Some Relations Between Glottal Air Flow and Vocal Fold Contact Area

Two variables relating to the measurement of glottal function during speech that can be recorded by relatively noninvasive techniques are the air flow at the glottis and the relative area of vocal fold contact. Though these variables are obviously related to each other when the supra-glottal vocal tract is open (less vocal fold contact area would generally correlate with more air flow), they emphasize different aspects of the vocal fold movements and their effects, and so can be considered to be complementary to a great degree, The glottal air flow primarily reflects vocal fold movements when the glottis is open, while the vocal fold contact area (VFCA) yields more information during the period of glottal closure, In this paper we look at some details of the correlation between these two variables, and some ways in which one variable can be helpful in interpreting the other.

Glottal Air Flow

Though the air flow waveform at the vocal folds is extremely difficult to measure directly during speech, it can be obtained from the air flow waveform at the mouth, or oral air flow, for brevity, by means of an 'inverse-filter', which removes the effect of the supra-glottal acoustic system (Rothenberg, 1977), Though the pressure waveform at the mouth can also be used for inverse-filtering, it does not supply an adequate representation of the low frequency components, including the baseline or zero air flow level. Therefore, we have generally been using the oral air flow waveform, as recorded by a specially-constructed pneumotachograph mask.

However, whether inverse-filtering air flow or pressure, the primary problem in this method is the proper setting of the inverse-filter parameters. For a non-nasalized or slightly nasalized vowel, the vocal tract configuration most amenable to inverse-filtering; the parameters to be set are the frequency and damping of the complex zeros (antiresonances) that cancel the complex poles of the lowest two or three resonances of the supra-glottal vocal tract or formants. The problem inherent in the setting of these parameters stems from the tact that the proper settings should match the formants with the glottis closed and not the formants that actually exist during the glottal cycle. With a normal voiced glottal cycle there is a significantly long period in which the glottis is either closed, or sufficiently closed so that the glottal impedance is high enough to satisfy this condition. Thus, the inverse-filter parameters could be set to match the vocal tract resonances during this period. Any procedure for inverse-filtering which averages over the entire glottal cycle will therefore be subject to some error, especially in the damping, and to a lesser extent the frequency of the first formant.

To avoid this problem, we have been adjusting the inverse-filter parameters by observing the inverse-filtered waveform during a repetitive playback of a few glottal cycles, and adjusting to

minimize or remove any oscillations at the formant frequencies that occur during the relatively flat portion of the waveform at or near zero air flow that corresponds to the most closed portion of the glottal cycle. This procedure has been used by a number of other investigators, and works well as long as the frequency of the first formant (F_1) is at least four or five times as large as the fundamental frequency (F_o). Thus, for higher values of F_o, as might be common in singing or in some speech styles, or for vowels having a lower than average value of F_1, there are often more than one set of parameter values that can result in a relatively flat segment of the inverse-filtered waveform at or near zero air flow. Of course, only one of these parameter sets will result in the correct glottal flow waveform. In this paper we show how information about the period of glottal closure, as obtained from the vocal fold contact area waveform, can be used to resolve such ambiguities, and greatly extend the usefulness of the inverse-filtering technique.

Vocal Fold Contact Area

Variations in vocal fold contact can be monitored by measuring the transverse electrical impedance through the tissues of the neck at the level of the vocal folds. In this method, the impedance between two surface electrodes, positioned on either side of the thyroid cartilage, is measured by means of a small electrical cur rent passed between the electrodes. A relatively high frequency is usually used, in order to keep the impedance between the contactors and the subcutaneous tissue low without the use of a special conductive paste. The unit we have been using for the measurement of trans- verse electrical impedance is called a Laryngograph, by the manufacturer, and operates at about three megahertz (Fourcin, 1974).

The primary limitation in this type of VFCA monitor is the large amount of noise that can be present in the resulting signal. This noise varies greatly between speakers, and is generally least with adult male subjects in which the thyroid cartilage is prominent and easily encompassed between the two electrodes. With subjects for which the signal is small, there is a broad-band noise originating in the electronics. However, with all subjects there is some low-frequency noise due to extraneous components added by movements of the larynx and other nearby structures. Unless care is taken in filtering out such low-frequency noise, the filtering can greatly distort the VFCA waveform during the glottal cycle. Commercial analog high-pass filters can cause significant phase distortion at frequencies over 10 times the cutoff frequency. Linear phase high-pass filtering, usually accomplished digitally, can reduce this distortion. However, if the signal is very weak with respect to the noise, some such distortion becomes unavoidable. Also, noise that is multiplicative rather than additive, as might be found when the vertical movements of the larynx in and out of the field of the monitoring electrodes, cannot be removed by ordinary linear filtering.

PROCEDURE

Data Collection

As shown in the system diagram in Figure 1, the waveforms in this paper were obtained by recording simultaneously on the FM tape recorder the oral airflow signal from a

circumferentially-vented pneumotachograph mask and the output of a modified Laryngograph. The mask covered only the mouth (and not the nose) and was mounted in the wall of a cubic enclosure 2 feet on each side, so that the subject spoke into the box through the mask. This enclosure was vented to the outside air, and was sound absorbent enough to not significantly affect the signals picked up by the mask. The box was built for another experiment, and not strictly required for these tests, however, it was used because a thermostatically-controlled heater inside the box kept the mask transducer near body temperature. This greatly reduced the drift that occurs when exhaled air changes the temperature of the diaphragm of the transducer. The only negative effect of the box was to muffle the auditory feedback to the talker. However, vocalizations could be monitored afterward with better fidelity by replaying the output of a microphone located within the box. This microphone signal was recorded on a third track of the tape. The Laryngograph used was the basic oscillator-detector unit that is found as an integral part of all of the Laryngograph analyzers now marketed. The unit contains two mechanisms for reducing low-frequency noise and drift that tend to distort the VFCA waveform, and therefore were partially bypassed at different points in the data collection. One such mechanism is an automatic gain control (AGC) feature in which the short time averaged amplitude of the detector output is fed back to the oscillator circuit to reduce the oscillator amplitude. Though this feature effectively equalizes the unit's output amplitude over a wide range of speakers and electrode placements, and greatly reduces low frequency drift problems, it can cause some distortion of the voicing waveform at low voice fundamental frequencies if the averaging time constant in the feedback circuit is not long enough.
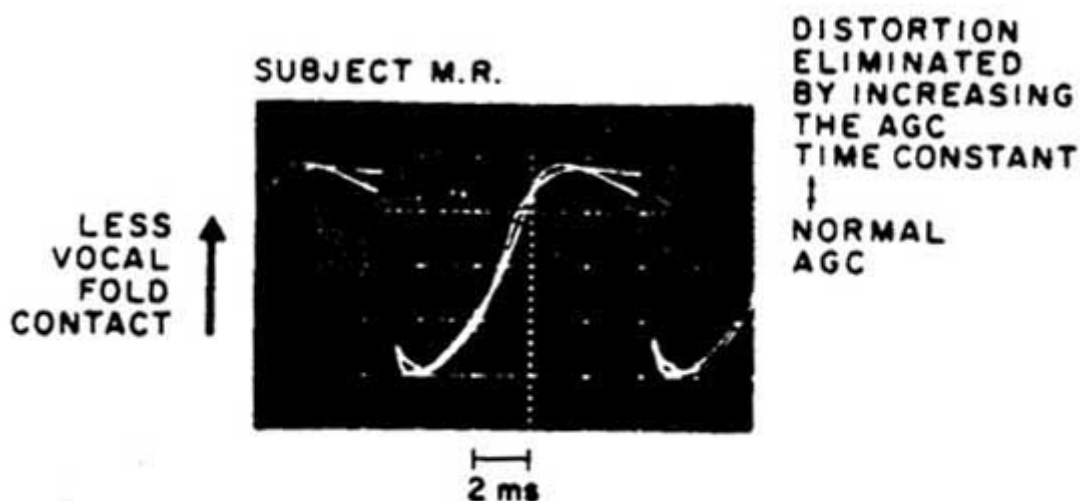

FIGURE 1. Distortion of the Laryngograph waveform caused by the AGC circuit. Distortion due to the coupling time constant was eliminated in both traces by increasing the coupling time constant.

The distortion obtained at F_o levels in the range of an adult male speaker is illustrated in Figure 2. In this figure, as well as in those presented below, the VFCA waveform is shown with an increase representing less vocal fold contact or a more open glottis, and is therefore referred to as the inverse VFCA waveform when describing waveform features. We have found that this

polarity facilitates comparison between vocal fold contact area and glottal air flow. The two VFCA waveforms superimposed in the figure were obtained by retriggering a storage oscilloscope during the same continuous vocalization, with the time constant in the AGC loop increased by a factor of 200 in the upper waveform. This increase in time constant was found to be more than enough to eliminate the distortion at fundamental frequencies as low as 50 Hz. Because of the nonlinear action of the AGC circuit, the amount of distortion cannot be predicted directly from the time constant used for the AGC control signal; the distortion must be determined experimentally. With normal AGC, the distortion consisted primarily of the decrease that occurs during the long flat portion of waveform that corresponds to the open glottal phase.
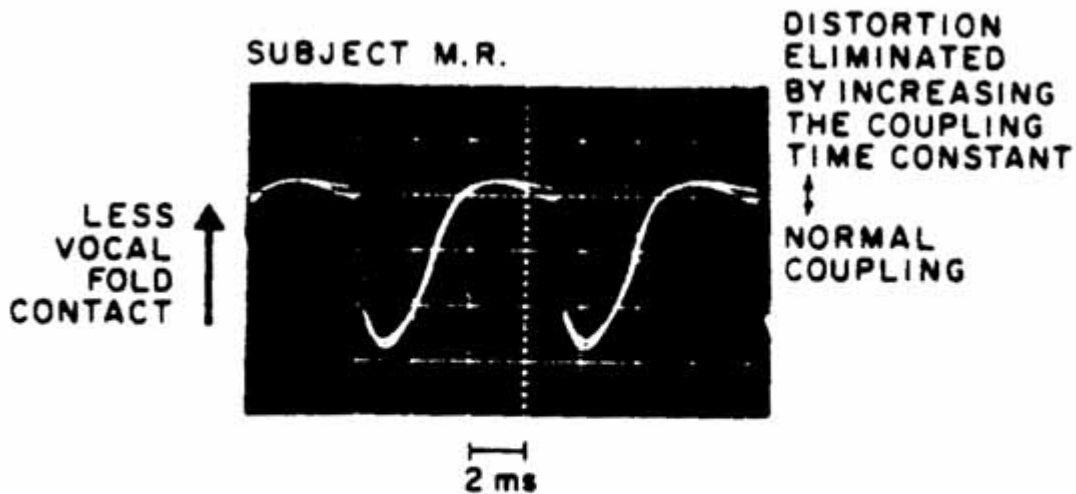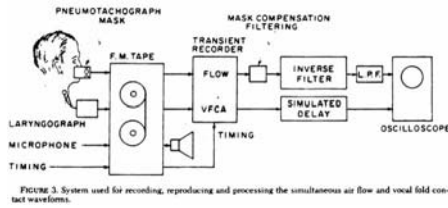

FIGURE 2. Distortion of the Laryngograph waveform caused by coupling time constant. Distortion due to the AGC action was eliminated in both traces by increasing the AGC time constant.

The second feature of the Laryngograph unit that was partially bypassed was the high-pass roll-off (6 dB/octave) due to the coupling capacitor in the final amplifier. Though this roll-off further reduces drift and low-frequency noise, it causes the waveform distortion shown in Figure 3. The distortion was essentially removed in the upper trace by increasing the coupling time constant from 4 ms to 40 ms.

It should be noted that with this speaker an adult male supplying a strong signal, it was possible to record the VFCA waveform during a single normal glottal cycle quite accurately by modifying the Laryngograph circuit. However, even for this subject the overall pattern in vocal fold contact during an abductory or adductory gesture, including the variation in the base line or zero level, could not be obtained nearly as accurately, since the two analyzer time constants described above would have to be increased to such a degree to accomplish this, that the low-frequency noise and drift would make the performance very erratic.

Though not as significant as the low-frequency modification, the high frequency roll-off at 3.3 kHz that was built into the final amplifier in our unit was extended to 6 kHz by another

modification of the circuit.

FIGURE 3. System used for recording, reproducing and processing the simultaneous air flow and vocal fold contact waveforms.

Finally, 2 Hz and 20 Hz timing signals (pulse trains) were also included on the FM tape to be used in locating any desired segment by means of an electronic pre-set counter.

Data Analysis

To produce simultaneous glottal flow and VFCA waveforms from the tape recorded data, a 40 ms segment was recorded on a transient recorder for repetitive playback (see Figure 1). Both the transient recorder and the FM recorder had a response flat to almost 5 kHz on each channel. During the repetitive playback, the air flow signal was processed by an analog inverse-filter of the type described previously (Rothenberg. 1977) having frequency and damping adjustments for F_1 F_2, and F_3, and a linear-phase low pass filter to partially compensate for formants above the third, For an adult male speaker, the low pass compensation for higher order formants should be -3 dB at about 1050 Hz.

The low pass filtering in our system was formed by a combination of an eight-pole Bessel filter, -3 dB at 1300 Hz. a six-pole Butterworth filter. -3 dB at 2500 Hz, a four-pole Bessel filter, -3 dB at 3200 Hz, and a number of real poles at frequencies above 5 kHz that were introduced by the inverse-filter stages for F_1, F_2, and F_3. The net low pass filtering produced by this system approximated a Bessel response of high order and was -3 dB at about 875 Hz and -6 dB at 1200 Hz. This total filter could be looked at as comprising a compensation filter for higher order formants, -3 dB at about 1050 Hz, and an additional linear-phase low pass filter that served to attenuate signal components outside of the range of mask fidelity. This second filter would be -3 dB at roughly 2 kHz. Low pass filter frequencies (except for the fixed real poles) were raised about 20% for the female speaker. The overall system response time in the flow channel, as limited by the low pass filtering, was roughly .2 ms.

The mask compensation filter shown in Figure 1 consists of three components. The most significant of these is the simple one-pole RC low pass filter that we have shown will compensate for the attenuation of the pressure outside of the pneumotachograph mask before it reaches the rear of the pressure transducer diaphragm (Rothenberg, 1977). The time constant we used for this filter was .2 ms.

A second component of the mask compensation filter was a 3500 Hz antiresonance of the same type used for the formant inverse-filter stages. This filter compensated for the resonance of the diaphragm of the mask transducer, which was near 3500 Hz in our mask.

The last component of the mask compensation filter is more difficult to explain because we have

not clearly identified the effect for which it compensates. In our previous attempts to inverse-filter oral volume velocity we have sometimes noted some apparent distortion of the waveform when the frequency response of the system was extended much beyond 1000 Hz. This distortion would often occur as a brief (± .5 ms) "overshoot" after the glottal closing phase, or as a damped oscillation, similarly located, and was found to be due to a moderately damped resonance at about 1250 Hz that was added to the normal formant pattern by our measurement system. This extra resonance appears to be an acoustic affect added by some portion of the pneumotachograph mask, since it was not traceable to the pressure transducer or electronics, and could be increased in frequency by introducing helium into the mask. In the waveforms reported below, this resonance was removed by an additional (fifth) antiresonance circuit. During the inverse-filter adjustment, this filter was set initially at 1250 Hz with moderate damping. However, the settings for frequency and damping were re-touched slightly when this would improve the natural-ness of the "closed" portion of the waveform. The Laryngograph signal was smoothed only by a four-pole Bessel low pass filter, -3 dB at 3200 Hz. and had a rise time of about .1 ms. The simulated delay shown in Figure 3 was selected to match the delay in the air flow channel caused by the obligatory low-pass filter action of each inverse-filter stage, the low-pass filter action of the mask compensation, and the three additional low-pass filters described above. Alternatively, the minimum delay that must be introduced by the inverse-filter stages can be considered equivalent to the glottis-to-mask transmission delay in the vocal tract. The compensatory delay in the VFCA channel was not effected by an actual time delay, but by electronically shifting the VFCA waveform on the oscilloscope screen by the equivalent distance. Since the accuracy of this compensatory delay is important to the interpretation of the VFCA waveforms, the computed delay was verified by measuring the delay of the system elements. Both computations and measurements yielded a compensatory delay of 1.05 ms ± .0.5 ms. Finally, this value of delay was tested by recording short glottal pulses on the tape and Comparing the VFCA waveform with the inverse-filtered flow. The pulses were obtained by producing an ingressive low-frequency voicing with a tightly closed glottis. It was found that pulses widths of as little as 1 ms could be produced in this fashion.

One such pulse is shown in Figure 4. The VFCA waveform is shown "delayed" by 1.05 msec. It can be seen that there is a close correlation between the onset of the pulse (the first increase in air flow) and an increase in the slope of the inverse-VFCA waveform. The precise timing of these waveform features is discussed further in the results.

In the inverse VFCA waveform shown in Figure 4, a large part of the exponential rise to a neutral value during the long period of glottal closure is due to the action of the AGC circuit of the Laryngograph, which was left on to improve the signal to noise ratio with the small VFCA signal obtained in this type of laryngeal maneuver.

After observing a number of ingressive-pulse and nonnal-voicing waveforms, as well as from the tolerance limits in our computations and from measurements of the delay in the flow channel, we estimate that the time synchronization between the glottal flow and VFCA waveforms is better than .2 ms, and probably within .1 ms.

The general shape of the response to a short pulse shown in Figure 4 also verifies that our air flow system response time is about .2 msec. The system was consistently able to show a pulse

rise or decay that occurred in little more than that value. The fine ripple following the pulse is the remanent F_4, that was highly attenuated, but not eliminated, by our system.
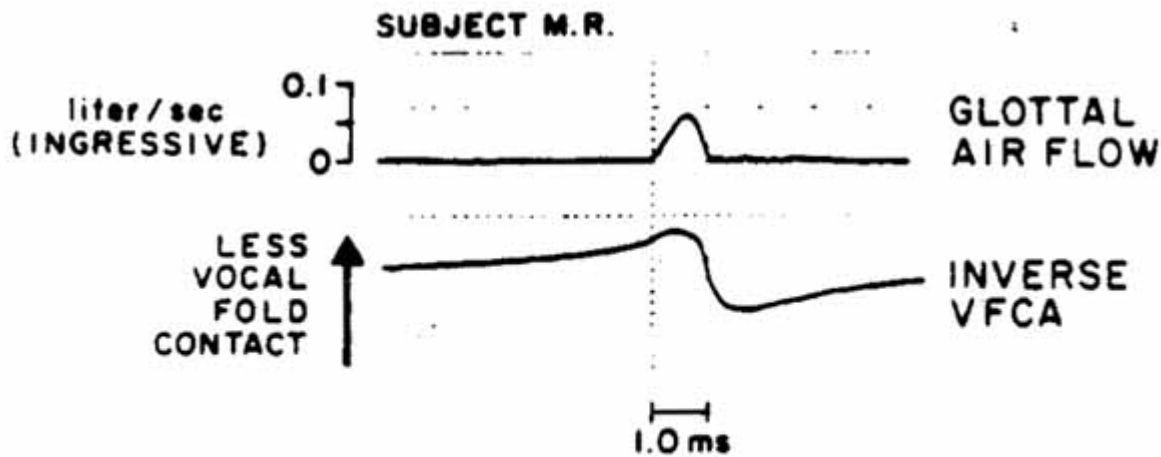

FIGURE 4. Test of the compensatory delay used for the VFCA waveform, using short, ingressive glottal pulses. Vocal tract shape was similar to that for [ɑ].

RESULTS

The VFCA Waveform During Normal Voicing

When the distortion produced by the AGC circuit and the high pass filtering is removed, the VFCA waveform produced by the Laryngograph during normal voicing tends to have a relatively flat portion roughly corresponding to the period in which the vocal folds are open. In the glottal air flow waveform, this period corresponds to the duration of what is sometimes referred to as the "glottal pulse". In this paper we will use a definition of the glottal pulse based on the air flow waveform, namely, the period form the first sign of an increase in air flow associated with the glottal opening phase to the instant at which the negative slope during the closing phase is interrupted and a period of near zero slope begins. This definition of the glottal pulse is illustrated in the example of Figure 5, which was very typical of waveforms noted for three speakers tested, two males and one female, during a variety of vowels. The vowel here is /ae/, from the nonsense syllable /b ae/. The /b/ in this test syllable was used to help keep a good velopharyngeal closure during the vowel. It also provided a zero air flow reference before each syllable. Since there was a small amount of drift and low frequency noise, and a flow component due to articulator movement, the zero level should be considered accurate to only about ± 20 milliliters/second. For this sample, both time constants of the Laryngograph unit were increased enough to eliminate waveform distortion. The lower photograph in the figure is an expanded version of one pulse in the upper photograph, to help in identifying the features of the glottal closing phase.
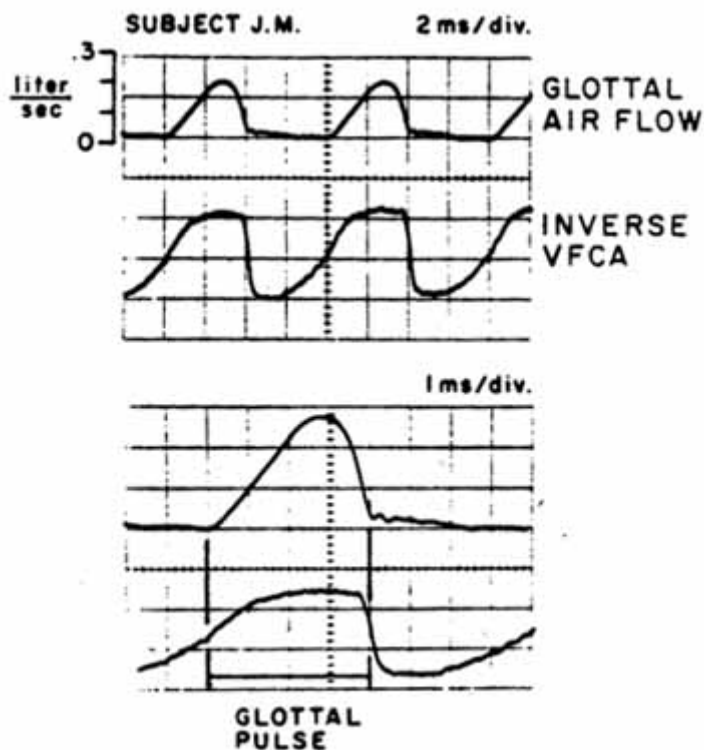
FIGURE 5. Glottal air flow and vocal fold contact area during /æ/ in the syllable /b æ/.

Shown in Figure 6 are idealized versions of the glottal air flow and the inverse VFCA waveforms, with our interpretation of the features of the VFCA waveform during normal. non-breathy vocalic speech. from the samples we have observed so far. The figure shows how these features appear to be related to the glottal pulse and to the sequence of physiological events that comprise the glottal cycle. This interpretation is in general agreement with the observations of others made from a frame-by-frame analysis of simultaneous motion pictures of the glottis (e.g., Fourcin, 1974, and Lecluse, Brocaar, & Verschuure 1975).

As illustrated in the figure, the termination of the glottis pulse is typically accompanied by the onset of a sharp drop in the inverse VFCA waveform, presumably due to the vocal folds coming into contact. During normal voicing, the vocal folds are usually observed to first come into contact near their lower margins, i.e., closest to the trachea (time t1), and then quickly close over the rest of their area (from t_7 to t_1). Thus the end of the glottal pulse comes near the beginning of the sharp drop in the inverse VFCA waveform, at the closing of the lower mar- gins of the vocal folds (t_1). This part of the VFCA waveform, segment t_7 to t_1, is the most invariant feature. Most other features to be described can vary markedly in distinctiveness between speakers or even within samples from the same speaker.
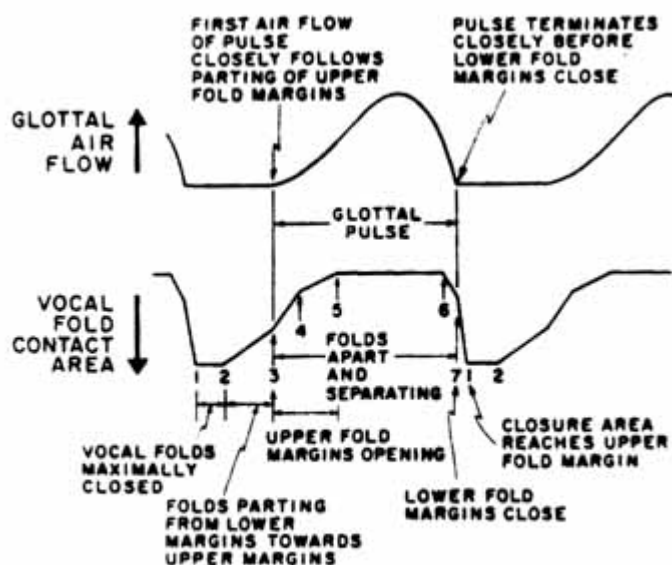
FIGURE 6. Idealized VFCA waveform during normal, non-breathy voicing, and its relation to the glottal air flow pulse.

More accurately referred to as the "most-closed" portion of the cycle, since there may still be a significant air flow indicating that the glottis is not completely sealed, During the "closed" portion of the glottal cycle, t_1 to t_3, the glottal air flow waveform is rather flat, at or near its minimum flow during the cycle, In the inverse VFCA waveform produced by a Laryngograph there may be a relatively flat portion, t_1 to t_2, initiating the closed period during which the vocal folds are being compressed without much change in (contact area. However, during most of this part of the glottal cycle, the waveform rises continuously (t_2 to t_3). This rise of the inverse VFCA waveform during the closed portion of the glottal cycle appears to be at least partially due to the slow separation of the lower margins of the vocal folds as they roll apart from below. However, it must be kept in mind that at the high electrical frequency used in the Laryngograph unit, the electrical impedance might be affected to some extent by the capacitance between folds when they are very close over a wide area, though not quite touching. It might be interesting in this respect to vary the drive frequency over a wide range, keeping the other conditions constant, to see the effect this has on the waveform. A previous attempt to do this (Lecluse, et al, 1975) compared devices that were somewhat different from each other, and so the results cannot be interpreted unambiguously.

The instant at which there is the beginning of a rise in air flow signaling the start of the next glottal pulse can usually be correlated with a discontinuity in the slope of the inverse VFCA waveform as it rises (t_3). At this instant the inverse VFCA waveform will begin to rise more quickly as the upper fold margins separate. The slower change in VFCA between t_4 and t_5 is probably due to phase differences along the length of the vocal folds. In other words, our present hypothesis is that when segment t_4 - t_5 is well-defined, the upper margins of the vocal folds begin to separate rather suddenly at one region (from time t3 to t4) and then proceed to separate more gradually along the rest of their length. During the period t_8 to t_7 we assume that the inverse process is occurring as the bottom margins of the vocal folds begin to approximate. The period

between t_5 and t_8 is associated with fully parted vocal folds. Though the distance between the vocal folds is varying in this interval, there is little change in contact area.

The model in Figure 6 describes our observations for what we have referred to as normal, non-breathy voicing during vocalic speech. It would not necessarily apply to "voicing produced with other laryngeal adjustments, such as falsetto or creaky voice. In breathy voicing, of the type produced by a medial abduction of the vocal folds, the period of glottal closure decreases, and the various distinctions we make during the period of glottal closure become progressively less identifiable as the vocal folds are abducted. On the other hand, period t_5 - t_6, associated with fully parted vocal folds, becomes progressively larger and better defined.*

Figure 7 and Figure 8 show the glottal air flow and inverse VFCA waveforms for typical productions by the two other subjects tested. The VFCA waveform in Figure 7 was obtained with the AGC circuit unmodified, in order to reduce the noise in the waveform. notice that the normally flat portion of the waveform, t_5 to t_8 in our model, shows a decay that is presumably due to the AGC action, and should therefore be neglected. If one ignores this decay, the limits of the glottal pulse, t_3 and t_7 are easily identifiable in the VFCA waveform, and correlate with the glottal flow waveform predicted by the model.

In Figure 8, however, the VFCA waveform is not as easily interpreted, because of the high level of noise and distortion. For this subject, a 50-year-old trained female singer, with considerable subcutaneous tissue surrounding the larynx, it was necessary to press the Laryngograph electrodes deeply into the neck to get a useable trace. Even then. the low-frequency noise was so high that it was necessary to add a one-pole high-pass filter between 10 Hz and 100 Hz. in order to keep the waveform in the range of our recording instruments. Unfortunately, the precise frequency of the filter used for the trace in Figure 8 was not recorded. This high-pass filter would cause a considerable amount of distortion, especially the decay that can be seen between t_4 and t_8. However. the termination of the glottal pulse, t_7, can be clearly identified. and this viewer believes there to be a tendency for an increase in slope after the onset of the glottal pulse (t_3). A better estimate of the VFCA waveform could be obtained if transient averaging was used to replace visual "averaging", with the glottal air flow signal used as the timing signal for the averager.

*The waveform during breathy voice and creaky voice is described in more detail by A. Fourcin in his contribution to these proceedings.
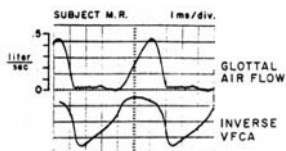

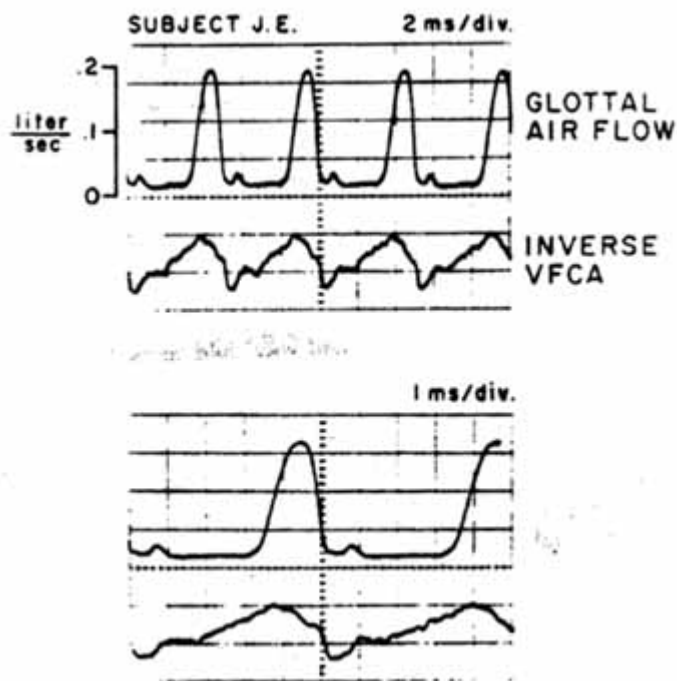FIGURE 7. Glottal air flow and vocal fold contact area during /æ/ in the syllable /b æ/.

FIGURE 8. Glottal air flow and vocal fold contact area during /æ/ in the syllable /b æ/.

The Use of VFCA in Inverse-Filtering

As discussed earlier, the accurate inverse-filtering of oral pressure or flow requires an approximate identification of the interval of glottal closure. We have found that the VFCA waveform can be very helpful in this regard and can extend the inverse-filtering procedure to a much larger range of voice qualities, F_o values, and vowel types.

Since the ease of inverse-filtering depends on a high ratio of F_1 to F_o, the VFCA waveform can be expected to be helpful when F_o is high or F_1 is low, As an example, Figure 9 shows the inverse VFCA waveform (bottom) and the inverse-filtered oral air flow (top two traces) for a number of glottal cycles during the vowel /i/ from the nonsense syllable /b/. The speaker was an adult male. and the fundamental frequency about 130 Hz. Since F_1 was slightly under 400 Hz during this vowel. the ratio of F_1 to F_o was only about three. Two settings of the inverse-filter parameters, shown in traces A and B, yielded a plausible flat interval near zero flow that could represent a period of glottal closure. However. from the VFCA waveform we can clearly see that only waveform B could be close to the actual glottal flow, since the closure interval in waveform A is too far displaced from the closure interval indicated by the VFCA wave, as identified by the double headed arrow in the figure. This closure interval was taken from the VFCA waveform using the analysis of Figure 6. The end of the closure interval. i.e., the beginning of the glottal pulse, is clearly indicated by the sudden increase in the slope of inverse VFCA wave, while the end of the glottal pulse is indicated by the onset of the rapid decline in the waveform.

The accuracy of the choice of waveform B is also verified by the formant values used in this adjustment, namely 390 Hz. 2100 Hz and 2500 Hz. They were within the general range of the adult male values for /i/ reported by Peterson and Barney (1952) and others. (They were actually closer to male /i/ values. however this was probably due to our sample being taken from the end of the transition from the consonant.) Waveform A, however, had an "extra" antiresonance at 960 Hz that would not normally be associated with and /i/ vowel, and an FI that was somewhat high for /i/ at 560 Hz.
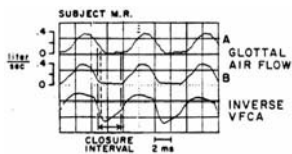

FIGURE 9. Example of the use of the VFCA waveform for corroborating a glottal air flow waveform obtained by inverse-filtering oral air flow or pressure. Vowel is /i/.

Note that both the VFCA wave and the correct estimate of glottal air flow (trace B) show signs of a second, less rapid and smaller, interval of glottal closure starting at the dashed line. In this way, the VFCA waveform can help verity that perturbations in the air flow waveform during the period of glottal closure are actually due to glottal activity and are not just an artifact of the inverse-filter procedure. Another example of this can be seen in Figure 8, where a small air flow pulse occurring during the beginning of the glottal period is roughly correlated with a perturbation in the VFCA waveform.

CONCLUSIONS

Simultaneous recordings of the glottal airflow, obtained by inverse-filtering oral air flow, and the vocal fold contact area, as derived from the transverse electrical conductance of the larynx, have suggested a seven stage model for the vocal fold contact area waveform in voiced speech. These stages have been given a physiological interpretation that agrees with the air flow waveform in the samples tested to date and with published descriptions of vocal fold action during normal chest voice. However, they are presented as a basis for future discussion rather than a final model, since their usefulness will need to be tested in applications involving many speakers, both normal and pathological. Since not all significant features of the vocal fold movements are indicated unambiguously by the two measurements we have been using, further corroboration by means of high-speed or stroboscopic motion pictures or x-ray films, or using computer simulations such as those presented by Titze at this conference would be helpful. It would also be desirable to study the applicability of the model to other voice qualities, including the affect of vocal fold adduction and abduction, and to determine the effect on the VFCA waveform of variations in transglottal pressure.

For most speakers, the VFCA waveform obtained from a laryngograph is sufficiently strong for the measurement of voice periodicity. However, for many speakers the signal is too weak to permit a detailed waveform analysis. To better define the range of clinical usefulness, it would be desirable to obtain an estimate of the range of speakers for which the signal is strong enough for the observation of waveform features such as those discussed here.

We have also shown that the VFCA waveform, by providing an estimate of the period of glottal closure, can be an aid in the manual inverse filtering of oral air flow or pressure. It remains to be seen whether the VFCA waveform, recorded simultaneously with either air flow or pressure, can be useful as part of an algorithm for high quality, automated inverse filtering.

ACKNOWLEDGMENTS

REFERENCES

FOURCIN, A.J. *Laryngograph examination of the vocal fold vibration, ventilation, and phonation; control mechanisms*, In B. Wyke (Ed.), London: Oxford University Press, 1974.

LECLUSE, F.L.E., BROCAAR, M.P., & VERSCHUURE, J. The electroglottography and its relation to glottal activity, *Folia Phoniatrica* 1975, 27, 215-244.

PETERSON, G.E., & BARNEY, H.L. Control methods used in a study of the vowels, *Journal of the Acoustical Society of America*, 1952, 24, 175-184.

ROTHENBERG, M. Measurement of air flow in speech. *Journal of Speech and Hearing Research* 1977, 20, 155-176.

ROTHENBERG DISCUSSION

Dr. Titze: The concept of inverse filtering has to be viewed in a different light when we're discussing high effort phonation. Take, for example, the waveforms you presented on the male singer. When using inverse filtering techniques, it is assumed that there is a closed glottis with a rigid boundary. However, you do not have this condition in high effort phonation. In such cases, the upper laryngeal cavity couples tightly with the vocal folds, and, according to Sundberg,[1] not as well with the remaining tract. A cavity resonance is generated right in the upper larynx. Also, when pulmonary pressure as high as 50 cm of water are produced, tissue strains on the order of 100% are generated in the mucosal tissues. This means that a half a millimeter mucosal layer will vary in thickness between one quarter and 1 mm. Thus, there is no such thing as a fixed boundary in this situation, rather, it is a yielding one, making it next to impossible to separate sound from source. Those wiggles that occur after closing might well be a result of tissue deformations. It would be a mistake to try to filter them out. Thus the assumption used in inverse filtering may not apply during high phonatory effort.

I don't know if you're aware of Sundberg's approach, but it seems to contain a paradox; he claims a great degree of cavity resonance in the upper larynx while at the same time using a linear source-system approach for synthesis of sung vowels.

[1. Sundberg, I. An articulatory interpretation of the singing formant. STL-QPSR, Royal Institute of Technology,

Dr. Rothenberg: The oral presentation included a brief review of some results obtained by inverse-filtering oral air flow. One previously unpublished slide (see [Figure 10](#)), illustrates the sharp decrease in airflow that we have sometimes found during the glottal closing phase of the waveform from a trained singer. This slide is referred to by Dr. Titze in his comment and also illustrates the offset from zero flow mentioned by Dr. Fujimura in his comment. Concerning your conjecture that some oscillations removed by the inverse-filtering procedure might actually originate at the glottis. I agree that one must be careful. especially at high phonatory levels. For example. as we extended the frequency response of the mask we use to over 3000 Hz. we found oscillations in the waveform at about 3200 Hz that appeared to come from a resonance that did not fit into the normal formant pattern. The oscillations near 3200 Hz that occur in the traces from the singer's voice were probably from this resonance, possibly amplified by a nearby 'singers formant' that was not removed by our 3-formant filter. However, the oscillations that we find in these frequency range do not appear to have been generated by oscillatory movements of any part of the vocal folds.

Dr. Titze: Why?

Dr. Rothenberg: This frequency seems to me to be too high for the tissue masses involved. and further testing indicated that it probably was a resonance introduced by the mask. So now we have added another stage to the inverse filter to remove that resonance also. However, this example might illustrate that a strictly automated inverse-filtering procedure could produce an error by removing a spectral peak that actually originates at the glottis. One can't eliminate the possibility of such oscillations at below, say, 1000 Hz.
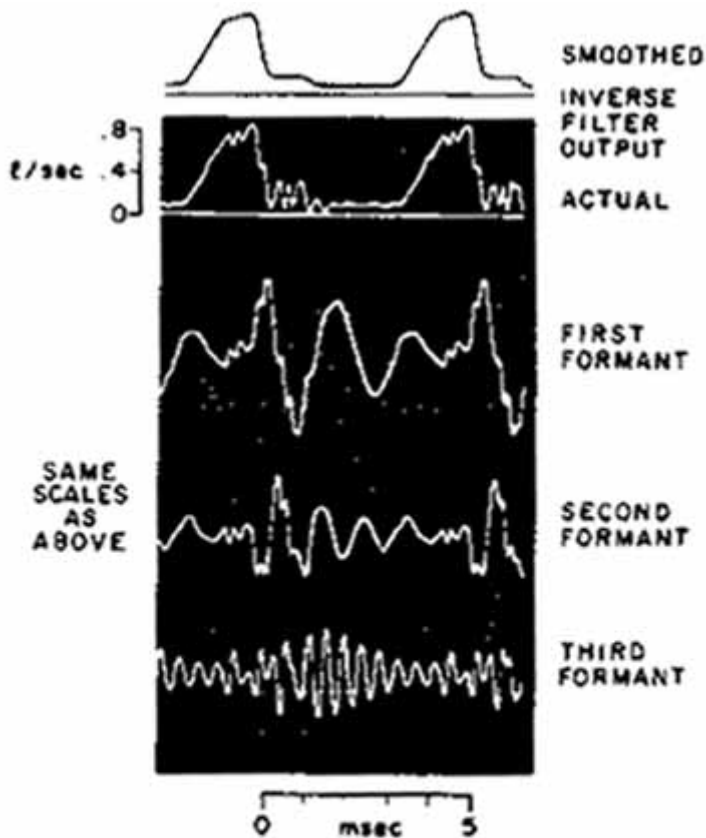
FIGURE 10. Components of the volume velocity waveform at the mouth for a vowel A song at 196 Hz by a trained male singer in the operatic-style "ring" mode. The three lower traces show the components removed by the three inverse filter stages, so that the four lower waveforms would sum to the measured volume velocity. The trace at the top was drawn to show what the inverse filter output would be like if the resonance at about 3200 Hz that can be seen in the waveform were removed by high order linear phase low pass filtering above about 2500 Hz. The recording was taken with the assistance of Dr. R. Colton and J. Estill of the New York State Upstate Medical Center.

Dr. Fujimura: Is the offset you discuss, Dr. Rothenberg, related to the vertical movement of the glottis? Do you have any estimate of the amount of contribution due to the net average vertical movement occurring in the glottis?

Dr. Rothenberg: I don't think that a significant offset of the waveform during the entire closed portion of the glottal cycle could be caused by vertical movement of the vocal folds. As Dr. Titze pointed out, smaller air flow components or components occurring over a shorter period of time, as during the glottal closing phase, could be due to vertical movements. But a significant offset existing over the whole closed period is not likely to have come from vertical motion. For example, assuming a closed period of 3 ms and a tissue area of 1 cm^2 that is moving vertically, the vertical motion required for a flow component of .1 liter/sec during the closed period would be about .3 cm. Assuming that the vertical movement extended somewhat outside of the closed period, the total vertical motion would be at least 1/2 cm. Thus, in the trained singer's waveform. the general offset from zero could be assumed to come from an incomplete glottal closure,

probably between the arytenoid cartiledge, while the 1 m/sec long 'shelf' just after the glottal closing period could conceivably have been due to a vertical motion.