A UNIFIED FRAMEWORK FOR LPC EXCITATION REPRESENTATION IN RESIDUAL SPEECH CODERS

E. Ofer, D. Malah[1], and A. Dembo[2]

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

ABSTRACT

In this paper the efficient representation of the excitation signal to an LPC synthesis filter by means of a vector expansion of the residual signal is examined. According to this approach the excitation signal is represented as a linear combination of a small number of vectors taken from a given vector set, known at both ends of the transmission channel. It is demonstrated that this approach provides a unified framework for describing and analyzing a wide range of residual speech coders, from Multipulse LPC and CELP to Residual Transform Coders and leads to generalization of some of these schemes. Optimality conditions based on the singular value decomposition (SVD) of the impulse response matrix of the perceptually weighted LPC synthesis filter are given. A resulting simplified Predictive Transform Coder is proposed and examined by computer simulations.

1. INTRODUCTION

Residual speech coders are typically designed to operate at transmission rates of 4.8 - 16 Kbps. These coders have the feature of coding both the LPC residual, which is used as the excitation to the synthesis filter, and the LPC parameters which make up the synthesis filter [1].

In recent years, some new residual coders were proposed. In the Multipulse scheme [2], the residual signal is represented by a small number of pulses. The location and amplitude of each pulse are coded and transmitted in addition to the LPC parameters. Following the multipulse scheme other schemes have appeared which attempt to represent the residual in a simpler manner. Some examples are: Maximum Residual Magnitude (MRM) [3], Regular Excitation [4,5], and Thinned-Out Residual [6].

Another recently proposed residual coder is the CELP (Code Excited Linear Prediction) coder [7] in which each segment of the residual signal is represented by an appropriate block of white Gaussian noise selected from a given dictionary (code-book). In another recent residual speech coder, the speech residual is first transformed by a DCT and then coded [8].

In this work, we present a unified mathematical framework for the above coders which leads to some generalizations of these schemes and to a better understanding of optimality conditions. This framework is based on a vector-expansion of the LPC residual signal and its representation by a linear combination of a small number of vectors taken from a given vector set. This can be seen as an extension of the unified description used in [9] for some of the above coders (although the work presented here is based on our earlier efforts in this direction [10]).

The paper is organized as follows: In section 2, a mathematical presentation of the vector-expansion approach is given. Section 3 demonstrates that several known residual speech coding schemes are included in the unified framework, and provides generalization to some of these schemes. Section 4 presents optimality consideration in choosing the vector set used in the proposed vector-expansion scheme, and Section 5 presents a simplified residual transform coding scheme which was examined in computer simulations.

1. Presently on sabbatical leave at the Department of Signal Processing Research, AT&T Bell Laboratories, Murray Hill, NJ 07974.

2. Presently with the Information Systems Laboratory (ISL) - Stanford University, Stanford, CA 94305.

2. VECTOR EXPANSION OF THE LPC RESIDUAL

In the proposed vector-expansion scheme, the excitation vector, whose elements are the input to the synthesis filter, is represented by a linear combination of a small number of vectors, taken from a given vector set. The vector set is known at both ends of the transmission channel, and the transmitted parameters are typically the indices of the selected vectors from the set and the coefficients needed to linearly recombine the vectors.

Let $\boldsymbol{V}$ be a set of M vectors of length N, each ($N \leq M$). Given the original speech vector $\mathbf{s}$ - constructed from consecutive N samples of the input speech (frame), and the LPC all-pole synthesis filter coefficients, we seek an excitation vector $\mathbf{u}$ which will be a linear combination of only a small number of vectors from $\boldsymbol{V}$, i.e.,

$$\mathbf{u} = \sum_{i=1}^{k} x_i \mathbf{v}_i , \quad \mathbf{v}_i \in \boldsymbol{V}, \quad k < N . \tag{1}$$

The vectors $\mathbf{v}_i$ are to be selected from the given set $\boldsymbol{V}$ such that the weighted squared error, between the original and the reconstructed speech signals in the given frame is minimized. The weighted squared error is defined here as

$$E_w \triangleq \sum_n [(s_n - \hat{s}_n) * w_n]^2 , \tag{2}$$

where s_n is the original speech signal, $\hat{s}_n$ is the reconstructed speech signal, w_n is the weighting filter impulse response, and the summation is assumed here to be over the N samples of the frame. The transfer function of the weighting filter is given by [2]:

$$W(z) = \frac{A(z)}{A(z/\gamma)} = \frac{1 + \sum_{k=1}^{p} a_k z^{-k}}{1 + \sum_{k=1}^{p} a_k \gamma^k z^{-k}} , \tag{3}$$

where the a_k's are the linear prediction coefficients, and the constant γ, $0 \leq \gamma \leq 1$, controls the shaping of error spectrum so as to match it to the frequency masking properties of the human ear [2]. Representing the reconstructed signal as:

$$\hat{s}_n = u_n * \tilde{h}_n + l_n , \tag{4}$$

where u_n is the excitation, $\tilde{h}_n$ is the synthesis filter impulse response (corresponding to $\tilde{H}(z) = 1/A(z)$), and l_n is a 'hang-over' signal generated by the filter memory from the previous frame (and needs not to be approximated, as it is also known at the receiver). Thus, from (2):

$$E_w = \sum_n [(e_n - u_n) * h_n]^2 , \tag{5}$$

where e_n is the residual signal resulting from passing $(s_n - l_n)$ through the inverse filter $A(z)$ with zeroed-out memory, and h_n is the impulse response of the weighted synthesis filter which is $\tilde{h}_n * w_n$.

Defining $\mathbf{R}$ to be the $N \times N$ matrix with elements r_{ij} given by:

$$r_{ij} = \sum_n h_{n-i} h_{n-j} \qquad 0 \leq i \leq N-1, \quad 0 \leq j \leq N-1 \tag{6}$$

results in

$$E_w = (\mathbf{e} - \mathbf{u})^T \mathbf{R} (\mathbf{e} - \mathbf{u}) , \tag{7}$$

CH2673-2/89/0000-0041 $1.00 © 1989 IEEE

where e is the residual vector and u is the excitation vector.

Now, let $\mathbf{Q}$ be an $N \times k$ matrix, having as its columns the selected k vectors from the set $\boldsymbol{V}$. Hence, from (1), $\mathbf{u}$ can be expressed as

$$\mathbf{u} = \mathbf{Q}\mathbf{x} \ , \tag{8}$$

where the elements of the vector $\mathbf{x}$ are the linear combination coefficients x_i, $i = 1, 2, \ldots, k$. (This is similar to the 'gain-shape' excitation representation in [9]). Thus, for a given matrix $\mathbf{Q}$, we obtain from (7) and (8) the following expression for the weighted squared error:

$$E_w^Q = (\mathbf{e} - \mathbf{Q}x)^T\mathbf{R}(\mathbf{e} - \mathbf{Q}\mathbf{x}) \ . \tag{9}$$

Solving for $\mathbf{x}$ which minimizes E_w^Q, results in:

$$\mathbf{x}_{opt} = (\mathbf{Q}^T\mathbf{R}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{R}\mathbf{e} \ . \tag{10}$$

Substituting (10) into (9) gives the following expression for the minimal error, given the matrix $\mathbf{Q}$,

$$E_{w\,\min}^Q = \mathbf{e}^T\mathbf{R}\mathbf{e} - (\mathbf{e}^T\mathbf{R}\mathbf{Q})(\mathbf{Q}^T\mathbf{R}\mathbf{Q})^{-1}\ (\mathbf{Q}^T\mathbf{R}\mathbf{e}) \ . \tag{11}$$

Thus, assuming that $\boldsymbol{V}$ is given, the basic procedure is to compute $\mathbf{e}$ and $\mathbf{R}$, for each given speech frame, and then to find $\mathbf{Q} \in \boldsymbol{V}$ such that ΔE_w^Q will be *maximal*, where:

$$\Delta E_w^Q = (\mathbf{e}^T\mathbf{R}\mathbf{Q})(\mathbf{Q}^T\mathbf{R}\mathbf{Q})^{-1}(\mathbf{Q}^T\mathbf{R}\mathbf{e}) \ . \tag{12}$$

3. RELATION TO PARTICULAR CODING SCHEMES

The residual expansion scheme presented above provides a unified framework for describing a range of residual speech coders, some of which were recently proposed, and some are generalizations which follow naturally from the presented framework. These coders, which we will briefly review some of them here, differ from each other by the chosen vector set $\boldsymbol{V}$ and in the way the expansion basis vectors (i.e. the columns of $\mathbf{Q}$) are selected from it.

It should be emphasized that we do not address here the problem of efficiently coding the coefficient-vector $\mathbf{x}$ but only treat the problems of choosing the vector set $\boldsymbol{V}$ and of finding $\mathbf{Q}$, so as to minimize (7). For some coders this is not the best approach, under given data rate constraints, since by using a bit allocation scheme to quantize a full-dimension vector $\mathbf{x}$ (i.e., having N elements and not just k), a lower error can be obtained. However, even then, issues concerning the choice of the vector set $\boldsymbol{V}$ are still of interest.

3.1 Multipulse LPC

Let $\boldsymbol{V}$ be chosen to consist of N unit vectors so that a matrix $\mathbf{V}$, having the N vectors in $\boldsymbol{V}$ as its columns, can be constructed to be the $N \times N$ identity matrix, i.e.,

$$\mathbf{V} = \mathbf{I}_N \ . \tag{13}$$

In this case the excitation signal is simply a linear combination of k unit vectors (i.e., it is represented by k pulses). Since for this choice of $\boldsymbol{V}$, eq. (12) does not reduce to a form which allows us to select all the k vectors in $\mathbf{Q}$ simultaneously, and an exhaustive search for the best k vectors in $\boldsymbol{V}$ is usually prohibitive, an iterative suboptimal solution is typically used. In the iterative scheme the positions of the pulses in the frame and their amplitudes are calculated iteratively - one pulse at a time. This scheme is known as Multipulse LPC and was first suggested by Atal & Remde [2]. Using (12) shows that in the j-th iteration of the algorithm we should let $\mathbf{Q} = \mathbf{v}_i$, and select $\mathbf{v}_i$ from $\boldsymbol{V}$ such that

$$\Delta E_w^{(j)} = (\mathbf{e}^{(j)T}\mathbf{R}\mathbf{v}_i)^2/r_{ii} \tag{14}$$

is maximized (over all vectors in $\boldsymbol{V}$), where i indicates the pulse position in the selected unit vector, r_{ii} is the i-th diagonal element of $\mathbf{R}$, and $\mathbf{e}^{(j)}$ is the residual vector at the j-th iteration, updated to take into account all the chosen vectors up to this point. Thus,

$$\mathbf{e}^{(j+1)} = \mathbf{e}^{(j)} - x_i\mathbf{v}_i \ ; \qquad \mathbf{e}^{(0)} = \mathbf{e} \ . \tag{15}$$

Implementing (14) is simple since multiplying the matrix $\mathbf{R}$ by $\mathbf{v}_i$ means choosing its i-th column, and x_i in (15) is simply given by the square root of the numerator of the r.h.s. of (14) (following maximization).

3.2 CELP

In the CELP scheme [7], a large vector set is typically used (called a dictionary). The vectors here are usually blocks of white Gaussian noise and only one vector is chosen per frame. Using (12) shows that the vector $\mathbf{v}$ to be chosen for a given frame should be selected from $\boldsymbol{V}$ (i.e. $\mathbf{Q} = \mathbf{v}$) such that

$$\Delta E_w^v = (\mathbf{e}^T\mathbf{R}\mathbf{v})^2/(\mathbf{v}^T\mathbf{R}\mathbf{v}) \tag{16}$$

will be maximized over all the possible vectors in $\boldsymbol{V}$. Thus, the process of selecting a vector in each iteration of the Multipulse scheme and in CELP is quite similar (except, of course, that a much larger amount of computations is needed in CELP to compute (16), since $\mathbf{v}$ is not a unit vector as in the Multipulse LPC scheme, and the set $\boldsymbol{V}$ usually contains many more vectors then the size of the frame) It should also be noted that since the vector set used in CELP consists of white Gaussian noise vectors, the residual signal e_n should also be whitened. This is approximately achieved by adding a pitch loop to the usual LPC scheme, i.e. removing long-term correlation in the input speech signal [7,9] (in which case h_n in (5) and (6) should be modified appropriately).

3.3 Generalized Maximum-Residual-Magnitude (GMRM)

Examination of (12) shows that if we use a vector set $\boldsymbol{V}$ having the property that for any $\mathbf{Q}$ constructed from it the matrix $(\mathbf{Q}^T\mathbf{R}\mathbf{Q})$ is diagonal, then we can select all the vectors in $\mathbf{Q}$ simultaneously (as will be exemplified in the sequel).

Assuming that $\mathbf{R}$ is calculated by a 'covariance' type method (i.e., in (6) the range of the summation is the frame length N), $\mathbf{R}$ can then be written as:

$$\mathbf{R} = \mathbf{H}^T\mathbf{H} \ , \tag{17}$$

where $\mathbf{H}$ is an $N \times N$ lower triangular Toeplitz matrix (known also as the impulse-response matrix), such that an (i,j) element in $\mathbf{H}$ is given by h_{i-j}, for $i \geq j$, and is zero for $i < j$. Now, let the matrix $\mathbf{V}$ (which represents the vector set $\boldsymbol{V}$) be given by

$$\mathbf{V} = \mathbf{H}^{-1} \ . \tag{18}$$

Then, since $\mathbf{Q}$ is a column-submatrix of $\mathbf{V}$, we obtain that $\mathbf{Q}^T\mathbf{R}\mathbf{Q} = \mathbf{I}_k$ (the $k \times k$ identity matrix), and hence, from (12):

$$\Delta E_w^Q = (\mathbf{e}^T\mathbf{R}\mathbf{Q})(\mathbf{Q}^T\mathbf{R}\mathbf{e}) = \|\ \mathbf{Q}^T\mathbf{R}\mathbf{e}\ \|^2 \ , \tag{19}$$

where $\|\cdot\|$ denotes the usual Euclidean vector norm. This expression will be maximized by choosing the vectors composing $\mathbf{Q}$ according to the indices in $\mathbf{V}$ which correspond to the k largest magnitude elements in the vector

$$\mathbf{y} \triangleq \mathbf{V}^T\mathbf{R}\mathbf{e} = \mathbf{H}\mathbf{e} \ . \tag{20}$$

Hence, $\mathbf{y}$ can simply be found by passing the elements of the residual vector $\mathbf{e}$ through the weighted synthesis filter $H(z) = \tilde{H}(z)W(z) = 1/A(z/\gamma)$ (with zeroed-out memory). From (10), the coefficient-vector $\mathbf{x}$ is given here by:

$$\mathbf{x}_{opt} = \mathbf{Q}^T\mathbf{R}\mathbf{e} \ . \tag{21}$$

Thus, the k elements of $\mathbf{x}_{opt}$ are simply given by the k largest magnitude elements in $\mathbf{y}$ of (20). Note that choosing a value of $\gamma \neq 0$ results in peak picking from the *colored* residual signal (by $1/A(z/\gamma)$), or the weighted input signal, whereas if $\gamma = 0$ is used, the peak picking is done from the residual signal itself.

We name this scheme 'Generalized Maximum Residual Magnitude' (GMRM) since it is a generalization of the simple residual coders proposed in the literature - MRM [5], and TOR (Thinned-Out Residual) [7]. In these coders $\gamma = 0$ is used, and indeed the peak picking is done there directly from the residual signal - e_n.

3.4 Residual Transform Coding

The vector set chosen in (18) can be made more general by letting:

$$\mathbf{V} = \mathbf{H}^{-1}\mathbf{T} \ , \tag{22}$$

where $\mathbf{T}$ is an $N \times N$ unitary matrix, satisfying $\mathbf{T}^T\mathbf{T} = \mathbf{I}$, where in this paper, if a matrix is complex, superscript T denotes *complex-conjugation and transposition*. Thus, since the property $\mathbf{V}^T\mathbf{R}\mathbf{V} = \mathbf{I}$ is maintained, we have again $\mathbf{Q}^T\mathbf{R}\mathbf{Q} = \mathbf{I}_k$, and as before $\mathbf{Q}$ is selected from $\mathbf{V}$ according to the indices which correspond to the k largest magnitude elements in

$$y = V^T Re = T^T He = T^T s_w \ , \tag{23}$$

where s_w denotes the weighted speech vector, i.e., $s_w \triangleq He$.

We see from (23) that with this more general selection of $\boldsymbol{V}$, y can be considered to be a *transform* of s_w, where the transform is defined by the unitary matrix T^T. Thus, as common in transform techniques, data compression is achieved by zeroing out the $(N-k)$ smallest magnitude elements in y, producing $\hat{y}$, from which the reconstructed weighted speech vector $\hat{s}_w$ is obtained by inverse transformation, i.e.,

$$\hat{s}_w = T\hat{y} \ . \tag{24}$$

Applying inverse weighting, by passing the elements of $\hat{s}_w$ through $1/W(z) = A(z/\gamma)/A(z)$ the reconstructed speech signal $\hat{s}_n$ can be obtained (care must be taken to add back the 'hangover' signal l_n). Alternatively, one can first find Q, by selecting the columns in V corresponding to the indices of the nonzero elements in $\hat{y}$, and then determine the excitation vector u from $u = Qx$, (where the elements of the vector x are the k nonzero elements of $\hat{y}$). Passing the elements of the excitation vector u through the LPC synthesis filter $1/A(z)$ (without zeroing-out its memory, so that (4) is satisfied) results in the reconstructed speech signal $\hat{s}_n$.

Note again that if $\gamma = 0$ is used, $A(z/\gamma) = 1$, and hence $s_w = e$ and the LPC residual signal is the one being transformed by the unitary transform represented by T^T. On the other hand, if $\gamma = 1$ is used, no weighting is performed ($W(z) = 1$), and the transform is applied directly to the input speech vector (after the removal of the 'hangover' sequence l_n). Thus, it may appear that conventional transform coding [11] also falls into this framework. However, since here the transform is applied to the weighted input speech vector, and only after the 'hangover' signal from the previous frame is removed, this scheme is different. It has the advantage of reducing frame-to-frame redundancy and also alleviating frame-to-frame transition effects (by adding back the 'hangover' signal at the receiver which smoothes these transitions), thus eliminating the need for overlapping frames - as commonly done in transform coders [12]. Because of this strong link with the LPC model of the input speech signal, we term here this coding scheme 'predictive transform coding' (PTC). The issue of selecting the unitary transform T and a simplified practical PTC are discussed in the following sections.

4. OPTIMALITY CONSIDERATIONS

In this section we consider the problem of finding the best vector set $\boldsymbol{V}$ for minimizing the error E_w in (7), under the constraint that only the knowledge of H may be assumed in constructing the vectors in the set. In terms of the weighted input speech vector s_w and its reconstruction $\hat{s}_w$, E_w in (7) can be written as:

$$E_w = (s_w - \hat{s}_w)^T(s_w - \hat{s}_w) = \| s_w - \hat{s}_w \|^2 \ . \tag{25}$$

Substituting (10) into (8) and using

$$\hat{s}_w = Hu \tag{26}$$

we obtain that $\hat{s}_w$ can also be expressed as

$$\hat{s}_w = \hat{H}e \ , \tag{27}$$

where $\hat{H}$ is an $N \times N$ matrix of rank k. Thus, (25) can be written as

$$E_w = \| (H - \hat{H})e \|^2 \ , \tag{28}$$

and by a well known norm inequality [13] we have,

$$E_w = \| \Delta_H e \|^2 \le \| \Delta_H \|^2 \| e \|^2 \ , \tag{29}$$

where $\Delta_H \triangleq H - \hat{H}$, and the *matrix norm* assumed here is the Forbenius norm [13], given by the square root of the sum of squares of all the elements in the matrix, (or equivalently, by the square root of the trace of the given matrix times its transpose). Since in choosing $\boldsymbol{V}$ we may only assume knowledge of H, we consider, in view of (29), choosing it such that the upper bound on the error (given by the right hand side of (29)) could be minimized. Let us consider first, therefore, the problem of finding $\hat{H}$, from the given matrix H, such that the norm of Δ_H will be minimized. The solution to this problem is given through the singular value decomposition (SVD) [13,14] of H, as follows:

The SVD of the $N \times N$ nonsingular matrix H is given by

$$H = \Psi D \Phi^T \tag{30}$$

where Ψ and Φ are unitary $N \times N$ matrices satisfying:

$$R\Phi = \Phi D^2 \ ; \qquad R'\Psi = \Psi D^2 \ , \tag{31}$$

where

$$R \triangleq H^T H \ ; \qquad R' \triangleq HH^T \ , \tag{32}$$

and $D = Diag(d_1, d_2, \cdots, d_N)$, with the elements on the main diagonal being the singular values of H (i.e., the positive square root of the eigenvalues of R, or R'). The important result which we are going to use here is that if the singular values are arranged in (30) in decreasing order of their magnitudes then the best least-squares approximation of the matrix H, by a lower rank ($k < N$) matrix $\hat{H}$, (i.e., minimizing $\| \Delta_H \|$ defined above, over all possible rank k matrices), is given by [14,15]

$$\hat{H} = \hat{H}_o \triangleq \Psi D_k \Phi^T \ , \tag{33}$$

where $D_k = Diag(d_1, \cdots, d_k, 0, 0, \cdots, 0)$.

Using this result we can find now a vector set $\boldsymbol{V}$ which corresponds to this approximation. Substituting (33) in (27), replacing e by $H^{-1}s_w$, and using $H^{-1} = \Phi D^{-1} \Psi^T$ (from (30)), we obtain

$$\hat{s}_w = \Psi P_k \Psi^T s_w \ , \tag{34}$$

where $P_k = Diag(1, 1, \cdots, 1, 0, 0, \cdots 0)$ has k nonzero elements on the main diagonal. In (34), $\hat{s}_w$ can interpreted as the result of the following operations: First, s_w is transformed by the unitray transform matrix Ψ^T, then the $N-k$ elements in the transformed vector which correspond to the smallest magnitude singular values are zeroed-out (by P_k), and finally, the resulting vector is inverse transformed by Ψ. Thus, we can determine the corresponding vector set V (represented by the matrix V), by letting $T = \Psi$ in (22), resulting in

$$V = H^{-1}\Psi = \Phi D^{-1} \ . \tag{35}$$

Now that V has been found, we can minimize the error (for the given vector set $\boldsymbol{V}$), as was done in section 3.4, by zeroing out the smallest magnitude $N-k$ elements in the transform of s_w (whose indices do not necessarily coincide with the indices of the smallest magnitude singular values). The transformed vector coincides then with y in (23), and (24) can be used for reconstruction.

Using $s_w = He$ and substituting (30) for H, we obtain from (23) (with $T = \Psi$),

$$y = \Psi^T s_w = D\Phi^T e = D\bar{e} \ , \tag{36}$$

where $\bar{e}$ denotes the transformed residual vector by the unitary transform Φ^T. Thus, given H, one can either use its left singular matrix Ψ for transforming s_w (resulting in y), and obtain from it $\hat{s}_w$ as discussed above, or one can use the right singular matrix Φ of H, for transforming e, and obtain y from the r.h.s. of (36), and then continue as above.

Furthermore, since $\hat{s}_w = Hu$ (see (26)) it could be more convenient to determine first the excitation vector u. To do this we need not find Q and x as required in (8), but can find it directly from the unitary transform of e , $\bar{e}$, by zeroing-out the elements in $\bar{e}$ according to the indices of the smallest magnitude elements in $D\bar{e}$ (i.e. in y) resulting in a vector $\bar{u}$ from which the excitation vector u is obtained by inverse transformation: $u = \Phi\bar{u}$. That this minimizes the error in (7) for the selected transform can also be seen by substituting for R in (7) the expression (from (31)) $R = \Phi D^2 \Phi^T$ and hence,

$$E_w = \sum_{i=1}^{N} d_i^2 (\bar{e}_i - \bar{u}_i)^2 \ , \tag{37}$$

where $\bar{e}_i$, $\bar{u}_i$ are the the elements of the transformed vectors $\bar{e}$,$\bar{u}$, respectively.

This is actually the approach taken in [8], with one difference, that instead of obtaining $\bar{u}$ by zeroing elements in $\bar{e}$, as discussed above, $\bar{\mathbf{u}}$ is a quantized version of $\tilde{\mathbf{e}}$, using vector quantization, with (37) being the error measure used for searching the code-book.

In the above discussion, the vector set $\boldsymbol{V}$ and the corresponding transforms, which were derived from the SVD of $\mathbf{H}$, were shown to ensure an error which is less or equal to the minimum value of the upper bound in (29), but are not necessarily optimal for the given $\mathbf{H}$. A sufficient condition for optimality is obtained from (29) by the following observation: If all the elements in the transformed residual vector $\Phi^T \mathbf{e}$ have the same magnitude, i.e. $\mathbf{e}$ is 'white' with respect to the transform Φ^T, then

$$\| \Delta_H \mathbf{e} \|^2 = \frac{1}{N} \| \Delta_H \|^2 \| \mathbf{e} \|^2 , \qquad (38)$$

and hence minimizing $\| \Delta_H \|$, by using $\hat{\mathbf{H}}$ in (33), also minimizes the error and hence the vector set corresponding to (35) and the corresponding transforms (Ψ^T for transforming $\mathbf{s}_w$ and Φ^T for transforming $\mathbf{e}$) are optimal, for the given $\mathbf{H}$.

To complete this discussion we consider briefly the case in which a statistical measure, instead of the deterministic measure E_w in (7) or (25) is used. We specifically consider the weighted mean square error (WMSE) given by the expected value of E_w, i.e., $\varepsilon^2 \triangleq E\{E_w\}$. The classical result is that the best transform for transforming $\mathbf{s}_w$ is the Karhunen-Loeve (KL) transform, Γ^T, given by the orthonormalized eigenvectors of the autocorrelation matrix [11]

$$\mathbf{R}_{s_w s_w} = \mathrm{E}\{\mathbf{s}_w \mathbf{s}_w^T\} = \mathbf{H}\mathrm{E}\{\mathbf{e}\mathbf{e}^T\}\mathbf{H}^T = \mathbf{H}\mathbf{R}_{ee}\mathbf{H}^T . \qquad (39)$$

Hence, if the residual signal, e_n, is spectrally flat, $\mathbf{R}_{ee} = \mathbf{I}$, and we obtain that $\mathbf{R}_{s_w s_w} = \mathbf{H}\mathbf{H}^T = \mathbf{R}'$. By (31), Ψ is the unitary matrix which diagonalizes $\mathbf{R}'$, and hence the KL transform is given then by $\Gamma^T = \Psi^T$. Thus, the KL transform coincides with the transform considered above for the deterministic case and which was found to be optimal if $\mathbf{e}$ is 'white' with respect to the transform derived from the right singular matrix Φ of $\mathbf{H}$.

5. A SIMPLIFIED PREDICTIVE TRANSFORM CODER

In view of the results and the discussion in Sections 3.4 and 4, we have considered simulating a simple predictive transform coder and compare its performance to the Multipulse LPC scheme. The main issue is the selection of the transform. Although $\mathbf{H}$ can be computed at the receiver from the transmitted LPC coefficients, finding its SVD is a computationally demanding task. Hence, as common in such situations, we have considered using a transform which is independent of $\mathbf{H}$. Noting that if the effective length of h_n is much less then N (as is particularly the case when $\gamma < 1$, since $h_n = \gamma^n \tilde{h}_n$ decays fast), the matrix $\mathbf{H}$ is banded and by adding only a small number of terms at the upper right corner can be approximated by circulant matrix (in fact $\mathbf{H}$ is asymptotically equivalent to a circular matrix [16]). Since any circulant matrix is diagonalizable by the DFT matrix [16], we consider the DFT as a candidate transform (note also that because a circular matrix is normal $\Psi = \Phi$ in its SVD decomposition). Similarly, using the arguments in [8] (based on the approximation of $\mathbf{R}$ by a Toeplitz matrix, for large N) the DCT approximately diagonalizes $\mathbf{R}$ and hence is also a candidate transform.

We examined first a coding system in which the weighted speech signal s_w is transformed by a DFT or a DCT and a small number of $k<N$ of the largest magnitude elements in the transformed vector are selected (and coded) for transmission. While providing relatively high SNR (with the DCT being preferable), this approach was found to result in audioable 'ringing' (tonal noise) when the transmission rate was 16 Kbps and below. To reduce this ringing effect we have attempted to use a scheme in which the bits available for coding in a given frame are distributed among more transform elements, as commonly done in Adaptive Transform Coders (ATC) [11,12]. However, here we used the scheme discussed in Section 4, in which the residual signal itself is transformed (using the DCT), but instead of picking k peaks according to the largest magnitude elements in $\mathbf{y}$, to obtain $\bar{\mathbf{u}}$, the elements of $\tilde{\mathbf{e}}$ are quantized using a bit assignment which varies from frame to frame according to the spectral envelope of $H(z) = 1/A(z/\gamma)$. This is based on (37), in which it seen that the squared singular values provide the weighting of the quantization error, and hence, since these are the eigenvalues of $\mathbf{R}$ (see (31)) they correspond to the spectral envelope of $H(z)$ (note also that the DFT and DCT provide identical spectral envelopes [12]). At the receiver, decoding and inverse transformation of $\bar{\mathbf{u}}$ provides the excitation vector $\mathbf{u}$ whose elements are input to the the LPC synthesis filter $\tilde{H}(z)$ (without zeroing-out its memory so as to add the 'hangover' signal) to obtain the reconstructed speech signal s_n.

The above simplified PTC scheme largely reduced the ringing effect and provided a very high quality reconstructed speech at 16 Kbps (a segmental SNR of 18.1 dB was obtained at this rate) and better then communication quality at 9.6 Kbps (with a segmental SNR of 12.9 dB). Both the quality and segmental SNR were higher, as compared with a corresponding multipulse LPC scheme (i.e., having no pitch loop), which achieved segmental SNR values of only 16.1 dB and 11.2 dB, respectively, for the same input speech data.

Further improvement at 9.6 Kbps could be achieved if the residual will be whitened (by using a pitch loop), before its transformation, as this will provide a better approximation of the DCT (or DFT) to the optimal transform, as discussed in section 4. Ultimately, adding the pitch loop and replacing the scalar quantization used here by vector quantization, the scheme in [8] is obtained.

REFERENCES

[1] B.S. Atal, "Predictive Coding of Speech at Low Bit Rates", IEEE Trans. on Comm., vol. COM-30, pp. 600-614, April 1982, pp. 600-614.

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

[3] S.T. Alexander, "A Simple Noniterative Speech Excitation Algorithm Using the LPC Residual", IEEE Trans. on ASSP, vol. ASSP-33, No. 2, April 1985.

[4] E.F. Deprettere and P. Kroon, "Regular Excitation Reduction for Effective and Efficient LP-Coding of Speech", Proc. Int. Conf. ASSP, 1985, pp. 965-968.

[5] P. Kroon, E.F. Deprettere and R.J. Sluyter, "Regular-Pulse Excitation - A Novel Approach to Effective and Efficient Multipulse Coding of Speech", IEEE Trans. on ASSP, vol. ASSP-34, No. 5, Oct. 1986, pp. 1054-1063.

[6] A. Ichikawa, S. Takeda and Y. Asakawa, "A Speech Coding Method Using Thinned-Out Residual, Proc. Int. Conf. ASSP, 1985, pp. 961-964.

[7] M.R. Schroeder and B.S. Atal, "Code-Excited Linear Prediction (CELP): High-Quality Speech At Very Low Bit Rates", IEEE Proc. Int. Conf. on ASSP, 1985, pp. 937-940.

[8] T. Moriya, M. Honda, "Transform Coding of Speech With Weighted Vector Quantization", Proc. Int. Conf. on ASSP, 1987, pp. 1629-1632.

[9] Coders for High Quality Speech Coding at Rates Between 4.8 and 16 kbits/s", IEEE Journal on Selected Areas In Commun., vol. 6, No. 2, Feb. 1988, pp. 353-363.

[10] E. Ofer, A. Dembo and D.malah, "A Unified Framework for Residual Coding of Speech", Proc. IEEE Seminar on Digital Signal Processing for Communication, Herzlia, Israel, Dec. 1987, pp. 13-19; and E. Ofer, *Speech Coding By Vector Expansion of the LPC Residual Signal*, M.Sc. Thesis (in Hebrew), Technion - Israel Institute of Technology, Haifa, Israel, Sept. 1986.

[11] N.S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice-Hall, NJ, 1984 (Ch. 12).

[12] J.M. Tribolet and R.E. Crochiere, "Frequency Domain Coding of Speech", IEEE Trans. on ASSP, vol. ASSP-27, No. 5, Oct. 1979, pp. 512-530.

[13] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, (2nd Ed.), Academic Press, NY, 1985.

[14] E.F. Deprettere, Ed., *SVD and Signal Processing*, North-Holland, Amsterdam, 1988. (Pt. I)

[15] H.C. Andrews and C.L. Patterson, III, "Singular Value Decomposition (SVD) Image Coding", IEEE Trans. on Commun., vol. 24, April 1976, pp. 425-432.