Adaptive post-filtering technique based on the modified Yule-Walker filter

Article *in* Proceedings - ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing · January 1999

DOI: 10.1109/ICASSP.1999.758096

Azhar Kassim Mustapha
Nervesis Sdn Bhd

AN ADAPTIVE POST-FILTERING TECHNIQUE BASED ON THE MODIFIED YULE-WALKER FILTER

Azhar Mustapha   *Suat Yeldener*

COMSAT Laboratories, 22300 Comsat Drive, Clarksburg, Maryland, USA
E-mail: [ad_cool , yeldener]@ctd.comsat.com

ABSTRACT

This paper presents an adaptive time-domain post-filtering technique based on the modified Yule-Walker filter. Conventionally, post-filtering is derived from an original LPC spectrum [1]. In general, this time-domain technique produces unpredictable spectral tilt that is hard to control by the modified LPC synthesis, inverse and high pass filtering and causes unnecessary attenuation or amplification of some frequency components that introduces muffling in speech quality. This effect increases when voice coders are tandemed together. Another approach of designing a post-filter was developed by McAulay and Quatieri [2] which can only be used in sinusoidal based speech coders. We have also developed another new time-domain post-filtering technique. This technique eliminates the problem of spectral tilt in speech spectrum that can be applied to various speech coders. The new post-filter has a flat frequency response at the formant peaks of speech spectrum. Instead of looking at the modified LPC synthesis, inverse, and high pass filtering in the conventional time-domain technique, we gather information about the poles of the LPC spectrum in the new technique. This post-filtering technique has been used in a 4 kb/s Harmonic Excitation Linear Predictive Coder (HE-LPC) and a subjective listening tests have indicated that this technique outperforms the conventional one in both one and two tandem connections.

1. INTRODUCTION

A perfect post-filtering technique should not alter the formant information and should attenuate null information in the speech spectrum in order to achieve noise reduction and hence produce better speech quality. Conventionally, time-domain post-filtering techniques use modified LPC synthesis, inverse, and high pass filters that are derived from an LPC spectrum and are configured by the constants: α (for modified synthesis filter), β (for modified inverse filter) and μ (for high pass filter) [1]. However, it is very hard to adapt these coefficients from one frame to another and still produce a post-filter frequency response without spectral tilt. Conventional time-domain post-filtering produces varying spectral tilt from one frame to another affecting speech quality. Another problem with conventional time-domain post-filtering is that, when two formants are close together, the frequency response may have a peak rather than a null between the two formants hence altering the formant information. Yet another effect is that in the original speech, the first formant may have a much higher peak than the second formant, however, the frequency response of the post-filter may have a second formant with a higher peak than the first formant. These phenomena are completely undesirable because they affect the output speech quality. Given the difficulties explained above, we have followed a different approach to design a post-filter which uses the pole information in the LPC spectrum and finds the relation between poles and formants. The new post-filtering technique uses adaptive multi band pass filtering based on the modified Yule-Walker filter [3] that can be designed to achieve good post-filtering and hence better speech quality than the conventional technique.

2. POLES AND FORMANTS

Generally, pole angles in an LPC spectrum have information about formant locations and associated bandwidths. Given that an LPC spectrum is defined as $1/(1 - A(z))$ where $A(z) = \sum_{i=1}^{M} a_i z^{-i}$, a_i is the i-th LPC coefficient, and M is the order of the LPC predictor, we can find the poles by solving for roots of $1 - A(z)$. In solving for the roots, $1 - A(z)$ is turned into a companion matrix [4]. The companion matrix is used to find the eigenvalues which are the roots of $1 - A(z)$. In finding the eigenvalues, QR (Q = Orthogonal Columns and R = Upper Triangular) algorithm for real Hessenberg matrices can be implemented. Further explanation can be found in [4].

Naturally, poles exist in conjugate pairs although two real poles might exist. If two real poles exist, they always have an angle of 0 and π. Noting this symmetrical property, the poles can be divided into a group of positive angles and a group of negative angles. For each group, the radii can be arranged in descending order so that r_1 is the longest radius in the positive group and r_8 is the longest radius in the negative group. Notice also that the longest radius has the shortest distance to the unit circle since all the radii are less than 1. With this arrangement, r_1 and r_8 have the same radius and occur in conjugate angles.

To analyze the relation between poles and formants, a typical LPC spectrum is plotted with the pole angles located on the normalized frequency axis as shown in Fig. 1.

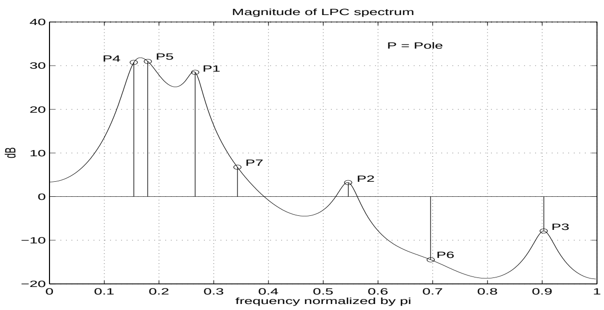

Figure 1: Relation between poles and formants

In this figure, the locations of poles 1 through 7 are noted by P1 through P7. Poles P1, P2 and P3 indicate the exact locations of the formant peaks. However, the first 3 poles are not always located at the peaks as shown in this example. In general, a wide formant bandwidth has two or three poles that are close together. This fact can be observed in Fig. 1 where the bandwidth of the first formant is wider than the second formant. The first formant has pole 4 and pole 5 that are close together while the other formants only have a single pole. By observation in the example, 5 poles need to be considered to estimate the locations of formants and associated bandwidths. However, we still consider pole 6 and pole 7 because these poles might be a part of a formant themselves. With knowledge of the locations of the seven poles, we can start estimating formants and nulls.

3. ESTIMATING FORMANTS AND NULLS

In order to estimate formants and nulls, the following steps are followed. First, the positive angles of the poles are arranged in ascending order. The negative angles are omitted due to the symmetrical property of the angles as mentioned previously. The magnitude response for any given angle, ω is then computed as:

$$H(\omega) = \Pi_{i=1}^{14} \sqrt{1 + r_i^2 - 2r_i \cos(\phi)} \quad (1)$$

where r_i is radius of pole P_i and $\phi = \theta_i - \omega$; ω is any given angle and θ_i is the angle of the pole P_i. In the next step, the backward and forward slopes of the neighboring angles are computed as:

$$\begin{aligned} m_1 &= H(\theta_i + \delta\omega) - H(\theta_i) \\ m_2 &= H(\theta_{i+1}) - H(\theta_{i+1} - \delta\omega) \end{aligned} \quad (2)$$

where m_1 and m_2 are the i^{th} forward and $(i+1)^{th}$ backward slopes of the two neighboring angles, respectively and $\delta\omega$ is a perturbation factor for each angle. The computed slopes of the neighboring angles are then compared. If $m_1 < 0$ and $m_2 > 0$, then it is assumed that a null between two angles exist and these two poles are treated as two independent formants. If the above condition is not satisfied, then the magnitude responses of the angles are compared. In this case, if $|H(\theta_i) - H(\theta_{i+1})| < 3$ dB, then both of these poles are treated as one formant. Otherwise, the pole with larger magnitude response is treated as a formant. 3dB was determined experimentally to be the optimal threshold. This process is repeated throughout all positive angles and hence all formants and nulls are estimated.

4. MODIFIED YULE-WALKER POST-FILTER

Estimated formant locations and number of poles for each formant are used to compute the bandwidths of the formants and eventually the frequency response of the desired post-filter. In the case of a formant with a single pole, the bandwidth of the corresponding formant is set to be $2\delta b$, where $\delta b = 0.04\pi$. For example, if the formant pole is assumed to be at θ_1, then the bandwidth of the corresponding formant will cover the frequency range from $\theta_1 - \delta b$ to $\theta_1 + \delta b$. In the example shown in Fig. 1, poles P1, P2 and P3 are the single pole formants.

In the case of a formant with multiple poles (2 or 3 poles), the bandwidth of the corresponding formant should cover the all corresponding pole locations. According to the example given in Fig. 1, poles P4 and P5 correspond to the first formant of the spectrum and the bandwidth of this formant ranges from $\theta_4 - \delta b$ to $\theta_5 + \delta b$, where θ_4 and θ_5 are the locations of poles P4 and P5 respectively. During estimation of formants and their bandwidths, the bandwidth of 2 formants might overlap each other when 2 formants are very close. This overlapping creates a problem in designing this post-filter. In order to avoid this problem, the bandwidths of these two formants are combined together to form only one band.

In this post-filter, the aim is to preserve the formant information. Therefore, the post-filter will have a unity gain on the formant regions of spectrum. Outside of the formant regions, the aim is to have some controllable attenuation factor, τ that controls the depth of the post-filtering. In our example, we set $\tau = 0.6$. However, τ can be adapted from one frame to another depending on how much post-filtering is needed and the type of speech coder used. The frequency response of the desired post-filter is shown in Fig. 2 for the envelope illustrated in Fig. 1.

In order to design a post-filter to have the features mentioned above, an adaptive multi band pass filter is required. Such an adaptive multi band pass filter can be implemented using a modified Yule-Walker (MYW) recursive filter. The form of this filter can be formulated as:

$$\frac{B(z)}{A(z)} = \frac{b(1) + b(2)z^{-1} + ... + b(N)z^{-(N-1)}}{1 + a(1)z^{-1} + ... + a(N)z^{-(N-1)}} \quad (3)$$

where N is the order of the MYW filter. The (MYW) filter coefficients are estimated using a least squares fit in the time domain. The denominator coefficients of the filter ($a(1)$, $a(2)$, ..., $a(N)$) are computed by the Modified Yule-Walker equations using non-recursive correlation coefficients computed by inverse Fourier transformation of the specified frequency response of the post-filter [3]. The numerator coefficients of the filter ($b(1)$, $b(2)$, ..., $b(N)$) are computed by a 4 step procedure: first, a numerator polynomial corresponding to an additive decomposition of the power frequency response is computed. The complete frequency response corresponding to the numerator and denominator polynomials is then evaluated. As a result, a spectral factorization technique is used to obtain the impulse response of the filter. Finally, the numerator polynomial is obtained by a least squares fit to this impulse response. A more detailed description of this algorithm is given by Friendlander and Porat [3].

This post-filter described above has a flat frequency response that overcomes the spectral tilt and other problems present in conventional post-filters as mention earlier in this paper. In order to view the differences between this and conventional post-filters, the frequency responses of these filters applied to the LPC spectrum shown in Fig. 1, are given in Fig. 2.

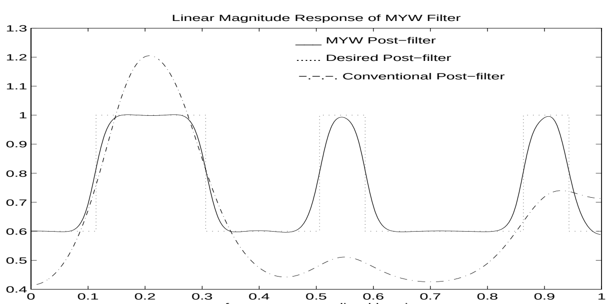

Figure 2: Frequency response of post-filters

The conventional post-filter uses $\alpha = 0.8, \beta = 0.5$ and $\mu = 0.5$ as suggested by Chen in [1]. From Figure 2, it is clear that the formant peaks are maintained to be flat in the frequency response of the new MYW post-filter. However, the conventional post-filter is not flat at formant peaks. The new and the conventional post-filtered LPC spectra are shown in Fig. 3: For the conventional post-filter, it is clear that there is a spectral tilt compared with the original LPC spectrum. For the new post-filter, there is no any spectral tilt at all. The new filter preserves the formant peaks and attenuates the nulls which is the desired phenomenon. In

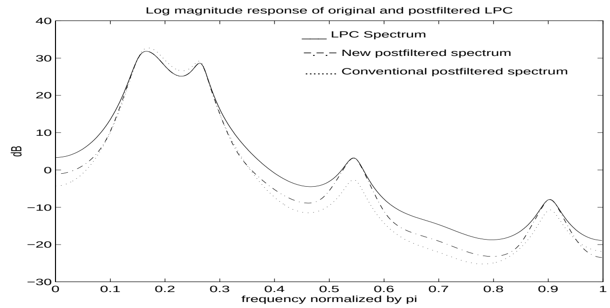

Figure 3: Post-filtered LPC spectra

addition, the attenuation of nulls can be more controllable in the new post-filter than in the conventional post-filter.

5. EXPERIMENTAL VALIDATION

This new post-filter is incorporated into our 4 kb/s Harmonic Excitation Linear Predictive Coder (HE-LPC). In the HE-LPC coder, the approach to represent the speech signals $s(n)$ is to use the speech production model in which speech is viewed as the result of passing an excitation, $e(n)$ through a linear time-varying filter (LPC), $h(n)$, that models the resonant characteristics of the speech spectral envelope [5][6]. The $h(n)$ is represented by 14 LPC coefficients which are quantized in the form of Line Spectral Frequency (LSF) parameters. In the HE-LPC speech coder, the excitation signal $e(n)$ is specified by a fundamental frequency or pitch, its spectral amplitudes, and a voicing probability. The voicing probability defines a cut-off frequency that separates low frequency components as voiced and high frequency components as unvoiced. The computed model parameters are quantized and encoded for transmission.

At the receiving end, the information bits are decoded, and hence, the model parameters are recovered. At the decoder, the voiced part of the excitation spectrum is determined as the sum of harmonic sine waves. The harmonic phases of sine waves are predicted using the phase information of the previous frames. For the unvoiced part of the excitation spectrum, a white random noise spectrum normalized to unvoiced excitation spectral harmonic amplitudes is used. The voiced and unvoiced excitation signals are then added together to form the overall synthesized excitation signal. The resultant excitation is then shaped by the linear time-varying filter, $h(n)$, to form the final synthesized speech. Finally, the synthesized speech is passed through the new and conventional post-filters, in order to evaluate the performance of each of these filters.

6. SUBJECTIVE LISTENING TESTS

In order to measure the subjective performance of the new and conventional post-filters, various listening tests were conducted at COMSAT Laboratories. For this purpose, two post-filters were separately used in the same 4 kb/s HE-LPC coder for subjective performance evaluation purposes. In the first experiment, an MOS test was conducted. In this test, 8 sentence pairs for 4 speakers (2 male and 2 female speakers) were processed by the two 4 kb/s coders. Altogether 24 listeners performed this test. Both one and two tandem connections of these coders are evaluated and the MOS results are given in Table 1.

Coder	**MOS Scores**	
	1 Tandem	*2 Tandem*
4 kb/s Coder With Conventional Post-filter	3.41	2.40
4 kb/s Coder With New Post-filter	3.55	2.75

Table 1: MOS scores for conventional and new post-filters

From these test results, it is clear that, the 4 kb/s coder with the new post-filter performed better than the coder with conventional post-filter. The improvement of speech quality attributable to the new post-filter is very substantial in the 2 tandem connection case.

To further verify the performance of the new post-filter, a pair-wise listening test was conducted to compare the 4 kb/s coders with the conventional and new post-filters. For this test, 12 sentence pairs for 6 speakers (3 male and 3 female speakers) were processed by the two 4 kb/s coders (for 1 and 2 tandem connection conditions) and the sentence pairs were presented to the listeners in a randomized order. Sixteen listeners performed this test. The overall test results for 1 and 2 tandem connections are shown in Tables 2 and 3, respectively.

Preferences		
No of Votes	**%**	*Preferred Coder*
21	10.9	New Post-filter (Strong)
60	31.3	New Post-filter
75	39.1	Similar
29	15.1	Conventional Post-filter
7	3.6	Conventional Post-filter (strong)

Table 2: Pair-wise test results for 1 tandem connection

Preferences		
No of Votes	**%**	*Preferred Coder*
30	15.6	New Post-filter (Strong)
79	41.1	New Post-filter
65	33.9	Similar
16	8.3	Conventional Post-filter
2	1.1	Conventional Post-filter (strong)

Table 3: Pair-wise test results for 2 tandem connection

The results are very conclusive. In the 1 tandem connection case, the new post-filter was found to be slightly better than the conventional post-filter. In the 2 tandem connection case, the new post-filter was found to be superior over conventional post-filter.

7. CONCLUSION

In this paper, a new post-filtering technique based on the modified Yule-Walker filter was described. In the designing steps, first the relation between poles and formants and then the estimation of the formants and their bandwidths were given. The information about the formants and their bandwidths were then used to design the modified Yule-Walker filter based on a least squares fit in time domain. Finally, subjective listening test results of the new and conventional post-filters were given. The test results indicated that the new post-filter outperforms the conventional post-filter in both 1 and 2 tandem connection cases of the voice coders.

Acknowledgment: Authors would like to acknowledge the valuable comments from Thomas F. Quatieri.

8. REFERENCES

[1] Juiun-Hwey Chen, Allen Gersho “Adaptive Post-filtering For Quality Enhancement of Coded Speech”, IEEE Trans. Speech & Audio Proc., vol.3, no.1, pp.59-71, 1995

[2] R. McAulay,T. Parks, T. Quatieri, M. Sabin “Sine-Wave Amplitude Coding At Low Data Rates”, Advances in Speech Coding, Kluwer Academic Pub., 1991, edited by B.S. Atal, V. Cuperman and A. Gersho, pp:203-214.

[3] B. Friedlander, B. Porat “The Modified Yule-Walker Method of ARMA Spectral Estimation” IEEE Trans. on Aerospace Electronic Systems. Vol: AES-20, No:2 pp:158-173, March 1984

[4] J.H. Wilkinson. C. Reinsch “Linear Algebra: Hand Book for Automatic Computation” Springer-Verlag New York Heidelberg Berlin 1971.

[5] S. Yeldener, A. M. Kondoz, B. G. Evans ”Multi-Band Linear Predictive speech coding at very low bit rates” IEE Proc. Vis. Image and Signal Processing, October 1994, Vol:141 No:5 pp:289-295.

[6] S. Yeldener, A. M. Kondoz, B. G. Evans ”Sine Wave Excited Linear Predictive Coding of Speech” Proc. Int. Conf. on Spoken Language Processing, Kobe, Japan, November 1990, pp. 4.2.1 - 4.2.4.