

PAPER

Weighted vector quantization of harmonic spectral magnitudes for very low-bit-rate speech coding

Masayuki Nishiguchi

Sony Corporation,

6-7-35 Kitashinagawa, Shinagawa-ku, Tokyo, 141-0001 Japan

(Received 10 May 2005, Accepted for publication 19 July 2005)

Abstract: Harmonic coding is a very powerful technique for the coding of speech at very low bit rates; and the efficient coding of spectral magnitudes sampled at harmonic frequencies is the key to obtaining good coded-speech quality. This paper presents a weighted vector quantization method for spectral vectors composed of a variable number of harmonic magnitudes. It is based on simple, efficient linear dimension conversion and employs a weighted distortion measure that exploits the human auditory sense. A codebook training algorithm using the weighting matrix is also presented. Finally, a low-complexity VQ codebook search technique based on pre-selection is described that reduces the computational complexity to less than 10% of that of an exhaustive search, without perceptible loss of quality. The proposed quantization scheme is used in Harmonic Vector eXcitation Coding (HVXC), which is a very low-bit-rate speech coding algorithm that combines harmonic and stochastic vector representations of LPC residual signals. Due to the high efficiency of this VQ scheme, HVXC provides good communication-quality speech at bit rates as low as 2–4 kbit/s, and was adopted as the ISO/IEC International Standard for MPEG-4 Audio.

Keywords: Harmonic coding, VXC, HVXC, CELP, Vector quantization, MPEG-4

PACS number: 43.72.Ar, 43.72.Gy [DOI: 10.1250/ast.27.43]

1. INTRODUCTION

Over the past decades, a great deal of effort has been directed at developing medium- to low-bit-rate speech coding algorithms. ADPCM [1] at 32 kbit/s and Code Excited Linear Prediction (CELP) [2], which is also known as Vector eXcitation Coding (VXC) [3], at around 4–16 kbit/s are used for many of the digital mobile phone and fixed telephone network standards. However, with CELP the quality of coded speech is poor at bit rates below 4 kbit/s because of the nature of waveform matching, which requires phase reconstruction.

On the other hand, for these low bit rates, a class of sinusoidal coding, including Harmonic Coding [4], Sinusoidal Transform Coding (STC) [5], and MultiBand Excitation (MBE) [6], provides good quality, mainly due to the smooth reconstruction of voiced signals. MBE adds band-pass noise for voiced excitations, and STC uses phase randomization for sinusoidal excitation to improve the quality of synthesized speech. To obtain natural quality, however, these coders have to use both phase information and spectral magnitudes, which results in higher bit rates of around 6–8 kbit/s.

To further lower the bit rate while ensuring natural quality, coding methods have been devised that efficiently combine Linear Predictive Coding (LPC) and sinusoidal coding, such as Mixed Excited Linear Prediction (MELP) [7], Waveform Interpolation (WI) [8], and Harmonic Vector eXcitation Coding (HVXC) [9]. These coders employ a parametric representation of LPC residual signals, and yield good speech quality at bit rates below 4 kbit/s. MELP mixes random noise with a periodic excitation signal to generate mixed voiced and unvoiced signals. The excitation signal in WI consists of two groups: one that changes rapidly and one that changes slowly in the time sequence of the power spectrum. HVXC combines the harmonic coding of the spectral magnitudes of LPC residual signals for voiced segments and the VXC algorithm for unvoiced segments. This combination provides good speech quality at bit rates of 2 and 4 kbit/s. In addition, 2-kbit/s decoding is possible in HVXC not only using a 2-kbit/s bitstream, but also using a subset of a 4-kbit/s bitstream.

The harmonic coding of an LPC residual signal is a natural and effective way of discarding the phase of residuals, and thus lowering the bit rate, without causing a discontinuity. To obtain good coded-speech quality with a

harmonic coder, it is crucial that an efficient quantizer for harmonic magnitudes be designed. This paper describes the variable-dimension weighted vector quantization of harmonic magnitudes using a multistage scheme, which provides high-performance bit-rate-scalable coding in HVXC.

This paper is organized as follows: Section 2 presents an overview of the encoder and decoder systems; Section 3 describes the vector quantization scheme for harmonic spectral magnitudes; and Section 4 explains codebook training and a fast search algorithm.

2. OVERVIEW OF ENCODER AND DECODER SCHEMES

2.1. Encoder

By way of background, an overview of the structure of a speech coder based on the proposed quantizer scheme is presented. Figure 1 shows the structure of the HVXC encoder. Speech input at a sampling rate of 8 kHz is formed into frames with a length of 256 samples and an interval of 160 samples. An LPC analysis of one frame is carried out using windowed input data. LPC residual signals are computed by inverse filtering the input data using quantized and interpolated LSP parameters. The residual signals are then fed into the pitch-and-spectral-magnitude estimation

block, which estimates the harmonic spectral envelope of the LPC residual signal as follows: A base spectrum representing the spectrum for one harmonic is gain-scaled and arranged with a spacing equal to the fundamental frequency, which is obtained from an open-loop pitch search. The gain scaling for each harmonic of the fundamental frequency and the fundamental frequency itself are adjusted simultaneously to minimize the difference between the synthesized power spectrum and the actual LPC residual spectrum. After the harmonic magnitudes and fundamental frequency are adjusted, the number of harmonic magnitudes is converted to a fixed number by a dimension converter in order to enable quantization of the set of harmonic magnitudes (= harmonic spectral envelope) with a fixed-dimension vector quantizer [10]. The harmonic spectral envelope for a voiced segment is then vector quantized with a weighted distortion measure. The quantization of harmonic spectral magnitudes is described in Section 3. For unvoiced segments, a closed-loop search for VXC is carried out.

2.2. Decoder

Figure 2 shows the structure of the HVXC decoder. The basic decoding process consists of four steps: the dequantization of parameters, the generation of excitation

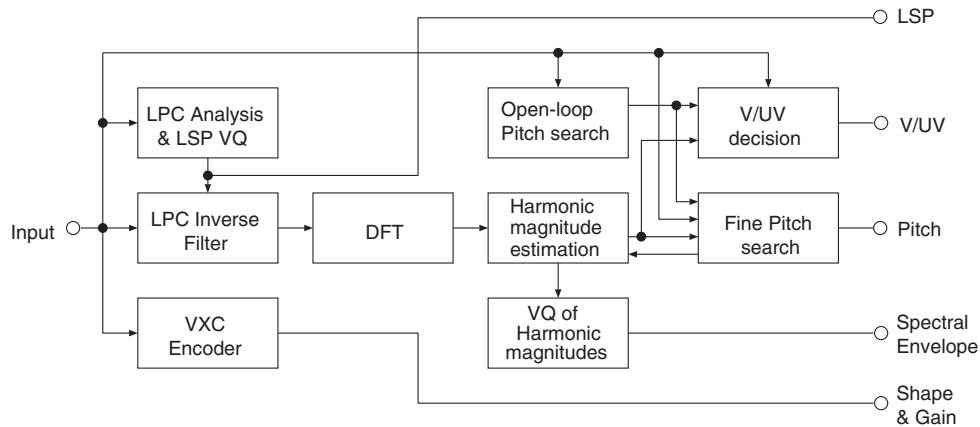


Fig. 1 Overall structure of the encoder.

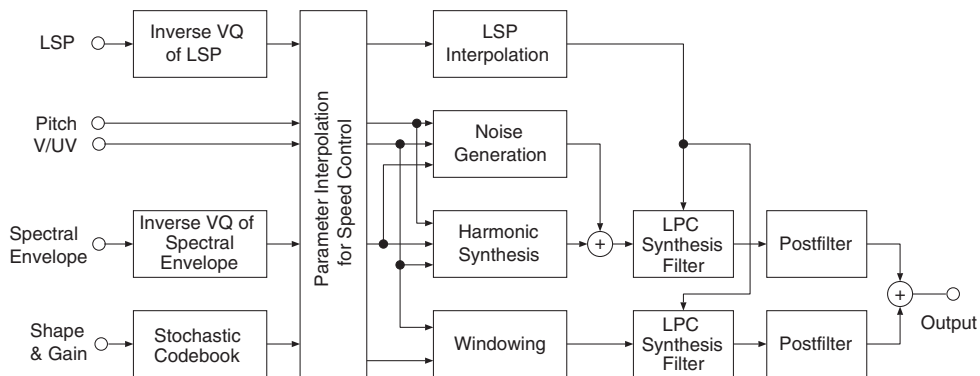


Fig. 2 Overall structure of the decoder.

signals for voiced frames by sinusoidal synthesis (harmonic synthesis) and the addition of a noise component, the generation of excitation signals for unvoiced frames by code-book look-up, and LPC synthesis. Spectral post-filtering enhances the quality of the synthesized speech.

For voiced frames, a fixed-dimension harmonic spectral vector obtained by dequantization of the spectral magnitude is first converted to one with the original dimension, which varies from frame to frame in accordance with the pitch. This is done by the dimension converter, which uses a band-limited interpolator to generate a set of spectral magnitudes at harmonic frequencies. The dimension converter in the decoder has the same structure as the one in the encoder (see Section 3.2). Using the spectral magnitude values at harmonic frequencies, a time domain excitation signal is then generated by the fast harmonic synthesis algorithm using an IFFT [11]. In order to make the synthesized speech sound natural, a Gaussian-noise spectral component covering a frequency range of around 2–3.8 kHz is also used. It is colored in accordance with the harmonic spectral magnitudes in the frequency domain, and its IDFT is added to voiced excitation signals in the time domain. The amount and bandwidth of the added noise is determined by a transmitted two-bit V/UV flag, which indicates one mode out of four; Unvoiced, Mixed voiced-1, Mixed voiced-2, and Full voiced modes. Voiced segments are declared to be one of either Mixed voiced-1, Mixed voiced-2, or Full voiced mode in the encoder according to the magnitude of the normalized maximum autocorrelation of the LPC residual signal. The addition of a noise component is done when Mixed voiced-1, Mixed voiced-2, or Full voiced mode is selected, and the scaling factors for the noise component are pre-determined for each of the modes. Harmonic excitation signals with added noise for voiced segments are fed into the LPC synthesis filter and then the postfilter.

For unvoiced segments, the usual VXC decoding algorithm is used, which multiplies the gain by the shape code-vector to generate an excitation signal. The result is fed into the LPC synthesis filter and then the postfilter. Finally, the synthesized speech components for voiced and unvoiced segments are combined to form the output signal.

3. VECTOR QUANTIZATION OF HARMONIC SPECTRAL MAGNITUDES

3.1. Framework

This section describes an algorithm for the harmonic coding of spectral magnitudes, which is the key technology in harmonic coding, such as HVXC. The basic coding process has three main steps:

- dimension conversion of the harmonic spectral magnitude vector,
- generation of the perceptual weighting matrix, and

- vector quantization of the fixed-dimension harmonic spectral magnitude vector.

Extended operation in the higher-bit-rate mode requires two more steps for quantization in the enhancement layer:

- dimension conversion of the quantized harmonic spectral magnitude vector back to the original dimension, and
- vector quantization of the difference between the original and quantized harmonic magnitudes.

Let us start with the dimension-conversion algorithm.

3.2. Dimension Converter

3.2.1. Interpolation

The number of points constituting the spectral envelope varies depending on the pitch, since the spectral envelope is a set of estimated magnitudes for each harmonic. The number of harmonics ranges from about 9 to 70. In order to vector quantize the spectral envelope with a fixed-dimension VQ, the coder has to convert the variable number to a fixed number. Meuse [12] showed that band-limited interpolation yields better results than 1st-order linear interpolation in the conversion of the sampling frequency to obtain fixed-dimension spectral vectors. The reason is that the number of points representing the shape of the harmonic spectral envelope needs to be modified without changing the shape (Fig. 3). Since band-limited dimension conversion using FFT or IFFT [12] suffers from a relatively high computational complexity, we devised a dimension-conversion method that combines an FIR low-pass filter and a 1st-order linear interpolator (Fig. 4) [10]. First, the

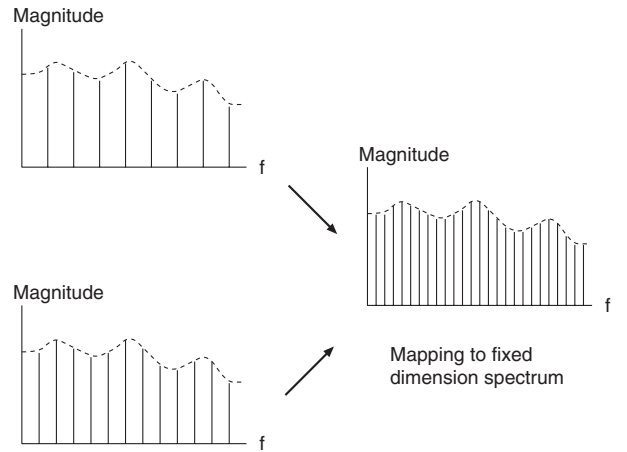


Fig. 3 Dimension conversion of spectral envelope.

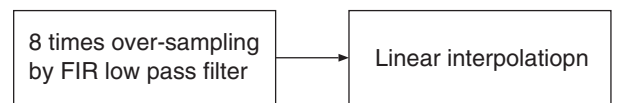


Fig. 4 Dimension conversion using an oversampling filter in combination with linear interpolation.

FIR low-pass filter for 8-times oversampling is designed as follows: The coefficients for the 65th-order oversampling filter $coef[i]$ are obtained from a sinc ($\sin x/x$) multiplied by a raised cosine window:

$$coef[i] = \frac{\sin \pi(i-32)/8}{\pi(i-32)/8} (0.5 - 0.5 \cos 2\pi i/64) \quad (0 \leq i \leq 64) \quad (1)$$

The interpolation can be efficiently implemented using polyphase filters [13]. For each offset, k , with respect to the original sampling grid, the coefficients of the corresponding polyphase filter of length 8 are:

$$c_k[j] = coef[8 \times j + k] \quad (1 \leq k \leq 7, \quad 0 \leq j \leq 7) \quad (2)$$

In the second stage (Fig. 4), 1st-order linear interpolation is applied to the 8-times over-sampled data to compute spectral magnitudes at the fixed harmonic frequencies $n \times \omega_0$ ($1 \leq n \leq N$), where ω_0 is the fundamental frequency or frequency spacing of the components in the fixed-dimension vector and N is the number of harmonics that comprise the fixed-dimension spectral vector. 1st-stage oversampling FIR filtering allows for decimated computation, in which only the points used in the 2nd-stage linear interpolation are computed; specifically, these are the points directly next to the frequencies $n \times \omega_0$ ($1 \leq n \leq N$). The number of multiplies and adds required for FIR low-pass filtering for 8-times oversampling is $N(\text{number of outputs}) \times 2(\text{directly surrounding points}) \times 8(\text{filter order})$, which is less than 1/40 of that needed in the direct FFT-IFFT method [12]. In this way, a spectral vector of fixed dimension N is obtained.

3.2.2. Processing of ends of a spectral vector

The 1st oversampling stage, which has an FIR low-pass filter, employs 4 data points each on the left and right sides of the target point. To compute the oversampling points at frequencies lower than the original lowest harmonic frequency and those at frequencies higher than the original highest harmonic frequency, appropriate data must be appended. In order to minimize the overshoot of the transient response of the oversampling filter, the values of these data have to be carefully determined. Since it is assumed that a high-pass filter with a cut-off frequency of around 120Hz is used in the pre-processing of speech, harmonic magnitudes are suppressed towards zero as the frequency decreases. So, we append zeros for the lower-frequency data. For the higher frequency data, we repeat the harmonic magnitude at the highest harmonic frequency for the higher-frequency data.

3.2.3. Performance of the proposed dimension conversion method

To evaluate the performance of the dimension converter, we measured the segmental SNR between speech segments synthesized with and without the dimension

converter. With the dimension converter, the original harmonics are dimension converted to a fixed dimension, and immediately recovered to the original dimension. In neither case was spectral quantization used. The average segmental SNR over 8 typical speech segments was about 30dB; and in a listening test, the two results were indistinguishable.

3.3. Generation of Perceptual Weighting Matrix

Now we have a fixed-dimension harmonic spectral vector to be quantized. Prior to harmonic magnitude quantization, the transfer functions of the perceptual weighting filter, $W(z)$ [14], and LPC synthesis filter, $H(z)$, are defined so that the frequency responses of $W(z)$ and $H(z)$ can be used for weighted vector quantization of the harmonic spectral envelope. $H(z)$ and $W(z)$ are defined as follows:

$$H(z) = \frac{1}{1 + \sum_{i=1}^P \alpha_i z^{-i}} \quad (3)$$

$$W(z) = \frac{1 + \sum_{i=1}^P \alpha_i \lambda_b^i z^{-i}}{1 + \sum_{i=1}^P \alpha_i \lambda_a^i z^{-i}}, \quad (4)$$

where α_i ($1 \leq i \leq P$) are the LPC coefficients of the current frame. It should be noted that the LPC coefficients of the speech frame being processed have already been obtained. λ_a and λ_b are constants that determine the shape of the quantization noise, and are usually given the values 0.5 and 0.8, respectively. Sampling the magnitudes $|H(e^{j\omega})|$ and $|W(e^{j\omega})|$ at the harmonic frequency points $\omega = n \times \omega_0$ yields the diagonal components of the weighting matrices \mathbf{H} and \mathbf{W} , which are used in Section 3.4.

3.4. Vector Quantization

A fixed-dimension vector is then quantized using the weighting matrices \mathbf{H} and \mathbf{W} in accordance with the scheme illustrated in Fig. 5. Table 1 lists the dimensions and sizes of the VQ codebook. For coding at a bit rate of 2 kbit/s, to reduce memory requirements and search complexity while achieving high performance, a two-stage vector quantization scheme is employed for the spectral shape together with a scalar quantizer for the gain. The weighted distortion measure, D , is used throughout the design and search of both the shape and gain codebooks:

$$D = \left\| \frac{1}{\|\mathbf{x}\|} \mathbf{W} \mathbf{H} (\mathbf{x} - g(s_0 + s_1)) \right\|^2, \quad (5)$$

where \mathbf{x} is the set of dimension converted harmonic spectral magnitudes with a fixed vector dimension (44); s_0 is the output of the spectral envelope (SE) Shape-0

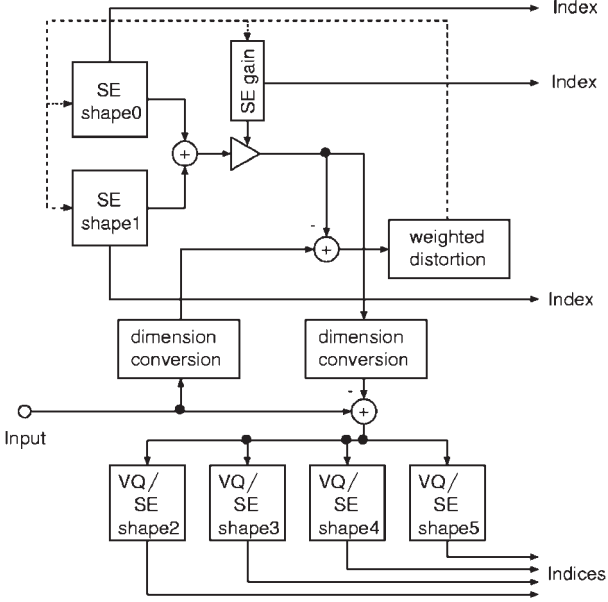


Fig. 5 Vector quantization scheme for harmonic spectral magnitudes.

Table 1 Codebook size and dimension.

Bit rate (kbit/s)	Codebook	Dim.	Number of bits
2.0/4.0	SE Shape-0	44	5
2.0/4.0	SE Shape-1	44	5
2.0/4.0	SE gain	1	5
4.0	SE Shape-2	2	7
4.0	SE Shape-3	4	10
4.0	SE Shape-4	4	9
4.0	SE Shape-5	4	6

codebook; s_1 is the output of the SE Shape-1 codebook; and g is the output of the SE gain codebook. H and W are the weighting matrices derived in Section 3.3. Codebook design [11] and a fast search technique for a two-stage VQ [15] are discussed in Section 4.

For the 4-kbit/s mode, an additional quantizer stage is used. In this stage, quantized harmonic magnitudes from the 2-kbit/s mode with a fixed dimension (44) are first converted to the dimension of the original harmonics by the band-limited interpolation method described in Section 3.2. The difference between the original harmonics and these harmonics is computed; and the resulting quantization error vector, which has the original dimension, is then quantized by the additional vector quantizers. In this case, however, only the components corresponding to the lower 14 harmonics are quantized using split VQ. The codebooks used in this stage are SE Shape-2, SE Shape-3, SE Shape-4, and SE Shape-5, as shown in Fig. 5. The codebook search of split VQ employs the same weights as those used in the first two-stage VQ. This multistage structure produces scal-

able bitstreams because the quantizer output of the 2-kbit/s mode is always used, regardless of whether or not the vector quantizers in the additional stage for the 4-kbit/s mode are used.

4. CODEBOOK TRAINING AND CODEBOOK SEARCH

4.1. Codebook Training

Using the distortion measure (Eq. (5)), the codebooks are trained by an optimal MSVQ design [16,17] method based on the generalized Lloyd algorithm (GLA) [18]. Summing the distortion from all the frames that use the j -th shape vector, $s_{k,j}$, in SE Shape- k ($k = 0, 1$), taking the derivative of the total distortion with respect to $s_{k,j}$, and setting the result to zero, we obtain the new centroid, $u_{k,j}$, for $s_{k,j}$:

$$u_{k,j} = \frac{\sum_{i \in j} g_i A_i^T A_i (x_i - g_i s_{1-k,i})}{\sum_{i \in j} g_i^2 A_i^T A_i} \quad (6)$$

where

$$A_i = \frac{W_i H_i}{\|x_i\|}; \quad (7)$$

the suffix i denotes the frame number that selects $s_{k,j}$ from SE Shape- k ; x_i , W_i , and H_i are the source vector and associated weights; and g_i and $s_{1-k,i}$ are the codewords selected in frame i . Similarly, the new centroid, v_j , for the j -th gain codeword is given by

$$v_j = \frac{\sum_{i \in j} x_i^T A_i^T A_i (s_{0,i} + s_{1,i})}{\sum_{i \in j} \|A_i (s_{0,i} + s_{1,i})\|^2}, \quad (8)$$

where the suffix i denotes the frame number that selects the j -th gain codeword.

4.2. Codebook Search

An efficient codebook search algorithm was developed for spectral vector quantization [15]. The usual sequential search process is first described, and the improved algorithm is then explained. Let x_n be the input vector to be quantized and W_n , H_n be the associated weights.

Step1: Find a set of shape vectors, $s_{0,i}$ and $s_{1,j}$, that maximize

$$CC_{ij} = \frac{(x_n^T A_n^T A_n (s_{0,i} + s_{1,j}))^2}{\|A_n (s_{0,i} + s_{1,j})\|^2}, \quad (9)$$

where

$$A_n = \frac{W_n H_n}{\|x_n\|} \quad (10)$$

Step2: Find the gain codeword g_1 in the gain codebook that is closest to

$$g_{\text{opt}} = \frac{\mathbf{x}_n^T \mathbf{A}_n^T \mathbf{A}_n (s_{0,i} + s_{1,j})}{\|\mathbf{A}_n (s_{0,i} + s_{1,j})\|^2} \quad (11)$$

Suppose that the two shape codebooks and the gain codebook all have a size of 5 bits, and that the vectors \mathbf{x} , s_0 , and s_1 have a dimension of 44. Then, the computational speed required to carry out the multiplies and adds in *Step1* is approximately 9 MOPS, which is very high.

So, we developed a pre-selection algorithm to reduce the amount of computation in *Step1* by generalizing the pre-selection algorithm in CELP coding [19] to a codebook search for the frequency domain. Since g_1 is designed to have only positive values, we can maximize the square root of CC_{ij} instead of CC_{ij} to do *Step1*. Now, we can say that

$$\frac{\mathbf{x}_n^T \mathbf{A}_n^T \mathbf{A}_n s_{0,i} + \mathbf{x}_n^T \mathbf{A}_n^T \mathbf{A}_n s_{1,j}}{\|\mathbf{A}_n s_{0,i}\| + \|\mathbf{A}_n s_{1,j}\|} \leq \frac{\mathbf{x}_n^T \mathbf{A}_n^T \mathbf{A}_n (s_{0,i} + s_{1,j})}{\|\mathbf{A}_n (s_{0,i} + s_{1,j})\|} \quad (12)$$

If we search for the optimal $s_{0,i}$ and $s_{1,j}$ independently, ignoring the correlation between the two shape vectors, the left side of the above inequality can be maximized by finding the $s_{0,i}$ and $s_{1,j}$ that maximize C_{0i} and C_{1j} , which are given by

$$C_{0i} = \frac{\mathbf{x}_n^T \mathbf{A}_n^T \mathbf{A}_n s_{0,i}}{\|\mathbf{A}_n s_{0,i}\|} \quad (13)$$

$$C_{1j} = \frac{\mathbf{x}_n^T \mathbf{A}_n^T \mathbf{A}_n s_{1,j}}{\|\mathbf{A}_n s_{1,j}\|}. \quad (14)$$

Code-vectors are pre-selected from the two shape codebooks by taking those several $s_{0,i}$ and $s_{1,j}$ that best maximize C_{0i} and C_{1j} , respectively. The value of the weight \mathbf{A}_n changes from frame to frame as a function of the input vector \mathbf{x}_n . Once we compute the value of $\mathbf{x}_n^T \mathbf{A}_n^T \mathbf{A}_n$, the numerators of C_{0i} and C_{1j} are obtained simply by computing the inner products. The denominator is obtained by calculating the weighted norm of each of the shape code-vectors using the value of \mathbf{A}_n for every frame. However, this step is pre-selection by approximation. So, we used a fixed value for the typical weight \mathbf{A}_n^* . For all the code-vectors, the inverse of the weighted norm,

$$L_{0,i} = \frac{1}{\|\mathbf{A}_n^* s_{0,i}\|}, \quad L_{1,j} = \frac{1}{\|\mathbf{A}_n^* s_{1,j}\|}, \quad (15)$$

is computed off-line and stored with the code-vectors in the codebooks. Then, the approximations of $C_{0,i}$ and $C_{1,j}$ are obtained by computing one inner product per code-vector entry. In this coding scheme, spectral quantization is only used for voiced segments. This allows us to use the typical weight \mathbf{A}^* , for which most of the spectral energy is concentrated in the low-frequency region. \mathbf{A}^* is computed from a long training sequence:

$$\mathbf{A}^* = \frac{1}{M} \sum_{m=1}^M \mathbf{A}_m, \quad (16)$$

where \mathbf{A}_m are weights computed for frames for which the voiced/unvoiced decision is voiced.

In this manner, P_s code-vectors are pre-selected in *Step1* from each of the shape codebooks with the size of 5 bits. The final selection is conducted by finding the $s_{0,i}$ and $s_{1,j}$ that maximize the CC_{ij} from the $P_s \times P_s$ combinations of the pre-selected shape code-vectors. For $P_s = 10$, this pre-selection procedure reduces the computational complexity of the shape codebook search to less than 1/10 that for an exhaustive search, and requires a processing speed of only about 0.85 MOPS.

The segmental SNR between speech synthesized using the pre-selection algorithm ($P_s = 10$) and the reference speech synthesized using an exhaustive search was measured. For 17 pairs of typical sentences used in MPEG-4 core experiments, the average segmental SNR is about 23 dB. Six trained listeners compared the same 17 pairs of sentences in paired comparison tests, and no difference in quality was detected between speech synthesized with and without pre-selection.

5. TEST RESULTS AND CONCLUSIONS

A low-complexity high-performance vector quantization scheme for coding harmonic spectral magnitudes was developed. This scheme made it possible to build a very low-bit-rate speech coder, the HVXC, which provides communication-quality speech at a bit rate of 2 kbit/s.

In the MPEG-4 standardization process, official subjective tests were conducted to evaluate the candidate coders [20]. 11 Japanese items and 6 English or German items were evaluated by 35 Japanese, 8 American, and 10 German listeners. The Japanese listeners evaluated only the Japanese items, and American and German listeners evaluated both the English and German items. ACR tests with a 5 point grading scale were used. Table 2 lists the items used in the tests. Table 3 shows the mean opinion score (MOS) for the HVXC and FS1016. FS1016 4.8-kbit/s CELP was used as a reference coder. The MOS scores are the average for all the items, including background noise, background music, and two-speaker items. The MOS for

Table 2 Speech files used in tests.

	Japanese	English/German
Male voice	4	2
Female voice	4	2
Two speakers	1	1
Background noise	1	1
Background music	1	—
Total	11	6

Table 3 MOS for 2-kbit/s HVXC.

Codec	Japanese	English/German	Total
HVXC@2.0 kbit/s	2.44	2.56	2.46
FS1016@4.8 kbit/s	2.29	2.44	2.32

the HVXC was the best among all the proposed coders, and was better than that for the FS1016 CELP coder.

After core experiments were carried out as part of the standardization process, HVXC was chosen to be the ISO/IEC International Standard for MPEG-4 Audio [21].

REFERENCES

- [1] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video* (Prentice Hall Inc., Englewood Cliffs, N.J., 1984).
- [2] M. R. Schroeder and B. S. Atal, "Code-excited linear predictive (CELP): High-quality speech at very low bit rates," *Proc. ICASSP85*, pp. 937–940 (1985).
- [3] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression* (Kluwer Academic Publishers, Boston, 1992).
- [4] J. S. Marques, L. B. Almeida and J. M. Tribolet, "Harmonic coding at 4.8 kb/s," *Proc. ICASSP90*, pp. 1–17–20 (1990).
- [5] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust. Speech Signal Process.*, **34**, 744–754 (1986).
- [6] D. W. Griffin and J. S. Lim, "Multiband excitation vocoder," *IEEE Trans. Acoust. Speech Signal Process.*, **36**, 1223–1235 (1988).
- [7] A. V. McCree and T. P. Barnwell III, "A mixed excitation LPC vocoder model for low bit rate speech coding," *IEEE Trans. Speech Audio Process.*, **3**, 242–250 (1995).
- [8] W. B. Kleijn and J. Haagen, "A speech coder based on decomposition of characteristic waveforms," *Proc. ICASSP95*, pp. 1–508–511 (1995).
- [9] M. Nishiguchi, K. Iijima and J. Matsumoto, "Harmonic vector excitation coding of speech at 2.0 kbps," *IEEE Workshop on Speech Coding* (1997).
- [10] M. Nishiguchi, J. Matsumoto, S. Ono and R. Wakatsuki, "Vector quantized MBE with simplified V/UV division at 3.0 kbps," *Proc. ICASSP93*, pp. II-151–154 (1993).
- [11] M. Nishiguchi and J. Matsumoto, "Harmonic and noise coding of LPC residuals with classified vector quantization," *Proc. ICASSP95*, pp. 1–484–487 (1995).
- [12] P. C. Meuse, "A 2400 bps multi-band excitation vocoder," *Proc. ICASSP90*, pp. 9–12 (1990).
- [13] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing* (Prentice Hall Inc., Englewood Cliffs, 1983).
- [14] S. Miki, T. Moriya, K. Mano and H. Ohmuro, "Basic algorithm of pitch synchronous innovation CELP(PSI-CELP) speech coding," *NTT R&D*, **43**, 363–372 (1994).
- [15] M. Nishiguchi, K. Iijima and J. Matsumoto, "Low bit-rate speech coding by harmonic vector excitation coding," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, 1–2–4 (1997).
- [16] T. Moriya, "Two-channel conjugate vector quantizer for noisy channel speech coding," *IEEE JSAC*, **10**, 866–874 (1992).
- [17] W.-Y. Chan, *Product Code Vector Quantization Methods with Application to High Fidelity Audio Coding*, PhD Thesis, University of California at Santa Barbara (1991).
- [18] Y. Linde, A. Buzo and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, **COM-28**, 84–95 (1980).
- [19] T. Moriya, A. Kataoka and S. Hayashi, "Complexity reduction of codebook search in CELP coding," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, 2–6–1 (1993).
- [20] "MPEG-4 Audio Test Results (MOS Tests)," Audio Sub-group ISO/IEC JTC1/SC29/WG11 N1144, Jan. (1996).
- [21] ISO/IEC 14496-3:1999, "Coding of Audiovisual Objects - Part 3: Audio," Version 1, December (1999).



Masayuki Nishiguchi received his B.E. degree in Physical Electronics from the Tokyo Institute of Technology in 1981 and his M.S. degree in Electrical and Computer Engineering from the University of California, Santa Barbara in 1989. Since 1981, he has been with the Sony corporation, where he is engaged in the development of speech and audio coding, and signal processing algorithms. He is currently a general manager of the Audio Codec Development Department of Sony. He is a member of IEEE.