Correspondence

An Approach to Parameter Reoptimization in Multipulse-Based Coders

M. Fratti, G. A. Mian, and G. Riccardi

***Abstract*— An algorithm for LPC parameter optimization in MP-LPC–based speech coders is presented. It is shown that by taking into account the nature of the MP-excitation signal into LPC parameter computation, it is possible to improve the effectiveness of the LPC model. This results in a better quality of the reconstructed signal in terms of both objective and subjective criteria. The implementation details of the algorithm are discussed and experimental results are presented. In particular, a comparison with standard MP-LPC techniques is given.**

I. Introduction

Multipulse LPC-based techniques are effective and efficient methods for high-quality speech coding at medium–low bit rates. Using a recently developed method for optimal multipulse (MP) amplitude computation [1], it is possible to obtain high performance from both the objective and subjective points of view. The power of this method allows even an application to hi-fi music signal coding at low bit-rates (i.e., 128 kbit/s), as it has been recently shown in [2].

There is, however, an intrinsic limit to the corresponding analysis-by-synthesis coding scheme. Present LPC analysis procedures assume that the input to the all-pole filter is white. In order to greatly improve the signal reconstruction process, it is important to optimize LPC parameters taking into account the particular nature of the excitation signal.

In this paper we present the results obtainable with a method that deals with the problem of optimizing both LPC parameter identification and excitation modeling. It results in a performance improvement of MP-LPC coders, revealing its effectiveness in predictor coefficients and pulse sequence coding. In addition, an efficient solution to LPC parameter reoptimization is given.

II. Parameter Reoptimization

With reference to Fig. 1, the ultimate aim of a MP-LPC scheme is to find a pulse sequence $u(n)$ and a set of filter parameters a_k that minimize a perceptually weighted MSE with respect to the reference signal $s(n)$.

This gives rise to the synthesized signal

$$\hat{s}(n) = \sum_{k=1}^{p} a_k \hat{s}(n-k) + u(n) \tag{1}$$

with p the predictor order. The predictor coefficients a_k and the excitation signal $u(n)$ are to be determined in such a way that

$$\sum_n (s(n) - \hat{s}(n))^2 \tag{2}$$

Manuscript received March 24, 1992; revised November 30, 1992. The associate editor coordinating the review of this paper and approving it for publication was Dr. C.-H. Lee.

M. Fratti is with Alcatel-Telettra, V. Trento 30 Vimercate, 20059 Milano, Italy.

G. A. Mian and G. Riccardi are with the Dipartimento di Elettronica Informatica, V. Gradenigo 6, 35131 Padova, Italy.

IEEE Log Number 9210880.

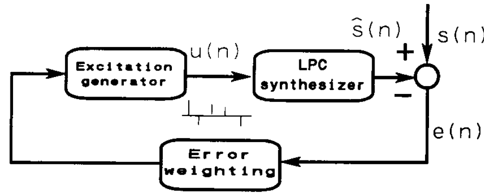


Fig. 1. MP-LPC synthesis scheme.

be minimum. The corresponding problem is highly nonlinear and almost irresolvable for real-time applications In the commonly used MP-LPC schemes, a set of LPC parameters a_k is first determined (assuming a white input), and then a pulse sequence $u(n)$ is derived. Singhal and Atal [3] proposed a sequential strategy to find a (possibly) less approximate solution to the minimization of (2). (Alternative solutions are suggested in [5], [6]). In particular they proposed the reoptimization of a_k on the basis of the MP sequence, $u(n)$, found at the first synthesis step, and the successive determination of a new input $u'(n)$ (eventually the procedure could be reiterated). In [3] an interpretation in terms of spectral envelope of the reiterated procedure was given.

This work was aimed to test the ultimate effectiveness of such a procedure, and to find an effective way for LPC parameter reoptimization. Starting from the expression of the decoded signal given by (1), an approximation is made, namely, $\hat{s}(n-k) \approx s(n-k)$. This leads to the following equation for the "reconstructed" signal:

$$\hat{s}(n) = \sum_{k=1}^{p} a_k s(n-k) + a_{p+1} u(n) + \hat{s}_0(n) \tag{3}$$

The term $\hat{s}_0(n)$ represents the free-response corresponding to the predictor computed at the first step. The insertion of $\hat{s}_0(n)$ makes analysis equation (3) similar to synthesis equation (1) and it was found to slightly improve the performance.

By inserting (3) into (2) and minimizing with respect to $a_k, k = 1 \cdots p+1$, the following system is obtained:

$$\left[\begin{array}{c|c} \Phi & \xi \\ \hline \xi^t & \Psi_{p+1,p+1} \end{array}\right] \begin{bmatrix} \mathbf{a}_{p,new} \\ a_{p+1,new} \end{bmatrix} = \begin{bmatrix} \zeta \\ \zeta_{p+1} \end{bmatrix} \tag{4}$$

where $\mathbf{a}_{p,new}$ is the vector of predictor coefficients, Φ is the covariance matrix, ξ is the vector containing the cross-correlations between the pulse sequence $u(n)$ and the reference signal $s(n)$, $\Psi_{p+1,p+1}$ is the energy of the pulse sequence, ζ is the vector of the cross-correlation between $s(n)$ and its delayed versions, while ζ_{p+1} is the cross-correlation between $s(n)$ and $u(n)$ (the symbols used are the same as in [3]). Therefore, after the initial LPC identification step and the MP excitation search, the LPC parameters can be determined by solving (4). A new set of excitation pulses is then recomputed. In the following we shall call $\{LPC_1, MP_1\}$ the LPC parameters and the excitation sequence determined in the first analysis and synthesis step and $\{LPC_2, MP_2\}$ the corresponding reoptimized quantities.

It is worth noticing that the determination of the LPC_2 parameters, through the solution of (4), can be efficiently carried out taking into account the fact that the inversion of the system matrix can be done using the inversion lemma of a partitioned matrix [4]. In particular

1063–6676/93$03.00 © 1993 IEEE

the new solution $\boldsymbol{a}_{p,new}$ for LPC_2 consists of a linear combination of the old solution $\boldsymbol{a}_{p,old}$ (corresponding to $\xi = \mathbf{0}$) and an additional term that takes into account the excitation signal MP_1. We shall now focus our attention on the expression of the additional term, which is a vector containing the variations of the predictor coefficients. The solution of the partitioned system is

$$\boldsymbol{a}_{p,new} = \boldsymbol{a}_{p,old} + \frac{\xi^t \boldsymbol{a}_{p,old} - \zeta_{p+1}}{\xi^t \delta - \psi_{p+1,p+1}} \delta = \boldsymbol{a}_{p,old} + K\delta \qquad (5)$$

where the p-dimensional vector δ containing the coefficient variations is given by the solution of linear system

$$\Phi\delta = \xi. \qquad (6)$$

This is efficiently accomplished via the Cholesky factorization of the covariance matrix Φ already used in the first step. As a result the vector $\boldsymbol{a}_{p,new}$ is the sum of $\boldsymbol{a}_{p,old}$ and of the correction vector $\delta' = K\delta$:

$$\boldsymbol{a}_{p,new} = \boldsymbol{a}_{p,old} + \delta' \qquad (7)$$

Using the relationship $\boldsymbol{a}_{p,old} = \Phi^{-1}\zeta$ and (6), (7) can be put into the form

$$\boldsymbol{a}_{p,new} = \Phi^{-1}(\zeta + K\xi) = \Phi^{-1}\zeta' \qquad (8)$$

where the vector ζ' has components

$$\zeta'_i = \sum_n (s(n) + Ku(n))s(n-i) \quad i = 1 \cdots p \qquad (9)$$

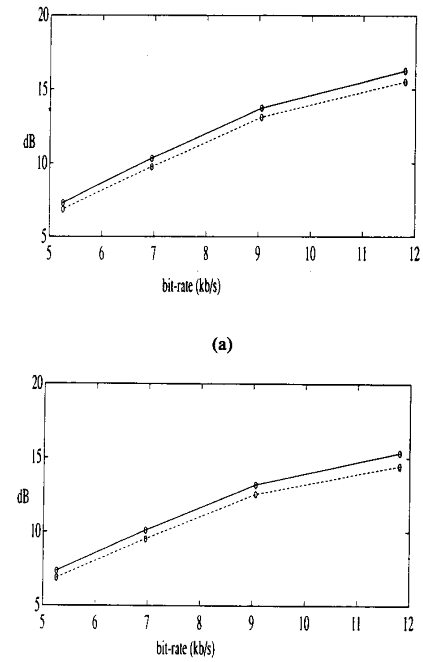
Equation (9) shows that the $\boldsymbol{a}_{p,new}$ vector is obtained by taking into account, as reference input, the sum of the signal $s(n)$ and a weighted pulse sequence $u(n)$.

It is important to point out that the effectiveness of reoptimization is expected to be more evident when the vector ξ is distant from the null vector. Really, the more the input pulse sequence tends to be white, the more ξ and, consequently, δ approaches the null vector: this way the new solution resembles the old one. Since the excitation tends to be white for unvoiced speech and colored (almost periodic) for voiced speech, we should expect the reoptimization procedure to be more effective in the case of voiced speech.

Once the LPC_2 parameters are determined, a new excitation sequence MP_2 has to be computed. (In fact it has been verified that in order to improve performance, i.e., S.N.R., it is not sufficient to compute only LPC_2, as the couple $\{LPC_2, MP_1\}$ is usually outperformed by $\{LPC_1, MP_1\}$). Then, a choice is made between the two sets of MP-LPC parameters according to the best perceptual matching to the reference signal. Notice that there is no mathematical proof that $\{LPC_2, MP_2\}$ corresponds to an increase in SNR. However, this was found to be true in 85% of the cases. It is worth noting that the final choice allows one to cope also with the eventual and rare (i.e., less than 2% of the cases) stability problems that may occur in the reestimation of LPC_2.

Summarizing, the main steps of the procedure are as follows:

- generation of a reference signal, $s(n) - \hat{s}_0(n)$;
- determination of LPC_1 via the stabilized convariance method [7] and of MP_1 through the usual analysis-by-synthesis scheme;
- efficient recomputation of the LPC_2 parameters through the solution of (4) and, then, of MP_2;
- choice between $\{LPC_1, MP_1\}$ and $\{LPC_2, MP_2\}$ according to perceptual SNR and to eventual stability problems occurring in LPC_2.



(a)

(b)

Fig. 2. SNRSEG vs. bit-rate for CO-MPLPC (continuous line) and MPLPC (dotted line), (a) Female speaker, (b) Male speaker.

III. Experimental Results

The performance of the reoptimized MP-LPC system (referred to as Cholesky optimized multipulse LPC, or CO-MPLPC) corresponding to $\{LPC_2, MP_2\}$ was tested and compared with that of the MP-LPC codec described in [1]. Six different short English sentences (i.e. about 2 s each) spoken by 4 male and 4 female English speakers were used. The speech database was sampled at 8 kHz. A 20-ms frame length, a 10-order LPC predictor, and various pulse-per-frame densities were used. Objective performance was evaluated using the segmental SNR (SNRSEG), while subjective one was assessed by informal listening tests.

In order to test the interaction between coding and reoptimization, the dependance of the segmental SNR on the pulse density was evaluated in the following three cases:

1) no coding;
2) coding of only MP_1 and MP_2 sequences;
3) coding of the reflection coefficients and of both MP_1 and MP_2 sequences.

In case 1) CO-MPLC gives some performance improvements, in terms of segmental SNR, with respect to MP-LPC. In addition, the lower the pulse density is, the bigger the gap between MP-LPC and CO-MPLPC techniques is: the SNRSEG difference ranges from 0.8 dB at 8 pulses/frame to 0.2 dB at 30 pulses/frame.

In case 2) the pulse sequences were coded according to the following scheme: exact quantization of the pulse locations, a 7-bit logarithmic quantizer for the maximum amplitude, and a 3-bit optimized vector quantizer for the normalized amplitudes. The vector quantizer was built on a training set of 30 minutes of different speakers with the LBG algorithm. For each pulse density, a proper vector quantizer was designed with the same number of bits (i.e., 3 bits). In the CO-MPLPC the second step used as input the quantized version of the excitation sequence, MP_1. The comparison with the

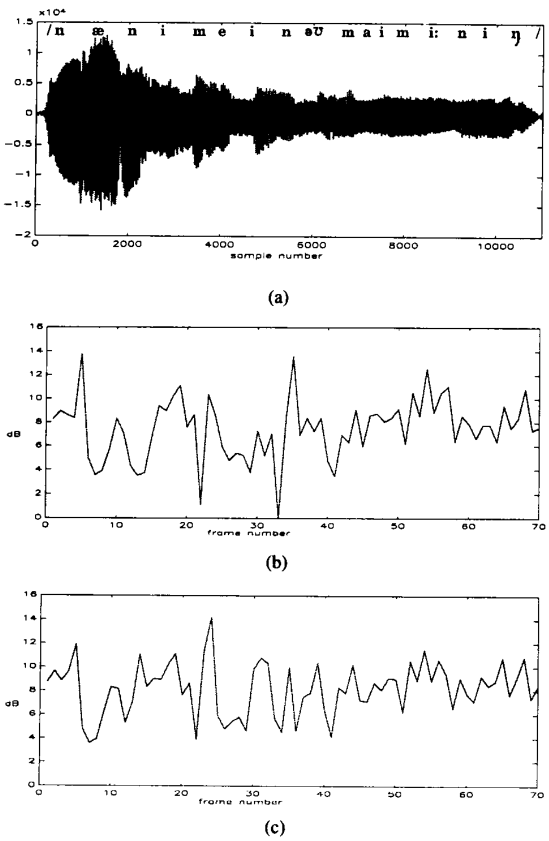


Fig. 3. Example of SNR behavior for the sentence "Nanny may know my meaning." (a) Waveform. (b) MP-LPC. (c) CO-MPLPC.

results corresponding to case 1) reveals a SNRSEG decrease for both male and female speakers ranging from 0.9 dB at 8 pulses/frame to 0.2 dB at 30 pulses/frame. In particular, when the pulse sequence is coded, female voices slightly outperform the male ones. This fact can be attributed to the greater sensitivity of low-pitched signals to multipulse sequence quantization errors. Moreover, there is some increase in the difference between the SNRSEG corresponding to the two methods. This fact is to be attributed to the use of the quantized version of MP_1 in the reoptimization procedure.

Case 3) corresponds to the overall codec implementation. A 34-bit statistically optimized scalar quantizer was used for reflection coefficients quantization. It was characterized by 1.5 dB average cepstral distance and 15% outliers beyond 4 dB. Fig. 2 gives (separately for male and female speakers) the performance of the resulting coder in terms of SNRSEG for bit-rates 5.25, 6.95, 9.05, 11.8 kb/s, corresponding to 8, 13, 20, 30 pulses/frame, respectively. As expected, the reflection coefficients quantization gives rise to a SNRSEG decrease (0.5 dB on the average). At the same time, however, the difference between CO-MPLPC and MP-LPC performance remains as in case 2) or slightly increases.

It is worth noticing that notwithstanding the SNRSEG difference of about 10% at 8 or 13 pulses/frame, the real effectiveness of CO-MPLPC relies on its capability to reduce SNRSEG variance. In fact, the corresponding SNRSEG behavior is "smoother" than the one corresponding to MPLPC and without very low minima. An example is given in Fig. 3, which compares the SNRSEG of CO-MPLPC and MP-LP at 13 pulses/frame, in the sentence "Nanny may know my meaning." The absence of sharp changes and very low minima in CO-MPLPC is evident. The subjective correlate of this fact is a synthesized speech that is perceived "smoother" and without occasional "clicks." The example of Fig. 3 might seem an *ad hoc* example. As observed by a reviewer it is strongly nasalized, and correspondingly, the all-pole predictor does not work very well. As such, it is conceivable that once the zeros (or excitations) have been determined, at the second step of the algorithm a better pole fit can be found. However, it has to be stressed that in all frames of the speech databases used, the SNR corresponding to COMPLC was found greater than the one of MPLPC.

IV. Conclusions

The effectiveness of the reoptimization (along the lines suggested in [3]) of LPC model in a multipulse coder has been evaluated and an efficient solution to predictor parameter reoptimization has been presented.

The main findings of the experiments carried out are that such a technique gives some improvements in terms of average SNRSEG and mainly for speech sounds (e.g., nasals) not adequately represented by an all-pole model.

References

[1] S. Singhal and B. Atal, "Amplitude optimization and pitch prediction in multipulse coders," *IEEE Trans. Acoustic, Speech and Signal Processing,* vol. 37, pp. 1001–1008, Mar. 1989.
[2] S. Singhal, "High quality audio coding using multipulse LPC," in *Proc. ICASSP 1990,* pp. 1101–1104.
[3] S. Singhal and B. Atal, "Optimizing LPC filter parameter for multipulse excitation," in *Proc. ICASSP 1983,* pp. 781–784.
[4] F. Gantmacher, *The theory of matrices.* New York: Chelsea, 1960.
[5] F. F. Tzeng, "An analysis-by-synthesis linear predictive model for narrowband speech coding," in *Proc. ICASSP 1990.* pp. 209–212.
[6] M. Niranjan, "CELP coding with adaptive output-error model identification," in *Proc. ICASSP 1990,* pp. 225–228.
[7] B. Atal, "Predictive coding of speech at low bit rates," *IEEE Trans. Communication,* vol. COM-30, pp. 600–614, Apr. 1982.

Reducing Redundant Computation in HMM Evaluation

J. R. Deller, Jr., and R. K. Snider

***Abstract*— Redundant computations occur when a set of HMM's is evaluated with respect to an observation string. A formal restructuring of the HMM allows the redundancy to be identified and removed. An $\mathcal{O}([1-\kappa]NT)$ complexity HMM results with the "compression index" $\kappa \in [0,1)$ depending upon several factors. Isolated-word recognition experiments illustrate.**

Manuscript received July 27, 1992; revised January 20, 1993. The associate editor coordinating the review of this paper and approving it for publication was Dr. X. Huang.

The authors are with the Department of Electrical Engineering, Control, Systems, and Signal Processing Group, Speech Processing Laboratory, Michigan State University, East Lansing, MI 48824-1226.

This work was sponsored in part by the Whitaker Foundation and in part by the National Science Foundation under Grant No. MIP-9016734. J. Deller was also partially supported by an Ameritech Fellowship.
IEEE Log Number 9210876.

1063-6676/93$03.00 © 1993 IEEE