An 800bps VQ Based LPC Voice Coder

Huan Chen, Xianglin Wang and C.-C. Jay Kuo

Electrical Engineering Department, University of Southern California, Los Angeles, 90089

Abstract: The idea of applying vector quantization (VQ) to linear prediction coding (LPC) to realize a very low bit-rate speech coder with reasonable speech quality is investigated in this research. Instead of encoding LPC parameters directly, line spectral frequency (LSF) coefficients are obtained through a transformation of LPC and then encoded via VQ in the proposed scheme. In order to utilize the inter-frame correlation, three frames of speech data are processed together. MSE is selected as the distortion measure in building the split VQ codebooks. With this new coder, the coding bit-rate can be reduced to 800 bps with quality similar to that obtained by the LPC 2.4kbps speech coder.

INTRODUCTION

With the increasing need on wireless communication, very low bit rate speech coding gains more attention nowadays. Low bit rate speech coding is important in wireless multimedia communication since only about a bandwidth of 12kbps is available for the transmission of all types of data. By considering synchronization and error protection, the bandwidth available for speech signals is even more limited. In this work, we modify the standard 2400bps LPC-10 coding algorithm (also named as FS1015)(1) by applying split VQ (Vector Quantization) to LPC coding so as to realize a very low bit-rate speech coder with reasonable speech quality.

CODING SCHEME DESCRIPTION

The 800bps voice coder is built upon 2400bps LPC algorithm. In the LPC-10 coding standard, speech is sampled at 8kHz. The sampled data are separated into frames with 180 samples each (corresponding to 22.5ms). Each frame is processed independently except for the smoothing of the pitch and voicing parameters. After LPC analysis, parameters to be encoded include 1 pitch, 1 voicing, 1 RMS and 10 reflection coefficients in each frame. All parameters are encoded via scalar quantization into 54 bits per frame (corresponding to 2400bit/s).

Effort can be traced back to 20 years ago (2)(3) or even earlier to further reduce the coding bit-rate while maintaining a comparable speech coding quality to that of LPC-10. In those coding algorithms, VQ was introduced as a very effective way to lower the entropy of LPC parameters. In addition, several frames are processed together so that inter-frame correlation could also be utilized (2). However, the LPC parameters were encoded directly and complicated error distortion measures were developed to match the characteristics of LPC reflection coefficients (4). An alternative coding strategy is to use the line spectrum frequency (LSF) (5) that could be considered as a frequency representation of LPC parameters (6). After transforming LPC into LSF, there is a strong correlation between LSF of adjacent frames as well as neighboring parameters in the same frame. Meanwhile, a simple distortion measures such as MSE could be directly applied.

Based on the above analysis, the following features are incorporated into the 800bps voice coder:

a) LPC-to-LSF transformation
b) Split vector quantization of LSF
c) Multi-frame processing

The encoding diagram is shown as in Figure 1. Since VQ is utilized and fast processing is desired, no entropy coder is required in this voice coder. LSF are encoded by VQ while other parameters such as voicing, pitch and RMS are encoded by scalar quantization.

Another advantage of using VQ to encode LSF is that it is possible to build separate VQ codebooks for different sets of parameters according to their different characteristics and significance in speech reconstruction. As a result, the VQ process can be speeded up since small codebooks are produced and the coder has a better error resilient property because of the use of shorter codewords. The error resilient property is specifically desirable in wireless
communication applications. In our work, codebooks are built separately for LSF in voiced and unvoiced frames. In voiced frames, all ten LSF components are encoded and separate codebooks are built for the first 4 LSF components and the last six LSF components. In unvoiced frames, only the first 4 LSF components are encoded. In all these cases, the length of codeword is limited to 7 bits (corresponding to a codebook size of 128).

To improve the coding efficiency of pitch, voicing and RMS, three frames are processed together. Therefore, there are three pitch, voice and RMS parameters, and predictive coding is adopted to encode these parameters. Let pitch\(_i\), voicing\(_i\), and RMS\(_i\) denote parameters for frame \(i\) (\(i=1,2,3\)), respectively. The bit allocation for these parameters is given below:

\[
\begin{align*}
\text{(a) When frame 1 is a voiced frame} & : & \text{Voicing}_1 & (1 \text{ bit}) & \text{Pitch}_1 & (6 \text{ bits}) & \text{Pitch}_2 - \text{Pitch}_1 & (3 \text{ bits}) & \text{Pitch}_3 - \text{Pitch}_1 & (3 \text{ bits}) \\
\text{(b) When frame 1 is an unvoiced frame} & : & \text{Voicing}_1 & (1 \text{ bit}) & \text{Pitch}_2 & (6 \text{ bits}) & & & \text{Pitch}_3 & (6 \text{ bits}) \\
\text{(c) } & & \text{RMS}_1 & (6 \text{ bits}) & \text{RMS}_2 - \text{RMS}_1 & (3 \text{ bits}) & \text{RMS}_3 - \text{RMS}_1 & (3 \text{ bits})
\end{align*}
\]

FIGURE 2. Bit allocation of pitch, voicing and RMS parameters

Differential values in Figure 2 are encoded through table look-up. In Figure 2(a), it should be mentioned that one index needs to be used specifically to indicate the case that frame 1 is voiced but frame 2 or 3 is unvoiced. Therefore, only seven indices are available for encoding the pitch differential values. According to such a bit allocation scheme, each frame encoded contains 18 bits on the average which is one third of the bits in each frame produced by the standard 2400bps LPC-10 algorithm.

This coder has been tested with both male and female voices. Although the coding bit rate is significantly reduced, the 800bps voice coder still achieve coding quality similar to that of the 2400bps LPC-10 algorithm, especially with male voices. Meanwhile, this coder has a similar computation complexity in comparison with that of the standard LPC-10 algorithm due to the small codebook size of split VQ.

REFERENCES