

EPOCH EXTRACTION FROM LINEAR PREDICTION RESIDUAL

T.V. Ananthapadmanabha and B. Yegnanarayana

Department of Electrical Communication Engineering
Indian Institute of Science, Bangalore 560 012

ABSTRACT

An interpretation of linear prediction (LP) residual is presented by considering the effect of following factors: shape of glottal pulse, phase angles of formants at the instant of excitation, inaccurate estimation of formants and bandwidths, zeroes in vocal tract system transfer function. Effect of improper phase cancellation on the accuracy of estimated epoch position is also discussed. A method for unambiguous identification of epochs from LP residual is presented.

INTRODUCTION

Instants at which significant excitation of the resonances of vocal tract system occur during phonation are referred to as epochs. Epochal information is useful in accurate estimation of pitch periods [1] and in identifying the closed glottis interval. This latter information is useful in accurate estimation of the frequency response of vocal tract system [2,3]. In spite of its importance in the analysis of voiced speech, there is no satisfactory technique for extracting epochal information from speech wave. It is known that epochs coincide with the instants of slope discontinuities of glottal pulses. But extraction of glottal pulses from connected speech is as yet an unsolved problem. Sobakin [4] has proposed a measure for the linear predictability of a signal over a given interval. Strube [5] has used this measure with some modification to identify the instant of glottal closure, i.e., the epoch. However, Strube's method is computationally complex that it is suitable for analysis of single or a few pitch periods of speech wave. The error signal obtained in LP analysis [6], LP residual is known to contain information pertaining to epochs. A large value of LP residual, within a pitch period, is supposed to indicate the epoch. But epoch identification directly from LP residual is not recommended [5] due to the following reasons. LP analysis assumes an all pole model for representing the combined effect of impulse response of vocal tract system and glottal pulse shape. The all pole model implicitly assumes a minimum phase characteristic for speech signal. If this is

not valid, the phase response of vocal tract system is not compensated exactly by the digital inverse filter. Phase compensation is also incorrect when formants and bandwidths are inaccurately estimated. Effect of uncompensated phase on LP residual is not known. Moreover, the inverse filter does not compensate for zeroes which may be introduced due to nasal coupling or glottal pulse shape. These factors cause multiple peaks of either polarity to occur around epochs and make unambiguous identification of epochs from LP residual difficult. Further, there is an uncertainty in the estimated epoch position since the inverse filter does not compensate the phase response of vocal tract system exactly.

INTERPRETATION OF LP RESIDUAL

A. Model for Voiced Speech. We shall develop a new model for voiced speech for the purpose of interpreting LP residual. Let $g(t)$ be the quasi-periodic sequence of glottal pulses exciting vocal tract system $V(s)$, producing voiced speech signal $s(t)$. Alternatively, $s(t)$ can be considered to be the response of system $V_1(s)$ to second derivative of $g(t)$; here

$$V_1(s) = V(s)/s^2 \quad (1)$$

Second derivative of $g(t)$ can be divided into two components: (i) $e(t)$ which consists of only impulses in the second derivative, and (ii) $r(t)$ which is obtained after removing $e(t)$ from the second derivative of $g(t)$. It may be noted that there are no glottal zeroes in the system $V_1(s)$. Also the envelope of short-time spectrum of $s(t)$ corresponds nearly to the frequency response of $V_1(s)$.

B. Glottal Pulse and LP Residual. To study the effect of glottal pulse shape on LP residual, let us assume that there are no (nasal) zeroes and that the LP coefficients (LPCs) are accurately estimated, i.e.,

$$V_1(z) A(z) = 1 \quad (2)$$

Where $V_1(z)$ is the digital equivalent of $V_1(s)$ and

$A(z)$ is the digital inverse filter. The z-transform of LP residual is given by

$$T(z) = S(z)A(z) = [E(z) + R(z)]V_1(z)A(z)$$

$$T(z) = E(z) + R(z) \quad (3)$$

Hence the LP residual $t(nT)$ will be the second derivative of $g(t)$.

LP residual for a segment of synthetic vowel waveform and the second derivative of glottal pulses used in the generation of the vowel are shown in Fig. 1. The signal $e(t)$ and $r(t)$ can be identified from the second derivative of $g(t)$. The correlation between LP residual and second derivative of $g(t)$ may be noted.

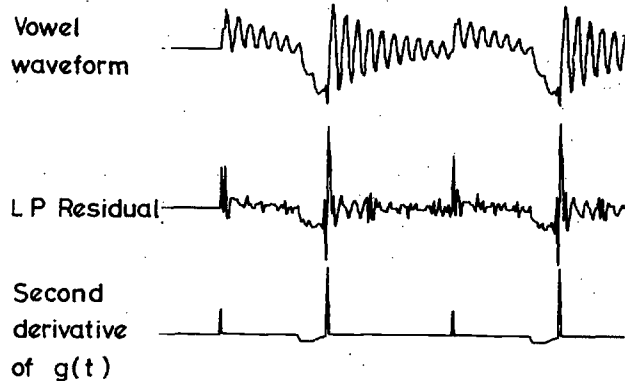


Fig. 1 Glottal pulse and LP residual

C. Effect of Phase. Consider the following form for the impulse response of vocal tract system:

$$v(nT) = \sum_{k=1}^M C_k [\cos(\theta_k) v_k(nT) + \sin(\theta_k) v_{kH}(nT)]$$

where $v_k(nT)$ is the minimum-phase part of impulse response of k -th formant, and $v_{kH}(nT)$ is its Hilbert transform, and M is the number of formants. This form takes into account the various phase angles with which the formants appear at the instant of excitation. We shall consider a single resonator ($M=1$) case,

$$v(nT) = p v_1(nT) + q v_{1H}(nT)$$

where $p = C_1 \cos(\theta_1)$ and $q = C_1 \sin(\theta_1)$. The spectra of $v_1(nT)$ and $v_{1H}(nT)$ are identical. The inverse filter $A(z)$ estimates this spectrum but possesses a minimum phase characteristic. Hence we get for LP residual the following equation:

$$t(nT) = v(nT) * a_n = p_1 \delta(nT) + q \delta_H(nT),$$

where $\{a_n\}$ is the finite impulse response of $A(z)$, $\delta(nT)$ is the Kronecker delta function and $\delta_H(nT)$

is the Hilbert transform of $\delta(nT)$. LP residuals for three different choices of p and q are shown in Fig. 2. Only when $q=0$, $t(nT)$ will be an impulse. In general $t(nT)$ will possess samples of either

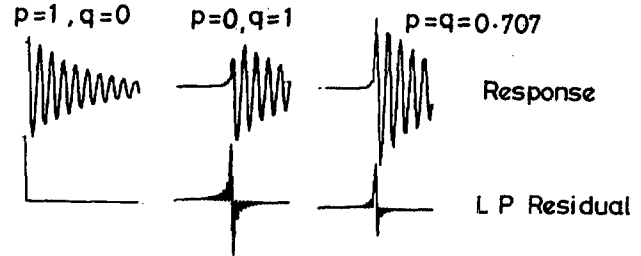


Fig. 2 Effect of phase angle of formant on LP residual

polarity and in the extreme case when $p=0$, $t(nT)$ will be zero at the instant of excitation. LP residual $t(nT)$ and its Hilbert transform $t_H(nT)$ for vowel sound /e/ is shown in Fig. 4. Although several other factors could have influenced LP residual, the effect of phase angles can be noted from a comparative examination of $t(nT)$ and $t_H(nT)$.

D. Effect of Inaccurate Estimation of $V_1(z)$. The inverse filter $A(z)$ is supposed to estimate the reciprocal of the frequency response of vocal tract system $V_1(z)$. Large errors in the estimated formant frequencies and bandwidths occur in the analysis of high pitched voiced sounds and when a zero (glottal or nasal) lies near a formant. Let us consider a single resonator case to study the effect of inaccurate estimation of $V_1(z)$ on LP residual. Let $f(nT)$ be the impulse response of a second order recursive filter having a damped cosine form;

$$f(nT) = -\hat{a}_1 f(n-1)T - \hat{a}_2 f(n-2)T,$$

with $f(0) = 1.0$, and with zero initial conditions. Let the estimated LPCs for $f(nT)$ be

$$a_1 = \hat{a}_1 + \tilde{a}_1 \quad \text{and} \quad a_2 = \hat{a}_2 + \tilde{a}_2$$

where \tilde{a}_1 and \tilde{a}_2 are errors in LPCs. These errors can be related to errors in the resonant frequency and bandwidth in the frequency response of $f(nT)$. LP residual for $f(nT)$ can be written as,

$$d(nT) = \delta(nT) + \tilde{a}_1 f(n-1)T + \tilde{a}_2 f(n-2)T.$$

It may be noted that $d(nT)$ consists of delayed and scaled versions of the original signal. Define $d_1(nT)$ as

$$d_1(nT) = \begin{cases} d(nT) & \text{for } n > p \\ 0 & \text{for } n = 0, 1, \dots, p \end{cases}$$

where p is the order of predictor which for signal $f(nT)$ is two. Signals $f(nT)$, $d(nT)$ and $d_1(nT)$, and

their log spectra are shown in Fig. 3. Log spectra of $d_1(nT)$ and $f(nT)$ are strikingly similar.

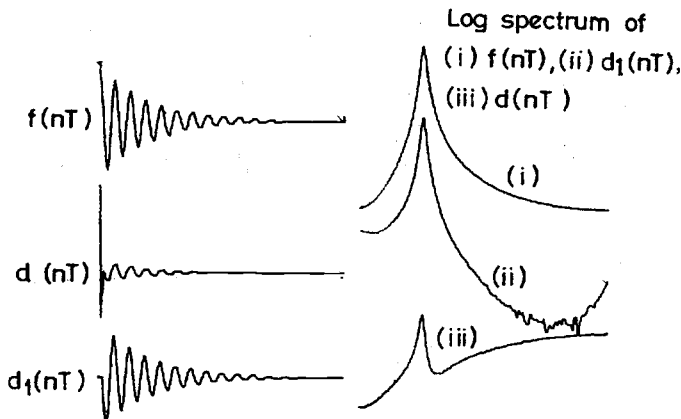


Fig. 3 Effect of inaccurate estimation of resonant frequency and bandwidth on LP residual.

E. Effect of Zeroes. Presence of zeroes affect the estimated frequency response of vocal tract system. This effect has already been discussed above. A complex zero has a response in time domain which can be represented by a sequence of the form, $(1, b_1, b_2)$. Since $A(z)$ does not compensate for zeroes, in place of an impulse, a finite sequence of impulses, corresponding to the zeroes, occurs in the LP residual.

F. Effect of Uncompensated Phase. Effect of uncompensated phase on the accuracy of identification of epochs will now be discussed. LP residual $t(nT)$ for the impulse response of vocal tract system should possess the largest value at $nT = 0$. The peak in $t(nT)$ occurs delayed by n_0T samples, if there is a linear phase term $(n_0\omega T)$ in the phase response $\phi(\omega)$ of $T(z)$. A linear phase term in $\phi(\omega)$ is equivalent to a constant term, C , in the derivative, $\phi^{(1)}(\omega)$ of $\phi(\omega)$. Hence

$$C = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi^{(1)}(\omega) d\omega$$

$$= \frac{1}{2\pi} [\phi(\pi) - \phi(-\pi)]$$

Change in phase for one complete rotation around the unit circle in z -plane is known (from complex variable theory [7,8]) to be $2\pi L$, where L is the number of zeroes outside the unit circle; this statement assumes that there are no poles outside the unit circle. Thus, irrespective of the nature of uncompensated phase, a large value in LP residual occurs at the correct epoch location provided, $T(z)$ does not contain any zero outside the unit circle. Thus, when the system $1/A(z)$ has stable impulse response, epoch identification from LP residual will be accurate for non-

nasalized sounds. In this context, it will be of interest to determine the location of nasal zeroes relative to unit circle.

EPOCH FILTERING OF LP RESIDUAL

Epoch filtering technique has been proposed for epoch extraction of voiced speech [9]. An epoch filter computes the Hilbert envelope of bandpass filtered speech signal. The bandpass filter is realized by using a frequency domain window function. The passband of the bandpass filter lies over a frequency range where the signal spectrum is nearly flat. We shall adopt epoch filtering technique for unambiguous identification of epochs from LP residual. LP residual can be considered to be the response of an all-pass filter. Hence the entire frequency range from dc ($\omega=0$) to folding frequency can be used in the epoch filter. The effect of glottal pulse shape is to introduce a low frequency signal $r(nT)$ in the LP residual, in addition to the impulse sequence $e(nT)$. Since our aim is to extract $e(nT)$ from LP residual, a frequency domain window which attenuates low frequency components can be chosen in the epoch filter. From considerations of noise, the frequency components near the folding frequency is given a low weightage. We have used a hanning window centered at half the folding frequency. LP residual $t(nT)$ and the windowed (in frequency domain) LP residual $t_1(nT)$ for vowel /e/ are shown in Fig. 4. The low frequency trend in $t(nT)$ due to $r(nT)$ is clearly absent in $t_1(nT)$. Epoch filter output, i.e., Hilbert envelope of $t_1(nT)$ shows unambiguous peaks at epochs. A secondary peak (at glottal opening), apart from the main peak (at glottal closure) is clearly evident in the epoch filter output.

CONCLUSIONS

An interpretation for the observed behaviour of LP residual has been given by considering the effect of various factors. The accuracy of estimated epoch position has been discussed in relation to the phase spectrum of LP residual. A method for obtaining unambiguous peaks at epochs has been presented.

REFERENCES

1. J.L. Flanagan, Speech Analysis Synthesis and Perception, Springer Verlag, Sec. Ed. 1972, Ch. 5, pp. 184-186.
2. E.N. Pinson, "Pitch-synchronous time-domain estimation of formant frequencies and bandwidths," J. Acoust. Soc. Am., Vol. 35, Aug. 1963, pp. 1264-1273.

3. K. Steiglitz and B. Dickinson, "The use of time-domain selection for improved linear prediction," IEEE Trans. Acoust., Speech and Sig. Proc., Vol. ASSP-25, Feb. 1977, pp. 34-39.
4. A. N. Sobakin, "Digital computer determination of formant parameters of the vocal tract from a speech wave," Soviet Phys. Acoust., Am. Inst. of Phys., Vol. 18, July-Sep, 1972.
5. H. W. Strube, "Determination of the instant of glottal closure from the speech wave," J. Acoust. Soc. Am., Vol. 56, Nov. 1974, pp. 1625-1629.
6. J. D. Markel and A. H. Gray, Linear Prediction of Speech, Springer Verlag, New York, 1976.
7. A. J. Berkhout, "On the minimum length property of one sided signals," Geophysics, Vol. 38, Aug. 1973, p. 661.
8. E. C. Titchmarsh, The Theory of Functions, Oxford University Press, London, 1932, p. 115.
9. T. V. Ananthapadmanabha and B. Yegnanarayana, "Epoch extraction of voiced speech," IEEE Acoust. Speech and Sig. Proc., Vol. ASSP-23, Dec. 1975, pp. 562-570.

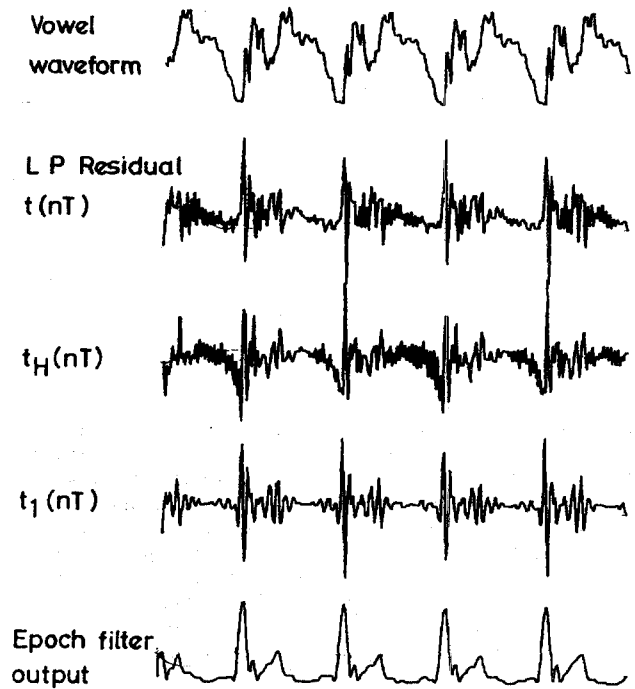


Fig. 4 Epoch filtering of LP residual.