Estimation of Glottal Source Parameters from Diverse Signals

Arlo Faria

***Abstract*—This is a study to investigate the parameters of the glottal source signal that are evidenced by three kinds of signals: an electroglottographic (EGG) signal, an acoustic signal, and an oral airflow signal. The glottal pulse waveform, whether intrinsically defined as glotal airflow or area, can be parameterized with attributes that vary across speakers and types of phonation. Some of these parameters are readily inferred from an EGG signal, while others are less apparent. Additionally, a variety of methods have been proposed for glottal source or parameter estimation from the acoustic or oral airflow signals. An informal experiment involving three subjects illustrates the use of EGG to observe the the open quotient. Then a method for deriving voice source parameters from a filtered acoustic waveform is considered. Finally, inverse filtering is briefly discussed with respect to the acoustic and oral airflow signals.**

I. Introduction

GLOTTAL source waveforms – whether representing air pressure, airflow, or glottal area – are very difficult signals to measure directly and non-invasively. There exist a variety of methods for estimating this signal from others such as an electroglottograph (EGG), acoustic signal, or oral airflow. For many applications involving voice quality, however an exact representation of the glottal waveform is unnecessary and certain descriptive parameters will suffice.

The classical glottal source parameters express time interval ratios of stages in one period of a vocal cord oscillation. One of these is the open quotient, the proportion of the period in which the vocal folds are not touching. Another classical parameter is the speed quotient, which relates the two phases of the closed interval: opening and closing. There is also a closing quotient. This paper focuses primarily on estimation of the critical time instants that mark the boundaries of the intervals defining these parameters:

$$\mathrm{OQ} = \frac{t_{open}}{t_{open} + t_{close}}$$

$$\mathrm{SQ} = \frac{t_{opening}}{t_{closing}}$$

$$\mathrm{ClQ} = \frac{t_{closing}}{t_{open} + t_{close}}$$

These four intervals can be calculated given accurate times for the moment of initial closure, maximum closure, and initial release of closure.

A common technique to indirectly observe the glottal waveform is to use an EGG, passing a small amount of electric current across the larynx. This measures the electrical impedance of the tissues, such that the signal's maximum values indicate glottal closure and the minima represent open glottis state. But this one-dimensional information correlates with vocal fold contact area, not the area of the glottal opening, so it does not faithfully represent the glottal pulse shape. Nonetheless, it is still possible to derive parameters from the EGG signal.

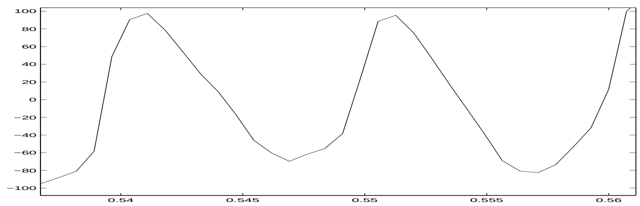

Fig. 1. EGG signal for a male.

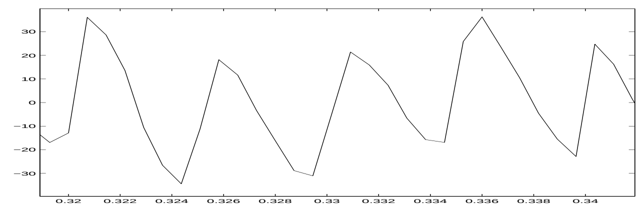

Fig. 2. EGG signal for a female.

The task of glottal analysis from an acoustic signal, the second part of this paper, can be done in a number of manners. One of these is discussed and implemented: tracking the attenuation of upper resonances in the acoustic signal. It proves successful in locating the crititical time instants that separate the open and closed phases of the glottal pulse.

Lastly, approaches with inverse filtering are given a cursory examination.

II. Experimental Procedure

Three subjects (two male, one female) were connected to three input channels: a regular headset microphone signal, an EGG signal with DC frequency response, and an oral airflow signal with DC frequency response[1]. One set of recordings was made for all three simultaneous channels, and another set was recorded with just the audio and EGG signals[2]. The vowel examined was [a] in the frame sentence "how <u>are</u> you?" Data were collected using the PCQuirer software, exported[3], and then analyzed in Matlab. The audio channel was sampled at 11,000 Hz and the DC channels at 1,375 Hz.

[1] The low-frequency components of the EGG signal, indicating laryngeal movement, are not used in this study. The DC response is neccessary for the airflow signal which will later be inverse filtered

[2] With the microphone outside the Rothenberg mask, the audio signal would be significantly affected and unsuitable for inverse filtering.

[3] The audio channel was saved in WAV format and the DC signals were written to log files as delimited ASCII text

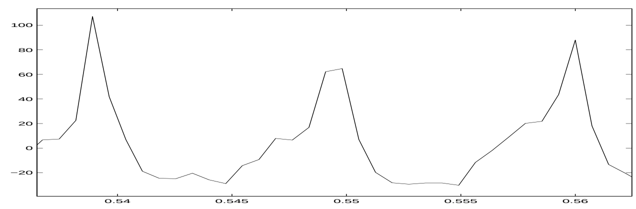

Fig. 3. First derivative of the EGG signal for a male.

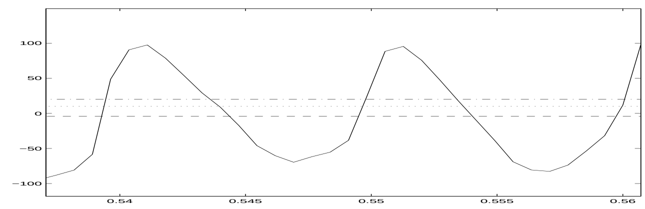

Fig. 4. Criterion thresholds (35%, 50%, equal-area) for male EGG.

III. Electroglottographic Signals

Figures 1 and 2 show EGG signals after being highpass filtered to remove drift from very low-frequency components (laryngeal movement). Males generally have less breathy voice than females, as these plots indicate by the nearly sinusoidal shape of the female voice source. It is worth noting that at this low DC sampling rate, voice source parameters will be more accurately estimated from the male's signal because its longer pitch period allows more samples per period.

It is thought that the EGG signal sufficiently represents the approximation and closure of the vocal cords, but not the dynamics of the open phase. The maximum of the waveform corresponds to the maximum closure of the vocal folds, so it marks the time dividing the closing and opening sub-intervals of the closed phase. Yet, parsing these shapes into the open and closed intervals is not so easy – it entails locating the instants at which the glottal opening and closure occur.

A. Derivative of the glottal waveform

Taking the derivative of the waveform helps to determine these time markers. It is widely believed that the strong positive peak in the derivative signal coincides with the closing of the glottis. It is also regarded, although questionably, that the weak negative peak indicates the moment of glottal opening. Figure 3 shows the first-order derivative of the signal in Figure 1, where positive and negative peaks are rather prominent and give the critical time instants of closure and release. Given the moment of maximum closure from Figure 1, all three glottal parameters can be calculated from these three times. Here the open quotient is seen to be about 0.5, the speed quotient is roughly 1.5 and the closing quotient is about 0.2.

B. Threshold Criteria

There is another way to estimate the moments of closure and opening from the glottal waveform: using a threshold criterion dependent on the amplitude variations. The portion of the waveform above the threshold corresponds to the closed phase, and below the threshold is the open phase. This empirically derived threshold can vary from 35-50% of the maximum amplitude, or even be set as an equal-area criterion. Each of these is shown in Figure 4, and it is apparent that the specific choice of threshold has a noticeable effect on the estimates of the open-quotient, which ranges from about 0.5 to 0.6.

IV. Acoustic Signals

In the majority of settings, the only practical speech input signal is the acoustic microphone signal. Nonetheless, some information about the glottal source might be recoverable.

A. Direct Inspection of the Acoustic Waveform

It has been suggested that some evidence of glottal opening and closing instants can be seen in the acoustic amplitude signal, particularly the high-frequency microphonics. At the moment of glottal closure, the vocal tract is a tube closed at the glottis and open at the mouth, and there is an impulse excitation. Though the glottis is closed, the vocal tract continues to resonate at high frequencies (such as F4), each reverberation attenuating with a decay envelope that can be seen in the filtered waveform. When the vocal cords separate, there is another excitation, smaller in amplitude, but resonating in a now longer cavity. The acoustic waveform should display another set of reverberations, possibly decaying at a different rate. These expectations were compared to the experimental data, with the result shown in Figure 5 for the same segment of the male voice source examined in Figures 1 and 3. The acoustic signal was adjusted by 0.8ms to account for the speed of sound to reach the microphone, and bandpass filtered to exhibit only the frequency range of the fourth formant.

It is interesting to compare Figure 5 with the the derivative of the EGG signal in Figure 3. Notice that the peaks of the derivative signal – indicating the instants of closure and opening – coincide perfectly with the minima of the resonance decay envelopes in the filtered acoustic waveform. There also appears to be a distinction at the moment of maximum closure, at the peak acoustic amplitude, although this is more tenuous. It was not possible to replicate these results as successfully with female speech, because the short pitch period did not allow enough time for the resonances to decay.

One problem with this signal is that while the time instants of closure and release are very clearly marked, it is not so apparent which one is which. However, in practice, it is easy to evaluate other sources of evidence – such as spectrograms or even the raw sound pressure signal – to determine the acoustically salient moment of glottal closure. The most important result here is that the moment of vocal fold opening is clearly marked.

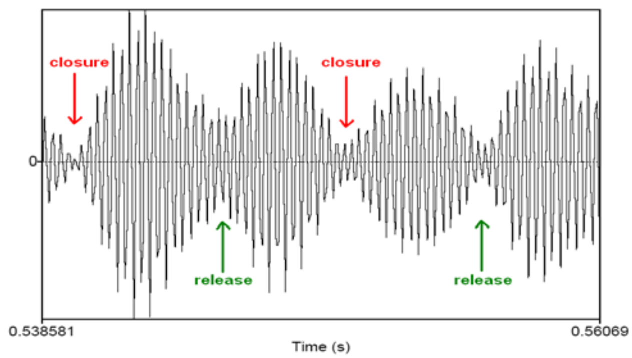

Fig. 5. Acoustic waveform filtered at fourth formant (male).

B. *Inverse filtering of the Acoustic Signal*

Inverse filtering obtains the input signal by applying the inverse of the vocal tract transfer function to the output signal. There are a variety of ways in which computational speech models can learn a vocal tract function for a given speaker. Generally speaking, the vocal tract's frequency response follows the formant peaks for a given acoustic spectrum, but without the harmonics. Inverse filtering seeks to recreate the glottal source signal by using the inverse of the transfer function, but in practice this is not perfect for several reasons: the procedure is not terribly robust given an output signal that includes noise, and the vocal tract's filtering characteristics are difficult to model precisely.

Consequently, the inverse-filtered signal will not always resemble the true glottal (air pressure) signal. But still parameters can be estimated from the signal and its derivative by treatment in the fashion of the EGG. A different method which is supposedly robust involves correlating open quotient with the difference between the first two harmonics of the inverse-filtered signal: $H1^* - H2^*$. Another solution is to create an alternative parameterization different from the classical parameters – artifical critical time instants are located on the inverse-filtered signal. Such measures may not reflect the classical parameters, but are still useful for many applications where voice qualities are under comparison.

V. Oral Airflow Signals

The acoustic signal recorded by a microphone can generally be considered equivalent to the first order derivative of the airflow signal at the mouth. Because it retains DC components, the oral airflow signal is sometimes inverse filtered to estimate the glottal airflow signal, where the slow laryngeal movements are under concern. It is not necessarily advantageous for estimation of more fine resolution parameters such as those mentioned previously. Nonetheless, inverse-filtering of the airflow signal is only effective if the DC signal is sampled at a high rate. As seen in Figure 6, an airflow signal sampled at 1,375 Hz does not capture the high-frequency microphonics that are needed to characterize the shape of the glottal pulse.

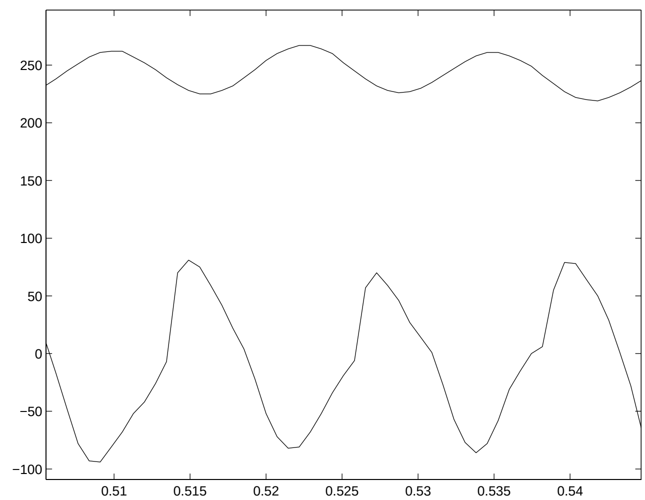

Fig. 6. Oral Airflow (top) and EGG signals (male).

VI. Conclusion

A variety of methods were illustrated for accessing the glottal source signal and estimating its parameters. Principally, this involves locating the moment of initial vocal fold contact, the moment of maximum contact, and the moment of release. The EGG signal clearly shows the instant of maximum contact, and its derivative indicates the other two critical time instants. Using criterion thresholds to divide the EGG signal into open and closed phases depends greatly on the empirical limits selected. An investigation of the acoustic waveform showed that open and closed phases could be extracted from the decaying waveform of a signal bandpass filtered at the F4 resonant frequencies.

VII. References

Many references were consulted for this project, but the most significant are informally listed below:

- An excellent EGG tutorial by Krzysztof Marasek: `www.ims.uni-stuttgart.de/phonetik/EGG`
- Paavo Alku's slides from VOQUAL03: `www.limsi.fr/WkG/VOQUAL/Slides/Paavo_ALKU/`
- Various publications from Martin Rothenberg: `www.rothenberg.org`
- Various publications from Nathalie Henrich: `www.lam.jussieu.fr/Individu/Henrich/`
- Numerous papers on inverse filtering.