An enhanced analysis-by-synthesis waveform interpolative speech coder able to operate at 2.8 kbps. Novel features include dual-predictive analysis-by-synthesis quantization of the slowly-evolving waveform, efficient parametrization of the rapidly-evolving waveform magnitude, and analysis-by-synthesis vector quantization of the rapidly evolving waveform parameter. Subjective quality tests indicate that it exceeds G.723.1 at 5.3 kbps, and of G.723.1 at 6.3 kbps.

18 Claims, 6 Drawing Sheets
OTHER PUBLICATIONS


Oded Gottesman, “Dispersion Phase Vector Quantization for Enhancement of Waveform Interpolative Coder,” Signal Compression Laboratory, Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106, USA, pp. 1-4.

Oded Gottesman et al., “Enhanced Waveform Interpolative Coding at 4 KBPS,” Signal Compression Laboratory, Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106, USA, pp. 1-3.


Oded Gottesman et al., “Enhanced Analysis-by-Synthesis Waveform Interpolative Coding at 4 KBPS,” Signal Compression Laboratory, Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106, USA, pp. 1-4.


* cited by examiner
FIG. 1

FIG. 2

VQ

\[ R(\omega) \]

VECTOR OFREW SPECTRA

\[ \hat{R}(\xi, \omega) \]

\[ \bar{R}(\xi, \omega) \]

\[ \min ||\hat{R}(\xi, \omega) - \bar{R}(\xi, \omega)||^2 \]

\[ \hat{R}(\xi, \omega) \]

\[ j \]

\[ \hat{\xi}_j \]

VECTOR QUANTIZER CODEBOOK

\[ \bar{R}(\xi, \omega) \]

\[ \hat{R}(\xi, \omega) \]

\[ \hat{\xi}_j \]

VECTOR QUANTIZER CODEBOOK

\[ VQ^{-1} \]

\[ \hat{R}(\omega) \]

VECTOR OF QUANTIZEDREW SPECTRA

FREQUENCY \( \omega \) [radians]
FIG. 3

GAIN g(m)
LPC a(m)
REW POLYNOMIAL COEFFICIENTS γ(m)

PREDICTOR CODEBOOK
VECTOR QUANTIZER CODEBOOK

P_i

\hat{C}_{ij}(m)

SYNTHESIS FILTER \frac{1}{1 - P_i z^{-1}} \hat{\xi}(m)

\tilde{\gamma}(\hat{\xi})

SPECTRAL WEIGHTING

\|\ast\|^2

TEMPORAL WEIGHTING

min(*)
FIG. 7

![Graph showing output SNR vs bits for mean-removed and output SEW.]

FIG. 8

![Bar chart showing output weighted SNR for voiced, intermediate, and unvoiced with different harmonics ranges.]
FIG. 9

HARMONICS RANGE

MEAN-REMOVED WEIGHTED SNI

VOICED INTERMEDIATE UNVOICED

9-14
15-19
20-24
25-29
30-35
36-69

FIG. 10

VOICED RANGE

INTERMEDIATE RANGE

UNVOICED RANGE

REw PREDICTOR
SEw PREDICTOR

REw PREDICTOR
SEw PREDICTOR

HARMONICS
REW PARAMETRIC VECTOR QUANTIZATION AND DUAL-PREDICTIVE SEW VECTOR QUANTIZATION FOR WAVEFORM INTERPOLATIVE CODING

CROSS REFERENCE TO RELATED APPLICATION

This application claims the benefit of Provisional Patent Application No. 60/190,371 filed Mar. 17, 2000, which application is herein incorporated by reference. This application is a divisional of U.S. patent application Ser. No. 09/811,187, filed Mar. 16, 2001 now U.S. Pat. No. 7,010,482.

BACKGROUND OF THE INVENTION

The present invention relates to vector quantization (VQ) in speech coding systems using waveform interpolation.

In recent years, there has been increasing interest in achieving toll-quality speech coding at rates of 4 kbps and below. Currently, there is an ongoing 4 kbps standardization effort conducted by an international standards body (The International Telecommunications Union-Telecommunication (ITU-T) Standardization Sector). The expanding variety of emerging applications for speech coding, such as third generation wireless networks and Low Earth Orbit (LEO) systems, is motivating increased research efforts. The speech quality produced by waveform coders such as code-excited linear prediction (CELP) coders degrades rapidly at rates below 5 kbps; see B. S. Atal, and M. R. Schroeder, (1984) “Stochastic Coding of Speech at Very Low Bit Rate”, Proc. Int. Conf. Comm., Amsterdam, pp. 1610-1613.


SUMMARY OF THE INVENTION

The present invention describes novel methods that enhance the performance of the WI coder, and allows for better coding efficiency improving on the above 1999 Gottesman and Gerhso procedure. The present invention incorporates analysis-by-synthesis (AbS) for parameter estimation, offers higher temporal and spectral resolution for the REW, and more efficient quantization of the slowly-evolving waveform (SEW). In particular, the present invention proposes a novel efficient parametric representation of the REW magnitude, an efficient paradigm for AbS predictive VQ of the REW parameter sequence, and dual-predictive AbS quantization of the SEW.

More particularly, the invention provides a method for interpolative coding input signals, the signals decomposed into or composed of a slowly evolving waveform and a rapidly evolving waveform having a magnitude, the method incorporating at least one various, preferably combinations of the following steps or can include all of the steps:

(a) AbS VQ of the REW;
(b) parametrizing the magnitude of the REW;
(c) incorporating temporal weighting in the AbS VQ of the REW;
(d) incorporating spectral weighting in the AbS VQ of the REW;
(e) applying a filter to a vector quantizer codebook in the analysis-by-synthesis vector-quantization of the rapidly evolving waveform whereby to add self correlation to the codebook vectors; and
(f) using a coder in which a plurality of bits therein are allocated to the rapidly evolving waveform magnitude.

In addition, one can combine AbS quantization of the slowly evolving waveform with any or all of the foregoing parameters.

The new method achieves a substantial reduction in the REW bit rate and the EWI achieves very close to toll quality, at least under clean speech conditions. These and other features, aspects, and advantages of the present invention will become better understood with regard to the following detailed description, appended claims, and accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a REW Parametric Representation;
FIG. 2 is a REW Parametric VQ;
FIG. 3 is a REW Parametric Representation AbS VQ;
FIG. 4 is a REW Parametric Representation Simplified AbS VQ;
FIG. 5 is a REW Parametric Representation Simplified Weighted AbS VQ;
FIG. 6 is a block diagram of the Dual Predictive AbS SEW vector quantization;
FIG. 7 is a weighted Signal-to-Noise Ratio (SNR) for Dual Predictive AbS SEW VQ;
FIG. 8 is an output Weighted SNR for the 18 codebooks, 9-bit AbS SEW VQ.
REW Parametric Representation

Direct quantization of the REW magnitude is a variable dimension quantization problem, which may result in spending bits and computational effort on perceptually irrelevant information. A simple and practical way to obtain a reduced, and fixed dimension representation of the REW is with a linear combination of basis functions, such as orthonormal polynomials; see W. B. Kleijn, Y. Shoahm, D. Sen, and R. Haagen, (1996), IEEE ICASSP'96, pp. 212-215; Y. Shoahm, (1997), IEEE ICASSP'97, pp. 1599-1602; Y. Shoahm, (1999), International Journal of Speech Technology, Kluwer Academic Publishers, pp. 329-341. Such a representation usually produces a smoother REW magnitude, and improves the perceptual quality. Suppose the REW magnitude, $R(\omega)$, is represented by a linear combination of orthonormal functions, $\psi_j(\omega)$:

$$R(\omega) = \sum_{j=0}^{L-1} \gamma_j(\omega) \psi_j(\omega), \quad 0 \leq \omega \leq \pi$$

where $\omega$ is the angular frequency, and $I$ is the representation order. The REW magnitude is typically an increasing function of frequency, which, can be coarsely quantized with a low number of bits per waveform without significant perceptual degradation. Therefore, it may be advantageous to represent the REW magnitude in a simple, but perceptually relevant manner. Consequently we model the REW by the following parametric representation, $\tilde{R}(\omega, \xi)$:

$$\tilde{R}(\omega, \xi) = \sum_{j=0}^{L-1} \tilde{\gamma}_j(\xi) \psi_j(\omega), \quad 0 \leq \omega \leq \pi, \quad 0 \leq \xi \leq 1$$

where $\tilde{\gamma}(\xi)=[\tilde{\gamma}_0(\xi), \ldots, \tilde{\gamma}_{L-1}(\xi)]^T$ is a parametric vector of coefficients within the representation model subspace, and $\xi$ is the “voicing” parameter which is zero for a fully voiced spectrum, and one for a fully unvoiced spectrum. Thus $R(\omega, \xi)$ defines a two-dimensional surface whose cross sections for each value of $\xi$ give a particular REW magnitude spectrum, which is defined merely by specifying a scalar parameter value.

A simple and practical way for parametric representation of the REW is, for example, by a parametric linear combination of basis functions, such as polynomials with parametric coefficients, namely:

$$R(\omega, \xi) = \sum_{j=0}^{L-1} \gamma_j(\xi) \psi_j(\omega), \quad 0 \leq \omega \leq \pi, \quad 0 \leq \xi \leq 1$$

For practical considerations assume that the parametric representation is a piecewise linear function of $\xi$, and may therefore be represented by a set of $N$ uniformly spaced spectra, as illustrated in FIG. 1.

REW Parametric Vector Quantization

One can observe the similarity between successive REW magnitude spectra, which may suggest a potential gain by VO of a set of successive REWs. FIG. 2 illustrates a simple parametric VO system for a vector of REW spectra. The input is an $M$ dimensional vector of REW magnitude spectra,

$$R(m)=[R_1(m), R_2(m), \ldots, R_M(m)]^T$$

FIG. 9 is a mean-removed SEW’s Weighted SNR for the 18 codebooks, 9-bit AbS SEW VQ; and FIG. 10 are predictors for three REW parameter ranges.

DETAILED DESCRIPTION


Also, since the sequence of SEW magnitude evolves slowly, successive SEWs exhibit similarity, offering opportunities for redundancy removal. Additional forms of redundancy that may be exploited for coding efficiency are: (a) for a fixed SEW/REW decomposition filter, the mean SEW magnitude increases with the pitch period and (b) the similarity between successive SEWs, also increases with the pitch period. In this work we introduce a novel “dual-predictive” AbS paradigm for quantizing the SEW magnitude that optimally exploits the information about the current quantized REW, the past quantized SEW, and the pitch, in order to predict the current SEW.

Introduction to REW Quantization

The REW represents the rapidly changing unvoiced attribute of speech. Commonly in WI systems, the REW is quantized on a waveform by waveform base. Hence, for low rate WI systems having long frame size, and a large number of waveforms per frame, the relative bitrate required for the REW becomes significantly excessive. For example, consider a potential 2 kbps system which uses a 240 sample frame, 12 waveforms per frame, and which quantizes the SEW by alternating bit allocation of 3 bit and 1 bit per waveform. The REW bitrate is then 24 bit per frame, or 800 kbps which is 40% of the total bitrate. This example demonstrates the need for a more efficient REW quantization.

Efficient REW quantization can benefit from two observations: (1) the REW magnitude is typically an increasing function of the frequency, which suggests that an efficient parametric representation may be used; (2) one can observe a similarity between successive REW magnitude spectra, which may suggest a potential gain by employing predictive VQ on a group of adjacent REWs. The next two sections propose REW parametric representation, and its respective VQ.
and the VQ output is an index, $j$, which determines a quantized parameter vector, $\xi$:

$$\xi = (\xi_1, \xi_2, \ldots, \xi_M)^T$$

which parametrically determines a vector of quantized spectra:

$$\hat{R}(\omega) = \hat{R}(\omega, \xi) = [\hat{R}(\omega, \xi_1), \hat{R}(\omega, \xi_2), \ldots, \hat{R}(\omega, \xi_M)]^T$$

The encoder searches, in the parameter codebook $C_j(\xi)$, for the parameter vector which minimizes the distortion:

$$\hat{\xi} = \arg\min_{\xi \in C_j(\xi)} \sum_{m=1}^{M} \left( \int_{0}^{\pi} |R_m(\omega) - \hat{R}(\omega, \xi)|^2 d\omega \right)$$

For example, suppose the input REW magnitude is represented by an l-th dimensional vector of function coefficients, $\gamma$, given by:

$$\gamma = [\gamma_1, \ldots, \gamma_l]^T$$

For a set of $M$ input REWs, each is of which represented by a vector of polynomial coefficients, $\gamma_m$, which form a $P \times M$ input coefficient matrix, $\Gamma$:

$$\Gamma = [\gamma_1, \gamma_2, \ldots, \gamma_M]^T$$

The inverse VQ output is a vector of $M$ quantized REWs, which form the quantized function coefficient matrix:

$$\hat{\gamma} = [\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_M]^T$$

which is used by the decoder to compute the quantized spectra.

A. Quantization Using Orthonormal Functions


$$R(\omega) = \sum_{n=0}^{N} c_n \psi_n(\omega), 0 \leq \omega \leq \pi$$

which is modeled using the parametric representation:

$$\hat{R}(\omega, \xi) = \sum_{n=0}^{N} c_n \psi_n(\omega), 0 \leq \omega \leq \pi, 0 \leq \xi \leq 1$$

and the VQ parameter is then given by:

$$\hat{\xi} = \arg\min_{\xi \in C_j(\xi)} \left( \int_{0}^{\pi} |R(\omega) - R(\omega, \xi)|^2 d\omega \right)$$

In VQ case, the quantized parameter vector is given by:

$$\hat{\xi} = \arg\min_{\xi \in C_j(\xi)} \left( \sum_{m=1}^{M} \left( |\gamma_m - \hat{\gamma}_m| \right)^2 \right)$$

B. Piecewise Linear Parametric Representation

In order to have a simple representation that is computationally efficient and avoids excessive memory requirements, we model the two dimensional surface by a piecewise linear parametric representation. Therefore, we introduce a set of $N$ uniformly spaced spectra, $\{R(\omega, \xi_n)\}_{n=0}^{N-1}$. Then the parametric surface is defined by linear interpolation according to:

$$\hat{R}(\omega, \xi) = (1 - \alpha) \hat{R}(\omega, \xi_{n-1}) + \alpha \hat{R}(\omega, \xi_n)$$

$$\hat{\xi}_n \leq \xi \leq \hat{\xi}_{n+1}; \alpha = \frac{\xi - \hat{\xi}_{n+1}}{\hat{\xi}_n - \hat{\xi}_{n+1}}$$

Because this representation is linear, the coefficients of $\hat{R}(\omega, \xi)$ are linear combinations of the coefficients of $\hat{R}(\omega, \hat{\xi}_n)$ and $\hat{R}(\omega, \hat{\xi}_{n+1})$ Hence,

$$\hat{\gamma} = (1 - \alpha) \hat{\gamma}_{n+1} + \alpha \hat{\gamma}_n$$

where $\hat{\gamma}_n$ is the coefficient vector of the n-th REW magnitude function representation:

$$\hat{\gamma}_n = [\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_M]^T$$

In this case, the distortion may be interpolated by:

$$D[\hat{R}(\omega, \hat{\xi}'], \hat{R}(\omega, \hat{\xi}')] = \int_{0}^{\pi} \left( |R(\omega) - (1 - \alpha)R(\omega, \hat{\xi}_{n-1}) - \alpha R(\omega, \hat{\xi}_n)|^2 \right) d\omega = \left( |\gamma - (1 - \alpha)\hat{\gamma}_{n-1} - \alpha \hat{\gamma}_n|^2 \right)$$

The above can be easily generalized to the parameter VQ case. The optimal interpolation factor that minimizes the distortion between two representation vectors is given by:

$$\alpha_{opt} = \frac{\|\gamma_\alpha - \gamma_{n+1}\|^2}{\|\gamma_n - \gamma_{n+1}\|^2}$$

and the respective optimal parameter value, which is a continuous variable between zero and one, is given by:

$$\hat{\xi} = (1 - \alpha_{opt}) \hat{\xi}_{n+1} + \alpha_{opt} \hat{\xi}_n$$

This result allows a rapid search for the best unvoicing parameter value needed to transform the coefficient vector to a scalar parameter, followed by the corresponding quantization scheme, as described in the section 4.
C. Weighted Distortion Quantization

Commonly in speech coding, the magnitude is quantized using weighted distortion measure. In this case the quantized Rew parameter is then given by:

\[ \xi = \arg \min_{\xi \in \mathbb{R}} \int_{0}^{\infty} [R(\omega) - R(\omega, \xi^o)]^2 W(\omega) d\omega \]  

(21)

and the orthonormal function simplification, given in equation (13), cannot be used. In this case, the weighted distortion between the input and the parametric representation modeled spectra is equal to:

\[ D_w(R, \tilde{R}|\xi) = \int_{0}^{\infty} [R(\omega) - \tilde{R}(\omega, \xi^o)]^2 W(\omega) d\omega = \|y - y(\xi^o)\|^2 \Phi(W(\omega))(y - y(\xi^o)) \]  

(22)

where \( \Phi(W(\omega)) \) is the weighted correlation matrix of the orthonormal functions, its elements are:

\[ \Phi_{ij}(W(\omega)) = \int_{0}^{\infty} W(\omega) \Phi_i(W(\omega)) \Phi_j(W(\omega)) d\omega, \]  

(23)

\( y \) is the input coefficient vectors, and \( \tilde{y}(\xi) \) is the modeled parametric coefficient vector. In VQ case, the quantized parameter vector is given by:

\[ \hat{\xi} = \arg \min_{\xi \in \mathbb{R}} \sum_{n=1}^{N} D_w(R_n, \tilde{R}(\xi \omega_n)) \]  

(24)

D. Weighted Distortion—Piecewise Linear Parametric Representation

Again, for practical considerations assume that the parametric representation is piecewise linear, and may be represented by a set of \( N \) spectra, \( \{\tilde{R}(\omega, \xi_n)\}_{n=1}^{N} \). For the piecewise linear representation, the interpolated quantized coefficient vector is:

\[ \tilde{y}(\xi) = (1 - \alpha)\tilde{y}_{\xi-1} + \alpha \tilde{y}_{\xi+1} \leq \xi \leq \tilde{y}_{\xi+1}, \alpha = \frac{\xi - \tilde{y}_{\xi-1}}{\tilde{y}_{\xi+1} - \tilde{y}_{\xi-1}} \]  

(25)

In the case where parameter VQ is employed, the interpolation allows for a substantial simplification of the search computations. In this case, the distortion can be interpolated:

\[ D_w(R, \tilde{R}|\xi) = \|y - y(\xi^o)\|^2 \Phi(W(\omega))(y - y(\xi^o)) \]  

(26)

Note that no benefit is obtained here by using orthonormal functions, therefore any function representation may be used. The above can be easily generalized to the parameter VQ case. The optimal parameter that minimizes the spectrally weighted distortion between two representation vectors is given by:

\[ a_{opt} = \frac{(\tilde{y}_n - \tilde{y}_{n-1})^T \Phi_W(y_n - y_{n-1})}{(\tilde{y}_n - \tilde{y}_{n-1})^T \Phi_W(\tilde{y}_n - \tilde{y}_{n-1})} \]  

(27)

and the respective optimal parameter value, which is a continuous variable between zero and one, is given by equation (20). This result allows a rapid search for the best unvoicing parameter value needed to transform the coefficient vector to a scalar parameter, for encoding or for VQ design. Alternatively, in order to eliminate using the matrix \( \phi \), the scalar product may be redefined to incorporate the time-varying spectral weighting. The respective orthonormal basis functions then satisfy:

\[ \int_{0}^{\infty} W(\omega) \phi_i(\omega) \phi_j(\omega) d\omega = \delta(i - j) \]  

(28)

where \( \delta(i - j) \) denotes Kroneker delta. The respective parameter vector is given by:

\[ \gamma = \int_{0}^{\infty} W(\omega) R(\omega) \phi(\omega) d\omega \]  

(29)

where \( \Phi(W(\omega)) = [\Phi_1, \Phi_2, \ldots, \Phi_{\omega}] \) is an \( \omega \)-th dimensional vector of time-varying orthonormal functions.

REW Parameter Analysis-by-Synthesis VQ

This section presents the A\( \text{bS} \) VQ paradigm for the Rew parameter. The first presentation is a system which quantizes the Rew parameter by employing spectral based A\( \text{bS} \). Then simplified systems, which apply A\( \text{bS} \) to the Rew parameter, are presented.

A. Rew Parameter Quantization by Magnitude A\( \text{bS} \) VQ

The novel Analysis-by-Synthesis (A\( \text{bS} \)) Rew parameter VQ technique is illustrated in FIG. 3. An excitation vector \( \tilde{e}(m) (m=1, \ldots, M) \) is selected from the VQ codebook and is fed through a synthesis filter to obtain a parameter vector \( \tilde{\xi}(m) \) (synthesized quantized) which is then mapped to quantized a representation coefficient vectors \( \gamma(\tilde{\xi}(m)) \). This is compared with a sequence of input representation coefficient vectors \( \gamma(m) \) and each is spectrally weighted. Each spectrally weighted error is then temporally weighted, and a distortion measure is obtained. A search through all candidate excitation vectors determines an optimal choice. The synthesis filter in FIG. 3 can be viewed as a first order predictor in a feedback loop. (While shown here is an auto-regressive synthesis filter, in other arrangements moving-average (MA) synthesis filter may be used.) By allowing the value of the predictor parameter \( P \) to change, it becomes a “switched-predictor” scheme. Switched-prediction is introduced to allow for different levels of Rew parameter correlation.

The scheme incorporates both spectral weighting and temporal weighting. The spectral weighting is used for the distortion between each pair of input and the quantized spectra. In order to improve SEW/REW mixing, particularly in mixed voiced and unvoiced speech segments, and to increase speech crispiness, especially for plosives and onsets, temporal
weighting is incorporated in the AbS REW VQ. The temporal weighting is a monotonic function of the temporal gain. Two codebooks are used, and each codebook has an associated predictor coefficient, $P_i$ and $P_2$. The quantization target is an $M$-dimensional vector of REW spectra. Each REW spectrum is represented by a vector of basis function coefficients denoted by $\gamma(m)$. The search for the minimal WME is performed over all the vectors, $\hat{c}(m)$, of the two codebooks for $i=1, 2$. The quantized REW function vector, $\hat{y}(m)$, is a function of the quantized parameter $\hat{z}(m)$, which is obtained by passing the quantized vector, $\hat{c}(m)$, through the synthesis filter. The weighted distortion between each pair of input and quantized REW spectra is calculated. The total distortion is a temporally-weighted sum of the M spectrally weighted distortions. Since the predictor coefficients are known, direct VQ can be used to simplify the computations. For a piecewise linear parametric representation, a substantial simplification of the search computations may be obtained by interpolating the distortion between the representation spectra set, as explained in sections 3.1, 3.2, and 3.3.

A sequence of quantized parameter, such as $\hat{c}(k)$, is formed by concatenating successive quantized vectors, such as $[\hat{c}(m)]_{m=1}^M$. The quantized parameter is computed recursively by:

$$\hat{z}(k) = P_k \hat{z}(k-1) + \hat{c}(k)$$  \hspace{1cm} (30)

where $k$ is the time index of the coded waveform.

B. Simplified REW Parameter AbS VQ

The above scheme maps each quantized parameter to coefficient vector, which is used to compute the spectral distortion. To reduce complexity, such mapping, and spectral distortion computation, which contribute to the complexity of the scheme, may be eliminated by using the simplified scheme described below. For a high rate, and a smooth representation surface $\hat{R}(\omega, \hat{z})$, the total distortion is equal to the sum of modeling distortion and quantization distortion:

$$\sum_{m=1}^M D_m(R(m), \hat{R}(\hat{z}(m))) = \sum_{m=1}^M D_m[R(m), \hat{R}(\hat{z}(m))]$$

(31)

The quantization distortion is related to the quantized parameter by:

$$\sum_{m=1}^M D_m[R(m), \hat{R}(\hat{z}(m))] = \sum_{m=1}^M [\hat{z}(m) - \hat{y}(m)]^T \Psi[w(m)](\hat{z}(m) - \hat{y}(m))$$

(32)

which, for the piecewise linear representation case, is equal to

$$\sum_{m=1}^M D_m[R(m), \hat{R}(\hat{z}(m))] = \frac{1}{\Delta^2} \sum_{m=1}^M [\hat{z}(m) - \hat{y}(m)]^T \Psi[w(m)](\hat{z}(m) - \hat{y}(m))$$

(33)

which is linearly related to the REW parameter squared quantization error, $(\hat{z}(m) - \hat{y}(m))^2$, and, therefore, justifies direct VQ of the REW parameter.

B.1. Simplified REW Parameter AbS VQ—Non Weighted Distortion

FIG. 4 illustrates a simplified AbS VQ for the REW parametric representation. The encoder maps the REW magnitude to an unvoicing REW parameter, and then quantizes the parameter by AbS VQ. Initially, the magnitudes of the M REWs in the frame are mapped to coefficient vectors, $[\hat{y}(m)]_{m=1}^M$. Then, for each coefficient vector, a search is performed to find the optimal representation parameter, $\hat{z}(\gamma)$, using equation (20), to form an $M$-dimensional parameter vector for the current frame, $[\hat{z}(\gamma(m))]_{m=1}^M$. Finally, the parameter vector is encoded by AbS VQ. The decoded spectra, $[\hat{R}(w, \hat{z}(m))]_{m=1}^M$, are obtained from the quantized parameter vector, $[\hat{z}(m)]_{m=1}^M$, using equation (15). This scheme allows for higher temporal, as well as spectral REW resolution, compared to the common method described in W. Klein, et al., IEEE ICASSP '95, pp. 508-511 (1995), since no downsampling is performed, and the continuous parameter is vector quantized in AbS.

B.2. Simplified REW Parameter AbS VQ—Weighted Distortion

The simplified quantization scheme is improved to incorporate spectral and temporal weightings, as illustrated in FIG. 5. The REW parameter vector is first mapped to REW parameter by minimizing a distortion, which is weighted by the coefficient spectral weighting matrix $\Psi$, as described in section 3.1. Then, the resulted REW parameter is used to compute a weighting, $w_i(\hat{z}(m))$, which we choose to be the spectral sensitivity to the REW parameter squared quantization error, $(\hat{z}(m) - \hat{y}(m))^2$, given by:

$$w_i(\hat{z}(m)) = \frac{\partial^2 \Psi}{\partial \hat{z}} (\hat{z})$$

(35)

For the piecewise linear representation case, using equation (33), the following equation is obtained:

$$w_i(\hat{z}(m)) = \frac{\partial^2 \Psi}{\partial \hat{z}} (\hat{z})$$

(35)

$$\frac{1}{\Delta^2} \sum_{m=1}^M [\hat{z}(m) - \hat{y}(m)]^T \Psi[w(m)](\hat{z}(m) - \hat{y}(m))$$

(35)

The above derivative can be easily computed off line. Additionally, a temporal weighting, in form of monotonic function of the gain, denoted by $w_i(g(m))$, is used to give relatively large weight to waveforms with larger gain values. The AbS REW parameter quantization is computed by minimizing the combined spectrally and temporally weighted distortion:

$$D_i[\hat{z}(m)]_{m=1}^M = \sum_{m=1}^M w_i(g(m)) [\hat{z}(m) - \hat{y}(m)]^T$$

(36)

The weighted distortion scheme improves the reconstructed speech quality, most notably in mixed voiced and unvoiced speech segments. This may be explained by an improvement in REW/SEW mixing.
Dual Predictive AbS SEW Quantization

FIG. 6 illustrates a Dual Predictive SEW AbS VQ scheme which uses two observables, (a) the quantized REW, and (b) the past quantized SEW, to jointly predict the current SEW. Although we refer to the operator on each observable as a "predictor", in fact both are components of a single optimized estimator. The SEW and the REW are complex random vectors, and their sum is a residual vector having elements whose magnitudes have a mean value of unity. In low bit-rate WI coding, the relation between the SEW and the REW magnitudes was approximated by computing the magnitude of one as the unity complement of the other. Suppose \( \mathbf{f}_{\omega} \) denotes the spectral magnitude vector of the last quantized REW in the current frame. An "implied" SEW vector, is calculated by:

\[
\mathbf{f}_{\omega}^{\text{imp}} = 1 - \mathbf{f}_{\omega}
\]


Subjective Results

A subjective A/B test was conducted to compare the 2.8 kbps EWI coder of this invention to G.723.1. The test data included 24 modified intermediate reference system (M-IRS) filtered speech sentences, 12 of which are of female speakers, and 12 of male speakers; see ITU-T, (1996), "Recommendation P.830, Subjective Performance Assessment of Telephone Band and Wideband Digital Coders", Annex D, ITU, Geneva. Twelve listeners participated in the test. The test results, listed in Table 2 and Table 3, indicate that the subjective quality of the 2.8 kbps EWI exceeds that of G.723.1 at 5.3 kbps, and it is slightly better than that of G.723.1 at 6.3 kbps. The EWI preference is higher for male than for female speakers.

Table 2 shows the results of subjective A/B test for comparison between the 2.8 kbps EWI coder to 5.3 kbps G.723.1. With 95% certainty the result lies within +/-5.53%.
Table 3 shows the results of subjective A/B test for comparison between the 2.8 kbps EWI coder to 6.3 kbps G.723.1. With 95% certainty the result lies within ±4.5.59%. It should, of course, be noted that while the present invention has been described in terms of an illustrative embodiment, other arrangements will be apparent to those of ordinary skills in the art. For example:

While in the disclosed embodiment in FIG. 3 have described auto-regressive (AR) synthesis filter, in other arrangements moving-average (MA) filter may be used.

While in the disclosed embodiment was related to waveform interpolative speech coding, in other arrangements it may be used in other coding schemes.

While in the disclosed embodiment temporal weighting, and/or spectral weighting are described, they are optional, and in other arrangements any or both of them may not be used.

While in the disclosed embodiment switch prediction having two predictors is described, in other arrangements no switch, or more than two predictor choice may be used.

While in the disclosed embodiment illustrated in FIG. 6 mean vectors are subtracted from the vector, this may be viewed as optional, and in other arrangements any or all of such mean vectors may not be used.

While in the disclosed embodiment the pitch range and/or the voicing parameter values were partitioned into subranges, and codebooks were used for each subrange, this may be viewed as optional, and in other arrangements any or all of such subranges may not be used, or other number or type of subranges may be used.

While in the disclosed embodiment describes prediction matrices were diagonal, in other arrangements non diagonal prediction matrices may be used.

The following references are each incorporated herein by reference:


The invention claimed is:

1. A computerized system for coding a digital input signal representative of an audio waveform, comprising:
   a vector quantizer module comprising an input accepting a vector or rapidly evolving waveform spectra determined from the digital input signal;
   a computer readable codebook comprising a plurality of arrays of quantized parametric vectors;
   computer readable program instructions for searching said parameter codebook for one of said plurality of arrays of quantized parametric vectors which minimizes distortion;
   and
   a computer program instructions for outputting a computer readable indicator that points to the one of said plurality of arrays of quantized parameter vectors which minimizes distortion.

2. The system of claim 1 wherein the indicator determines a quantized vector of the rapidly evolving waveform.

3. The system of claim 2 further comprising a slowly evolving waveform vector quantizer, said slowly evolving waveform quantizer in communication with the vector quantizer for inputting the quantized vector of the rapidly evolving waveform, said slowly evolving waveform quantizer comprising instructions for determining an implied slowly evolving waveform as a function of the quantized rapidly evolving waveform.
4. The system of claim 1 further comprising a synthesis filter coupled to the output of the codebook wherein said synthesis filter outputs a synthesized parameter vector.

5. The system of claim 1 further comprising computer readable instructions for spectrally weighting a plurality of coefficient vectors mapped from synthesized parameter vectors.

6. The system of claim 1 wherein the audio waveform comprises speech.

7. The system of claim 6, wherein said speech comprises mixed voiced and unvoiced speech segments.

8. The system of claim 1 further comprising a converter that samples an analog input signal to create the digital input signal.

9. The system of claim 1, wherein the system outputs a coded representation of the digital input signal and the computer readable indicator is embedded in that coded representation.

10. A computer implemented method for coding a digital input signal representative of an audio waveform, comprising:
    determining a vector of rapidly evolving waveform spectra from the digital input signal;
    searching a computer readable codebook comprising a plurality of arrays of quantized parametric vector for one of said plurality of arrays of quantized parameter vectors which minimizes distortion; and
    outputting a computer readable indicator that points to the one of said plurality of arrays of quantized parameter vectors which minimizes said distortion.

11. The method of claim 10 wherein the indicator determines a quantized vector of the rapidly evolving waveform.

12. The method of claim 10 further comprising synthesis filtering the output of the codebook and outputting a synthesized parameter vector.

13. The method of claim 10 further comprising spectrally weighting a plurality of coefficient vectors mapped from synthesized parameter vectors.

14. The method of claim 10 further comprising determining an implied slowly evolving waveform as a function of the quantized rapidly evolving waveform.

15. The method of claim 10 wherein the digital input signal comprises speech.

16. The method of claim 15, wherein said speech comprises mixed voiced and unvoiced speech segments.

17. The method of claim 10 further comprising sampling an analog input signal to create said digital input signal.

18. The method of claim 10, wherein the method outputs a coded representation of the digital input signal and the computer readable indicator is embedded in that coded representation.
On the Title Page:

The first or sole Notice should read --

Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 124 days.

Signed and Sealed this
Fourteenth Day of September, 2010

David J. Kappos
Director of the United States Patent and Trademark Office