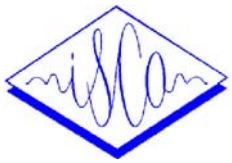

INTERSPEECH 2004 - ICSLP
8th International Conference on Spoken Language Processing
ICC Jeju, Jeju Island, Korea
October 4-8, 2004

INNER PRODUCT BASED-MULTIBAND VECTOR QUANTIZATION FOR WIDEBAND SPEECH CODING AT 16KBPS

Joon-Hyuk Chang[1], *Seung Yeol Lee*[2] *and Nam Soo Kim*[2]

[1]Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA, USA.
[2]School of Electrical Engineering, Seoul National University, Seoul, Korea.

ABSTRACT

This paper describes a multiband vector quantization (VQ) technique for wideband speech coding at 16 kb/s. Our approach consists of splitting the input speech into two separate bands and then applying an independent coding scheme for each band. A code excited linear prediction (CELP) coder is used in the lower band while a transform based coding strategy is applied in the higher band. The spectral components in the higher frequency band are represented by a set of modulated lapped transform (MLT) coefficients. The higher frequency band is divided into three subbands, and the MLT coefficients construct a vector for each subband. For the vector quantization (VQ) of these vectors, an inner product-based distance measure is proposed.

1. INTRODUCTION

There have been numerous advances in the field of speech coding in the past twenty years, allowing a considerable decrease in the bit rate necessary to transmit speech on the telephone bandwidth (300-3400 Hz). The 300-3400 Hz bandwidth, requiring a sampling frequency of 8 kHz, provides a speech quality referred to as toll quality. This is sufficient for telephone communications, but for emerging applications such as teleconferencing, multimedia services, and internet telephony, an improved quality is required. For that reason, a research on wideband speech coding with a frequency range from 300 to 7500 Hz is much more needed.

Recently, a number of practical wideband speech coding algorithms have been proposed. Most of these algorithms are based on the code excited linear prediction (CELP) model, or on the transform/subband coding strategies. The bit rate in general varies from 16 to 32 kb/s. In [1], a subband coder which addresses the problem of low delay coding of wideband speech as well as general audio coding at 32 kb/s is announced. At the same bit rate of 32 kb/s, [2] presented a transform coder based on the Fourier transform. At lower bit rates, [3] developed a 16 kb/s subband CELP coder. A distinguished feature of this algorithm is that the excitation signal is split into lower frequency and higher frequency bands. In recent years, several new algorithms using the base CELP coder [4], [5] have been introduced. The use of base coders is due to the fact that they can be efficiently implemented while retaining a good performance.

We propose a 16 kb/s wideband speech coder with high speech quality comparable to G.722 at 48 kb/s [6]. An overview of our wideband speech codec and a novel algorithm for the transmission of higher frequency band information of speech are presented. Specifically, for the lower band speech, we apply the conventional CELP codec (G.729A at 8 kb/s) [7]. For the higher band information which does not have sufficient harmonic components, the coefficients generated by the modulated lapped transform (MLT) are used to represent the spectral characteristics. The quantization and encoding are carried out in each band separately. The higher frequency band (from 3.6 to 7.1 kHz) is divided into three subbands. In the proposed algorithm, MLT coefficients extracted from each subband are first normalized by the root mean square (RMS) value and then vector quantized. It is noted that an inner product based distance measure is incorporated to achieve a good performance of vector quantization (VQ). For the purpose of evaluation, we carried out subjective quality tests. Results from the test confirm that our approach is much better in performance than the G.722 at 48 kb/s.

2. OVERVIEW OF THE PROPOSED CODING METHOD

We apply a split-band coding approach, using a high-quality narrowband CELP codec for the lower band and a VQ coding scheme for the higher band MLT coefficients. First, the input signal, sampled at 16 kHz, is lowpass filtered with a cutoff frequency of 4 kHz and then downsampled to 8 kHz sampling rate. The tap size of the lowpass filter (LPF) used for downsampling is 31. It is known that the use of the lowpass filter for changing the sampling rate leads to an algorithmic delay.

The lower band speech is passed through the G.729A

This work was partly supported by the critical technology program of Korean ministry of science and technology.

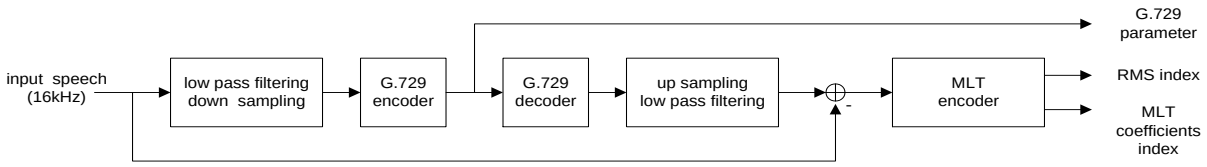

Fig. 1. Encoding part architecture.

coder and constructs a stream of coded parameters. Besides transmitting the lower band coded parameters, the residual speech in the higher band is computed from the error signal between the original speech and the reconstructed speech at the encoding stage. For reconstructing the decoded speech at 16 kHz sampling frequency, the coded parameters which represent the lower band are decoded by the G.729A decoder and then upsampled to 16 kHz. The operation of our wideband speech coding algorithm is depicted in Fig. 1.

The residual signal which includes both the quantization error in the lower band and the spectral components in the higher band is fed to the MLT encoder. In the MLT encoder, we focus on the base codec error in the frequency range of 3.6-4 kHz which is responsible for the perceptual quality degradation, and the high frequency (4-7.1 kHz) speech. Experimentally, it is found that the speech in 3.6-4 kHz should be included since the G.729A cannot cover the highest frequency range in the lower band perfectly, which may result in a loss of boundary information between the two split bands.

3. HIGHER BAND SPEECH CODING ALGORITHM

For efficient transmission of the higher band information (3.6-7.1 kHz) including the error term of the base coder, we divide the frequency region into non-uniform three subbands. The frequency range of the three subbands are i=1 (3.6-4.3 kHz), i=2 (4.4-5.1 kHz), and i=3 (5.2-7.1 kHz).

3.1. Modulated Lapped Transform

The MLT is commonly used to implement a block transform coding algorithm in video and audio compression. A serious drawback of this block transform is the introduction of blocking artifacts in the reconstructed signal. This is mainly due to the independent blocking process of quantization where the error will make discontinuities at the block boundaries. The MLT is a form of the lapped orthogonal transform (LOT), which was extensively investigated in [8] for eliminating discontinuities in the reconstructed signal at the block boundaries. One of the features of the MLT is to apply a 50 per cent of overlap between adjacent MLT frames. As shown in Fig. 2, 80 samples of the current frame are overlapped with the previous frame. This overlap leads to 80 samples of algorithmic delay at the decoding stage.

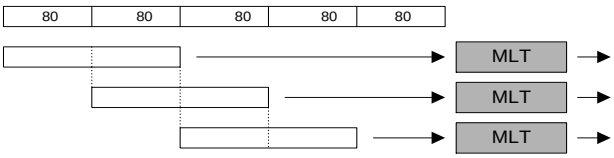

Fig. 2. Forward MLT embedded in this system.

Two consecutive subbands are windowed by a rectangular window of length $2M$, and the MLT coefficients are computed as follows:

$$MLT(k) = \sum_{n=0}^{2M-1} x(n)p_{n,k} \quad (1)$$

for k = 0, 1, $\cdots$, M - 1, where

$$p_{n,k} = h(n)\sqrt{\frac{2}{M}}cos\Big[\frac{(2n+M+1)(2k+1)\pi}{4M}\Big] \quad (2)$$

and $h(n)$ is a lowpass half band filter. For perfect reconstruction of the original signal, we choose $h(n)$ such that

$$h(n) = sin\Big[\Big(n+\frac{1}{2}\Big)\Big(\frac{\pi}{2M}\Big)\Big]. \quad (3)$$

Note that only M MLT coefficients are computed from $2M$ samples.

3.2. RMS Normalization and Quantization for MLT coefficients

To reduce the dynamic range of each MLT coefficient , we normalize each MLT coefficient by the corresponding RMS value in each subband such that

$$RMS(i) = \sqrt{\frac{1}{D(i)}\sum_{\forall k \in SB(i)}\big(MLT(k)\cdot MLT(k)\big)} \quad (4)$$

where $D(i)$ denotes the number of MLT coefficients in the ith subband and $SB(i)$ represents the set of coefficient indices in the ith subband. For the scalar quantization of $RMS(i)$, we assign 5 bits for each subband and perform the quantization procedure as follows:

$$\Big[\frac{\widehat{RMS}(i)-0.5}{2}\Big] < log_2(RMS(i)) < \Big[\frac{\widehat{RMS}(i)+0.5}{2}\Big] \quad (5)$$

$$, \text{for} -10 \le \widehat{RMS}(i) \le 21$$

where the minimal(-10) and maximal(21) bounds of $\widehat{RMS}(i)$ are chosen as the experimentally optimized values. Consequently, the RMS normalized MLT coefficients vector, $\widetilde{MLT}(i)$ is given by

$$\widetilde{\mathbf{MLT}}(i) = \frac{\mathbf{MLT}(i)}{RMS(i)\sqrt{D(i)}} \quad (6)$$

where

$$\begin{aligned} \mathbf{MLT}(1) &= [MLT(36), ..., MLT(43)] \\ \mathbf{MLT}(2) &= [MLT(44), ..., MLT(51)] \\ \mathbf{MLT}(3) &= [MLT(52), ..., MLT(71)]. \end{aligned}$$

3.3. Vector Quantization for the Normalized MLT coefficients

Generally, when establishing the codebook, an appropriate choice of the distance measure is inevitable [9]. In this subsection, we propose a distance measure based on the inner product rather than the conventional Euclidean distance. Let x, y be two different vectors with unit magnitude. Then, the inner product of x and y represents $cos\theta$ where θ is the angle of intersection as shown in Fig. 3.

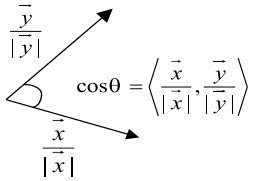

Fig. 3. Inner product.

Since the MLT coefficients vector in each subband is normalized to have unit magnitude, the inner product based distance measure is considered desirable for efficient quantization of the spectral shape. Let $\tilde{x}$ and $\tilde{y}$ be two normalized MLT coefficients vectors. Then the distance $d(\tilde{x}, \tilde{y})$ is given by

$$d(\tilde{x}, \tilde{y}) = 1 - < \tilde{x}, \tilde{y} > \quad (7)$$

where $< \cdot, \cdot >$ indicates the inner product of the two vectors.

We assign 8, 8, and 9 bits to the first, second and third subbands, respectively, for the quantization of normalized MLT coefficients.

4. DESCRIPTION FOR DECODER

At the decoder, we can reconstruct the wideband speech by summing the lower band and higher band decoded speech. The G.729A decoder makes the lower band speech while the higher band speech is obtained by the transmitted RMS energy and the vector quantized MLT coefficients in each subband.

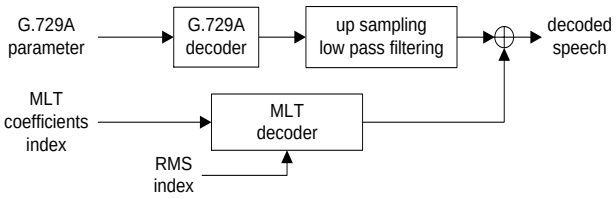

Fig. 4. Decoding part architecture.

4.1. Lower Band Decoder

The base G.729A decoder constructs the 4 kHz bandlimited speech from the lower band coding parameters. The use of upsampling by a factor of 2 and the LPF following the G.729A decoder is shown in Fig. 4.

4.2. MLT-VQ Decoder

Based on (5), $RMS(i)$ can be recovered from the RMS index such that

$$RMS^*(i) = 2^{\left(\frac{2}{RMS(i)}\right)}. \quad (8)$$

The RMS normalized MLT vectors are reconstructed by substituting the VQ index with the corresponding codeword as follows:

$$\widetilde{\mathbf{MLT}}^*(i) = \mathbf{C}_i(M_i). \quad (9)$$

in which $\mathbf{C}_i(M_i)$ represents the codeword corresponding to the VQ index M_i in the ith subband. Finally, according to (6), the higher band MLT coefficients are given by

$$\mathbf{MLT}^*(i) = RMS^*(i) \cdot \sqrt{D(i)} \cdot \widetilde{\mathbf{MLT}}^*(i). \quad (10)$$

To generate the decoded speech $x(n)$, one requires not only the MLT coefficients of the current block but also those obtained in the previous block as shown in Fig. 5. The operation of inverse MLT (IMLT) is decomposed into windowing and overlap-addition given as follows:

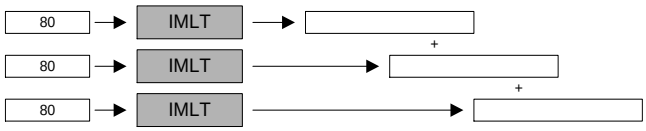

Fig. 5. Inverse MLT embedded in this system.

$$x(n) = \sum_{k=0}^{M-1} \left[MLT^*(k) p_{n,k} + MLT^{p*}(k) p_{n+M,k} \right] \quad (11)$$

where $MLT^{*}(k)$ and $MLT^{p*}(k)$ represent the MLT coefficients of the current and previous blocks, respectively with

$$p_{n,k} = h(n)\sqrt{\frac{2}{M}}cos\Big[\frac{(2n+1)(2k+1)\pi}{4M} + (2k+1)\frac{\pi}{2}\Big]. \quad (12)$$

5. RESULTS

5.1. Bit Allocation

A total of 160 bits are required for the encoding of each frame (10 msec), and the detailed bit allocation is summarized in Table I.

TABLE I Bit allocation of proposed 16 kb/s wideband speech coding algorithm.

Parameter	Frequency Range (kHz)	Frequency Size (kHz)	Bits per Frame (bits)
CELP	0.1-4.0	4.0	80
RMS	3.6-4.3	0.8	10
	4.4-5.1	0.8	10
	5.2-7.1	2.0	10
MLT	3.6-4.3	0.8	16
	4.4-5.1	0.8	16
	5.2-7.1	2.0	18
Total	0.1-7.1	7.1	160

5.2. Subjective Evaluation

To evaluate the performance of the wideband speech coding algorithm, we conducted blind preference tests on a number of clean and noisy speech samples. A male and a female speakers provided 10 sentences which were sampled at 16 kHz. Two kinds of typical noise sources were applied to simulate the noisy environments. They were the babble and car noises from the NOISEX-92 database. The listening tests were performed by ten male and ten female listeners. The blind preference test informs how much the speech generated by the method A is perceived better than that by the method B. The degree of preference is scored such that 3 (Much better), 2 (Better), 1 (Slightly better), 0 (About), -1 (Slightly worse), -2 (Worse), -3 (Much worse). The preference test scores both for the clean and noisy conditions are shown in Table II.

TABLE II Blind Preference Test Result (A:proposed algorithm, B:G.722 at 48 kb/s).

Condition	Clean	Babble SNR 15dB	Car SNR 15dB
Preference Result	0.346	0.099	0.423

6. CONCLUSIONS

In this paper, we have proposed a novel approach to the wideband speech coding at 16 kbp/s. One of the distinguished feature of our approach is to apply the VQ technique for the quantization of the RMS normalized MLT coefficients in the higher frequency band. For this, we have devised an inner product based distance measure which offers an efficient way to quantized the constrained set of MLT vectors. The performance of the proposed algorithm has been found superior to that of the G.722 48kb/s ITU-T standard codec through a number of the blind preference tests.

7. REFERENCES

[1] T. Wuppermann and F. de Bont, "Feasability study of 32 kb/s wideband speech and music coding with a low-delay filterbank," *IEEE workshop on speech coding for Telecommunications"*, pp. 11-12, 1993.

[2] S. Quackenbush, "A 7 kHz bandwidth, 32 kbps speech coder for ISDN," *Proc. of Int. Conf. Acoust., Speech, Signal Processing*, pp. 1-4, 1991.

[3] G. Roy and P. Kabal, "Wideband CELP speech coder at 16 kbits/sec," *Proc. of Int. Conf. Acoust., Speech, Signal Processing*, Toronto, pp. 17-20, 1991.

[4] A. McCree, "A 14 kb/s wideband speech coder with a parametric highband model," *Proc. of Int. Conf. Acoust., Speech, Signal Processing*, Istanbul, pp. 3109-3112, 2000.

[5] K. Koishida, V. Cuperman, and A. Gersho, "A 16-kbit/s bandwidth scalable audio coder based on the G.729 standard," *Proc. of Int. Conf. Acoust., Speech, Signal Processing*, Istanbul, pp 1444-1447, 2000.

[6] ITU-T Recommendation G.722, "7 kHz audio-coding within 64 kbit/s," 1988.

[7] ITU-T Recommendation G.729, "Coding of speech at 8 kbit/s using conjugate structure algebraic code-excited linear prediction (CS-ACELP)," 1996.

[8] H. S. Malvar, *Singal Processing with Lapped Transforms.* Boston, MA: Artech House, 1992.

[9] Y. Linde, A. Buzo, R. M. Gray, "An algorithm for vector quantization design," *IEEE Trans. Communications*, Jan. 1980.