

CHAPTER 21

ACOUSTIC INTERACTION BETWEEN THE GLOTTAL SOURCE AND THE VOCAL TRACT

Martin Rothenberg
Syracuse University,
Syracuse, New York, USA

INTRODUCTION

Though the glottal sound source is often considered to have a volume velocity waveform independent of the supralaryngeal configuration during vowel-like sounds, it has long been suspected that the separation of sound source and vocal tract can lead to a significant error in the estimation of voice quality (Fant, 1960; Flanagan 1968). It is generally realized that there can be appreciable first formant energy absorbed by the glottis during the open phase of the glottal cycle, and that this energy can cause oscillations on the glottal flow (volume velocity) waveform and a change in the frequency and damping of the formant. However, it is not generally recognized that this interaction can also have a strong effect on the overall waveshape of the glottal pulse, and, in particular, on the amount of high frequency energy generated at the instant of vocal fold closure.

High speed and stroboscopic motion pictures of the glottis during chest voice have generally yielded a rather symmetrical, triangular waveform for the projected glottal area, as, for example, in the samples shown by Dr. Hiroto in Chapter 1 of this book. On the other hand, measurements of the glottal flow waveform by inverse-filtering the sound pressure or the flow at the mouth have often shown a markedly unsymmetrical waveform, with a slowly-rising glottal opening phase and a sharply terminating glottal closing phase (for example, see Miller, 1959; Holmes, 1962; and Rothenberg, 1973). The glottal flow waveforms in simulation studies have usually shown these

same characteristics, as we have seen in the contribution to this conference by Dr. Ishizaka and Dr. Titze (Chapters 17 and 18 of this book). This dissymmetry of the glottal flow waveform can be an important determinant of voice quality in that it increases the high frequency energy of the waveform, as compared to the projected area waveform, and concentrates the energy in the glottal closed phase, during which the vocal tract is most efficient (Fant, 1979).

Of the possible causes for this flow dissymmetry, the following three appear to me to be the most likely to be significant:

- (1) a non-invariant relationship between projected glottal area and glottal flow conductance (the reciprocal of flow resistance) due to the different vocal fold configurations that exist during the opening and closing phases,
- (2) the small component of air flow which is due to the air displaced by vocal fold motion (Rothenberg, 1973; Rothenberg and Zahorian, 1977),
- (3) the effect of the supraglottal acoustic impedance.

The relationship between projected glottal area and flow conductance is likely to be significant in determining voice quality under some conditions. We will not consider this factor further in this paper, however, and we make the common first-approximation assumption that the relation between glottal conductance and projected glottal area is invariant.

Since the air displaced by vocal fold motion tends to decrease the glottal flow as the folds separate, and increase the flow as the folds come together (as in a hand clap), this component will tend to cause a dissymmetry of the type we are discussing. A simple calculation shows that the component will be small, but not necessarily negligible (Rothenberg, 1973. See also the simulation result by Flanagan and Ishizaka (1978).) However, in this paper we will not consider this component further, and we will study, instead, the effect of the third factor. We will consider the effect of supraglottal loading on the glottal waveform, using a model which is valid at the fundamental frequency and lowest order harmonics, since it is these components which most strongly influence the overall waveshape of the glottal pulse. Our comments will be restricted to acoustic interaction and not include the effect that the supraglottal pressure variations might have on the motion of the vocal folds.

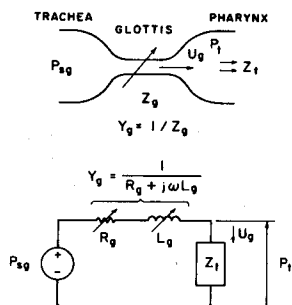


Fig.1. Simplified linear model for the glottal source.

A MODEL FOR SOURCE-TRACT ACOUSTIC INTERACTION

Figure 1 shows a simple linear, lumped-parameter model for the glottis and the supraglottal vocal tract. The elements in the model are in standard electrical circuit form, but are defined acoustically as follows:

- Z_g = glottal impedance (The dissipative part of the glottal impedance is termed the glottal resistance.)
- Y_g = glottal admittance (The dissipative part of the glottal admittance is termed the glottal conductance.)
- U_g = glottal volume velocity
- P_{sg} = subglottal pressure
- Z_t = impedance of the supraglottal vocal tract as seen by the glottis
- P_t = supraglottal pressure

In this paper, we will consider the subglottal pressure to be constant and the glottal inertance L_g to be zero. The glottal admittance will then be a pure conductance that is the inverse of the glottal resistance R_g and is equal to $U_g/(P_{sg}-P_t)$. In the following theoretical development, we will initially consider this admittance to have the symmetrical, triangular shape shown at the upper right in Fig.2, as the vocal folds open and close during the glottal air pulse. Though this representation does not properly reflect the flow dependency of the glottal resistance (which, for larger glottal areas causes the differential or small-signal flow resistance to be somewhat higher

than the resistance defined by pressure/flow, and causes both measures of resistance to increase with volume velocity for a given glottal area), it should yield a good approximation to the actual glottal flow for small values of Z_t , if the variation in projected glottal area is approximately triangular (Flanagan 1958).

In the common assumption of independence between the glottal source and the vocal tract, the supraglottal impedance Z_t is considered negligible compared to Z_g , and therefore, as shown in the curve for $R_t = 0$ in Fig.2, the glottal air flow would have the same shape as the glottal admittance function Y_g . In the figure, the pressure, admittance and time scales are normalized to arbitrary units ($P_{sg} = 1$, Y_g maximum = 1, pulsewidth = 2).

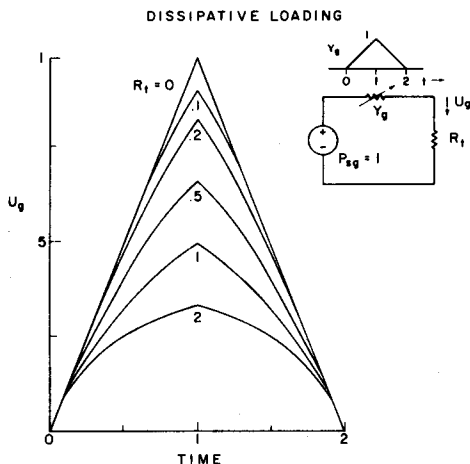


Fig.2. Computed glottal air flow with a dissipative supraglottal impedance and a triangular variation in glottal admittance.

In looking at the effect of a non-negligible Z_t , let us first assume Z_t to be purely dissipative, as shown in Fig.2, for some representative values of R_t . In this case, the air flow U_g is given by

$$U_g(t) = \frac{P_{sg}}{R_g(t) + R_t} = \frac{P_{sg}}{\frac{1}{Y_g(t)} + R_t} \quad (1)$$

At the initiation and termination of the air flow pulse, R_g is high and dominates R_t , and the resulting flow is similar to that which would occur with no R_t . However, as R_g decreases (and Y_g increases) the flow becomes limited by R_t . For high values of R_t , the flow U_g approaches a rectangular pulse of amplitude P_{sg}/R_t , as it is "switched" on and off by the onset and offset of Y_g . This rectangular pulse will have a much higher ratio of high-frequency energy to low-frequency energy than would be the case with R_t equal to zero (the non-interactive model).

A reactive supraglottal impedance will also cause an increase in the proportion of high frequency energy, The way in which this change occurs, however, is different from that for the dissipative case. If the lowest supraglottal resonance (the first formant F_1) is higher than the voice fundamental frequency F_0 , as is usually the case, the supraglottal loading will be inertive at frequencies between F_0 and F_1 . In addition, in a more accurate representation the inertive component of the glottal impedance would be added to the load impedance in determining the flow, thus increasing the importance of the effect of inertive loading. In this paper, we will present a mathematical analysis for only the case of an inertive reactance, assuming, for simplicity, that Z_t is due to a pure inertance L_t , so that $Z_t(\omega) = j\omega L_t$.

Figure 3 shows our simplified model with an inertive supraglottal loading of magnitude L_t . The flow U_g for this model is governed by the following differential equations:

$$L_t \frac{dU_g}{dt} + \frac{U_g}{t} = 1, \quad \text{for } 0 \leq t \leq 1, \quad (2A)$$

$$L_t \frac{dU_g}{dt} - \frac{U_g}{t-2} = 1, \quad \text{for } 1 \leq t \leq 2, \quad (2B)$$

with the added condition that $U_g(0) = 0$.

As can be verified by substitution in the equations, the solution for this set of differential equation is

$$U_g = \frac{t}{L_t + 1} \quad \text{for } 0 \leq t \leq 1. \quad (3A)$$

For $1 \leq t \leq 2$ the solution has two forms, depending on the magnitude of L_t :

$$U_g = \begin{cases} \frac{2L_t}{L_t^2 - 1} (2-t)^{1/L_t} - \frac{2-t}{L_t - 1} & \text{for } L_t \neq 1 \\ \left[\frac{1}{2} - \ln(2-t) \right] (2-t) & \text{for } L_t = 1 \end{cases} \quad (3B)$$

These functions are plotted in Fig.3 for some representative value of L_t . It can be seen that inertive loading would tend to give the flow pulse its often-noted skew to the right and rapid termination. Since the high frequency energy produced at a discontinuity of slope tends to be proportional to the change of slope at the discontinuity, the high frequency energy produced at the termination of the glottal pulse can be greatly increased by inertive loading. The terminal slope increases rapidly with L_t , even at values of L_t which do not drastically reduce the peak amplitude of the flow waveform. This differs from the dissipative case, in which a steep termination is obtained only by limiting the pulse height, i.e., in which the proportion of high frequency energy is increased by reducing the low frequency energy (which varies with the pulse height). In the inertive case, there is an actual increase in high frequency energy caused by the loading.

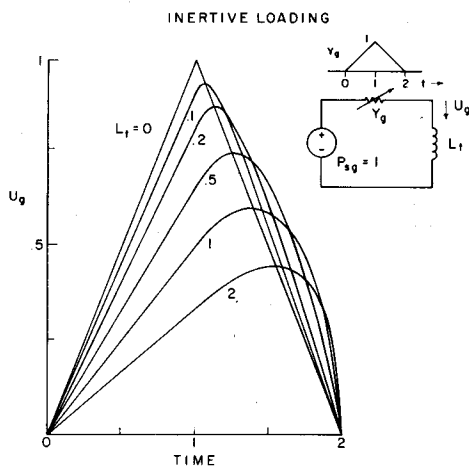


Fig.3. Computed glottal air flow with an inertive supra-glottal impedance and a triangular variation in glottal admittance.

It can also be seen in Fig.3 that the response between $t = 0$ and $t = 1$ is a linear increase in flow with a smaller slope than for the non-interactive case. Thus the high frequency energy produced at the onset of the glottal pulse is reduced. In the limit, with high values of L_t , U_g approaches a "sawtooth" or "ramp" waveform with a linear increase from zero flow at $t = 0$ to a peak flow of $2/(L_t+1)$ at $t = 1$, followed by a sudden decrease to zero flow at $t = 2$. The spectrum of such a pulse falls off at only 20 dB/decade compared to the 40 dB/decade fall-off for the triangular pulse that occurs with $L_t = 0$.

However, the increase of high frequency energy produced by the glottal closure is not just an asymptotic effect that is present only with high values of L_t . The increase of terminal slope occurs with relatively small values of L_t , and in fact, the terminal slope becomes infinite for values of L_t greater than unity. This result can be seen from the time derivative of $U_g(t)$. For $1 \leq t \leq 2$ the derivative is

$$\frac{dU_g}{dt} = \begin{cases} \frac{1}{L_t - 1} - \frac{2}{L_t^2 - 1} (2-t) \left(\frac{1}{L_t} - 1\right) & \text{for } L_t \neq 1 \\ \frac{1}{2} + \ln(2-t) & \text{for } L_t = 1 \end{cases} \quad (4)$$

For $0 \leq t \leq 1$ the slope is $1/(L+1)$ for all values of L . Thus the slope at the onset of air flow decreases gradually with L_t , while the terminal slope increases rapidly for values of L_t approaching unity as the term $1/(L_t-1)$ increases and dominates the second term. For $L_t > 1$ the second term becomes infinite as t approaches 2, since $(\frac{1}{L_t} - 1)$ is negative.

Inertive loading also delays the peak of the volume velocity flow. For the simple model of Fig.3, the time at which U_g is maximum can be calculated by setting the time derivative equal to zero and solving for t . This procedure yields:

$$t \text{ at } U_g \text{ maximum} = \begin{cases} 2 - \left[\frac{L_t + 1}{2} \right]^{\frac{L_t}{1-L_t}} & \text{for } L_t \neq 1 \\ 2 - e^{-1/2} = 1.39 & \text{for } L_t = 1 \end{cases} \quad (5)$$

To show the effect of inertive loading with an admittance waveform that is more rounded near the peak than is

a triangle, Fig.4 gives the flow that would result if Y_g were one-half cycle of a sinusoid. Since we have not derived a closed-form solution for this case, the curves were obtained by means of a digital simulation of the differential equation. With the sinusoidal admittance, the peaks of the flow waveforms become more rounded than for the triangle, especially for small values of L_t . In addition, the peak of the flow waveform is delayed somewhat more, as a function of L_t , than for the triangular admittance function, and the value of L_t at which the derivative at $t = 2$ becomes infinite appears to be somewhat less than unity. Otherwise, the responses for the sinusoidal and triangular cases are very similar. The insensitivity of the general form of the volume velocity to the details of the glottal admittance function with higher values of L_t supports our assumption that a rough approximation to the actual admittance can be informative.

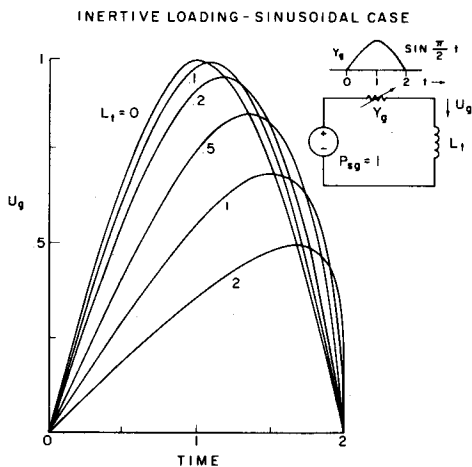


Fig.4. Computed glottal air flow with an inertive supraglottal impedance and sinusoidal variation in glottal admittance.

Another simplification that we have made in modeling the glottal admittance waveform has been to assume that the admittance before and after the glottal pulse is zero. Inverse-filtering studies and motion pictures of the glottis during voicing have shown that there is often a patent

air path between the arytenoid cartilages. This flow path would tend to show up in our model as a fixed impedance (or admittance) in parallel with the time-varying component. In the non-interactive model, the result would be primarily to offset the flow waveform from zero flow, without much change in the shape of the glottal pulse. However, with inertive loading assumed, the added admittance would alter the form of the differential equation and therefore affect the pulse waveshape, primarily near the onset and offset of the glottal pulse. To see in more detail what the effect of such glottal "leakage" might be, we have, in Fig.5, inserted a fixed admittance, Y_{g-l} , of 0.2 times the amplitude of the time-varying component of Y_g , Y_{g-ac} , in parallel with the time varying component. In doing this, we have again assumed a linear model for the glottal admittance and neglected the inertive component. Offsets of this magnitude are commonly observed during inverse-filtering of oral air flow, and may be much larger in pathological voice, or, naturally, during breathy voice. The general result of this glottal leakage is to cause a gradual onset of the glottal flow pulse, and a more gradual offset, with less high frequency energy produced at both locations.

The reduction of high frequency energy at the instant of glottal closure is of special interest because of its strong potential affect on voice quality. Without "leakage," the high frequency energy is produced by the sealed glottis forcing the flow to zero, in opposition to the influence of the inertance, which acts to continue the flow. With leakage, the decrease in Y_g stops when U_g is still decreasing slowly. At that instant the glottal flow pattern changes to a more gradual, exponential decay, with a time constant ($Y_{g-l}L_t$). It may be significant that, according to this simple model, the stronger the source-tract interaction (the higher the value of L_t) the greater is the degradation of the high frequency energy caused by Y_{g-l} . In other words, according to this model, inertive loading of the glottal source will cause an increase in high frequency energy on glottal closure only if there is very little air leakage during the interval of vocal fold closure.

Before turning to the experimental corroboration of acoustic loading effects on the glottal waveform, we should note that the above theoretical development has been for a symmetrical glottal admittance pulse. Though a symmetrical pulse (whether triangular or sinusoidal)

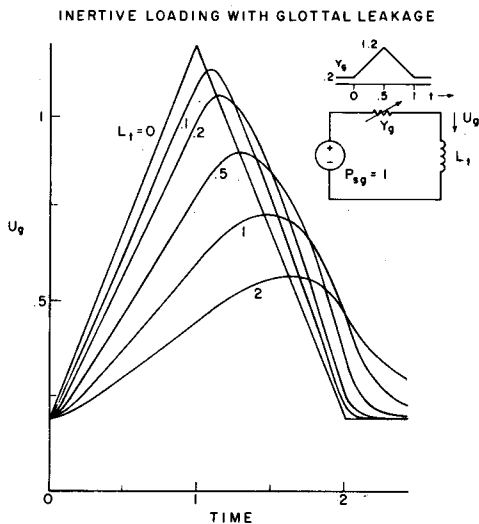


Fig.5. Computed glottal air flow with an inertive supraglottal impedance and a triangular variation in glottal admittance. An incomplete glottal closure is simulated by adding a fixed glottal conductance to the triangular variation.

may be a good first approximation to admittance functions derived from projected area measurements, there may be a good deal of dissymmetry in individual cases. To see how such dissymmetry could interact with loading effects, we solved the differential equation resulting from the triangular admittance function pulse used for Fig.3 (having a total width or duration of two, a peak value of unity, and no "leakage"), but with the peak occurring at some arbitrary time T_m , instead of at $t = 1$. If T_m is between zero and one, the admittance function tilts to the left, and if T_m is between one and two, the tilt is to the right, with $T_m = 1$ being the case illustrated above in Fig.3. The differential equations for U_g would then be:

$$L_t \frac{dU_g}{dt} + \frac{T_m}{t} U_g = 1, \quad \text{for } 0 \leq t \leq T_m \quad (6A)$$

$$L_t \frac{dU_g}{dt} + \frac{2-T_m}{2-t} U_g = 1, \quad \text{for } T_m \leq t \leq 2 \quad (6B)$$

As long as L_t is not exactly equal to unity (since $L_t = 1$ results in an indeterminacy in the solution form) the solution to these equations is

$$U_g = \frac{t}{L_t + T_m} \quad \text{for } 0 \leq L \leq T_m \quad (7A)$$

$$U_g = K(2-t) \frac{2-T_m}{L} + \frac{2-t}{L-2+T_m} \quad \text{for } T_m \leq t \leq 2, \quad (7B)$$

where

$$K = \frac{2L(2-T_m) \frac{T_m-2}{L}}{L^2 + 2(T_m-1)L + T_m^2 - 2T_m} \quad (7C)$$

As for the special case of $T_m = 1$ (a symmetrical pulse), the onset of the pulse is linear, and the offset is steeper than the onset. However, the skew of the admittance pulse is now a function of both L_t and T_m . We illustrate this by showing that the "critical" value of L_t , i.e., the value above which the derivative becomes infinite at $t = 2$, will vary with T_m . The relationship can be derived by first differentiating U_g in the interval $T_m \leq t \leq 2$:

$$\frac{dU_g}{dt} = \frac{K(T_m-2)}{L_t} (2-t) \frac{2-T_m-L_t}{L_t} + \frac{1}{L_t-2+T_m} \quad (8)$$

As t approaches two, this function becomes infinite for values of T_m and L_t that make the exponent in the first term negative. This occurs for values of L_t greater than $(2 - T_m)$. Thus, if the admittance pulse is skewed to the right ($T_m > 1$) the flow termination will become infinite at smaller values of L_t , and vice-versa. For example, if $T_m = 4/3$, for a rather high but possibly attainable opening time to closing time ratio of two, the critical value of L_t is reduced to $2/3$ as compared to the value of unity that holds for the symmetrical admittance function.

THE MEASUREMENT OF SOURCE-TRACT ACOUSTIC INTERACTION

Though an inertive loading of the glottal source would produce flow waveforms very much like those observed by

using standard inverse-filtering techniques, given a symmetrical variation of glottal admittance, it still remains to be shown that this type of glottal-supraglottal interaction is significant in such activities as speech or singing. There are a number of ways that we can study this during actual vocalizations. The most direct method might be to measure Z_t somehow, but this requires the measurement of the pressure just above the glottis, and is difficult to implement. (However, see Dr. Koike's Chapter 14 in this book.)

Using another approach, we have attempted to implement a "nonlinear inverse filter" in which some of the effects of the glottal-supraglottal interaction are removed (Rothenberg and Zahorian, 1977). The resulting waveform, shown in Fig.6, is more directly related to Y_g than is the actual flow U_g . To aid in visualizing the asymmetry of each waveform, the connected lines above the waveform were drawn in to match the maximum slopes of the rising and falling segments, neglecting the oscillations at the frequency of the first formant in the case of the linear filter. It can be seen that in this case the supraglottal impedance did cause an alteration of U_g similar to that produced by inertive loading in our model.

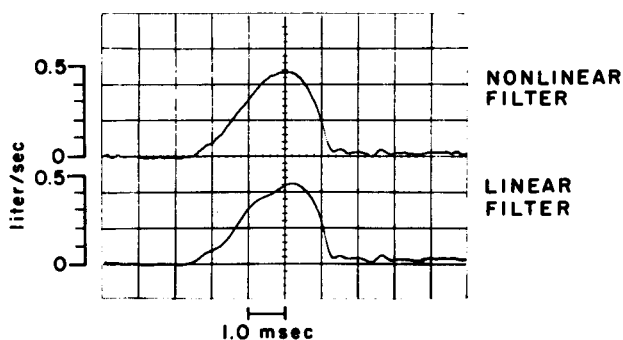


Fig.6. Nonlinear inverse filtering of a spoken vowel [æ]. The top trace shows the output of the time-varying nonlinear inverse filter. The bottom trace shows the output of the time-invariant inverse filter. (From Rothenberg and Zahorian, 1977).

Another approach in measuring the affect of Z_t is to change Z_t while keeping the glottal area function approximately invariant. One way this can be done is by changing the vowel value. In the example in Fig.7, the glottal flow was measured by inverse filtering oral air flow, while vocal fold movements were monitored by simultaneously recording the vocal fold contact area (VFCA) waveform obtained from a Laryngograph (Fourcin, 1974). The VFCA waveform is shown inverted in the figure, since the Inverse VFCA waveform tends to rise and fall with the glottal air flow. The samples shown in the figure were from the center of the vowel in the nonsense syllable /b Vowel p/ in the syllable sequence /b a p b æ p ---/. Each row in the figure represents one such sequence, with the first two sequences (top two rows) spoken slowly enough so that each vowel had a distinct steady-state segment, while the last sequence (bottom row) was spoken at a natural rate. Vocal effort was at a moderate conversational level. The stop consonants help assure a good velopharyngeal closure during the vowel, which is important for accurate inverse filtering. Filter parameters corresponding to the frequency and damping of the first three formants were adjusted manually during a repetitive playback of the vowel sample, using the VFCA waveform as an aid in defining the glottal closed and open periods (Rothenberg, 1979). The low-frequency response limitation of the standard automatic amplitude control and high-pass filtering in the Laryngograph probably caused a slight falling of the Inverse VFCA trace during the glottal open phase, but this distortion should be similar for all samples.

Except perhaps for the first (topmost) repetition of /a/ and the second (middle) repetition of /æ/, a constancy of the VFCA waveform between samples suggests that the differences in the U_g waveform were not caused by different vibratory patterns of the vocal folds. We refer here to the general shape of the waveform and not the amplitude, since the waveform amplitude depends on the larynx position relative to the electrodes, and therefore can vary with the vowel articulation.

If the two samples having a non-representative VFCA waveform are ignored, the figure shows that, of the vowels tested, those with a constriction closer to the glottis and a higher first formant (at the left in the figure) tended to have a glottal flow waveform which was more skewed to the right, with a steeper flow termination, and therefore might be expected to have more high frequency

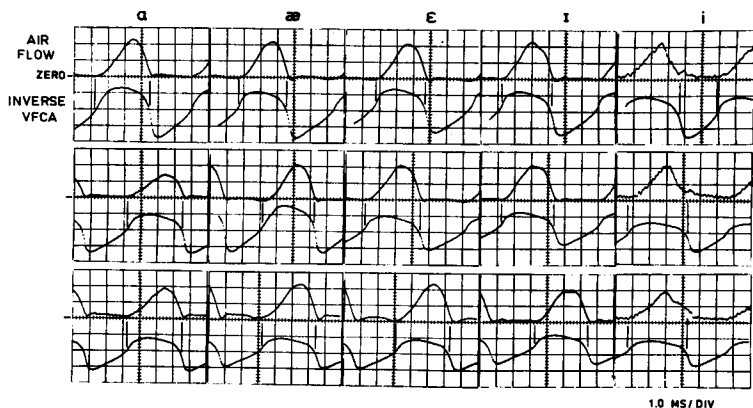


Fig.7. Glottal volume velocity as a function of vowel value for a single male adult speaker. The vertical lines indicate the approximate beginning and end of the primary glottal opening movements, as measured from the vocal fold contact area (VFCA) waveform using the model presented in Rothenberg (1979). Upper curves: Glottal air flow obtained by inverse filtering oral air flow. Lower curves: Inverse vocal fold contact area from a Laryngograph. The different rows represent different repetitions of the vowel sequences. Note: A slight negative overshoot after the glottal closing phase may be due to a damped mask resonance near 1200 Hz. This artifact can be removed with a more complex inverse filter.

energy generated by the termination of the glottal closure. In fact, for the /i/ samples shown, the sharpest discontinuity in waveform slope, and therefore greatest production of high-frequency energy, appears to be at the peak of the waveform, and not at the instant of glottal closure. The overall shapes of the flow waveforms in Fig.7 are roughly similar to those produced by the model in Figs. 3 and 4, as inertive loading is varied, except for the added oscillations at the first formant frequency. The oscillatory component of the interaction is not included in our simple model. Though there was considerable waveform variability between speakers, the general trend toward more dissymmetry with vowels closer to /a/ was also found

in the samples from two other speakers tested, one male and one female.

These results are at least roughly consistent with calculations made from a simple model of the vocal tract. For a lossless supraglottal vocal tract 17 cm long, with a uniform area of 5 cm², the impedance Z_t will be inertive for frequencies below the first formant frequency of 500 Hz (or, more precisely, below a frequency just slightly under the first formant). The magnitude of this impedance will be approximately $8 \tan(\omega/2000)$ in either acoustic ohms or units of cm H₂O per liter/sec. In order to represent the actual vocal tract impedance below F_1 by a pure inertance, it is necessary to find a linear approximation to the actual impedance function. Assuming that in our example the frequencies between 125 Hz and 375 Hz, representing an F_0 of 125 Hz and the second and third harmonics, are of most interest, we can make a linear approximation by estimating L_t as the derivative of $|Z_t|$ with respect to ω at some intermediate frequency, say 250 Hz:

$$L_t \approx \frac{d|Z_t|}{d\omega} = \frac{d}{d\omega} [8 \tan \frac{\omega}{2000}] \text{ at } \omega = 250(2\pi) \quad (6)$$

$$= 0.008 \text{ at } \omega = 250(2\pi)$$

To normalize L_t to the scales used for Figs.2, 3, and 4, it is necessary to multiply by the actual maximum glottal admittance (or conductance, ignoring L_g) and divide by the actual time for one-half the glottal pulse (approximately 2 ms at $F_0 = 125$ Hz). In calculating interactive effects, the differential conductance $dU_g/d(P_{sg} - P_t)$ should be used. For a flow of 500 ml/sec and a maximum glottal area of 0.16 mm², typical adult male values, the maximum glottal conductance would be approximately .05 in cgs units, as computed from $A^2/\rho U_g$ (Flanagan, 1958; Fant, 1960). Defining \bar{L}_t as the value of L_t normalized to the scales used for Figs.2,3, and 4, we obtain:

$$\bar{L}_t \text{ (normalized)} = L_t \text{ (actual)} \times \frac{.05}{.002} = 0.2 \quad (7)$$

Vowels with a higher F_1 would be expected to yield a value of \bar{L}_t that is valid over a wider frequency range, since more harmonics would be included before F_1 is surpassed and the reactance becomes compliant. In addition, vocal tract configurations in which the pharynx is more constricted than for a neutral vowel might be expected to

lead to a higher value of \bar{L}_t . Since the value of \bar{L}_t also depends upon the maximum glottal admittance, a vocal fold vibratory pattern that resulted in a wider than average glottal opening would also increase \bar{L}_t , as would the inertive component of an unusually pronounced constriction at the entrance to the larynx or at the false vocal folds. Thus a value of \bar{L}_t of at least 0.5 for a back vowel such as /a/ in some speakers is not inconceivable. Any glottal inance would add to this figure.

Finally, we have also tested for the presence of supraglottal loading effects by comparing vocalizations made with air and with a large proportion of helium mixed with the air. By reducing the acoustic inance in the vocal tract, the helium would be expected to reduce any supraglottal loading effect, if present. (Though, unlike the change of vowel value, the use of inspired helium also affects the glottal inance. To affect only the acoustic loading, it would be necessary to introduce the helium only into the supraglottal pathway.) Figure 8 shows the general result we obtain. Both waveforms in the figure were obtained by inverse filtering the oral volume velocity for a male adult holding a vocal tract position for an /æ/ vowel. The fundamental frequency in each case was about 110 Hz. It can be seen that the symmetry of the waveform increases significantly with helium displacing some of the air.

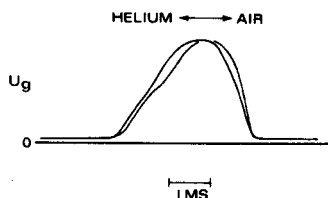


Fig.8. Glottal waveforms obtained by inverse filtering glottal air flow with air and with a large proportion of helium mixed with the air.

CONCLUSIONS

It appears that the often-noted skewing of the glottal flow waveform which results in the primary vocal tract

excitation being at the instant of glottal closure can be caused by a combination of factors, including dissymmetry in the glottal admittance function, air displaced by vocal fold motion, and acoustic interaction with the supraglottal impedance. The inertance of the glottal slit may also be a factor, at least at the smaller glottal openings, when this inertance is largest.

To study the effect of acoustic interaction, we have defined a normalized supraglottal inertance \bar{L}_t which approximates the actual loading of the glottis at frequencies between F_0 and F_1 , assuming F_1 is appreciably larger than F_0 . This model for the supraglottal loading indicates that for a symmetrical triangular admittance pulse the effect of acoustic interaction becomes large if \bar{L}_t is close to or exceeds unity, since at $\bar{L}_t = 1$ the derivative of the glottal flow becomes infinite at the instant of glottal closure. This critical value of \bar{L}_t will vary with the symmetry and general shape of the admittance pulse, but appears to remain in a range of roughly 0.5 to 2.0 for the type of admittance variation to be found in normal speech. The model also indicates that at higher values of \bar{L}_t the volume velocity waveform tends to have a characteristic shape which is rather insensitive to the details of the glottal admittance function.

A rough analysis of the magnitude of the supraglottal vocal tract impedance shows that values of \bar{L}_t of at least 0.5 are not implausible, with the value in any given case being a function of the vowel value, the maximum admittance of the glottis during the glottal cycle and the configuration of the larynx above the vocal folds.

A simple representation of the glottal admittance function that would result from an incomplete vocal fold closure during the glottal cycle indicates that this factor is also important, in that it can appreciably modify the effect of source-tract acoustic interaction. However, small leakage paths may not be as significant as our results in Fig.5 suggest, since we have neglected the glottal inertance, a factor that could be relatively large with a small opening.

Our results demonstrate that acoustic interaction can cause the glottal source waveform to vary widely as a function of vowel value and F_0 , since the first formant must be high compared to F_0 in order for the supraglottal loading to be inertive at an appreciable number of glottal harmonics, and because the magnitude of the impedance below F_1 varies as a function of vowel value. It is also

possible that different pronunciations (allophonic variations) of the same vowel phoneme can result in appreciably different source spectra because of differences in the impedance seen by the glottis. The degree to which this phenomenon can explain differences in voice quality in speech and singing would be an interesting subject for future research.

ACKNOWLEDGMENTS

The work described here was supported by research grant NS08919 from the National Institutes of Health. James T. Mahshie assisted in the experimental work, and the author was fortunate to have the help of Wilbur R. LePage in the solution of the differential equations.

REFERENCES

- Fant, G. (1960) Acoustic theory of speech production. s-Gravenhage: Mouton.
- Fant, G. (1979) Glottal source and excitation analysis. Royal Inst. of Tech. (Stockholm): Speech Trans. Lab., Quart. Prog. and Stat. Rep. 1/1979, 85-105.
- Flanagan, J.L. (1958) Some properties of the glottal sound source. Journ. Speech and Hearing Res. 1, 99-116.
- Flanagan, J.L. (1968) Source-system interaction in the vocal tract. Ann. N.Y. Acad. Sci. (Sound Production in Man), 155, 9-17.
- Flanagan, J.L. and Ishizaka, K. (1978) Computer model to characterize the air volume displaced by the vibrating vocal cords. J. Acoust. Soc. Am. 63, 1559-1565.
- Fourcin, A.J. (1974) Laryngographic examination of vocal fold vibration. In Ventilatory and phonatory control mechanisms, B. Wyke (ed.) Oxford: Oxford University Press, 315-333.
- Holmes, J.N. (1962) An investigation of the volume velocity waveform at the larynx during speech by means of an inverse filter. In Proc. IV Int. Congr. Acoust,

Copenhagen, Denmark, Aug. 1962.

Miller, R.L. (1959) Nature of the vocal cord wave. J. Acoust. Soc. Am. 31, 667-679.

Rothenberg, M. (1973) A new inverse-filtering technique for deriving the glottal airflow waveform during voicing. J. Acoust. Soc. Am. 53, 1632-1645.

Rothenberg, M. (1979) Some relations between glottal air flow and vocal fold contact area. In National Institutes of Health, Proceedings of the Conference on the Assessment of Vocal Pathology (to be published).

Rothenberg, M. and Zahorian, S. (1977) Nonlinear inverse filtering technique for estimating the glottal area waveform. J. Acoust. Soc. Am. 61, 1063-1071.