지구물리 제 4권 제 4호, p. 231-238 (2001년 12월)

Frequency Selective Recursive LP of Harmonic Spectra

Seung Hyon Nam
Dept. of Electronic Engineering, Paichai University

ABSTRACT : In this paper, an efficient LP method for discrete harmonic spectra is proposed and discussed. A new efficient LP method is a combination of recursive and frequency selective LP. While the recursive LP provides better spectral matching in spectral hills, frequency selective LP eliminates numerical instability and improves spectral matching when the harmonics are confined in the low frequency region. The proposed LP method is applied to the HILN coder. Simulation results using a verification model(VM) software for real audio signals show a definite trend of significant improvement.

1. Introduction

As computational speed and data transfer rates of internet and mobile communications improve sharply, requests for necessary technologies to compress audio and speech signals also increase. In this respect, sinusoidal modeling of audio and speech signals has gained lots of research interests during last two decades[1]-[6].

In sinusoidal modeling, audio signals are assumed to be composed of many sinusoids and residual noise. Sinusoids are easily parameterized by their amplitudes, frequencies, and phases. Further, a set of sinusoids is often treated as a harmonic set which can be modeled as a group. Use of a harmonic set improves coding efficiency while preserving harmonic structures of audio signals.

In modeling harmonic sets; efficient modeling and coding of discrete harmonic magnitudes are crucial. One widely used approach in speech coding is to use vector quantization(VQ) of harmonic spectra[2]. The dimension of VQ is fixed while the number of harmonic sinusoids varies. For this reason, discrete harmonic spectra has to be interpolated into the fixed number of harmonic set before applying VQ. In this case, dimension conversion suffers from the loss of around 3dB in case of speech signals.

Another approach is to use all-pole modeling which approximates discrete harmonic spectra using a spectral envelope of the all-pole filter at each harmonic frequency[3],[6]. Linear prediction (LP) is a popular way to obtain an all-pole filter that minimizes the spectral distance between the signal spectrum and the all-pole spectrum[4].

When the signal spectrum is continuous (or smooth enough), LP provides stable all-pole filters which are accurate enough. When the signal spectrum is available at only a set of harmonic frequencies, however, LP suffers from several shortcomings. One of such shortcomings is the poor estimation accuracy evident in the spectral hills which is perceptually important. This problem is addressed by El-Jaroudi and Makhoul[5]. They proposed DAP(Discrete All-Pole) modeling that employs the Itakura-Saito error measure, to solve the problem. Recently, however, both LP and DAP may suffer from numerical instability due to ill-conditioned autocorrelation matrices[7]. In addition, it is analyzed and illuminated that LP can be easily stabilized by applying LP only to limited frequency region selectively.

In this paper, LP modeling of discrete harmonic spectra is discussed and apply to the existing audio coding standard. First of all, linear prediction of discrete harmonic spectra is shortly reviewed in terms of accuracy. A new recursive LP method is then proposed for audio coding. Next, numerical instability inherited in the bandlimited harmonic spectra is addressed and frequency selective LP is proposed. The proposed frequency selective recursive LP is then applied to the existing audio coding standard as part of the MPEG-4 standard. A brief description of the HILN coder and application of the proposed LP method is given. Finally, advantages of the proposed method are discussed through simulation results.

2. Linear Prediction of Discrete Harmonic Spectra

The signal composed of L harmonics can be expressed as

$$s[n]=\sum_{k=1}^{L}A_k\cos(k\omega_0 n+\phi_k)$$

(1)

where $\omega_0=2\pi f_0$ is the fundamental (or pitch) frequency. The power spectrum for this signal is expressed by

$$P(\omega)=\sum_{k=1}^{L}\frac{|A_k|^2}{2}\{\delta(\omega+k\omega_0)+\delta(\omega-k\omega_0)\},$$

(2)

and the autocorrelation coefficient of lag m is given by

$$r_m=\sum_{k=1}^{L}\frac{|A_k|^2}{2}\cos(k\omega_0 m). \tag{3}$$

Most often, information on discrete harmonic spectra $P(k\omega_o)$ are unknown and to be estimated from the signal spectrum which is a mixture of harmonics and noise. Given an p^{th} order all-pole filter

$$H(z)=\frac{1}{\sum_{k=0}^{p}a_k z^{-k}}, \tag{4}$$

the all-pole envelope is defined as

$$\hat{P}(\omega) = \frac{G^2}{\left| \sum_{k=0}^{p} a_k e^{-j\omega k} \right|^2} \quad (5)$$

where G is the gain factor. Then both LP and DAP try to find the best all-pole envelope which matches best to the signal spectrum $P(\omega_k)$.

In LP, the all-pole filter is obtained by minimizing the MSE criterion

$$E_{LP} = \frac{1}{L} \sum_{k=1}^{L} \frac{P(\omega_k)}{\hat{P}(\omega_k)} \quad (6)$$

with respect to a_k. This is accomplished by solving the normal equation

$$\boldsymbol{R}_p \, \boldsymbol{a}_p = - \, \boldsymbol{r}_p \quad (7)$$

where

$$\boldsymbol{R}_p = \begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & r_0 & r_1 & \cdots & r_{p-2} \\ r_2 & r_1 & r_0 & \cdots & r_{p-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & r_0 \end{bmatrix}$$
$$\boldsymbol{a}_p = [a_1 \; a_2 \; a_3 \cdots a_p]^T$$
$$\boldsymbol{r}_p = [r_1 \; r_2 \; r_3 \cdots r_p]^T$$
(8)

Since the error criterion (6) emphasizes the region where $P(\omega) > \hat{P}(\omega)$, spectral error is larger for smaller discrete harmonic spectra. In the following, two methods for improving conventional LP are introduced.

2.1 Recursive LP

To overcome this shortcoming, the following error criterion is considered.

$$E_{RLP} = \frac{1}{L} \sum_{k=1}^{L} \left| \log \frac{P(\omega_k)}{\hat{P}(\omega_k)} \right| . \quad (9)$$

The error criterion is very similar to the one discussed in [4]. The error criterion emphasizes both spectral valleys and hills equally. This property improves the poor estimation accuracy of LP in the spectral peaks which is perceptually important. As a matter of fact, the error criterion (9) becomes identically zero at each discrete frequency ω_k only if $P(\omega_k) = \hat{P}(\omega_k)$. Therefore, if the signal spectrum comes from an all-pole model, there must be a unique solution which is identically equal to the signal spectrum. This property is a clear advantage of the proposed criterion.

Minimization of E_{RLP} with respect to a_n (omitting the factor G^2 for convenience) yields a recursive equation

$$\hat{\boldsymbol{R}}_p \, \boldsymbol{a}_p = - \, \hat{\boldsymbol{r}}_p \quad (10)$$

where

$$\hat{r}_m = \sum_{k=1}^{L} \hat{P}(\omega_k) \cos(k\omega_0 m).$$

This equation is the same as one for LP given in (7) except that autocorrelations are replaced by those computed from estimated harmonic spectra. The equation (10) can be solved recursively with an LP solution as an initial value. Normally, it takes about five recursive steps for convergence to a solution.

2.1 Frequency Selective LP

Since LP requires inversion of the autocorrelation matrix, conditioning of the autocorrelation matrix affects numerical stability of a solution critically. Recently, it is pointed out that LP may suffer from numerical instability if discrete harmonic spectra to be modeled is bandlimited[7]. In practice, discrete harmonic spectra for various audio signals are frequently band limited to the lower frequency region. Further, it is shown that numerical instability may be severe as the model order increases. This numerical instability problem is also common to recursive LP-like methods.

As a solution to this problem, frequency selective LP is proposed. Frequency selective LP is done by redefining the maximum signal frequency correspond to the maximum harmonic frequency $\omega_{\max} = (k+1)\omega_0$. We may call the frequency region from zero up to $\omega_{\max}$ a harmonic bandwidth. Specifically, a set of discrete frequencies $\{\omega_k,\ 1 \le k \le L\}$ is mapped into $\{\tilde{\omega}_k, 1 \le k \le L\}$ by

$$\tilde{\omega}_k = \pi \frac{\omega_k}{\omega_{\max}}. \tag{11}$$

Such frequency redefinition has been proposed originally by Makhoul in order to improve estimation accuracy by applying all-pole filters selectively to various parts of the signal spectrum[4]. In [7], however, its conditioning property that resolves the numerical instability of linear prediction has been explained and illuminated. Clearly, frequency selective LP provides both estimation accuracy and improved stability over full band LP.

3. Application of Frequency Selective Recursive LP to An Audio Coder

Two proposed LP methods in the previous section are combined into frequency selective recursive LP. In order to demonstrate the advantage of the proposed LP method, it is applied to the HILN(Harmonic and Individual Line plus Noise) coder in the MPEG-4 standard[6]. As the name implies, the HILN coder is the parametric audio coder that utilizes sinusoidal models in forms of harmonics and individual tones. The HILN coder offers coding of speech and audio signals at bit rates 4k-16kbps. In the following, the encoding procedure of the HILN is briefly explained[6].

Figure 1 shows the overall encoding operation of the HILN audio coder. The input signal is divided into consecutive frames. The main objective of the encoder is to model the framed signal best and quantize/code within a budget of constrained bits. For this goal, the framed signal is first preprocessed for estimating the envelope and the coarse fundamental frequency. The coarse fundamental frequency is estimated using the complex cepstrum. Later, this coarse estimation is used to find more accurate fundamental frequency. Decomposition of the framed audio into harmonics, individual lines, and noise are done by extracting sinusoidal components in a closed loop manner via analysis-by-synthesis. A psychoacoustic model is employed to find the most perceptually significant sinusoidal component in each iteration. After extracting all the sinusoidal components in the signal, the residual is treated as noise. Parameters for harmonics, individual lines, and noise are estimated and quantized. Parameters for extracted sinusoids may be estimated using a procedure[8]. Quantized parameters are finally encoded and transmitted to the decoder.

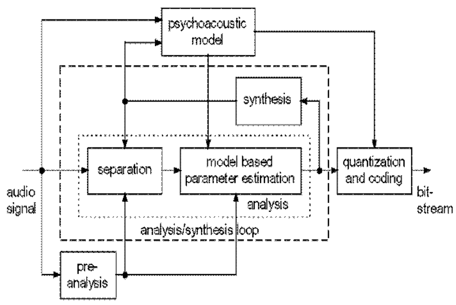

Fig. 1. Block diagram of the HILN encoder.

In audio coders employing a sinusoidal model, extraction of exact sinusoidal components is very critical. Treating noisy peaks as sinusoids or vice versa will affect perceptual quality of synthesized audio signals. In this respect, harmonics should be exactly defined in order to avoid abuse of tonal components. In order words, it is necessary to define a harmonic bandwidth exactly from the valid harmonics. Perceptually significant valid harmonics may be properly defined from their spectral magnitudes compare to masking threshold. In the HILN, LP is applied to the discrete harmonic spectra identified as a harmonic set. The LP order is variable according to the number of discrete spectra. The LP parameters are converted into the LAR(Log Area Ratio), quantized, and transmitted. A harmonic set is regarded as continued one from previous frame if the fundamental frequency is changed within 15%. If a harmonic set is continued from the previous frame, only the difference of power and fundamental frequency is quantized and transmitted.

No additional information is required to be transmitted for using frequency selective recursive LP. The harmonic bandwidth can be recovered easily from received information on the number of harmonics and the fundamental frequency. The LP spectrum is then calculated at each harmonic frequency.

4. Simulation Results

The HILN coder with the proposed LP method is tested for various audio signals at 16kHz sampling frequency. Harmonics, individual lines, and noise modes are turned on at the bit rate of 16kbps. It should be noted that the HILN coder used in the simulation is the non-optimized VM reference software[9]. It does, however, provide meaningful results on the performance of various LP methods.

Spectral distances between the input signals and estimated audio signals are calculated along the frames for the cases LP, recursive LP(RLP), frequency selective LP(FSLP), and frequency selective recursive LP(FSRLP). The spectral distance is defined by the average value of spectral distance at each harmonic frequency. That is,

$$SD = \sum_{k=1}^{L} \left| 10 \log_{10}(P(k\omega_0)) - 10 \log_{10}(\hat{P}(k\omega_0)) \right|.$$

Simulations are performed on an audio item (a cappella song by a female). Figure 2 shows the variation of harmonic bandwidth over the frames. Clearly, harmonic bandwidths are much smaller than the full band in some frames and even shrinks to a quarter of the full bandwidth.

Spectral distance during these frames for various LP method are shown in Fig. 3 and 4. In Fig. 3, we note that RLP yields better spectral matching than LP for all frames. This can be explained as follows. As pointed out in the previous section, the conventional LP becomes unstable from time to time and a stabilization mechanism should be employed. Here, this numerical instability is treated by stopping step-up of order in the Levinson-Durbin algorithm as soon as the error hits some minimum threshold level. In case of RLP, however, model order is allowed to increase as recursive spectral matching continues. FSLP improves spectral matching by providing stable LP process. The simulation result also shows that FSLP does not always better than RLP does. This is the case when the harmonic bandwidth is set to full. Clearly, both LP and FSLP do the same in these frames. Therefore, it is natural to expect the best performance from FSRLP. It is clearly seen from Fig. 4.

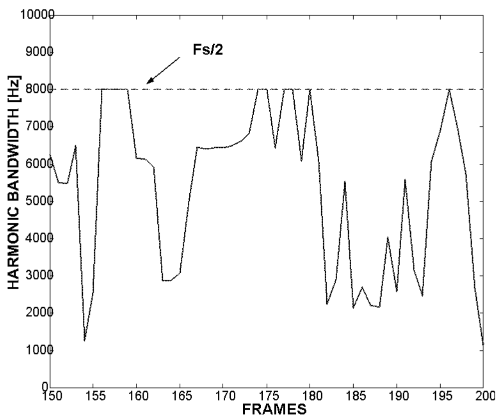

Fig. 2. Variation of harmonic bandwidth over frames.

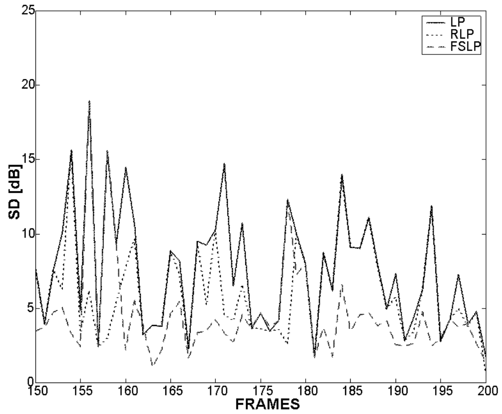

Fig. 3. Spectral distance of harmonics: LP, RLP, FSLP

In order to compare performances of various LP methods, average spectral distance over frames are computed for various audio items as shown in Table 1. It is clear that FSRLP provides the best performance of all LP methods tested in terms of spectral distance. Improvement is larger for speech signals which has the smallest harmonic bandwidth.

In TABLE II, bit consumption for harmonics are compared for various LP methods and audio items. Interestingly, recursive LP demands more bits. This is due to the fact that the harmonic bandwidth is made as large as possible to get better perceptual quality. This policy allows harmonics of small magnitude into a valid harmonic set. This makes LP unstable and stops at lower model order which requires less bits to encode. Therefore, a tradeoff between perceptual quality and bit rate is required.

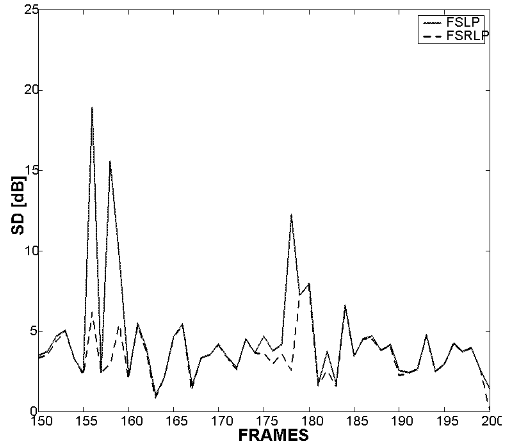

Fig. 4. Spectral distance of harmonics: FSLP and FSRLP.

Table I. Average spectral distance(dB) of harmonics for various audio items.

audio item	vocal	female speech	male speech	clarinet
LP	6.80	7.55	9.00	6.34
RLP	5.31	6.91	8.38	5.58
FSLP	4.21	4.11	5.65	4.79
FSRLP	3.63	3.75	5.28	4.60

Table II. Average bit rates(kbps) of harmonics for various audio items.

audio item	vocal	female speech	male speech	clarinet
LP	2.34	1.97	1.94	1.92
RLP	3.50	2.78	2.66	2.28
FSLP	2.37	2.03	1.97	1.94
FSRLP	3.13	2.56	2.50	2.09

4. Conclusions

In this paper, an efficient LP method for discrete harmonic spectra has been proposed and discussed. A new efficient LP method is a combination of recursive and frequency selective LP. While the recursive LP provides better spectral matching in spectral hills, frequency selective LP eliminates numerical instability and improves spectral matching when the harmonics are confined in a low frequency region. The proposed LP method is applied to the HILN coder. Simulation results using the VM software for real audio signals shows a definite trend of significant improvement. Simulation results show that harmonics generally consumes less bits. Although the proposed LP improves spectral matching significantly, further investigation is necessary to optimize the tradeoff between perceptual quality and bit rate.

References

1. R. J. McAulay and T. F. Quatieri, Speech analysis/synthesis based on a sinusoidal representation, *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP--34, pp. 744-754, 1986.
2. M. Nishiguchi, J. Matsumoto, R. Wakatsuki, and S. Ono, Vector quantized MBE with simplified V/UV division at 3.0kbps, *Proc. of ICASSP*, vol. II, pp. 151-154, 1993.
3. A. Das, A. V. Rao, and A. Gersho, Enhanced multichannel excitation coding of speech at 2.4Kbps with discrete all-pole spectral modeling, *IEEE Global Telecommunications Conference*, pp. 863-866, 1994.
4. J. Makhoul, Spectral linear prediction: Properties and applications, *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 283--296, 1975.
5. A. El-Jaroudi and J. Makhoul, Discrete all-pole modeling, *IEEE Trans. Signal Processing*, vol. 39, pp. 411-423, 1991.
6. ISO/IEC 14496-3:1999, Coding of audio-visual objects. Part 3: Audio (MPEG-4 Audio Version 1), International Standard, 1999.
7. S. H. Nam, Satbilizing discrete spectral modeling of audio signals, *IEEE Signal Processing Letters*, vol. 9, pp. 292-294, Sept. 2002.
8. G. B. George and M. J. T. Smith, Speech Analysis/Synthesis and Modification Using an Analysis-by-Synthesis/Overlap-Add Sinusoidal Model, *IEEE Trans. Signal Processing*, vol. 5, pp. 389-406, Sept. 1997.
9. MPEG-4 Reference Software Edition 2001 (ISO/IEC 14496-5:2001)