THE GLOTTAL VOLUME VELOCITY WAVEFORM DURING LOOSE AND TIGHT VOICED GLOTTAL ADJUSTMENTS

MARTIN ROTHENBERG

In the study of the speech process, there has been a continuing interest in methods for obtaining the waveform of the air flow, or volume velocity, at the glottis. One of the methods used previously is illustrated at the top of Figure 1. In this technique, the radiated pressure waveform from the microphone is processed by a filter having a transfer characteristic the inverse of that of the vocal tract. This INVERSE-FILTER has a zero (antiresonance) for every pole (resonance) of the vocal tract in the frequency range of interest, including a pole at zero frequency. The pole at zero frequency, which is an integrator in the time domain, is a fundamental source of inaccuracy in this method. Since the integrator increases the low frequency response of the filter, this method is very sensitive to low frequency noise. Secondly, since it is impossible to use a true integrator, the dc level of the glottal waveform is lost in the process. And lastly, the amplitude calibration depends on factors such as the distance of the microphone from the lips and cannot be made accurately (Miller 1959, Holmes 1962, and Miller and Mathews 1963).

In our work, we use the VOLUME VELOCITY at the mouth, instead of the PRESSURE, as the input to an inverse-filter, as shown at the bottom of Figure 1. The inverse-filter for volume velocity does not require an integrator and, therefore, provides an accurate zero level and is not sensitive to low frequency noise. There is also no problem with amplitude calibration. The volume velocity at the mouth is shown sensed by a standard wire screen pneumotachograph mask, of the type used in respiratory measurements.
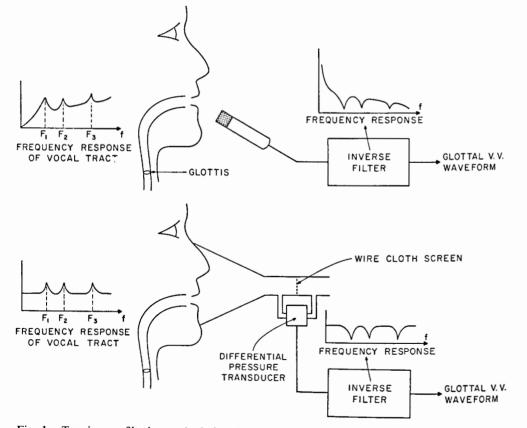

Fig. 1. Two inverse-filtering methods for obtaining the glottal volume velocity waveform: above; inverse-filtering the radiated pressure waveform, below; inverse-filtering the volume-velocity waveform at the mouth.
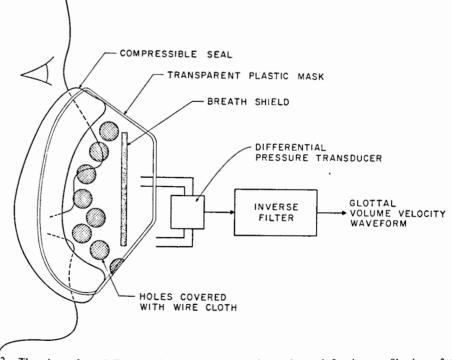

Fig. 2. The circumferentially-vented pneumotachograph mask used for inverse-filtering of the volume velocity at the mouth. The differential pressure transducer was formed by a pair of modified Sennheiser MKH 110 microphones connected to a differential amplifier. One microphone sensed the pressure inside the mask and one microphone sensed the pressure outside the mask.

A pneumotachograph mask more suitable for inverse-filtering is shown in Figure 2. To reduce speech distortion caused by acoustic loading of the vocal tract, and improve the frequency response of the pneumotachograph, the wire screen is distributed around the circumference of the mask, as close to the face as possible. To keep the pressure difference across the screen linearly related to volume velocity, the transducer port sensing internal pressure is shown protected from the direct impact of the breath stream. We use a disc of polyurethane foam for this purpose.

If the inverse filter is adjusted accurately, the fidelity of the inverse-filtered glottal waveform is limited by the pneumotachograph mask. The response time of our mask was about $\frac{1}{2}$ msec., limiting the system to glottal waveforms with a fundamental frequency no more than 150 to 200 Hz.

For a reasonably good approximation to the glottal waveform, formant frequency and damping settings are not critical if measurements are made during vowels having a first formant at least three or four times higher than the fundamental frequency. For this reason we favored vowels such as /a/ and /æ/. The inverse-filter was normally adjusted from spectrogram measurements of the first two formants, and the settings retouched, if necessary, by observing the inverse-filtered waveform (Holmes 1962). The third and higher formants were removed by low pass filtering.

Figure 3 shows some typical results when inverse-filtering voiced vowels of an adult male speaker of American English. The waveforms were recorded during productions of the English vowel /a/ (as in 'hot') in successive repetitions of the non-sense phrase /bap/, as /bapbapbap.../. The measurements were made at a relatively steady-state portion of the vowel. The sub-glottal pressure was estimated from the intraoral pressure during the labial closure for the /pb/. Intraoral pressure was transmitted to a pressure transducer by means of a small tube inserted at the corner of the mouth (Rothenberg 1968).

Note that in the lowest trace there was not a complete closure of the vocal cords; during the most closed portion of the glottal cycle, where the flow is minimum, and the waveform is flat, the waveform was not at the level of zero air flow. When inverse-filtering the PRESSURE wave, such effects are not detectable, since there is no zero level indication in that method. This air flow during the most closed portion of the glottal cycle, when the waveform is relatively flat, is probably due to an incomplete closure of the vocal folds posteriorly, between the arytenoid cartilages.

In applying volume velocity measurements to the development of phonetic theory, we have been especially interested in the ways in which the glottal air flow waveform reflects the openness or the tightness of the glottal adjustment. For the same trans-glottal pressure and mode of vocal fold vibration, a larger peak air flow generally indicates a looser adjustment of the vocal folds (vocal folds not pressed together as tightly), and vice versa.

We have previously hypothesized that a maximally-fast CYCLIC GLOTTAL OPENING MOVEMENT is important in the description of stop consonants (Rothenberg 1968). This movement can be from a VOICED glottal adjustment to BREATHY-VOICED and
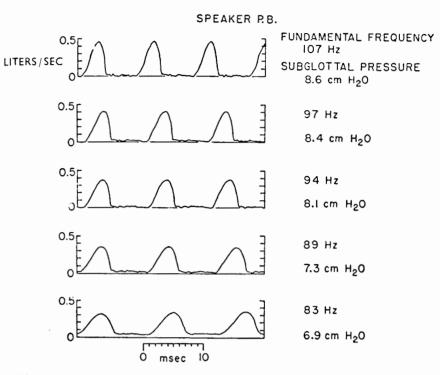
Fig. 3. Glottal volume velocity waveforms obtained by inverse-filtering the volume velocity at the mouth. From five successive repetitions of the phrase /bap/ (from top to bottom). Time in each trace runs from left to right. Waveforms traced from original.

back to VOICED, or from VOICED to OPEN to VOICED. As shown in Figure 4, the dynamic limitations in such a movement can be measured by recording the air flow during an /h/ phoneme in English when unstressed, or during rapid speech. The same, or similar glottal opening movements can occur during unstressed /p/, /t/ and /k/ phonemes in English. The changes in glottal adjustment during a /p/ can be monitored if the closure at the lips is bypassed by a short tube of a suitably large diameter (Rothenberg 1968). We used here a tube about 3/8" in diameter and about 3/8" long, located at the corner of the mouth. During these traces, formant energy was removed with a 6-pole linear-phase low-pass filter, and not by inverse-filtering. Although the low-pass filtering method obscures the details of the waveform, the general shape is correct enough for seeing the time course of the glottal opening movement. A BREATHY-VOICED adjustment of the vocal folds can be defined empirically to be indicated by an increase by a factor of about two in the peak air flow during the glottal cycle, and an OPEN adjustment by a termination of quasi-periodic oscillations of the air flow waveform. As observed previously by a less quantitative method (Rothenberg 1968) the traces in Figure 4 illustrate that for this speaker a maximally-fast movement from VOICED to BREATHY-VOICED to VOICED takes about 100 msec., and a movement from VOICED to OPEN to VOICED takes about 125 msec.
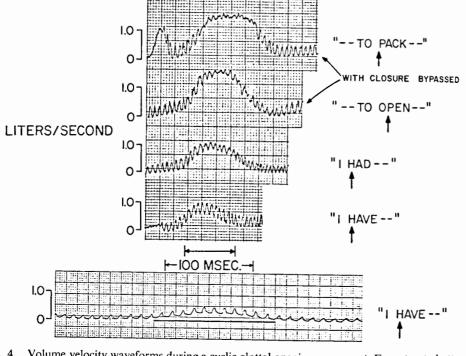
Fig. 4. Volume velocity waveforms during a cyclic glottal opening movement. From top to bottom:
voiced — open — voiced
voiced — open — voiced (minimum duration)
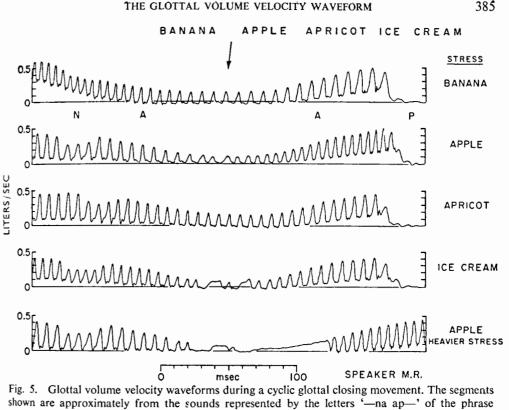voiced — breathy — voiced
voiced — breathy — voiced (minimum duration)
voiced — breathy — voiced (in vocal fry)

In other such traces it was observed that a speedup in the rate of speech only results in the target glottal adjustment not being reached. The last trace in Figure 4 was produced by a phonetically-trained speaker, using an intended "vocal fry" voicing (Hollien *et al.* 1966).

Also of interest for describing the phonetic characteristics of speech are the dynamic constraints in the production of a CYCLIC GLOTTAL CLOSING MOVEMENT, from VOICED to TIGHTLY-VOICED to VOICED, or from VOICED to STOPPED to VOICED. As illustrated in Figure 5, this type of movement can be found in English at word boundaries, under certain patterns of stress. The traces show the inverse-filtered glottal volume velocity waveforms during a cyclic glottal closing movement at the intervocalic word boundary between 'banana' and 'apple' during the phrase 'banana apple apricot ice cream', as spoken by a native speaker of American English. The TIGHTLY-VOICED glottal adjustment results in glottal pulses that are much smaller than during normal voicing and that are often irregular. The glottal closing movement appeared to be most consistently strong when the primary stress immediately followed the juncture.

In some languages, such as Somali and other Cushitic and Semitic languages, this type of glottal closing movement attains phonemic significance (Bell 1953). Some

Fig. 5. Glottal volume velocity waveforms during a cyclic glottal closing movement. The segments shown are approximately from the sounds represented by the letters '—na ap—' of the phrase 'banana apple apricot ice cream'. Primary stress was on the word indicated at the right. The apparent small negative air flows occurring during some repetitions were probably due to the displacement of air caused by the moving articulators. Waveforms traced.
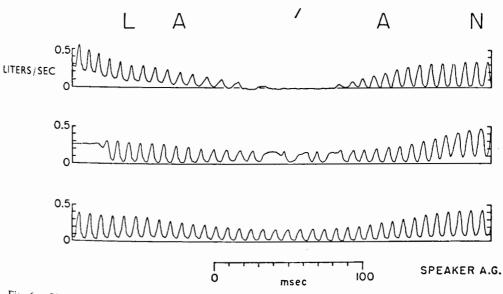

Fig. 6. Glottal volume velocity waveforms during three occurrences of a glottal stop phoneme in Somali phrases containing the term /la'an/ ('lacking'). The speaker was an adult male native speaker of Somali. Waveforms traced.

sample traces are shown in Figure 6 for the air flow during the phrase /la'an/ ('lacking') in various contexts. We have found that in both English and in Somali, there is often no complete glottal stop, with only a glottal adjustment of TIGHTLY-VOICED attained during the movement. From measurements made so far, we estimate that a cyclic glottal movement that reaches a CLOSED or STOPPED adjustment takes at least 140 msec., and a movement that reaches a TIGHTLY-VOICED adjustment, defined empirically by a peak flow reduced by about ½, must take at least 110 msec.

Department of Electrical and Computer Engineering
and Department of Linguistics
Syracuse University
Syracuse, New York

REFERENCES

Bell, C.R.V.
1953   *The Somali Language* (London, Longmans, Green and Co.).
Hollien, H., P. Moore, R. Wendahl, and J. Michel
1966   "On the Nature of Vocal Fry", *Journal of Speech and Hearing Research* 9:245-247.
Holmes, J.N.
1963   "An Investigation of the Volume Velocity Waveform at the Larynx During Speech by Means of an Inverse Filter", in *Proceedings of the Speech Communication Seminary, Stockholm 1962, Vol. I*, paper B-4 (Royal Institute of Technology, Stockholm).
Miller, J.E. and M.V. Mathews
1963   "Investigation of the Glottal Waveshape by Automatic Inverse Filtering", *Journal of the Acoustical Society of America* 35:1876(A).
Miller, R.L.
1959   "Nature of the Vocal Cord Wave", *Journal of the Acoustical Society of America* 31:667-679.
Rothenberg, M.
1968   "The Breath-Stream Dynamics of Simple-Released-Plosive Production", *Bibliotheca Phonetica VI* (Basel, Karger).

DISCUSSION

DANILOFF (Champaign, Ill.)

1. Have you used your experimental technique to examine glottal behavior during various kinds-levels of juncture?

2. Have you any idea of how severe the error is in measured glottal source during open vs. closed glottal conditions, i.e., tract response differs during open and closed glottal phases?

ROTHENBERG

1. Relating to juncture, we have investigated only the type of situation reported in the paper, in which we have verified that a glottal closing or tightening movement can be used to signal a word boundary in English, and perhaps even help signal the stress pattern. I will be looking for other examples where there is reason to believe

that changes of glottal parameters other than fundamental frequency are phonetically significant.

2. There is indeed an error due to the change in vocal tract acoustics when the glottis opens, though I believe this error to be small in the waveforms shown. It can be shown that the slightly s-shaped curvature in the rising or opening phase of the topmost waveform in Figure 4 (at 107 Hz) is due to this cause. For this type of glottal waveform, the affect generally becomes worse at higher fundamental frequencies. It may be possible to design an inverse-filter that takes this affect into account, though the increased complexity would probably require the use of a digital computer for the implementation of the filter.

HOLLIEN (Gainesville, Fla.)
How do you separate the poles and zeros produced glottally from those created in the vocal tract?

ROTHENBERG
If you mean the poles and zeros introduced by the subglottal system when the glottis opens during the glottal cycle, then my reply to Prof. Daniloff also holds for your question.

If you mean the zeros in the glottal wave spectrum, as described some time ago by Flanagan, then the answer is that we remove the poles due to the vocal tract and leave the zeros associated with the glottal waveform. This assumes, of course, that we know the locations of the vocal tract poles. For some combinations of vowel articulation and glottal quality, the poles are hard to locate accurately, sometimes due to the glottal zeros. However, if we are interested mainly in the action of the glottis, we can usually change the vowel to one for which the formants are easier to measure. This is why we favor the vowels /a/ and /æ/.

PERKELL (Cambridge, Mass.)
Could you explain what you mean in detail by 'tightly-voiced'?

ROTHENBERG
TIGHTLY-VOICED is used as a comparative term. It indicates that the voicing was produced by a glottal adjustment in which the vocal folds are pressed together more than in a neighboring 'voiced' segment, as, for example, if the lateral cricoarytenoid muscles were more strongly activated. However, I do not refer to any one adjustment, but a class of such adjustments that might be expected to function together linguistically in any given language in which 'tightly-voiced' is used in opposition to 'voiced'.

SCULLY (Leeds)
The opening-closing time of the vocal folds which your results show is a good deal shorter than the 250 msec time indicated by the data of e.g., D.H. Klatt for voiceless

fricatives ("Articulatory Activity and Air Flow During the Production of Fricative Consonants", *M.I.T. Quarterly Progress Report* 84, 257-260 [1967]). My own aerodynamic data indicates times in agreement with Klatt for English voiceless stops and fricatives in post-stressed position.

ROTHENBERG

The numbers quoted in my paper are for maximally-fast changes in glottal adjustment. Considerably slower movements can and do occur, especially in a stressed position and in slow speech. Also, I prefer to define the duration of an opening-closing movement as the shortest time interval that includes about 90% of the area under the waveform of the increase in volume-velocity, or the increase in glottal area. If instead, one measures the duration of the opening movement from the instant that the vocal folds first begin to separate, to the first instant that the vocal folds have clearly returned to normal non-breathy voicing, the durations measured would be somewhat higher in magnitude.