On the Appearance of a Positive Real Pole in the Results of Glottal Closed Phase Linear Prediction

Alan O’Cinneide  
*Technological University Dublin*, alan.ocinneide@tudublin.ie

David Dorran  
*Technological University Dublin*, david.dorran@tudublin.ie

Mikel Gainza  
*Technological University Dublin*, Mikel.Gainza@tudublin.ie

Eugene Coyle  
*Technological University Dublin*, Eugene.Coyle@tudublin.ie

Follow this and additional works at: https://arrow.tudublin.ie/argcon

Part of the *Signal Processing Commons*

**Recommended Citation**


*This Conference Paper is brought to you for free and open access by the Audio Research Group at ARROW@TU Dublin. It has been accepted for inclusion in Conference papers by an authorized administrator of ARROW@TU Dublin. For more information, please contact yvonne.desmond@tudublin.ie, arrow.admin@tudublin.ie, brian.widdis@tudublin.ie.*

This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/).
ON THE APPEARANCE OF A POSITIVE REAL POLE IN THE RESULTS OF GLOTTAL CLOSED PHASE LINEAR PREDICTION

Alan Ó Cinnéide, David Dorran, Mikel Gainza and Eugene Coyle

Audio Research Group, Dublin Institute of Technology
Kevin Street, Dublin 8, Republic of Ireland
phone: + (353) 1 402 4726, email: alan.ocinneide@dit.ie
web: www.audioresearchgroup.com

ABSTRACT

Often when performing glottal closed phase covariance linear prediction, a positive real pole can appear in the resulting filter transfer function. The commonly adopted approach is to discard this pole, as it does not fit with the usual model of the all-pole vocal tract filter. However, this real pole describes some aspect of the speech signal; this paper provides a novel perspective on its occurrence. This viewpoint has a useful implication to the speech community, especially from the perspective of fitting a glottal pulse to the inverse filtered signal, as the real pole describes the return phase of the glottal flow for certain voice types that adhere to a reasonable criterion. Tests with synthetic signals are performed to validate this approach.

1. INTRODUCTION

Glottal closed phase linear prediction [1] is a speech analysis technique for estimating the parameters of the vocal tract filter based on the linear source filter theory of speech production. The method has been shown to have effective formant tracking abilities when compared to some other inverse filtering methods [2] and found application in voice quality analysis [4], speaker identification [3] and analysis of spoken prosody [5]. The technique assumes that there exists a region within the speech signal where the glottis is closed and leaks no contribution into the speech signal.

However, in practice the resulting solution rarely yields the vocal tract parameters directly. Following analysis, the filter polynomial is factorized to determine the locations of its poles on the Z-plane, whereupon any pole that appears on the positive real axis is removed [1] [4] [6]. In [6], Alku et al. remark that “poles on the positive real axis of the Z-plane are] unrealistic from the point of view of Fant’s source tract theory of vowel production and its underlying theory of tube modeling”. Because the theory cannot rationally associate this pole with the vocal tract, it is discarded and the remaining poles are recomposed into the vocal tract filter of reduced filter order. Failure to remove this pole can lead to distortions in the time domain signal around the instant of glottal closure called “jags” [1] [6] (see Figure 1). Indeed, the development of DC-constrained closed phase linear prediction [6] was in part motivated to increase the likelihood that closed phase analysis will yield pole locations at more realistic Z-domain coordinates.

Wong et al. [1] offer a number of explanations for the appearance of these real poles. They may appear due to the intrusion of low frequency recording noise, a non-zero mean in the analysis windows and/or the over-specification of the filter order. In the case of an ideal closed phase, these reasons cannot be disputed. However, this paper will illustrate that for reasons related to the identification of the location of the glottal closed region, certain voice types will also cause the appearance of such a pole. While the usual approach to discard this pole is appropriate when only the parameters of the vocal tract are desired, it is shown here that the pole has a useful significance in the parameterization of the glottal source.

This paper is outlined as follows: the following section gives the necessary background of the acoustic theory of speech production, glottal closed phase linear prediction and its implementation.

It will be shown that, in practical application, a positive real pole in closed phase analysis should be expected for certain voice types. The third section discusses relationship between the Z-plane position of the pole and the return phase of glottal models, with specific focus on the Liijencrants-Fant (LF) model [7]. Experiments which validates the theory are described in the fourth section. The fifth section discusses the results yielded by these experiments. Conclusions are drawn in the final section, which also outlines some directions for future research.

2. BACKGROUND

2.1 Acoustic Theory of Speech Production

The acoustic theory of speech production [8] views speech as the convolution of glottal flow signal with a vocal tract filter which is then radiated at the lips. In the Z-domain, the process can be represented as follows:

\[ S(z) = G(z)V(z)L(z) \]

where \( S(z) \) represents the speech waveform, \( G(z) \) the glottal flow, \( V(z) \) the vocal tract filter, and \( L(z) \) represents lip radiation.

As lip radiation \( L(z) \) is usually modeled as a differentiating filter and the relationship between the speech chain components assumed linear, it is often combined with the glottal flow \( G(z) \) to form the derivative glottal flow \( G'(z) \). This reduces the number of elements in the speech production process to two:

\[ S(z) = G'(z)V(z) \] (1)
tion, the build-up of air pressure from the lungs sets the glos-
tis into the speech signal [1]. Therefore, during this inter-

cessive glottal pulses, there exists no exogenous contribu-
tion from within the larynx into a quasi-periodic cycle of opening and clos-

Wong [1] and is briefly recapitulated here. During voiced pho-

2.2 Glottal Closed Phase Covariance Linear Prediction

Closed phase inverse filtering was first theoretically outlined by

Figure 2: The figure above shows a synthetic speech signal \( s(n) \) generated by an all-pole filter excited by a derivative glottal pulse train \( g'(n) \). The closed phase of a pulse and the corre-

sponding region in the speech has been highlighted.

2.3 Detecting the Closed Phase of the Glottal Cycle

As illustrated in Figure 3 by the LF model, the glottal cycle is often

described in three phases [15]:

- the open phase, during which the glottis opens.
- the return phase, the interval when the glottis proceeds to close.
- the aforementioned closed phase, representing the time during

which the glottis is closed and there is no glottal excitation.

One of the main difficulties with closed phase inverse filtering re-
lates to the determination of this closed phase from the speech signal [6]. This problem can be overcome by the analysis of the signal of an electroglottograph (EGG) which has been recorded in tandem with the speech signal [2]. In the absence of such a signal, estimates of the glottal closed phase can be determined from the speech signal directly by a number of different methods [12] [13] [14].

However, rather than the instant of glottal closure, closed phase
detection methods often search for the instant of greatest excitation of the speech signal; this instant is sometimes called the speech epoch and is marked in Figure 3 as \( t_e \). For those voice types that exhibit an instantaneous closure, the closed phase of the glottal cy-

cle will indeed begin in the sample following this point. However, for other voice types, the sample after this point can often signify the beginning of the glottal cycle’s return phase. Should the return phase be inadvertently included in the interval for closed phase anal-

ysis, its time domain shape will affect the vocal tract filter param-

ers and introduce unanticipated elements. In many cases, these deviations will appear as a positive real pole, as discussed below.

3. THE GLOTTAL RETURN PHASE AS A SINGLE POLE IIR FILTER

The return phase of the cycle is an important perceptual aspect of the glottal flow [15] as it determines the spectral slope of the source and thus the amount of high-frequency energy present in the spectrum. An instantaneous closure would cause the most sudden time domain clip into the signal, imparting the maximum high frequency content. Similarly, glottal pulses with more gradual closures exhibit more attenuated upper harmonics. Many glottal models represent their return phases as an asymptotic exponential decay from a negative maximum to zero; this type of segment is fully parameterized by the impulse response of a single pole, low pass filter.

During glottal closed phase analysis, linear prediction does not differentiate between the source and filter contributions of speech. The results of any attempt to model the speech signal during an analysis interval that may also include a portion of the return phase will be affected in some way by all signal elements. Thus, in the cases where the return phase of the signal can be modeled as an asymptotic exponential decay, it is unsurprising to see a positive real pole appear in the result of the linear predictive analysis.

The real pole describing the return phase of glottal flow can be used to infer the parameters used in the formulation of glo-
tal models as the Z-plane position of the pole is a direct indica-
tion of the graduality of glottal closure. Section 3.1 will illustrate the mathematical relationship between the real pole and the return phase whose return phases approximate an exponential function is described. As an illustrative example, a method for determining the return phase parameter of the prevalent LF model is also given. Similar relationships can be established to the return phase param-

eters of other glottal models which obey the same basic premise.

3.1 The Glottal Return Phase as the Impulse Response of a Positive Real Pole

Exponential type functions are often used to model the return phase of glottal models, e.g. the LF and the KLGLOTT88 models. Referring to the time domain formulation of the return phase as \( g_{ret}' \), the normalized return phase of such a model can be expressed mathe-
matically by the following equation:

\[
g_{ret}'(n) = \mu^n u(n)
\]

(2)

where \( u(n) \) represents the unit step function and \( \mu \) is the base of the exponential. Arbitrarily beginning the return phase at \( n = 0 \), its Z-transform \( G_{ret}'(z) \) can be shown to be:

\[
G_{ret}'(z) = \sum_{n=0}^{\infty} \mu^n u(n) z^{-n} = 1 + \mu z^{-1} + \mu^2 z^{-2} + \mu^3 z^{-3} + \ldots = \frac{1}{1 - \mu z^{-1}} = \frac{z}{z - \mu}
\]

Thus, the return phase of derivative glottal signals where (2) holds can be modeled as the impulse response of a single pole IIR filter. The Z-plane amplitude of the pole \( \mu \) therefore reflects the rate of exponential decay of the return phase.

The signal amplitude independence of the relationship between the pole amplitude \( \mu \) and the return phase parameters implies the inclusion of any part of the return phase segment will theoretically yield the same positive real pole in the analysis results so long as (2) is valid.

3.2 Method to Determine the Return Phase Parameter of the LF Model

The LF model [7] represents the general flow shape of the glo-
tal flow derivative over one glottal cycle and whose shape can be
The total number of samples in the pulse is the pitch period, referred to as the return phase length and the exponential base to ensure that the curve reaches null at the point zero at a rate determined by the steepness of the slope of the tangent at an arbitrary setting of the value of \( t_e \) to be 0, the equation of the tangent line can then be shown to be:

\[
y = (-E_e \ln \mu)x - E_e
\]

Thus, the value of \( T_0 \) can be calculated by solving (5) at \( y = 0 \), which yields the following identity:

\[
T_0 = -\frac{1}{\ln \mu}
\]

### 4. EXPERIMENT

An experiment was undertaken to validate the theory that the return phase parameter \( T_0 \) can be accurately estimated from the real pole which appears in the analysis results of glottal closed phase covariance linear prediction. Using a sampling rate of 10kHz, various vocal tract filters were convolved with an LF model pulse train of varied configurations to create voiced synthetic speech segments, in accord with the acoustic theory of speech production given in (1). Synthetic speech pulses were used for validating the theory due to the inherent lack of reference parameters in actual speech.

The LF model pulses were generated using all parameter combinations given in Table 1. The relationships between the utilized shape parameters and the LF model timing parameters given here:

\[
O_q = \frac{t_e}{t_0}, \quad \alpha_0 = \frac{t_p}{t_e}, \quad Q_0 = \frac{T_0}{(1 - O_q)T_0}
\]

where \( O_q \) is the open quotient of the pulse, \( \alpha_0 \) its asymmetry coefficient, and \( Q_0 \) its return phase coefficient.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_0 )</td>
<td>80 : 20 : 200 (Hz)</td>
</tr>
<tr>
<td>( O_q )</td>
<td>0.3 : 0.05 : 0.9</td>
</tr>
<tr>
<td>( \alpha_0 )</td>
<td>0.67 : 0.05 : 0.9</td>
</tr>
<tr>
<td>( Q_0 )</td>
<td>0.01 : 0.05 : 1</td>
</tr>
</tbody>
</table>

Table 1: All LF model parameter configurations used for synthetic testing.

Following the recommendations of [7], when producing the LF pulses, the return phase spanned from \( t_e \) to \( t_0 \), of the following pulse such that the instants \( t_e \) and \( t_c \) coincide between adjacent pulses.

Covariance linear predictive analysis was performed according to the guidelines laid down by Wong [1]. Care was taken to ensure that the interval to be minimized extends from one sample following the instant of glottal closure \( n_c \) to one sample before glottal opening \( n_o \). The instant of glottal closure in all cases was chosen to be the point marking the beginning of the return phase. However, unlike Wong [1], a pre-emphasis operation is not performed for two reasons. Firstly, it was noted in [2] that pre-emphasis makes very little difference to the derived parameters. Secondly, and more acutely in the context of this work, a pre-emphasis operation alters the return phase in such a way that the positive real pole located after analysis would not bear the same relationship to the return phase as the one outlined in this paper.

The value of \( p \) is usually chosen by a “rule of thumb” derived from acoustic tube modeling [9] [11]. This rule follows from the relationship between \( p \) and the length of the average male vocal tract, the speed of sound \( c \) and sampling frequency \( f_s \):

\[
p = \frac{f_s}{1000}
\]

Thus, for a sampling rate of 10kHz, \( p \) is 10. However, as the analysis is also intended to capture the real pole describing the return phase, the \( p \) is incremented by one to 11.
Of the derivative glottal pulse trains generated, some were excluded from the results for two reasons:

- At least $p$ linear equations are required to ensure convergence of the linear prediction equations; any glottal source wave configuration that prevented this requirement could not undergo analysis. This excluded 13.19% of configurations.
- Occasionally the covariance analysis did not yield a single positive real pole, meaning that $T_a$ could not be estimated by the outlined method. This excluded 5.38% of configurations.

In all, a total of 185,260 experiments yielded analyzable results.

The error measured by the signal is the percentage error of the $T_a$ parameter, calculated according to the formula:

$$E_{T_a} = 100 \frac{T_{a_{est}} - T_a}{T_a}$$  \hspace{1cm} (7)

where $T_{a_{est}}$ has been calculated according to (6).

The routine utilized to generate the LF model pulses requires integer period lengths. Because of this, fundamental frequency values are slightly different than the ones listed: period lengths are rounded to the nearest integer $T_0 = \text{int}(\frac{f_0}{f})$.

In order to test the relationship in more realistic situations, two other scenarios were also tested: the case where the system is corrupted by varying levels of amplitude modulated Gaussian noise, and the scenario where the system is noiseless but the glottal closing instant is offset by a certain number of samples.

### 5. DISCUSSION OF RESULTS

The graph shown in Figure 4 shows the distribution of the $T_a$ parameter percentage error determined by the method described in Section 3.2. The diagram shows that for small return phase coefficients, i.e. small $T_a$ values, the percentage error is low, with a large number of cases exhibiting errors near 0%.

The graph indicates that as the return phase coefficient of the tested pulses increases the percentage error in the estimation of the $T_a$ coefficient also grows larger. This positive correlation was expected as the relationship given by (6) was derived from the premise that the return phase is an exponential function. As previously mentioned, this segment is not a true exponential, due to the offset required for a null value at $t_a$. The value of this offset can be directly calculated from the LF model return phase (3):

$$LF_{offset} = \frac{-E_a e^{-\alpha(T_e-T_c)}}{\epsilon T_a}$$  \hspace{1cm} (8)

From (8) above, it can be seen that large offsets occur when the length of the return phase (calculated as $(T_e - T_c)$) is short in duration and the $T_a$ parameter is large. In these extreme cases, it seems that the return phase would be more appropriately modeled by some other mathematical function, rather than an exponential.

As is evidenced by the diagram, the method outlined in this work is most successful when parameterizing return phases resulting from small $Q_a$ values. It has been noted in [16] that, for real speech, normal $T_a$ values tend to be small. In order to support this claim, the $Q_a$ values for several voice types were calculated from the typical parameter values given in [17]. These values were obtained from the analysis of data and speech synthesis experiments in voice conversion, and can be seen in Table 2:

<table>
<thead>
<tr>
<th>Voice</th>
<th>$Q_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>modal</td>
<td>0.001</td>
</tr>
<tr>
<td>vocal fry</td>
<td>0.08</td>
</tr>
<tr>
<td>breathy</td>
<td>0.07</td>
</tr>
<tr>
<td>falsetto</td>
<td>0.4</td>
</tr>
<tr>
<td>harsh</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 2: The $Q_a$ coefficients of several voice types, from [17].

Modal, vocal fry, breathy and harsh voice types all have small $Q_a$ parameters; Figure 4 suggests that these values would introduce little percentage error. Only falsetto voices types, where $Q_a = 0.4$ could errors become significant. Informal perceptual testing performed by the first author confirmed this finding.

Inappropriately modeling the return phase influences the ability of the model to correctly determine the vocal tract parameters. In cases where the return phase is significantly different from an exponential, false poles at very low frequencies (usually less than 200Hz) may appear. This low resonance shifts the estimated vocal tract formant center frequency and bandwidth values in a manner that is difficult to predict. This observation seems to contradict the heuristical rule mentioned in [4] where any root below 250Hz is removed. In those experimental scenarios where low poles were observed to occur in this work, the remaining resonances are not representative of the vocal tract. Although, as this paper confirms, it is reasonable to remove a single positive real pole, it is not obvious why removing other low frequency resonances would be reasonable. However, the appearance of such poles may be an indication that closed phase covariance linear prediction is a technique unsuited to the analysis of that particular speech period.

The results given in Figures 5 and 6 reflect the sensitivity of the relationship between the real pole and the $T_a$ parameter to noise and location of the analysis interval to the glottal closing instant respectively. As would be expected, the greater the noise level within the signal the less applicable the method, though still offering reasonable accuracy for certain voice types with high signal-to-noise ratios (SNRs). In the cases where the analysis interval is misplaced due to an inaccurately detected glottal closing instant, the relationship is fairly robust to any positive offset. However, as can be seen in lower sub-figures of Figure 6, any misplacement which includes any portion of the open phase offers renders the method virtually useless: any real pole detected tends to be quite near the unit circle and not representative of the slope of the signal. In fact, oftentimes two real poles or a very low frequency complex conjugate pair appear - anomalies of this type may be used as an indication of misplacement.

### 6. CONCLUSIONS AND FUTURE WORK

In this work, the issue of the positive real pole that sometimes appears in the results of closed phase covariance linear prediction is
highlighted. Because this pole is unexpected from the point of view of the acoustic theory of speech production, the common approach to deal with this pole is to simply discard it. Rather than discard the speech signal information implied by the presence of this pole, this paper has illustrated that it is an indication of time domain shape of the return phase of the glottal cycle in the common situation where the phase can be approximated by an exponential function and a portion of the return phase is wrongly attributed to the glottal closed region. Specific focus was placed upon the prevalent LF model of derivative glottal flow, whose $T_c$ parameter was shown to be modeled by a simple mathematical relationship with the pole position.

In order to validate this theory behind this relationship, various experiments were undertaken for a diverse set of synthetic voice types. Due to the difficult nature of applying the theory developed within this work to real speech, the experiments have been necessarily confined to synthetic speech. However, as the LF model has shown to be a suitable model for realistic derivative glottal source waveforms [7], the techniques described within this work can theoretically be applied to real-world scenarios. Depending on the characteristics of the voice, it was shown that the return phase of the underlying glottal model can be parameterized with high accuracy.

The techniques outlined in this work also implies that the glottal return phase can be parameterized by applying a single order covariance method linear prediction directly. Future work includes using the techniques above to develop a novel method of glottal source parameterization, and to extend the technique in order to more successfully handle return phases that deviate from an ideal exponential function. This would included an exploration the failure of closed phase inverse filtering in the cases which produces a very low frequency pole in the analysis results.

REFERENCES


Figure 5: The percentage error of the $T_c$ estimate predicted by the real pole found by analysis in the case where Gaussian noise is added to and modulated by the source signal at different SNRs.

Figure 6: The percentage error of the $T_c$ estimate predicted by the real pole found by analysis in the case the instant of glottal closure is offset by a number of samples.