A NEW APPROACH TO MULTIPULSE LPC CODER DESIGN

A. Dembo and D. Malah

Department of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel

ABSTRACT

Two new ideas for the design of multipulse excited LPC coders are presented in this paper. The first idea relates to improving the extraction of the all-pole filter parameters by taking into account the error weighting function. This function, which takes advantage of the noise masking properties of the ear, is ignored in the conventional covariance and autocorrelation methods. The new approach leads to an iterative algorithm, in which the first iteration is essentially the covariance method. Each new iteration involves estimation of the residual and increases a likelihood function, taking into account the error weighting function. The second idea relates to improving the derivation of the excitation parameters. A recursive algorithm for the optimal choice of the i-th pulse location, given the previously obtained (i-1) pulse locations, is presented. The low complexity of the new algorithm enables its combination with a tree search algorithm, thus giving a solution which is better than the solution of the algorithms reported earlier , when the error weighting function is used.

I. INTRODUCTION

The multipulse model of LPC Excitation [1,3,5] provides a method for producing natural sounding speech.

The model is based on two sets of parameters:
(A) The coefficients of an all-pole filter used to approximate the vocal tract transfer function.
(B) The excitation signal parameters, composed of pulse locations and amplitudes.

Simultaneous optimization of all the parameters is highly difficult. Thus, all previous works obtained the all-pole filter parameters using standard LPC methods (the auto-correlation method, or the stabilized covariance method [1]). These methods assume either a single pulse or white noise as an excitation signal. After the LPC parameters are set, optimization of an error criterion with respect to the excitation parameters is performed. In [5], an iterative scheme is presented in which the LPC parameters are re-optimized under a previously obtained excitation signal. The basic multipulse LPC model is as follows:

$$s(n) - \sum_{i=1}^{p} a_i s(n-i) = e(n) - \sum_{i=1}^{p} a_i \gamma^i e(n-i) + u(n) \quad (1)$$

where $s(n)$ is the speech signal, $e(n)$ is the input noise sequence, and $u(n)$ is the multipulse excitation sequence. In each frame of the analysis (of duration M), $u(n)$ contains only $K \ll M$ non-zero elements at locations $\{n_i\}_{i=1}^{K}$, with amplitudes $\{b_i\}_{i=1}^{K}$. The parameters of the multipulse LPC coder, which generally vary from frame to frame, are the values of $\{a_i\}_{i=1}^{P}$, $\{n_i\}_{i=1}^{K}$ and $\{b_i\}_{i=1}^{K}$. The value of γ is a priori set and in the range $0 \leq \gamma < 1$. Conventional LPC coders use $\gamma = 0$. However, due to the masking property of the ear it was found that an optimal value of γ is about 0.7 [1].

The synthesis of the speech signal is performed by exciting the all-pole model by the multipulse sequence $u(n)$. For each frame, the P initial values needed for the synthesis by an all-pole filter, are obtained from the synthesized speech of the previous frame. The synthesis error can be measured by the energy of the weighted residual sequence $e(n)$, in the given frame, i.e.:

$$E_R \triangleq \sum_{n=0}^{M-1} e(n)^2 \quad (2)$$

For every set of parameters in the current analysis frame, (1) and (2) are used to calculate the error measure E_R. This is done, using the P last samples of speech and weighted residual, of the previous frame, as initial values in (1). The optimal set of parameters is defined to be the set which yields the minimum value of E_R.

As stated earlier, direct minimization of the error measure with respect to all the parameters is not practicable . Therefore, two minimization problems are solved. The first is the minimization of E_R with resepct to the all-pole filter coefficients, and the second one is its minimization with respect to the excitation signal parameters. In both minimization steps, the parameters that are not optimized are assumed to be known. The minimization steps are therefore:

$$\min_{\{a_i\}_{i=1}^{P}} E_R\left(\{a_i\}_{i=1}^{P} \mid \{b_i^{(r)}, n_i^{(r)}\}_{i=1}^{K}\right) \quad (3A)$$

$$\min_{\{b_i, n_i\}_{i=1}^{K}} E_R\left(\{b_i, n_i\}_{i=1}^{K} \mid \{a_i^{(r)}\}_{i=1}^{P}\right) \quad (3B)$$

where the superscript r denotes the iteration number .

CH2118-8/85/0000-0949 $1.00 © 1985 IEEE

II. EXTRACTING THE LPC PARAMETERS

As mentioned earlier we extract the LPC parameters assuming that the excitation signal parameters are known (either from previous iterations, or are set to zero for the first iteration). For $\gamma = 0$ (i.e. not taking advantage of the masking property of the ear), the solution of (3A) is by means of the conventional covariance method for LPC parameters, as in [5]. Therefore, the vector $\underline{a}$ (with elements a_i, $i = 1,2,\ldots,P$, representing the LPC coefficients) is obtained by solving the following set of linear equations:

$$R\,\underline{a} = \underline{\rho} \tag{4}$$

where the (i,j) element of the correlation matrix R is:

$$R(i,j) = \sum_{n=0}^{M-1} s(n-i)s(n-j) \tag{5}$$

and the j-th element of the correlation vector $\underline{\rho}$ is:

$$\rho(j) = \sum_{n=0}^{M-1} (s(n)-u(n))\cdot s(n-j)\ . \tag{6}$$

However, this simple solution is not optimal for $\gamma \neq 0$. Direct minimization of the error measure for $\gamma \neq 0$ leads to a non-linear set of equations which are very difficult to solve. We will use therefore, an alternate method to define and extract the optimal set of LPC parameters. For this purpose, we now modify the basic multipulse LPC model, by assuming that there are two independent input noise sequences. The first is e(n) which appears in the basic model represented by (1), and the second input noise sequence, denoted by v(n) is a white noise sequence, having zero mean and variance $\sigma_v^2 \triangleq \varepsilon^2\cdot\sigma_e^2$.

The modified model (which replaces (1)) is therefore:

$$s(n) - \sum_{i=1}^{P} a_i s(n-i) = e(n) - \sum_{i=1}^{P} a_i\gamma^i e(n-i) + u(n) + v(n) \tag{7}$$

However, again direct minimization of the error measure leads to a non-linear set of equations which are difficult to solve. We, therefore, modify the error measure, and instead of minimizing the energy of e(n) over all possible LPC parameters, we seek the set of parameters for which the probability of generating the given speech samples is maximal. This is actually a maximum likelihood (ML) formulation of the parameter extimation problem. However, to solve the ML problem we can use an iterative algorithm [2], in which each iteration involves a solution of a set of linear equations, similar to the covariance method equations (4)-(6). Thus, each iteration is quite simple, and in this iterative algorithm the likelihood function can be shown [2] to be non-decreasing from iteration to iteration.

We choose a Gaussian model for both input noise sequences (thus, assuming also a Gaussian model for the speech signal), since under this model the likelihood function of the speech is a function of the weighted residual energy E_R, and therefore the ML set of parameters is closely related to the solution of the original problem. We use vector notation, to simplify the equations. The vectors $\underline{e}$, $\underline{v}$ and $\underline{s}$ are of length M, and refer to the weighted residual, white noise and speech samples in the current frame, respectively. Now, the likelihood function is given by:

$$\text{Prob}(\underline{s}|\underline{a},\sigma_e) = \int_{\underline{e}\in\mathbb{R}^M} \left(\frac{1}{2\pi\sigma_e\sigma_v}\right)^M \cdot$$

$$\cdot \exp - \frac{1}{2}\left[\frac{1}{\sigma_e^2}(\underline{e}^T\cdot\underline{e}) + \frac{1}{\sigma_v^2}(\underline{v}^T\cdot\underline{v})\right] d\underline{e} \left[\triangleq L(\underline{a},\sigma_e)\right] \tag{8}$$

This is, of course, a function of the parameter vector $\underline{a}$, since the vector $\underline{v}$ in (8) is given by (7), and thus depends on the LPC parameters. The ML problem is therefore the following maximization problem:

$$\text{Max} \quad \{L(\underline{a},\sigma_e)\} \tag{9}$$

The gradient of the function $L(\cdot)$ with respect to the unknowns, is a complicated non-linear function. Therefore, as stated earlier, a direct solution of (9) is not practicable. In the iterative algorithm of [2], the r-th iteration involves a solution of the following maximization problem:

$$\underset{\underline{a}\in\mathbb{R}^P,\sigma_e}{\text{Max}} \quad \{Q(\underline{a},\sigma_e|\underline{a}^{(r)},\sigma_e^{(r)})\} \tag{10}$$

where the function $Q(\cdot)$ is defined as:

$$Q(\underline{a},\sigma_e|\underline{a}^{(r)},\sigma_e^{(r)}) \triangleq \int_{\underline{e}\in\mathbb{R}^M} \ell n[\text{Prob}(\underline{e},\underline{v}|\underline{a},\sigma_e)] \cdot$$

$$\cdot\ \text{Prob}(\underline{e},\underline{v}|\underline{a}^{(r)},\sigma_e^{(r)})\,d\underline{e} \tag{11}$$

Since the logarithm of the joint probability density of both input noises is a p.s.d. quadratic form in the unknowns $\underline{a}$ and σ_e, the global maximum of $Q(\cdot)$ in each iteration can be found analytically as well as the new set of parameters, for the (r+1)-th iteration. The new set of LPC parameters $\underline{a}^{(r+1)}$ is the solution of the following set of P linear equations:

$$R^{(r)}\cdot\underline{a}^{(r+1)} = \underline{\rho}^{(r)} \tag{12}$$

where the (i,j) element of the correlation matrix $R^{(r)}$ in the r-th iteration is:

$$R^{(r)}(i,j) \triangleq \sum_{n=0}^{M-1} (s(n-i)-\gamma^i e^{(r)}(n-i)) \cdot$$

$$\cdot (s(n-j)-\gamma^j e^{(r)}(n-j)) + \gamma^{(i+j)} \sum_{n=0}^{M-1} B^{(r)}(n-i,n-j) \tag{13}$$

and the j-th element of the correlation vector $\underline{\rho}^{(r)}$ in the r-th iteration is:

$$\rho^{(r)}(j) \triangleq \sum_{n=0}^{M-1} (s(n) - u(n) - e^{(r)}(n)) \cdot$$

$$\cdot (s(n-j) - \gamma^j e^{(r)}(n-j)) + \gamma^j \sum_{n=0}^{M-1} B^{(r)}(n-j,n) \tag{14}$$

The values of $e^{(r)}(n)$ and $B^{(r)}(n,m)$ depend on the set of parameters that were obtained in the r-th iteration, and their exact value is given in the sequel. The optimal gain in the (r+1)-th iteration is given by:

$$\sigma_e^{(r+1)^2} \triangleq \frac{1}{2M}\left\{\sum_{n=0}^{M-1} e^{(r)}(n)^2 + \sum_{n=0}^{M-1} B^{(r)}(n,n) + \right.$$
$$\left. + \frac{1}{\varepsilon^2}\left[\sum_{n=0}^{M-1} v^{(r+1)}(n)^2 + \sum_{n=0}^{M-1}\left(B^{(r)}(n,n) - \sum_{i=1}^{P} a_i^{(r+1)}\cdot\gamma^i\cdot B^{(r)}(n-i,n)\right)\right]\right\} \quad (15)$$

$$v^{(r+1)}(n) \triangleq \left(s(n) - \sum_{i=1}^{P} a_i^{(r+1)} s(n-i)\right) - \left(e^{(r)}(n) - \sum_{i=1}^{P} a_i^{(r+1)}\cdot\gamma^i\cdot e^{(r)}(n-i)\right) - u(n) \quad (16)$$

In (13)-(16) the values of $e^{(r)}(n)$ for $-P \leq n \leq -1$ are independent of the iteration number, as they are known from the previous frame (the P last samples of the resulting weighted residual of the previous frame). The values of $B^{(r)}(n,m)$ for negative indices are zero, and their appearance in (13)-(15) is in order to simplify these equations. Therefore, the dependence of the (r+1)-th iteration in the result of the (r)-th iteration, is via the M×M matrix $B^{(r)}$ (whose (n,m) element is $B^{(r)}(n,m)$), and the vector $\underline{e}^{(r)} \in \mathbb{R}^M$ (whose n-th element is $e^{(r)}(n)$). Now, $\underline{e}^{(r)}$ is given by:

$$\underline{e}^{(r)} \triangleq B^{(r)}\cdot\underline{\omega}^{(r)} \quad (17)$$

where

$$\omega^{(r)}(n) \triangleq \frac{1}{\sigma_e^{(r)^2}\cdot\varepsilon^2}\left[z^{(r)}(n) - \sum_{i=1}^{P(n)} a_i^{(r)}\cdot\gamma^i\cdot z^{(r)}(n+i)\right] \quad (18)$$

$$z^{(r)}(n) \triangleq s(n) - u(n) - \sum_{i=1}^{P} a_i^{(r)} s(n-i) - \sum_{i=(n+1)}^{P} a_i^{(r)}\cdot\gamma^i\cdot e(n-i) \quad (19)$$

$B^{(r)}$ is the inverse of the matrix $A^{(r)}$ of dimension M×M, whose elements are:

$$A^{(r)}(n,n+d) = A^{(r)}(n+d,n) = \frac{1}{\sigma_e^{(r)^2}\cdot\varepsilon^2}\left[\sum_{i=d}^{P(n)} \gamma^{(2i-d)}\cdot a_i^{(r)}\cdot a_{i-d}^{(r)} + \varepsilon^2\delta(d=0)\right] \quad (20)$$

In (18)-(20) $0 \leq n \leq (M-1)$, $d \geq 0$, $P(n) \triangleq \mathrm{Min}[M-1-n, P]$ and $a_o^{(r)} \triangleq -1$.

Equations (12)-(20) summarize the r-th iteration, and their derivation is by computing the gradient of Q(·) with respect to the unknown parameters.

The initial set of parameters $a^{(1)}$ is obtained by using $\underline{e}^{(0)} = \underline{0}$, $B^{(0)} = 0$ in (12)-(16). This set of parameters coincides with the set of parameters obtained by the conventional covariance method. Each iteration of the new method increases the likelihood function (10), thus the new method results in a set of LPC parameters that are more likely than the set of parameters obtained by conventional methods. It can be shown that for $\varepsilon \to 0$ the iterative algorithm become singular, and stops at any initial set of parameters. Thus, it is recommended not to choose very small values of ε.

III. EXTRACTING THE EXCITATION PARAMETERS

The optimal set of excitation parameters is the solution of the minimization problem (3B). Given the optimal pulse locations $\{n_i\}_{i=1}^K$, the optimal amplitudes are the solution of a set of K linear equations, as shown in [3]. However, there are $\binom{M}{K}$ possible ways of setting the pulse locations. For $\gamma \neq 0$ (i.e., taking advantage of the masking property of the ear), the error measure E_R depends on all K pulse locations so that the only known way of optimal pulse locations setting involves an exhaustive search over all the $\binom{M}{K}$ settings. We present here a simple algorithm (of complexity $o(m^2)$) for computing the value of the minimal error corresponding to the location of the (m+1)-th pulse, given the previously selected m pulse locations, without explicitly calculating the optimal pulse amplitudes. This algorithm, which has lower complexity than the reported algorithms [3], is combined with tree search methods in order to find a reasonable set of excitation parameters. Although the search algorithm is sub-optimal, it is guaranteed that its result has lower value of E_R than the result of the algorithm in [3], while preserving its complexity.

We first describe the tree search method. We arrange all possible pulse locations in a tree of depth K. The root of this tree represents the initial state, in which the excitation signal is zero, and it has M sons, each of them represents a different possible location of the first pulse. Each of these nodes (denoted as nodes of the first level), has (M-1) sons (one for every possible value of the second pulse location). In general, each node at the m-th level has (M-m) sons, and there are $\binom{M}{m}\cdot m!$ nodes at the m-th level. Each route from the root to one of the leaves of this tree represents a complete set of K pulse locations, and there are exactly K! different routes. An exhaustive search over all $\binom{M}{K}$ possible pulse locations, is equivalent to a search over all the (different) leaves of this tree and selection of the leave with the minimal error value. Using the new algorithm, we can calculate the minimal error value on the leaves of this tree, by calculating at first the minimal error value on the nodes in the first level, then the minimal error value on the nodes in the second level, etc. This process has complexity of $o\left(\sum_{m=0}^{K} \frac{M!}{(M-m)!}\cdot m^2\right)$, and thus is not feasible. The minimal error value at each node is an upper bound on the minimal error value at each of the offsprings of this node. Therefore, the following sub-optimal search algorithm seems to be a reasonable alternative of the exhaustive search:

(i) Find the minimal error value of all M nodes at the first level, select the R nodes which have the lowest error values.

(ii) Given R nodes at the m-th level (m=1,...,K-1), find the minimal error values at their sons, then select the R nodes having the minimal

error values out of these R·(M-m) nodes of the (m+1)-th level.

(iii) Reaching the K-th level (i.e., the leaves of the tree), select the leave with minimal error value out of the R leaves.

The complexity of this algorithm is $O(M \cdot R \cdot K^3)$, while the algorithm in [3], has complexity of $O(M \cdot R \cdot K^4)$. Because of this complexity, the value of R=1 was used in [3]. Here, for the same complexity, we can use R=K. Thus, the new algorithm results in a smaller error than obtained in [3] for the same complexity.

We describe now the new algorithm for calculating the minimal value of E_R over all possible amplitudes of the (m+1) pulses. This algorithm is used in the tree search in order to calculate the value of the minimal error.

Let $\{e_o(\ell)\}_{\ell=0}^{M-1}$ be the weighted residual samples due tothe previous frame (i.e., initial conditions), and let $\{h(\ell)\}_{\ell=0}^{N-1}$ be the impulse response of the filter $1/A(z/\gamma)$, both are assumed to be known from the solution of problem (3A). For each selection of m pulse locations $\{n_i\}_{i=1}^{m}$ and m pulse amplitudes $\{b_i\}_{i=1}^{m}$, the resulting weighted residual samples are:

$$e_m(\ell) = e_o(\ell) - \sum_{i=1}^{m} b_i \cdot h(\ell - n_i) \quad 0 \leq \ell \leq (M-1) \tag{21}$$

The error E_R is therefore given by:

$$E_R(\{b_i;n_i\}_{i=1}^{m}) = \left(\sum_{\ell=0}^{M-1} e_o(\ell)^2\right) - 2 \cdot \sum_{i=1}^{m} b_i \left(\sum_{\ell=n_i}^{M-1} e_o(\ell) h(\ell - n_i)\right) + \sum_{i=1}^{m}\sum_{j=1}^{m} b_i b_j \sum_{\ell=0}^{M-1} h(\ell-n_i)h(\ell-n_j) \tag{22}$$

This is a p.s.d. quadratic form with respect to the pulse amplitudes, and its minimal value for a given set of pulse locations is:

$$\hat{E}_R(\{n_i\}_{i=1}^{m}) \triangleq \min_{b_i} E_R(\{b_i;n_i\}_{i=1}^{m}) = \left(\sum_{\ell=0}^{M-1} e_o^2(\ell)\right) - \sum_{i=1}^{m}\sum_{j=1}^{m} v_o(n_i) \cdot v_o(n_j) \cdot q(i,j)$$

where (23)

$$v_o(\ell) \triangleq \sum_{k=\ell}^{M-1} e_o(k) h(k-\ell) \tag{24}$$

and $q(i,j)$ is the (i,j)-th element of the matrix Q which is the inverse of the matrix R whose (i,j) element is $r(n_i,n_j)$, defined below:

$$r(n,\ell) = \sum_{k=\max(n,\ell)}^{M-1} h(k-n) \cdot h(k-\ell) \tag{25}$$

The initial operations needed are the calculation of the M^2 values of $r(n,\ell)$ and M values of $v_o(\ell)$. This is done with a complexity of $O(M^2)$. The calculation of the sequence $h(\cdot)$ from the solution of (3A) is done with a complexity of $O(M \cdot P)$

We assume that the matrix Q and the value of $\hat{E}_R$ for the m pulse locations $\{n_i\}_{i=1}^{m}$ are already known, and present an efficient algorithm to update both Q and $\hat{E}_R$ for the (m+1) pulse locations $\{n_i\}_{i=1}^{m+1}$. The efficient algorithm uses a matrix inversion theorem from [4], in order to avoid the need to invert the matrix R for updating the matrix Q. We denote by $Q^{(m)}$ the m×m matrix Q of the m pulse locations, and by $Q^{(m+1)}$ the (m+1)×(m+1) matrix Q of the (m+1) pulse locations including n_{m+1} as the new pulse location. Let denote by $\underline{r}(n_{m+1})$ the m dimensional vector whose i-th element is $r(n_i,n_{m+1})$, and by $\underline{v}_o^{(m+1)}$ the (m+1) dimensional vector whose i-th element is $v_o(n_i)$. We also denote by $\hat{E}_R^{(m)}$ the minimal error value of the set of m pulse locations. With this notation the new algorithm is given by:

(i) Compute:

$$\eta \triangleq \underline{r}(n_{m+1})^T \cdot Q^{(m)} \cdot \underline{r}(n_{m+1}) \tag{26}$$

$$\underline{\varphi} \triangleq \begin{bmatrix} Q^{(m)} \cdot \underline{r}(n_{m+1}) \\ \hline -1 \end{bmatrix} \in \mathbb{R}^{m+1} \tag{27}$$

(ii) Now:

$$Q^{(m+1)} = \left[\begin{array}{c|c} Q^{(m)} & \underline{0} \\ \hline \underline{0}^T & 0 \end{array}\right] + \frac{1}{(r(n_{m+1},n_{m+1})-\eta)} \underline{\varphi} \cdot \underline{\varphi}^T \tag{28}$$

$$\hat{E}_R^{(m+1)} = \hat{E}_R^{(m)} - \frac{1}{(r(n_{m+1},n_{m+1})-\eta)} (\underline{v}_o^T \cdot \underline{\varphi})^2 \tag{29}$$

It can be readily verified that $Q^{(m+1)}$ and $\hat{E}_R^{(m+1)}$ obtained by (28)-(29), match their definition in (23)-(25), and that the overall complexity of (26)-(29) is $O(m^2)$ while direct calculation of $Q^{(m+1)}$ requires $O(m^3)$ operations.

REFERENCES

[1] B.S. Atal and J.R. Remde, "A New Model of LPC Excitation for Producing Natural Sounding Speech at Low Bit Rates", Proc. ICASSP, 1982, Paris, pp. 614-617.

[2] L.E. Baum et al., "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains", The Ann. of Math. Stat., Vol. 41, No.1, 1970, pp. 164-171.

[3] S. Singhal and B.S. Atal, "Improving Performance of Multipulse LPC Coders at Low Bit Rates", Proc. ICASSP 1984, pp. 1.3.1-1.3.4.

[4] M.S. Bartlett, "An Inverse Matrix Adjustment Arising in Discriminant Analysis", The Ann. of Math. Stat., Vol. 20, 1951, pp. 107-111.

[5] V.K. Jain and R. Hangartner, "Efficient Algorithm for Multi-Pulse LPC Analysis of Speech", Proc. ICASSP 1984, pp. 1.4.1-1.4.4.