FREQUENCY DOMAIN LINEAR PREDICTION BASED ON TEMPORAL ANALYSIS

Ravi R. Shenoy [1, 2]

Nokia Research Center [1]
Nokia India Pvt. Ltd.,
Bangalore-560103, India
Email: rashenoy@ee.iisc.ernet.in

Chandra Sekhar Seelamantula [2]

Department of Electrical Engineering [2]
Indian Institue of Science
Bangalore-560012, India
Email: chandra.sekhar@ieee.org

ABSTRACT

Frequency-domain linear prediction (FDLP) is widely used in speech coding for modeling envelopes of transients signals, such as voiced and unvoiced stops, plosives, etc. FDLP fits an auto regressive model to the discrete cosine transform (DCT) coefficients of a sequence. The spectral prediction coefficients provide a parametric model of the temporal envelope. The prediction coefficients are obtained by solving the set of Yule–Walker equations expressing the relationship between lagged spectral autocorrelation values. A limitation of the direct approach of computing the spectral autocorrelation values is that the sequence has to be padded with a large number of zeros for the autocorrelation estimates to be reasonably accurate. This comes at the cost of increased computational complexity. We present an efficient and accurate method for computing the spectral autocorrelation samples. We show that the spectral autocorrelation can be computed as cosine-weighted temporal centroids, where the weighting function is dependent on time-index of the samples.

Index Terms— frequency-domain linear prediction, temporal envelope, temporal centroid

1. INTRODUCTION

Autogressive (AR) modeling provides an efficient way of predicting signal samples from past observations. Correlations that exist in periodic signals, and signals with rich harmonic structure are exploited by the linear prediction model. A thorough review of the theory of linear prediction (LP) and its applications to voiced speech signals can be found in the review article by Makhoul [1, 2]. The magnitude response of the all-pole filter derived from the prediction coefficients gives a parametric representation of the envelope of the magnitude spectrum, also known as spectral envelope. The parametric spectral envelope is the bane of source-filter model [3]. The linear prediction model fits well for the analysis and synthesis of voiced speech signals as the statistics of such signals can be assumed to be approximately stationary over a duration of 20 ms [4]. For transient signals such as stop consonants that are inherently non stationary, linear prediction analysis in the time domain is not useful as the signal lacks strong temporal correlations. However, the temporal localization of transient signals manifests as strong correlations in both real and imaginary parts of the spectrum. For example, consider a Kronecker impulse. The real and imaginary parts of the spectrum are sinusoids and the autocorrelation functions of real and imaginary parts of the spectrum are sinusoids, which are perfectly predictable. Athineos and Ellis showed that spectral correlations can be exploited to perform linear prediction of the discrete cosine spectrum (DCT) [5]. The spectral prediction model is referred to as frequency-domain linear prediction (FDLP). It has been shown that autocorrelation of the DCT sequence and the Hilbert envelope of the analytic representation of the sequence form a Fourier transform pair [6, 7, 8]. The prediction coefficients in the spectral domain in turn give a parametric representation of the Hilbert envelope [7]. The order of the predictor controls the degree of smoothness in the obtained envelopes. The Hilbert envelope of a real signal can also be obtained without explicitly computing the analytic signal [9]. For a signal that can be expressed as a superposition of transients and harmonic signals, linear prediction models applied to both time- and frequency-domains have been shown to be most suitable [7, 10].

The parametric temporal envelope model finds applications in speech coding, to suppress perceptible pre-echoes occurring before the onset of transients using temporal noise shaping [11], computation of robust features for automatic speech recognition [12], high-fidelity audio compression [13], speech recognition in reverberant conditions [14, 15], etc. In all of these applications, the temporal envelope is often computed by first obtaining the DCT spectrum from the observed signal following which sample spectral correlations are estimated. The prediction coefficients are then computed as a solution to the Yule–Walker equations constructed from sample autocorrelation values. Computing the spectrum and corresponding autocorrelation values are the most computationally intensive steps in FDLP. In this paper, we show an efficient way of computing the sample spectral autocorrelations without computing the spectrum explicitly. Starting from first principles, we show that the spectral autocorrelation can be expressed as a weighted temporal centroid, where the weighting is a cosine function of the time index. Given these sample spectral correlations one can then solve for spectral prediction coefficients using the Levinson–Durbin recursion procedure.

1.1. Contributions of this paper

The highlights of this paper are as follows. We derive the relationship between type-I odd-length DCT and discrete Fourier transform (DFT) of an even symmetric sequence constructed from the original sampled sequence. We then establish a relationship between spectral autocorrelation values and cosine weighted temporal centroids. We also compare computational complexities of both methods. For validation we show the application of proposed method to compute the temporal envelope of a Castanet click. The temporal centroid based method is applied on filtered sequence for improving the resolution of the temporal envelope. Improvement in temporal resolution by the application of a window in the spectrum in presence of two clicks within the observation window is shown.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

2. RELATING DFT AND TYPE-1 ODD DCT

We first show that, for a zero-mean sequence, the type-I odd DCT and DFT of an even symmetric version of the signal constructed from the original sequence match up to a scale factor. To illustrate this aspect, consider a real zero-mean N_o-point discrete-time sequence $x[n]$, $0 \leq n \leq N_0 - 1$. The sequence is zero padded to make it an N-point sequence. Without loss of generality, we assume that $x[0] = 0$. We construct an M-point ($M = 2N - 1$) even symmetric sequence from $x[n]$ as follows:

$$y[n] \quad = \begin{cases} x[n] & 0 \leq n \leq N-1 \\ x[M-n] & N \leq n \leq M-1. \end{cases}$$

The M-point DFT of $y[n]$ is given by

$$Y[k] = \sum_{n=0}^{M-1} y[n] e^{-\mathrm{j}\frac{2\pi nk}{M}},$$

where $0 \leq k \leq M-1$. For an even-symmetric sequence, the imaginary part of the DFT is 0, that is $Y[k]$ is real [16]. Since $x[0] = 0$ the lower limit of the summation is changed to $n = 1$. We split the summation in two terms

$$Y[k] = \underbrace{\sum_{n=1}^{N-1} y[n] \cos\left(\frac{2\pi nk}{M}\right)}_{\text{A}} + \underbrace{\sum_{n=N}^{M-1} y[n] \cos\left(\frac{2\pi nk}{M}\right)}_{\text{B}}.$$

Substituting $n = M - n$ in B and taking the 2π periodicity of the complex exponential into account, B is simplified to

$$\text{B} = \sum_{n=1}^{N-1} y[M-n] \cos\left(\frac{2\pi nk}{M}\right).$$

Replacing $y[n] = x[n]$ and $y[M-n] = x[n]$ for $n \in [0, N-1]$ (by construction), we observe that A and B are identical. The DFT of $y[n]$ is expressed as

$$Y[k] \quad = \begin{cases} 0 & k = 0, \\ 2\displaystyle\sum_{n=1}^{N-1} x[n] \cos\left(\frac{2\pi nk}{M}\right) & \text{otherwise.} \end{cases} \tag{1}$$

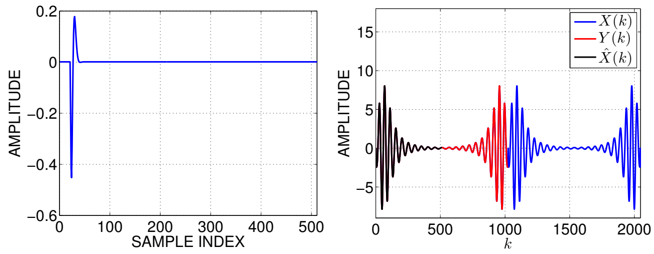
The N-point type-I odd discrete-time cosine transform (henceforth referred to as DCT-Io to distinguish it from the often used DCT type II) of $x[n]$, denoted as $X[k]$, is given by

$$X[k] = 4 \sum_{n=0}^{N-1} c_{n,k} x[n] \cos\left(\frac{2\pi nk}{M}\right),$$

where $0 \leq k \leq N-1$ and

$$c_{n,k} \quad = \begin{cases} \frac{1}{2} & \text{for } n, k = 0, \\ \frac{1}{\sqrt{2}} & \text{when only one of } n \text{ or } k = 0, \\ 1 & \text{otherwise.} \end{cases}$$

$X[k]$ is a scaled version of the orthogonal DCT with a factor of $2\sqrt{M}$. DCT-Io is the only trigonometric transform that is directly linked to the DFT of the even symmetric sequence [17, 7]. It should be observed that all the terms in the summation with coefficient of multiplication $c_{n,k} = 1/\sqrt{2}$ are zero since $x[0] = 0$ and mean of



(a) (b)

Fig. 1. (In Color) a) Synthetic transient constructed from β-density function; b) Symmetric mirror boundary periodicity of DCT-Io of $x[n]$ overlaid on DFT of even symmetric version of $x[n]$.

$x[n]$ is zero. $x[0] = 0$ also allows one to start the summation with index $n = 1$. Hence, $X(k)$ can be expressed as

$$X[k] \quad = \begin{cases} 0 & k = 0, \\ 4\displaystyle\sum_{n=1}^{N-1} x[n] \cos\left(\frac{2\pi nk}{M}\right) & \text{otherwise.} \end{cases} \tag{2}$$

Thus,

$$X[k] = 2Y[k], \qquad 0 \leq k \leq N-1 \tag{3}$$

Figure 1(a) and 1(b) show a synthetic transient and its DCT-Io, respectively. Figure 1(b) also shows the DFT of the $2N-1$ point even symmetric sequence constructed from $x[n]$.

2.1. Limitations of computing FDLP directly from the DCT coefficients

FDLP is often computed by taking the DCT of the windowed sequence. The discrete set of transform coefficients are then used to obtain an estimate of the spectral autocorrelation samples either by autocorrelation or covariance method. The autocorrelation method is preferred because the matrix of autocorrelation samples is Toeplitz and the corresponding AR filter is stable. Computation of the Toeplitz matrix using the autocorrelation method directly from the DCT coefficients is shown below:

$$\hat{R}_{pxp} = X^T X, \text{ where} \tag{4}$$

$$X = \begin{pmatrix} X(0) & 0 & 0 & \cdots \\ X(1) & X(0) & 0 & \cdots \\ X(2) & X(1) & X(0) & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ X(N-1) & X(N-2) & \cdots & X(0) \\ 0 & X(N-1) & \cdots & X(1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X(N-1) \end{pmatrix}.$$

The prediction coefficients obtained from a Toeplitz matrix is efficiently computed using a Levison-Durbin recursion formulation.

For achieving higher resolution in the temporal envelope, one should choose the prediction order p to be sufficiently large. The autocorrelation values computed with larger values of τ are based on fewer non-zero samples. This limitation arises from the situation that when τ is large, fewer terms in the inner product of columns are non-zero; that is, there are $\tau - 1$ zeros in both column 1 and column p

of X and since they are non-overlapping the inner product is formed from $N-2\tau-2$ samples. This limitation can be circumvented by computing the autocorrelation samples using covariance method. However, the resulting matrix is no longer Toeplitz and stability of the filter is not guaranteed. Alternatively, the number of samples in $X(k)$ can be increased to be sufficiently larger than p [1, 18]. This is achieved by either computing extended $X(k)$, denoted as $\hat{X}(k)$, for values of $0 \leq k \leq 2N-1$. Figure 1(b) shows the symmetric mirror boundary property of extended DCT-Io. Both of these steps can be combined and the spectral autocorrelation values can be obtained from temporal centroid as shown in Section 3.

3. COMPUTING SPECTRAL COVARIANCE FROM TEMPORAL CENTROIDS

For a given sequence $X[k]$, the autocorrelation function and power-spectral density form a Fourier transform pair (Wiener–Khintchine theorem). In practice, one estimates a sampled power spectral density from a finite-length autocorrelation as follows

$$S_{XX}[k] = \sum_{\tau=0}^{N-1} r_X[\tau] e^{-\mathrm{j}\frac{2\pi\tau k}{N}},$$

where $r_X[\tau]$ is the biased first-order autocorrelation defined as:

$$r_X[\tau] = \frac{1}{N}\sum_{k=0}^{N-1} X^*[k]X[k-|\tau|].$$

$r_X[\tau]$ is even, that is, $r_X[\tau] = r_X[-\tau]$. To overcome the limitations described in Section 2.1, we compute $r_X[\tau]$ using extended DCT-Io; that is

$$r_{\hat{X}}[\tau] = \frac{1}{M}\sum_{k=0}^{M-1} \hat{X}^*[k]\hat{X}[k-|\tau|]. \tag{5}$$

We observe from Figure 1(b) that $\hat{X}(k)$ and $Y(k)$ are identical except for the scaling factor of 2. Substituting (3) in (5), we get that

$$r_{\hat{X}}[\tau] = \frac{4}{M}\sum_{k=0}^{M-1} Y^*[k]Y[k-|\tau|]. \tag{6}$$

From the time-shifting property of Fourier transform, shift in one domain results in a modulation in the conjugate domain,

$$Y[k] \leftrightarrow y[n] \text{ then } Y[k-|\tau|] \leftrightarrow y[n]e^{\mathrm{j}\frac{2\pi n|\tau|}{M}}.$$

The DFT operation can be expressed in the matrix form as

$$\begin{aligned} \mathrm{Y} &= \mathbf{\Phi}\mathrm{y}, \text{ where} \\ \mathrm{y} &= [y[0], y[1], \cdots, y[M-1]]^T \\ \mathrm{Y} &= [Y[0], Y[1], \cdots, Y[M-1]]^T, \end{aligned} \tag{7}$$

and $\mathbf{\Phi}$ is the Vandermonde matrix. $\mathbf{\Phi}$ is a Hermitian matrix $\mathbf{\Phi}^H\mathbf{\Phi} = \mathrm{I}$, where ^{H} represents conjugate transpose. For the shifted spectrum, the DFT of the modulated signal is expressed in matrix form as

$$\begin{aligned} \mathrm{Y}_\tau &= \mathbf{\Phi}\mathrm{y}_\tau, \text{ where} \\ \mathrm{y}_\tau &= \left[y[0], y[1]e^{\mathrm{j}\frac{2\pi|\tau|}{M}}, \cdots, y[M-1]e^{\mathrm{j}\frac{2\pi(M-1)|\tau|}{M}}\right]^T, \\ \mathrm{Y}_\tau &= [Y[0-|\tau|], Y[1-|\tau|], \cdots, Y[(M-1)-|\tau|]]^T. \end{aligned} \tag{8}$$

Expressing (6) as the inner product of Y^H and Y_τ

$$r_{\hat{X}}[\tau] = \frac{4}{M}\mathrm{Y}^H\mathrm{Y}_\tau,$$

and substituting (7) and (8) for Y^H and Y_τ, respectively we get

$$r_{\hat{X}}[\tau] = \frac{4}{M}\mathrm{y}^H\mathbf{\Phi}^H\mathbf{\Phi}\mathrm{y}_\tau.$$

Since $\mathbf{\Phi}^H\mathbf{\Phi} = \mathrm{I}$, $r_{\hat{X}}(\tau)$ is expressed as

$$r_{\hat{X}}[\tau] = \frac{4}{M}\mathrm{y}^H\mathrm{y}_\tau. \tag{9}$$

Substituting for y and y_τ in (9), and noting that $\mathrm{y}^H = \mathrm{y}^T$ for real y

$$r_{\hat{X}}[\tau] = \frac{4}{M}\sum_{n=0}^{M-1} y^2[n]e^{\mathrm{j}\frac{2\pi n|\tau|}{M}}.$$

We start the summation from $n=1$ as $y[0]$=0. Since $y[n]$ is even symmetric, $y[n] = y[M-n]$, we combine terms corresponding to n and $M-n$ as follows

$$\begin{aligned} r_{\hat{X}}[\tau] &= \frac{4}{M}\sum_{n=1}^{N-1} y^2[n]\left(e^{\mathrm{j}\frac{2\pi n|\tau|}{M}} + e^{\mathrm{j}\frac{2\pi(M-n)|\tau|}{M}}\right) \\ &= \frac{4}{M}\sum_{n=1}^{N-1} y^2[n]\left(e^{\mathrm{j}\frac{2\pi n|\tau|}{M}} + e^{-\mathrm{j}\frac{2\pi n|\tau|}{M}}\right) \\ &= \frac{8}{M}\sum_{n=1}^{N-1} y^2[n]\cos\left(\frac{2\pi n|\tau|}{M}\right). \end{aligned}$$

Since, $y[n] = x[n]$ for $n = 1, 2, \cdots, N-1$, $r_{\hat{X}}(\tau)$ is expressed as

$$r_{\hat{X}}[\tau] = \frac{8}{M}\sum_{n=1}^{N-1} x^2[n]\cos\left(\frac{2\pi n|\tau|}{M}\right). \tag{10}$$

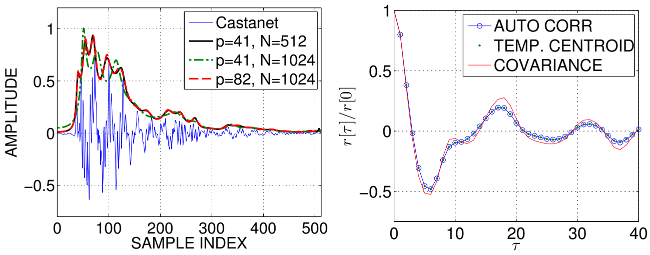
$r_{\hat{X}}(\tau)$ is the autocorrelation function of $\hat{X}(k)$; that is the DCT-Io of $x[n]$. As shown in (10), $r_X(\tau)$ is estimated from the temporal centroid with a cosine weighting that is a function of time-index. Since $x[n] = 0$ for $N_0 < n \leq N-1$, $r_{\hat{X}}(\tau)$ for $\tau = 0, 1, 2, \cdots, p$, can be computed as a matrix-vector product as shown below

$$\begin{pmatrix} r_{\hat{X}}[0] \\ r_{\hat{X}}[1] \\ \vdots \\ r_{\hat{X}}[p] \end{pmatrix} = A \begin{pmatrix} x^2[1] \\ x^2[2] \\ \vdots \\ x^2(N_0-1) \end{pmatrix},$$

where

$$A = \frac{8}{M}\begin{pmatrix} 1 & \cdots & 1 & 1 \\ \cos(1\gamma) & \cos(2\gamma) & \cdots & \cos((N_0-1)\gamma) \\ \cos(2\gamma) & \cos(4\gamma) & \cdots & \cos(2(N_0-1)\gamma) \\ \vdots & \vdots & \vdots & \vdots \\ \cos(p\gamma) & \cos(p2\gamma) & \cdots & \cos(p(N_0-1)\gamma) \end{pmatrix},$$

and $\gamma = \frac{2\pi}{2N-1}$. $\hat{R}$ is then constructed by creating a Toeplitz matrix from $[r_{\hat{X}}[0], r_{\hat{X}}[1], \cdots, r_{\hat{X}}[p]]$. Since the Toeplitz matrix is constructed after normalizing the autocorrelation samples by $r_{\hat{X}}[0]$, the scale factor does not affect the computation of prediction coefficients.



(a) (b)

Fig. 2. (In Color) a) An example Castanet click extracted from a recording with an overlay of FDLP envelopes obtained for a set of values for p and N; and b) Normalized autocorrelation function for a single Castanet click. It can be observed that the autocorrelation function obtained using the autocorrelation method and temporal centroid method are identical.

Direct approach	Temporal centroid
$\mathcal{O}\left(N\log(N)\right) + Np$	$(N_0 + 1)p$

Table 1. Comparison of computational complexities for the direct approach and the proposed temporal centroid method.

3.1. Advantages

Computing $\hat{R}$ using the temporal centroids method allows one to overcome the limitations described in Section 2.1. Given a sequence, let the DCT coefficients be computed for an arbitrarily large N such that $N - \tau + 1 \gg \tau$. As shown in (10), the additional zeros $N_0 \leq n \leq N - 1$ do not contribute to the value of $r(\tau)$. Instead, zero padding allows one to obtain values of the spectrum at finer resolution. For a given set of samples the modification needed for computing $\hat{R}$ is the value of γ that is dependent on N. The computational complexity of obtaining the transform coefficients is of the order of $\mathcal{O}\left(N\log(N)\right)$. Though $x[n]$ is zero for values of $N_0 \leq n$, $X[k]$ is not necessarily zero for all k. Computation of p spectral autocorrelation values therefore requires Np multiplications. On the other hand, computing the temporal centroid requires N_0 multiplications for computing sample squares and $N_0 p$ multiplications for a matrix vector product. A comparison is given in Table 1.

4. EXPERIMENTAL VALIDATION

We next show the experimental validation of the proposed method of computing spectral correlations and the corresponding temporal envelope obtained from linear prediction in the spectral domain. To illustrate this aspect, we consider a single Castanet click obtained from the European Broadcasting Union Sound Quality Assessment Material (EBU-SQAM) database. We select $N = 512$ samples of the recording at a sampling frequency of 48 kHz. We compute a 40^{th}-order temporal envelope using FDLP technique. Figure 2(a) shows the Castanet clip on which the estimated temporal envelope is overlaid. Figure 2(b) shows the plot normalized spectral autocorrelation values for $\tau = 0, 1, 2, \cdots, 40$ with $N = N_0$. It can be observed that the sample autocorrelation values computed from the temporal centroid method matches with the autocorrelation values computed using the autocorrelation method. Figure 2(b) also highlights the difference in the autocorrelation function obtained using the covariance method.

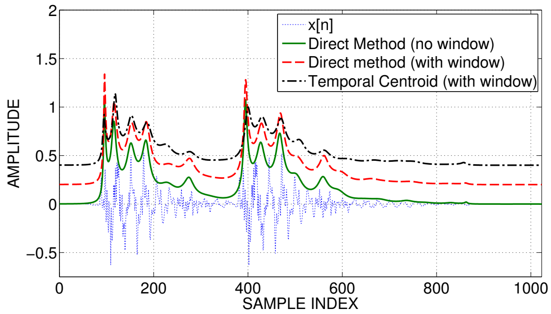


Fig. 3. (In Color) Highlighting the change in temporal resolution of the DCT-Io and Gaussian windowed DCT-Io for a sequence consisting of two Castanet clicks within a window of $N = 1024$ samples with $p = 41$. For temporal centroids method, the centroids were computed on the filtered sequence. The envelope plots are offset by a constant mainly to ensure readability.

We next focus on obtaining higher temporal resolution FDLP envelope. Ganapathy and Hermansky showed that envelope computed on windowed DCT-Io coefficients improves the temporal resolution of the FDLP envelope [8]. Any smooth window such as Gaussian, Hamming, etc. may be chosen. We show the adaptation our technique for computing the autocorrelation function of the windowed DCT-Io. Since multiplication in spectrum is equivalent to convolution in time, the inverse DCT-Io operation is employed on the window function to obtain the impulse response in the time domain. The filter coefficients may be truncated to a lower order as the window function is smooth. The original sequence is filtered with the impulse response and the temporal centroids are computed on the filtered sequence. The autocorrelation function of the windowed spectrum and the temporal centroids of the filtered sequence match very closely. In order to compare the envelopes, consider an audio clip of two Castanet clicks from the recording described earlier. The length of the clip is $N = N_0 = 1024$ samples. In the direct method, the envelope is computed from Gaussian windowed ($\sigma = 0.5$) DCT-Io. The estimated envelopes using the direct method without any window, direct method with a Gaussian window, and temporal centroid method with Gaussian window are shown in Figure 3. We observe that employing a window increases the peakiness of the envelope. The envelope of the windowed spectrum computed using direct and temporal centroid methods match very closely.

5. CONCLUSIONS

We presented an efficient method of computing the autocorrelation function from either the real or imaginary parts of the spectrum. We showed that the spectral autocorrelation is a temporal centroid whose weighting is dependent on the time index. For signals with zero-mean and sample at origin being zero, type-I odd DCT of the signal and the Fourier spectrum of the even symmetric signal constructed from the original sequence are identical up to a scale factor. The relation allows one to obtain the sample autocorrelation values of the DCT of a signal by computing cosine-weighted temporal centroids of the signal. Based on this relation, one can perform frequency-domain linear prediction (FDLP) at a significantly reduced computational cost. In the case where the temporal resolution of the FDLP envelope needs to be improved using a windowed DCT sequence, the temporal centroid method may be employed on a filtered sequence.

6. REFERENCES

[1] J. Makhoul, "Linear prediction: A tutorial review," *IEEE Proc.*, vol. 63, no. 4, pp. 561–580, 1975.

[2] P. P. Vaidyanathan, *The Theory of Linear Prediction*, Morgan and Claypool Publisher, 2008.

[3] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Pearson Education, Delhi, India, first indian reprint edition, 2004.

[4] T. F. Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practice*, Pearson Education, Delhi, India, first indian reprint edition, 2001.

[5] M. Athineos and D. P. W. Ellis, "Frequency-domain linear prediction for temporal features," in *IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, 2003, pp. 261–266.

[6] R. Kumaresan and A. Rao, "Model-based approach to envelope and positive instantaneous frequency estimation of signals with speech applications," *J. Acoust. Soc. Am.*, vol. 105, no. 3, pp. 1912–1924, 1999.

[7] M. Athineos and D. P. W. Ellis, "Autoregressive modeling of temporal envelopes," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5237–5245, 2007.

[8] S. Ganapathy and H. Hermansky, "Temporal resolution analysis in frequency domain linear prediction," *J. Acoust. Soc. Am.*, vol. 132, no. 5, pp. EL436–EL442, 2012.

[9] R. Kumaresan, "An inverse signal approach to computing the envelope of a real valued signal," *IEEE Signal Process. Letters*, vol. 5, no. 10, pp. 256–259, 1998.

[10] Han-Wen Hsu and Chi-Min Liu, "Autoregressive modeling of temporal/spectral envelopes with finite-length discrete trigonometric transforms," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3692–3705, 2010.

[11] J. Herre and J. D. Johnston, "Enhancing the performance of perceptual audio coders by using temporal noise shaping (TNS)," in *Proc. Audio Engg. Soc. 101st Conv.*, Nov 1996.

[12] S. Ganapathy, S. Thomas, and H. Hermansky, "Robust spectro-temporal features based on autoregressive models of Hilbert envelopes," in *IEEE Intl. Conf. Acoust. Speech and Signal Process.*, 2010, pp. 4286–4289.

[13] S. Ganapathy, P. Motlicek, and H. Hermansky, "Autoregressive models of amplitude modulations in audio compression," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 18, no. 6, pp. 1624–1631, 2010.

[14] S. Thomas, S. Ganapathy, and H. Hermansky, "Recognition of reverberant speech using frequency domain linear prediction," *IEEE Signal Process. Letters*, vol. 15, pp. 681–684, 2008.

[15] S. Ganapathy, S. Thomas, P. Motlicek, and H. Hermansky, "Applications of signal analysis using autoregressive models for amplitude modulation," in *Applications of Signal Process. to Audio and Acoust. (WASPAA), IEEE Workshop on*, 2009, pp. 341–344.

[16] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing Principles, Algorithms, and Applications*, Prentice Hall, fourth edition, 2009.

[17] Z. Wang and B. R. Hunt, "The discrete W transform," *Appl. Math. Comput.*, vol. 16, no. 1, Jan. 1985.

[18] D.W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, no. 9, pp. 975–989, 1982.