# GLOTTAL SOURCE ESTIMATION: METHODS OF APPLYING THE LF-MODEL TO INVERSE FILTERING

*Edward L. Riegelsberger* *Ashok K. Krishnamurthy*

Department of Electrical Engineering
The Ohio State University
Columbus, OH 43210

## ABSTRACT

This work focuses on incorporating the LF-model [1] into closed-phase inverse filtering based estimation of glottal source waveforms. The LF-model is a five-parameter model representing the effective voice source over open and closed glottal phases. Three techniques for LF-model fitting are described: two based on Prony's method and one using gradient descent. The estimation methods are evaluated using glottal flow waveforms obtained synthetically and through inverse filtering. Under appropriate conditions, all of the estimation methods produce good fits of the LF-model to the inverse filtered waveform. In noise, gradient descent produces more consistent fits than the more tractable Prony-based techniques.

## 1. INTRODUCTION

In simple speech synthesis and speech coding systems, the glottal source in voiced speech is frequently modeled as a series of impulses at the fundamental frequency. Recent research [2, 3] has indicated that more accurate modeling of the glottal source waveform results in more natural sounding synthesis and coding. Appropriate glottal source waveforms have been shown to help distinguish between breathy, pressed, or normal phonations and incorporate speaker distinctive qualities.

Measuring the glottal source waveform and its interaction with the vocal tract is not an easy task. There are no known methods to directly measure volume velocity through the glottal opening. Techniques do exist for measuring vocal fold motion and determining the glottal opening area such as electroglottography and photoglottography. While these measurement techniques can be used to infer characteristics of the glottal volume velocity, they do not produce direct estimates of glottal volume velocity flow.

If the transfer function of the vocal tract is known, glottal volume velocity can be determined through deconvolution of the sound pressure waveform. This is the principle of inverse filtering in speech analysis [4, 5]. For many reasons, inverse filtering is a difficult and inherently underdetermined task. Although many varieties of inverse filtering techniques exist, the vocal tract transfer function is generally estimated over the closed phase interval of the pitch period to obtain an all-pole model. The precision of vocal tract transfer function estimates is therefore limited by the finite closed phase duration. This limitation is particularly serious for high pitched voices, which have very short closed phase durations. Since the estimates are all-pole, vocal tract configurations with dominant spectral zeros (such as nasals) are poorly represented. Additionally, the inverse filtering paradigm assumes a time-invariant vocal tract over each pitch-period. This assumption neglects source-tract interaction produced by the coupling of the sub- and supraglottal tracts through the time-varying glottal opening. Despite these problems, inverse filtering is very popular and researchers have successfully used this technique in many applications.

Clearly, the problem of accurately and robustly modeling speech production is far from solved. The limited amount of information from which estimates must be made necessarily limits the accuracy and robustness of our estimates. It may be possible to improve upon the results of inverse filtering by further parameterizing the speech production system. One approach to further parameterizing the speech production system is to simultaneously estimate the vocal tract and glottal source parameters. This has been done in an iterative manner in [6]. Another possibility is to parameterize the results of inverse filtering [5]. The remainder of this document discusses such a parameterization of the glottal source waveform using the LF-model [1]. Electroglottography (EGG) is used to determine the instant of closure and information about opening location. Inverse filtering is performed based on this information and then the LF-model is fit to the inverse filtered waveform using three different fitting techniques.

## 2. MODELING THE INVERSE FILTER WAVEFORM

A number of glottal flow models have been proposed [6] of varying detail and complexity. Several model the smoothed derivative of the glottal pulse waveform, referred to as differentiated glottal volume velocity (DGVV). The models describe the DGVV with two curves, for the open and closed glottal phases.

The LF-model of Fant, Liljencrants and Lin [1] describes the DGVV waveform in terms of an exponentially growing sinusoid in the open phase and a decaying exponential in the closed phase.

$$g(t) = \begin{cases} E_0 e^{\alpha t} \sin \omega_g t, & t < T_e \\ -\frac{E_0}{\epsilon T_a}\left[e^{-\epsilon(t-T_e)} - e^{-\epsilon(t-T_e)}\right], & T_e < t < T_c \end{cases} \quad (1)$$

The model is expressed in terms of five parameters. A



0-7803-0946-4/93 $3.00 

slightly modified version of the model [7] expressed in discrete-time is as follows.

$$q(n) = \begin{cases} A_{go}e^{\alpha_{go}n}\sin(\omega_{go}n + \phi_{go}), & n = 0,\ldots,N-1 \\ -A_{gc}e^{-\alpha_{gc}(n-N)}, & n = N,\ldots,M-1 \end{cases} \quad (2)$$

The addition of the phase term $\phi_{go}$ to the open phase sinusoid enables the model to be expressed neatly in terms of complex exponentials

$$q(n) = \begin{cases} C_o z_{go}^n + C_o^*(z_{go}^*)^n, & n = 0,\ldots,N-1 \\ C_c z_{gc}^{n-N}, & n = N,\ldots,M-1 \end{cases} \quad (3)$$

where

$$C_o = 0.5A_{go}e^{j(\phi_{go}-\pi/2)}, \quad (4)$$

$$z_{go} = e^{\alpha_{go}+j\omega_{go}}, \quad (5)$$

$$C_c = -A_{gc}, \quad (6)$$

and

$$z_{gc} = e^{-\alpha_{gc}} \quad (7)$$

We expect the estimated DGVV obtained through inverse filtering to resemble the shape of the LF-model waveform. Therefore, by fitting the LF-model to the inverse filter output, noise may be reduced and other inaccuracies inherent in the inverse filtering technique and its simplifying assumptions may be improved. Also, we get a tractable parametric model for the glottal source useful in synthesis and coding.

## 3. LF-MODEL PARAMETER ESTIMATION

### 3.1. Prony-Based Techniques

In 1795, Prony [8] proposed a method for solving for the parameters of equations of the form

$$x(n) = \sum_{i=0}^{N} \alpha_i e^{\beta_i n} \quad (8)$$

The technique, now known as Prony's method, decomposes the problem into two sets of linear equations. First, the modes $\beta_i$ of the signal are calculated, and then the residues $\alpha_i$ are calculated. Both $\alpha_i$ and $\beta_i$ are complex. The original method of Prony does not perform well in noise. As a result, much work has been done [9, 10] to improve upon Prony estimation in noise. These methods improve estimation typically by employing some form of noise cleaning prior to Prony estimation, and overestimating the initial number of modes.

The modified LF-model of Equation 3 is in the form of a sum of complex exponentials in both the closed and open phases. Therefore, Prony techniques can be applied in each phase to fit these equations to the inverse filter output.

Prony's method applied directly to the open and closed phases of the vowel /a/ inverse filtered using EGG data to identify the closed phase intervals is shown in Figure 1. The direct use of Prony's method provides a good fit around the glottal opening although noticeable errors occur immediately before the estimated glottal opening. In the final three pitch periods, there are noticeable discontinuities in the LF-model fit at glottal opening. The presence of the phase term $\phi_{go}$ in the modified equations (2) allows for a

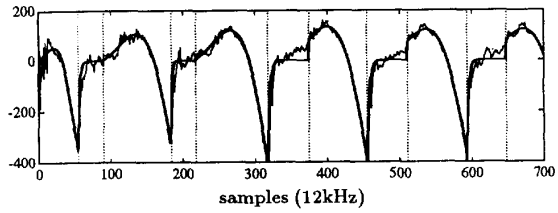


Figure 1. Direct Prony Estimation: LF-model fit (dark line) to inverse filter waveform.

non-zero phase at the estimated glottal opening location resulting in a discontinuous closed/open phase boundary. These discontinuities are most likely the result of poor opening location estimates from the EGG.

During closed phase estimation, the direct use of Prony's method can produce a complex mode. Since the LF-model requires only one real mode for the closed phase, a complex mode is unacceptable. In the majority of cases, the angle of the estimated mode is very small or zero and the absolute value of the mode is therefore used, with the angle set to zero.

The Prony-based method assumes that the instant of opening is accurate. This may not be the case. Notice that direct Prony estimation provides a very good fit over the open phase. If the estimated opening location is disregarded and the damped sinusoid is simply extend backwards in time until it reaches zero, the good model fit in the estimated open phase will remain and perhaps a better fit of the inverse filter waveform at the opening "knee" will result. This technique, hereafter referred to as the extended Prony method, produces results for the sample vowel as seen in Figure 2.

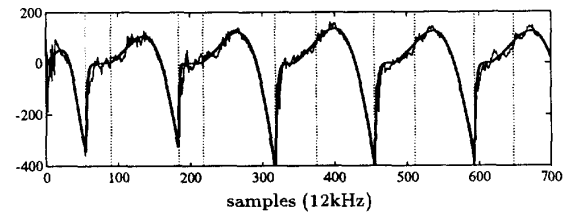


Figure 2. Extended Prony Estimation: LF-model fit (dark line) to inverse filter waveform.

### 3.2. Gradient Descent Technique

Extended Prony estimation clearly provides the most aesthetic fit for the given vowel segment, but as yet we have no evidence that the opening locations implicitly determined by the method are accurate. If we know or choose the opening location prior to estimation, then neither of the above methods will guarantee a suitable fit. In situations such as this, iterative estimation methods are required.

Gradient descent, a very common iterative search technique, can be used to estimate the open phase LF-model parameters that best fit the inverse filter waveform in a least squares sense. As implemented in this work, glottal closing location must be specified. In other implementa-

tions, it may be possible to search opening location as well. One of the major advantages of gradient descent is that since it is a search technique, the parameter range can be constrained. We know the range of acceptable LF-model parameters and we can constrain the search to only these ranges regardless of the noise content of the inverse filter waveform.

Gradient descent is applied to the sample waveform to produce the fit shown in Figure 3. Gradient descent per-

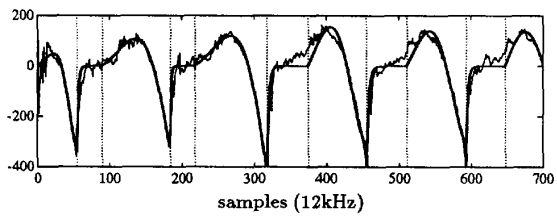


Figure 3. Gradient Descent: LF-model fit (dark line) to inverse filter waveform.

forms well and produces results consistent with the original LF-model equations 1. Some hand tuning of the gradient descent solution is necessary to avoid convergence to poor local minima. It is tempting to say that the EGG opening locations for the first two pitch periods are near the correct ones since the gradient descent solution fits rather well. This is also supported by the fact that the implicit opening location found by extended Prony estimation coincides with this location. The results appear to suggest that extended Prony estimation may be a good technique for estimating opening locations, but the application of this method to diverse sounds tends to discredit this conclusion.

## 4. EVALUATION

The previous section presented three techniques for estimating LF-model parameters from an inverse filter waveform. The following section attempts to evaluate the utility of these three methods for the range of inverse filter waveforms that may be encountered.

### 4.1. Synthetic Glottal Flow Waveforms

Inverse filtering of speech is known to produce noisy estimates of glottal flow. In cases where the assumptions upon which inverse filtering is based are not valid, inaccurate waveforms can result. Due to the extreme variability in inverse filtering estimates, it is useful to study the estimation of LF-model parameters independent of errors introduced by inverse filtering. Synthetic glottal flow waveforms that incorporate source-tract interaction effects can be used as "clean" inverse filter results.

We generated a number of synthetic DGVV waveforms for the vowel /i/ that included a varying amount of source-tract interaction using the technique described in [11]. An example of a synthetic waveform and its LF-model fits using the three methods can be seen in Figure 4. It is clear that the LF-model does not represent the phenomena of source-tract interaction well. Even waveforms without source-tract interaction differ noticeably in the open phase from the exponentially growing sinusoid of the LF-model. Regardless,

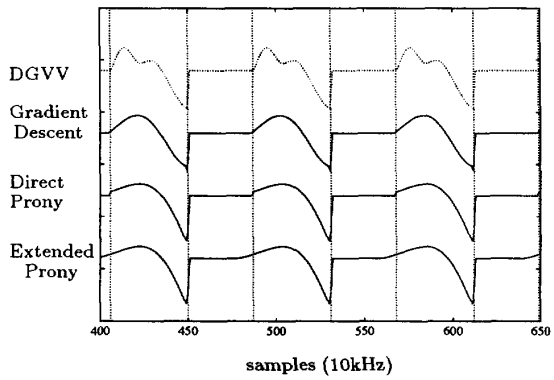


Figure 4. LF-model fit to synthetic glottal flow waveform. (with source-tract interaction effects)

the LF-model does very simply represent the general shape of the differentiated glottal volume velocity waveform. So the question still remains; do the three estimation methods produce a good fit to the waveform? In general, for different open quotient values and degrees of source-tract interaction, the fits are good. Inadequacies do occur at the end of the open phase where the angle of the sinusoid extends into the fourth quadrant. While this is acceptable within the scope of the LF-model, it is not appropriate for this application. Constraints placed on the gradient descent technique can be used to ensure that this does not occur. The extended Prony technique in general, extrapolates near the correct opening location although slightly before the EGG derived opening location.

### 4.2. Examples on Continuous Speech

With some idea of how the estimation techniques perform on clean glottal flow waveforms, inverse filtered glottal flow waveforms will be studied. Obviously, an estimation technique should produce good fits for clean glottal flow waveforms, but what performance is desirable for waveforms that clearly do not resemble the expected glottal flow shape? The answer to that question is application specific.

Figures 5 and 6 demonstrate LF-model fits to reasonably "clean" inverse filter waveforms. Gradient descent clearly produces the most pleasing fit. It is important to reiterate that gradient descent requires some human interaction to avoid convergence to poor local minima while the two Prony methods are purely automatic. The figures illustrate that the Prony-based methods' problems arise from the additional phase term in the LF-model sinusoid. The inverse filter waveform typically diverges from the expect LF-model shape around the opening location causing the problems with the Prony techniques. The figures do demonstrate a few periods where the extended Prony method does break down. In these cases the estimated open phase waveform never crosses zero before reaching the closing location of the previous pulse. In these cases, no extension is performed and the extended Prony produces the same result as direct Prony.

When the inverse filter waveform is very noisy, Prony

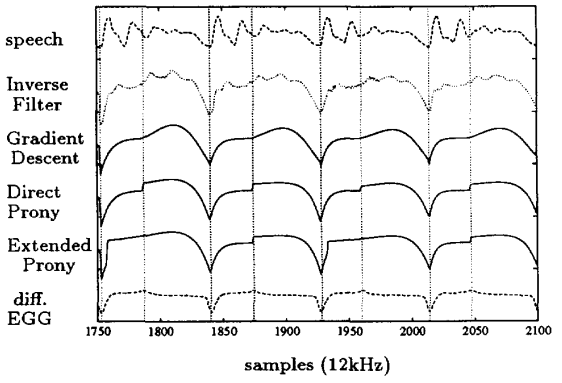


Figure 5. LF-model fit to continuous speech segment by a male speaker.

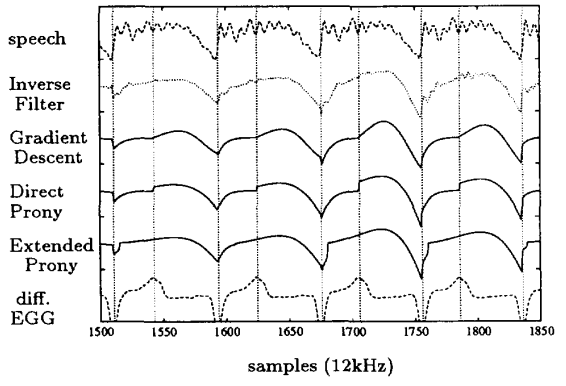


Figure 6. LF-model fit to continuous speech segment by a different male speaker.

techniques can generate high frequency modes that are clearly incorrect. The use of more noise resistant Prony techniques allows this problem to be avoided in many cases. By overestimating the number of modes, modes that are too high in frequency can be removed and the best lower frequency modes kept.

## 5. CONCLUSIONS

Although Prony-based techniques provide an appealing analytic way for fitting the LF-model to inverse filtered waveforms, they are outperformed by gradient descent techniques. As experiments with synthetic waveforms suggest, both techniques produce reasonable fits to clean glottal flow waveforms. But in real speech, where inverse filtering is known to produce waveforms that marginally represent that of the LF-model, the results of Prony techniques are inferior to that of gradient descent.

## REFERENCES

[1] G. Fant, J. Liljencrants, and Q. Lin, "A four parameter model of glottal flow," *STL-QPSR 4/1985*, pp. 1–13, 1985. Presented at French-Swedish Symposium, Grenoble, April 22-24, 1985.

[2] D. G. Childers and C. K. Lee, "Vocal quality factors: Analysis, synthesis, and perception," *Journal of the Acoustical Society of America*, vol. 90, no. 5, pp. 2394–2410, 1991.

[3] D. H. Klatt and L. C. Klatt, "Analysis, synthesis and perception of voice quality variations among male and female talkers," *Journal of the Acoustical Society of America*, vol. 87, pp. 820–857, February 1990.

[4] D. J. Wong, J. D. Markel, and A. H. Gray, "Least squares glottal inverse filtering from the acoustic speech wave," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-27, pp. 350–355, August 1979.

[5] J. de Veth, B. Cranen, and H. Strik, "Extraction of control parameters for the voice source in a text-to-speech system," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 301–304, April 1990.

[6] H. Fujisaki and M. Ljungqvist, "Proposal and evaluation of models for the glottal source waveform," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, (Tokyo, Japan), pp. 1605–1608, 1986.

[7] A. K. Krishnamurthy, "Glottal source estimation using a sum-of-exponentials model," *IEEE Transactions on Signal Processing*, vol. 40, pp. 682–686, March 1992.

[8] G. de Prony, "Essai expérimental et analytique: sur les lois de la dilabilité de fluides élastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alkool á différentes températures," *Journal École Poytechnique*, vol. 1, no. 2, pp. 22–76, 1795.

[9] R. Kumaresan, D. W. Tufts, and L. L. Scharf, "A prony method for noisy data: Choosing the signal components and selecting the order in exponential signal models," *Proceedings of the IEEE*, vol. 72, pp. 230–233, February 1984.

[10] M. A. Rahman and K.-B. Yu, "Total least squares approach for frequency estimation using linear prediction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-35, pp. 1440–1454, October 1987.

[11] T. V. Anathapadmanabha and G. Fant, "Calculation of true glottal flow and its components," *Speech Communication*, vol. 1, pp. 167–184, 1982.