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.

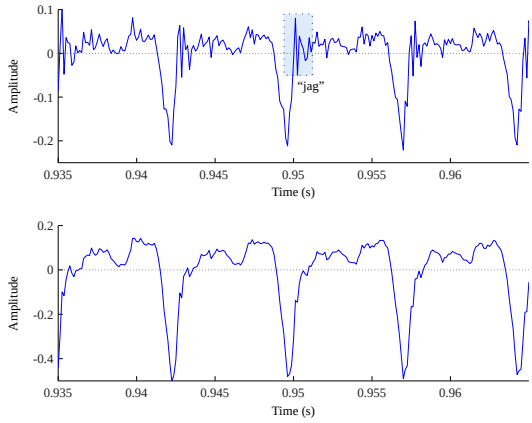

Figure 1: A diagram illustrating the "jags" that occur when the positive real pole remains in the speech signal. Above, glottal derivative source with "jag" of middle pulse highlighted. Below, glottal source signal without "jag" distortion.

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 Liljencrants-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:

$$S(z) = G'(z)V(z) \quad (1)$$

 ISSN 2076-1465

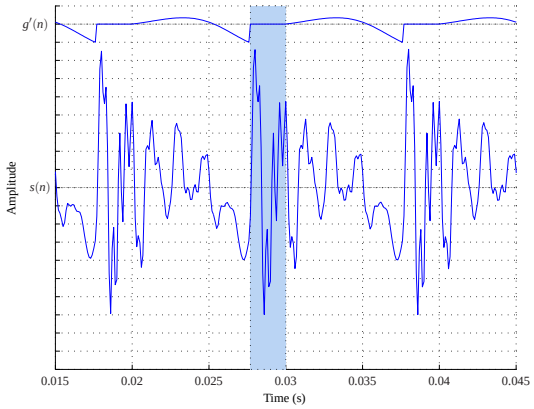

Figure 2: The figure above shows a synthetic speech signal $s(n)$ generated by an all-pole filter excited by a derivative glottal pulse train $g'(n)$. The closed phase of a pulse and the corresponding region in the speech has been highlighted.

2.2 Glottal Closed Phase Covariance Linear Prediction

Closed phase inverse filtering was first theoretically outlined by Wong [1] and is briefly recapitulated here. During voiced phonation, the build-up of air pressure from the lungs sets the glottis within the larynx into a quasi-periodic cycle of opening and closing. Individual pulses of air excite the vocal tract and radiate at the lips, which generates speech.

During the ideal glottal closed phase condition i.e. between successive glottal pulses, there exists no exogenous contribution from the glottis into the speech signal [1]. Therefore, during this interval, the speech signal results solely from the decaying vocal tract resonances, as in Figure 2. As these resonances are assumed to be the result of an all-pole system, the speech signal's closed phase can theoretically be fully described by these all-pole coefficients during this interval.

A suitable method for determining an all-pole filter's coefficients from its output is covariance method linear prediction [11]. This technique ascertains the filter parameters over finite intervals by minimizing the energy of the residual of the analyzed signal.

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 relates 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 cycle 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 analysis, its time domain shape will affect the vocal tract filter parameters 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 glottal models as the Z-plane position of the pole is a direct indication 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 parameters 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 mathematically by the following equation:

$$g'_{ret}(n) = \mu^n u(n) \tag{2}$$

where $u(n)$ represents the unit step function and μ 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:

$$\begin{aligned} G'_{ret}(z) &= \sum_{-\infty}^{\infty} \mu^n u(n) z^{-n} \\ &= 1 + \mu z^{-1} + \mu^2 z^{-2} + \mu^3 z^{-3} + \cdots \\ &= \frac{1}{1-\mu z^{-1}} = \frac{z}{z-\mu} \end{aligned}$$

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 μ therefore reflects the rate of exponential decay of the return phase.

The signal amplitude independence of the relationship between the pole amplitude μ 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 glottal flow derivative over one glottal cycle and whose shape can be

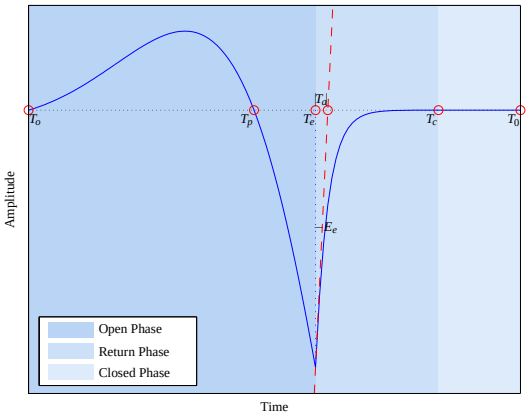

Figure 3: An LF model of derivative glottal flow, with timing parameters $(T_o, T_e, T_a, T_c, T_p)$ and amplitude parameter E_e. Also marked are the different phases of the glottal cycle and the tangent at $(T_e, -E_e)$ which defines T_a.

uniquely described with four parameters. The mathematical formula describing the LF model is a piece-wise function, consisting of two segments, the evolution of which can be seen in Figure 3. The first segment is an exponentially increasing sine function, characterizing the glottal flow derivative from the instant of glottal opening t_o, through the time axis at t_p, to the instant of maximum negative extreme at t_e. At this point the second segment of the LF model, often referred to as the return phase, begins. This portion models the glottal closure as a modified exponential function which returns to zero at a rate determined by the steepness of the slope of the tangent to the function at t_e. The distance of this tangent's time axis intercept from t_e is called T_a, and is referred to as the effective duration of the return phase. The total number of samples in the pulse is the pitch period, referred to as T_0.

In order to correctly place the pulse in time, the timing instants are calculated to be relative to the instant of glottal opening, i.e. $T_o = 0$, $T_p = t_p - t_o$, $T_e = t_e - t_o$ and $T_c = t_c - t_o$. Below are the mathematical equations describing time domain LF model shape using these parameters:

$$u_{LF}(n) = \begin{cases} E_0 e^{\alpha n} \sin \omega_g n & \text{for } 0 \le n < T_e \\ \frac{-E_e}{\varepsilon T_a}(e^{-\varepsilon(n-T_e)} - e^{-\varepsilon(T_c-T_e)}) & \text{for } T_e \le n \le T_c \\ 0 & \text{for } T_c \le n < T_0 \end{cases} \quad (3)$$

The return phase of the model deviates from a true exponential function in that an offset is added to the value of the exponential to ensure that the curve reaches null at the point t_c. Depending on the return phase length and the exponential base μ, the value of this offset can be negligibly small such that an exponential function very closely approximates that the LF return phase. Indeed, Fant [7] notes that in practice, it is convenient to set T_c equal to T_0 as the energy difference between the exponential during this interval and the ideal closed phase is negligible. In those cases, the general LF return phase can be assumed to be equivalent to a scaled version of the exponential function given above in (2).

$$\begin{aligned} u_{LF}(n) &= -E_e g'_{ret}(n) \text{ for } T_e \le n < T_c \\ &= -E_e \mu^n \end{aligned} \quad (4)$$

In order to determine T_a of such a return phase, calculus and linear geometry can be used. First, differentiating (4) yields the slope of the general tangent to the exponential return phase:

$$m = -E_e \mu^n \ln \mu$$

Referring to the time and amplitude axes as the x and y axes respectively, the slope and y-intercept of the tangent at the point $(t_e, -E_e)$ can be determined by substituting the values into the line equation. Arbitrarily setting the value of t_e to be 0, the equation of the tangent line can then be shown to be:

$$y = (-E_e \ln \mu)x - E_e \quad (5)$$

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

$$T_a = \frac{-1}{\ln \mu} \quad (6)$$

4. EXPERIMENT

An experiment was undertaken to validate the theory that the return phase parameter T_a 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:

$$O_q = \frac{t_e}{T_0}, \quad \alpha_m = \frac{t_p}{t_e}, \quad Q_a = \frac{T_a}{(1-O_q)T_0}$$

where O_q is the open quotient of the pulse, α_m its asymmetry coefficient, and Q_a its return phase coefficient.

Parameter	Range
f_0	80 : 20 : 200 (Hz)
O_q	0.3 : 0.05 : 0.9
α_m	0.67 : 0.05 : 0.9
Q_a	0.01 : 0.05 : 1

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_o of the following pulse such that the instants t_o 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.

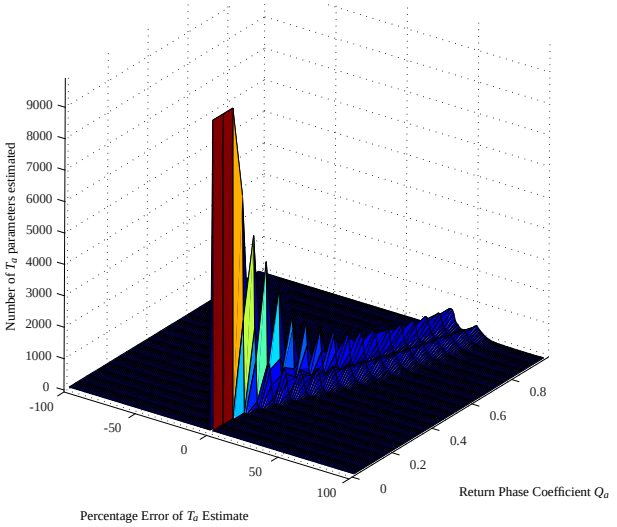

Figure 4: A 3-dimensional histogram displaying the number of waveforms determined at a particular percentage error of T_a against the return phase coefficient Q_a.

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:

$$E_{T_a} = 100\frac{T_a^{est} - T_a}{T_a} \tag{7}$$

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 $T_0 = int(\frac{f_0}{f_s})$.

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 do the offset required for a null value at t_c. The value of this offset can be directly calculated from the LF model return phase (3):

$$LF_{offset} = \frac{-E_e e^{-\varepsilon(T_c - T_e)}}{\varepsilon T_a} \tag{8}$$

From (8) above, it can be seen that large offsets occur when the length of the return phase (calculated as $(T_c - T_e)$) 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 parameters 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.

Voice	Q_a
modal	0.001
vocal fry	0.08
breathy	0.07
falsetto	0.4
harsh	0.01

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 bandwidths 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 are detected - 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 appear in the results of closed phase covariance linear prediction is

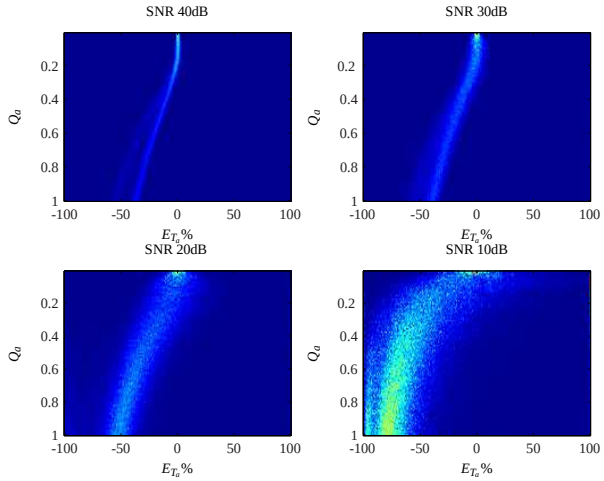

Figure 5: The percentage error of the T_a estimate predicted by the real pole found by analysis in the case where Gaussian noise is added to and modulated by the source signal at different SNRs.

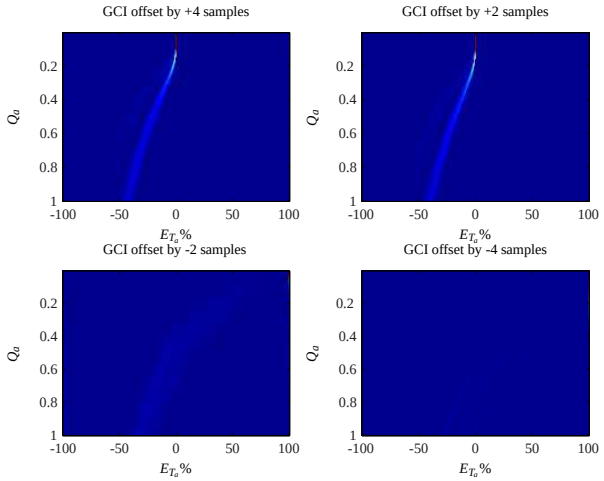

Figure 6: The percentage error of the T_a estimate predicted by the real pole found by analysis in the case the instant of glottal closure is offset by a number of samples.

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_a 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

[1] D. Wong et al., "Least squares glottal inverse filtering from the acoustic speech," *IEEE T. Acoust. Speech*, 1979, pp. 350-355.

[2] A. Krishnamurthy and D. Childers, "Two-channel speech analysis," *IEEE T. Acoust. Speech*, vol. 34, 1986, pp. 730-743.

[3] Plumpe, M., et al., "Modeling of the glottal flow derivative waveform with application to speaker identification," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol.: 7, pp. 569-586, 1999.

[4] Childers, D. G., Lee, C. K., "Vocal quality factors: Analysis, synthesis, and perception," *The Journal of the Acoustical Society of America*, vol. 90(5), pp.2394-2410, November 1991.

[5] Cummings, K.E., Clements, M.A., and Hansen, J.H.L., "Estimation and comparison of the glottal source waveform across stress styles using glottal inverse filtering," in *Southeastcon '89. Proceedings. Energy and Information Technologies in the Southeast, IEEE*, vol.2, pp.776-781, 1989.

[6] Alku, P., et al., "Closed phase covariance analysis based on constrained linear prediction for glottal inverse filtering," *The Journal of the Acoustical Society of America*, vol. 125, pp. 3289-3305, 2009.

[7] Fant, G., J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow," *STL-QPSR*, vol. 26, pp. 1-13, 1985.

[8] G. Fant, *Acoustic theory of speech production*, Walter de Gruyter, 1970.

[9] Markel, J.E. and A.H. Gray, *Linear Prediction of Speech*. Ney York: Springer-Verlag, 1982.

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

[11] Rabiner, L.R. and R.W. Schafer, *Digital Processing of Speech Signals*. Prentice-Hall, 1978.

[12] P.A. Naylor et al., "Estimation of glottal closure instants in voiced speech using the DYPSA algorithm," *IEEE T. Audio Speech*, vol. 15, 2007, pp. 34-43.

[13] Murty, K. and Yegnanarayana, B., "Epoch Extraction From Speech Signals", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 16, pp. 1602-1613, 2008.

[14] Drugman, T. and Dutoit, T., "Glottal Closure and Opening Instant Detection from Speech Signals," *Interspeech09*, Brighton, U.K., 2009.

[15] Doval, B. and d'Alessandro, C., "The spectrum of glottal flow models." Notes et document LIMSI, num. 9907, 1999.

[16] Gobl, C., Ní Chasaide, A., "Acoustic characteristics of voice quality", *Speech Communication*, vol. 11, pp. 481490, 1992.

[17] D.G. Childers, Speech Processing and Synthesis Toolboxes, Wiley, 1999.