Scalable Multimode Tree Coder with Perceptual Pre-weighting and Post-weighting for Wideband Speech Coding

Ying-Yi Li and Jerry D. Gibson
Department of Electrical and Computer Engineering, University of California, Santa Barbara, USA
Email: yingyi_li@umail.ucsb.edu, gibson@ece.ucsb.edu
Review Topic: Speech, Image and Video Processing

***Abstract*—A scalable Multimode Tree Coder with perceptual pre-weighting and post-weighting filters for wideband speech is presented. The average bit-rate of the Mutlimode Tree Coder operates at 30%-60% of the bit-rate of G.722. In addition, the algorithmic delay of the Multimode Tree Coder is 12.375 ms, which is about half of AMR-WB, and computational complexity is about a third of AMR-WB. Therefore, the Multimode Tree Coder is a low delay, low complexity, and moderate bit-rate speech codec for wideband speech. The scalable wideband Multimode Tree Coder provides a variable bit rate option not available before, while maintaining good performance, low delay, and low complexity.**

I. Introduction

G.722 [1] is an ITU-T standardized codec that provides good performance for wideband speech with low delay and low complexity at 48, 56, and 64 kbps. AMR-WB [2], [3] is another standardized speech codec for wideband speech. Even though it produces better performance at lower rates, the complexity and delay are substantially increased. In order to develop a low delay, low complexity, and low bit-rate speech coder for wideband speech, we extend the Multimode Tree Coder [4], [5] to wideband speech.

The Multimode Tree Coder is based on Multimode classification and tree coding [4], [5]. Tree coding is an encoding procedure where speech samples are coded effectively based on the best long term tree-structured fit to the input waveform [6]. In our wideband Multimode Tree Coder, we split the wideband speech into two sub-bands, and the lower sub-band signal is classified into five modes: Voiced (V), Unvoiced (UV), Onset (ON), Hangover (H), and Silence (S), and each mode is coded at a suitable bit-rate.

The tree coder with G.727 ADPCM as the Code Generator is used in the lower sub-band of the Multimode Tree Coder while the higher sub-band is coded by a tree coder with G.722 higher sub-band ADPCM as the Code Generator. In the lower sub-band, Multimode coding is also used with perceptual error weighting filters. In order to reduce the computational complexity of the distortion calculation in the lower sub-band signal, pre- and post-weighting filters are used instead.

In communications and computer networks, bit-rate scalability provides acceptable speech quality based on the current network condition. The enhancement bits may be added or dropped depending on available transmission bit-rate. In this paper, G.727 embedded ADPCM is used as the Code Generator in the lower sub-band tree coder. Since G.727 provides a bit-rate scalability option, our wideband Multimode Tree Coder is also bit-rate scalable.

The paper is organized as follows. Section II describes the details of the perceptual pre-weighting and post-weighting Multimode Tree Coder, and the details of bit-rate scalability are provided in Section III. The performance of our wideband Multimode Tree Coder is compared with G.722 and AMR-WB in Section IV. Finally, conclusions are presented in Section V.

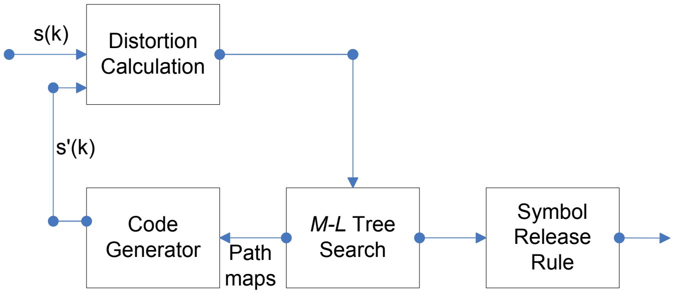

Fig. 1. Tree coder

II. Perceptual Pre-weighting and Post-weighting Multimode Tree Coder

A. Tree Coding

A tree coder includes a Code Generator, a Tree Search algorithm, a distortion measure and a path map symbol release rule as shown in Figure 1. The Tree Search algorithm, in combination with the Code Generator and appropriate distortion measure, chooses the best candidate path to encode the current input sample. The symbol release rule decides the symbols on the best path to encode. For simplicity, we use G.727 as the Code Generator in the lower sub-band and G.722 higher sub-band coder as the Code Generator in the higher sub-band since they are low delay and low complexity ADPCM coders. In order to reduce the computational complexity, we use the *M-L* Tree Search as the Tree Search algorithm. The *M* paths with minimum cumulative distortion are chosen and extended along their children. The distortion of each path is calculated,

This research has been supported by NSF Grant Nos. CCF-0728646 and CCF-0917230.

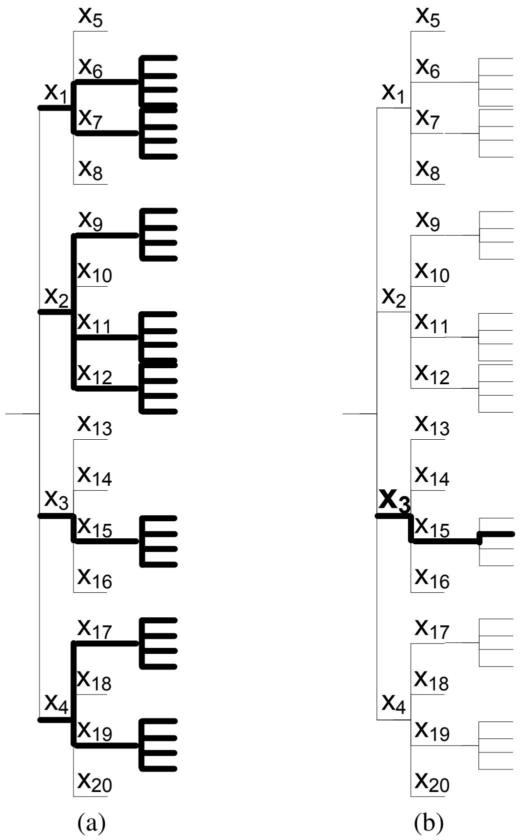

Fig. 2. An example of 2 bits/sample tree (a) Search paths of the *M-L* Tree Search algorithm for $L = 3$ and $M = 8$ (b) the minimum cumulative distortion path

and the symbol corresponding to the first node in the minimum cumulative distortion path is transmitted.

For example, there are 4^L paths of a tree generated with a 2 bits/sample ADPCM coder as shown in Figure 2. Assume $L = 3$ and $M = 8$ for the *M-L* Tree Search; the 8 minimum cumulative distortion paths, $x_1 \rightarrow x_6$, $x_1 \rightarrow x_7$, $x_2 \rightarrow x_9$, $x_2 \rightarrow x_{11}$, $x_2 \rightarrow x_{12}$, $x_3 \rightarrow x_{15}$, $x_4 \rightarrow x_{17}$, and $x_4 \rightarrow x_{18}$, with their children are marked as search paths in Figure 2 (a). Based on the *M-L* Tree Search algorithm, we only maintain M minimum cumulative distortion paths instead of 4^L paths, which saves computational complexity for tree searching. In Figure 2 (b), the minimum cumulative distortion path, $x_3 \rightarrow x_{15}$, is marked. By the single symbol release rule, the symbol x_3 is released and encoded.

B. Mode Decision

The mode decision of the Multimode Tree Coder is a low delay and low complexity method based on G.727 ADPCM coder state parameters, step-size scale factor and long-term average magnitude of weighted quantization level, frame energy, and number of zero-crossings. A speech frame of 11.25 ms is classified into one of these five modes: Voiced (V), Onset (ON), Unvoiced (UV), Hangover (H), and Silence (S). Each mode is coded at a suitable bit-rate.

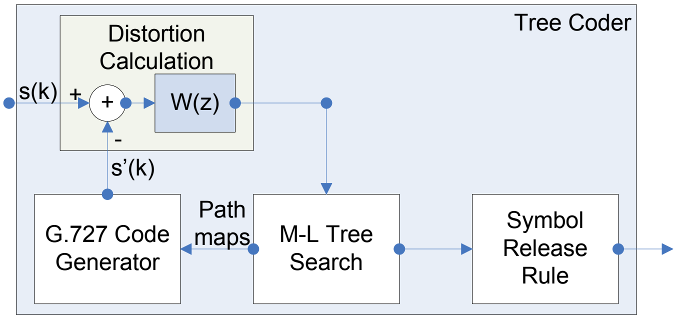

Fig. 3. Perceptual weighting in the tree coder

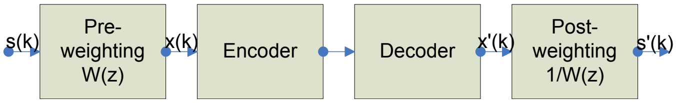

Fig. 4. Pre-weighting and post-weighting filters for speech codecs

C. Perceptual Pre-weighting and Post-weighting

The distortion of each path is calculated with a perceptual weighting filter, which helps to choose a path where the noise is masked by the speech spectrum. The weighting filter is

$$W(z) = \frac{1 - \sum_{i=1}^{N} a_i z^{-i}}{1 - \sum_{i=1}^{N} \mu^i a_i z^{-i}}, \tag{1}$$

where μ is 0.86, N is 5, and the a_i's are the short term predictor coefficients calculated from the current speech frame [4]. The computational complexity with the perceptual weighting filter inside the tree search loop, shown in Figure 3, is high. Assume that the computational complexity of $W(z)$ is C operations, and B is the number of siblings of the tree such as $B = 4$ for the 2 bits/sample tree, then the complexity of releasing one symbol is $M \cdot B \cdot L \cdot C$ operations. Schuller, Yu, Huang, and Edler [7] have employed adaptive pre-filtering and post-filtering in lossless audio coding. They showed that lossless audio coding with pre- and post-filtering still keeps the high quality. In addition, Shetty and Gibson [8] employed perceptual pre-weighting and post-weighting in a G.726 ADPCM codec and a modified AMR-NB CELP codec. They showed that the performance of lossy coding with pre- and post-weighting also performs well. As shown in Figure 4, the computational complexity of our Multimode Tree Coder is reduced to $2C$ operations for releasing one symbol by using pre-weighting and post-weighting filters.

The pre-weighting filter $W(z)$ and post-weighting filter $\frac{1}{W(z)}$ are designed to mask the reconstruction error at the output by the input spectrum. Let $S(z)$ be the input speech, $X(z)$ be the pre-weighted speech, $X'(z)$ be the pre-weighted speech output, and $S'(z)$ be the output speech after post-weighting. From Figure 4, the relation of $S(z)$ and $X(z)$ is

$$S(z)W(z) = X(z), \tag{2}$$

and the relation of $S'(z)$ and $X'(z)$ is

$$X'(z)\frac{1}{W(z)} = S'(z). \tag{3}$$

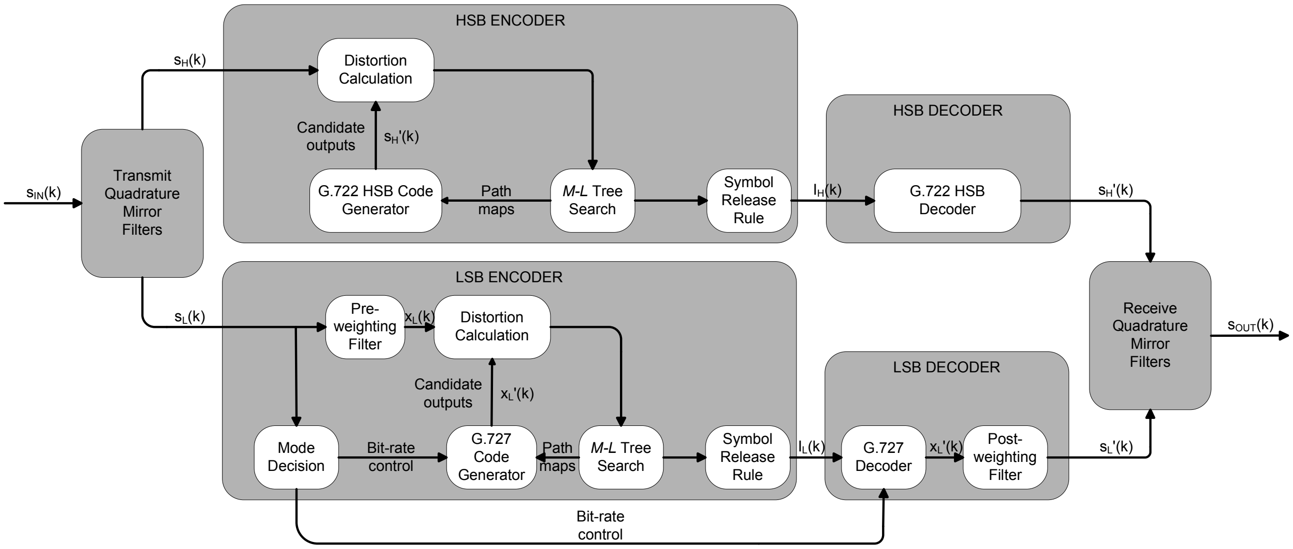

Fig. 5. Perceptual Pre- and Post-weighting Multimode Tree Coder for wideband speech coding

Let $E(z)$ denote the coding error for the pre-weighted speech. From Eq. (2) and Eq. (3), the coding error $E(z)$ will be

$$\begin{aligned} E(z) &= X(z) - X'(z) \\ &= W(z)S(z) - W(z)S'(z) \\ &= W(z)[S(z) - S'(z)], \end{aligned} \tag{4}$$

where $W(z)$ is used to shape the reconstruction error [8].

The objective is to match the frequency response of the perceptual error weighting filter generated with the 5th order LPC coefficients, Eq. (1), with the frequency response of the filter generated with ADPCM predictor coefficients. The post-weighting filter of the 5th order LPC coefficients is

$$\frac{1}{W(z)} = \frac{1 - \sum_{i=1}^{5} (0.86)^i a_i z^{-i}}{1 - \sum_{i=1}^{5} a_i z^{-i}}, \tag{5}$$

while the post-weighting filter generated with ADPCM pole-zero coefficients is

$$H_{post}(z) = \frac{1 + \sum_{i=1}^{6} m_2^i b_i z^{-i}}{(1 + \sum_{i=1}^{6} m_3^i b_i z^{-i})(1 - \sum_{i=1}^{2} m_1^i a_i z^{-i})}, \tag{6}$$

where a_i's are pole coefficients, b_i's are zero coefficients, $m_1 = 0.2$, $m_2 = 1.0$, and $m_3 = 0.85$ in both pre- and post-weighting filters in our experiments.

D. Wideband Perceptual Pre-weighting and Post-weighting Multimode Tree Coder

The block diagram of the Multimode Tree Coder for wideband speech is shown in Figure 5. The input, s_{IN}, is split into two sub-bands: the lower sub-band (LSB), s_L, and the higher sub-band (HSB), s_H, by transmit quadrature mirror filters (QMFs). The LSB input signal, s_L, is coded by perceptual pre-weighting and post-weighting Multimode Tree Coder [4] with G.727 [9] as the Code Generator, and the HSB input signal, s_H, is coded by a tree coder with G.722 HSB as the Code Generator without perceptual weighting. The LSB symbol, I_L is decoded by the G.727 decoder followed by a post-weighting filter, and the HSB symbol, I_H, is decoded by the G.722 HSB decoder. Finally, the receive QMFs interpolate the outputs, s'_L and s'_H, of the LSB and HSB decoders and produces an output, s_{OUT}.

The LSB input signal, s_L, sampled at 8 kHz is filtered by the pre-weighting filter, and used for mode decisions. The Code Generator, a G.727 codec, codes the pre-weighted sample, x_L, at a suitable bit-rate based on the results of the mode decision. Then the distortion between candidate output, x'_L, and pre-weighted input, x_L, is calculated via the *M-L* Tree Search, a tree search with depth L and M retained paths. Finally, the first symbol, I_L, in the minimum distortion path is encoded. In the LSB decoder, the coded symbol, I_L, is decoded by a G.727 decoder. Since a pre-weighting filter is used in the LSB encoder, a post-weighting filter is required after the decoder. Therefore, the output of the G.727 decoder, x'_L, is filtered by the post-weighting filter, and the reconstructed signal, s'_L, is produced. The HSB encoder is similar to the LSB encoder. However, there is no pre- and post-weighting filters in the HSB encoder, and the bit-rate is 2 bits/sample. The HSB decoder is a G.722 HSB decoder.

III. Bit-rate Scalability

In communications and computer networks, it is advantageous to allow the network to drop least significant bits in a transmitted data word without the source being re-encoded. Embedded coding was developed to be used in these situations. Embedded ADPCM algorithms are variable bit rate coding algorithms with the capability of bit dropping outside the encoder and decoder blocks, which is bit-rate scalable. The decision levels of the lower rate quantizers are subsets of the quantizer at the highest rate. This allows bit reductions at any point in the network without the need of coordination between the transmitter and the receiver. The block diagram of an embedded DPCM system is shown in Figure 6. There is a second quantizer inside the prediction loop. Quantizers Q_1 and

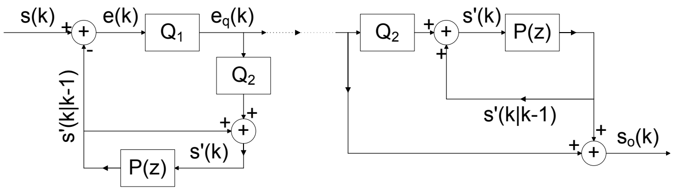

Fig. 6. Block diagram of embedded DPCM [10]

Q_2 are embedded quantizers, where Q_2 has fewer levels than Q_1. The prediction loop always uses the coarsely quantized prediction error signal so that dropping bits does not change the operation of the feedback loop. This is also true at the decoder, and so dropping least significant bits only affects the accuracy of the prediction error signal added in to reconstruct the speech at the decoder.

Embedded coding is different from variable-rate coding where the encoder and decoder must use the same number of bits in each sample. In both cases, the decoder must be told the number of bits to use in each sample. Embedded algorithms produce code words which contain enhancement bits and core bits. The Feed-Forward (FF) path utilizes enhancement and core bits, while the Feedback (FB) path uses core bits only. The inverse quantizer and the predictor of both the encoder and the decoder use the core bits. With this structure, enhancement bits can be discarded or dropped during network congestion. However, the number of core bits in the FB paths of both the encoder and decoder must remain the same.

Since G.727 is an embedded ADPCM coder and we use G.727 as the Code Generator in the lower sub-band, our wideband Multimode Tree Coder is also bit-rate scalable.

IV. Results

In this section, we compare the performance of our Multimode Tree Coder with two standardized wideband speech codecs, AMR-WB and G.722, and test the bit-rate scalability of the Multimode Tree Coder. The performance of the Multimode Tree Coder with AMR-WB at 23.05 kbps and G.722 at 64 kbps is shown in Section IV-A. Section IV-B shows the bit-rate scalable results of the Multimode Tree Coder. Finally, the algorithmic delay, computational complexity, WPESQ and average bit-rate of the Multimode Tree Coder are compared with those of AMR-WB and G.722 in Section IV-C.

A. Comparison with Standardized Speech Codecs

In order to compare the performance of the Multimode Tree Coder with AMR-WB [2], [3], and G.722 [1], the Wideband Perceptual Evaluation of Speech Quality (WPESQ) [11] is used for evaluating the quality of the wideband speech codecs. Table I shows the WPESQ and average bit-rate of the Multimode Tree Coder for wideband speech, AMR-WB, and G.722. Three female and three male English sequences are used for testing [12]. In our experiments, the tree depth, L, is 10, and M is 4 for the *M-L* Tree Search for both lower sub-band and higher sub-band tree coders. The V and ON modes in the lower sub-band are coded at 4 core bits/sample, the UV and H modes are coded at 3 core bits/sample. The higher sub-band tree coder is coded at 2 bits/sample for non-Silence frames. The S mode is coded by Silence coding at 1.19 kbps for full-band. AMR-WB in Table I is coded by the 23.05 kbps mode, and Source controlled rate operation (SCR) is enabled. G.722 in Table I is coded at 64 kbps. The results in Table I show that the WPESQs of our Mutlimode Tree Coder for wideband speech coding are between fair and good quality with that of AMR-WB at 23.05 kbps somewhat higher, and G.722 at 64 kbps best of all. The average bit-rate of the Multimode Tree Coder is higher than that of AMR-WB. However, it is 30%-60% lower than that of G.722.

TABLE I
COMPARISON OF WPESQ AND AVERAGE BIT-RATE OF THE MULTIMODE TREE CODER WITH STANDARDIZED SPEECH CODECS

Sequence	Multimode Tree Code V,ON:(4,0);UV,H:(3,0)		AMR-WB 23.05 kbps		G.722 64 kbps	
	WPESQ	bit-rate (kbps)	WPESQ	bit-rate (kbps)	WPESQ	bit-rate (kbps)
F1	3.348	30.00	3.689	13.09	4.336	64
F2	3.442	25.65	3.733	12.45	4.236	64
F3	3.444	31.34	3.739	14.69	4.219	64
M1	3.609	34.59	4.136	16.55	4.430	64
M2	3.485	29.44	4.171	14.26	4.309	64
M3	3.497	31.52	4.159	16.02	4.391	64
Average	3.471	30.42	3.938	14.51	4.320	64

B. Bit-rate Scalability

Bit-rate scalability provides acceptable speech quality to each user based on the current network condition. There is no need to send extra side information for controlling the bit-rate. Since G.727 is an embedded ADPCM coder, the lower sub-band of the Multimode Tree Coder with G.727 Code Generator is also scalable. Table II shows the WPESQ and average bit-rate of the scalable Multimode Tree Coder. The UV and H modes are coded at 3 core bits/sample in the lower sub-band, the higher sub-band tree coder is coded at 2 bits/sample for non-Silence frames, and the S mode is coded at 1.19 kbps for full-band. The V and ON modes are scalable in the lower sub-band, and are coded at 3 core bits/sample with different enhancement bits. Table III shows the WPESQ of G.722 at 64 kbps, 56 kbps, and 48 kbps. The results in Tables II and III show that the average bit-rate of the scalable Multimode Tree Coder is from 19 kbps to 37 kbps, while the bit-rate of G.722 is from 48 kbps to 64 kbps. The highest average bit-rate of the Multimode Tree Coder is lower than the lowest bit-rate of G.722. Even though the WPESQs of the Multimode Tree Coder shown in Table II are lower than those of G.722, it can be improved by increasing the number of core bits.

C. Comparison of the Performance, Average Bit-rate, Delay, and Complexity

The delay of the Multimode Tree Coder is caused by the mode decision and the *M-L* Tree Search algorithm. The mode decision is made every 11.25 msec. Since L is 10, there is a 9 samples delay for look-ahead. Therefore, the total delay of the Multimode Tree Coder is 12.375 ms. The comparison of WPESQ, average bit-rate, algorithmic delay, and computational complexity of the three speech codecs are shown in Table IV. While the WPESQ values of the Multimode Tree

TABLE II
WPESQ AND AVERAGE BIT-RATE OF THE SCALABLE MULTIMODE TREE CODER WITH 3 CORE BITS AND DIFFERENT ENHANCEMENT BITS FOR V AND ON MODES IN LSB

Sequence	Multimode Tree Coder V,ON:(3,2);UV,H:(3,0)		Multimode Tree Coder V,ON:(3,1);UV,H:(3,0)		Multimode Tree Coder V,ON:(3,0);UV,H:(3,0)	
	WPESQ	bit-rate (kbps)	WPESQ	bit-rate (kbps)	WPESQ	bit-rate (kbps)
F1	3.415	34.64	3.432	30.00	3.110	25.37
F2	3.401	29.59	3.411	25.65	3.255	29.13
F3	3.478	36.11	3.490	31.34	3.283	19.20
M1	3.684	40.05	3.596	34.59	3.461	21.71
M2	3.524	34.01	3.513	29.44	3.349	22.91
M3	3.512	36.26	3.564	31.52	3.467	26.79
Average	3.502	35.11	3.501	30.42	3.321	24.19

TABLE III
WPESQ OF G.722 AT DIFFERENT BIT-RATE

Sequence	G.722 64 kbps	G.722 56 kbps	G.722 48 kbps
F1	4.336	4.157	3.777
F2	4.236	4.226	4.041
F3	4.219	4.198	3.990
M1	4.430	4.403	4.146
M2	4.309	4.264	4.015
M3	4.391	4.347	4.127
Average	4.320	4.266	4.016

TABLE IV
COMPARISON OF THE SPEECH CODER ATTRIBUTES OF THE MULTIMODE TREE CODER, AMR-WB, AND G.722

	Scalable Multimode Tree Coder	AMR-WB 23.05 kbps	G.722
WPESQ	3.1-3.7	3.6-4.2	4.2-4.5
Average Bit-rate(kbps)	19.20-36.26	12.45-16.55	64
Algorithmic Delay(ms)	12.375	25	1.625
Computational Complexity (WMOPS)	12.74	39.0	<10

Coder are lower, the average bit-rate of the Multimode Tree Coder saves at least 40% compared to G.722. In addition, the algorithmic delay of the Multimode Tree Coder is about half of the AMR-WB, and the computational complexity is about a third of the AMR-WB.

V. CONCLUSIONS

The proposed Multimode Tree Coder reduces the average bit-rate by Multimode coding. Even though tree coder is a delayed coding, the frame delay and look-ahead delay are low, 12.375 ms. In addition, the perceptual pre-weighting and post-weighing tree coder with G.727 ADPCM coder as the Code Generator combined with the *M-L* Tree Search algorithm is also computationally efficient. Last but not least, bit-rate scalability also can be achieved by our Multimode Tree Coder.

While the performance of the scalable Multimode Tree Coder with perceptual pre-weighting and post-weighting filters is lower than AMR-WB at 23.05 kbps and G.722 at 64 kbps, the algorithmic delay and computational complexity of the Multimode Tree Coder are much lower than those of AMR-WB. The delay is about half of the AMR-WB and the computational complexity is about a third of the AMR-WB. The bit-rate scalability results of the Multimode Tree Coder shows that the highest average bit-rate of the Multimode Tree Coder is lower than the lowest bit-rate of G.722, and that the low delay and low complexity of the Multimode Tree Coder are still retained. We have thus added a source-controlled and network-controlled variable bit rate option important in many applications.

REFERENCES

[1] ITU-T Recommendation G.722, "7 kHz Audio-Coding within 64 kbits/s ," Nov. 1988.

[2] 3GPP, "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding functions," 3rd Generation Partnership Project (3GPP), TS 26.190, Mar. 2011.

[3] ITU-T Recommendation G.722.2, "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)," July 2003.

[4] P. Ramadas, Y.-Y. Li, and J. D. Gibson, "Multimode Tree Coding of Speech with Perceptual Pre-weighting and Post-weighting," *the 12th IASTED International Conference on Signal and Image Processing*, 2010.

[5] P. Ramadas and J. D. Gibson, "Phonetically Switched Tree coding of speech with a G.727 Code Generator," *the 43rd Annual Asilomar Conference on Signals, Systems, and Computers*, Nov. 1-4, 2009.

[6] H. C. Woo and J. D. Gibson, "Low delay tree coding of speech at 8 kbit/s," *IEEE Trans. on Speech and Audio Processing*, vol. 2, no. 3, pp. 361 –370, July 1994.

[7] G. Schuller, B. Yu, D. Huang, and B. Edler, "Perceptual Audio Coding using Adaptive Pre- and Post-Filters and Lossless Compression," *IEEE Trans. on Speech and Audio Processing*, vol. 10, no. 6, pp. 379 –390, Sept. 2002.

[8] N. Shetty and J. D. Gibson, "Perceptual Pre-weighting and Post-inverse weighting for Speech Coding," *the 41st Annual Asilomar Conference on Signals, Systems, and Computers*, Nov. 4-7, 2007.

[9] ITU-T Recommendation G.727, "5-, 4-, 3- and 2-bit/sample embedded Adaptive Differential Pulse Code Modulation (ADPCM)," Dec. 1990.

[10] J. D. Gibson, T. Berger, T. Lookabaugh, D. Lindbergh, and R. L. Baker, *Digital compression for multimedia: principles and standards*. San Francisco, CA: Morgan Kaufmann Publishers Inc., 1998.

[11] ITU-T Recommendation P.862.2, "Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs," Nov. 2007.

[12] ITU-T Series P Supplement 23, "ITU-T coded-speech database," Feb. 1998.