SPECTRAL MAGNITUDE QUANTIZATION BASED ON LINEAR TRANSFORMS FOR 4 KB/S SPEECH CODING

Çağrı Özgenç Etemoğlu and Vladimir Cuperman

Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106
E-mail:[cagri, vladimir]@scl.ece.ucsb.edu

ABSTRACT

This paper presents a matching pursuits sinusoidal speech coder which incorporates new techniques including a novel vector quantization (VQ) technique used for the weighted quantization of spectral magnitude vector, and an inter-frame quantization of spectral magnitudes using an interpolation matrix that minimize the weighted interpolation error. The paper describes a novel vector quantization technique, wherein the quantized vector is obtained by applying a linear transformation selected from a first codebook to a codevector selected from a second codebook. The transformation is selected from a family of linear transformations, represented by a matrix codebook. Vectors in the second codebook are called residual codevectors. In order to avoid high complexity during the search for the best linear transformation, each linear transformation is assigned a representative vector, such that the search can be done employing the representative vectors. The VQ design algorithm is based on joint optimization of the linear transformation and the residual codebooks. The introduced techniques are general enough to be used in any sinusoidal speech coding scheme. In this work we incorporated the techniques into the matching pursuits sinusoidal model to achieve high quality speech using sinusoidal speech coder at 4 kbps. Subjective tests indicate that the proposed coding model at 4 kbps has quality comparable to that of G.729 at 8kbps.

1. INTRODUCTION

There is a growing demand to develop toll quality speech coders at rates of 4 kbps and below. The successful use of analysis-by-synthesis search procedure combined with a perceptually weighted error measure has enabled the waveform coders such as *code-excited linear predictive* CELP coder [1] to achieve toll or nearly toll quality speech at rates above 5 kbps. However rapid degradation of speech quality below 5 kbps show that waveform matching criteria does not work well at low rates. On the other hand, parametric coders such as the *sinusoidal-transform coder* (STC) [2], the *waveform-interpolative* (WI) coder [3], the *multiband-excitation* (MBE) coder [4], and the *mixed excitation linear prediction* (MELP) coder [5] try to find a parametric representation of the speech without the constraint of preserving the waveform and can synthesize good quality speech at rates as low as 2 kbps, but they do not achieve toll quality.

This paper presents a matching pursuits sinusoidal speech coder based on the sinusoidal model described in [6]. In general sinusoidal speech coders belong to the parametric coding category, so they can only be successful if the model parameters are quantized efficiently. Since in sinusoidal coders the variable dimension spectral magnitude vector usually consume a large amount of available bits, and their faithful reproduction is crucial for toll quality, their efficient quantization is needed. To meet this goal, we propose a novel VQ scheme for the weighted quantization of spectral magnitude vector, and a matrix based interpolation of spectral magnitudes that minimize the weighted interpolation error for the inter-frame coding of magnitudes. In the past, the idea of using a linear transform in vector quantization in a different context is suggested in [7], but no design rules are derived for optimal quantization.

In this paper, the variable dimension spectral vector is first transformed into a fixed dimension vector, and then the fixed dimension vector is quantized efficiently using the proposed VQ technique. The proposed VQ approach reconstructs the input vector by applying a linear transformation selected from a first codebook to a codevector selected from a second codebook. The transformation is selected from a family of linear transformations, represented by codebook of matrices. Vectors in the second codebook are called residual codevectors. In order to avoid high complexity during the search for the best linear transformation, each linear transformation is assigned a representative vector, such that the search can be done employing the representative vectors. The design algorithm is based on joint optimization of the linear transformation and the residual codebooks. Subjective tests using the 4 kbps matching pursuit sinusoidal coder shows that it has quality comparable to that of G.729 at 8kbps.

2. PROPOSED CODER SCHEME

Figure 1 shows the overall structure of the encoder and the decoder. In the encoder, the input LP residual is classified into one of the three classes: voiced, transition, or unvoiced speech. Then depending on the class the corresponding matching pursuit analysis is performed to determine sinusoidal model parameters. Extracted spectral magnitude and phase (only in transition class) information are quantized. The decoder uses overlap-and-add synthesis model to reconstruct the LP residual. The processing flow in the whole encoder/decoder system is the same as that in described in [6]. Therefore, only the new techniques are discussed in this paper.

This work was supported in part by the National Science Foundation under grant no. NCR-9314335, the University of California MICRO program, Cisco Systems, Inc., Dialogic Corp., Fujitsu Laboratories of America, Inc., General Electric Co., Hughes Network Systems, Intel Corp., Lernout & Hauspie Speech Products, Lockheed Martin, Lucent Technologies, Inc., Qualcomm, Inc., Rockwell International Corp., Panasonic Technologies, Inc., and Texas Instruments, Inc.

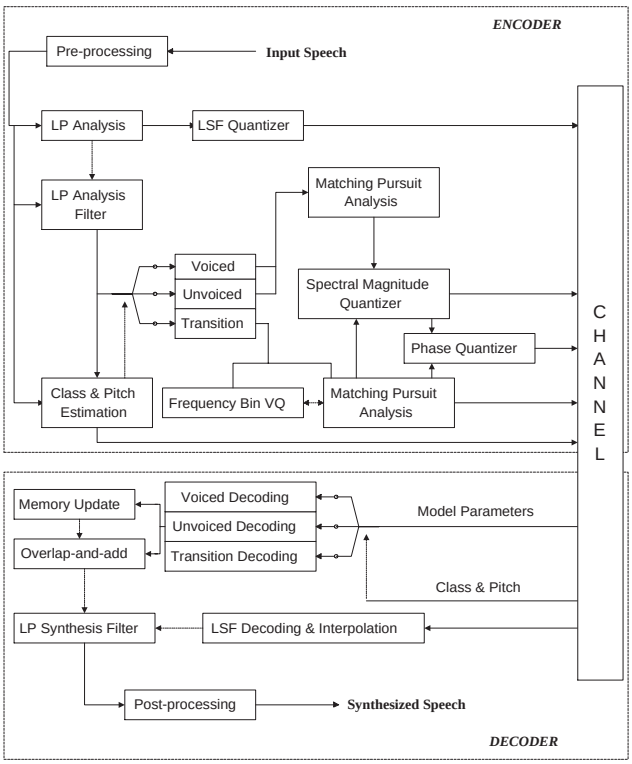

Figure 1: Block diagram of the Encoder and Decoder

2.1. Spectral Magnitude Quantization

In sinusoidal coders, generally, the spectral magnitudes are obtained by sampling the spectrum of either the speech or the LP residual at frequencies corresponding to pitch harmonics. In the proposed sinusoidal coder, matching pursuits [8] is used in a pitch dependent manner [6] to determine the spectral magnitudes. Both of these procedures generate a variable dimension vector since the number of pitch harmonics changes in time.

In the proposed scheme the spectral magnitudes are determined twice per frame. The variable dimension spectral vectors of dimension less than 48 are transformed into one of the three possible fixed dimensions ($M = 24, 36, 48$) by zero padding. The spectral vectors of dimension greater than 48 are truncated to a fixed dimension of $M = 48$. To exploit the high correlation of magnitudes between the 1st and 2nd subframe, only the magnitudes corresponding to 2nd subframe are quantized. The 1st subframe magnitudes are estimated using a linear interpolation of the quantized 2nd subframe magnitudes in current and previous frames. Then the interpolation error is quantized. This section describes the weighted VQ technique, and the interpolation scheme used in inter-frame coding of magnitudes is described in the next section.

Let $\mathbf{x}$ be an M-dimensional input vector. According to the proposed approach, the quantized vector $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = \hat{\mathbf{T}}\hat{\mathbf{c}} \tag{1}$$

where $\hat{\mathbf{T}}$ represents the linear transformation matrix selected from the matrix codebook $\mathcal{C}_T$ and $\hat{\mathbf{c}}$ represents a residual codevector which is a member of the residual codebook $\mathcal{C}_r$.

The quantization distortion criterion is defined by the distance between the original spectral vector and the quantized spectral vector weighted by the spectral magnitude of the LP synthesis filter and a perceptual weighting filter. The weighting matrix $\mathbf{W}$ is diagonal and ith diagonal element corresponding to the ith spectral sample at the frequency ω_i is given by

$$w_i = |\frac{A(z/\gamma_1)}{A(z)A(z/\gamma_2)}|^2 z = \exp(j\omega_i) \qquad 0 \le \gamma_2 < \gamma_1 \le 1 \tag{2}$$

The average distortion D on a set of N vectors $\{\mathbf{x_k}\}$ with weighting matrices $\{\mathbf{W_k}\}$ is

$$D = \frac{1}{N}\sum_{k=0}^{N-1} \|\mathbf{x_k} - \hat{\mathbf{x}}_\mathbf{k}\|^2_{\mathbf{W_k}} \tag{3}$$

or, based on (1):

$$D = \frac{1}{N}\sum_{k=0}^{N-1} \|\mathbf{x_k} - \hat{\mathbf{T}}_\mathbf{k}\hat{\mathbf{c}}_\mathbf{k}\|^2_{\mathbf{W_k}} \tag{4}$$

where $\hat{\mathbf{T}}_\mathbf{k}$ is the transformation matrix and $\hat{\mathbf{c}}_\mathbf{k}$ is the residual codevector corresponding to the input vector $\mathbf{x_k}$.

The objective here is to design the codebooks $\mathcal{C}_T$, $\mathcal{C}_r$ that minimize (4) and to develop an efficient coding rule for this VQ technique.

2.1.1. Encoding/Decoding

Given the linear transformation codebook $\mathcal{C}_T$ and the residual codebook $\mathcal{C}_r$, the optimal pair $(\hat{\mathbf{T}}, \hat{\mathbf{c}})$ for encoding the vector $\mathbf{x}$ is simply given by

$$(\hat{\mathbf{T}}, \hat{\mathbf{c}}) = \arg \min_{\hat{\mathbf{T}}\in\mathcal{C}_T, \hat{\mathbf{c}}\in\mathcal{C}_r} \|\mathbf{x} - \hat{\mathbf{T}}\hat{\mathbf{c}}\|^2_{\mathbf{W}} \tag{5}$$

The minimization required in (5) is computationally intensive if an exhaustive search in both codebooks is employed. To avoid high search complexity, a sequential search is employed whereby the linear transformation $\mathbf{T}$ is determined first.

To simplify the search of the linear transformation codebook $\mathcal{C}_T$, we map this codebook into a set of codevectors $\{\mathbf{t_j}\}$ stored in $\mathcal{C}_t$, so that the ith matrix of $\mathcal{C}_T$ namely, $\mathbf{T_i}$, is mapped into a corresponding codevector, $\mathbf{t_i}$ in $\mathcal{C}_t$. The codebooks $\mathcal{C}_T$ and $\mathcal{C}_t$ are related such that the linear transformation to be assigned to the input vector $\mathbf{x}$ will be given by the code-matrix i iff

$$\mathbf{t_i} = \arg \min_{\mathbf{t_j}\in\mathcal{C}_t} \|\mathbf{x} - \mathbf{t_j}\|^2_{\mathbf{W}} \tag{6}$$

Note that the search in (6) has the same computational complexity as the usual VQ search. However, the use of transforms associated with the vectors $\{\mathbf{t_j}\}$ allow us to trade-off a larger memory (required for storing the transforms) for improved performance.

Once the vector $\mathbf{t_i}$ is determined, the associated linear transformation $\hat{\mathbf{T}} = \mathbf{T_i}$ is employed to search the second stage by choosing $\hat{\mathbf{c}}$ to minimize

$$\min_{\hat{\mathbf{c}}\in\mathcal{C}_r} \|\mathbf{x} - \hat{\mathbf{T}}\hat{\mathbf{c}}\|^2_{\mathbf{W}} \tag{7}$$

The quantized vector is given by $\hat{\mathbf{x}} = \hat{\mathbf{T}}\hat{\mathbf{c}}$. Depending on the memory and complexity requirements the search in (7)

can be done by either generating the reconstruction vectors using matrix multiplication at the time of search, or storing pre-computed reconstruction vectors. In the former case, the complexity is larger than MSVQ, while in the latter case the computational complexity is practically the same as in MSVQ.

2.1.2. Joint Codebook Optimization

In order to jointly optimize the codebooks, we use an iterative sequential optimization. The algorithm iterates between optimizing linear transformation codebook $\mathcal{C}_T$ and the associated $\mathcal{C}_t$ for a given residual codebook $\mathcal{C}_r$ and optimizing the residual codebook for the given linear transformation codebook.

In order to sequentially optimize the codebooks, the input vector space is partitioned with respect to the codebook whose entries are being optimized. Let $R_{i,j}$ denote the set of input vectors whose assigned indices are i for the codebook $\mathcal{C}_T(\mathcal{C}_t)$, and j for the codebook $\mathcal{C}_r$. Given $R_{i,j}$, the set of input vectors assigned to the ith entry of the codebook $\mathcal{C}_T(\mathcal{C}_t)$ is given by

$$U_i = \bigcup_{j=1}^{N_r} R_{i,j} \tag{8}$$

and the set of vectors assigned to the jth entry of residual codebook $\mathcal{C}_r$ is

$$V_j = \bigcup_{i=1}^{N_T} R_{i,j} \tag{9}$$

where N_r is the size of $\mathcal{C}_r$ and N_T is the size of both $\mathcal{C}_T$ and $\mathcal{C}_t$.

2.1.3. Design of the Linear Transformation Codebook For a Given Residual Codebook

Given the fixed residual codebook and the partition U_i, our objective is to compute $\mathbf{T_i}$ for $i = 1, \ldots, N_T$ to minimize (4). In other words, $\mathbf{T_i}$ is obtained as the solution of the optimization problem

$$\mathbf{T_i} = \arg\min_{\hat{\mathbf{T}}} \sum_{k:\mathbf{x_k}\in U_i} \|\mathbf{x_k} - \hat{\mathbf{T}}\hat{\mathbf{c}}_\mathbf{k}\|^2_{\mathbf{W_k}} \tag{10}$$

The solution of the above minimization problem may not be unique. The jth row of $\mathbf{T_i}$, $\mathbf{r_{ij}}$ will be chosen as the solution with the minimum norm and is given by

$$\mathbf{r}^\mathbf{T}_\mathbf{ij} = \mathbf{z}^\mathbf{T}_\mathbf{ij}\mathbf{Y}^+_\mathbf{ij} \quad j = 1, \ldots, M \tag{11}$$

where

$$\mathbf{z}^\mathbf{T}_\mathbf{ij} = [w^{1/2}_{1,j}x_{1,j} \cdots w^{1/2}_{\|U_i\|,j}x_{\|U_i\|,j}] \tag{12}$$

$$\mathbf{Y_{ij}} = [w^{1/2}_{1,j}\hat{\mathbf{c}}_\mathbf{1} \cdots w^{1/2}_{\|U_i\|,j}\hat{\mathbf{c}}_{\|\mathbf{U_i}\|}] \tag{13}$$

$\mathbf{Y^+_{ij}}$ denotes the pseudoinverse of $\mathbf{Y_{ij}}$, $w_{k,j}$ denotes jth diagonal element of the kth weight matrix, and $\|U_i\|$ denotes the cardinality of U_i.

Experimental evidence shows that a good way of designing $\mathbf{t_i}$ for $i = 1, \ldots, N_T$, is to update $\mathbf{t_i}$ as the weighted Euclidean centroid of the reconstructed vectors $\hat{\mathbf{x}}_\mathbf{k}$ whose input vectors $\mathbf{x_k} \in \mathbf{U_i}$;

$$\mathbf{t_i} = (\sum_{k:\mathbf{x_k}\in U_i} \mathbf{W_k})^{-1} \sum_{k:\mathbf{x_k}\in U_i} \mathbf{W_k}\hat{\mathbf{x}}_\mathbf{k} = (\sum_{k:\mathbf{x_k}\in U_i} \mathbf{W_k})^{-1} \sum_{k:\mathbf{x_k}\in U_i} \mathbf{W_k}\mathbf{T_i}\hat{\mathbf{c}}_\mathbf{k} \tag{14}$$

There is a simple analytical justification for this approach. In the case of high bit rate quantization or highly clustered input vectors, for an input vector $\mathbf{x_n}$ which has $\mathbf{t_i} = \arg\min_{\mathbf{t_j}\in\mathcal{C}_t} \|\mathbf{x_n} - \mathbf{t_j}\|^2_{\mathbf{W_n}}$, the weighted Euclidean distance between $\mathbf{t_i}$ and $\hat{\mathbf{x}}_\mathbf{n}$ will be small due to (14). Furthermore using the triangle inequality the Euclidean distance between $\mathbf{x_n}$ and $\hat{\mathbf{x}}_\mathbf{n}$ can be upper bounded as

$$\|\mathbf{x_n} - \hat{\mathbf{x}}_\mathbf{n}\|_{\mathbf{W_n}} \le \|\mathbf{x_n} - \mathbf{t_j}\|_{\mathbf{W_n}} + \|\mathbf{t_j} - \hat{\mathbf{x}}_\mathbf{n}\|_{\mathbf{W_n}} \tag{15}$$

The right hand side of (15) is expected to have a low value at $j = i$, because the first term is minimized by the choice $j = i$ and the second term corresponds to the weighted distance between a vector and its weighted centroid. This shows that by employing the sequential encoding rule given by (6) we can obtain a low value for the upper bound on the quantization error.

2.1.4. Design of the Residual Codebook For a Given Linear Transformation Codebook

Given the fixed linear transformation codebook and the partition V_j, we will compute $\mathbf{c_j}$ for $j = 1, \ldots, N_r$ to minimize (4). So $\mathbf{c_j}$ will be given by

$$\mathbf{c_j} = \arg\min_{\hat{\mathbf{c}}} \sum_{k:\mathbf{x_k}\in V_j} \|\mathbf{x_k} - \hat{\mathbf{T}}_\mathbf{k}\hat{\mathbf{c}}\|^2_{\mathbf{W_k}} \tag{16}$$

The minimum norm solution of the above minimization is the centroid equation for the jth centroid and computed as

$$\mathbf{c_j} = \mathbf{A}^+\mathbf{b} \tag{17}$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{W}^{\mathbf{1/2}}_\mathbf{1}\hat{\mathbf{T}}_\mathbf{1} \\ \vdots \\ \mathbf{W}^{\mathbf{1/2}}_{\|\mathbf{V_j}\|}\hat{\mathbf{T}}_{\|\mathbf{V_j}\|} \end{bmatrix} \mathbf{b} = \begin{bmatrix} \mathbf{W}^{\mathbf{1/2}}_\mathbf{1}\mathbf{x_1} \\ \vdots \\ \mathbf{W}^{\mathbf{1/2}}_{\|\mathbf{V_j}\|}\mathbf{x}_{\|\mathbf{V_j}\|} \end{bmatrix} \tag{18}$$

2.1.5. Joint Codebook Design

The main design algorithm can now be stated by using the centroid computations and the sequential encoding rule described in earlier sections.

Once the codebooks are initialized, the main design algorithm performs the following steps:

1. Partition the training set to obtain $R_{i,j}$.
2. Compute the overall distortion, if termination criterion is satisfied then stop else continue.
3. Compute the optimum codebook $\mathcal{C}_T$ using (11), update the codebook $\mathcal{C}_t$ using (14).
4. Partition the training set to obtain a new $R_{i,j}$.
5. Compute the optimum codebook $\mathcal{C}_r$ using (17).
6. Go to 1.

While steps 3 and 5 of this algorithm always decrease the overall distortion, the partitioning steps 1 and 4 may increase the distortion due to the suboptimal sequential encoding rule. Hence, the algorithm does not guarantee strict descent, however, in practice the distortion generally decreases. The termination criterion adopted in this algorithm is to stop when the relative change in the distortion is less than a given threshold.

2.2. Inter-frame Quantization of Spectral Magnitudes

As mentioned before, the variable dimension spectral vectors are transformed into one of the three possible fixed dimensions ($M = 24, 36, 48$). Using a linear interpolation of the quantized 2nd subframe magnitudes in current and previous frames, an estimate of the 1st subframe magnitudes is constructed as

$$\tilde{\mathbf{x}}_{\mathbf{1,k}} = \mathbf{P}\hat{\mathbf{x}}_{\mathbf{2,k-1}} - (\mathbf{I} - \mathbf{P})\hat{\mathbf{x}}_{\mathbf{2,k}} \tag{19}$$

where $\mathbf{x_{i,k}}$ denote the fixed dimension vector corresponding to ith subframe of frame k for $i = 1, 2$ and P denotes the interpolation matrix.

For each possible fixed dimension, an optimum interpolation matrix is determined that minimize the weighted interpolation error (weights are defined in section 2.1);

$$\begin{aligned}\mathbf{P} &= \arg\min_{\mathbf{P}} \sum_{k:\mathbf{x}_{\mathbf{1,k}} \in R_M} \|\mathbf{x}_{\mathbf{1,k}} - \tilde{\mathbf{x}}_{\mathbf{1,k}}\|^2_{\mathbf{W_k}} \qquad (20)\\ &= \arg\min_{\mathbf{P}} \sum_{k:\mathbf{x}_{\mathbf{1,k}} \in R_M} \|\mathbf{x}_{\mathbf{1,k}} - \hat{\mathbf{x}}_{\mathbf{2,k}} - \mathbf{P}(\hat{\mathbf{x}}_{\mathbf{2,k-1}} - \hat{\mathbf{x}}_{\mathbf{2,k}})\|^2_{\mathbf{W_k}}\end{aligned}$$

where R_M denotes the set of 1st subframe vectors of dimension M. To simplify the notation, the vectors $\mathbf{f_k}$ and $\mathbf{g_k}$ are defined as

$$\mathbf{f_k} = \mathbf{x}_{\mathbf{1,k}} - \hat{\mathbf{x}}_{\mathbf{2,k}} \tag{21}$$

$$\mathbf{g_k} = \hat{\mathbf{x}}_{\mathbf{2,k-1}} - \hat{\mathbf{x}}_{\mathbf{2,k}} \tag{22}$$

The minimization problem can be reformulated as

$$\mathbf{P} = \arg\min_{\mathbf{P}} \sum_{k:\mathbf{x}_{\mathbf{1,k}} \in R_M} \|\mathbf{f_k} - \mathbf{P}\mathbf{g_k}\|^2_{\mathbf{W_k}} \tag{23}$$

The solution of the above minimization problem may not be unique. The jth row of $\mathbf{P}$, $\mathbf{p_j}$ will be chosen as the solution with the minimum norm and is given by

$$\mathbf{p_j^T} = \mathbf{q_j^T}\mathbf{V_j^+} \qquad j = 1, \ldots, M \tag{24}$$

where

$$\mathbf{q_j^T} = [w_{1,j}^{1/2} f_{1,j} \cdots w_{\|R_M\|,j}^{1/2} f_{\|R_M\|,j}] \tag{25}$$

$$\mathbf{V_j} = [w_{1,j}^{1/2} \mathbf{g_1} \cdots w_{\|R_M\|,j}^{1/2} \mathbf{g}_{\|\mathbf{R_M}\|}] \tag{26}$$

$\mathbf{V_j^+}$ denotes the pseudoinverse of $\mathbf{V_j}$, $w_{k,j}$ denotes jth diagonal element of the kth weight matrix, and $\|R_M\|$ denotes the cardinality of R_M.

3. BIT ALLOCATION

The bit allocation for basic types of speech frame is given in Table 1. The frame size is 20 ms.

We used the described techniques specifically in voiced frames for which accurate quantization of spectral magnitudes is needed to achieve high perceptual quality. In voiced frames, the variable dimension spectral magnitudes corresponding to the second subframe are quantized using a 3-stage (7+7+7) VQ with the first two stages designed using weighted linear transformations method introduced, and a last stage designed to quantize the residual error. The interpolation error defined in Section 2.2 corresponding to 1st subframe spectral magnitudes is quantized using 7 bits. Finally the gain is quantized using 6 bits resulting in a total of 34 bits. The remaining 6 bits are used to quantize spectral magnitudes corresponding to the aperiodic components of the voiced frame [6].

Parameter	Transition	Voiced	Unvoiced
LSF	18	24	18
Magnitude	27	40	34
Frequency	$2 \times 3 = 6$	0	0
Pitch	0	$8 + 5 = 13$	0
Linear Phase	4	0	0
Dispersion Phase	$2 \times 6 = 12$	0	0
Envelope	$2 \times 5 = 10$	0	$2 \times 9 = 18$
Classifier	3	3	3
Total	80	80	73
Bit-rate	4kbps	4kbps	3.65kbps

Table 1: Bit allocation

4. SUBJECTIVE RESULTS

We have conducted a preference listening test to compare the subjective performance of the proposed matching pursuits sinusoidal coder with the G.729 standard. The test data included 16 MIRS speech sentences, 8 from female speakers and 8 from male speakers. Eight listeners participated in the test. The subjective test results presented in Table 2, indicate that the proposed coder at 4 kbps have quality comparable to that of G.729 at 8kbps.

Speakers	Proposed coder	G.729	Same
Female	21.88%	43.75%	34.37%
Male	43.75%	18.75%	37.50%
Total	32.81%	31.25%	35.94%

Table 2: Preference test results

REFERENCES

[1] M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (celp): High-quality speech at very low bit rates," in *Proc. of ICASSP*, pp. 937–940, 1985.

[2] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 34, pp. 744–754, August 1986.

[3] W. B. Kleijn, "Encoding speech using prototype waveforms," *IEEE Trans. on Speech and Audio Processing*, vol. 1, pp. 386–399, October 1993.

[4] D. W. Griffin and J. S. Lim, "Multiband excitation vocoder," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 36, pp. 1223–1235, August 1988.

[5] A. V. McCree and T. P. B. III, "A mixed excitation lpc vocoder model for low bit rate speech coding," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 242–249, July 1995.

[6] C. O. Etemoglu, V. Cuperman, and A. Gersho, "Speech coding with an analysis-by-synthesis sinusoidal model," in *Proc. of ICASSP*, pp. 1371–1374, 2000.

[7] D. H. Lee, D. L. Neuhoff, and K. K. Paliwal, "Cell-conditioned multistage vector quantization," in *Proc. of ICASSP*, pp. 653–656, 1991.

[8] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. on Signal Processing*, vol. 41, pp. 3397–3415, December 1993.