ADAPTIVE TRANSFORM CODING OF SPEECH INCORPORATING TIME DOMAIN ALIASING CANCELLATION

A.W. Johnson and A.B. Bradley

Department of Communication and Electronic Engineering
Royal Melbourne Institute of Technology

ABSTRACT - A new transform coder that incorporates time domain aliasing cancellation is described. It contains 56 uniform 62.5 Hz channels that are adaptively quantized. Objective test results for this coder at bitrates of 16kbit/sec, 12kbit/sec and 9.6 kbit/sec are presented for two analysis/synthesis windows. Both windows are designed to meet the stated time domain aliasing cancellation constraints. The first window is a maximum length 128 point window. The second window is a 68 point trapezoidal window similar to that used in existing transform coders.

INTRODUCTION

Adaptive Transform Coding (ATC) involves block transformations of windowed input segments of a speech signal. The resulting transform coefficients are individually quantized according to the relative power of the transform coefficients to the overall block power. At the receiver the transform coefficients are decoded and an inverse block transformation performed to reconstruct the original block of windowed speech signal. This coder employs adaptive bit allocation in the frequency domain as the bits available for transmission are allocated to transform coefficients (or bands) on a block to block basis.

Present ATC designs employ no time or frequency domain aliasing cancellation. Because of this fact window designs are limited to being rectangular or trapezoidal with a small degree of overlap. This limitation in the window results in a wide channel bandwidth and hence a large amount of spectral interaction between bands.

In recent literature (Princen, J.P. & Bradley, A.B. 1986) a new filter-bank has been developed which generates both time and frequency domain aliasing in the analysis process and cancels the time domain aliasing in the synthesis process. The Time Domain Aliasing Cancellation (TDAC) filter-bank is the time domain dual of filter-banks that generate both time and frequency domain aliasing in the analysis process and cancel the frequency domain aliasing in the synthesis process. These frequency domain aliasing cancellation filter-banks (sometimes known as Quadrature Mirror Filters) are used in the coding technique known as subband coding.

The TDAC filter-bank offers advantages when implemented in an ATC. The restrictions on the window shape no longer applies as overlap between adjacent analysis/synthesis windows is permitted. This allows for a degree of flexibility in the design of the windows and therefore in the frequency response of the system's bands. This paper explores the application of TDAC filterbank to an adaptive transform coder.

DESCRIPTION OF THE CODER

The block diagram of the coder is shown in Figure 1. The incoming signal x(n) is speech, sampled at 8000 bits per second. The transform splits the spectrum of this signal into 64 unique bands.

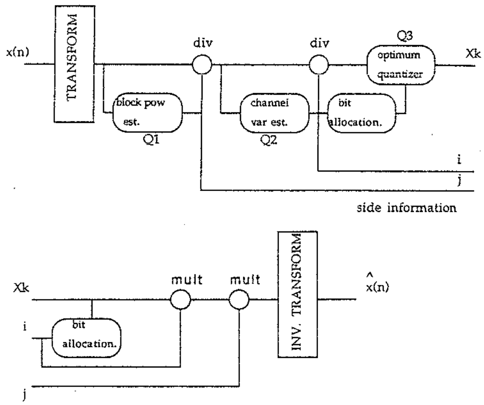

Figure 1 The block diagram of the TDAC ATC.

THE TRANSFORM

The transform operation for the TDAC coder can be developed from the single sideband analysis/synthesis filter-bank equations (Princen, J.P. & Bradley, A.B. 1986).

$$Xk\,ssb(m) = \mathrm{Re}\left[\sum_{n=0}^{K-1} h(mM-n)x(n)e^{-J2\pi(k+1/2)(n+nø)/K}\, e^{jm\pi/2}\right] \qquad 1$$

$$\hat{x}(n) = \mathrm{Re}\left[\sum_{m=-\infty}^{\infty} f(n-mM)\,\frac{1}{K}\sum_{k=0}^{K-1} Xk\,ssb(m)\, e^{J2\pi(k+1/2)(n+nø)/K}\, e^{-jm\pi/2}\right] \qquad 2$$

$$k = 0,1,2.\,.\,.\,.\,K-1$$

With an odd channel stacking arrangement there are K/2 independant channels having centre frequencies:

$$Wk = 2\pi(k+1/2)/K$$

$$k = 0,1,2.\,.\,.\,.\,K-1$$

This filter-bank is said to be critically sampled when the decimation factor M is equal to the number of independant channels. It can be shown (Princen, J.P. et al 1986) that the forward and inverse transform operations have the form:

$$yk = \sum_{n=0}^{K-1} y(n) \cos(2\pi(k+1/2)(n+n\varnothing)/K) \qquad 3$$

$$\hat{y}(n) = 1/K \sum_{k=0}^{K-1} yk \cos(2\pi(k+1/2)(n+n\varnothing)/K) \qquad 4$$

$$k = 0,1,2.\dots K\text{-}1$$

where y(n) represents a single block of K samples (x(n)) weighted by the analysis window. By substituting 3 into 4 it can be shown that:

$$\hat{y}(n) = 1/2K \left[\sum_{r=0}^{K-1} y(r) \left[\cos(\pi(r+n+2n\varnothing)/K) \sum_{k=0}^{K-1} \cos(2\pi(r+n+2n\varnothing)/K) + \cos(\pi(r-n)/K) \sum_{k=0}^{K-1} \cos(2\pi(r-n)/K) \right]\right] \qquad 5$$

Recognising that:

$$\sum_{k=0}^{K-1} \cos(2\pi n/K) = \begin{cases} K \text{ when } n = sK & s \text{ is an integer} \\ 0 \text{ otherwise} \end{cases}$$

then equation 5. reduces to:

$$\hat{y}(n) = \frac{(-1)^s\ y(sk+n)}{2} + \frac{(-1)^s\ y(sk-n-2n\varnothing)}{2} \qquad 6$$

If we place the following restrictions on the windows:

(i) $2n\emptyset = K/2 + 1$,
(ii) only adjacent windows are permitted to overlap,
(iii) adjacent windows overlap and add so the result is flat,
(iv) the windows are even and symmetric,

then the recovered signal y(n) over the interval r = 0..K-1 can be pictorially represented (Figure 2) as the sum of weighted terms of the original sequence x(n).

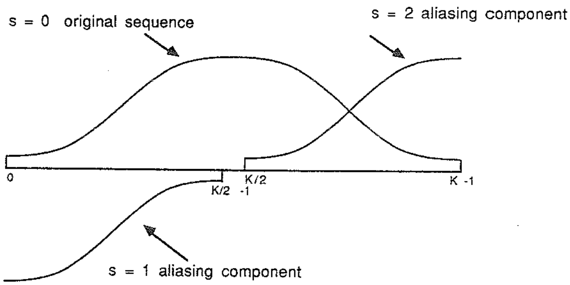

Figure 2 Weighting of x(n) in the recovered sequence.

The recovered sequence is the sum of the original sequence y(n) and aliasing components. When the signal y(n) is weighted by a synthesis window identical to the analysis window, the time domain aliasing cancellation property of this filterbank is evident. If the weighted recovered sequence of two adjacent block times, m and m+1 are overlapped and added then the aliasing component S=2 in block time m exactly cancels with the aliasing component S=1 in block time m+1.

CODING OF THE TRANSFORM COEFFICIENTS

The quantization is dynamic in blocktimes of 24 ms. Each block corresponds to 3 overlapped transform operations. Referring to Figure 1, there are 3 quantizers in the coder. Q1 is a logrithmic 4 bit quantizer - used to quantize the block power estimate. Q2 is a 3 bit logarithmic quantizer used to quantize the variance of the transform coefficients over the 24 ms block time. Q3 is an optimal quantizer (Max, J. 1960) used to quantize the transform coefficients.

To reduce the side information overhead, not all the variance information associated with the transform coefficients is transmitted to the decoder. Only the first 56 bands are considered for encoding and only 31 of these bands have their variance transmitted as side information. The variance of the remaining 25 bands is estimated at both the encoder and decoder by linearly interpolating the log quantized transmitted values. The selection of the 31 unequally spaced side frequencies follows the same procedure as the reference Tribolet, J.M. & Crochiere, R.M. (1979). The bit allocation alogorithm is the majority vote rule using the variance estimate for bands 1 to 56 (Ramstead, T.A. 1982). A maximum of 5 bits per sample and minimum of 0 bits per sample can be allocated when quantizing the transform coefficients.

CODER COMPARISONS

A Pascal program was written to simulate the TDAC ATC. Two windows were experimented with in the simulation. Both windows meet the previously defined requirements for time domain aliasing cancellation. The first window is a maximum length 128 point window optimized to achieve the best frequency response. The second window is a 68 point trapezoidal window similar to windows used in existing transform coders.

Simulations were performed at bit rates of 16 kbit/sec, 12 kbit/sec and 9.6 kbit/sec. Table 1 lists the objective signal to segmental noise ratios for the coder using the above mentioned windows. A comparison is also made with the 128 band speech specific adaptive transform coder (Tribolet, J.M. & Crochiere, R.M. 1979).

CODER	SIGNAL TO SEGMENTAL NOISE RATIO		
	9600 bits/sec	12000 bits/sec	16000 bits/sec
64 band TDAC ATC 128 point window	12.6 db	15.1 db	17.7 db
64 band TDAC ATC 68 point window	11.2 db	13.6 db	16.6 db
128 band speech specific ATC	12.8 db	14.5 db	18.0 db

Table 1 Comparison of objective performances.

The results of Table 1 indicate the maximum window length, 64 band TDAC ATC is objectively superior to the same coder employing the shorter trapezoidal window and comparable with the more complex speech specific ATC. This is due to the narrower bandwidth and therefore larger coding gain than can be achieved by using the longer window function.

CONCLUSION

In this paper a new technique of adaptive transform coding incorporating time domain aliasing cancellation has been described. Preliminary results indicate this 64 band coding scheme is comparable to more complex 128 band adaptive transform coders. Further improvements can be expected by reducing the side information overhead.

REFERENCES

CROCHIERE, R.E. & RABINER, L.R. (1983) "Multirate Digital Signal Processing", Ch 7 (Prentice Hall Inc.:New Jersey).

ESTEBAN, D. & GALAND, C. (1977) "Application of Quadrature Mirror Filters to Split Band Voice Coding Scheme", ICASSP-77, Hartford, Connecticut, pp 191-195.

MAX, J. (1960) "Quantizing for Minimum Distortion", IEEE Transactions Information Theory March 1960.

PRINCEN, J.P. & BRADLEY, A.B. (1986) "Analysis/Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-34, No. 4 (to be published).

PRINCEN, J.P., JOHNSON, A.W. & BRADLEY, A.B. (1986) "Subband/Transform Coding Using Filterbank Designs Based on Time Domain Aliasing Cancellation", submitted to ICASSP 1987.

RAMSTEAD, T.A. (1982) "Subband Coder with a Simple Adaptive Bit Allocation Algorithm. A Possible Candidate for Digital Mobile Telephoning", IEEE ASSP-1982 pp. 203-207.

TRIBOLET, J.M. & CROCHIERE, R.E. (1979) "Frequency Domain Coding of Speech", IEEE Transactions on Acoustics Speech and Signal Processing, Vol. ASSP-27, No.5, October 1979, pp 512-530.