THE SECOND ISRAEL SYMPOSIUM ON
CIRCUITS, SYSTEMS AND CONTROL
ISCSC '88

May 30 - June 1, 1988
Hotel Dan Accadia/Daniel Tower - Diplomat Hall
DESIGN OF UNIFORM DFT FILTER BANKS
OPTIMIZED FOR SUB-BAND CODING OF SPEECH
A. Sab and D. Mollah
Department of Electrical Engineering
Technion — Israel Institute of Technology
Haifa 32000, Israel

ABSTRACT
A new approach for designing uniform DFT analyzable/analyzable filter-banks, optimized for sub-band coding (SBC) of speech, is presented. The design is performed by an iterative algorithm, which consists of solving two sets of linear equations in each iteration, aiding to minimize a spectral-domain distortion function. This function takes into account the desired characteristics of analyzable filter-banks, with decimation, quantization, and interpretation, used in SBC. Using the new design approach a 16 Kbps SBC is simulated and is found to achieve similar subjective and objective (SNR) performance, as that of a conventional G723-based SBC, with only about 60% of the computation.

1. INTRODUCTION
Sub-band coding is a well known method for digital speech coding at medium rates (e.g. 16 Kbps). In a sub-band coder (SBC) the signal is divided into separate bands typically eight by using an analysis filter bank. Usually, each filter signal is quantized by a gain-adaptive scalar quantizer. The speech is reconstructed from the quantized band-signal using a synthesis filter bank. A design of this type is referred to as "medium complexity", with its most complex part being the filter bank (1).

Two common types of filter banks are the Quadrature Mirror Filter (QMF) bank (2), and the uniform DFT (MDCT) bank (3). The QMF bank is designed to completely cancel the aliasing due to the decimation of the band-signal (in the absence of quantization), and it is widely used in sub-band speech coders. The QMF-based SBC obtains good quality at medium bit rates. In drawback, however, it is relatively high implementation complexity.

On the other hand, the DFT bank can be implemented efficiently using the Weighted Overlap Add (WOA) algorithm (4), in which case it is of much lower complexity than the QMF bank, for similar band separation. However, the performance (in terms of subjective quality) of DFT-based SBC, obtained by known filter design methods, was found to be much lower. In a recent work (4), a new design approach was presented, but was not found sufficient to improve the subjective quality.

In this work we present a method for designing filters for the uniform DFT filter-banks, which are optimized for sub-band coding of speech. This is achieved by minimizing a spectral-domain distortion function which takes into account the following desired characteristics of an analyzable filter-bank:

a. Good band-separation in the analysis stage, in order to enhance the redundancy removal by providing uncorrelated band-signals and thereby allowing the design of quantizers which are better matched to the non-quantizer properties of the band-signals (in particular, the gain in each band).

b. Good band-separation in the synthesis stage, for exploiting the auditory masking effect (3). The quantization noise in each frequency band is "masked" by a stronger band-signal in the same band, i.e. reducing the loudness as which the noise is perceived. It is important to minimize the leakage of quantization noise from one frequency band to other bands (including adjacent bands), since their band-sizes are not narrow (3), since its masking in other bands can not be constant, as it depends on signal intensity in those bands.

Minimizing aliasing effects (due to the decimation of the band-signals) in the reconstructed speech signal, and providing a time approximation to a unity system, i.e. the overall analysis/synthesis transfer function (in the absence of quantization) should be close to a pure delay.

The paper is organized as follows: section 2 presents the distortion function to be minimized, section 3 presents the iterative design algorithm used, section 4 presents a design example, and section 5 the implementation of a 16 Kbps SBC using the optimized filter banks. Section 6 presents simulation results with new SBC system and, for comparison, also with a conventional G723-based SBC.

2. DISTORTION FUNCTION
The basic equivalent structure of a complex uniform DFT filter bank is shown in Fig. 1. In the analysis stage, the input signal is decomposed by exp(j2\pi n/N) and filtered by the FIR analysis low-pass filter h(n), and quantized by R to produce the M band signals \hat{x}_m(n), k=1,...,M. The M complex signals, have a bandwidth of 2\pi/M. In the synthesis stage, the band signals are interpolated by the synthesis low-pass filters f(n), modulated by exp(j2\pi n/N), and summed up to produce the output signal x(n).

Using well known transform expressions for decimated and interpolated signals, the correlation of the output signal, which we quantify is approximated by given by:

\hat{X}_k = \sum_{n=0}^{N-1} x(n) f(n) \frac{e^{-j2\pi kn}}{M} \sum_{k=0}^{N-1} \frac{e^{j2\pi kn}}{M} \frac{e^{j2\pi kn}}{M}

(3)

The terms which include E(K|\hat{K}) for k=0 are the aliasing components in the output signal. Hence, we define the aliasing distortion as:

\hat{E}_a = \frac{1}{N} \sum_{n=0}^{N-1} |x(n)|^2 - \frac{1}{M} \sum_{k=0}^{M-1} \sum_{m=0}^{M-1} \frac{e^{j2\pi k n}}{M} \frac{e^{j2\pi k m}}{M}

(4)

The mean squared error (MSE) relative to unit transformation, i.e. pure delay of \tau samples is given by:

\hat{E}_a = \frac{1}{N} \sum_{n=0}^{N-1} |x(n)|^2 - \frac{1}{M} \sum_{k=0}^{M-1} \sum_{m=0}^{M-1} \frac{e^{j2\pi k n}}{M} \frac{e^{j2\pi k m}}{M}

(5)

As given in Eq. 2, the ideal frequency responses of the filters, are:

\hat{H}(\omega) = \frac{1}{\sqrt{M}} \sum_{k=0}^{M-1} \frac{e^{j2\pi k \omega}}{M}

(6)

where \theta denotes "don't care" regions. The energy of the side-lobe regions of the analysis and synthesis filters is: 1
\[ E_T = \frac{1}{2} \sum \left[ y(n) - d(n) \right]^2 \]

\[ E_f = \frac{1}{2} \sum \left[ \hat{x}(n) - d(n) \right]^2 \]

\[ Q_n = \left\{ \begin{array}{ll} 0, & |s| < \alpha M \\ \text{if } \sin \theta, & |s| < K - \alpha M \\ 1, & \sin \theta < K - \alpha M \\ \end{array} \right. \]

The detection function \( w(n) \) is introduced in order to give the following weighted sum:

\[ D = w(n)E_T + n_w(n)E_f \]

where the weights \( w(n) \) and \( n_w(n) \) are non-negative constants. Using Parseval's theorem, \( D \) can be represented as a positive-semi-definite (PSD) quadratic form, in terms of the analysis and synthesis filter coefficients:

\[ D = 1 - 2 \Re \{ f^H q \} \text{wss terms which are independent of } f \]

and are expressed in the form of \( Q \) and \( Q \) for the vectors of PSD matrices, respectively.

3. ITERATIVE ALGORITHM FOR OPTIMAL FILTER-BANK DESIGN

When the analysis filter \( h(n) \) is given, the synthesis filter \( f(n) \) is computed by minimizing \( D \), and vice versa. This problem is iteratively solved for the optimal filters, respectively:

\[ Q_{n+1} = Q_n \] \[ Q_{n+1} = Q_n \]

Assuming \( w_0 > 0 \) and \( w_1 > 0 \) (since we wish to limit the frequency accuracy in the stop-band, hence for \( h(n) \) and \( f(n) \)), and following the derivation in the Appendix, we conclude that \( Q \) and \( Q \) are the optimal solutions for the PSD matrices, respectively:

\[ Q_{n+1} = Q_n \]

and its magnitude is plotted in Fig. 4.

5. SBC IMPLEMENTATION

The SBC consists of the above analysis and synthesis filter-banks, and adaptive quantizer with dynamic bit allocation.

By increasing the weights \( w_0 \) and \( w_1 \) in the detection function defined in (8), the amplitude of the windows obtained by minimizing \( D \), decreases. In order to reduce the amplitude of the synthesised signal, without affecting the output SNR, \( f(n) \) is scaled by a factor \( c_{\text{opt}} \), which minimizes the MSE relative to uniform transmission:

\[ E_f = \frac{1}{2} \sum \left[ \hat{x}(n) - d(n) \right]^2 \]

\[ c_{\text{opt}} = \arg \min \{ E_f(\hat{x}(n)) \} \]

An expression for the factor \( c_{\text{opt}} \) in terms of the filter coefficients, is given in the Appendix.

The iterative algorithm consists of computing the analysis and synthesis filters alternately, minimizing the same distortion function. Hence, \( D \) is monotonically decreasing from iteration to iteration. The algorithm converges to a local minimum of \( D \), which depends on the initial filter chosen.

4. DESIGN EXAMPLE

In this section, we present a filter design example, for a sub-band coder with \( M = 6 \) complex filters. To achieve maximum coding efficiency, it is desirable to optimize the hard-signals by the critical ratio \( R = M \). However, for designed filters, for the critical ratio, they were found to have the desired characteristics mentioned in section 3, unless they are of very long duration. Decreasing slightly the detection ratio, using \( R = M = 15 \), enables the design of filters which obtain in high performance SBC, but still of quite lower complexity than a conventional QMF coder of similar performance.

The impulse and frequency responses of the optimal filters, with \( M = 16 \) bands, decimation ratio \( R = 9 \), and length of 256 taps each, are shown in Fig. 3. The weight factors were chosen to be \( w_0 = 10, w_1 = 40, \) and the resulting SBCs provide good hard-separation and a high performance SBC. The transfer function of the system in the above example (obtained from (1)) by excluding the aliasing component is given by:

\[ T(s) = \frac{1}{M} \sum_{i=0}^{M-1} H(e^{j2\pi i/M})F(e^{j2\pi i/M}) \]

and its magnitude is plotted in Fig. 4.
the estimated variances $\hat{\sigma}^2_i = \sum_{i=1}^{M-2} \hat{\sigma}^2_i$ by 2, before computing the allocation. A procedure for this allocation is given in [7]. The allocation is also updated every $N=16$ samples of each of $X_{m}(n)$, $m=0, \ldots, M-1$.

The computation rate required for implementing the analysis filter bank is on the following: $L$ multiplies, $L$-M adds and an FFT of length $M$. In the analysis stage for producing $M$ complex samples [2, Fig.3] where $n$ is the filter length. The same number of operations is required for the synthesis stage [2]. For a decomposition rate of $M=12$, the implementation of both the analysis and synthesis filter banks requires 36 multiplies and 40 adds per input sample.

6. SIMULATION RESULTS

As SBC which utilizes the above optimized filters, and an SBC which is based on an 8-band QMF bank, were simulated for a transmission rate of 16 Kbit/s. The implementation of the QMF bank requires 98 multiplies and 102 adds per input sample (using 32-bit filters). The DFT-based coder and the QMF-based coder were found to have similar subjective performance. Both coders yield high quality synthesized speech, decoded only by very slight heariness. The objective performance of the QMF-based coder is slightly higher, 22.69 vs. 19.85 dB for the QMF SBC, as compared to 18.9 dB for the DFT-based coder. An attempt to reduce the QMF bank complexity to 48 multiplies per input sample, using shorter filters, caused noticeable degradation in the synthesized speech quality.

Another filter-bank structure is the generalized parallel QMF bank [18, 9, 10]. The filter bank described in [9] has 8 bands, and requires about 75% of the number of multiplies and 15% of the additions as compared to the above uniform DFT filter bank. However, in that filter bank [9], adjacent bands are not well separated, and the coder using this filter bank can be expected to have similar performance to the above DFT-based coder, only if longer filters will be used (thus increasing its complexity). The filter bank presented in [9] splits the signal into 16 bands (also not too well separated), and requires 12 multiplies and 22 adds per input sample. Both papers [9], and [10], do not present comparisons to QMF-based or other coders. The SBC presented in [10], includes 5-band filter banks, using longer filters (60 taps) to reduce the interband aliasing.

7. SUMMARY

A new approach for designing FIR filters for uniform DFT filter banks, optimized for sub-band coding of speech was presented. A 16 Kbit/s SBC which utilizes the optimized filter was simulated and was found to achieve similar subjective performance as that of the QMF-based SBC but effecting about 60% reduction in computations as compared to the QMF bank.

APPENDIX

1. Aliasing Distortion:

From (2), using Parseval's theorem:

$$E_{\text{NL}} = \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} |f(n)\tilde{w}(n)|^2 + |f(n)w(n)\omega_{m}|^2$$

Using the identity:

$$\sum_{n=0}^{M-1} w(n)^2 = M - \sum_{n=0}^{M-1} w(n)\phi(n)\mod M$$

the following expression for $E_{\text{NL}}$ is obtained:

$$E_{\text{NL}} = \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} \frac{1}{2} |f(n)\tilde{w}(n)|^2 + \left|f(n)w(n)\omega_{m}\right|^2$$

For simplicity we choose the filters $h(n)$ and $f(n)$ to be of equal length.
5. Expression for $e_k$ (eqn. (9))
Using Parseval’s theorem and (14), $E_k(e)$ can be expressed as:

$$E_k(e) = 1 - 2eF_q + e^2 F_q f$$

(32)

Minimizing $E_k(e)$ results in:

$$e^* = (F_q f)^{-1} (F_q q)$$

(33)

REFERENCES


Fig. 1 The uniform DFT filter bank

Fig. 4 The transfer function $f(x)$

Fig. 3 The impulse and frequency responses of the optimal filter

Fig. 2 The ideal filter frequency responses