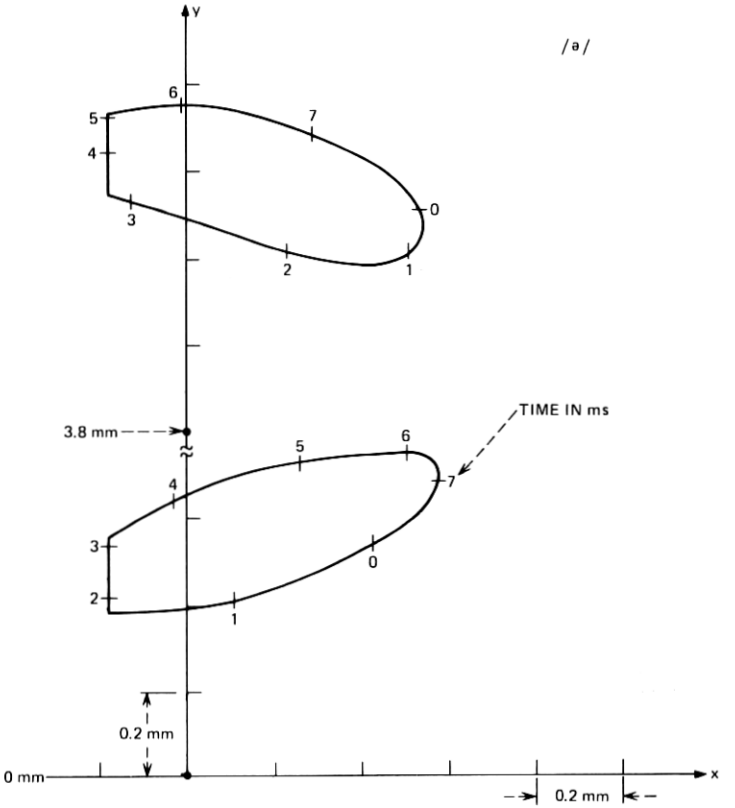

Fig. 16(a)—X-Y trajectory for the vowel /ə/.

trachea-bronchi-lung system is removed and the lung pressure applied directly to the cord input (as a zero-impedance source; i.e., P_{sg} becomes a pressure "battery" equal to P_s), then the temporal structure previously reflected in P_{sg} is eliminated, the transglottal pressure excursions simply equal P_t, and the longitudinal component of cord displacement is lessened. This is illustrated by the x-y trajectories for oscillation buildup and steady state shown in Figs. 13 and 14. For this case, k_y is reset to the typical value 80 kdynes/cm. Note the slight lowering of the fundamental frequency to 123 Hz.

4.6 Effect of vowel configuration

It also is instructive to consider the influence of vowel configuration upon the cord model, as presently formulated. Such studies were made in detail in the original work.[1] We therefore compare synthesized results for the vowel /ə, a and i/. (In this case the longitudinal stiffness constant

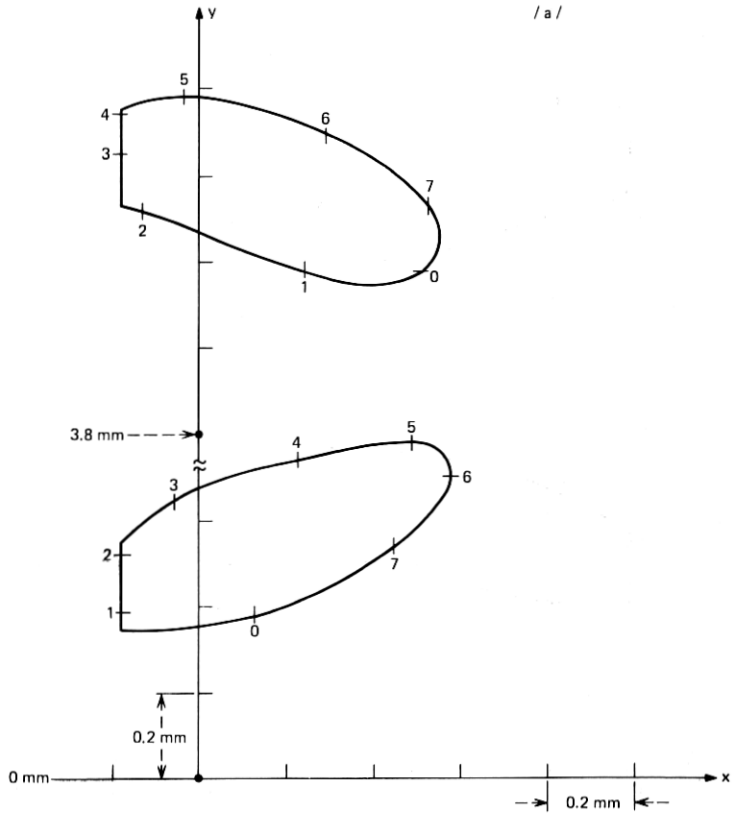

Fig. 16(b)—X-Y trajectory for the vowel /a/.

was set to $k_y = 88$ kdynes/cm through an inadvertent keypunch error. Because the differences are so small, it did not seem worthwhile to recompute the data for $k_y = 80$ kdynes/cm.)

The y-displacements are shown in Fig. 15, and the x-y trajectories for one period of steady-state oscillation are shown in Fig. 16a, b, and c. The longitudinal displacement is not greatly affected by vowel configuration, but the constricted articulations /a, i/ clearly lead to slightly greater longitudinal peak-to-peak excursions than does the open-pipe (neutral) vowel /ə/. This typically is owing to the greater acoustic interaction at the eigenfrequencies for the configurations with higher acoustic impedance levels, which in turn leads to greater transglottal pressure differences. This is well reflected in the acoustic variables resulting from this calculation. The corresponding acoustic quantities are shown in Figs. 17 thru 20. In these data, note especially how the tract eigenfrequencies are manifest, including in the synthetic output sound pressure from the

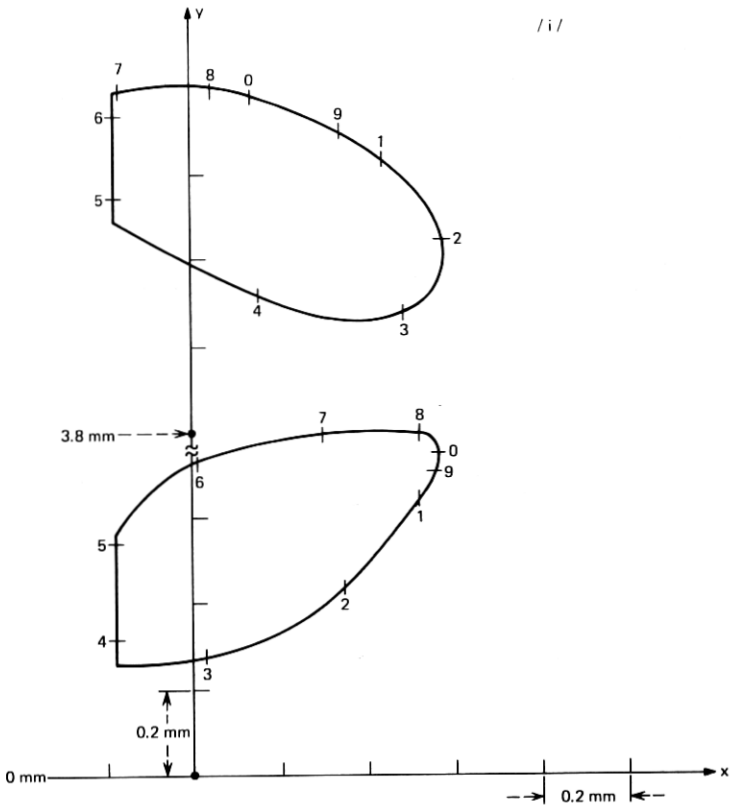

Fig. 16(c)—X-Y trajectory for the vowel /i/.

mouth, P_{out}. Note, too, that the resulting steady-state pitch frequencies are about 125 Hz for /ə/ and /a/ and about 120 Hz for /i/.

V. SIGNIFICANCE OF LONGITUDINAL DISPLACEMENT

Having established that the cord oscillator (with lateral and longitudinal degrees of freedom) appears to behave realistically, we consider the next questions:

(*i*) Is the longitudinal motion significant or necessary for proper self-oscillatory operation of the model?

(*ii*) Is the acoustic volume velocity contributed by the longitudinal motion physically or perceptually significant?

So far as the purposes of speech synthesis are concerned, we answer both of these questions in the negative.

What we wish to do, therefore, is compare the mechanical and acoustical behavior with and without longitudinal displacement. Because the longitudinal effects are coupled only through U_{y1} and U_{y2}, the lat-

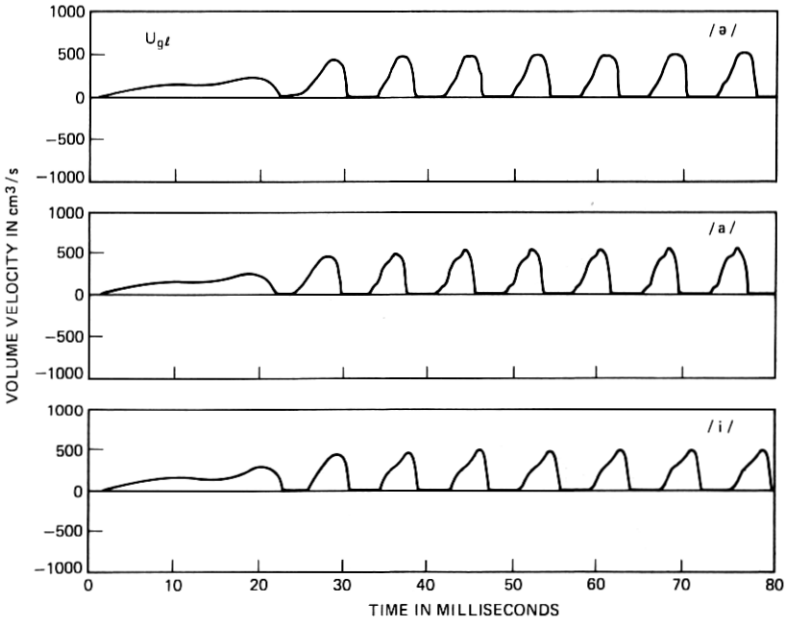

Fig. 17—$U_{g\ell}$ for the vowels /ə/, /a/, and /i/.

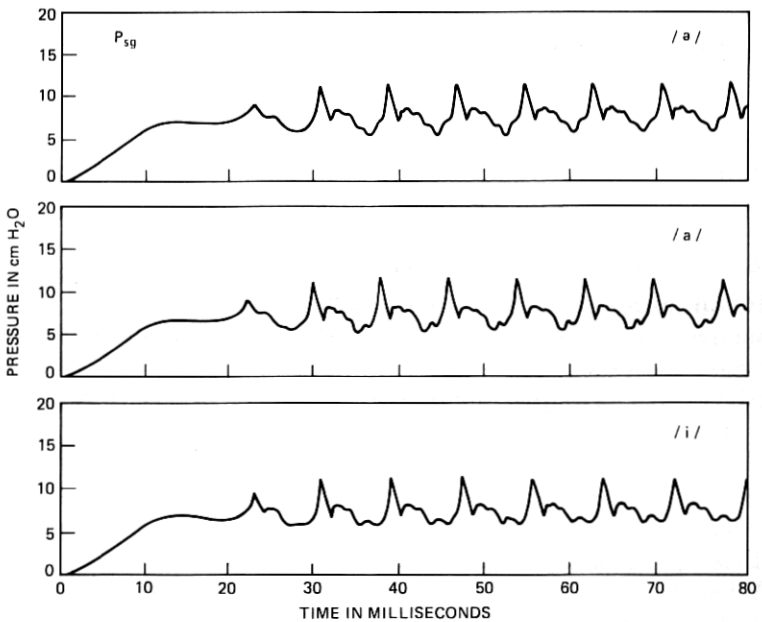

Fig. 18—P_{sg} for the vowels /ə/, /a/, and /i/.

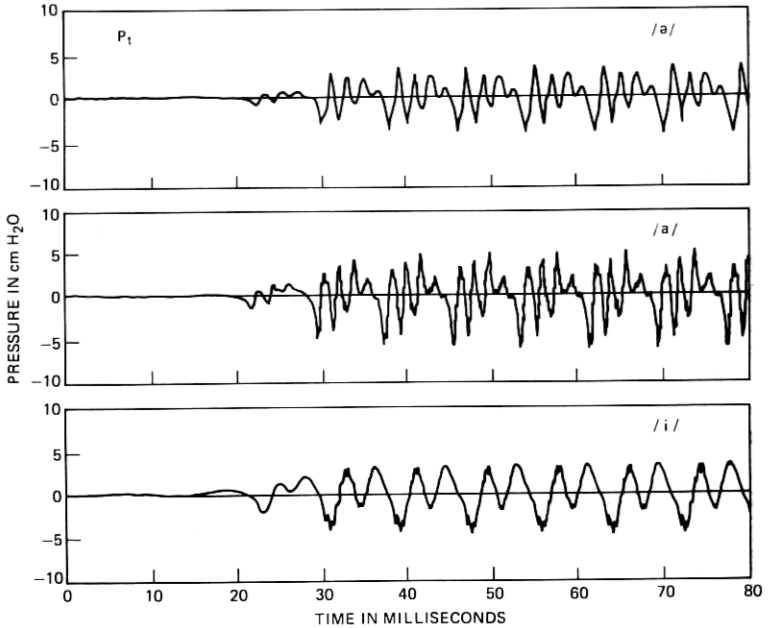

Fig. 19—P_t for the vowels /ə/, /a/, and /i/.

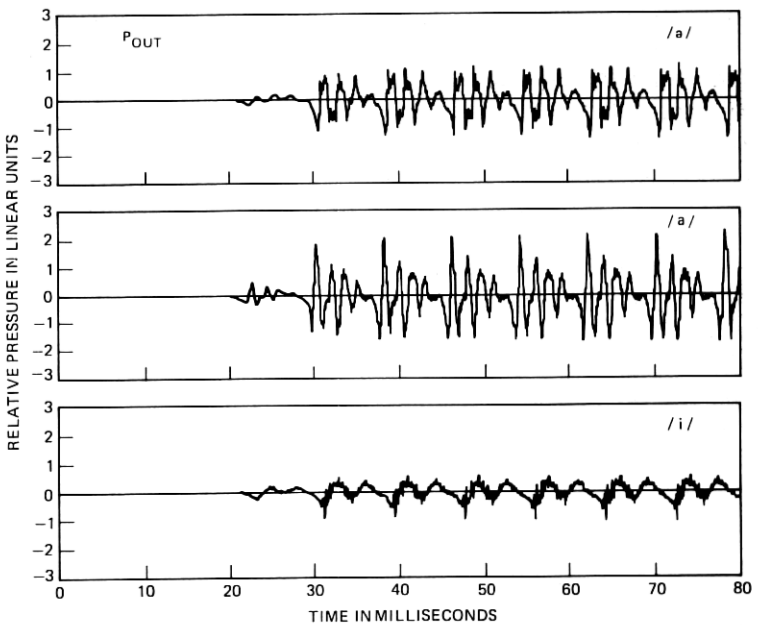

Fig. 20—P_{out} for the vowels /ə/, /a/, and /i/.

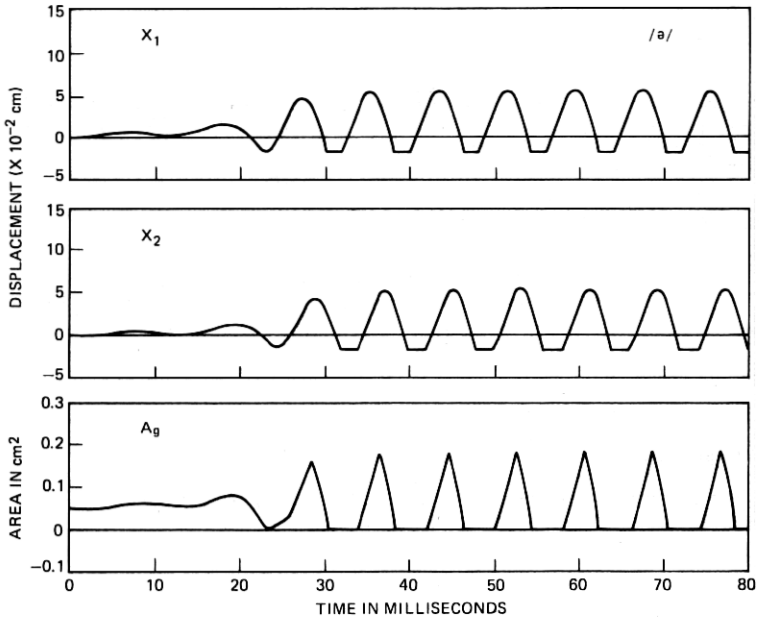

Fig. 21—Computed mechanical behavior of the cord/tract model without y-displacement.

eral-motion-only condition (that is, the original formulation of the model) is conveniently realized by setting $U_{y1} = U_{y2} = 0$. This condition, with no longitudinal displacement flow, yields the mechanical results shown in Fig. 21. These results are essentially identical (at least to three significant figures) to the corresponding quantities in Fig. 4. In particular, note that the crucial A_g waveforms are virtually identical.

We can now examine pertinent acoustic quantities. Calculation of the conditions with y-displacement yields the result shown in Fig. 22. The figure shows one cycle of the steady-state oscillation. Recall that $U_{g\ell}$ is the total volume velocity at the larynx tube entry to the vocal tract. U_g is the flow component through the actual opening of the glottis. Note that $U_{g\ell}$ is non-zero during the time the cords are actually closed, corresponding to an upward (vertical, longitudinal) displacement of air volume that adds positively to U_g. Similarly, later in the cycle, downward longitudinal displacement subtracts from U_g. The difference between these volume velocities is

$$(U_{g\ell} - U_g) = U_{y2} = U_{y1}$$

by virtue of the assumption of cophasic longitudinal motion. This difference is plotted on an X10 enlarged scale in Fig. 22. The peak value

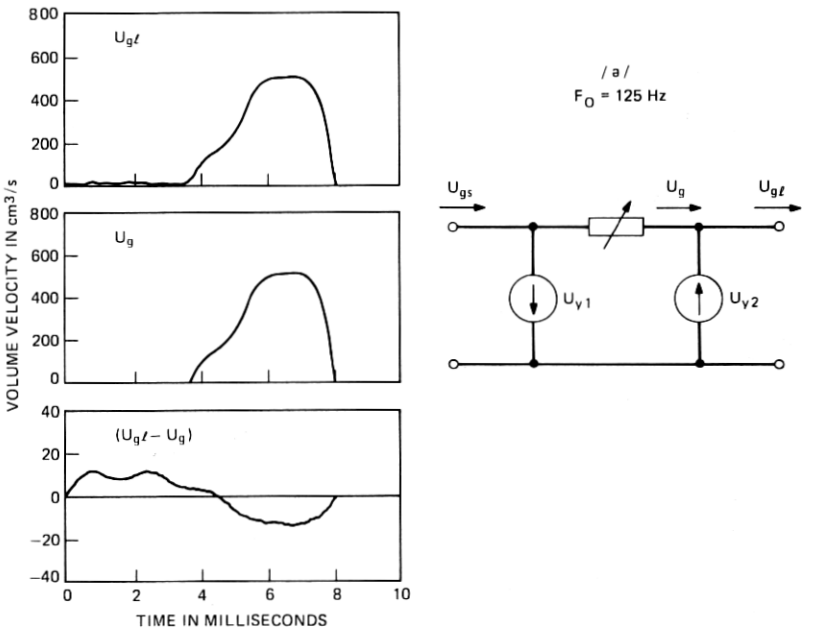

Fig. 22—Glottal volume velocities calculated with y-displacement (refer to Fig. 3).

of the difference for this condition is on the order of 10 cm^3/s, or about $1/50$ of the peak value of the total $U_{g\ell}$. This result is not just peculiar to this range of volume velocity, but rather it scales comparably at louder and softer phonation. For example, if the lung pressure P_s is doubled say to 16 cm H_2O, the longitudinal displacement current increases because the transglottal pressure and the longitudinal displacement increase. But the U_g flow also increases and remains far and away the dominant quantity.

The amplitude spectra of these quantities provide convenient correlation with auditory percepts. The spectra for $U_{g\ell}$ and U_g are shown in Figs. 23a and b. A close comparison shows the differences to be less than 2 dB, an amount that is not significant perceptually. The more relevant comparison is obtained when the effect of y-motion is eliminated (by removing U_{y1} and U_{y2}). The corresponding glottal waveform for no y-displacement is illustrated in Fig. 24. It is denoted $U_{g\ell}^*$. Also reproduced is the $U_{g\ell}$ with y-displacement. Further, the difference between the longitudinal displacement and lateral-only conditions $(U_{g\ell} - U_{g\ell}^*)$ is shown on an X10 enlarged scale. During the glottis-closed time, this difference is identical to the $(U_{g\ell} - U_g)$ difference of Fig. 22, because $U_g = 0$. During the glottis-open time, the $(U_{g\ell} - U_{g\ell}^*)$ difference differs from the $(U_{g\ell} - U_g)$ difference. In other words, U_g differs from $U_{g\ell}^*$

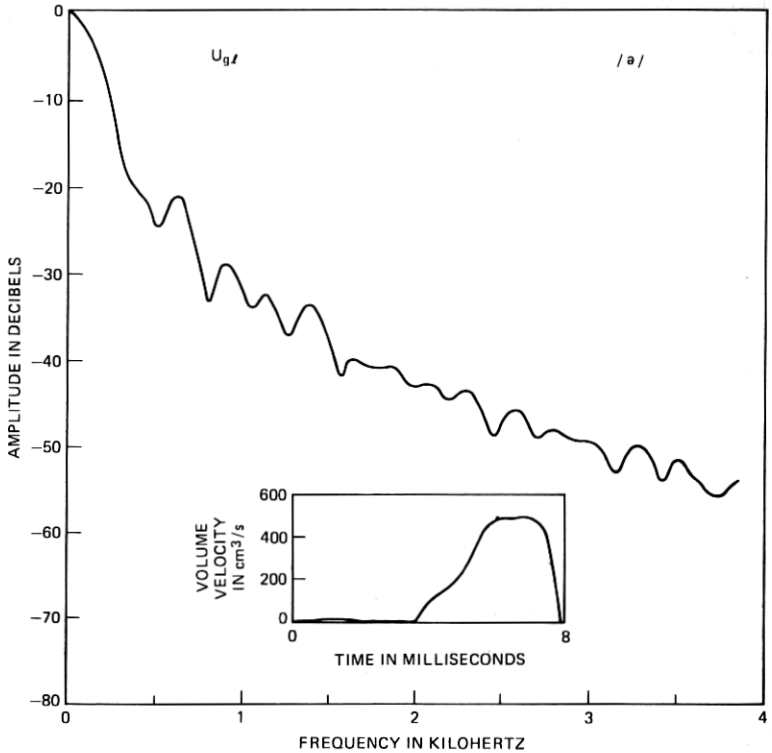

Fig. 23(a)—Amplitude spectrum of $U_{g\ell}$, which includes the effects of y-displacement.

essentially by the influence that U_{y1} and U_{y2} have upon the transglottal pressure difference $(P_{sg} - P_t)$.

Again, the more perceptually relevant comparison is to the amplitude spectrum. The spectrum of $U_{g\ell}^*$ is given in Fig. 25. A close comparison to the $U_{g\ell}$ spectrum of Fig. 23 shows the differences to be less than about 2 dB. Auditory assessment of the output synthetic vowels shows them to be indistinguishable even in close comparison.

VI. CONCLUSION

In view of these results, we conclude that realistic acoustic behavior (which is needed in speech synthesis) can be obtained in the cord model without the additional complexity of longitudinal displacement. Longitudinal displacement is not necessary for realistic self-oscillation of the model. The important vertical phase differences in the two-mass motion are adequately duplicated by lateral displacement only, as is the significant acoustic interaction between vocal tract and vocal cords. Further, the rate of volume displacement owing to longitudinal motion

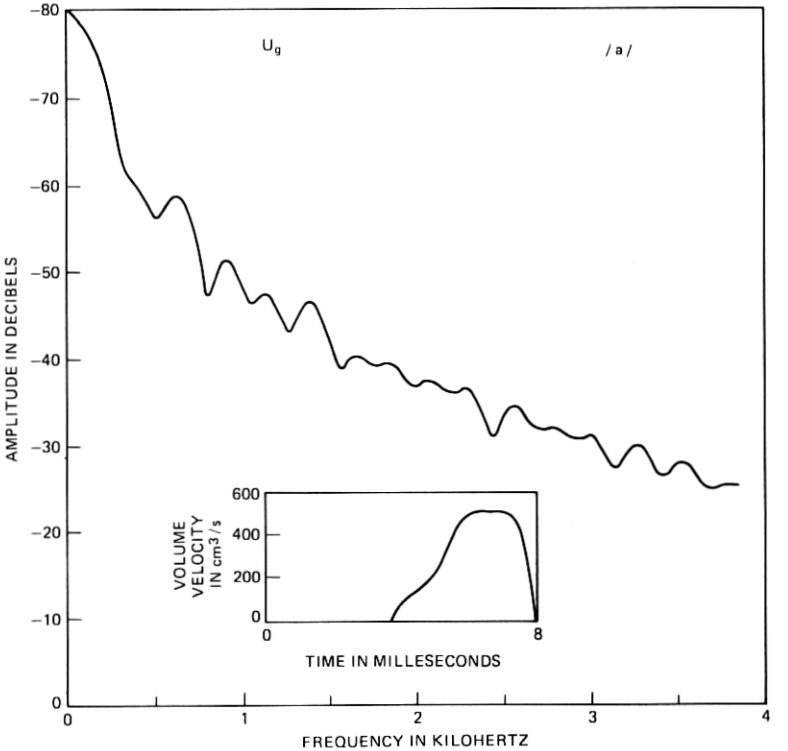

Fig. 23(b)—Amplitude spectrum of U_g calculated with y-displacement.

is clearly perceptually not significant and need not be represented with added detail.

These conclusions about the mechanical and acoustic behavior have a corollary in a companion study on the rate of displacement of air volume owing to lateral motion only.[3] This contribution was examined by making use of the shunt branches in Fig. 2 that include U_{x1}, U_{x2}. Calculations and computer simulations showed that the contribution to glottal volume velocity of the air extruded from the glottal port by lateral tissue displacement is barely discriminable in a differential auditory comparison. In fact, the perceptual effect for the lateral volume displacement is just slightly larger than for the longitudinal displacement. Both are quite second-order in importance.

We have found in the present study that proper acoustic and oscillatory behavior of the model does not depend significantly upon longitudinal displacement. The longitudinal motion is relatively insensitive to acoustic loading and to changes in longitudinal stiffness. The longitudinal motion influences fundamental frequency only slightly. What,

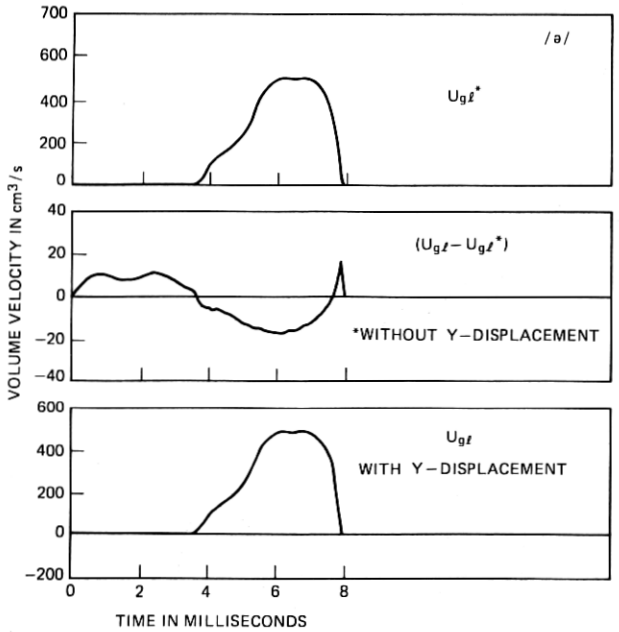

Fig. 24—Comparison of waveforms for $U_{g\ell}$, which includes y-displacement current, and $U_{g\ell}^*$, which is calculated without y-displacement.

then, are the critical and sensitive parameters of the cord model? In other words, what parameters are most influential upon the perceptual attributes of $U_{g\ell}$, since the end product—the output sound—depends directly upon $U_{g\ell}$? The results of our earlier work can be combined with the insights obtained here to consider this question.

The original study showed that the intra-glottal pressure distribution, and the fluid flow laws used to deduce it, are quite important to proper oscillatory behavior, to proper generation of the $U_{g\ell}$ flow, and to realistic acoustic interaction between the vocal tract and vocal cords. To a large extent this pressure distribution determines how the pitch frequency varies with subglottal pressure and with articulatory configuration. The mass-stiffness product (i.e., the natural frequency of the mechanical system) is quite dominant in determining pitch range. Subglottal pressure, assuming it to be above an initiation threshold of several cm H_2O, is primarily correlated with sound intensity, a relatively noncritical factor for voiced sounds. Mechanical parameters such as cord thickness, damping ratio, and nonlinearity are relatively noncritical except as they influence duty factor and "flow chopping" at collision (which yields a broad-spectrum $U_{g\ell}$ function). None of the mechanical variables, lateral or longitudinal, reflects the temporal fine structure of the acoustic variables, but both must and do reflect the open-close cycles of the vi-

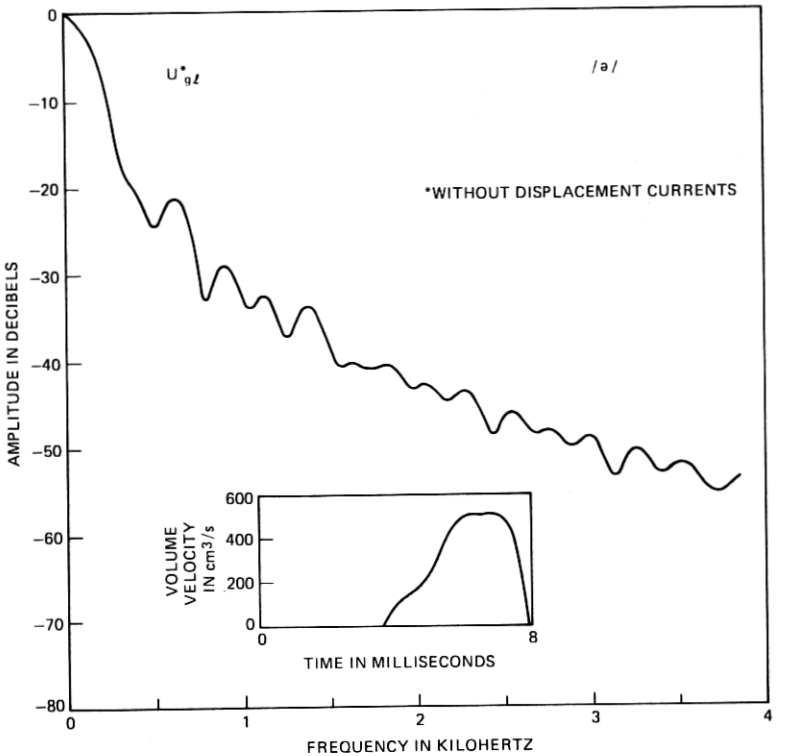

Fig. 25—Amplitude spectrum of $U_g^*\ell$, the glottal volume velocity without y-displacement.

brating cords. The realistic phase differences in motion of the upper and lower edges of the cords (m_2 and m_1 in the model) allow phonation smoothly over a wide range of input impedances to the vocal tract (both inductive and capacitive), and this behavior can be obtained satisfactorily by permitting lateral displacement only of the stiffness-coupled masses. The computational complexity of anything more detailed does not seem necessary from the standpoint of duplicating realistic acoustic behavior, which is the objective in speech synthesis.

On the other hand, if the objective were a detailed study of tissue deformation (as might be the case in simulations for clinical diagnosis or for representing pathological conditions) then the computational complexity of longitudinal displacement might be considered. In such a case, the vocal-cord model should be treated as a more distributed system. For the representation and synthesis of normal speech, however, these details do not appear perceptually significant and are not needed to represent the dominant properties of vocal-cord vibration.

REFERENCES

1. K. Ishizaka and J. L. Flanagan, "Synthesis of Voiced Sounds from a Two-Mass Model of the Vocal Cords," B.S.T.J. *50,* (July–August 1972), pp. 1233–1268.
2. J. L. Flanagan and K. Ishizaka, "Automatic Generation of Voiceless Excitation in a Vocal-cord/Vocal-tract Speech Synthesizer," IEEE Trans. ASSP, *24* (April 1976), pp. 163–170.
3. K. Ishizaka and J. L. Flanagan, "Acoustic Characterization of the Air Volume Displaced by the Vibrating Vocal Cords," Proc. (British) Inst. Acoust. (Sept. 1976).
4. J. L. Flanagan, K. Ishizaka, and K. L. Shipley, "Synthesis of Speech From a Dynamic Model of the Vocal Cords and Vocal Tract," B.S.T.J., *54,* No. 3 (March 1975), pp. 485–506.
5. K. Ishizaka, J. C. French, and J. L. Flanagan, "Direct Determination of Vocal Tract Wall Impedance," IEEE Trans. ASSP, *23,* No. 4 (August 1975), pp. 370–373.
6. K. Ishizaka, M. Matsudaira, and T. Kaneko, "Input Acoustic-Impedance Measurement of the Subglottal System," J. Acoust. Soc. Am., *60,* No. 1 (July 1976), pp. 190–197.
7. T. Baer, "Investigation of Phonation Using Excised Larynxes," Ph.D. thesis, M.I.T., Cambridge, Mass., 1976.
8. M. Sawashima, H. Hirose, and Y. Kurauchi, "Subglottal Pressure during Phonation," Jpn. J. Logopedics and Phoniatrics, *5,* No. 2 (August 1964), pp. 84–85 (in Japanese).
9. T. Kaneko, H. Asano, H. Miura, and K. Ishizaka, "Biomechanics of the Vocal Cords—On Stiffness," Practica Otologica (Kyoto), *64,* 1229–1235, 1971.