Nonsquare Transform Vector Quantization

Peter Lupini and Vladimir Cuperman

Abstract—**Several techniques for low-rate speech coding have emerged, requiring quantization of the spectral magnitudes. The set of spectral magnitudes may be considered as a variable-dimension vector with dimension dependent on the pitch period. In this letter, we present a technique called nonsquare transform vector quantization (NSTVQ). This technique addresses the problem of variable-dimension vector quantization by combining a fixed-dimension vector quantizer with a variable-sized nonsquare transform. The experimental results presented show that NSTVQ outperforms existing harmonic magnitude quantization techniques.**

I. Introduction

In recent years, several techniques for speech coding at rates of 4 kb/s and lower have emerged, requiring quantization of spectral magnitudes at a set of frequencies that are harmonics of the fundamental pitch period of the talker [1]–[3]. Because the pitch period is time varying, the number of components to be quantized changes from frame to frame, causing a variable-dimension vector quantization problem. In an attempt to solve this problem, the IMBE codec [1] uses a complicated encoding scheme with variable-bit assignments and hybrid scalar/vector quantization. Recently, a technique called variable dimension vector quantization (VDVQ) [4] has been proposed. VDVQ has been shown to perform better than the IMBE quantization scheme and all-pole modeling.

In this letter, we present a quantization technique called nonsquare transform vector quantization (NSTVQ). This technique addresses the problem of variable-dimension vector quantization by combining a fixed-dimension vector quantizer with a variable-sized nonsquare transform. Experimental results comparing NSTVQ to all-pole modeling, VDVQ, and IMBE magnitude quantization are provided.

II. The Nonsquare Transform

Let $\vec{y}$ be a vector of length N where N is variable. For example, if $\vec{y}$ is a vector of harmonic magnitudes for a frame of speech, then N depends on the pitch period for that frame. Given a known $N \times M$ matrix $\boldsymbol{A}$ (to be specified below), we want to find a fixed-length M-dimensional vector $\vec{z}$ that can be used to compute an estimate of $\vec{y}$ using the transformation $\vec{y}_m = \boldsymbol{A}\vec{z}$. For any given $\boldsymbol{A}$, our goal is to minimize the mean squared error distortion criterion D_m with respect to $\vec{z}$ where $D_m(\vec{y}, \vec{y}_m) = 1/N||\vec{y}_m - \vec{y}||^2$. It can be shown that the vector $\vec{z}_{opt}$ that minimizes $D_m(\vec{y}, \vec{y}_m)$ is obtained as the solution to the set of linear equations

$$(\boldsymbol{A}^T\boldsymbol{A})\vec{z}_{\text{opt}} = \boldsymbol{A}^T\vec{y}. \tag{1}$$

A solution to this equation can always be found (regardless of the rank of $\boldsymbol{A}$) using one of the linear algebra techniques for inverting ill-conditioned matrices, for example, singular value decomposition (SVD). If $N \geq M$ and the M columns of $\boldsymbol{A}$ are linearly independent, the matrix $\boldsymbol{A}^T\boldsymbol{A}$ is of full rank and, therefore, has an explicit inverse that gives a unique solution. Furthermore, if $N \geq M$ and the columns of $\boldsymbol{A}$ are orthonormal, the optimal solution vector is simply $\vec{z}_{\text{opt}} = \boldsymbol{A}^T\vec{y}$. For the case of $N < M$, (1) is underdetermined and therefore has no unique solution. It was found experimentally that a zero-padded solution works well when combined with vector quantization. The zero-padded solution is obtained by using N orthonormal vectors for the first N columns of $\boldsymbol{A}$ and setting the last $M - N$ columns to zeros. Using orthonormal columns and zero-padding, the general solution for $\vec{z}_{opt}$ can be written as

$$\vec{z}_{opt} = \boldsymbol{A_p}^T\vec{y}. \tag{2}$$

$\boldsymbol{A_p}$ is defined as

$$\boldsymbol{A_p} = \begin{cases} (\vec{a}_1\,\vec{a}_2\,\cdots\,\vec{a}_M) & \text{if } N \geq M \\ (\vec{a}_1\,\vec{a}_2\,\cdots\,\vec{a}_N|O) & \text{if } N < M \end{cases} \tag{3}$$

where $\vec{a}_i$ are orthonormal column vectors, and O is an $N \times (M - N)$ all-zero matrix. For our tests, we chose the columns of $\boldsymbol{A_p}$ using variable-length basis functions derived from the discrete cosine transform (DCT). In this case, the elements $a_i[n]$ for $n = 1 \cdots N$ are given by

$$a_i[n] = \left(\frac{2}{N}\right)^{1/2} C_i \cos\left(\frac{[2(n-1)+1]\pi(i-1)}{2N}\right) \tag{4}$$

where $C_i = 1$ when $i \neq 1$, and $C_i = 1/\sqrt{2}$ when $i = 1$.

III. Vector Quantizer Design

The nonsquare transformation derived in Section II transforms a variable-length vector $\vec{y}$ into $\vec{z}$ which can be encoded using a fixed-dimension VQ. The quantized fixed-length vector $\vec{z}_q$ is then transformed into the quantized variable-length vector $\vec{y}_q$ using $\vec{y}_q = \boldsymbol{A_p}\vec{z}_q$. The vector quantizer should be designed to minimize the distortion $D_t(\vec{y}, \vec{y}_q) = 1/N||\vec{y} - \vec{y}_q||^2$. It can be shown that

$$D_t(\vec{y}, \vec{y}_q) = \frac{1}{N}||\boldsymbol{A_p}\vec{z} - \vec{y}||^2 + \frac{1}{N}||\boldsymbol{A_p}(\vec{z}_q - \vec{z})||^2. \tag{5}$$

The first term of (5) is the modeling distortion due to the nonsquare transform, and the second term is the quantization

Manuscript received December 1, 1994. The associate editor coordinating the review of this paper and approving it for publication was Dr. B. S. Atal.

The authors are with the School of Engineering Science, Simon Fraser University, Burnaby V5A 1S6, Canada.

Publisher Item Identifier S 1070-9908(95)01172-8.

distortion due to the VQ. The fact that these distortions can be separated shows that, once we have chosen an orthonormal transformation matrix $\boldsymbol{A_p}$, we need not consider it during training. The distortion measure for vector quantizer error minimization is given by

$$D_q(\vec{z}, \vec{z}_q) = \frac{1}{N} \sum_{i=1}^{\min(M,N)} (z[i] - z_q[i])^2 \qquad (6)$$

where $z[i]$ is the ith element of the vector $\vec{z}$.

We trained our vector quantizers using the generalized Lloyd algorithm (GLA). A training set of size L consists of fixed-length vectors $\vec{z}_l$ and corresponding vector lengths N_l, where $l = 1 \cdots L$. Given an initial codebook of size K with entries $\vec{c}_k$, $k = 1 \cdots K$, we encode the training set by assigning each vector $\vec{z}_l$ to partition $\mathcal{S}^i$ if $D_q(\vec{z}_l, \vec{c}_i)$ is minimum over all codebook entries. The centroid rule for computing the new kth codebook entry $\vec{c}'_k$ is given by

$$c'_k[n] = \frac{\sum_{l \in \mathcal{S}^k} \frac{1}{N_l} z_l[n] p_l[n]}{\sum_{l \in \mathcal{S}^k} \frac{1}{N_l} p_l[n]} \quad \text{for } n = 1 \cdots M \qquad (7)$$

where $z_l[n]$ is the nth element of the lth training vector. $p_l[n]$ are the components of a vector that eliminates zero-padded elements from the distortion calculation and are defined as

$$p_l[n] = \begin{cases} 1 & \text{if } n \le \min(N, M) \\ 0 & \text{otherwise.} \end{cases} \qquad (8)$$

An extensive discussion on the training of vector quantizers can be found in [5].

IV. Experimental Results

In order to evaluate the NSTVQ method for spectral magnitude quantization, we compared the objective performance of NSTVQ with three other methods: all-pole modeling [6]; the combination scalar/vector quantization scheme of IMBE [1]; and the direct VQ approach of VDVQ [4]. For all methods, the spectral log-magnitudes to be quantized and the associated pitch periods were obtained exactly as specified in the IMBE standard. A set of 40 000 vectors was used for training the quantizers, and a set of 12 000 vectors outside the training set was used for evaluation. For the comparison with IMBE, we implemented a version of NSTVQ with predictive vector quantization (NSTPVQ), which uses vector prediction to exploit intervector correlations. Prediction is made simpler with NSTVQ because the quantized vectors are of fixed length. Other methods that use vector prediction, including IMBE, must use interpolation prior to prediction.

The fixed vector length for NSTVQ was chosen by comparing the performance using values of M ranging from 10 to 60. As M increases, the modeling distortion decreases but the quantization distortion increases. It was found that, depending on the number of bits available to encode each spectrum, increasing M beyond 30 or 40 resulted in no performance improvement. Unless otherwise stated, $M = 40$ was used. Because the complexity increases with M, there may be applications that would benefit by trading performance for lower M.

TABLE I
Spectral Distortion (in Decibels) for LPC-10, VDVQ, and NSTVQ (30 B/Spectrum)

METHOD	FEMALE	MALE	BOTH
LPC-10	4.42	5.05	4.73
VDVQ	2.94	3.58	3.25
NSTVQ	2.86	3.46	3.17

The distortion criterion used to evaluate performance is the root mean square spectral distortion (SD). The spectral distortion between the unquantized and quantized harmonic magnitude vectors $\vec{m}$ and $\vec{m}_q$ is given by

$$D = \sqrt{\frac{1}{i_2 - i_1} \sum_{n=i_1}^{i_2-1} \left[10 \log_{10} \left(\frac{|m[n]|^2}{|m_q[n]|^2} \right) \right]^2} \qquad (9)$$

where i_1 and i_2 are chosen such that only harmonics within the frequency range of interest are included in the distortion calculation.

We first compared NSTVQ to an all-pole modeling technique (LPC-10) similar to the algorithm used in [6] and to VDVQ [4]. We quantized the LPC-10 model coefficients using a 24-b multistage VQ (MSVQ) for the 10 LSP values and a 6-b scalar gain quantizer. VDVQ uses two 10-b mean-removed VQ's to encode harmonics lying in the range of 64–1500 and 1500–3600 Hz respectively, and another 10-b VQ to encode the means (actually the log-gains). Splitting the spectrum in this way may improve subjective quality by using more bits to encode the lower frequency harmonics, but objective performance is often reduced. Because of this, we kept the comparison with VDVQ fair by using exactly the same spectral splitting and mean-removal, with NSTVQ (M= 14) applied to each half-spectrum separately. Table I shows the results of this comparison. For both male and female speakers, the NSTVQ system outperformed the LPC-10 system by approximately 1.6 dB. The improvement with respect to VDVQ is small (approximately 0.1 dB) and may be affected by the choice of the training and test databases (the databases were used for all systems, however the training requirements may be different). NSTVQ shows an advantage over VDVQ in terms of complexity and codebook storage requirements. For this configuration, the complexity of the NSTVQ system was estimated to be almost half that of the VDVQ system, even when the worst-case complexity was assumed for the NST transforms. It is expected that the use of fast-transform algorithms for the DCT can further reduce the NSTVQ complexity. The codebook storage requirement for NSTVQ was about one-quarter that of the VDVQ algorithm.

The next test compared IMBE scalar/vector quantization with NSTVQ. For this test, we applied NSTVQ to the entire spectral range of 64–3600 Hz without splitting. We used a 6-b per stage MSVQ structure with M-L search [7]. By training 11 stages and dropping the stages sequentially, we were able to obtain results for systems ranging from 6–66 b/spectrum. The same structure was used for our predictive

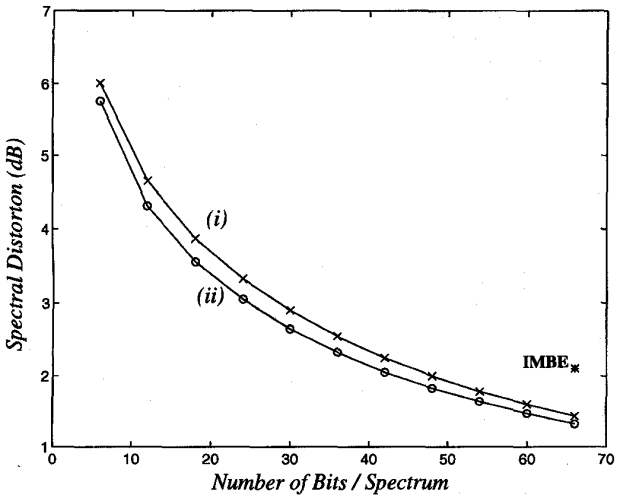

Fig. 1. Spectral distortion versus number of bits per spectrum for (i) NSTVQ and (ii) NSTPVQ. IMBE spectral magnitude quantization using an average of 66 bits/spectrum is indicated with an asterisk.

system, NSTPVQ. IMBE uses vector prediction and a variable-bit assignment scheme, which obtained an average rate of 66 b/spectrum on our test data. Fig. 1 shows that performance equivalent to 66-b IMBE can be obtained using 46-b NSTVQ, or 41-b NSTPVQ. An IMBE codec with NSTVQ magnitude quantization could save 20–25 b/frame, or 1000–1250 b/s. Furthermore, the NSTVQ system shows a smooth drop in performance as the number of bits per spectrum is reduced.

We have incorporated NSTVQ in a 2.4 kb/s harmonic speech coder and found that the subjective performance is consistent with the objective results presented here. The 2.4 kb/s coder with NSTVQ scored 3.2 in an informal MOS test compared with a score 3.4 for the 4.15 kb/s IMBE codec and 1.8 for the 2.4 kb/s LPC-10e standard.

V. Conclusion

Motivated by the problems encountered when encoding variable-length vectors such as harmonic magnitudes, we have introduced a quantization technique called NSTVQ that uses a variable-size nonsquare transform combined with a fixed-dimension vector quantizer. The technique is shown experimentally to outperform all-pole modeling. Performance comparable to VDVQ is achieved with lower complexity and storage requirements. NSTVQ is also shown to provide significant bit-reduction potential when compared to IMBE magnitude quantization. Furthermore, the performance of NSTVQ is shown to degrade gracefully as the bit rate is reduced, making it a good candidate for very low bit-rate systems.

References

[1] Digital Voice Systems, *Inmarsat-M Voice Codec, Version 2.* Inmarsat, Feb. 1991.

[2] R. McAulay, T. Parks, T. Quatieri, and M. Sabin, "Sine-wave amplitude coding at low data rates," in *Proc. IEEE Workshop Speech Coding Telecommun.,* Vancouver, Canada, 1989.

[3] Y. Shoham, "High-quality speech coding at 2.4 to 4.0 kbps based on time-frequency interpolation," in *Proc. ICASSP,* Minneapolis, MN, 1993.

[4] A. Das, A. V. Rao, and A. Gersho, "Variable-dimension vector quantization of speech spectra for low-rate vocoders," in *Proc. Data Compression Conf.,* 1994, pp. 421–429.

[5] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression.* Boston, MA: Kluwer, 1992.

[6] M. S. Brandstein, "A 1.5 Kbps multi-band excitation speech coder," Dept. Elect. Eng. Comput. Sci., Mass. Inst. of Technol., Cambridge, USA, 1990.

[7] B. Bhattacharya, W. LeBlanc, S. Mahmoud, and V. Cuperman, "Tree searched multi-stage vector quantization of LPC parameters for 4 kb/s speech coding," in *Proc. ICASSP,* 1992, pp. 105–108.