COMPOSITE PERMUTATION CODING WITH SIMPLE INDEXING FOR SPEECH/AUDIO CODECS

Shinya Abe, Kei Kikuiri, and Nobuhiko Naka

Research Laboratories, NTT DoCoMo, Inc.
3-5 Hikari-no-oka, Yokosuka, Kanagawa, 239-8536 Japan

ABSTRACT

This paper proposes a vector quantization (VQ) method based on composite permutation coding for transform audio coding. VQ is widely used for audio data compression. It requires mean square error computation or a similar metric for finding the nearest neighbor in the codebook, which generally incurs a lot of operations. To reduce such operations, we focus on the permutation representation and easy indexing of vectors in the codebook. The proposal consists of constrained composite permutation codes, which are distinguished by the number of components quantized into each quantization level. This scheme makes the output bit stream take the same form as a parallel array of scalar quantization (SQ). Simulation results show that the proposal almost matches the performance of VQ at 2 bit/scalar bit-rates with lower computational complexity. Its structure yields the efficient representation of tones that are important for auditory perception.

Index Terms— Permutation codes, Vector quantization, Transform coding, Speech coding, Audio coding

1. INTRODUCTION

Internet Protocol (IP) telephony is a technique that transmits speech/audio signals between all platforms connected via an IP network. It makes replacement of the codec easy, and the codec design can be flexible. However, at the same time, complexity of codecs becomes important especially for low-power devices.

Broadband channels of IP networks also make it possible to consider higher rate and higher quality audio codecs than conventional speech codecs such as G.711 [1] and G.729A [2]. For instance, G.722.1 [3] is a coding for 7 kHz bandwidth audio signals at 24 kbps or 32 kbps. To represent not only speech but also music and ambient sounds, many audio coders [3]-[5] are based on transform coding instead of CELP coding [2][6]. We consider the basic transform coding algorithm widely used, that is,

1) Transform the input signal to a spectrum coefficient vector, e.g. Modified Discrete Cosine Transform (MDCT).
2) Decompose the spectrum coefficient vector into sub-vectors.
3) Quantize and encode spectrum envelop information and flattened sub-vectors separately.

To achieve high quality at low bit-rates with this type of codec, we must consider the characteristics of audio signals. Most sounds in nature have spectral peaks (harmonics), and human ears are very sensitive to tones. Thus it is effective to preserve these dominant components over the sub-vector.

To efficiently code the sub-vectors at low bit-rates, VQ [7] is a reasonable method. An optimal VQ codebook can be created when the distribution of target vectors is known. Therefore, many speech/audio codecs use the VQ technique for quantization of spectrum coefficients and other parameters [2][5].

The larger the dimension of the vector and/or the greater the number of codebook entries, the more complex VQ becomes to calculate the distance between the input vector and all candidate vectors in the codebook. To reduce the computational complexity of VQ, various algorithms have been proposed [8]-[10]. Composite permutation coding [11][12] is one of the alternatives to VQ. This scheme reduces the distance calculations by omitting the permutation; composite codevectors can represent various shapes.

We propose composite permutation coding with low complexity indexing. It is derived using two constraints. First, each codevector has the same number of quantization levels. Second, the codevectors are distinguished by the number of components quantized into each level.

This paper is organized as follows. In Section 2 we describe composite permutation coding. In Section 3 we implement the proposal for a 2 bit/scalar spectrum vector quantizer in a transform audio encoder. A performance comparison of the proposal and other quantizers is shown in Section 4. The paper concludes with discussion in Section 5.

2. COMPOSITE PREMUTATION CODING

Permutation coding is a block coding system of which there are two types. Variant I has a codebook of N-dimensional codevectors $y_1, y_2, \ldots, y_M$,

$$y_1 = (\underbrace{\mu_1,...,\mu_1}_{n_1}, \underbrace{\mu_2,...,\mu_2}_{n_2},..., \underbrace{\mu_K,...,\mu_K}_{n_K}) \quad (1)$$

$$\sum_{i=1}^{K} n_i = N \quad (2)$$

$$\mu_1 > \mu_2 > ... > \mu_K \quad (3)$$

where μ_i is the i-th quantization level and n_i is the number of components quantized to μ_i. The optimum values of μ_i and n_i are determined uniquely by the probability density function (PDF) of input vectors. y_2, y_3,…, y_M are generated by reordering the components of y_i. K is the number of the quantization levels which depends on the codebook size,

$$M = N!/\prod_{i=1}^{K} n_i! \quad (4)$$

where $M \le 2^R$ at R bits.

The n_i largest components of an input vector are replaced by μ_i one after another.

Variant II is the same as Variant I except that Variant II quantizes the sign information separately. This means that all μ_i are non-negative values, and sign bits are allocated for non-zero components. In Variant II, the codebook design is restricted, however, fewer operations are needed for error calculations than Variant I.

Permutation codes are suitable for large length random sources whose PDF is known. Unfortunately, they cannot well quantize short vectors since they have only one combination of quantization levels and so cannot handle the shape instability of the short vectors.

To solve this problem, the composite permutation coding [12] has multiple basic permutation codes such as y_i.

3. IMPLEMENTING THE PROPOSAL

This section describes the proposal, its implementation and a practical codebook optimization method.

Composite permutation coding usually incurs high computational complexity for indexing. We propose composite permutation coding with a low complexity indexing algorithm using integer bit/scalar quantization.

At B bit/scalar (B must be positive integer), the proposal sets the constraints. One is that K in (2) is determined by $K = 2^B$ for all basic permutation codevectors. The other is that the quantization levels of each basic permutation codevector are uniquely identified by $\{n_1, n_2,..., n_K\}$. These constraints enable each component of a selected codevector to be encoded in the same way as SQ, and eliminate the need for complex permutation indexing. Thus the proposal can represent various shapes with low indexing cost.

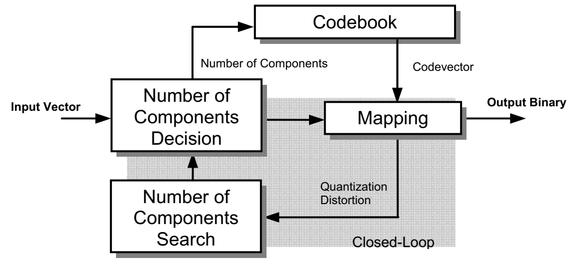

Figure 1. Encoder block diagram with the proposal

The block diagram of the encoder with the proposal is illustrated in Figure 1. First, it decides the number of components quantized into each level, i.e. the selection $\{n_1, n_2,..., n_K\}$. Second, a codevector is selected according to this selection. Next, locate the optimum codevector using the same approach as used by composite permutation coding.

We implement the proposal for the case of 2 bit/scalar quantization. According to [13], the marginal PDF of scalars in sub-vectors of the transform audio coding is very close to that of a Laplace distribution. To simplify the quantization, this paper assumes that the PDF of the input vectors is symmetric, and encode the sign information separately. In other words, 1bit/scalar is assigned for sign encoding and 1bit/scalar is for the permutation codes shape as in Variant II.

It is generally difficult to optimize the codebook because we need to find the optimum codebook and the optimum $\{n_1, n_2,..., n_K\}$ at the same time. To reduce the valid space for finding the optimum codes, the proposal first normalizes the target vector and then encodes it. In this case, equation (2) can be rewritten as

$$\sum_{i=1}^{K} n_i \mu_i^2 = N . \quad (5)$$

For codebook optimization, we use an exhaustive search algorithm that finds the best relation between μ_i and n_i.

To describe the relation in simple terms, we focus on the variance of codevector using

$$\frac{1}{N}\sum_{i=1}^{K} n_i \mu_i^2 - \left(\frac{1}{N}\sum_{i=1}^{K} n_i \mu_i\right)^2 = s^2 . \quad (6)$$

In the case of $K = 2$, the quantization levels μ_1 and μ_2 are obtained from (5) and (6) as follows.

$$\mu_1 = \sqrt{1-s^2} + \sqrt{\left(1-s^2\right)\left(1-\frac{N}{n_1}\right) + \frac{n_2}{n_1}} \quad (7)$$

$$\mu_2 = \frac{1}{n_2}\left(N\sqrt{1-s^2} - n_1\mu_1\right) \quad (8)$$

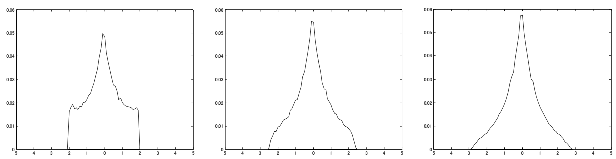

Figure 2. PDF of normalized Laplace distributed noise (N=4,6,8)

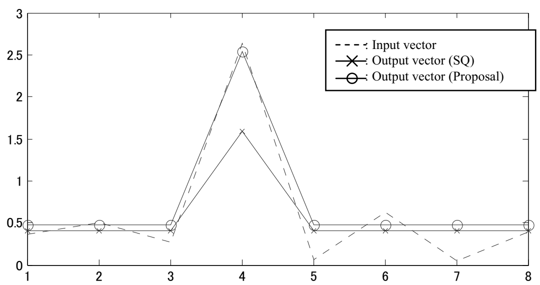

Figure 3. Quantization results for peaky input vector

Here s^2 changes according to n_1, so codebook optimization is equivalent to finding the combination of s^2 for $n_1 = 1, 2,..., N$ that minimizes the quantization distortion. Once the optimum set of quantization levels is obtained, the codebook can be made from the basic permutation codevectors, $\{\underbrace{\mu_{n_1 1},...,\mu_{n_1 1}}_{n_1},\underbrace{\mu_{n_1 2},...,\mu_{n_1 2}}_{n_2}\}_{n_1=1}^{N}$.

4. PERFORMANCE COMPARISON

4.1. Laplace distributed noise simulation

To clarify the performance of our proposal, we evaluated the quantization distortion and operating time of SQ, VQ, and the proposal. VQ performance is better than any composite permutation coding Variant II implementation. Note that VQ also quantizes the sign information separately. The input vector was normalized Laplace distributed noise. The dimensions of input vector N were 4, 6 and 8. Figure 2 shows the PDFs of the input vector.

The proposal found the optimum codebook by 2-stage resolution search to increase speed. At the first stage, resolution was 0.05 on s^2, and in the second it was 0.01.

Table 1 lists the optimum codebook generated by the proposal for an 8-dimensional vector. The quantization results by SQ (cross) and the proposal (circle) are shown in Figure 3. The broken line is an input vector that has a peak. In this case, an output vector of SQ is smoothed while the proposal preserves the peak. This shows that the proposal is suitable for representing tones that are important for perception.

Table 2 shows the SNR of normalized Laplace distribution noise. We can see that the proposal achieves near VQ performance. Note that the PDF difference in Figure 2 leads the lower SNR for the higher dimension.

Table 1. Optimum codebook of the proposal (N=8)

n_1	variance	μ_1	μ_2
1	0.47	2.542	0.469
2	0.40	1.870	0.409
3	0.34	1.565	0.361
4	0.27	1.374	0.335
5	0.19	1.238	0.337
6	0.14	1.143	0.279
7	0.07	1.064	0.264
8	-	1.000	1.000

Table 2. SNR of normalized Laplace distributed noise

	SNR (dB)		
	4-dim	6-dim	8-dim
SQ	11.44	10.36	9.80
VQ	14.22	12.76	11.97
Proposal	13.81	12.06	11.18

Table 3. The number of operations and required codebook size

dim		Add.	Multi.	Comp.	codebook size
4	SQ	8	8	4	2
	VQ	80	64	15	64
	Proposal	20	16	15	8
6	SQ	12	12	6	2
	VQ	448	384	63	384
	Proposal	42	36	35	12
8	SQ	16	16	8	2
	VQ	2304	2048	255	2048
	Proposal	72	64	63	16

To compare the computational complexity, we measure the number of operations (additions, multiplications and comparisons) and the required codebook size. The proposal performs sort processing, which can influence the number of comparisons. The evaluation result of the proposal in Table 3 is for the worst case.

4.2. Application to transform audio codec

We applied the proposal to a 64 kbps transform audio coder with 32 kHz sampling frequnecy. The algorithm is based on [4] but applies 256-point MDCT once a frame and the resulted transform coefficients below 4 kHz and the rest are decomposed into four- and eight-dimensional vectors, respectively.

Figure 4 shows the PDFs of normalized MDCT coefficients of speech and music signals. We can confirm that they are similar to those of Laplace distributed noise as [13] pointed out. Then we built codebooks of SQ, VQ and the proposal optimized for this distribution. The quantization results when 2 bit/scalar allocated for all the coefficients are summarized in Table 4. It was confirmed that similar results can be obtained for real speech and audio signals.

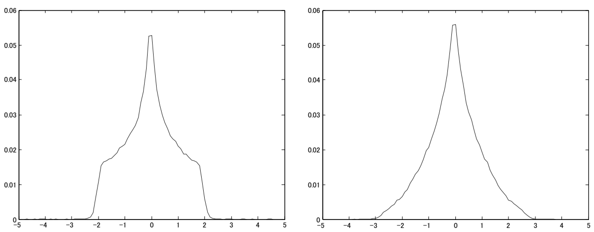

Figure 4. PDF of normalized MDCT coefficients (N=4,8)

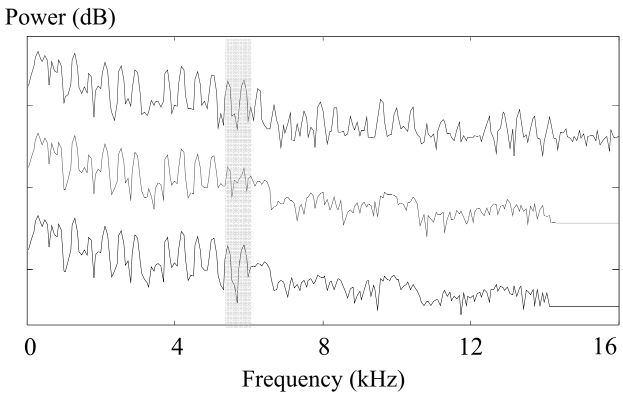

Figure 5. Spectra of jazz music
(Top: Original, Middle: SQ, Bottom: Proposal)

Table 4. SNR of normalized MDCT coefficients

	SNR (dB)	
	4-dim	8-dim
SQ	11.18	9.76
VQ	13.95	12.04
Proposal	13.31	11.17

Figure 5 shows spectra of the compressed music signal. The middle one is the result in the case quantizing MDCT coefficients by SQ, and the bottom one is that by SQ and the proposal. Note that in Figure 5, the proposal is effective only for the normalized sub-vector to be quantized at 2 bit/scalar. The proposal can preserve the peaks of harmonics better than SQ (shaded band). By an expert listening test, we confirmed that the proposal reduces the harsh noise in comparison with SQ.

5. CONCLUSION

This paper proposes a new vector quantization method for transform audio coding. It is based on composite permutation coding with a constraint on the number of quantization levels. SQ-like encoding allows the proposal to represent various shape vectors with low complexity. Simulations show that proposal achieves near VQ performance with lower computational complexity. We intend to implement and clarify the proposal's performance at more than 3 bit/scalar quantization.

ACKNOWLEDGMENT

We would like to thank Dr. S. A. Ramprashad for his advice and information on permutation codes.

REFERENCES

[1] ITU-T Recommendation G.711, "Pulse code modulation (PCM) of voice frequencies"

[2] ITU-T Recommendation G.729 Annex A, "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP), Reduced complexity 8 kbit/s CS-ACELP speech codec"

[3] ITU-T Recommendation G.722.1, "Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss"

[4] J-H. Chen, "A high fidelity speech and audio codec with low delay and low complexity," *ICASSP2000*, vol.2, pp. 1161-1164, Istanbul, Turkey, June 2000.

[5] T. Moriya and M. Honda, "Transform coding of speech using a weighted vector quantizer," *IEEE, JSAC-6*, pp.425-431, 1988.

[6] M.R. Schroeder and B.S. Atal, "Code-excited linear prediction (CELP): high-quality speech at very low bit rates," *Proc. ICASSP 85*, pp.937-940, 1985.

[7] R.M. Gray, "Vector quantization," *IEEE ASSP Magazine*, vol.73, pp. 4-29, 1984.

[8] B.H. Juang and A.H. Gray, "Multiple stage vector quantization for speech coding," *Proc. ICASSP 82*, pp.597-600, 1982.

[9] D.Y. Cheng and A. Gersho, "A fast codebook search algorithm for nearest neighbor pattern matching," *Proc. ICASSP 86*, pp261-264, 1986.

[10] M. Xie and J-P. Adoul, "Embedded algebraic vector quantization (EAVQ) with application to wideband speech coding," *Proc. ICASSP 96*, pp.240-243, 1996.

[11] T. Berger, F. Jelinek and J.K. Wolf, "Permutation Codes for Sources," *IEEE Trans. Inform. Theory*, vol.IT-18, No.1, pp.160-169, 1972.

[12] Luzheng Lu, G. Cohen and P. Godlewski, "Composite permutation coding of speech waveforms," *Proc.ICASSP 86*, pp.2359-2362, 1986.

[13] S.A. Ramprashad, "Understanding the quality losses of embedded speech and audio coders," in *IEEE Workshop on Speech Coding*, Ibaraki, Japan, October 2002, pp.11-13.