STUDIES OF GLOTTAL EXCITATION USING INVERSE FILTERING AND AN ELECTROGLOTTOGRAPH

Melvyn J. Hunt
National Research Council of Canada
National Aeronautical Establishment
Building U61, Montreal Road
Ottawa, K1A OR6, Canada

Abstract
Glottal excitation has been studied in steady vowels produced by three subjects, two male and one female. Electroglottograph waveforms are shown together with simultaneous glottal airflow waveforms and waveforms for individual formants, both derived by inverse filtering. Modal voice shows formant excitation concentrated on the instant of closure. Falsetto voice shows a triangular or sinusoidal airflow waveform. Breathy voice shows appreciable formant excitation both on closure and at the centre of the open phase. Creaky voice shows appreciable excitation at the start of the open phase as well as its end, and there is often an alternation in spectral content of the excitation from cycle to cycle causing the relative intensities of formants to vary. In examples of extreme creak the airflow waveforms are complex and difficult to interpret, but they are similar to the electroglottograph waveforms.

1. Introduction
The purpose of this paper is to present the results of a study of several modes of phonation. Glottal airflow waveforms derived by inverse filtering are compared with a waveform indicating the area of contact of the vocal folds. By leaving one formant uncanceled in the inverse filtering, the excitation of individual formants is also shown.

2. Method
Subjects spoke in an anechoic chamber, and two-channel recordings were made on a Revox A77 tape recorder of the output from a B&K condenser microphone and a laryngograph (electroglottograph) [1]. The laryngograph measures r.f. impedance across the larynx and hence the area of contact of the vocal folds [2]. The two signals were simultaneously digitized at 20kHz, sharply low-pass filtered at 5kHz without phase distortion and downsampled to 10kHz. Low-frequency phase distortions introduced by the recording process were removed from both signals by an automatic method [3].

The effect on the speech signal of radiation from the lips, which corresponds approximately to differentiation, was countered by integrating the signal. The filtering effect of the vocal tract on the glottal airflow was removed by interactive inverse filtering [4]. In this process, the formants in the speech signal are canceled by an equal number of antiresonances (typically five), whose frequencies and bandwidths are intended to match those of the corresponding formant exactly. The interactive system allows the parameters of one antiresonance to be adjusted at a time. The frequency and bandwidth of this antiresonance are determined by an a/d converter that frequently samples the values of two adjustable potentiometers. The speech signal filtered by the fixed and varying antiresonances is displayed on a graphics screen (DEC VT11). When fewer than a thousand samples are being plotted, the filter parameter update, the filtering, and the reploting can all be carried out within 50ms, giving the user the impression of a continuously changing display as he turns the knobs connected to the potentiometers. By adjusting the antiresonance corresponding to each formant in turn, the effect of the vocal tract can be tuned out in about one minute.

In modal voice at least, the spectrum of the glottal airflow waveform falls off at roughly 12dB per octave. Consequently, if the airflow is differentiated twice, the spectrum is flattened and the main feature in the waveform is an impulse at the instant of glottal closure. It is predominantly this impulse that excites the formants, which in this representation have roughly equal amplitude. When individual formants are shown in this paper, they are shown in a signal corresponding to the doubly differentiated airflow (or, equivalently, the singly differentiated speech signal).

3. Speech Material
Three speakers, two male (AF and MH) and one female (EA), each produced examples of steady unaspirated vowels with modal, falsetto, breathy, and creaky voice quality. Examples of several vowels were provided, but analysis has been concentrated largely on the schwa, since this vowel has formants that are well separated from each other and the first formant is fairly high in frequency and thus well separated from the fundamental frequency (F0).

4. Discussion of the Validity of the Method
It is sometimes alleged that inverse filtering is a subjective process in which the user obtains the airflow waveform he expects. There may be some truth to the allegation as far as fine detail in the flow is concerned, but we believe that the gross shape of the airflow waveform corresponds to the actual flow. The criterion normally said to be used to determine the antiresonance parameters is the flatness of the closed-glottis phase, since there is by definition no airflow during this period. This criterion is not useful for phonation modes in which the glottal closure is very brief. With the interactive method used here, however, it is still possible to estimate the parameters because a continuous range of parameters can be surveyed quickly and it becomes evident that only one set of values results in a maximally simple airflow. If the parameters of any one antiresonance are disturbed, the corresponding formant will appear in the waveform.

Further support for the validity of the method comes from the following considerations. Pairs of individuals working independently choose substantially the same sets of parameters and thus obtain the same airflow patterns. Modal-voice airflow waveforms from two recording sessions with the same speaker look similar. When there is an identifiable closed-glottis phase of suitable length, covariance-method linear predictive coding (LPC) [5] derives a very similar set of parameters and hence the same airflow waveform as an operator using the interactive method. In tests with synthetic speech, it is possible to recover the known excitation waveform exactly. The surprisingly complex airflow waveforms in the extreme creak illustrated in the next section show a strong similarity to their corresponding laryngograph waveforms. Finally, distinguishing features of the different phonation types are consistent across speakers.

5. Results
In the figures the laryngograph waveform is labeled as Ln, the glottal airflow as F0, the doubly differentiated flow as F0", and the n'th formant as Fm. Unless otherwise indicated in the figure cap-
ions, the vowel is a schwa and the duration of the waveforms is 20ms. The waveforms derived from the speech signal are delayed relative to the laryngograph signal by about 0.5ms because of the sound propagation time to the microphone.

I) Modal Voice

The waveforms for the two male speakers (Fig. 1,2) showed a prolonged closed phase, there was a clear impulse at closure in $F_2^\text{II}$, and formant excitation was concentrated at this point. The female speaker (Fig. 3) showed a much shorter closed phase, $F_2^\text{II}$ being more complex, but formant excitation still appeared to be concentrated on the instant of closure.

![Fig 1. MH modal voice. Excitation at instant of closure, appearing as an impulse in $F_2^\text{II}$.](image1)

![Fig 2. AF modal voice. Excitation concentrated on closure, though less clearly than for MH.](image2)

![Fig 3. EA modal voice. $F_2^\text{II}$ not impulse-like.](image3)

II) Falsetto

For all three speakers (Figs 4,5,6) $F_2$ has a triangular or sinusoidal waveform consisting with an excitation spectrum tending to $F_2^\text{II}$ with frequency more quickly than in modal voice. Any closed phase is very short. $F_2^\text{II}$ is reminiscent of the modal voice example from the female speaker, EA, suggesting that its form may be due to a high $F_2$ rather than to a property of falsetto voice.

![Fig 4. MH falsetto. $F_2$ is triangular. $F_2^\text{II}$ is not impulse-like.](image4)

![Fig 5. AF falsetto. $F_2$ and $L_2$ are sinusoidal. $F_2^\text{II}$ is not impulse-like.](image5)

![Fig 6. EA falsetto 10ms.](image6)

II) Breathy Voice

For all three speakers (Figs 7,8) $F_0^\text{II}$ is rhythmic, the duration of closure is less impulse-like than in modal voice. Single formant peaks for the two male speakers (Figs 7,8) indicate that excitation is strongest at two instants: at closure and at the point of maximum airflow in the centre of the open phase. Excitation for the female speaker is more noise-like (Fig 9), though it may peak on closure and at maximum flow.

![Fig 7. MH breathy. Note excitation in the centre of the open phase.](image7)

![Fig 8. AF breathy. Note excitation in the centre of the open phase.](image8)

![Fig 9. EA breathy. Excitation seems noise-like.](image9)

III) Creaky Voice and Creak

Subjects produced a range of phonation types ranging from a creaky voice whose $F_0$ was superficially similar in form to that for modal voice to a full creak with a glottal cycle so long that individual pulses are perceived rather than $F_0$ and with $F_0$ and $L_0$ waveforms that are complex and difficult to interpret. MH produced exclusively creaky voice, EA exclusively creak, and AF examples of both extremes as well as intermediate examples. In all cases, the waveforms varied from cycle to cycle in those other phonation modes.

Creaky voice often showed alternation between two glottal cycles, the cycles differing in their airflow waveform as well as in their duration. In one of the two cycles there was often strong excitation at the point of glottal opening (Fig. 10). The details of closure generally differed in the two cycles, leading to differences in the excitation spectrum and hence to the relative intensities of formants. In examples of creaky (c) from MH and AF (Figs 11,12) there were series of cycles in which a strong $F_0$ and a weak $F_0$ alternated with a weak $F_0$ and a strong $F_0$. Since the formant waveform from a previous excitation has not decayed to a negligible value when the next excitation comes along, there will be an interaction between the two waveforms that depends on their relative phases and hence on the formant frequency. This interaction could conceivably account for the observed weak/strong alternation. However, when a single excised piece of flow waveform containing just one excitation event is used to excite a resonance corresponding to $F_0$ or $F_0$ the effect is still seen.

It seems, then, that the alternation in formant intensities really is due to an alternation in the spectral content of the glottal airflow.

The schwa examples from AF and EA (Figs 13,14) are complex, extreme creak. As noted in Section 4, the $L_2$ and $F_0$ waveforms are similar. The $F_2$ waveform shows periods in which an impulse is followed by a long phase of little activity. Set of formants excited in this period showed a sharp decay (Fig 15), which offers an opportunity of accurate formant bandwidth determination. They seem to indicate bandwidths much narrower than values normally assumed.

Examples of (c) and (c) vowels from AF (Figs 16,17) are complex by intermediate between the creaky (c) and the extreme creak in the schwa. They perhaps offer a means of bridging the gap between the two extremes and thus interpreting the extreme creak waveforms.
6. Discussion of Results

Justifications of LPC usually assume that excitation of voiced speech is concentrated on the instant of glottal closure and that the excitation waveform approximates an impulse for speech preemphasized at 6dB/octave. The results presented here suggest that these assumptions may be valid only for modal voice, perhaps only for male modal voice. Pitch-synchronous LPC [6,7], in which a negative-going impulse in the \( L_x \) waveform is used to detect closure and hence excitation, would be likely to encounter difficulties with creaky or breathy phonations having excitation on opening or in the middle of the open phase. Tests of pitch-synchronous LPC with the speech discussed here support this expectation.

Some algorithms for determination of \( F_o \) detect excitation points, while others detect repetitions of patterns in the waveform or, equivalently, harmonic structure in the spectrum. These two approaches would give different results from each other with breathy or creaky phonations. Moreover, neither approach would consistently indicate the frequency of opening and closing of the glottis.

Acknowledgments

The work reported here was carried out mainly while I was employed at the U.K. Joint Speech Research Unit. I am grateful to John Bridle and to Prof. Adrian Fourcin for their help.

References