TREE-STRUCTURED SUBBAND CODING OF SPEECH WITH 2 DIFFERENT FILTER BANKS

Selma Özaydın[1] and Gözde Bozdağı[2]

[1]Department of Electronics Engineering, Gazi University,Ankara 06570 TURKEY
[2]Department of Electronics Engineering, Başkent University,Ankara 06530 TURKEY
email: ozaydin@mikasa.mmf.gazi.edu.tr
bozdagi@baskent.edu.tr

ABSTRACT

Subband coding (SBC) is a well-known method for digital speech coding at medium rates (e.g. 16 kbit/s). In the basic sub-band coding procedure, the speech is first split into frequency bands using a bank of bandpass filters. The individual bandpass signals are then decimated and encoded for transmission. At the receiver, the channel signals are decoded, interpolated and recombined to form the received signal. A subband coder system can be divided into two distinct subsystem, which includes all of the operations involved in the encoding and decoding of the subband signals; and the analysis/synthesis filter banks along with the down-sampling and up-sampling operations. In this paper, a speech analysis/synthesis system has designed based on a 5-band SBC system using two different filters . A 32-tap Johnston and a 32-tap Barnwell filter structures have been used. The speech bandwidth allowed in this five band system includes 4 khz range (This is assumed for telephone speech). Finally, both filters have compared according to their performances and mean square errors (MSE) only from analysis/synthesis systems of two filter structures are computed and plotted.

1. INTRODUCTION

SBC is a popular technique for compressing speech, images and video signals. With this technique, the input signal is first split into several bandpass components using several filters called an *analysis filter bank*. The analysis filter bank typically contains a lowpass filter, bandpass filter, and a highpass filter that collectively cover the frequency spectrum without any spectral gaps. The outputs of the filters are downsampled to their respective Nyquist rates and the downsampled signals, known as *subbands*, are then coded for transmission or storage. The recontruction procedure, which is called synthesis, consists of upsampling the subbands, filtering them with an appropriate filter, and summing the results together. This part of the subband coding system is called the *synthesis filter bank*. [2,3]

The number of bits used to code each subband is varied based on a perceptual importance of that subband (For a speech signal, in the lower frequency bands, where pitch and formant structure must be accurately preserved, a larger number of bits/sample can be used; whereas in upper frequency bands where fricative and noise-like sounds occur in speech fewer bits/sample can be used). Furthermore, the overall reconstruction error spectrum can be controlled as a function of frequency with different quantization levels. It has been shown for audio and visual signals that coding subbands generally results in better performance than coding the input signal directly.

Inavitably, frequency components in one sub-band will leak into other sub-bands since the filters cannot have sharp cutoffs at the edges of the passband. This generally introduces *aliasing* effect [3-5] that grade the reconstructed signal even in the absence of quantization. In particular, the analysis-reconstruction subsystem can introduce three separate types of distortions: aliasing, short-time phase distortion, and short-time frequency distortion. Two common types of filter banks are :

a) *Uniform DFT filter bank*, and
b) The *QMF bank*.

The Discrete Fourier Transform (DFT) bank [3,4] can be implemented efficiently using FIR filters and the Weighted Overlap Add method (WOLA) scheme, in which case it is of much lower complexity than the QMF bank, for similar band seperation. However, because known design techniques aim at minimizing the overall response error of the filter bank, using other deterministic or statistical error measure, the performance of the DFT-based SBC (in terms of subjective quality) was found to be much lower.

The quadrature mirror filter bank (QMF) [1-6] is designed to completely cancel the aliasing due to the decimation of the band signals (in the absence of quantization), and it is widely used in SBC. The use of QMF banks allows the aliasing to be removed in the reconstruction stage by achieving nearly *Perfect Reconstruction* (PR) [2,5,6]. The QMF based SBC obtains good quality at medium bit rates. Its drawback, however, is its relatively high implementation complexity.

2. DOWNSAMPLING & UPSAMPLING

The idea behind the filter bank theory is to separate the incoming signal (x) into frequency bands. Often it is used a bank of two filters. A lowpass filter C and a highpass filter D will split the signal into two parts. Those parts can be compressed and coded seperately (and efficiently). They can be transmitted and the signal can be recovered. When the synthesis bank recovers the input signal exactly (apart from a delay), it is a PR filter bank.

The outputs of Cx and Dx are **downsampled** [3,6] removing every other component. This operation is called *decimation*. Downsampling is represented by the symbol (↓). If y downsampled with 2, this operation is not invertible. Most vectors y cannot be recovered from (↓2)y. The odd-numbered components are lost. Recovery of x from (↓2)x is possible if the transform X(w) is zero over a half band of frequencies. Such a signal is "band-limited". It may be limited to the upper half band or the lower half band :

$$X(w)=0 \text{ for } 0\le|w|\le\frac{\pi}{2}, \text{ or } X(w)=0 \text{ for } \frac{\pi}{2}\le|w|\le\pi$$

Downsampling in the time domain is y = (↓2)x means y(k) = x(2k). The fourier transform of y is,

$$Y(w)=\sum x(2k)e^{-jkw}=\frac{1}{2}\left[x(\frac{w}{2})+x(\frac{w}{2}+\pi)\right]$$

For band limited signal we can recover the odd-numbered components from the even-numbered components:

Upsampling places zeros into the odd-numbered components. Its symbol is (↑). After downsampling aliasing can be occured. If x(w) is bandlimited to IwI < π/2, aliasing is prevented and original x(n) can be recovered from (↓2)x. Upsampling in the time domain is:

$$u=(\uparrow 2)y=\begin{cases}u(2k) &= y(k)\\ u(2k+1)=0\end{cases}$$

The fourier transform of u = (↑2)y is U(w) = Y(2w)

$$U(w)=\sum u(2k)e^{-j2wk}=\sum y(k)e^{-j2wk}=Y(2w)$$

where Y(w) has period 2π, the new function Y(2w) has period π. Upsampling creates an *imaging* effect! [3-5] This imaging is the opposite of aliasing. In aliasing, two input frequencies w and w+π give the same output. Upsampling produces imaging where downsampling produces aliasing. The transform of,

u = (↑2)y = (↓2) (↑2)x is
U(w) = (1/2).[X(w) + X(w+π)].

The aliasing X(w+π) comes from $(-1)^n x(n)$. When n is odd, this cancels x(n). In the frequency domain the proof has two steps:

down : V(w) = (1/2).[X(w/2)+x(w/2+π)]
then up : U(w) = V(2w) = (1/2).[X(w)+x(w+π)]

Conclusion : (↓2) (↑2) produces aliasing and also imaging. Figure 1 shows them both: the image of the alias overlaps the original. We can't determine the original X(w) from U(w).

3. QMF BANK (Perfect Reconstruction for SBC)

In a two channel filter bank, the analysis filters are normally lowpass and highpass [1-6]. Those are filters H0 and H1 at the start of Fig.2.

Because H0 and H1 are not ideal filters, there is aliasing in each channel. There is also amplitude and phase distortion. The synthesis filters F0 and F1 must be specially adapted to the analysis filters H0 and H1, in order to cancel the errors in this analysis bank.

For the PR, filter bank must be *biorthogonal* [3,6]. The synthesis filter bank, from F0 and F1 and (↑2), is the inverse of the analysis bank. Inverse matrices automatically involve biortogonality. The transform of (↓2)(↑2)H0.x and (↓2)(↑2)H1.x are:

First (↓2) produces

$Y0(z) = (1/2).[H0(z^{1/2}).X(z^{1/2}) + H0(-z^{1/2}).X(-z^{1/2})]$

$Y1(z) = (1/2).[H1(z^{1/2}).X(z^{1/2}) + H1(-z^{1/2}).X(-z^{1/2})]$

then (↑2) produces :

$$\hat{X}(z) = (1/2).[H0(z).F0(z) + H1(z).F1(z)].X(z) + (1/2).[H0(-z).F0(z) + H1(-z).F1(z)].X(-z)$$

The alasing term H0(-z)X(-z) is multiplied by F0(z) at the synthesis step. This alias has to cancel the alias F1(z)H1(-z)X(-z) from the other channel. So there is an alias cancelation condition in addition to a reconstruction condition and alias term must be zero :

Alias Cancelation : F0(z)H0(-z) + F1(z).H1(-z) = 0 (1)

In addition, for PR with l time delays, $\hat{X}(z)$ must be $z^{-l}X(z)$. So the "distortion term" must be z^{-l} :

No distortion : $F0(z)H0(z) + F1(z)H1(z) = 2z^{-l}$ (2)

Finally, A 2-channel filter bank gives PR when, in matrix notation,

$$\left[F0(z)\ F1(z)\right].\begin{bmatrix}H0(z) & H0(-z)\\ H1(z) & H1(-z)\end{bmatrix}=\left[2z^{-l}\ \ 0\right]$$

The use of FIR filters for QMF bank design, has the advantage of linear-phase characteristics which eliminate the problems of group delay distortions. This feature also allows the 2-band design of Fig 2 to be conveniently cascaded in tree structures without the need for phase compensation. However, effective FIR designs imply significant coding delays. For example, with a 32-tap design, the coding and encoding delays due to the first level of the QMF tree are 4 ms each, assuming 8 kHz input sampling; and subsequent levels of the QMF partition introduce corresponding additional delays.

3.1. Linear -Phase FIR QMF Bank.

One way of obtaining alias and imaging cancelation property in the QMF bank is to use filter *h0(n)* and *h1(n)*, which are respectively symmetrical and anti-symmetrical FIR designs with even numbers of taps, ie,

$h0[n] = h1[n]=0$ for $0 > n \geq N$;
$h0[n] = h0(N-1-n], \quad n = 0,1,\ldots,N/2-1$;
$h1[n] = -h1[N-1-n], \quad n = 0,1,\ldots,N/2-1$; (3)

In order to cancellation of aliasing effect, Esteban-Gland [7] chose alternating signs $H1(z) = H0(-z)$ .The resulting filter bank was called 'quadrature mirror filter' (QMF):

$h0(n) = h(n), \quad h1[n] = (-1)^n.h0[n] \quad n=0,1,\ldots,N-1$ (4)
$\Rightarrow H1(w) = H0(w-\pi) \Rightarrow H1(z) = H0(-z)$

The high pass response $|H1(e^{jw})|$ is a mirror image of the lowpass magnitude $|H0(e^{jw})|$ with respect to the middle frequency $\pi/2$ - the quadrature frequency.

Further, if the filter bank output y[n] is desired to be delayed replica of input x(n) (in the absence of channel errors), the filters hu[n] and hl[n] must also satisfy the condition:

$$|H0(e^{jw})|^2 + |H1(e^{jw})|^2 = 1 \quad (5)$$

This is simply the condition for an *allpass* [3-6] characteristic. If the allpass condition is included, the point of intersection of the two filter functions will be the -3dB point for each transfer function.

The filter requirement in (5) cannot be met exactly by the mirror image filters of (4) except when N=2 and when N approaches infinity. However, it can be very closely approximated for modest values of N. Filter designs which satisfy (3) and (4) and approximate the condition of (5) can be obtained with the aid of optimization algorithms.

For a Johnston QMF bank [5,11], filters are selected for alias cancellation in (1) as below:

For alias cancellation choose : $H0(z) = H(z)$
$H1(z) = H0(-z)$
$F0(z) = H1(-z)$, and
$F1(z) = -H0(-z)$

Define the product filter by $P0(z) = F0(z).H0(z)$ and then,

$P0(-z) = F0(-z).H0(-z) \Rightarrow$
$F0(z).H0(z) + F1(z).H1(z) = 2z^{-1}$ simplifies to,
$F0(z).H0(z) - F0(-z).H0(-z) = P0(z) - P0(-z) = 2z^{-1}$

The normalized product filter is (P(z) must be a "halfband filter") :

$P(z) = z^{l}.P0(z) \quad \Rightarrow \quad P(z) + P(-z) = 2$

This means that all even powers in P(z) are zero, except the constant term (which is 1). The odd powers cancels when P(z) combines with P(-z). Therefore the coefficients of odd powers in P(z) are design variables in 2-channel PR filter banks.

Johnston QMFs were earliest popular filters used in literature. These symmetrical filters constitute a non-PR-QMF bank. Their non-PR characteristics, particularly for longer duration filters, do not present practical significance for subband image coding.

Johnston designed several sets of QMFs based on filter lengths, transition bands, and stopband weighting parameters.

3.2. COF (Conjugate Quadrature Filter) bank (Halfband QMF Bank)

Smith and Barnwell [6,12] have shown that alias-free reconstruction using recursive and nonrecursive filters was possible where analysis / reconstruction subsystem had no frequency distortion or no phase distortion, but not both.In the development, the coefficient symmetry condition on the analysis filters is lifted, and exact reconstruction free of aliasing, phase distortion, and frequency distortion is shown to be possible using FIR filters. These filters are called "conjugate quadrature filters" or COF's [3,4,6].

Smith and Barnwell chose alternating flip (N is odd):

$H1(z) = -z^{-N}H0(-z^{-1})$.

This leads to orthogonal filter banks, when H0 is correctly chosen.(The Daubechies filters will fit this pattern.). For the alias cancellation :

The analysis filters are : $H_0(z) = H(z)$
$H_1(z) = -z^{-N}H_0(-z^{-1})$

The synthesis filters are :
$F0(z) = H1(-z) = z^{-N}H0(z^{-1})$ comes from, $(h(N), h(N-1), \ldots, h(0))$
$F1(z) = -H0(-z) = z^{-N}H1(z^{-1})$ comes from, $(-h(0), h(1), -h(2), \ldots, h(N))$

The synthesis matrices are the transposes of the analysis matrices. A shift by N delays make them causal. When we flip to get F0 and then alternate signs to get H1, we have the alternating flip from H0 to H1.

The alternating flip gives double-shift orthogonality between highpass and lowpass. As a result, when the design of H0 lead to PR in the alternating filter bank, it also leads to orthogonality.

With aliasing cancelled, we now look at the PR condition,

$F0(z)H0(z) + F1(z)H1(z) = 2z^{-1}$

The better choice is alternating flip. PR is definitely possible. Product filters H0F0 and H1F1 become,

$P0(z) = z^{-N}H0(z^{-1})H0(z)$ and $P1(z) = -z^{-N}H0(-z^{-1})H0(-z)$.

multiply by $z^{l}=z^{N}$ to center these filters. The normalized product filter is P(z) and the reconstruction condition is : $P(z)+P(-z) = 2$ with $P(z) = H0(z^{-1})H0(z)$

This spectral factorization of a halfband filter. A half-band filter is defined as a zero-phase FIR filter whose impulse response p(n) (of length 2N+1) satisfies the condition,

$$p(2n) = \begin{cases} constant, & n = 0 \\ 0, & n \neq 0 \end{cases}$$

On the unit circle $z=e^{jw}$, the product $H0(e^{-jw})H0(e^{jw})$ is a magnitude squared:

$$P(e^{jw}) = \sum_{-N}^{N} p(n)e^{-jnw} = \left|\sum_{0}^{N} h(n)e^{-jnw}\right|^2$$

The halfband coefficients are p(n) = p(-n) for odd n and p(n)=0 for even n (except p(0)=1). Where N is odd. We design P(z) and factor to find H0(z). This symmetric factorization coincides with the Smith-Barnwell alternating flip. It yields orthogonal banks with PR [3,4,6].

4. SBC OF SPEECH IN 5 BANDS

The efficiency of digital encoding of speech and audio [8-12] depends strongly on the degree to which the bit rate can be reduced (compressed) without impairing the quality of the decoded signal. Coding the signal in narrower subbands offers one possibility for controlling the distribution of quantizing noise across the signal spectrum and hence, for realizing an improvement in signal quality. The quality of the coded signal is improved over that obtained from a single full-band coding of the total spectrum.

In this paper, speech signal is restricted at 4 kHz and divided into 5 bands. (These bands can be quantized and coded depending on subband energy. In this study, in order to show the error results only from analysis/ synthesis system, quantization of speech didn't used.) For sampling at 8 kHz, the frequency bands of the tree are taken as [0-500, 500-1000, 1000-2000, 2000-3000, 3000-4000 Hz]. Fig. 3 shows frequency response characteristics for an N=32 tap (1) Johnston (2) Barnwell filter design used for 5-band SBC of speech in this paper.

Sampled speech signal is normalized and divided into short segments (For this work, one speech segment is taken as L=128 voice samples). Then, each segment is applied to analysis/synthesis system as shown in Fig.4.a and Fig.4.b, respectively. First, speech signal is analyzed/ synthesized with Johnston QMF bank and MSE results are plotted for 200 speech segment as shown in Fig.5.a. Then, the same procedure repeated for Barnwell QMF bank and results are plotted as shown in Fig. 5.b. It can be shown from figures that, speech is reconstructed with Barnwell QMF bank better than Johnston QMF's.

5. CONCLUSIONS

In this paper, we compared two different filter banks in an analysis/synthesis SBC system. To compare the performances, a 5-band SBC speech analysis/synthesis system is implemented. Sampled speech signals are applied to analysis/synthesis system and MSE results are calculated. This test is repeated for different speech signals. Finally, it is shown Barnwell QMF banks give better results than Johnston QMF banks when applied to speech analysis/sythesis.

REFERENCES

1. Aharon Satt, David Malah ,"Design of Uniform DFT Filter Banks Optimized for Subband Coding of Speech", IEEE Transactions on ASSP, Vol 37, N0 11, November 1989
2. Ali N. Akansu, Richard A. Haddad, "Multiresolution Signal Decomposition, Transforms - Subbands - Wavelets",Academic Press Inc, 1992, TK5102.5, A414, C.2
3. Gilbert Strang, Truong Nguyen- "Wavelets and Filter Banks, " Wellesley - Cambridge Press, 1996, TK.7872.F5 S79
4. John Proakis, Charles M. Rader, Fuyun Ling, Chrysostomos L. Nikias, "Advanced Digital Signal Processing", Maxwell Macmillan International Edition, 1992
5. R.E. Crochiere, "Sub-band Coding", The Bell System Technical Journal, Vol.60, No.7, September,1981
6. Mark J.T. Smith, Thomas P. Barnwell, " Exact Reconstruction Techniques for Tree-Structured Subband Coders", IEEE Transactions on ASSP, Vol.34, No 3, June 1986
7. D.Esteband, C.Galand, "Application of QMFs to splitband voice coding schemes",in. Proc.IEEE Int.Conf. Acoust. Speech, Signal Processing, Hartford, CT. May. 1977, pp.191-195
8. Ronald E. Crochiere, Richard V. Cox, James L. Johnston, " Real Time Speech Coding", IEEE Transactions on Communications , Vol.. 30 , No 4, April 1982
9. N.S.Jayant, V. Lawrence, D. Prezas, "Coding of Speech and Wideband Audio", AT&T Technical Journal, Vol. 69 (5), pp. 25-41, Sept./Oct.1990
10. R.E. Crochiere, S.A. Webber, J.L. Flanagan, "Digital Coding of Speech in Sub-bands", The Bell System Technical Journal, Vol.55, No.8, October 1976
11. J.D. Johnston, "A filter family designed for use in quadrature mirror filter banks", in Proc. 1980 IEEE Int. Conf. Acoust. Speech, Signal Processing, Denver, CO, Apr. 1980, p.291
12. T.P.Barnwell, "Subband Coder Design incorporating recursive quadrature filters and optimum ADPCM coders", IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP-30, pp.751-765, Oct. 1982

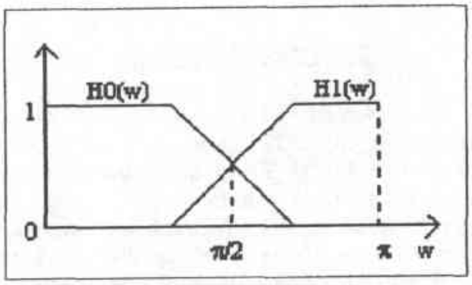

Figure 1. Filter characteristics for SBC

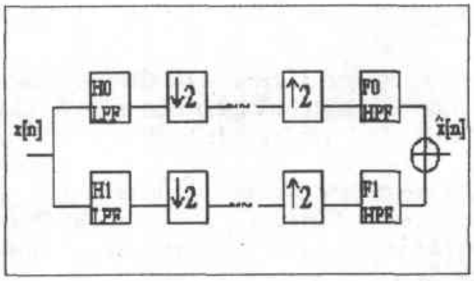

Figure 2. 2-band analysis/synthesis filter bank

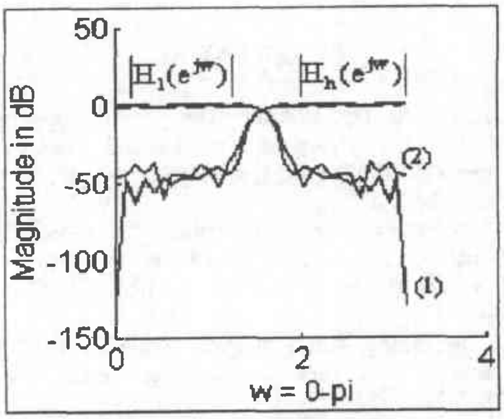

Figure 3.a. Frequency responses for 32-tap FIR
(1) Johnston
(2) Barnwell QMF design

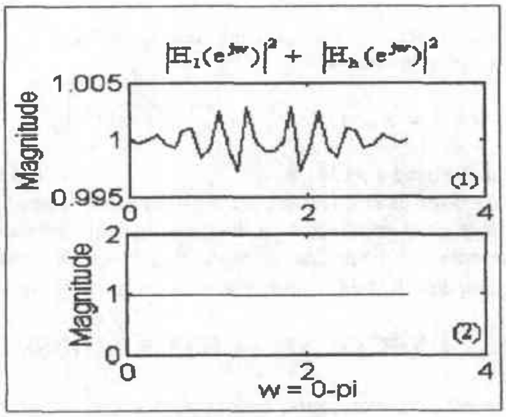

Figure 3.b. Allpass characteristics of
(1) Johnston
(2) Barnwell QMF design

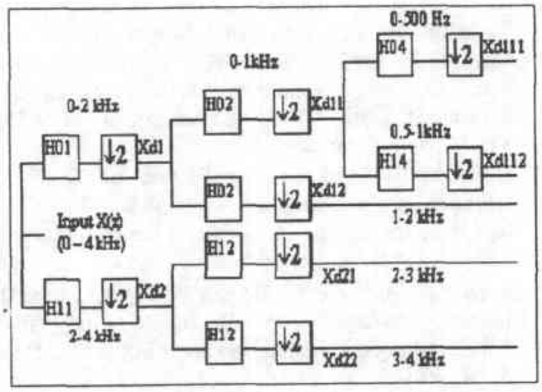

Figure 4.a 5-band SBC Transmitter (analysis) section

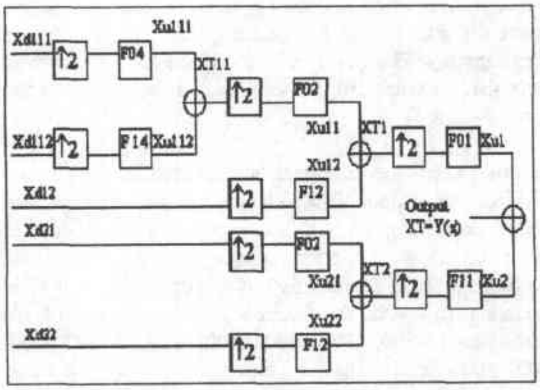

Figure 4.b 5-band SBC Receiver (synthesis) section

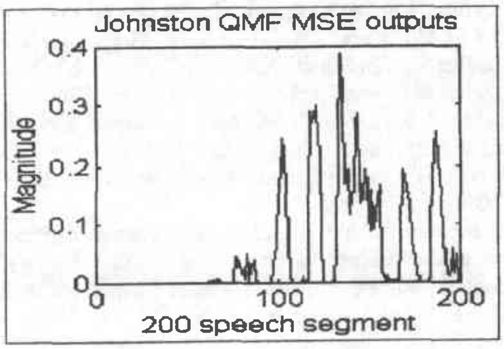

Figure 5.a Johnston filter MSE results for 5-band SBC

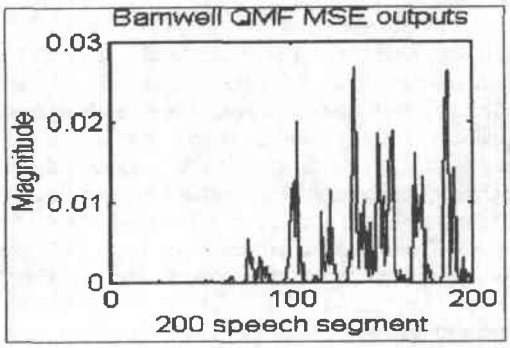

Figure 5.b. Barnwell filter MSE results for 5-band SBC