

MULTI-BAND CELP CODING OF SPEECH AND MUSIC

Anil Ubale and Allen Gersho

Department of Electrical Engineering
University of California, Santa Barbara, CA 93106, USA.

ABSTRACT

A novel low-delay wideband speech coder, called Multi-band CELP (MB-CELP), overcomes the major obstacles usually associated with two traditional CELP approaches to wideband speech coding, namely, fullband CELP and split-band CELP. The new MB-CELP coder employs a multi-band bank of off-line filtered excitation codebooks, fullband linear prediction synthesis, and minimization of the error between original and synthesized audio signals over the full frequency range. Two versions of MB-CELP coder at 16 and 24 kbps are described in this paper. Subjective comparison test results show that these coders perform better than G.722 coder at 48 and 56 kbps respectively.

1. INTRODUCTION

In February 1995, the ITU-T SG15 formalized a wideband (50-7000 Hz) speech coding effort by approving the terms of reference for a new dual-mode standard. Mode A is intended for general use, and will support multiple encodings which require low delay. We describe our MB-CELP coder which satisfies the requirements needed for a new ITU-T standard (Mode A) for speech input under error-free conditions. We also report our progress in applying the MB-CELP paradigm to code music signals under a low delay constraint.

2. MB-CELP CODING OF SPEECH

Among the traditional wideband CELP coding schemes, fullband CELP yields coded speech that seems to match the original extremely well in the low-frequency region, however, it fails to match the spectrum well in the high frequency region and it suffers from a high frequency hiss distortion [1]. The split-band CELP coding scheme has extra algorithmic delay due to the filter banks, and suffers in the range of frequencies where the two frequency bands overlap.

Our novel encoding scheme, MB-CELP avoids the drawbacks of both fullband and split-band CELP coders. This

This work was supported in part by the National Science Foundation under grant no. NCR-9314335, the University of California MICRO program, ACT Networks, Advanced Computer Communications, Stratacom, DSP Group, DSP Software Engineering, Fujitsu, General Electric Company, Hughes Electronics, Intel, Moseley Associates, National Semiconductor, Nokia Mobile Phones, Qualcomm, Rockwell International, and Texas Instruments.

Parameters	Bits/Frame	Bits/Frame
	16 kbps	24 kbps
LSP	28	40
Pitch Parameters	52	88
MB Codebook Indices	56	72
MB Codebook Gains	24	40
Total	160	240

Table 1: Bit allocation

is achieved by off-line filtered multi-band excitation codebooks, fullband LPC synthesis, and error minimization over the entire 8 kHz band.

The MB-CELP coder, shown in Figure 1, uses off-line filtered stochastic codebooks. The linear prediction analysis and synthesis is done over the full band. Furthermore, the error is minimized between the original and synthesized fullband signals. Due to off-line filtering of the excitation, no filterbank delay is introduced. Also, since the error minimization is over the full frequency band (unlike split-band CELP coders) the quality does not suffer in the overlap regions of the multi-band codebooks.

We have implemented two specific MB-CELP coder configurations (Table 1), each with only two bands, one at the rate of 16 kbps [2] and the other at 24 kbps. In each case, the frame size is 10 ms, and the subframe size is 2.5 ms. The look-ahead for the LP analysis is 3.75 ms, yielding an algorithmic delay of 13.75 ms. The short-term LP filter order is 16, and the perceptual weighting filter is of the type $A(z/\gamma_1)/A(z/\gamma_2)$.

The fixed (non-adaptive) codebook excitation is generated from two filtered codebooks with spectral bands from 0-4 kHz and 4-8 kHz. The excitation search is similar to ordinary multi-stage vector quantization and offers low complexity and high robustness. Furthermore, since the codevectors of the two codebooks, are nearly orthogonal (being restricted to different frequency bands), the sequential search of the codebooks provides almost the same performance as that of an optimal joint search of the codebooks.

3. MB-CELP CODING OF MUSIC

The key features of most of the state-of-the-art music coding algorithms are high resolution frequency-domain representation, adaptive bit allocation based on psychoacoustic

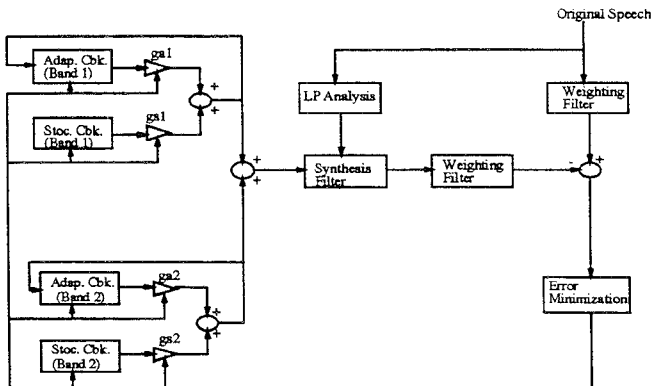


Figure 1: A general Multi-band CELP encoder structure.

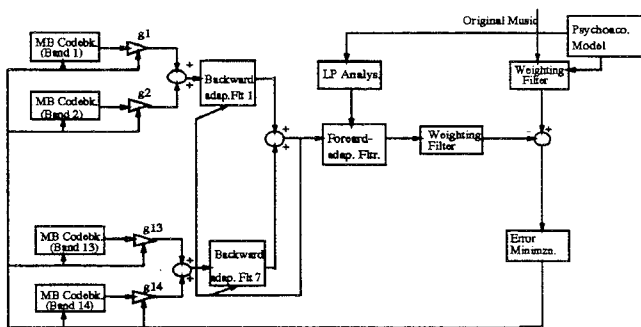


Figure 2: A Multi-band CELP encoder for music.

models, and entropy coding [3, 4].

To achieve high resolution for the frequency-domain decomposition, the transform coders use large number of samples per frame. At lower sampling rates, this implies a high algorithmic coding delay.

To achieve a delay of about 10 ms, the MB-CELP music coding algorithm as shown in Figure 2, attempts to exploit the time-domain correlations using linear prediction, and also psychoacoustic properties of human ear by using adaptive bit allocation and noise shaping.

The fine structure in the music spectrum has strong interframe correlation. In order to exploit this correlation, we use two-stage linear prediction. One stage operates in forward-adaptive mode and exploits time-domain correlations due to overall coarse spectral shape of music spectrum. The other stage models the fine structure in music spectrum and operates in backward-adaptive mode in order to achieve low bit-rate and relies on the interframe correlation of the fine structure in music spectrum.

We further utilize the properties of human ear, by using adaptive size allocation for 14 uniform bandwidth (500 Hz) multi-band codebooks. The weighting filter is derived using the ISO/MPEG [3] psychoacoustic model on input signal, to achieve perceptually optimum noise shaping.

The frame size is 10 ms and the subframe size for excitation search is 5 ms. The forward-adaptive LP order is 8. The dynamic codebook-size allocation for the multi-band excitation is derived from the locally decoded forward-

adaptive LP spectrum, and there is no need to send extra side information.

The backward-adaptive LP filter to remove the fine structure in the music signal, is updated every subframe. To obtain an accurate spectral fit to the fine structure of the music signal the backward-adaptive LP filter is implemented as a bank of seven parallel LP filters. We take 512-point FFT of the windowed reconstructed residual. The coefficients of each backward-adaptive LP filter (order 24) are derived by using the Durbin algorithm from 32 FFT lines falling in each 1000 Hz band. The LP coefficients are upsampled by a factor of 8, and used for the backward-adaptive filtering.

We use 16 bits to code forward-adaptive LP parameters, 56 bits for MB codebook gains, and 168 bits for MB codebook indices per frame.

4. RESULTS AND CONCLUSION

As reported earlier [2] our coder at 16 kbps was preferred over the G.722 coder at 48 kbps 66.80% to 33.20%.

Informal listening tests also indicate that our 24 kbps MB-CELP coder performs better than the G.722 coder at 56 kbps for clean speech inputs.

The MB-CELP music coder was tested on four types of music, namely classical music with vocal and instrumental content, classical music with more than one instruments, modern music with vocal and instrumental content and modern music with more than one instruments. At 24 kbps the quality of the classical music passages was judged very close to G.722 at 56 kbps by informal listening. However, the quality of the modern music passages was poor, and judged to be worse than G.722 at 48 kbps.

The main problem with modern music passages seems to be inefficient coding of transient signals (such as drum beats, gongs etc.). The time-domain prediction techniques provide little or no prediction gain in this case. Due to the low-delay constraint, even the frequency-domain techniques can not provide good coding efficiency.

The MB-CELP coder overcomes the drawbacks of full-band and split-band CELP coders while maintaining low-delay. For music, the 24 kbps version provides good performance for passages with a steady-state harmonic structure, while maintaining low delay. However, this work has revealed limitations of the proposed MB-CELP scheme for coding transient music signals. We expect that, at rates below 24 kbps and a delay of 10 ms, an effective mix of techniques to exploit time-domain correlations and frequency-domain masking properties will lead to a coder satisfying the performance objectives.

5. REFERENCES

- [1] E. Ordentlich, "Low Delay Code Excited Linear Predictive (LD-CELP) Coding of Wide Band Speech at 32 Kbit/sec.", in *MS Thesis*, EE Dept., MIT, Mar. 1990.
- [2] A. Ubale and A. Gersho, "A Multi-band CELP Wideband Speech Coder", *To appear in Proc. of ICASSP*, April 1997.
- [3] ISO/IEC SC29/WG11 (MPEG) Committee Draft, Nov. 1993.
- [4] A. Crossman, "A Fixed Bit Rate Wideband Codec for use in ISDN Videotelephony", *IEEE Workshop on Speech Coding for Telecommunications*, Sept. 1995, Annapolis, USA.