Linear Prediction of Deterministic Components in Hybrid Signal Representation

Elias Azarov
Computer engineering department
Belarusian State University
of Informatics and Radioelectronics
Minsk, Belarus

Alexander Petrovsky*)
Department of real-time systems,
Bialystok University of Technology
Wiejska 45A, 15-351 Bialystok, Poland
palex@bsuir.by

***Abstract*—The paper presents a framework for representing sinusoidal and harmonic components by means of linear prediction. In order to improve estimation accuracy parametric representation of the signal is used instead of its waveform. The main target application of the proposed technique is accurate estimation of spectral envelopes in hybrid (stochastic/deterministic) speech processing.**

I. INTRODUCTION

Sinusoidal and harmonic signals constitute an important class of audible sounds that can be effectively described using parametric representation. Sinusoidal modeling was introduced in [1] and has been used in many different audio and speech applications. Contemporary processing systems often use hybrid parameterization that implies signal separation into deterministic (quasistationary or periodic) and stochastic components [2]. As long as sinusoidal modeling can be applied directly to the input the stochastic component is often referred to as residual. Any sinusoid is determined by three parameters (amplitude, frequency and initial phase) that are estimated using a harmonic analysis technique. As has been reported there are several approaches that can provide instantaneous (different at each moment of time) sinusoidal parameters estimation [3-6]. Speech modeling by means of the hybrid approach is of special interest because speech segments of different nature (voiced and unvoiced) conform to different parts of the representation (deterministic and stochastic respectively). Voiced speech is modeled by harmonically related sinusoids (harmonic model) that can significantly reduce number of parameters needed for storage, processing or transmission.

Some applications, e.g. speech coders and voice convertors, require an efficient description of short-time spectral envelopes of the signal. One of the most popular representations is linear spectral frequencies (LSF) that are obtained from linear prediction (LP) coefficients. The coefficients can be obtained either by the autocorrelation or covariance method. Both of the methods are applied to samples of the signal in time domain, thus if one wants to use deterministic/stochastic separation of the signal in combination with LSF techniques it is necessary to synthesize deterministic part from its parametric representation first and then perform coefficients estimation. Such solution results in a significant accuracy loss both in time and frequency domain because of parameters averaging within the analysis frame and therefore exact match between LSF description and its deterministic counterpart, represented by sinusoidal parameters, cannot be achieved.

In this work a transformation technique is proposed that can be used for direct prediction coefficients estimation from sinusoidal parameters. The technique minimizes the difference between sinusoidal and LP spectral envelopes.

II. SINUSOIDAL REPRESENTATION AND LP

The sinusoidal model assumes that the signal $s(n)$ can be expressed as a sum of cosine functions with slowly varying parameters:

$$s(n) = \sum_{k=1}^{K} \mathrm{MAG}_k(n)\cos\varphi_k(n), \tag{1}$$

where $\mathrm{MAG}_k(n)$ - the instantaneous magnitude of the k-th sinusoidal component, K is the number of components and $\varphi_k(n)$ is the instantaneous phase of the k-th component. Instantaneous phase $\varphi_k(n)$ and instantaneous frequency $f_k(n)$ are related as follows:

$$\varphi_k(n) = \sum_{i=0}^{n} \frac{2\pi f_k(i)}{F_s} + \varphi_k(0), \tag{2}$$

where F_s is the sampling frequency and $\varphi_k(0)$ is the initial phase of the k-th component. The harmonic model states that frequencies $f_k(n)$ are integer multiples of the fundamental frequency $f_0(n)$ and can be calculated as:

$$f_k(n) = kf_0(n). \tag{3}$$

LP model assumes that a given signal sample $s(n)$ can be approximated as a linear combination of the p past samples that leads to the following equality:

$$s(n) = \sum_{i=1}^{p} a_i \, s(n-i) + Gu(n), \tag{4}$$

where $a_1, a_2, \ldots, a_p$ are prediction coefficients, $u(n)$ is a normalized excitation and G is the gain of the excitation [7]. In z-domain the following transfer function can be written:

$$H(z) = \frac{G}{1-\sum_{i=1}^{p} a_i z^{-i}} = \frac{G}{A(z)}. \tag{5}$$

The prediction error $e(n)$ is defined as the difference between the source and predicted samples:

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^{p} a_k s(n-k). \tag{6}$$

The basic problem of LP is to find the set of predictor coefficients that minimize the mean-square prediction error. There are two primary solutions to this problem: the autocorrelation and covariance methods.

The autocorrelation method assumes that $s(n) = 0$ outside the interval $0 \le n < N$ and minimizes the prediction error within infinity. It comes to the system of linear equations [8]:

$$\sum_{i=1}^{p} a_i r(|i-j|) = -r(j), \tag{7}$$

where $j = 1,2,\ldots,p$ and $r(l)$ is the autocorrelation function $r(l) = \sum_{n=0}^{N-1-l} s(n)s(n+l)$, $l \ge 0$. The covariance method uses the interval where the signal is defined and can be expressed as the system:

$$\sum_{i=1}^{p} a_i c_{ij} = -c_{0j}, \tag{8}$$

where $j = 1,2,\ldots,p$ and $c_{ij} = \sum_{n=p}^{N-1} s(n-i)s(n-j)$.

The autocorrelation method always produces a stable solution (i.e. all the roots of $A(z)$ are within the unit circle). Unlike it the covariance method does not and therefore it is less popular in speech processing though it may give more accurate results [8]. The frequency domain behaviour of the $A(z)$ can be derived by evaluating

$$H(e^{i\omega}) = \frac{G}{1-\sum_{k=1}^{p} a_k e^{-j\omega k}} = \frac{G}{A(e^{i\omega})}. \tag{9}$$

In terms of these two models the transformation problem can be defined as follows: given a spectral envelope, specified by frequency and amplitude vectors $\mathrm{MAG}_1(n), \ldots, \mathrm{MAG}_K(n)$ and $f_1(n), \ldots, f_K(n)$ estimate linear prediction coefficients $a_1(n), \ldots, a_p(n)$ that provide the closest possible spectral envelope, specified by the all-pole filter $1/A(e^{i\omega})$. An obvious solution is to synthesize the corresponding periodic signal and then apply a conventional LP analysis technique to it. Although this approach seems very simple it is not the simplest considering computation complexity and, moreover, is not the most accurate. If one uses the autocorrelation method it is necessary to analyze a long frame to provide good results. The covariance method theoretically can have better performance however it requires at least one period of the fundamental and is not used for spectral analysis because of specific spectrum interpretation [8].

III. Transformation of Sinusoidal Parameters into Prediction Coefficients

Let us consider a sinusoid with a constant amplitude MAG, constant frequency f and zero initial phase:

$$s(n) = \mathrm{MAG}\cdot\cos(fn). \tag{10}$$

The prediction error can be written in the following way:

$$\begin{aligned} e(n) &= s(n) - \tilde{s}(n) = \\ &= \mathrm{MAG}\cdot\cos(fn) - \sum_{i=1}^{p} \mathrm{MAG}\cdot a_i \cos\big(f(n-i)\big) = \\ &= \mathrm{MAG}\cdot\left[1 - \sum_{i=1}^{p} a_i \cos(fi)\right]\cos(fn) - \\ &- \mathrm{MAG}\cdot\left[\sum_{i=1}^{p} a_i \sin(fi)\right]\sin(fn). \end{aligned} \tag{11}$$

The prediction error is a sinusoid, which energy can be characterized by the squared amplitude:

$$E_a^2 = \mathrm{MAG}^2 \cdot \begin{pmatrix} \left[1 - \sum_{i=1}^{p} a_i \cos(fi)\right]^2 \\ + \left[\sum_{i=1}^{p} a_i \sin(fi)\right]^2 \end{pmatrix} \tag{12}$$

Thus, in the case of multicomponent sinusoidal or harmonic signal, it is possible to evaluate prediction coefficients that minimize the sum of squared amplitudes of the residual sinusoids. Given an instantaneous vector of sinusoidal amplitudes $\mathrm{MAG}_1(n), \ldots, \mathrm{MAG}_K(n)$ and frequencies $f_1(n), \ldots, f_K(n)$ the relative residual energy can be evaluated as the following sum:

$$E_a^2 = \sum_{k=1}^{K} \mathrm{MAG}_k(n)^2 \begin{pmatrix} \left[1 - \sum_{i=1}^{p} a_i \cos(f_k(n)i)\right]^2 \\ + \left[\sum_{i=1}^{p} a_i \sin(f_k(n)i)\right]^2 \end{pmatrix}. \tag{13}$$

In order to minimize E_a^2 it is possible to use the basic minimization approach by finding partial derivatives with respect to variables a_i and then solving the system of linear equations. Eventually the following system can be derived:

$$\sum_{i=1}^{p} a_i q(|i-j|) = -q(j) \tag{14}$$

where $j = 1,2,\ldots,p$ and $q(l) = \sum_{k=1}^{K} \mathrm{MAG}_k(n) \cos(f_k(n) l)$, $(l \ge 0)$.

The system is similar to the one in autocorrelation method and can be solved accordingly. There are several well-studied approaches that provide simple and effective solutions [9].
Rank of the matrix

$$Q = \begin{bmatrix} q(0) & \cdots & q(p-1) \\ \vdots & \ddots & \vdots \\ q(p-1) & \cdots & q(0) \end{bmatrix} \quad (15)$$

does not exceed $2K$ and therefore there is no need to choose $p > 2K$. However, if the prediction order p is to be more than $2K$ the system should be reduced to $2K$ order and then $p - 2K$ zeros should be added to the evaluated prediction coefficients.

Spectrum analysis of a sinusoidal signal is illustrated in Fig.1. The signal consists of two sinusoids with equal constant amplitudes $\mathrm{MAG}_1(n) = 1$, $\mathrm{MAG}_2(n) = 1$ and harmonically related normalized frequencies $f_1(n) = 0{,}05\pi$, $f_2(n) = 0{,}1\pi$. The results were obtained by different approaches with the same prediction order $p = 8$. In order to provide accurate results 400 samples of the signal weighted by the Hamming window were used for the autocorrelation method and one exact period of the f_1 (i.e. 40 samples) was used for the covariance method.

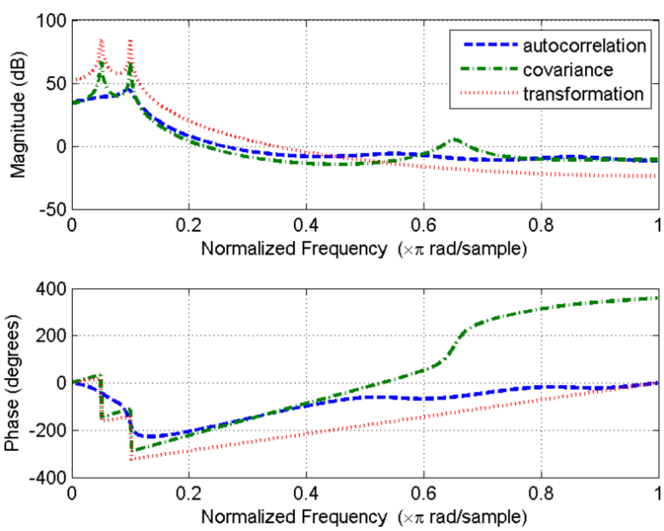

Figure 1. Spectral analysis of a two-component harmonic signal

Almost all signal processing applications that use LP require stability of the filter $1/A(z)$. The stability is guaranteed if all the roots of the polynomial $A(z)$ are inside the unit circle. The proposed technique may produce an unstable filter (when $p \geq 2K$) with roots on the circle. In this case a stabilizing parameter Δd can be used [10]. Equation (14) becomes

$$\sum_{i=1}^{p} a_i q(|i-j|) = -q(j) \quad (16)$$

where $j = 1,2,\dots,p$ and

$$q(l) = \begin{cases} \sum_{k=1}^{K} \mathrm{MAG}_k(n)\cos(f_k(n)l) + \Delta d, & l = 0 \\ \sum_{k=1}^{K} \mathrm{MAG}_k(n)\cos(f_k(n)l), & l > 0 \end{cases}.$$

The value of Δd specifies how close the roots $A(z)$ are to the unit circle. The closer Δd is to zero the closer filter is to instability. However large values of the parameter may degrade prediction accuracy.

It is known that LP spectral representation tends to model individual harmonic components instead of the spectral envelope when the order of prediction becomes high. Using derived transformation system it is possible to represent exactly the specified envelope as a high-order filter $1/A(z)$ by using amplitude and frequency vectors of infinite dimension.

The spectral envelope can be considered as a continuous function of frequency $\mathrm{MAG}(\omega)$, specified on the interval $[0,\pi]$. Then the matrix elements $q(l)$ of (14) can be derived as the integral

$$q(l) = \int_0^{\pi} \mathrm{MAG}(\omega)\cos(\omega l)d\omega. \quad (17)$$

If $\mathrm{MAG}(\omega)$ contain discontinues in points $\omega_d = (\omega_1, \omega_2, \dots, \omega_I)$, then (17) can be expressed as follows:

$$q(l) = \sum_{i=1}^{I+1} \int_{\bar{\omega}_{d,i}}^{\bar{\omega}_{d,i+1}} \mathrm{MAG}(\omega)\cos(\omega l)d\omega, \quad (18)$$

where $\bar{\omega}_d = (0, \omega_1, \omega_2, \dots, \omega_I, \pi)$.
If we specify the function $\mathrm{MAG}(\omega)$ in the following way:

$$\mathrm{MAG}(f) = \begin{cases} 1, & F_1 \leq \omega \leq F_2 \\ 0, & \omega \leq F_1, \omega \geq F_2 \end{cases}, \quad (19)$$

$0 \leq F_1 \leq F_2 \leq \pi$, then (18) becomes

$$q(l) = \begin{cases} \sin(F_2 l)/l - \sin(F_1 l)/l, & l \neq 0 \\ F_2 - F_1, & l = 0 \end{cases} \quad (20)$$

and (14) will result in a bandpass filter with the passband $F_1 \leq \omega \leq F_2$. In Fig.2 two frequency responses are shown - one of a filter designed using the window method (inverse Fourier transform of the ideal frequency response multiplied by Hamming window) and another of a filter designed as has been described above. Both filter were designed using the same bandpass $0{,}2\pi \leq \omega \leq 0{,}3\pi$ and the same filter order.

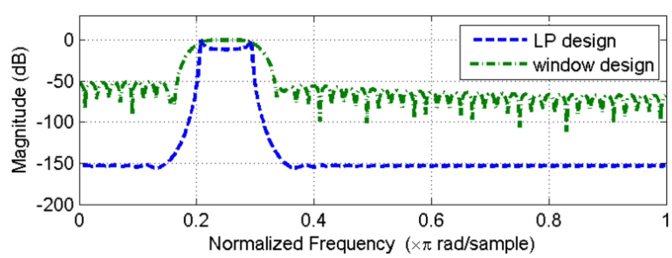

Figure 2. Frequeny responses of LP design and window design filters

Continuous spectral envelope can be estimated from amplitude and frequency vectors using linear interpolation. Single segments of the envelope $f_i \leq \omega \leq f_{i+1}$, $1 \leq i \leq K-1$ are described by linear equations of the form

$\mathrm{MAG}(\omega) = b_i\omega + c_i$. Parameters b_i and c_i are estimated from adjacent values of frequency and amplitudes. Thus from (18) elements of the system (14) can be derived in the following way:

$$q(l) = \sum_{i=1}^{K-1} D(l,i), \qquad (21)$$

where

$$D(l,i) = \begin{cases} \frac{b}{l^2}[\cos(f_{i+1}l) + f_{i+1}l\sin(f_{i+1}l)] + \frac{c}{l}\sin(f_{i+1}l) - \\ \quad -\frac{b}{l^2}[\cos(f_i l) + f_i l\sin(f_i l)] - \frac{c}{l}\sin(f_i l) & l \neq 0 \\ \quad \frac{1}{2}bf_{i+1}^2 + cf_{i+1} - \frac{1}{2}bf_i^2 - cf_i & l = 0 \end{cases}$$

Fig.3 shows how the proposed transformation technique compares with other methods.

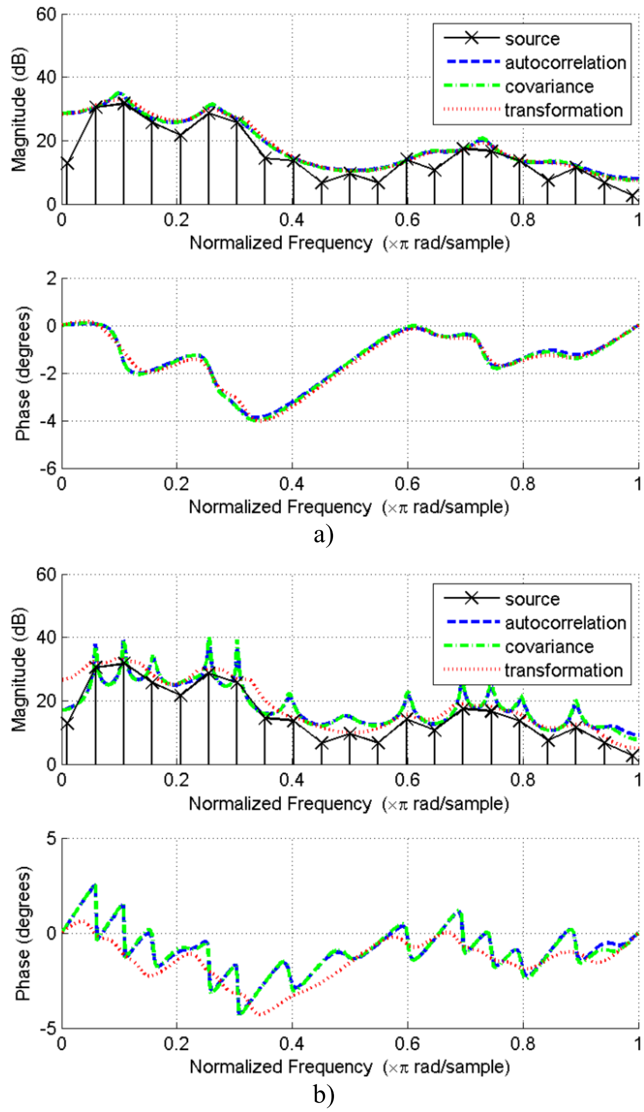

Figure 3. LP spectral envelope estimation of a harmonic signal (a – 14 prediction coefficients, b – 30 prediction coefficients)

In order to get estimation of prediction coefficients via autocorrelation and covariance methods 1024 samples of the periodical signal was generated using the source amplitude and frequency vectors. As can be seen from the picture the results are very close in the case of 14 prediction coefficients (Fig.3a), however when 30 coefficients are estimated (Fig.3b) the proposed technique provides much close approximation to the given spectral envelope.

IV. Conclusion

A transformation technique of sinusoidal parameters into prediction coefficients has been proposed. The distinguishing feature of the technique is accurate spectral envelopes estimation by means of LP. Unlike time domain methods (autocorrelation and covariance) exact shape of the envelope is approximated instead of single harmonics or single sinusoidal components. Taking into account that sinusoidal parameters can express instantaneous behavior of quasistationary periodic signals the composite approach (sinusoidal modeling with parameters transformation) is able to provide smooth LP spectral envelopes that are precisely localized in time. This can be very useful in speech processing applications like speech recognition, voice conversion and speech/audio coding.

Acknowledgment*)

This work was supported by the Polish Ministry of Science and Higher Education (MNiSzW) in years 2009-2011 (Grant no.N N516 388836).

References

[1] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation", IEEE Trans. On Acoustics, Speech and Signal Process., vol. 34, no. 4, pp. 744-754, 1986.

[2] X. Serra, "A system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition", Ph.D. thesis, Stanford University, Stanford, Calif, USA. – 1989.

[3] Boashash, B. "Estimating and interpreting the instantaneous frequency of a signal", Proceedings of the IEEE, Vol. 80, No. 4, 1992, pp. 520-568.

[4] T. Abe, M. Honda, "Sinusoidal model based on instantaneous frequency attractors", IEEE Trans. on Audio, Speech, and Language processing, vol. 14, no. 4, pp. 1292-1300, July 2006

[5] Maragos P., Kaiser J. F., Quatieri T. F., "Energy Separation in Signal Modulations with Application to Speech Analysis", IEEE Trans. on Signal Process., vol. 41, no. 10, pp. 3024-3051, 1993.

[6] Al. Petrovsky, E. Azarov, and A. Petrovsky, "Harmonic representation and auditory model-based parametric matching and its application in speech/audio analysis", AES 126th Convention, Preprint 7705, Munich, Germany, 2009. -13p.

[7] L. Rabiner, B. H. Juang, "Fundamentals of speech recognition" Prentice Hall, New Jersey 1993.

[8] J. D. Markel, A. H. Gray, Jr, "Linear prediction of speech", Springer-Verlag Berlin Heidelberg New York 1976.

[9] X. Huang, A. Acero, H.W. Hon, "Spoken language processing", Prentice Hall, New Jersey 2001.

[10] E. Azarov, A. Petrovsky, "Linear prediction of speech based on instantaneous harmonic parameters", Speech and Computer: proceedings of the 13-th intern. conf., Russia, Saint-Petersburg, June 21–25, 2009. pp. 215–220.