A MULTI-BAND CELP WIDEBAND SPEECH CODER

Anil Ubale and Allen Gersho

Dept. of Electrical and Computer 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 speech signal over the full frequency range. A 16 kbps version of MB-CELP coder with two equal bands, is described in this paper. Subjective comparison test results show that this coder performs better than the G.722 coder at the bit-rate of 48 kbps.

1. INTRODUCTION

Wideband speech has a bandwidth of roughly 50 to 7000 Hz thereby allowing a richer, more natural, and more intelligible speech signal. Compression of wideband speech makes it possible to achieve low bit rates while obtaining a decompressed speech signal output that resemble the audio quality of AM radio, sometimes referred to as *commentary* grade quality, as opposed to the usual telephone audio quality. In subjective tests, the mean opinion score (MOS) of wideband speech is usually higher than that of telephone speech by as much as 1 MOS unit. People tend to experience less fatigue in communicating with wideband speech compared to telephone speech.

In 1986, the ITU-T (then CCITT) standardized a subband-ADPCM wideband speech coder (referred to as the G.722 standard) at 64, 56, and 48 kbps [1]. In February 1995, the ITU-T Study Group 15 formalized a wideband 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. Mode B requires a low complexity but allows a higher delay, which is acceptable for point-to-point links. The quality targets for the proposed coders are to achieve at 24 kbps a quality equivalent to that of G.722 at 56 kbps, and at 16 kbps to obtain a quality equivalent to that of G.722 at 48 kbps. Our MB-CELP coder satisfies some of the requirements needed for a new ITU-T standard for Mode A and is already well suited for some important applications. We are currently optimizing it for music, and if this work is successful, MB-CELP may be a worthwhile candidate for submission to the ITU-T.

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.

2. BACKGROUND AND INTRODUCTION TO MULTI-BAND CELP

Contemporary approaches to wideband speech coding are typically either CELP or transform-based coders. Transform coders generally achieve efficient compression by adaptive bit allocation and entropy coding. On the other hand, effective transform coding requires a relatively high delay. Recently the ITU-T adopted an 8 kbps algorithm[2] for narrowband speech coding, called G.729. The fact that the G.729 standard achieves "toll" quality at 8 kbps suggests that it should be possible to reconstruct high quality wideband speech (with twice the telephone bandwidth) at the bitrate of 16 kbps using CELP coding. These considerations prompted us to explore the possibility of obtaining a very high quality wideband speech coder at 16 kbps based on the CELP approach to speech coding.

A popular approach to wideband speech coding has been to take a state-of-the art narrowband CELP speech coder and modify and tune it for wideband speech. Traditionally, such wideband speech coders belong to two classes: fullband CELP [3], and split-band CELP. The fullband CELP usually has a higher complexity (more than double that of the corresponding narrowband speech coder) and suffers from an intermittent background hiss (high frequency noise) in the decoded

0-8186-7919-0/97 $10.00

speech [4].

A split-band CELP coder is usually of lower complexity than a fullband one (about 1.5 times as complex as the corresponding narrowband speech coder), but has extra algorithmic delay due to the analysis and synthesis filterbanks. Also, the split-band CELP suffers from bad quality due to degradations in the frequency range where the filter frequency responses for low and high band overlap.

In this paper, we shall describe our novel encoding scheme, MB-CELP, and show that it avoids both these artifacts namely, the high frequency hiss of fullband CELP coders and the bad speech quality in the filterbank overlap region of split-band CELP coders. This is achieved by off-line filtered multi-band excitation codebooks, fullband LPC synthesis, and error minimization over the original speech signal over the entire 8 kHz band.

Fullband CELP coded speech seems to match the original extremely well in the low-frequency region, but it fails to match the spectrum well in the high frequency region. This is explained by the fact that the short-term frequency spectrum of wideband speech has a strong downward spectral tilt with a drop as much as 35 dB in the average energy of the high band compared to the low band. However, the perceptual importance is not reduced by the same proportion. Hence, unless this perceptual asymmetry in coding low and high band is accounted for, the fullband CELP coded speech suffers from a high frequency hiss distortion.

One solution to this problem is to design the perceptual weighting filter to give more weight to the error in high frequency region [4]. Another solution is to use a split-band CELP coder to allow better control of the high band SNR. This also offers a complexity advantage [5]. However, the speech quality suffers in the range of frequencies where the two frequency bands overlap. To minimize overlap, the filterbank frequency responses need to have a sharp cut-off, and hence the filters have to be of high order. This implies that the algorithmic delay of the coder increases. Thus, it is difficult to achieve good speech quality at low delay using split-band CELP.

MB-CELP does not suffer from the above drawbacks - namely, high frequency hiss, bad coding of the filterbank overlap region. It also has low delay and moderate complexity.

As shown in Figure 1 , MB-CELP coder uses the split-band excitation with the use of off-line filtered stochastic codebooks. The multi-band codebook sizes can be dynamically tailored in accordance with the perceptual importance of the frequency bands. The linear prediction analysis and synthesis is done over the full band. Further the error is minimized between the original fullband signal and the synthesized signal. 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 region of the frequency bands that correspond to the multi-band codebooks. Modeling the periodicity in speech signal accurately is very important for the overall speech quality of the coder. With the multiband structure, the pitch periodicity can be modeled independently in each of the frequency bands. Thus the general MB-CELP structure has an adaptive codebook for each frequency band. This allows the tracking of small variations in pitch periodicity with frequency, and also provides a frequency-dependent gain for the long-term prediction. However this also implies an increase in bitrate is needed to transmit the pitch information for each band. Hence, at low bitrates (16 kbps or lower), a single fullband adaptive codebook can be used.

In the following sections, we will describe a specific MB-CELP coder configuration which operates at the rate of 16 kbps and has only two bands. 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 $A(z/\gamma_1)/A(z/\gamma_2)$.

3. LP-ANALYSIS AND CODING

Short-term linear prediction (LP) analysis is performed once per input frame using the autocorrelation method with a 10 ms Hamming window. The autocorrelation values of the windowed speech are computed and a bandwidth expansion of 10 Hz is introduced by lag-windowing the autocorrelations. The LP coefficients are converted to line spectral frequencies (LSFs) and quantized using 28 bits for each frame. The LSFs are quantized using predictive multi-stage vector quantization. We use a second-order moving-average interframe predictor. The prediction residual is coded using multi-stage vector quantization (MSVQ) with 7 stages of 4 bits each. The distortion measure employed for predictive multi-stage vector quantization is a weighted mean-square-error (WMSE). The weights are proportional to the distance between the neighboring LSFs.

The multi-stage vector quantization scheme uses a multiple-survivor method for an effective trade-off between complexity and performance. Four residual survivors are retained from each stage and are tested by the next stage. The final quantization decision is made

at the last stage, and a backward search is conducted to determine the entries in all stages. The multi-stage vector quantization is design by a joint optimization procedure [6].

The LP coefficients are computed once per frame with the LP analysis window situated at the center of the fourth subframe. These LP coefficients are directly applied to the fourth subframe of each frame. For the first three subframes, the quantized LP coefficients are obtained by a linear interpolation of the corresponding parameters in the adjacent frames. Similarly, unquantized LP parameters are used for perceptual weighting in the fourth subframe of each frame. For the first three subframes, a perceptual weighting filter is derived by interpolating between the unquantized LSFs of adjacent frames. The perceptual weighting filter's parameters were tuned to the values of $\gamma_1 = 1.0$ and $\gamma_2 = 0.65$.

4. PITCH ANALYSIS AND CODING

Pitch prediction is implemented using the adaptive codebook method where pitch delays greater than the subframe length are searched in the range of (41-296) samples. Fractional pitch delays with nonuniform spacing are quantized using 9 bits per subframe. The highest resolution for pitch delay is equal to 1/4 of a sample. For the selected pitch delay, the pitch gain is closed-loop scalar quantized using 4 bits.

5. MULTI-BAND EXCITATION

The fixed (non-adaptive) codebook excitation is generated from two filtered codebooks with spectral bands from 0-4 kHz and 4-8 kHz. It must be noted that since the filtered codevectors are of finite duration (equal to the subframe length) there is frequency leakage between the low- and high-band codebooks. However, this is not a critical issue, since the error minimization is done over the fullband speech. The sizes of the multi-band codebooks can be dynamically selected for each frame. Such dynamic codebook-size allocation can be derived using a perceptual criterion based on a psychoacoustic model. With dynamic codebook-size allocation, multi-band CELP can prove to be an important paradigm for low-delay coding of input signals whose spectrum in different frequency bands varies greatly with time, e.g., music. In the case of wideband speech, we observed that the signal strength in each of the two bands varies relatively little with time. Therefore, a fixed codebook size allocation can be used with the MB-CELP algorithm for coding wideband speech. It should also be noted that with dynamic codebook-size allocation, side information specifying the bit allocation must either be transmitted to the decoder, or must be derived at both the encoder and decoder from other available coded parameters. After experimenting with different codebook sizes, we chose a lower band codebook size of 256 and a higher band codebook size of 64 for this particular coder. The excitation search is similar to ordinary multi-stage vector quantization and offers low complexity and high robustness. The high-band codebook search is carried out first and then the low-band codebook search. 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.

Parameters	Bits per Frame
LSP	28
Pitch Delay	36
Pitch Gain	16
Low-band Codebook Index	32
High-band Codebook Index	24
Multi-band Codebook Gains	24
Total	160

Table 1: Bit allocation

The fixed codebook gains are computed, and are vector quantized using predictive vector quantization with 6 bits. This quantization scheme is very similar to that in G.729 standard [7]. An eighth order moving-average vector predictor with fixed coefficients is used. The predictor predicts the energy of the fixed excitation contribution based on the sequence of previously selected fixed excitation vectors. The quantized gain is expressed as a product of the predicted gain based on previous fixed excitation codebook energies and a correction factor. The correction factors for both bands are vector quantized using 6 bits.

The bit allocation is summarized in Table 1.

6. MB-CELP DECODER

Figure 2 shows an MB-CELP decoder with a single fullband adaptive codebook. The function of the decoder is to unpack the encoded bit stream, and decode the parameters. Namely, LP parameters, pitch delay, pitch gain, MB codebook indices, and codebook gains. These parameters are used to compute the reconstructed speech signal. The quality of this reconstructed speech signal is extremely good, as the subjective test results of the next section indicate. Adaptive

postfiltering [8] has been found to be very useful for enhancing the speech quality for CELP based coders, particularly when multiple tandem encodings are considered. The adaptive postfilter we use is a cascade of three filters: a pitch postfilter, a short-term postfilter, and a tilt compensation filter. The postfiltering process also includes an adaptive gain control scheme. Presently, the postfilter is tuned for a single encoding of clean speech signals. With the postfilter, the quality of the MB-CELP coded speech at 16 kbps is very close to that of the original speech signal.

7. RESULTS

We conducted subjective tests of the MB-CELP wideband speech coder at 16 kbps (without postfiltering), and the G.722 coder at 48 kbps. The test was performed using forced-choice A/B pairwise comparisons with 9 male and 9 female listeners evaluating 12 sentence-pairs per listener. The listeners were unfamiliar with these coders. A practice session with 4 sentence pairs preceeded each test. The sentence-pairs were chosen according to the subjective test plan formalized by the ITU-T for the current wideband standardization program. Our coder at 16 kbps was preferred over the G.722 coder at 48 kbps 66.80% to 33.20%.

8. CONCLUSION

The new MB-CELP coder employs a multi-band bank of excitation codebooks, fullband LP synthesis, and error minimization over the fullband. It overcomes the drawbacks of conventional fullband and split-band CELP coders, while maintaining a low-delay. Listening tests show that a 2-band 16 kbps version of this coder performs better than the G.722 coder at the much higher bit-rate of 48 kbps.

9. REFERENCES

[1] X. Maitre,"7 kHz audio coding within 64 kbit/s", *IEEE Journal on Selected Areas in Comm.*, vol. 6, No. 2, pp. 283-298, Feb. 1988.

[2] R. Salami et. al., "Description of the proposed ITU-T 8-kb/s speech coding standard", *IEEE Speech Coding Workshop*, Annapolis, 1995.

[3] E. Harborg, J. E. Knudsen, A. Fuldseth and F. T. Johansen, "A Real-time Wideband CELP Coder for a Videophone Application", *Proc. of ICASSP, 1994*, pp. II-121 - II-124.

[4] 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.

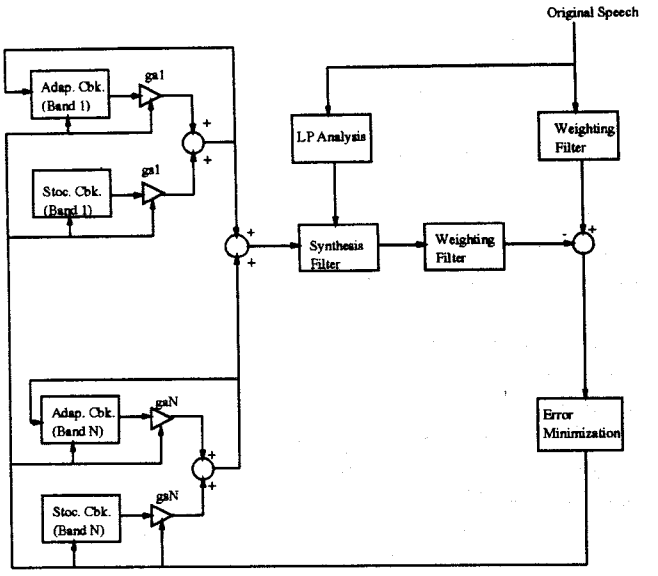

Figure 1: A general Multi-band CELP encoder structure.

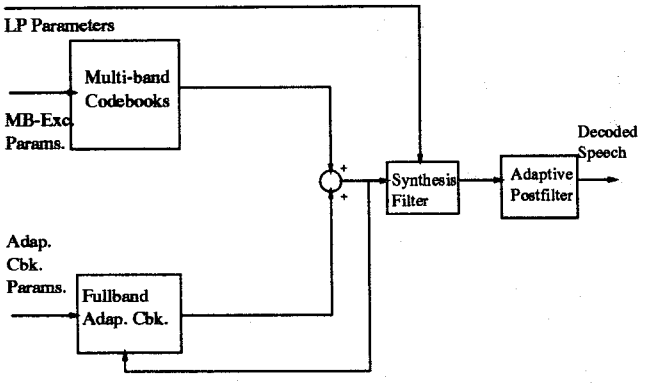

Figure 2: A Multi-band CELP decoder with a single fullband adaptive codebook.

[5] R. Drogo de Iacovo, R. Montagna, F. Perosino and D. Sereno, "Some Experiments of 7 kHz Audio Coding at 16 kbit/s", *Proc. of ICASSP, 1989*, pp. 192-195, Glasgow, May 1989.

[6] W. P. LeBlanc, B. Bhattacharya, S. A. Mahmoud and V. Cuperman, "Efficient Search and Design Procedure for Robust Multi-Stage VQ of LPC Parameters for 4 kb/s Speech Coding ", *IEEE Trans. Speech and Audio Processing*, vol. SAP-1, pp. 373-385, Oct. 1993.

[7] A. Kataoka, J. Ikedo and S. Hayashi, "LSP and Gain Quantization for the Proposed ITU-T 8-kb/s Speech Coding Standard", *IEEE Speech Coding Workshop*, Annapolis, pp. 7-8, 1995.

[8] J.-H. Chen and A. Gersho, "Adaptive Postfiltering for quality enhancement of coded speech", *IEEE Trans. Speech and Audio Processing*, vol. SAP-3(1), pp. 59-71, Jan. 1995.