An-Najah Univ. J. Res. (N. Sc), Vol. 16 (2), 2002

The Role of Pitch Filter in Pulse-by-Pulse Reoptimization of the LP Synthesis Filter

دور المرشح الخطي طويل المدى على طريقة النبضة – نبضة لتحسين أداء المرشحات الخطية التجميعية

Allam Mousa

Department of Electrical Engineering, An-Najah National University, Nablus, Palestine.

Received: (27/11/2000), Accepted: (29/12/2001)

Abstract

Two iterative analysis algorithms were developed for the reoptimization of the LP synthesis filter based on a pulse-by-pulse reoptimization manner. In this study, the use of the pitch filter in the analysis algorithms is introduced. Similar to the no pitch case, improvement in the gain is achieved. On the other hand, this gain has dropped compared to the no pitch case. Moreover, the number of pulses needed to reoptimize the LP filter found to be much less than that, in the no pitch case.

ملخص

لقد تم تطوير طريقتين تحليليتين من اجل تحسين أداء المرشحات الخطية المتوقعة التجميعية وذلك بالاعتماد على عدد النبضات المستخدمة في العملية التجميعية. وفي هذا البحث فلقد تمت دراسة تأثير وجود المرشح الخطي طويل المدى على أداء الخوارزميات حيث لوحظ أن سلوك الخوارزمية له نفس النمط السلوكي بوجود هذا المرشح أو عدمه إذ لوحظ تحقق تحسن في مقدار الكسب العام غير أن مقدار هذا الكسب قد انخفض عما هو عليه الحال مع عدم وجود هذا المرشح. و من ناحية أخرى فقد لوحظ أن عدد النبضات اللازمة من أجل تحسين أداء المرشحات هي أقل من تلك في الحالة المتمثلة بغياب هذا المرشح.

I. Introduction

The conventional cascade connections of the formant and pitch prediction filters were considered in [1], where the coefficients of these predictors are defined for one prediction filter then for the other one, sequentially. The formant filter is used to remove near-sample redundancies and the pitch filter is introduced to remove the distant-sample redundancies in the speech signal. The joint reoptimization of the formant and pitch filter coefficients was also discussed in several methods [2-6]. It has been shown in [2] that the joint reoptimization solution provides higher prediction gain compared to that of the conventional formant-pitch sequential one.

For improving the performance of the Analysis-by-Synthesis (A-b-S) speech coders, it has been shown that the reoptimization of the formant filter coefficients, jointly, with the excitation parameters, provides better performance as shown in [3-4]. A different approach for parameter reoptimization of the formant filter together with the excitation and considering the pitch filter in the analysis algorithm was also discussed in [5]. Another joint reoptimization of the Linear Prediction Coder (LPC) and pitch filter parameters in Code Excited Linear Prediction (CELP) was presented in [6].

A simplified analysis algorithm for joint reoptimization of the formant filter and the excitation was discussed in [7], where two iterative algorithms had given a considerable gain in performance. These two algorithms depend only on the pulse locations and not on their amplitudes as in [3-4]. Hence it is referred to as "pulse-by-pulse reoptimization analysis algorithms". However, the pitch filter was not considered in the analysis algorithms in [7] and the overall Segmental Signal-to-Noise Ratio (SEGSNR) was improved by 1.7 to 2.6 dB.

In this paper, the pitch filter is introduced to the pulse-by-pulse reoptimization algorithms discussed in [7], and the results are analyzed, with and without using the pitch filter, refereed to as (COD-WP) and (COD-NP), respectively. The former coder has shown similar behavior like the later in achieving an improvement in the SEGSNR, but the amount of gain has dropped. Another major difference is in the significant drop of the number of iterations required, by COD-NP, to reach saturation.

Computer simulations are performed using Multi-Pulse (MP) excitation to show the effect of including the pitch filter in the reoptimization process, and the complexity of the new algorithm is discussed.

II. Pulse-Based LP Analysis Including the Pitch Filter

The pulse-based LP analysis, including the pitch filter, was introduced in [5]. Assuming a block of data with N+p samples of speech is available, the predicted speech samples vector can be written as;

$$\hat{\mathbf{s}} = \mathbf{S}\overline{\alpha} + \mathbf{D}\overline{\beta} + \mathbf{HG} \tag{1}$$

or

$$\hat{\mathbf{s}} = \begin{bmatrix} \mathbf{S} & \mathbf{D} & \mathbf{H} \end{bmatrix} \begin{bmatrix} \overline{\alpha} \\ \overline{\beta} \\ \mathbf{G} \end{bmatrix} \tag{2}$$

Where $\overline{\alpha} = [\alpha_1\ \alpha_2 \ldots \alpha_p]^t$ is the vector of the Short Term Prediction (STP) coefficients, $\overline{\beta} = [\beta_1 \ \ \ldots \ \ \beta_{\mathrm{np}}]^t$ is the vector of the Long Term Prediction (LTP) coefficients, $\mathbf{G} = [\mathbf{g_1} \ \ \mathbf{g_2} \ \ \ldots \ \ \mathbf{g_M}]^t$ is the vector of the pulse amplitudes, $\mathbf{s} = [\mathrm{s}(0), \mathrm{s}(1), \mathrm{s}(2), \ldots, \mathrm{s}(\mathrm{N}-1)]^t$ is the vector of the speech samples, **S** is the Nxp data matrix and **H** is the NxM position matrix, given by S_{ij}=s(i-j), $H_{ij} = \delta(i - \tau_j)$, respectively, and **D** is the intermediate data matrix constructed from the formant residual which is obtained by inverse filtering.

The equation, which gives the minimum norm solution of Eq. (2), is

$$\begin{bmatrix} \overline{\alpha} \\ \overline{\beta} \\ \mathbf{g} \end{bmatrix} = \begin{bmatrix} \mathbf{S}^t\mathbf{S} & \mathbf{S}^t\mathbf{D} & \mathbf{S}^t\mathbf{H} \\ \mathbf{D}^t\mathbf{S} & \mathbf{D}^t\mathbf{D} & \mathbf{D}^t\mathbf{H} \\ \mathbf{H}^t\mathbf{S} & \mathbf{H}^t\mathbf{D} & \mathbf{H}^t\mathbf{H} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{S}^t\mathbf{s} \\ \mathbf{D}^t\mathbf{s} \\ \mathbf{H}^t\mathbf{s} \end{bmatrix} \tag{3}$$

It is a very difficult task to obtain the minimum norm solution as given in Eq. (3). In fact, direct attempt to solve Eq. (3) requires the inversion of a (p+M+np) x (p+M+np) matrix which is impractical for typical values of p and M. A formal description of the algorithm which solves, the STP, LTP and pulse gain, jointly, is discussed in [5]. Hereafter, this algorithm will be referred to as ALG1.

III. The Effect of Pitch Filter in the Pulse-by-Pulse LP Reoptimization

In the pulse-by-pulse linear prediction analysis algorithm descried in [7], two computationally efficient iterative algorithms were developed with no matrix inversion except the one at their initial steps. However, the pitch filter was not taken into consideration in the reoptimization process, and so the pulse excited linear prediction problem was stated as the determination of the filter parameters, $\overline{\alpha}_{(k)}$, k pulse amplitudes, $\mathbf{g}_{(k)}$, and k-th pulse location, τ_k, assuming that the previous pulse locations, $\tau_1, \tau_2, \ldots, \tau_{k-1}$, are known. However, in this section, the pitch filter is integrated with the pulse-by-pulse reoptimization analysis algorithm. This, in fact, jointly reoptimizes the LTP coefficients, $\overline{\beta}_{(k)}$, assuming that the LTP lag, T, is known.

At the k-th iteration, the prediction speech sample at time n can be written as

$$\hat{s}_{(k)}(n) = \sum_{i=1}^{p} \alpha_{(k)i} s(n-i) + \sum_{i=1}^{np} \beta_{(k)i} d(n-T-i+1) + \sum_{i=1}^{k} g_{(k)i} \delta(n-\tau_i) \qquad (4)$$

This can be written in matrix form as

$$\hat{\mathbf{s}}_{(k)} = \begin{bmatrix} \mathbf{S} & \mathbf{D} & \mathbf{H}_{(k)} \end{bmatrix} \begin{bmatrix} \overline{\alpha}_{(k)} \\ \overline{\beta}_{(k)} \\ \mathbf{g}_{(k)} \end{bmatrix} \qquad (5)$$

or

$$\hat{\mathbf{s}} = \begin{bmatrix} \Omega_{(k-1)} & \mathbf{h}_{(k)} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{(k)} \\ \mathbf{g}_{k(k)} \end{bmatrix} \qquad (6)$$

Where $\mathbf{C}_{(k)} = \begin{bmatrix} \overline{\alpha}_{(k)} & \overline{\beta}_{(k)} & g_{1(k)} & \ldots & g_{k-1(k)} \end{bmatrix}^t$ is the augmented coefficient-gain vector holding the STP and LTP coefficients and the pulses gain, $\Omega_{(k-1)} = \begin{bmatrix} \mathbf{S}\,\mathbf{D}\,\mathbf{H}_{(k-1)} \end{bmatrix}$, $\mathbf{H}_{(k-1)}$ is the pulse position matrix corresponding to τ_1, τ_2, … τ_{k-1}, $\mathbf{h}_{(k)}$ is the position vector corresponding to the unknown τ_k, and **S** and **D** are as defined before. Note that $\Omega_{(0)} = [\mathbf{S}\ \mathbf{D}]$.

The equation that gives the minimum norm solution of Eq. (6) is

$$\begin{bmatrix} \mathbf{C}_{(k)} \\ g_{k(k)} \end{bmatrix} = \begin{bmatrix} \Omega^t_{(k-1)}\Omega_{(k-1)} & \Omega^t_{(k-1)}\mathbf{h}_{(k)} \\ \mathbf{h}^t\Omega_{(k-1)} & \mathbf{h}^t_{(k)}\mathbf{h}_{(k)} \end{bmatrix}^{-1} \begin{bmatrix} \Omega^t_{(k-1)}\mathbf{s} \\ \mathbf{h}^t_{(k)}\mathbf{s} \end{bmatrix} \tag{7}$$

Defining, $\mathbf{R}_{(k)} = \mathbf{S}^t_{(k)}\mathbf{S}_{(k)}$, $\mathbf{r}_{(k)} = \mathbf{S}^t_{(k)}\mathbf{s}$, $\mathbf{w}_k = \mathbf{h}^t_{(k)}\mathbf{S}_{(k-1)}$, and noting that $\mathbf{h}^t_{(k)}\mathbf{h}_{(k)} = 1$ and $\mathbf{h}^t_{(k)}\mathbf{s} = s(\tau_k)$, Eq. (7) can be rewritten as

$$\begin{bmatrix} \mathbf{C}_{(k)} \\ g_{k(k)} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{(k-1)} & \mathbf{w}^t_{(k)} \\ \mathbf{w}_{(k)} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{r}_{(k)} \\ \mathbf{s}(\tau_k) \end{bmatrix} \tag{8}$$

However, it is more informative to write

$$\mathbf{w}_{(k)} = [\mathrm{s}(\tau_k - 1)\ \ \mathrm{s}(\tau_k - 2) \ldots \mathrm{s}(\tau_\mathrm{k} - \mathrm{p})\ \ \mathrm{d}(\tau_\mathrm{k} - \mathrm{T})\ \ \mathrm{d}(\tau_\mathrm{k} - \mathrm{T} + 1 - \mathrm{np})\ \ \Phi] \tag{9}$$

where Φ is a 1x(k-1) null matrix.

Solving and rearranging Eq. (8) give the two-coupled equations;

$$g_{(k)} = \mu^{-1} e_{(k-1)}(\tau_k) \tag{10}$$

$$\mathbf{C}_{(k)} = \mathbf{C}_{(k-1)} - g_{(k)}\mathbf{R}^{-1}_{(k-1)}\ \mathbf{w}^{-1}_{(k)} \tag{11}$$

where $\mu = 1 - \mathbf{w}_{(k)}\mathbf{R}^{-1}_{(k-1)}\mathbf{w}^t_{(k)}$, and $e_{(k-1)}(\tau_k)$ is the prediction error of the corresponding STP-LTP cascaded filter at time τ_k, which is given by;

$$e_{(k-1)}(\tau_k) = s(\tau_k) - \sum_{j=1}^{p}\alpha_{(k-1)j}s(\tau_k - j) - \sum_{j=1}^{np}\beta_{(k-1)j}d(\tau_k - T - j + 1) \tag{12}$$

Eq. (11) requires updating the inverse of $\mathbf{R}_{(k)}$ which can be written as;

$$\mathbf{R}^{-1}_{(k)} = \frac{1}{\mu}\begin{bmatrix} \mu\mathbf{R}^{-1}_{(k-1)} + \mathbf{Z}\mathbf{Z}^t & -\mathbf{Z} \\ -\mathbf{Z}^t & 1 \end{bmatrix} \tag{13}$$

where $\mathbf{Z} = \mathbf{R}_{(k-1)}^{-1}\mathbf{w}_{(k)}^{t}$, and, initially, the inverse of $\mathbf{R}_{(0)}$ is required which can be related to the inverse of the usual covariance matrix **R** constructed from **S** and **D**. Assuming the pulse locations $\tau_k, k = 1,2,...M$ are available, the algorithm based on the recursive Eqs. (10) and (11) can be stated as follows:

Step 0: Given; M, p, np, **s**, **d**, and $\{\tau_i\}_{i=1}^{M}$.

Step 1: Do the conventional covariance analysis to find the STP parameters, say $\overline{\alpha}_c$, and $\mathbf{R}^{-1}$ explicitly.

Step 2: Determine the LTP parameters as in the conventional method.

Step 3: Construct $\mathbf{R}_{(0)}^{-1}$.

Step 4: Obtain the parameters $\overline{\alpha}_c$ and $\overline{\beta}_c$ from the equation

$$\begin{bmatrix} \overline{\alpha}_c \\ \overline{\beta}_c \end{bmatrix} = \mathbf{R}_{(0)}^{-1}\mathbf{r}_{(0)}$$

Step 5: Obtain the residual e(n) and estimate the pulse locations, $(\tau_1, \tau_2, ..., \tau_M)$.

Step 6: Set $\mathbf{C}_{(0)} = \begin{bmatrix} \overline{\alpha}_c & \overline{\beta}_c \end{bmatrix}$, and k=1.

Step 7: Compute the residual sample $e_{(k-1)}(\tau_k)$ using Eq. (12).

Step 8: Determine $\mathbf{g}_{(k)}$ and $\mathbf{C}_{(k)}$ using Eqs. (10) and (11), respectively.

Step 9: Update $\mathbf{R}_{(k)}^{-1}$ using Eq. (13).

Step 10: Set k=k+1. If k<M go to step 7.

Step 11: Set $\mathbf{C} = \mathbf{C}_{(M)}$ and $\mathbf{g} = \mathbf{g}_{(M)}$.

Note that the solution at the M-th iteration corresponds to Eq. (3). Hereafter, this algorithm will be referred to as ALG2.

Similar to the no pitch case discussed in [7], a computationally more efficient, but suboptimal, algorithm can be derived if the problem is stated as the determination of the k-th pulse location τ_k, amplitude g_k and the predictor coefficients $\mathbf{C}_{(k)}$, given all pulse locations and amplitudes up to (k-1). The resulting equations are identical to those given in Eqs. (10) and (11) except that $\mathbf{R}_{(k-1)}^{-1}$ is replaced by $\mathbf{R}_{(0)}^{-1}$, that is:

$$\mathbf{C}_{(k)} = \mathbf{C}_{(k-1)} - g_{(k)}\mathbf{R}^{-1}\mathbf{w}_{(k)}^{-1} \tag{14}$$

where,

$$\mu = 1 - \mathbf{w}_{(k)}\mathbf{R}_{(k-1)}^{-1}\mathbf{w}_{(k)}^{t} \tag{15}$$

This drops step 9 from ALG2. Since, at each iteration, the previous pulse amplitudes are fixed, the algorithm is unable to give the result in Eq. (3) at the final iteration. That is why it is not optimal. A formal description of the algorithm (say ALG3) is as follows;

Step 0: Given; M, p, np, **s**, **d**, and $\{\tau_i\}_{i=1}^{M}$.

Step 1: Do the conventional covariance analysis to find the STP parameters, say $\overline{\alpha}_c$, and $\mathbf{R}^{-1}$ explicitly.

Step 2: Determine the LTP parameters as in the conventional method.

Step 3: Construct $\mathbf{R}_{(0)}^{-1}$.

Step 4: Obtain the parameters $\overline{\alpha}_c$ and $\overline{\beta}_c$ from the equation $\begin{bmatrix}\overline{\alpha}_c \\ \overline{\beta}_c\end{bmatrix} = \mathbf{R}_{(0)}^{-1}\mathbf{r}_{(0)}$

Step 5: Obtain the residual e(n) and estimate the pulse locations, ($\tau_1, \tau_2, \ldots, \tau_M$).

Step 6: Set $\mathbf{C}_{(0)} = \begin{bmatrix}\overline{\alpha}_c & \overline{\beta}_c\end{bmatrix}$, and k=1.

Step 7: Compute the residual sample $e_{(k-1)}(\tau_k)$ using Eq. (12).

Step 8: Determine $\mathbf{g}_{(k)}$ and $\mathbf{C}_{(k)}$ using Eqs. (10) and (14), respectively.

Step 9: Set k=k+1. If k<M go to step 7.

Step 10: Set $\mathbf{C} = \mathbf{C}_{(M)}$ and $\mathbf{g} = \mathbf{g}_{(M)}$.

IV. Experimental Results

To show the effect of including the pitch filter in the pulse-by-pulse reoptimization, experiments were conducted for the following MP excited coders:

1. conventional coder (COD1)
2. coder that uses only the pulse locations of the A-b-S excitation with ALG2 (COD2)
3. coder that uses only the pulse locations of the A-b-S excitation with ALG3 (COD3)

The speech database was taken from both male and female speakers. Speech was band limited then sampled at 8 kHz and quantified by 16 bits. The number of the LP coefficients, p, was fixed to ten and that of the LTP coefficients, np, was fixed to one, the LTP lag search was performed in the range 20 to 147 samples. No windowing or preemphasis was applied. The analysis frame length, N, was set to 200 samples and so was the parameter-updating rate. The covariance analysis method was used with COD1 and the stabilized covariance method was used with COD2 and COD3, the excitation frame was set to 50 samples and the LTP update frame was also 50 samples, the weighting factor, μ, was set to 0.8.

The Segmental Signal-to-Noise ratio (SEGSNR) was used as an objective test of the performance of the coders. Segmental SNR was calculated as:

$$\text{Segmental SNR (dB)} = \frac{1}{N_f}\sum_{j=1}^{N_f} SNR_j\,(dB)$$

Where,

$$SNR_j \text{ (dB)} = 10\log\left(\frac{\sum_{n=1}^{K} s^2(n)}{\sum_{n=1}^{K} (\hat{s}(n) - s(n))^2}\right)$$

is the Signal-to-Noise ratio of the j-th segment, $s(n)$ and $\hat{s}(n)$ are the actual and reconstructed speech samples respectively, N_f is the number of frames and K is the number of samples per segmental frame which is taken to be 100 samples.

Fig.1 shows the SEGSNR versus the number of pulses per frame for COD1, COD2 and COD3 with pitch (WP) filter included in the analysis algorithms. The improvement due to the reoptimization used in COD2 is clear. It can also be seen in Fig.2 that the loss in COD3 due to the suboptimal algorithm, ALG3, is small and even negligible. Hence, the simple analysis algorithm, ALG3, turned out to be efficient and it provides almost the same performance as ALG2 but with much less complexity.

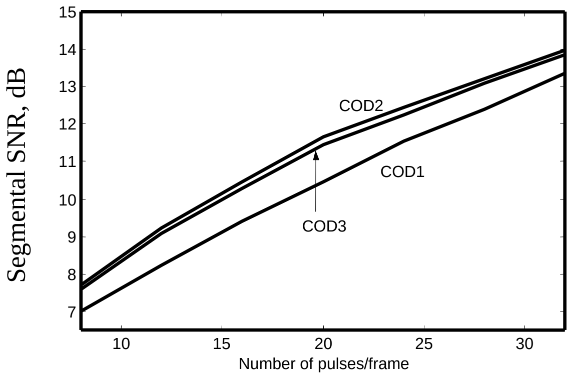

Figure 1: SEGSNR versus pulse rate for COD1, COD2 and COD3, with pitch filter included

Fig. 2 compares the performance of COD1 and COD3 for the two cases with and without a pitch filter. It is obvious in Fig.2 that the two cases have similar behavior, but with a drop in the gain obtained by the reoptimization when the pitch filter is included, COD3-WP, compared to the gain in the reoptimization of the STP parameters in the absence of the LTP filter, COD3-NP. For example, at the pulse rate of 20 pulses per frame, the gain obtained in the reoptimization process using COD3-NP is 1.8 dB, whereas it is only 1.2 dB in COD3-WP.

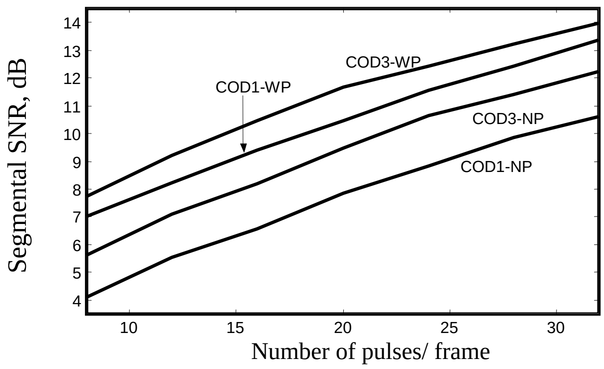

Figure 2: SEGSNR versus pulse rate for COD1 and COD3 with and without a pitch filter.

Another important difference in behavior of the reoptimization algorithm in COD3-WP compared to COD3-NP is shown in Fig. 3 which reflects the SEGSNR versus the number of pulses used in the reoptimization analysis algorithm, ALG3. The number of excitation pulses was chosen to be 20 pulses per frame, and so the pulse-by-pulse reoptimization analysis algorithm uses the pulses starting from no pulse (conventional method) up to 20 pulses, and the resulting SEGSNR was recorded for each number of pulses used. As it can be easily seen, COD3-NP performance is gradually increasing with the increasing number of pulses used in the analysis algorithm. However, when the pitch filter is used, the performance, rapidly, reaches saturation after using the first three pulses out of the 20 available pulses. This property can be used to significantly reduce the complexity of the analysis algorithm in the presence of the pitch filter.

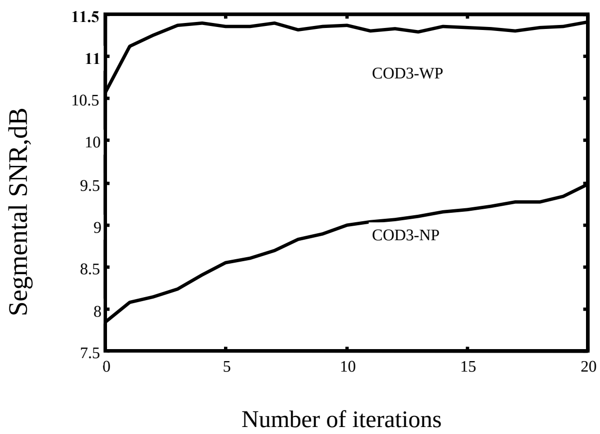

Figure 3: SEGSNR versus the number of analysis pulse in ALG3 with and without pitch filter.

Fig. 4 shows a short speech segment of a female speaker and the corresponding temporal variations of the SEGSNR obtained using COD1 and COD3 both with pitch filter included. The number of pulses used is 20 pulses per frame and the over all gain here is 1.8 dB. Similar to [7], the reoptimization in COD3 does not always outperform that in COD1; better results were obtained in 61% and 57% of the female and male speakers, respectively, compared to 67% and 61% when no pitch was introduced. This observation also shows another difference in performance of the reoptimization algorithm when the pitch filter is considered, and it gives a hint on the drop in gain observed in Fig. 2. Focusing on the successful and unsuccessful frames, it was found that the SEGSNR for female speaker is improved by 2.6 dB over the successful frames and dropped by 0.7 dB over the unsuccessful frames. The corresponding figures with no pitch used were 3.2 dB and 0.58 dB respectively. Similar results were obtained for the male speaker but with less gain for COD3-WP compared to that of COD3-NP.

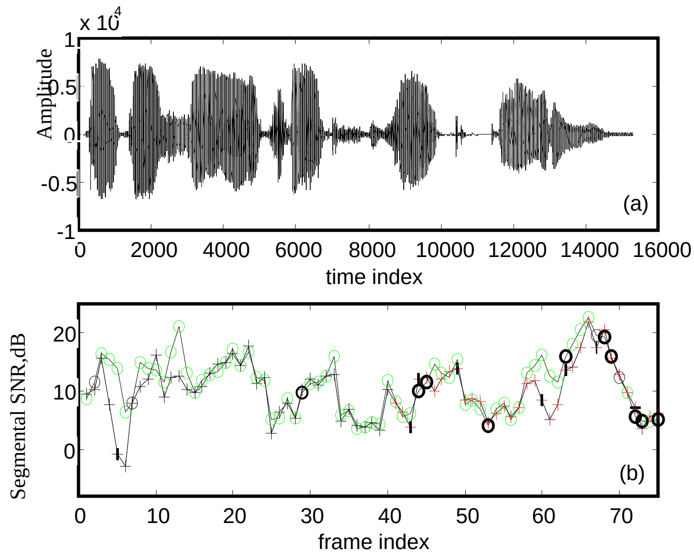

Figure 4: (a) Female speech segment, (b) SEGSNR performance, '+' COD1 and 'o' COD3.

To see the effect of reoptimization on the LP coefficients, Fig. 5 illustrates a typical LP envelope obtained using the conventional and the reoptimization schemes. It is clear that the reoptimization here smoothes the unnatural sharp peaks of the first and third formants obtained using the conventional analysis scheme.

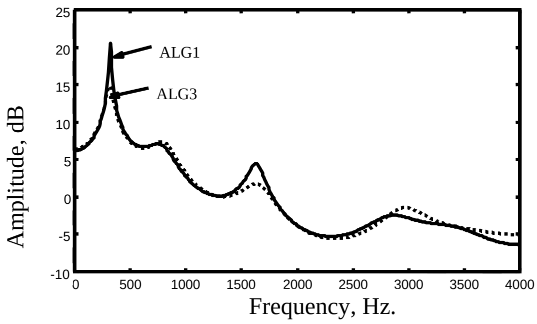

Figure 5: LP spectral envelopes using ALG1 and ALG3

V. Conclusion

The pitch filter has been introduced in the reoptimization of the LP synthesis filter in pulse excited A-b-S coders. This new approach has shown the same performance as that of the no pitch case. However, less gain has been obtained in reoptimizing the LP coefficients with pitch. Moreover, in pulse-by-pulse reoptimization method with no pitch filter, the SNR is increased gradually as a function of the number of used pulses, whereas it has almost reached saturation after the third pulse when pitch filter is used.

VI. References

1] R. Ramachandran and P. Kabal, "Pitch Prediction Filters in Speech Coding", *IEEE Trans. On ASSP*, **37(4)**, (1989), 467-479.

2] P. Kabal and R. P. Ramachandran, "Joint Optimization of Linear Predictors in Speech Coding", *IEEE Trans. on ASSP*, **37(5)**, (1989), 642-650.

3] S. Singhal and B. Atal, "Optimizing LPC Filter Parameters for Multi-Pulse Excitation", in *proc. ICASSP*, (1983), 781-784.

4] M. Fratti, G. Mian and G. Riccardi, "An Approach to Parameter Reoptimization in Multipulse-Based Coders", *IEEE Trans. On Speech and Audio Proc.*, **1(4)**, (1993), 463-465.

5] A. Hasib, K, Hacioglu, "Source Combined Linear Predictive Analysis in Pulse-Based Speech Coders", *IEE Proc., Vision, Image and Signal Processing*, **143(3)**, June (1996) 143-148.

6] M. Serizawa, A. Gersho, “Joint Optimization of LPC and Closed-Loop Pitch Parameters in CELP Coders”, *IEEE, Signal Processing Letters*, **6(3)**, (1999), 52-54.

7] K. Hacioglu, A. Hasib, "Pulse-by-Pulse Reoptimization of the Synthesis Filter in Pulse Based Coders", *IEEE Trans. On Speech and Audio Processing*, **6(2)**, March (1998), 180-185.