Proc. IEEE Workshop Speech Coding (Whistler, BC), pp. 12-14, Sept. 1991

High Quality Low-Delay Speech Coding at 12 kb/s

John Grass [1], Peter Kabal [2,1], Majid Foodeei [2] and Paul Mermelstein [3,1,2]

[1] INRS-Télécommunications
Université du Québec
Verdun, Quebec
Canada H3E 1H6

[2] Electrical Engineering
McGill University
Montreal, Quebec
Canada H3A 2A7

[3] BNR
3 Place du Commerce
Verdun, Quebec
Canada H3E 1H6

1. Introduction

A low-delay CELP algorithm operating at 16 kb/s has been proposed for CCITT standardization [1, 2]. An alternate coding structure operating at the same rate is based on a ML-Tree algorithm [3]. Both algorithms offer near-network quality with coding delays below 2 ms at 16 kb/s. We evaluate the potential of these two basic coder structures to operate at the reduced rate of 12 kb/s while retaining high speech quality.

2. Low-Delay Block-Based Coding

The low-delay CELP algorithm was modified to operate at 12 kb/s. The bit-rate of the block-based coder is determined by the sampling rate multiplied by the codebook size (number of bits) and divided by the vector length used in the codebook. The sampling rate was kept fixed at 8 kHz. A number of different combinations of the parameters were examined. The best of these combinations was found to be a 9 bit codebook and a 6-sample vector size (which corresponds to a coder delay of 0.75 ms). The codebook design uses a full search approach rather than partitioning into shape/gain sub-codebooks. The codebook was retrained for the lower bit-rate.

The modified coder operating at 12 kb/s maintains good quality for female talkers but the quality degrades somewhat for male speakers. This difference can be attributed to the ability of the 50th order predictor (autocorrelation with analysis updated every 24 samples) to capture some aspects of pitch for female talkers but not for male talkers. Higher order predictors were studied by Foodeei and Kabal [4]. High order (up to 80) covariance analysis allows for the capture of pitch redundancies associated with male talkers. Furthermore, the Cumani algorithm provides a numerically stable algorithm for determining the coefficients of the high-order filter [5].

Using the covariance-lattice predictor in the block-based coder at 12 kb/s instead of the autocorrelation predictor, the quality of the male speech is improved. The covariance-lattice predictor has been shown to increase prediction gain over 2 dB for male speakers [4]. In the 12 kb/s coder, the overall performance of the coder in terms of SNR did not change. Perceptually however, the covariance-lattice technique provides improvements in the coder for male speakers.

3. Low-Delay Tree Coder

The ML-Tree algorithm was originally used in a configuration with a 3-tap pitch predictor. The adaptive predictor, with dynamic determination of the pitch lag, suf-

fers from error propagation effects. Using an 8th order formant predictor and a simple gain adjustment procedure, the ML-Tree coder at 16 kb/s has speech comparable to that for LD-CELP at the same bit rate [6].

At 16 kb/s, the Coding Tree has a branching factor of 4 at each sample (2 bits per sample). Our strategy to lower the bit rate is to use combined vector-tree coding. The encoding delay is a function of the path length and the number of samples populating each node. The overall bit-rate is then given by the sampling rate divided by the number of samples considered at each node and multiplied by the number of bits to represent the branching factor. Two configurations were studied, one using 3 bits for the branching factor and 2 samples per node while in the second configuration 6 bits are used for the branching factor and 4 samples per node. The former structure was preferred.

3.1 Prediction Filter

The original implementation of the low-delay tree coder uses the generalized predictive coder configuration [3]. In this structure, the reconstruction error is given by $R(z) = Q(z)\frac{1-N_1(z)}{1-F(z)}$. $F(z)$ is the predictor filter, $N_1(z)$ is the noise feedback function and $Q(z)$ is the quantization error. $N_1(z)$ is set equal to $F(z/\mu_1)$. The feedback filter in the this structure provides a method to shape the noise spectrum.

An alternative configuration of the generalized predictive coder structure is that given by Atal and Schroeder [7]. In this closed-loop structure, the perceptual weighting takes the same form as that used in the block-based coder; $W(z) = \frac{1-N_2(z)}{1-N_1(z)}$ where $N_1(z)$ is set equal to $F'(z/\mu_1)$ and $N_2(z)$ to $F'(z/\mu_2)$. The noise feedback filter is no longer directly linked to the prediction filter. The weighting filter can be determined from the clean input speech signal. Furthermore, the prediction filter and perceptual filter no longer need be of the same order. This configuration was implemented with a 10th order adaptive lattice predictor. The resulting speech was significantly better than that for the original configuration of the low-delay tree coder.

3.2 Gain Adapter

Several gain adaptation schemes were evaluated in the context of the low-delay tree coder. Particular attention was given to the adaptive logarithmic gain update strategy originally used in the 16 kb/s LD-CELP. It was found that the simple gain adaptation scheme proposed by Iyengar [3] achieved SNR results similar to the more complex gain adapters. Perceptually, a slight preference is given to the LD-CELP gain update method.

3.3 Dictionary Training

The dictionary for the innovation tree of the coder can be populated in a random fashion [3]. However, improvements as large as 1.5 dB in the performance of the coder at 12 kb/s were achieved by a new training procedure of the the dictionary (training speakers and sentences were different than those used for testing).

The training procedure used a random populated initial codebook. In an iterative process, the coder is run, accumulating the unquantized prediction errors (residuals) associated with each released node of the tree (corresponding to an entry in the

dictionary). Note that due to the delayed nature of the tree coder, the unquantized residuals must be retained for the length of the delay. Further, the gain value used at each node of the tree must be kept so as the unquantized residual can be appropriately scaled. The centroid for the dictionary entry is taken and used to re-populate the dictionary. With an updated dictionary, this process is repeated for several iterations.

4. Discussion

The speech quality for the block-based coder operating at 12 kb/s is remarkably good. The principle difference when compared to 16 kb/s LD-CELP is a modest degradation for some male speakers. In comparing the two coders at 12 kb/s, the low-delay tree coder is slightly better perceptually than the block-based coder.

We noted a significant improvement in the low-delay tree coder with the change to a generalized perceptual weighting, with the weighting filter determined from the clean speech rather than the reconstructed speech. Further work is warranted to compare the noise feedback as used in the tree coder with the open-loop weighting used in block based coders. In addition, the use of high order covariance-lattice predictors in tree coders needs further investigation.

Both types of coders have potential for high quality speech at 12 kb/s. This work is part of an on-going investigation of low-delay speech coders operating at bit-rates of 8–16 kb/s with high quality performance.

References

1. AT&T contributions to CCITT Study Group XV and T1Y1.2 (October 1988–July 1989).
2. J.-H. Chen, "High-quality 16 kb/s speech coding with a one-way delay less than 2 ms", *Proc. Int. Conf. on Acoust. Speech, Signal Processing*, (Albuquerque, NM), April 1990, pp. 453–456.
3. V. Iyengar and P. Kabal, "A low-delay 16 kbits/sec speech coder", *IEEE Trans. Signal Processing*, vol. 39, May 1991, pp. 1049–1057.
4. M. Foodeei and P. Kabal, "Backward adaptive prediction: high-order predictors and formant-pitch configuration", *Proc. Int. Conf. on Acoust. Speech, Signal Processing*, (Toronto, Canada), May 1991, pp. 2405–2408.
5. A. Cumani, "On a covariance lattice algorithm for linear prediction", *Proc. Int. Conf. on Acoust. Speech, Signal Processing*, (Paris, France), 1982, pp. 651–654.
6. M. Foodeei and P. Kabal, "Low-delay CELP and Tree coders: comparisons and performance improvements", *Proc. Int. Conf. on Acoust. Speech, Signal Processing*, (Toronto, Canada), May 1991, pp. 25–28.
7. B. S. Atal and M. R. Schroeder, "Predictive Coding of Speech Signals and Subjective Error Criteria", *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-27, June 1979, pp. 247–254.