A WIDEBAND CODEC AT 16/24 KBIT/S WITH 10 MS FRAMES

R. Salami, R. Lefebvre, and C. Laflamme

Department of Electrical Engineering, University of Sherbrooke,
Sherbrooke, Québec, Canada J1K 2R1

ABSTRACT

This paper describes a wideband speech/audio codec at 16/24 kbit/s with 10 ms frames. The algorithm uses an ACELP model at 16 kbit/s and a switched ACELP/TCX model at 24 kbit/s. Adaptive preemphasis is used to improve the performance at high frequencies and a hybrid forward/backward LP filter is used to improve the performance of stationary signals. Subjective tests showed that for speech signals, the codec performance at 16 and 24 kbit is equivalent to G.722 at 48 and 56 kbit/s, respectively. For music signals, the performance of the codec at 24 kbit/s was equivalent to that of G.722 at 48 kbit/s.

1. INTRODUCTION

Compression of wideband speech and audio (7 kHz bandwidth) is increasingly needed in many applications such as videoconferencing and Internet. The ITU-T has recently started a standardization activity for a wideband codec at 16 and 24 kbit/s which is required to perform similar to G.722 at 48 and 56 kbit/s, respectively, in most operating conditions. Initially, two modes were proposed: Mode A with 25 ms delay (the delay is considered as twice the frame size plus the lookahead) and Mode B with longer delay (60 ms) and lower complexity (15 MIPS). This paper describes a codec which wad proposed for mode A standardization. Section 2 describes the codec principles and Section 3 discusses the codec's performance. The conclusions are given in Section 4.

2. CODEC PRINCIPLES

2.1. Coding model and bit allocation

The coder uses the algebraic code-excited linear predictive (ACELP) coding model [1] at 16 kbit/s and switch CELP/TCX (Transform coded excitation [2]) at 24 kbit/s. The coder uses 10 ms speech frames (160 samples at the sampling frequency of 16000 sample/s). An adaptive preemphasis procedure is performed before the encoding process (2 bits are used to quantize the preemphasis filter). A hybrid forward/backward linear prediction (LP) analysis is used. The short-term prediction parameters (or LP parameters) are transmitted every speech frame. 1 bit is used to determine LP mode (forward or forward/backward). The speech frame is divided into 2 subframes of 5 ms (80 samples). The pitch and algebraic codebook parameters are transmitted every subframe. The bit allocation of the coder is shown in Table 1. The LP parameters are quantized with 34 bits. The pitch lag is encoded with 9 bits in the first subframe and 6 bits in the second subframe. The pitch gain is quantized with 4 bits and the fixed codebook gain is quantized with 5 bits in each subframe. The only difference between the two bit rate modes is the size of the innovation codebook. In the 16 kbit/s mode, the innovation codebook index is encoded with 45 bits each subframe, while in the 24 kbit/s mode it is encoded with 85 bits each subframe. The algorithm can be switched dynamically from frame to frame between the two bit rates.

Parameter	*1st subfr.*	*2nd subfr.*	*Per frame*
LP mode			1
Preemphasis filter			2
ISPs			34
Pitch delay	9	6	15
Pitch gain	4	4	8
Innovation codebook	45	45	90
codebook gain	5	5	10
Total			160

Table 1. Bit allocation of the coding algorithm at 16 kbit/s. At 24 kbit/s the innovation codebook uses 85 bits instead of 45 bits.

2.2. Pre-processing

The pre-processing block performs adaptive preemphasis. Four possible 2nd order filters, $P(z)$, can be used (2 bits). The preemphasis filter is determined based on 2nd order LP analysis of the input signal. The use of preemphasis significantly improves the codec performance at high frequencies. This gave better performance than introducing a tilt filter into the perceptual weighting filter [3]. The second advantage of preemphasis is that it reduces the dynamic range of the input signal which facilitates the fixed-point implementation of the algorithm.

2.3. Short-term prediction

Short-term prediction, or linear prediction (LP), analysis is performed once per speech frame (with a lookahead of 5 ms). The 16th order LP parameters are quantized with 34 bits, and used for the second subframe while the first subframe uses interpolated filters. The quantization and interpolation are performed in the immittance spectral pair (ISP) domain [4]. Predictive split vector quantization of the ISPs is used.

To improve the quality in case of music signals, a hybrid forward/backward LP filter configuration is used. The LP filter is either a forward filter or a hybrid forward/backward filter (depending on stationarity criterion). 1 bit is used for the LP mode.

2.4. Long-term prediction analysis

The pitch parameters are the delay and gain of the pitch filter. In the first subframe, a fractional pitch delay is used with resolutions: 1/3 in the range $[29\frac{1}{3}, 158\frac{2}{3}]$ and integers only in the range [159, 281]. For the second subframe, a pitch resolution of 1/3 is always used in the range $[T_1 - 10\frac{2}{3}, T_1 + 9\frac{2}{3}]$, where T_1 is nearest integer to the fractional pitch lag of the first subframe.

To simplify the pitch analysis procedure, a two stage approach is used [5]. First, an open loop pitch is computed every frame (10 ms) using the weighted speech signal $s_w(n)$ to find a pitch estimate T_{op}. The weighted speech is low-pass filtered and decimated by 3 to simplify the search. Second, a closed-loop pitch analysis is performed around the open-loop pitch estimate on a subframe basis. In the first subframe the range $T_{op} \pm 9$, bounded

0-7803-4073-6/97/$10.00©1997IEEE.

by 30–281, is searched. For the second subframe, closed-loop pitch analysis is performed around the pitch selected in the first subframe as described earlier. The pitch delay is encoded with 9 bits in the first subframe and the relative delay of the second subframe is encoded with 6 bits.

The pitch gain is quantized using 4-bit scalar quantization.

2.5. Innovation codebook structure

At 16 kbit/s, a 45-bit algebraic codebook is used. The 80 positions in a subframe are divided into 5 interleaved tracks. The innovation vector contains 10 non-zero pulses, where 2 pulses are placed in each track. All pulses can have the amplitudes +1 or −1. The positions and signs of the two pulses in a given track are encoded with 9 bits. This gives a total of 45 bits. The codebook is search using the fast procedure described in [6].

At 24 kbit/s the innovation codebook is either based on an algebraic codebook structure or transform-coded excitation (TCX) structure. The former codebook is more suitable for transient frames and attacks while the latter codebook is used in case of stationary periods. In the algebraic codebook case, the innovation vector contains 20 non-zero pulses, where 4 pulses are placed in each one of the 5 tracks. All pulses can have the amplitudes +1 or −1. In the TCX case, the target vector for codebook search is quantized in the transform domain [2].

The fixed codebook gain is quantized using scalar quantization with 5 bits, after applying a 2nd order moving average (MA) prediction to the innovation energy in the logarithmic domain.

2.6. Decoder

The function of the decoder consists of decoding the transmitted parameters (LP parameters, adaptive codebook vector, algebraic code vector, and gains) and performing synthesis to obtain the reconstructed speech.

The ouput of the LP synthesis filter is passed through the post-processing block which performs an adaptive deemphasis procedure (the inverse of the preprocessing procedure) to restore the dynamic of the speech signal.

3. CODEC PERFORMANCE

The codec was tested in compliance with the qualification test plan set by the ITU [7]. The test consisted of three experiments. Experiment 1a tested the codec performance in case of speech (single talkers without background noise; Experiment 1b tested the performnace for music signals; and Experiment 2 tested the performance in case of speech with background noise. Table 2 gives some of the results of Experiment 1a for the nominal level of −26 dBov (26 dB below overload). The table gives the codec condition, the combined Mean Opinion Score (MOS), the standard deviation, the difference between the reference and candidate codec and the 95% confidence interval. The codec meets the requirements for speech and significantly better in case of bit errors. The coder was significatly better than G.722 at the lower input level of −36 dBov but didn't meet the requirement at higher input level of −16 dBov. This because the G.722 reference coder is level dependent whose performance increases with increasing the input level. G.722 shows a MOS variation of almost 1 between higher and lower levels while the MOS variation of the candidate codec is limited to 0.1. The results showed that performance is slightly below meeting the tandem requirement. Initially, the nominal level was at −32 dB at which the tandem requirement was met. Then it was increased to −26 dB which increased the MOS of G.722 by 0.3 for a single encoding (due to its level dependancy).

Condition	*MOSc*	*Sc*	*d*	C_{int}
G.722 48k	3.41	0.78		
codec 16k	3.33	0.99	0.08	0.18
G.722 56k	3.77	0.78		
codec 24k	3.71	0.87	0.06	0.17
G.722 48k 0.001 BER	2.36	0.91		
codec 16k 0.001 BER	2.68	1.02	-0.32	0.19
G.722 56k 0.001 BER	2.59	0.88		
codec 24k 0.001 BER	2.85	1.11	-0.26	0.20
G.722 48k 2 tandem	2.87	0.71		
codec 16k 2 tandem	2.55	0.98	0.32	0.17
G.722 56k 2 tandem	3.34	0.73		
codec 24k 2 tandem	3.09	1.02	0.25	0.18

Table 2. Test results from Experiment 1a.

In Experiment 1b (music), the coder didn't meet the requirement at 16 kbit/s, and at 24 kbit/s it was slightly worse that 56 kbit/s G.722. The test showed that for music, the performance at 24 kbit/s is better to G.722 at 48 kbit/s. The requirements for music are difficult to attain with the short frame size of 10 ms due to the lack of frequency resolution to perform perceptual transform coding.

In Experiment 2, the codec didn't meet the requirements in the presence of background noise. This is due to the discriminatory Comparison Category Rating procedure used in this experiment.

4. CONCLUSION

The article described a wideband speech codec operating at 16/24 kbit/s. The coder operates on 10 ms speech frames using an ACELP algorithm at 16 kbit/s and a switched ACELP/TCX algorithm at 24 kbit/s. Subjective test results showed that the codec meets most the performance requirements for clean speech (equivalent to 48/56 kbit/s G.722), while it is below the requirements for music signals and background noise conditions.

In the March 1997 meeting of SG 16, it was decided to keep only the longer delay mode (60 ms) for the wideband coding standard while allowing more complexity (single commercial DSP chip). The larger delay is essential in order to meet the requirements for music signals. The procedure for testing the background noise conditions has been changed, which is likely to make it less difficult to meet the requirements. The remaining difficulty will be meeting the requirement at −16 dBov. It is in fact illogical to test the level dependency of the candidate codec against a reference codec which is itself very level dependent.

REFERENCES

[1] C. Laflamme, J-P. Adoul, R. Salami, S. Morissette, and P. Mabilleau, "16 kbps wideband speech coding technique based on algebraic CELP," Proc. ICASSP'91, pp. 13–16.

[2] R. Lefebvre, R. Salami, C. Laflamme, and J.-P. Adoul, "High quality coding of wideband audio signals using Transform-Codec eXcitation (TCX)," Proc. ICASSP'94, pp. I-193–I-196.

[3] E. Ordentlich and Y. Shoham, "Low-delay code-excited linear-predictive coding of wideband speech at 32 kbps," Proc. ICASSP'91, pp. 9–12.

[4] Y. Bistritz and S. Peller, "Immittance spectral pairs (ISP) for speech encoding," Proc. ICASSP'93, pp. II-9–II-12.

[5] R. Salami, C. Laflamme, J-P. Adoul, and D. Massaloux, "A toll quality 8 kb/s speech codec for the personal communications system (PCS)," *IEEE Trans. Veh. Technol.*, vol. 43, no. 3, pp. 808–816, Aug. 1994.

[6] R. Salami et al, "Description of GSM enhanced full rate codec," Proc. ICC'97.

[7] "Subjective qualification test plan for the ITU-T wideband (7 kHz) speech coding algorithm," ITU-T, Version 3.1, November 1996.