한국통신학회 1984년도 춘계학술발표회 논문집 (1984. 5. 25)

음성 신호에 대한 트리 코우딩

84305

°김 경 수 이 상 욱
서울대학교 공과대학 제어계측공학과

Tree Coding of Speech Signals

°Kyeong Soo Kim Sang Uk Lee
Dept. of Control and Instrumentation Eng., Seoul National University

Abstact

In this paper, the tree coding using the (M,L) multi-path search algorithm has been investigated. A hybrid adaptation scheme which employs a block adaptation as well as a sequential adaptation is described for application in quantization and compression of speech signals. Simulation results with the hybrid adaptation scheme indicate that a relatively good speech quality can be obtained at rate about 8 kbps. All necessary parameters such as M,L and filter-order were found from simulation and these parameters turned out to be a good compromise between the complexity and overall performance.

1. Introduction

Speech compression systems usually fall into one of two general classes : waveform coders and source coders. Waveform coders include traditional coding schemes such as PCM(pulse code modulation), DM(delta-modulation), DPCM(differential pulse code modulation) and more advanced scheme such as fixed or adaptive tree coding [1-3]. Source coders parameterize the speech production process and transmit a digitized representation of these parameters rather than a waveform itself. LPC(linear predictive coding) and formant type vocoders are good examples of the source coders.

Source coders operate at lower bit rates - typically provide a reasonable good quality of speech at about 2.4 kbps, but are far more complex than the waveform coders. But the waveform coders are generally more robust against speaker variation, background noise and channel errors. Recently a medium band coding, typically operated at 7.2-9.6 kbps, have received increasing attention. Hence, waveform coders are of interest when the available channel bit rate is adequate, although of course one still wishes to keep the bit rate as low as possible.

In this paper, attempts have been made to use recent techniques from source coding to develop a waveform coder that operates at relatively low bit rate of about one bit per sample or 8 kbps. It is believed that the bit rate of 8 kbps is effectively the lowest rate at which there exist intelligible waveform coders with fair subjective fidelity. when applying tree coding to speech at 8 kbps, it is necessary to adapt the tree coding process to the short-time changes in speech production. So far, block adaptation [3] or sample-by-sample adaption [2] has been used previously. But we shall consider a tree coding using block adaptation as well as sample-by-sample adaptation. A main motivation of our approach is the use of recently developed speech coding idea to design a code for medium to low rate speech compression. Results obtained by simulation of the tree coding algorithm are presented in section 4.

2. Tree Codes

Actually all waveform coders function sequentially, i.e., a sequence of decisions are made uniformly in time and according to same rules. A DM, for example, predicts according to a fixed rule and then quantizes the prediction error in a sequential manner. As such, the possible decision outcomes of a waveform coder can be graphically listed on a code tree like that of fig.2.1. On each brach lies a letter $\hat{x}$ representing a decision outcome. Since all decisions are made identically, a fixed number b of branches stem fram each node.

After each sample x_t arrived, the modulator angments a staircase approximation with a positive and negative step of fixed size(s=1). The example in the figure 2.1 shows that without exact decision levels, a binary sequence called a path map (+1,+1,-1,-1,-1) uniquely reconstructs the decoded values by integrating the map sequence in the decoder.

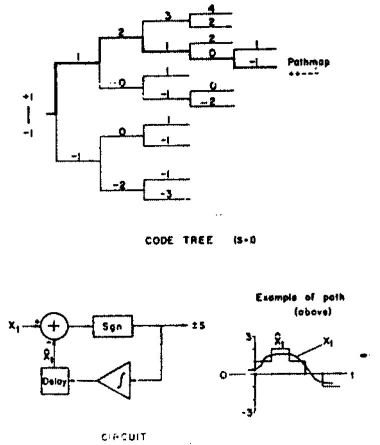

Fig. 2.1 Circuit, code tree, and example for Linear Delta Modulation(LDM)

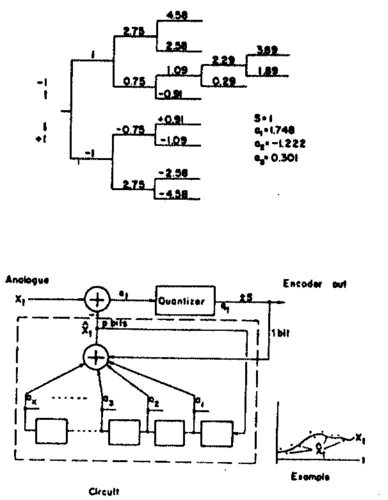

Fig. 2.2 Circuit, Code tree, and example for DPCM with k-tap predictor

The general convolutional tree code is shown in Fig. 2.2. Here $\hat{x}_t$ is generated at each time with the previous k samples. Sample-by-sample adaptation is achieved by the step size logic in the quantizer proposed by Jayant. For each node $b=2^l$ (l is number of bit in quantizer) branches stem out as a candidate for a reconstruction level. Hence we need an algorithm to select the best path, that is, a path which minimizes the error between the input sample and b-ary reconstruction levels. This algorithm is known to as search algorithm, and this is very important procedure for tree code to give a good quality.

The optimum choice of $\{a_i,\ i=1,2,\dots,k\}$ is equivalent to the predictor coefficients in DPCM [6] or speech analysis in LPC [4]. The mean square error (m.s.e.) of estimation is minimized [6] if $x=\sum_{i=1}^{k} a_i x_{t-i}$, where

$$A_{opt}=R^{-1}\ r \qquad (1)$$

at an estimation mse of

$$\sigma^2_{min}=1-A^T_{opt}r=1-r^T R^{-1} r \qquad (2)$$

where, r is the column vector of autocorrelation function $r(\tau)$, $\tau=0,1,\dots,k$ beginning with r(1), R is a matrix in the form of

$$R=\begin{bmatrix} 1 & r(1) & r(2)\dots\dots r(k-1) \\ r(1) & 1 & r(1)\dots\dots r(k-2) \\ r(2) & r(1) & 1 \qquad \vdots \\ \vdots & \vdots & \vdots \qquad \vdots \\ r(k-1) & r(k-2) & r(k-3)\cdots\cdots 1 \end{bmatrix} \qquad (3)$$

and A is a column vector of $\{a_i\}$.

In the block adaptation, A_{opt} is determined in a frame-to-frame basis and easily calculated using one of a variety techniques such as autocorrelation method or Levinson's algorithm, the covariance method and Burg algorithm.

3. (M.L) Search Algorithm

For the tree code to operate efficienly, there should be exist a good search algorithm. There exist many different search algorithms for a tree coding however, the most simple and efficient algorithm is believed to be the (M,L) algorithm proposed by Jayant [2]. This algorithm can be considered as a special case of Viterbi algorithm [8]. The Viterbi algorithm is optimum in the seme of searching but the computational burden is enormous.

The (M,L) algorithm is known as a very efficient and instrumentable multi-path tree search algorithm. In view of a multi-path search, the DPCM explained in section 2 can be considered as a single-path search case. The key property of a multi-path tree search is that the encoder, at a given sampling instant, does not make an irrevocable brach-decision, but keeps a multiplicity of possible branches in contention for a finite number of sampling instants : the idea of multi-path search is to obtain a better 'fit' to the input waveform than what is possible with the instantaneous decisions of a single-path search. Because of its delayed characteristic, multi-path tree search is also called as delayed encoding.

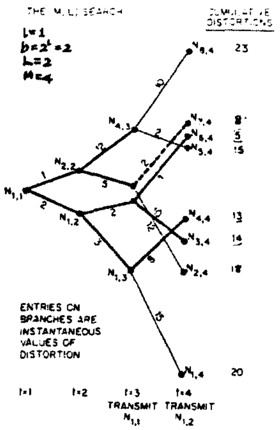

Fig. 3.1 Explanation of the (M,L) search for l=1

The (M,L) multi-path tree search can be described by two parameters. The first parameter is M, the maximum number of branches that can be kept in contention at a given point along the tree. (M=4 in Figure 3.1 : branches kept in contention are shown as heavy lines). The consequent branch selection is chosen by minimizing the cumulative mean square error difined in eq.(4)

$$E = \sum_{j=0}^{M-1} (\hat{x}_j - x_j)^2 \qquad (4)$$

where $\hat{x}_j$ is the reconstruction level, and x_j is the input speech sample at j-th time. The second parameter is the encoding delay L : at a given sample time, t, the encoder is forced to make a final decision on the node passed at time t-L, although up to M nodes might have been kept in contention (L=2 in Fig. 3.1 : for example, at t=4, the node at t=2 is obviously determined to be $N_{1,2}$). Note that the selection of $N_{1,2}$ can possibly invalidates some of the branches already in contention. ($N_{7,4}$ in the Figure 3.1)

4. Results of Computer Simulation

The speech used in simulation is 8KHz sampled and low pass filtered with the cut-off frequency of 3.2KHz. A signal-to-quantization ratio was used to evaluate the performance, and the cumulative error up to time t was used as an error criterion for branch selection of L-delayed node in multi-path search. The dynamic range of 40dB was chosen for adaptive quantization.

(1) Effect of block adaptation

Simulation results indicate that the block-length of 20-40 msec duration yields a good compromise between the rapid adaptation and the size of side information. The good code tree generation is a key proplem in multi-path search. Thus it is clear that higher order prediction is more powerful. In our simulation, the predictor coefficients were obtained using the autocorrelation method [5] and the order of 7 was found to be approximately saturating point in the sense of maximizing the SNR.

(2) Effect of M and L

Experiments have shown that as long as L is large enough to accomodate the delay dictated by the predictor used in the encoder, the actual value of L is not very critical to SNR. But longer L can search the optimum path better. If the delay L is too small, the gain in the multi-path search is alleviated. Therefore in case of 7-th order predictor, it is adequate to select L to 7 because the encoder need not have the additional memory for delay. The value of M was chosen 4. Throughout the simulation, we have found that this value is a good compromise between complexity and performance at given bit-rate requirement.

(3) Effect of Sequential adaptation

In a one bit quantizer we used an adaptation scheme originally proposed by Jayant. But we used slightly different parameters. The constant of 1.2 and 0.8 was found to provide better quality than that used the original Jayant's parameters (1.5 and 0.67) because of its multi-path nature. But it is believed that the overall performance is not significantly affected by changing of the adaptation parameters in the sequential adaptation.

5. Conclusion

Computer simulation has shown that M=4, L=7 is adequate, and 7-tap predictor is sufficient, and with these parameters 15-20dB SNR is obtained in voiced frames. This SNR is about 4-5dB improvement over the non-adaptive scheme.

It is intuitively sufficient to assign 7 bit for each coefficient of predictor. The side information is about 49 bits per each frame. When a frame length of 32 ms duration is used, and only b=2 branches per node(1 bit/sample) is allowed, the resulting total bit-rate is below 9.6kb/s.

Subjective test indicates that the tree coding based on the block and sequential adaptation provides relatively good speech but a slight background noise exists. Currently investigation is under way to improve the speech quality using application of tree coding to the residual signal and utilizing a noise shaping filter.

References

1. J.B. Anderson and J.B. Bodie, "Tree encoding of speech," IEEE Trans. Inform. Theory, vol. IT-20, pp.379-387, July 1975

2. N.S. Jayant and S.A. Christensen, "Tree-encoding of speech using the (M,L)-algorithm and adaptive quantization," IEEE Trans. Commun., vol.COM-26, pp.1376-1379, Sept. 1978

3. S.G. Wilson and S. Husain, "Adaptive tree encoding of speech at 8000 bits/s with a frequency weighted error criterion," IEEE Trans. Commun., vol.COM-27, pp.165-170, Jan. 1979

4. J.D. Markel and A.H. Gray, Linear prediction of Speech, Berlin : Springer, 1976

5. L.R. Rabiner and R.W. Schafer, Digital Processing of Speech Signals. Englewood Cliffs, NJ : Prentice Hall, 1978

6 J.B. O'Neal, and R.W. Stroh, "Differential PCM for speech and data signals," IEEE Trans. Commun., vol.COM-20, pp.900-912 Oct. 1972

7. J. Uddenfeldt and L.H. Zetterberg, "Algorithms for delayed encoding in delta modulation with speech-like signals," IEEE Trans. commun., vol.COM-22, pp.1195-1198, sept.1974

8. G.D. Forney, "The Viterbi algorithm," Proc. IEEE, vol.61, No.3, March 1973