**Publications of Dr. Martin Rothenberg:**
# Inverse Filtering on Your Laptop

## Abstract

Until recently, obtaining the glottal airflow waveform by inverse filtering oral airflow, using a circumferentially vented (CV) flow mask, required relatively expensive and cumbersome equipment. However, recent advances in computerized automated inverse filtering and computerized manual inverse filtering, and the development of a method for inputting the output of a CV mask into any Windows-based computer without the use of a special A-D converter, have made the process much more convenient and informative, and considerably less expensive. Some new techniques are described and the relationships between the electroglottograph signal and the glottal airflow waveform are illustrated. It is argued that the EGG and inverse filtered airflow are complementary, and an example from previous literature shown in which the combination of the two helped explicate source-tract interaction that may underlie the power in the upper ranges of a highly trained soprano, and protect the vocal folds from excessive airflow.

## Background

Under certain assumptions about the vocal tract, the waveform of the airflow pulses at the glottis during voiced speech or singing can be obtained by processing the waveform of the oral volume velocity (volume airflow at the lips) with an analog or digital filter having a transfer function (frequency response) which is the inverse of that of the vocal tract while the glottis is closed or almost closed. (Rothenberg, 1973, 1977) The most significant assumption is that vocal tract can be represented by a hard-walled tube of possibly non-uniform diameter, which is closed at the end representing the sound source (at the glottis) and open at the other, radiating end (at the mouth) These assumptions result in a transfer function with only poles (resonances or formants) and no zeroes (anti-resonances, such as that introduced by nasalization). There is also a implicit assumption of airflows and pressures throughout the vocal tract such that the laws of linear acoustics hold and that there are no significant sources of acoustic energy within the tract. Under these conditions, the transfer function of an inverse filter would consist of a series of zeroes or anti-resonances having frequencies and damping values that match those of the lowest poles or formants of the vocal tract.

The fact that an inverse filter can yield a very believable waveform having a flat (constant value) segment at or near zero flow during the glottal closed phase of normal, non-breathy voicing indicates that these assumptions, and others pertaining to the linearity and frequency response of the CV-mask and transducer system described below, are generally warranted. This period immediately following glottal closure is the greatest test of an inverse filter, since it is during this period that the acoustic energy to be removed is strongest.

The oral volume velocity waveform required for inverse filtering can be recorded using a circumferentially vented (CV) wire screen flow mask that serves to convert volume flow to a pressure differential, and an associated differential pressure transducer to record that pressure

differential. CV mask systems having a frequency response relatively flat to over 1000 Hz and usable to about 1500 Hz, and having tolerable linearity, drift and noise, have been marketed by Glottal Enterprises for over 25 years.
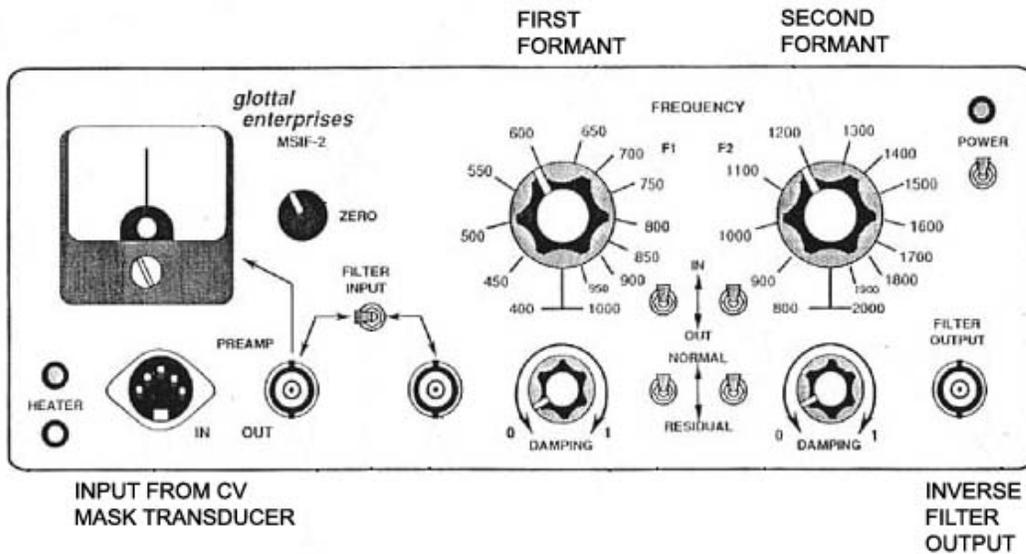
**Figure 1. A two-formant hardware inverse filter.**

Specially made masks that were designed to have a smaller internal volume by fitting within the lip opening (during a held vowel) have also been used to stretch the usable frequency response for measuring the glottal waveform with more precision during held sung vowels, as in Figure 4 below. (Approximate inverse filtering of a microphone signal is also possible, by adding an integration operation to the inverse filter, however the zero level is lost, amplitude calibration is difficult, and low frequency room noise such as from electrical equipment or an air conditioner is amplified by the integration operation.)

There have been a number of programs written to make the inverse filtering process completely automatic, and such algorithms can yield tolerable results for many combinations of voice pitch, voice quality and vowel value. However, it should be kept in mind that the inverse filter parameters (formant frequencies and damping values) must be set to remove formant energy only during periods in the glottal cycle in which the glottis is closed or nearly closed. Therefore, accurate settings usually require some manual adjustment, even if set initially by some automated algorithm. A good system for inverse filtering should allow for this type of manual adjustment.

## Techniques for setting inverse filter parameters

The problem of adjusting the inverse filter parameters has been shown to be made easier to solve by using vowels with a first formant (F1) much higher than that of the voice fundamental frequency (F0). As illustrated in Figure 2, these conditions result in a clear first formant

oscillation during the periods of glottal closure if the most important first formant is not cancelled correctly. If functioning of the voice source independent of articulation is being studied, the first formant can be kept high by using an open vowel , such as /a/ or /ae/. I usually prefer the English vowel /ae/ (as in 'bat'), since with this vowel, the second formant is better separated from the first than is the case with /a/.
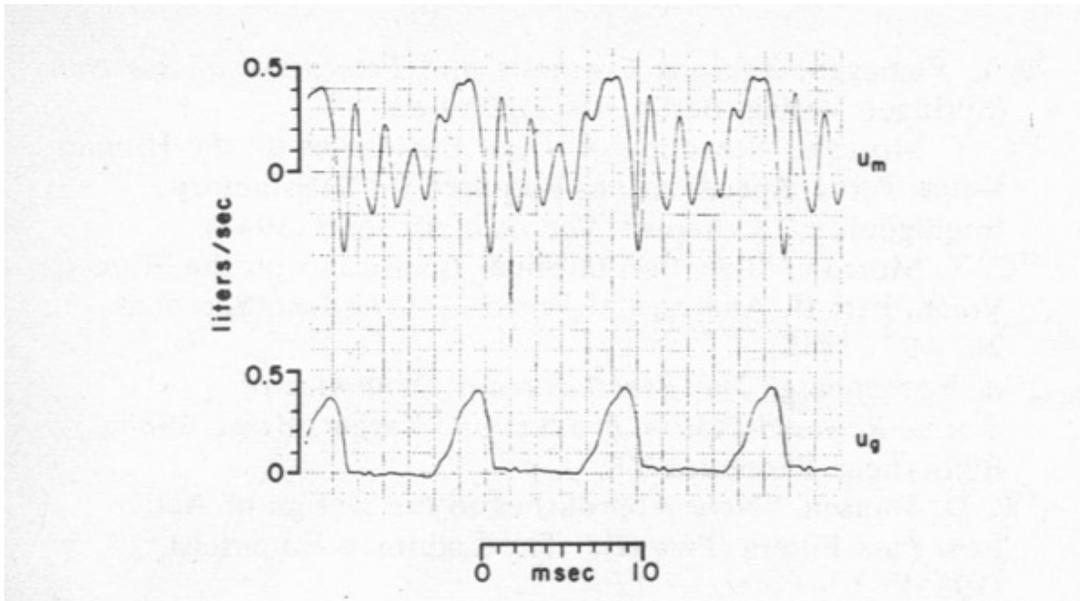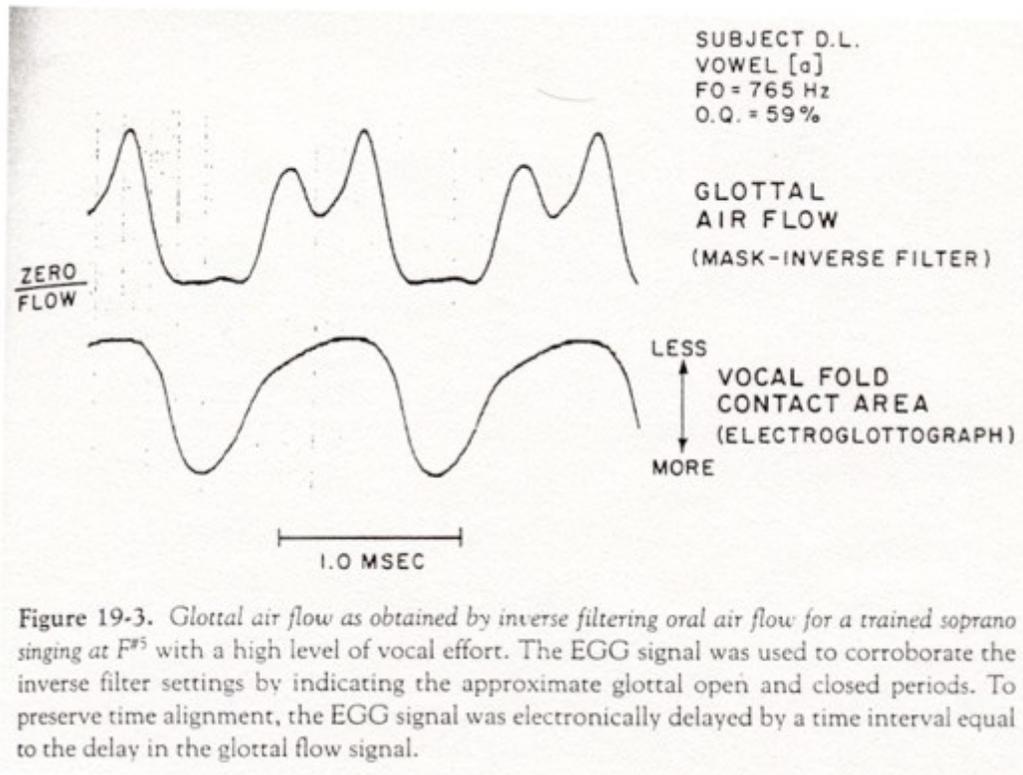


**Figure 2. Waveform from a CV mask for a held vowel /a/ by an adult male speaker, for which the value of F1 was much higher than F0, showing the strong oscillation at the frequency of F1 during the glottal closed phase [above], and the glottal airflow obtained by inverse filtering the waveform from the CV mask [below]. [from Rothenberg, 1973]**

When a good separation between F1 and F0 is not possible, as during the study of the interaction of articulation and pitch in tenor or soprano voices, we have used an electroglottograph (EGG) signal to help locate the period of glottal closure in the airflow waveform, as an aid in filter adjustment. Figure 3 shows an example taken from Rothenberg (1979), in which simultaneous EGG and inverse filtered airflow waveforms were used to corroborate each other. The periods of glottal closure indicated by each waveform were entirely consistent. In addition, in looking at many of such waveform pairs, one notes that the abruptness of glottal closure, an important factor in determining voice quality, is reflected equally in both waveforms.
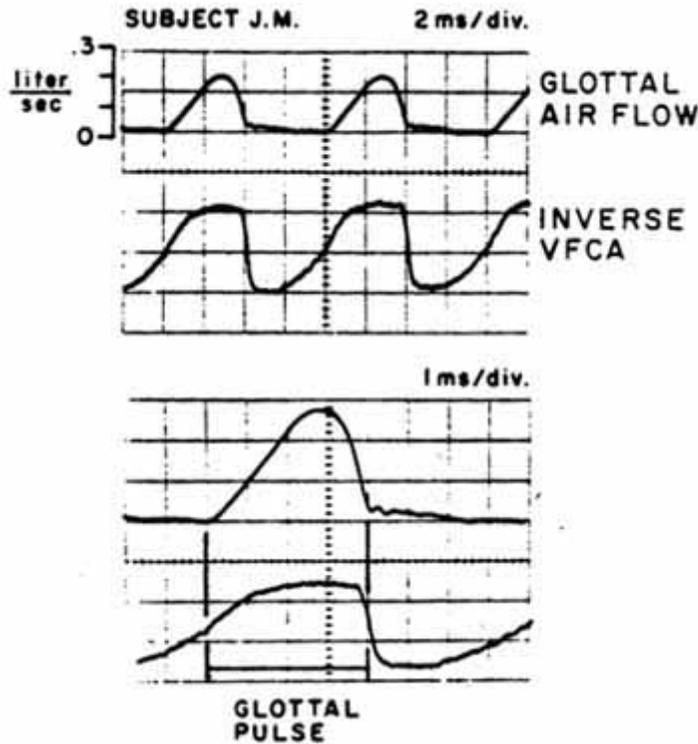
**Figure 3. Simultaneous glottal airflow waveform obtained by inverse filtering and electroglottograph waveform, showing the mutual corroboration of both the open glottis interval (the "glottal pulse") and period of glottal closure. [from Rothenberg, 1979]**

Figure 4 shows an important example of an inverse filter adjustment made possible by reference to a simultaneous EGG signal. It is from the paper "Cosi' Fan Tutte, and What it Means" (Rothenberg, 1986) in which it was shown that at some pitches (F0 approximately 765 Hz in this case) a highly trained soprano can both reduce the mean airflow and increase the level of the higher harmonics in the glottal waveform by tuning F1 to approximately match F0. With a properly tuned vocal tract, the pressure wave generated by the previous glottal airflow pulse returns to the glottis during its open phase, to depress the airflow and cause the dip in airflow seen in the figure. The strength of the higher harmonics are indicated by the relatively abrupt onset and offset of the airflow pulse. The fundamental frequency component, though repressed at the glottis by this tuning, is amplified acoustically in the tuned vocal tract, so that this production can be expected to have a final spectrum in the radiated acoustic pressure signal that is well-balanced and rich in tone. It may also be expected that the depression of peak airflow caused by the vocal tract tuning helps protect the laryngeal mucosa from the drying effects of high airflow.

**Figure 19-3.** *Glottal air flow as obtained by inverse filtering oral air flow for a trained soprano singing at F#5 with a high level of vocal effort. The EGG signal was used to corroborate the inverse filter settings by indicating the approximate glottal open and closed periods. To preserve time alignment, the EGG signal was electronically delayed by a time interval equal to the delay in the glottal flow signal.*

**Figure 4. Inverse filtered waveform with F1 approximately equal to F0, obtained using an electroglottograph waveform to identify the period of glottal closure during the adjustment of filter parameters. [Figure 19-3 from Rothenberg, 1986]**

In the example of Figure 4, proper tuning of the tuning of the inverse filter would have been extremely difficult, and probably could not have been accomplished, without a simultaneous EGG waveform to identify the period of vocal fold closure. As mentioned above, a small-volume mask was used to extend the frequency response of the flow-measurement system.

For inverse filtering with a high pitched voice, we have also had some success in using ingressive, glottalized air pulses, with a held vocal tract shape, to locate the formants, as used by Miller and his associates (1997, for example).

# New methods for inverse filtering

Figure 3 shows a hardware inverse filter from Glottal Enterprises having controls for varying the frequency and damping of each of two formants (model MSIF-2). Front panel switches also allow the inverse formants to be activated or removed selectively, as an aid in filter adjustment as well as an aid in visualizing the effect on the speech waveform of each formant. When used with an appropriate transient recorder capable of repetitive playback, it allows the user to vary filter parameters while observing the result.

It has long been feasible to implement as inverse filter digitally, using a slower-than-real-time post-processing of a captured airflow waveform. However, recent advances in processor speed have now made real-time processing possible to the extent that the real-time operation of a hardware filter can be emulated.

At Glottal Enterprises we have been developing a digital form of the MSIF-2 inverse filter. This software retains all the features of the hardware version and eliminates the need for a separate transient recorder for repetitive replay during filter adjustment.

The screen of the new digital filter is shown in Figure 5 with an airflow waveform from a vowel /ae/ spoken by an adult male speaker. The digital filter can cancel three formants and has an adjustable linear phase low pass filter for smoothing the inverse filtered trace and partially compensating for the high frequency emphasis caused by formants not canceled, in this case all formants above the third (Rothenberg, 1977). Formant parameters can be either set numerically or altered in small steps by clicking on, or holding down, the on-screen arrows. With processor speeds available in a modern home computer, the inverse filtered waveform changes essentially instantaneously in response to parameter changes.
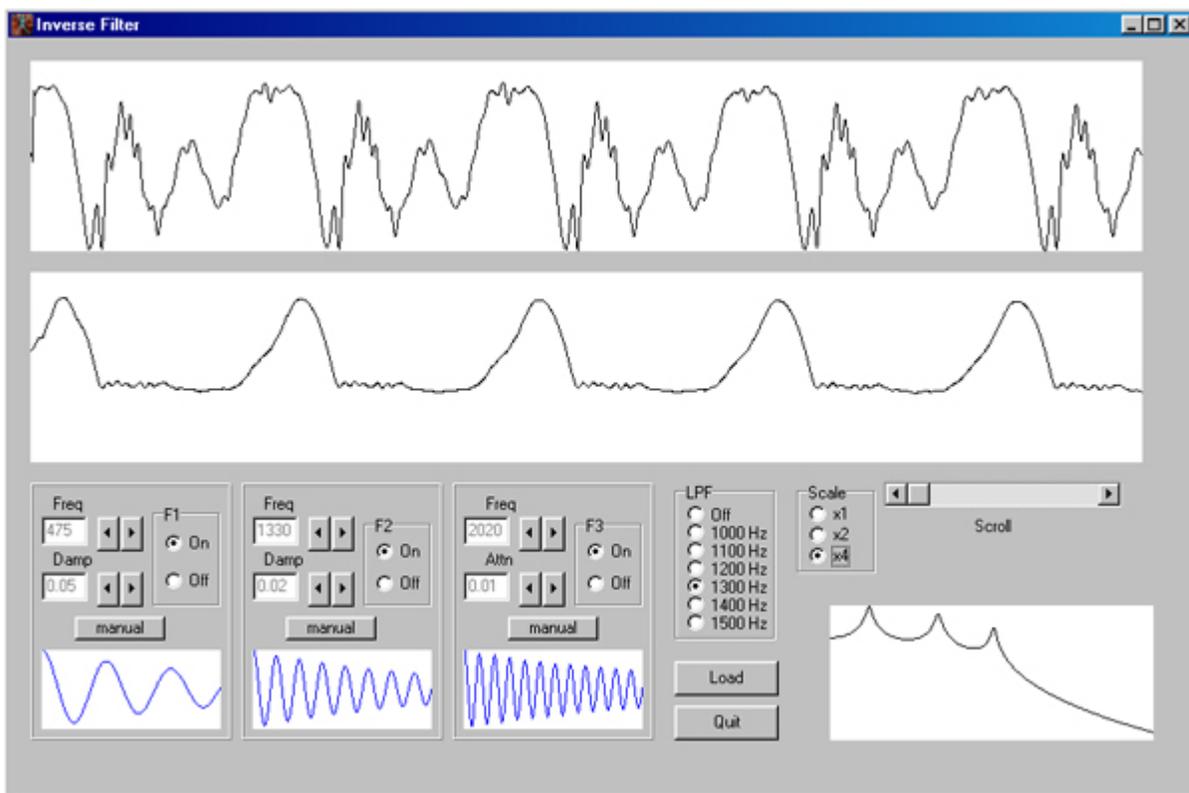


Figure 5. Screen layout for a digital inverse filter having all the functionality of a manual filter.

# REFERENCES

ROTHENBERG, M. "A New Inverse-Filtering Technique for Deriving the Glottal Airflow Waveform During Voicing", J. Acoust. Sec. Amer., 53, 1, pp.1632-1645, June 1973.

ROTHENBERG, M. "Measurement of Airflow in Speech", J. Speech Hear. Res., 20, 1, pp. 155-176, 1977.

ROTHENBERG, M. "Some Relations Between Glottal Air Flow and Vocal Fold Contact Area", in Proceedings of the Conference on the Assessment of Vocal Pathology, ASHA Reports No. 11, pp. 88-96, 1979.

ROTHENBERG, M. "Cosi' Fan Tutte and What It Means - or - Nonlinear Source-Tract Interaction in the Soprano Voice and Some Implications for the Definition of Vocal Efficiency", Vocal Fold Physiology: Laryngeal Function in Phonation and Respiration, T. Baer, C. Sasaki, and K. S. Harris, eds., College Hill Press, San Diego, pp. 254-263, 1986.

MILLER, D. & SCHUTTE, H., Comparison of vocal tract formants in singing and non-periodic phonation, J of Voice 1997, Vol. 1, pp. 1-11.