A 1.6 KB/S MELP CODER FOR WIRELESS COMMUNICATIONS

Alan McCree and Juan Carlos De Martin

Corporate Research and Development, Texas Instruments, Dallas, Texas
E-mail: [mccree|demartin]@csc.ti.com

ABSTRACT

This paper describes our new Mixed Excitation Linear Predictive (MELP) coder designed for wireless applications. This new coder, through algorithmic improvements and enhanced quantization techniques, produces better speech quality at 1.6 kb/s than the new U.S. Federal Standard MELP coder at 2.4 kb/s. A key feature of the coder is a novel Line Spectral Frequencies (LSF) quantization scheme, requiring only 20 bits per frame. With channel coding, the new MELP coder at 3.1 kb/s is capable of maintaining good speech quality even in very degraded channels.

1. INTRODUCTION

The Mixed Excitation Linear Predictive (MELP) coder [1] was recently adopted as the new U.S. Federal Standard at 2.4 kb/s. Although 2.4 kb/s is generally considered to be a low bit rate, there are a number of applications where an even lower bit rate is necessary. One such application is wireless digital transmission of speech, where channels with poor signal-to-noise ratios require the insertion of a considerable amount of redundancy in order to preserve acceptable speech quality, thereby reducing the number of bits available to the source coder.

In this paper, we describe a MELP coder which requires only 1.6 kb/s and delivers speech quality superior to that of the Federal Standard at 2.4 kb/s for both clean and noisy speech. Properly protected with convolutional codes and with adequate handling of frame-erasures, our new MELP coder is capable of preserving the base quality even in 5% random errors.

2. CODER DESCRIPTION

The 1.6 kb/s MELP coder is similar to the new Federal Standard with three significant differences: model improvements, more efficient quantization, and channel coding. The overall bit rate when channel coding is employed is 3.1 kb/s.

2.1. Model Improvements

First, the pitch and voicing estimation has been improved, particularly for IRS-filtered input speech. Second, a noise suppression front-end has been added to improve performance in acoustic background noise. This noise suppression uses the smoothed spectral subtraction technique described in [2]. Finally, the frame size has been decreased from 22.5 to 20 ms, resulting in an overall increase in speech quality.

2.2. Quantization

The major bit rate reduction in the new MELP coder comes from the new LSF quantization scheme, which lowers the number of bits needed to represent the LPC filter from 25 to 20 bits, at no extra cost in terms of storage or complexity. More efficient quantization of pitch, voicing, and gain saves an additional four bits per frame. In order to reduce the overall data rate, the Fourier series magnitudes transmitted in the Federal Standard coder are eliminated, saving 8 bits per frame.

2.2.1. LSF Quantization

We have designed a 20-bit switched predictive quantization scheme with slightly better performance than the 25-bit quantizer used in the Federal Standard. Most of this efficiency improvement is due to the use of predictive quantization, but there is additional performance gain from using a theoretically optimal LSF weighting function.

We use closed-loop optimization of switched predictive multistage vector quantization (MSVQ) of the LSF's. For training, we use an extension of the iterative sequential MSVQ training procedure [3], in which we alternate between training the predictor coefficients given the codebook and training the codebook given the predictor coefficients. We use one bit to signal which of two codebook/predictor pairs is selected for the current frame, where the selection is based on best quantizer performance. We have a found a significant advantage to using two different codebooks rather than sharing a single codebook, without any increase in complexity compared to the non-predictive case. Since each of the two 4-stage, 19-bit MSVQ codebooks is less than half the size of the 25-bit non-predictive version, both the storage and search complexity are actually reduced in the new scheme, and we can increase the search depth of our M-best MSVQ search from $M = 8$ to $M = 12$ for equivalent complexity.

In addition, we use a new LSF weighting function to approximate the frequency-weighted spectral distortion (SD_{fw}) defined in [1]. We have previously found this perceptual weighting function based on the Bark scale to better predict listener preference in the MELP coder, and we have recently developed an LSF weighting function which optimizes this form of SD.

At high rates, the optimal LSF weighting to minimize unweighted SD is the sensitivity matrix of the LSF's [4]:

$$\left.\frac{\partial^2 SD(a(\omega), a(\bar{\omega}))}{\partial\omega_k \partial\omega_l}\right|_{\omega=\bar{\omega}} = 4\beta j_{\omega_k}^T R_A j_{\omega_l}$$

where j_{ω_k} is the kth column of the Jacobian matrix for the LSF's, R_A is the autocorrelation matrix of the impulse response of the

Quantizer	SD_{fw} (dB)	$> 2dB$ (percent)
25-bit	1.06	2.4
20-bit switched	1.04	2.1

Table 1: LSF quantizer performance for flat input speech.

LPC synthesis filter, and β is a scale factor. Using the principles of linear filtering, it is straightforward to show that the optimal LSF weighting for a perceptually-weighted form of SD can be computed by replacing the matrix R_A with R'_A, the autocorrelation matrix of the perceptually-weighted impulse response of the LPC filter. In practice, we use an 8th order all-pole model approximation to the Bark weighting function $W_B(f)$ in [1]. We find experimentally that this optimal weighting function results in a consistent but modest improvement in SD_{fw} over the empirical weighting described in [5], and an improvement of more than 0.05 dB compared to the power-weighted LSF distance used in the Federal Standard.

The weighted spectral distortion for the Federal Standard quantization and the switched-predictive version is shown in Table 1. This test set is flat input speech, as was the majority of the training material. For this material, the 20-bit quantizer performs slightly better than the 25-bit version. We have also observed that for severely filtered speech, which is not well represented in the training set, the switched-predictive scheme outperforms the non-predictive version. This suggests that the use of prediction reduces the sensitivity of the quantizer to mismatches between training and test sets due to filtering of the speech material.

2.2.2. *Quantization of Remaining Parameters*

We have found that, if the Fourier Series magnitudes are not employed, 6 bits are sufficient to quantize the pitch. Two more bits are freed by selecting the band-pass voicing information from a catalog of four possible patterns. The aperiodic flag is replaced by a functionally equivalent pitch contour perturbation technique, which does not require explicit transmission. The gain is quantized with a four-bit switched-predictor.

2.3. Channel Coding

Forward-error correction (FEC) codes are used to improve the performance in channel errors. Every 40 ms, two frames worth of data are grouped and encoded with a convolutional code of rate 3/5. Counting a 4-bit CRC protecting the most significant bits and a 6-bit tail, the overall bit rate on the channel is slightly less than 3.1 kb/s. At the receiver side, a Viterbi decoder accepts soft inputs from the demodulator and performs Maximum-Likelihood decoding. If the CRC signals an error, a frame erasure algorithm extrapolates reasonable values for the parameters of the current frame from the past history.

3. TEST RESULTS

We conducted forced choice A-B comparison tests with 102 sentence pairs, uttered by 10 different speakers, and with the Federal Standard at 2.4 kb/s as reference coder. 38 sentences consisted of clean speech, either flat or IRS filtered, while the rest was corrupted by different kinds of noise (traffic, office, babble and truck). The pairs were randomized and presented to five different listeners. Overall the new 1600 b/s coder was preferred over the Federal Standard, with a clear preference in five of the six test conditions. Only for clean flat speech was the Federal Standard preferred, probably due to the presence of the Fourier Series magnitudes.

Parameters	Bits
LSF's	20
Gain	4
Pitch and overall voicing	6
Bandpass voicing	2
Total bits / 20 ms	32

Table 2: 1.6 kb/s MELP coder bit allocation

We also informally assessed the performance of the system in channel errors. For 5% random errors on the channel, the FEC scheme described in Section 2.3 results in a post-decoding error rate of only $3 \cdot 10^{-4}$ and a frame-erasure rate of 0.25%. This results in virtually flawless performance. Even at error rates as high as 7%, the quality is quite good, with very few annoying artifacts in the output speech. Only at error rates approaching 9% is the performance seriously degraded, since the rate 3/5 coding begins to fail.

4. CONCLUSIONS

We have presented a new MELP coder which, through model and quantization improvements, outperforms the new Federal Standard while requiring only 2/3 of the bit rate, making it an attractive candidate for wireless communications and other low data rate applications. A channel coding scheme based on convolutional codes preserves output quality even under very degraded channel conditions.

5. REFERENCES

[1] A. McCree, K. Truong, E. B. George, T. P. Barnwell, and V. Viswanathan, "A 2.4 kbit/s MELP Coder Candidate for the New U.S. Federal Standard," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, pp. 200–203, May 1996.

[2] L. Arslan, A. McCree, and V. Viswanathan, "New Methods for Adaptive Noise Suppression," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, pp. 812–815, May 1995.

[3] W. P. LeBlanc, B. Bhattacharya, S. A. Mahmoud, and V. Cuperman, "Efficient Search and Design Procedures for Robust Multi-stage VQ of LPC Parameters for 4 kb/s Speech Coding," *IEEE Transactions on Speech and Audio Processing*, vol. 1, pp. 373–385, October 1993.

[4] W. Gardner and B. Rao, "Theoretical Analysis of the High-Rate Vector Quantization of LPC Parameters," *IEEE Transactions on Speech and Audio Processing*, vol. 3, pp. 367–381, September 1995.

[5] R. P. Cohn and J. S. Collura, "Incorporating Perception into LSF Quantization – Some Experiments," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, April 1997.