THE EFFECT SPECTRAL MODIFICATIONS HAVE UPON THE PERFORMANCE OF FREQUENCY DOMAIN CODERS

A. W. Johnson * and A. B. Bradley ♠

* Telecom Australia Research Laboratories

♠ Department of Communication and Electrical Engineering
Royal Melbourne Institute of Technology

ABSTRACT - The effect of spectral modifications on the performance of low data rate frequency domain coders employing overlapped transform operations and the Ramstad bitallocation procedure is the introduction of distortion into the recovered signal. While this distortion cannot be removed certain aspects of it can be controlled. This theoretical analysis suggests the design of a new bitallocation procedure based upon the Ramstad majority vote rule [Ramstad 1984]. This is the subject of on going research which it is hoped will improve the overall subjective quality of speech recovered from a low data rate frequency domain coder.

INTRODUCTION

The long time averaged power spectral density of speech has a predominant low pass, non-flat spectral characteristic which frequency domain coders exploit to obtain data bit rate reductions in coding. The analysis operation of a frequency domain coder splits the sampled speech waveform into frequency components, or bands, and encodes the signal pertaining to each band using separate adaptive quantizers. The synthesis operation decodes the quantized data of each band and recombines the bands to reconstruct the full bandwidth discrete signal.

The technique of allocating bits to bands is integral to achieving high quality recovered speech in a low data rate frequency domain coder. Existing methods of determining the optimum bitallocation attempt to minimize the reconstruction error in a short time block for which speech can be assumed stationary. Whilst this produces the minimum distortion for frequency domain coders which employ a block transform operation (such as the Discrete Cosine Transform) to segment the speech spectrum into bands, error in the form of incomplete time domain aliasing cancellation will result if overlapping analysis/synthesis transforms are used to achieve spectral segmentation. Compared to block transform techniques, overlapping analysis/synthesis transforms are desired in frequency domain coders as the longer analysis/synthesis windows inherent with this technique produce less spectral interaction between bands of the system. This results in more efficient utilization of bits available for transmission. An understanding of the effect spectral modifications have upon the aliasing cancellation, and hence reconstruction properties of filter banks, is therefore an important consideration in the design of low data rate frequency domain coders.

FREQUENCY DOMAIN CODER FUNDAMENTALS

The block diagram of a forward adaptive frequency domain coder is shown in figure 1. The encoding operation takes a block of the discrete signal x(n) and segments this signal into frequency components using a transform/filter bank operation. Three quantizers are shown in the encoding operation. Q1 quantizes the block variance of the input signal. This quantized value of block power is transmitted to the decoder as side information and is used at the encoder to normalize the power of the frequency domain samples. Q2 quantizes the standard deviation of each band of the coder. This quantized value is transmitted to the decoder as side information and used at the encoder to normalise the signal of each band for that block time. The frequency domain samples are then assumed to have a Gaussian distribution with unit standard deviation. The quantized value of standard deviation is used to determine the bitallocation for that block being quantized. Quantization of the frequency domain coefficients is performed using quantizer Q3, an optimum quantizer [Max 1960].

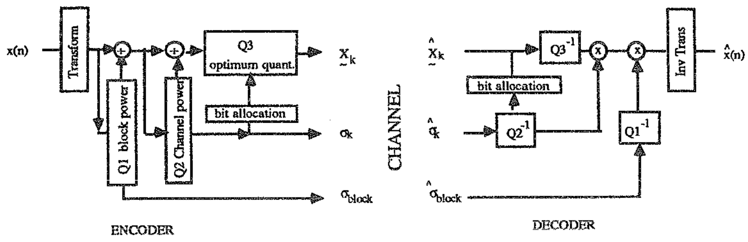

Figure 1 Block diagram of a frequency domain coder.

The bitallocation procedure generally used in the implementation of frequency domain coders is the Ramstad majority vote rule [Ramstad 1984]. This technique allocates bits to bands dependent upon the quantized values generated by Q2. It does not take into account the bitallocation generated for the quantization of the previous block. It attempts only to minimize the distortion resultant from the block being quantized. The identical bitallocation procedure is performed at the decoder so the quantized values of frequency domain samples can be deciphered from the transmitted data stream.

THE EFFECT SPECTRAL MODIFICATIONS HAVE UPON CODER PERFORMANCE

Subdividing the speech signal into spectral components can be achieved using a filter bank that incorporates aliasing cancellation. Time domain aliasing results from the use of analysis/synthesis windows which have duration longer than the number of unique bands the speech spectrum is segmented into. The analysis and synthesis equations for an oddly stacked single side band filter bank that incorporates time domain aliasing cancellation are shown in equations 1 and 2. [Crochiere and Rabiner 1984]

$$X_k(m) = \text{Real}\left[\sum_{n=0}^{K-1} h(P-1+mM-n)\, x(n)\, e^{\frac{-j2\pi(k+1/2)(n+n\o)}{K}}\, e^{\frac{+jm\pi}{2}}\right] \qquad 1$$

$$\hat{x}(n) = \text{Real}\left[\sum_{m=-\infty}^{+\infty} f(n-mM)\, \frac{1}{K}\sum_{k=0}^{K-1} X_k(m)\, e^{\frac{+j2\pi(k+1/2)(n+n\o)}{K}}\, e^{\frac{-jm\pi}{2}}\right] \qquad 2$$

The filter bank is said to be critically sampled when the decimation factor M is equal to the number of unique bands of spectral segmentation. The delay P introduced in the analysis equation is the duration of the analysis window h(n) and synthesis window f(n). The time shift nø is required to achieve aliasing cancellation. Complete aliasing cancellation can be achieved if appropriately designed analysis/synthesis windows are used and 2nø is equal K/2 + 1, where K/2 equals M

[Princen & Bradley 1986] [Princen Johnson & Bradley 1987]. The variable m refers to the particular transform operation being processed.

An understanding of the effect spectral modifications have upon the performance of a frequency domain coder can be achieved by examining the recovered sequence following an analysis and then synthesis operation. The analysis and synthesis equations can be converted from a fixed time frame to a sliding time frame by making the change of variable r = n - mM [Crochiere and Rabiner 1984] . The analysis equation rewritten for a sliding time frame has the form of equation 3.

$$X_k(m) = \sum_{r=0}^{K-1} (-1)^{mk} \; h(P-1-r)\; x_m(r) \cos\left[\frac{2\pi\,(k+1/2)\,(r+n\phi)}{K}\right] \qquad 3$$

where

$$x_m(r) = x(r+mM)$$

The recovered sequence following an analysis and then synthesis operation for block time m_0 has the form of equation 4. This is determined by first converting equation 2 to a sliding time frame, substituting $X_k(m)$ (equation 3) into equation 2 and then simplifying the resultant equation with the application of simple trigonometric functions.

$$\hat{x}_{m0}(r) = f(r)\; y_{m0}(r) \qquad 4$$

where

$$y_{m0}(r) = \frac{1}{2K}\sum_{p=0}^{K-1} h(P-1-r)\; x_{m0}(p)\left[\left[\cos\left[\frac{\pi(p+r+2n\phi)}{K}\right]\sum_{k=0}^{K-1}\cos\left[\frac{2\pi k(p+r+2n\phi)}{K}\right]\right] + \left[\cos\left[\frac{\pi(p-r)}{K}\right]\sum_{k=0}^{K-1}\cos\left[\frac{2\pi k(p-r)}{K}\right]\right]\right]$$

The sequence $y_{m0}(r)$ represents the recovered sequence following an analysis and then synthesis operation, but prior to synthesis filtering being performed. The sequence $y_{m0}(r)$ can be rewritten (equation 5) making it more convenient for gaining an understanding of the effect spectral modifications have upon the recovered signal .

$$y_{m0}(r) = \sum_{p=0}^{K-1} h(P-1-r)\; x_{m0}(r) \sum_{S=-\infty}^{\infty} \Omega(SK - r - 2n\phi - p) + \Omega(SK + r - p) \qquad 5$$

where

$$\Omega(x) = \frac{1}{2K}\cos\left[\frac{\pi x}{K}\right]\sum_{k=0}^{K-1}\cos\left[\frac{2\pi k x}{K}\right] \qquad 6$$

Inspection of equation 5 reveals that the sequence $y_{m0}(r)$ is the result of the summation of two convolution operations, these being:

$$y_{m0}(r) = \sum_{S=-\infty}^{\infty}\left[h(P-1-r)\; x_{m0}(r) \otimes \Omega(SK - 2n\phi - r) + h(P-1-r) \otimes \Omega(SK + r)\right] \qquad 7$$

$$r = 0..\, K-1$$

When all bands are made available to the synthesis operation, $\Omega(x)$ of equation 6 reduces to a periodic train of impulses having alternate polaritity. The sequence $y_{m0}(r)$ for this case reduces to the form of equation 8.

$$y_{m0}(r) = \sum_{S=-\infty}^{\infty} \frac{(-1)^S}{2}\; h(P-1-SK+2n\phi+r)\; x_{m0}(SK-2n\phi-r) + \frac{(-1)^S}{2}\; h(P-1-SK-r)\; x_{m0}(SK+r) \qquad 8$$

Equation 8 shows that the recovered signal prior to synthesis filtering for the case of all band data made available to the synthesis operation, consists of two periodic sequences having alternate polarity. The first component of equation 8 is the original data sequence windowed in the analysis operation, inverted in time. This component of $y_{mo}(r)$ represents time domain aliasing and must be cancelled in the process of summing overlapped transform operations to achieve a signal free of aliasing distortion. The second component of equation 8 is a forward in time component of the original windowed data sequence of the analysis operation.

Frequency domain coding of speech at low bit rates does not make available to the synthesis operation all bands of the speech spectrum. Bands containing a small percentage of the full bandwidth energy are not quantized in an effort to reduce the transmission data rate. The effect the removal of bands in the analysis operation has upon the recovered speech signal can be determined by including in the summation over k of equation 6 only those bands which are available to the synthesis operation. Table 1 lists the values of $\Omega(x)$ of equation 6 for an eight band (K = 16) filter bank, when all bands are not included in the summation operation. Of note with the data contained in table 1 is that :

1. As more bands are removed from the summation, the function element $\Omega(0)$ gets progressively smaller.

2 The value of $\Omega(0)$ is dependent only on how many bands are removed from the spectrum and not their location in the full bandwidth spectrum.

3. Removal of bands from the synthesis operation will result in the time expansion of the funtion $\Omega(x)$, the degree of expansion being dependent upon the number of bands removed and the position of these bands within the speech spectrum.

	$\Omega(x)$ for Bands Removed from synthesis operation					
	zero	6 and 7	4 and 5	3 and 4	5, 6 and 7	4, 5, 6, and 7
$\Omega(0)$	0.5000	+ 0.0375	+0.0375	+0.0375	+ 0.3125	+ 0.2500
$\Omega(1)$	0.0000	+ 0.1089	+0.0451	0.0000	+ 0.1423	+ 0.1541
$\Omega(2)$	0.0000	- 0.0697	+0.0674	+0.0985	- 0.0493	0.0000
$\Omega(3)$	0.0000	+ 0.0274	- 0.0663	0.0000	- 0.0149	- 0.0389
$\Omega(4)$	0.0000	0.0000	0.0000	- 0.0441	+ 0.0221	0.0000
$\Omega(5)$	0.0000	- 0.0082	+0.0198	0.0000	- 0.0044	+ 0.0116
$\Omega(6)$	0.0000	+ 0.0049	- 0.0049	+0.0070	- 0.0035	0.0000
$\Omega(7)$	0.0000	+ 0.0008	- 0.0003	0.0000	+ 0.0011	- 0.0012
$\Omega(8)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\Omega(9)$	0.0000	- 0.0008	+0.0003	0.0000	- 0.0011	+ 0.0012
$\Omega(10)$	0.0000	- 0.0049	+0.0049	- 0.0070	+ 0.0035	0.0000
$\Omega(11)$	0.0000	+ 0.0082	- 0.0198	0.0000	+ 0.0044	- 0.0116
$\Omega(12)$	0.0000	0.0000	0.0000	- 0.0441	- 0.0221	0.0000
$\Omega(13)$	0.0000	- 0.0274	+0.0663	0.0000	+ 0.0149	+ 0.0389
$\Omega(14)$	0.0000	+ 0.0697	- 0.0674	- 0.0985	+ 0.0493	0.0000
$\Omega(15)$	0.0000	- 0.1089	- 0.0451	0.0000	- 0.1423	- 0.1541

Table 1 Tabulated values of the function $\Omega(x)$ of equation six for an eight unique band filter bank for the case of bands removed in the summation operation.

A method of determining the effect spectral modifications have upon the recovered sequence is to consider the sequence $y_{mo}(r)$ determined in the presence of spectral modificatoins as the result of convolving the sequence $y_{mo}(r)$ determined in the absence of spectral modifications, with one period of the sequence $\Omega(x)$ determined for the particular spectral modifications performed. This is expressed mathematically as

$$y_{mo}^{*}(r) = y_{mo}(r) \otimes \Omega(r) \qquad 9$$

Figure 2 is a plot of equation 9 when bands six and seven have been removed from the synthesis operation. The sequence $\Omega(x)$ for this figure is column 2 of table 1. For the sake of clarity, only values of $|\Omega(x)|$ greater than 0.05 have been included.

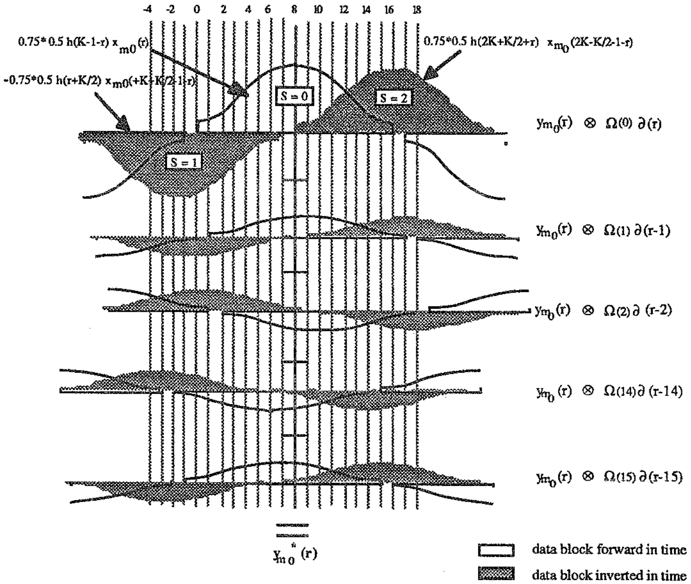

Figure 2 Components making up the sequence $y_{mo}^{*}(r)$ of an eight unique band filter bank having bands six and seven made unavailable to the synthesis operation.

Figure 2 shows that the effect of spectral modifications on the recovered sequence $y_{mo}^{*}(r)$, is the introduction of delayed and weighted components of the sequence $y_{mo}(r)$ of equation 8. The delay of each component of $y_{mo}^{*}(r)$ corresponds to the value of x for $\Omega(x)$, while the weighting to each component corresponds to the actual value of $\Omega(x)$. The components of $y_{mo}^{*}(r)$, corresponding to values of $\Omega(x)$ for x equalling one to K - 1, introduce aliasing distortion in the recovered sequence which can not be removed. This is because the antisymmetry of the time inverted components of $y_{mo}(r) \otimes (\Omega(x)\, \partial(r\text{-}x))$ of $y_{mo}^{*}(r)$ for x equals 1 to K - 1 does not occur about the midway point of the analysis/synthesis window. This requirement is necessary to achieve aliasing cancellation [Princen, Johnson and Bradley 1987]. When the component of $y_{mo}^{*}(r)$ corresponding to $y_{mo}(r) \otimes (\Omega(0)\, \partial(r))$ is considered, it is seen that the time inverted part is antisymmetric about the midway point of the analysis/synthesis window. If the value of $\Omega(0)$ for adjacent transform operations is equivalent, then the time inverted part of $y_{mo}(r) \otimes (\Omega(0)\, \partial(r))$ will cancel exactly with their time inverted counterparts of previous and subsequent transform operations. If the value of $\Omega(0)$ for adjacent transform operations differs, which will be the case when the bitallocation procedure results in the adjacent transform operations having a different number of bands allocated bits, then complete aliasing cancellation will no longer result. As the difference between $\Omega_{mo}(0)$ and $\Omega_{mo+1}(0)$ increases, the distortion in the recovered signal will also increase due to the lack of aliasing cancellation.

AN ALTERNATIVE BITALLOCATION PROCEDURE

For a speech signal which experiences rapid changes in the distribution of energy in the frequency domain, the allocation of bits to bands can result in large differences in the number of bands allocated bits between adjacent transform operations. This applies when the Ramstad majority vote rule is implemented for the bitallocation procedure. Whilst the Ramstad majority vote rule minimizes distortion for a particular block being processed, it does not attempt to minimize distortion in the process of summing the result of overlapped synthesis transform operations. Poor performance of low data rate frequency domain coders employing overlapped transform operations and the Ramstad bitallocation procedure will therefore result for segments of speech which exhibit rapid spectral variations. Subjective testing of speech processed by low data rate frequency domain coders, indicates that the subjective performance is determined not by the good quality of well coded segments of speech, but rather by the poor quality of poorly coded segments, even if their rate of occurrance is fewer in number. To improve the subjective quality of the recovered speech, an alternative bitallocation procedure is therefore suggested which first attempts to minimize distortion for a particular block being processed, and then attempts to maintain a smooth transition between blocks processed, on the number of bands of the speech spectrum which are allocated bits for quantization. A smooth change in the number of bands encoded will result in a smooth change to the value $\Omega_m(0)$ and therefore better aliasing cancellation properties at the overlapping block boundary. The bitallocation procedure should therefore incorporate some knowlege on the number of bands encoded in the previous block.

CONCLUSION

This theoretical investigation of the effect spectral modifications have upon the performance of frequency domain coders suggests the use of an alternative bitallocation procedure to the Ramstad technique, when aliasing cancellation filter banks which incorporate overlapping analysis /synthesis windows are used. This new technique of bitallocation should result in a subjective improvement in the recovered speech of low data rate coders. This is due to more complete aliasing cancellation which will result when the speech spectrum experiences rapid changes in the distribution of energy in the frequency domain. The degree of performance improvement is the subject of current and on going research.

REFERENCES

Crochiere, R.E. and Rabiner, L.R. (1983), *Multi-rate Digital Signal Processing,* Prentice Hall, Englewood Cliffs, NJ.

Max, J. (1960), "Quantizing for minimum distortion," *IRE Transactions on Information Theory,* Vol IT-6, pp. 7-12, March 1960.

Princen, J.P. and Bradley, A.B.(1986), "Analysis/synthesis filter bank design based on time domain aliasing cancellation," *IEEE Transaction on Acoustics Speech and Signal Processing,* Vol. ASSP-34, pp 1153-1161, October 1986.

Princen, J.P., Johnson, A.W. and Bradley, A.B.(1987), "Subband/transform coding using filter bank design based upon time domain aliasing cancellation," *Proceedings 1987 IEEE International Conference Acoustics Speech and Signal Processing,* pp.2161-2164.

Ramstad, T.A. (1982), "Subband coder with a simple adaptive bitallocation algorithm. A possible candidate for digital mobile telephony?," *Proceedings 1982 IEEE International Conference Acoustics Speech and Signal Processing,* pp.203-207.