

SUBBAND-DOMAIN FILTERING OF MPEG AUDIO SIGNALS

Chris A. Lanciani and Ronald W. Schafer

Center For Signal and Image Processing
Georgia Institute of Technology
Atlanta, GA 30332
{lanciani,rws}@ece.gatech.edu

ABSTRACT

The cosine modulated filter bank is commonly used for the time-frequency decomposition of audio signals. For example, it is a basic element of the MPEG-1 and MPEG-2 audio coding standards. While this filter bank is not perfectly-reconstructing, it does provide for the cancelation of aliasing components that are introduced during the analysis decomposition. If the subband signals are to be processed, care must be taken to preserve the properties of the subband signals such that the aliased terms will be canceled successfully in the synthesis filter bank despite the modification of the subband signals. In this paper, a framework is provided for the generation and application of arbitrary FIR filters to signals that have been decomposed using the MPEG filter bank.

1. INTRODUCTION

The MPEG audio standard, which was originally introduced in 1992 and was updated in 1994 and 1997 [1], has recently gained popularity for both the storage and the transmission of high-quality audio signals. One of the fundamental elements of the MPEG coder is the filter bank, which is used for the time/frequency transformation of the time-domain audio signal. The filter bank that is used in Layers 1, 2, and 3 of both MPEG-1 and MPEG-2 is a pseudo-QMF, cosine modulated filter bank [2, 3]. The benefits of this filter bank include efficient implementation, flexibility in the prototype design, and the cancelation of aliasing components introduced during the analysis decomposition.

Several authors have discussed the filtering of MPEG audio signals in the compressed domain [4, 5, 6]. In the case of MPEG signals, this implies that the filtering is done on subband signals that have been dequantized and rescaled, but not resynthesized into time-domain signals. The advantage of this type of processing is that a modified MPEG-compliant signal can be produced without having to convert the signal to the time domain for processing and then back to the MPEG domain for continued transmission or storage. These conversions require the use of the filter bank, which, while efficiently implemented, still consumes a large portion of the operations in both the MPEG encoder and the decoder. The removal of the analysis and synthesis banks results in significant computational savings during the processing of the MPEG signal.

Since the subband signals are formed using a filter bank that introduces aliasing into the signals, any filtering that is done in the subband domain must not destroy the ability of the synthesis filter

bank to cancel the aliasing terms. In [4], the solution is to constrain the subband filters to have smooth responses across the subband boundaries. This preserves the relationship between the amplitudes of the aliasing components resulting from the overlapping frequency responses of the analysis filters. In [5], aliasing errors are introduced, but the nature of the subband processing masks the distortion caused by the modifications. Finally, in [6], only simple filters such as lowpass or highpass filters are used, and they are applied using a zero-phase technique to preserve the phase relationships between the samples in adjacent subbands. In this paper, we present a framework that allows for the application of FIR filters with arbitrary frequency responses to the subband signals that result from the decomposition performed in the MPEG Layer 1 and 2 encoders. Quantization issues regarding the processing of MPEG signals, discussed in [6], are not covered in this paper.

2. FILTERING SUBBAND SIGNALS

The objective of filtering subband signals is to apply filters to the subband samples in such a way that, when the modified subband signals are input to the synthesis filter bank, the output signal will be equivalent to the signal that is obtained by reconstructing the unmodified signal, filtering it in the time domain, and then recoding into the subband domain. If this technique is to be applied to an MPEG signal, the analysis and synthesis filter banks cannot be modified since the filtered signal must remain MPEG-compliant. To accommodate this, the method described in this paper requires neither modification of the analysis and synthesis filters nor post-processing at the output of the synthesis bank.

The general structure of the cosine modulated filter bank with additional subband filtering is shown in Figure 1. The subbands of the filter bank are indexed by k , which ranges from 0 to $M - 1$ where $M = 32$ for the MPEG filter bank. $H_k(z)$ is the analysis filter in subband k and $G_k(z)$ is the corresponding synthesis filter. The analysis filter outputs are filtered by $C_{sb}(z)$, which consists of filters $C_k(z)$ for $0 \leq k \leq M - 1$. While the outputs of each $C_k(z)$ remain in subband k , $C_k(z)$ actually operates on samples from subband k as well as those from adjacent subbands. This will be explained further in Section 2.3.

2.1. Matrix Representation of the MPEG Filter Bank

The matrix formulation of the system in Figure 1 provides a useful framework for producing the subband filters from an arbitrary time-domain filter. Let $p(n)$ be the impulse response of an N -length FIR filter, and let \mathbf{x} and \mathbf{y} be vectors of length N and

This work was supported in part by the Hewlett-Packard Company and the Georgia Research Alliance.

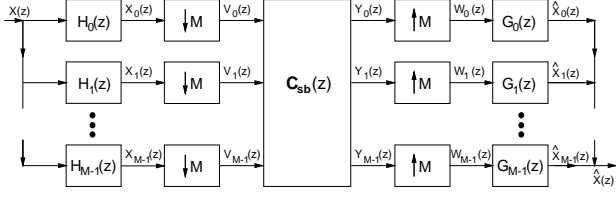


Figure 1 General structure of the MPEG filter bank with additional subband filtering.

$2N - 1$ that represent blocks of the input and output signals $x(n)$ and $y(n)$:

$$\mathbf{x} = [x(0), x(1), x(2), \dots, x(N-1)]^T, \quad (1)$$

$$\mathbf{y} = [y(0), y(1), y(2), \dots, y(2N-2)]^T. \quad (2)$$

The filtering of $x(n)$ by $p(n)$ can be represented by

$$\mathbf{y} = \underbrace{\begin{bmatrix} p(0) & 0 & \dots & 0 \\ p(1) & p(0) & \dots & 0 \\ p(2) & p(1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ p(N-1) & p(N-2) & \dots & p(0) \\ 0 & p(N-1) & \dots & p(1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & p(N-1) \end{bmatrix}}_{\mathbf{P}} \mathbf{x}, \quad (3)$$

where \mathbf{P} is the $(2N - 1) \times N$ convolution matrix [7]. The filter $p(n)$ can also be represented by the $N \times N$ filter matrix \mathbf{P} , which consists of the first N rows of \mathbf{P} . In this and the following sections, the filter matrix will generally be used in theoretical descriptions while the convolution matrix will be used to explain implementation details.

The analysis filter bank, consisting of the M filters $H_k(z)$ and M downsamplers, can be represented by the $2N \times N$ convolution matrix \mathbf{H} :

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_{M-1} \end{bmatrix}, \quad (4)$$

where \mathbf{H}_k is given by

$$\begin{bmatrix} h_k(0) & 0 & \dots & 0 \\ h_k(M) & h_k(M-1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_k(N-M) & h_k(N-M-1) & \dots & 0 \\ 0 & h_k(N-1) & \dots & h_k(1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & h_k(N-M+1) \end{bmatrix} \quad (5)$$

and $N = 512$ for the MPEG filters. The matrix \mathbf{H}_k is $\frac{2N}{M} \times N$ and is a downsampled (by M) version of the convolution matrix for the filter $H_k(z)$, which has the structure shown in (3). The

$N \times N$ filter matrix, \mathbf{H} , has the same structure as \mathbf{H} , with each \mathbf{H}_k consisting of the first $\frac{N}{M}$ rows of \mathbf{H}_k .

The synthesis filter bank can be represented by the $2N \times N$ matrix \mathbf{G} :

$$\mathbf{G} = [\mathbf{G}_0 \ \mathbf{G}_1 \ \dots \ \mathbf{G}_{M-1}], \quad (6)$$

where

$$\mathbf{G}_k = \begin{bmatrix} \begin{bmatrix} \mathbf{g}_k \\ 0 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ \mathbf{g}_k \\ 0 \end{bmatrix} & \dots & \begin{bmatrix} 0 \\ 0 \\ \mathbf{g}_k \end{bmatrix} \end{bmatrix}. \quad (7)$$

In (7), \mathbf{G}_k is $2N \times \frac{N}{M}$, \mathbf{g}_k is an N -point vector containing the coefficients of $G_k(z)$, and each $\mathbf{0}$ represents a vector of M zeroes. The filter matrix \mathbf{G} is $N \times N$ with the same structure as \mathbf{G} . The submatrix \mathbf{G}_k consists of the first N rows of \mathbf{G}_k .

Based on the matrix formulation, the system in Figure 1 with no subband filtering can be represented by

$$\hat{\mathbf{x}} = \mathbf{G}\mathbf{H}\mathbf{x}. \quad (8)$$

2.2. Generation of the Subband Filters

The subband filters $C_k(z)$ can also be represented by a convolution matrix, \mathbf{C}_{sb} , and a filter matrix, \mathbf{C}_{sb} . To obtain these matrices using the representation developed in the previous section, it is useful to consider the purpose of the subband filtering process. The objective of the subband-domain filter can be summarized using Figure 2. In this figure, the case shown in (a) represents the operation of the MPEG encoder and decoder (in the absence of quantization) followed by time-domain filtering of the decoded signal. The case in (b) represents the subband-domain filtering of an MPEG-encoded signal, again in the absence of quantization. The goal is to obtain equivalence between the outputs $\hat{X}_{sb}(z)$ and $\hat{X}_d(z)$. This

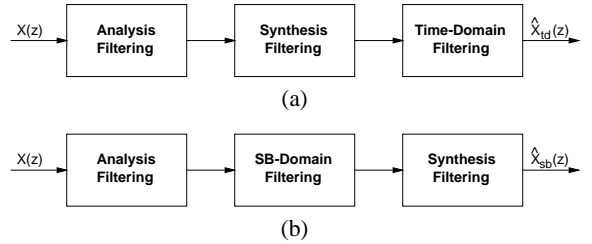


Figure 2 Overall operation in the (a) desired case and (b) subband filtering case.

amounts to finding the appropriate set of subband filters.

For an input vector \mathbf{x} , the systems in Figure 2 are described by

$$\hat{\mathbf{x}}_d = \mathbf{C}\mathbf{G}\mathbf{H}\mathbf{x} \quad (9)$$

$$\hat{\mathbf{x}}_{sb} = \mathbf{G}\mathbf{C}_{sb}\mathbf{H}\mathbf{x}, \quad (10)$$

where \mathbf{C} is the filter matrix for the time domain filter $C(z)$ and \mathbf{C}_{sb} is the filter matrix for the subband filters $C_k(z)$ for $0 \leq k \leq$

$M - 1$. To find the appropriate \mathbf{C}_{sb} , $\hat{\mathbf{x}}_{sb}$ is set equal to $\hat{\mathbf{x}}_d$. Since \mathbf{x} appears in both expressions, the equality becomes

$$\mathbf{G}\mathbf{C}_{sb}\mathbf{H} = \mathbf{C}\mathbf{G}\mathbf{H}. \quad (11)$$

Pre-multiplying both sides of (11) by \mathbf{G}^{-1} and post-multiplying both sides by \mathbf{H}^{-1} give

$$\mathbf{C}_{sb} = \mathbf{G}^{-1}\mathbf{C}\mathbf{G}. \quad (12)$$

For the analogous calculation using the convolution matrices \mathbf{C}_{sb} , \mathbf{C} , and \mathbf{G} , the matrices and Equation (12) must be modified. First, since \mathbf{G} is rectangular and therefore difficult to invert, \mathbf{G}^{-1} is replaced by \mathbf{H} . This replacement gives an inexact solution, but the nearly-perfect reconstructing property of the MPEG filter bank allows for this substitution without introducing significant errors. Additionally, for \mathbf{C}_{sb} to be $2N \times N$, the \mathbf{H} and \mathbf{C} matrices must be extended. The subband-domain convolution matrix is then obtained using the following equation:

$$\mathbf{C}_{sb} = \mathbf{H}'_{2N \times 2N} \mathbf{C}'_{2N \times 2N} \mathbf{G}_{2N \times N}. \quad (13)$$

In (13), \mathbf{H}' and \mathbf{C}' are formed by extending the rows and columns of \mathbf{H} and \mathbf{C} as necessary. The extension of \mathbf{C} is shown in the following equation:

$$\mathbf{C}' = \begin{bmatrix} & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ \begin{bmatrix} \mathbf{C} \end{bmatrix} & & & & & \\ & c(0) & 0 & \cdots & 0 \\ & c(1) & c(0) & \cdots & 0 \\ & \vdots & \vdots & \ddots & \vdots \\ & c(N-2) & c(N-3) & \cdots & 0 \\ \mathbf{0}_{1 \times N} & c(N-1) & c(N-2) & \cdots & c(0) \end{bmatrix}, \quad (14)$$

where \mathbf{C} has the structure shown in (3). The convolution matrix \mathbf{H}' is obtained by similarly extending each \mathbf{H}_k .

2.3. Application of the Subband Filters

The subband-domain filtering is performed in a blockwise manner using the convolution matrices described in the previous section. The output samples in each block are combined with samples from the previous block using the overlap-add method of block convolution [8]. For subband signals, the last half of each subband vector from the previous block is added to the first half of the subband vector in the current block. For time-domain signals, the last half of the entire output vector from the previous block is added to the first half of the output vector in the current block. The details of each filtering operation are described below.

Referring to Figure 1, the analysis filter bank matrix \mathbf{H} is multiplied by \mathbf{x} to give the subband signal \mathbf{v}' :

$$\mathbf{v}' = \mathbf{H}\mathbf{x}, \quad (15)$$

where \mathbf{x} is the $N \times 1$ vector containing $x(n)$ and

$$\mathbf{v}' = [\mathbf{v}'_0^T \quad \mathbf{v}'_1^T \quad \cdots \quad \mathbf{v}'_{M-1}^T]^T. \quad (16)$$

The $N \times 1$ subband output vector, \mathbf{v} , is obtained from the $2N \times 1$ vector \mathbf{v}' using the subband overlap-add technique on each \mathbf{v}'_k .

The subband processing is represented by

$$\mathbf{y}' = \mathbf{C}_{sb}\mathbf{v}, \quad (17)$$

where

$$\mathbf{y}' = [\mathbf{y}'_0^T \quad \mathbf{y}'_1^T \quad \cdots \quad \mathbf{y}'_{M-1}^T]^T, \quad (18)$$

$$\mathbf{v} = [\mathbf{v}_0^T \quad \mathbf{v}_1^T \quad \cdots \quad \mathbf{v}_{M-1}^T]^T, \quad (19)$$

and

$$\mathbf{C}_{sb} = \begin{bmatrix} \mathbf{C}_{0,0} & \mathbf{C}_{0,1} & \cdots & \mathbf{C}_{0,M-1} \\ \mathbf{C}_{1,0} & \mathbf{C}_{1,1} & \cdots & \mathbf{C}_{1,M-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_{M-1,0} & \mathbf{C}_{M-1,1} & \cdots & \mathbf{C}_{M-1,M-1} \end{bmatrix}. \quad (20)$$

In (20), \mathbf{C}_{sb} is $2N \times N$ and consists of M rows and M columns of $\frac{2N}{M} \times \frac{N}{M}$ submatrices $\mathbf{C}_{m,n}$, where m indexes the subbands of \mathbf{y}' and n indexes the subbands of \mathbf{v} . The vector \mathbf{y} is obtained from \mathbf{y}' using the same overlap-add technique that was used for \mathbf{v} .

The synthesis filtering is accomplished by applying \mathbf{G} to the filtered subband signal \mathbf{y} to obtain $\hat{\mathbf{x}}'$:

$$\hat{\mathbf{x}}' = \mathbf{G}\mathbf{y}, \quad (21)$$

where $\hat{\mathbf{x}}'$ is the $2N \times 1$ vector containing the delayed reconstructed signal $\hat{x}(n)$ and \mathbf{y} is a composite $N \times 1$ vector with the same structure as \mathbf{v} in (19). The time-domain overlap-add technique is used to obtain $\hat{\mathbf{x}}$ from $\hat{\mathbf{x}}'$.

Equations (17) and (20) show that, for each subband k , the output samples are obtained by filtering the samples in every subband, not just those in subband k . This preserves the relationships between adjacent subbands that are necessary for aliasing cancellation to occur in the synthesis filter bank. However, the frequency response of the analysis filter in subband k only overlaps those of subbands $k - 1$ and $k + 1$ significantly. Consequently, the relationships of interest to subband k are those with subbands $k - 1$ and $k + 1$, and the procedure can be simplified to filter only the samples in the subbands directly adjacent to k . This is shown in the following equation, which describes the filtering in subband k :

$$\mathbf{y}'_k = \mathbf{C}_k \mathbf{v}'_k, \quad (22)$$

where

$$\mathbf{C}_k = [\mathbf{C}_{k,k-1} \quad \mathbf{C}_{k,k} \quad \mathbf{C}_{k,k+1}] \quad (23)$$

and \mathbf{v}'_k is a composite vector consisting of \mathbf{v}_{k-1} , \mathbf{v}_k , and \mathbf{v}_{k+1} (\mathbf{v}'_k is three times the length of \mathbf{v}_k). With this simplification, \mathbf{C}_{sb} retains only the submatrices $\mathbf{C}_{m,n}$ for which $m - 1 \leq n \leq m + 1$.

3. EXAMPLE OF SUBBAND-DOMAIN FILTERING

The subband filtering of MPEG audio signals has several applications. Perceptual effects such as head-related transfer functions can be applied to MPEG signals by filtering them in the subband domain. Filtering can be applied to an MPEG signal for sampling rate conversion, either to prevent aliasing before downsampling or to interpolate samples after upsampling. Additionally, MPEG signals can be equalized in frequency using subband filters. In this section, we will give an example of equalization filtering in the subband domain.

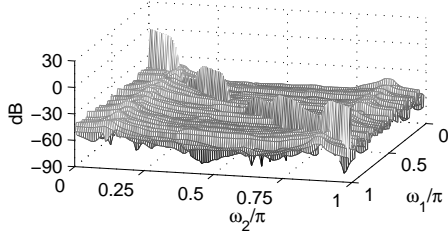


Figure 3 Magnitude of the estimated bifrequency system function for time-domain filtering.

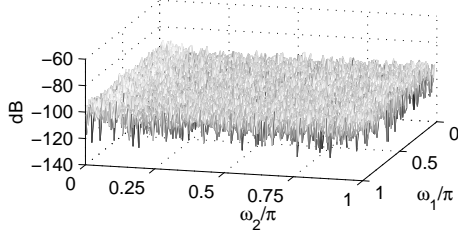


Figure 4 Magnitude of the difference between the estimated bifrequency system functions for time-domain and subband-domain filtering.

To display the results, we will use the bifrequency system function, $K(e^{j\omega_2}, e^{j\omega_1})$, which gives the frequency-domain relationship between the input to a system, $X(e^{j\omega_1})$, and the corresponding output signal, $Y(e^{j\omega_2})$. The time-invariant behavior of the system is expressed along the main diagonal of $K(e^{j\omega_2}, e^{j\omega_1})$ (corresponding to $\omega_1 = \omega_2$). Time-varying behavior, such as aliasing distortion, appears on other diagonals. Further details and a description of the method used to obtain our results are given in [9].

The frequency response of an equalization filter is shown in Figure 5. The subband convolution matrix for this filter is generated using the method described in Section 2.2 and applied to the subband samples as detailed in Section 2.3. The estimated bifrequency system function for time-domain filtering is shown in Figure 3, while the difference between the estimates for time-domain filtering and subband-domain filtering is displayed in Figure 4. The time-invariant behavior of the two systems is shown in Figure 5, along with the frequency response of the desired equalization filter. These results show that both the time-invariant and time-varying parts of the bifrequency system functions are very similar for the time-domain and subband-domain equalization.

For MPEG-encoded signals, subband-domain filtering is performed on the unpacked subband signals. The filtered signal is then requantized and re-packed into the MPEG format. Time-domain filtering requires decoding, filtering, and re-encoding the signal. Neglecting quantization and psychoacoustic modeling, this would require synthesis filtering, time-domain filtering, and analysis filtering. On one implementation on a 166 MHz Pentium PC, these filtering processes take 0.57, 0.98, and 0.46 seconds per second of audio. Using these values, the time-domain filtering method requires 2.01 seconds per second of audio, while subband-domain filtering takes 0.71 seconds per second of audio. In an actual MPEG implementation, the difference between these times would be increased due to the application of the psychoacoustic

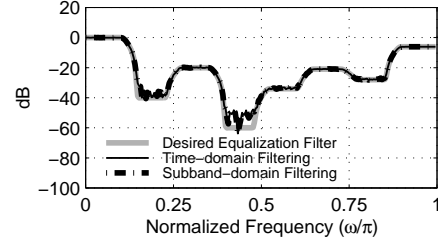


Figure 5 Spectrum of the equalization filter and main diagonals of the bifrequency system function magnitudes.

model and bit allocation algorithms in the MPEG encoder, and the fact that with subband-domain filtering, subbands with no allocated bits are not processed.

4. CONCLUSION

We have presented a framework for the filtering of signals in the subband domain. Specifically, this technique can be used to modify MPEG signals without having to decode, filter, and re-encode the signal. The method enables the application of FIR filters with arbitrary frequency responses to MPEG signals while preserving the alias-canceling properties of the MPEG filter bank. Additionally, this technique does not require modifications of the analysis and synthesis filter banks. As a result, this framework can be used within an MPEG processing system to produce MPEG-compliant signals that can be accepted by any compatible decoder.

5. REFERENCES

- [1] K. Brandenburg and G. Stoll, "Overview of MPEG Audio: Current and Future Standards for Low-Bit-Rate Audio Coding," *Journal of the Audio Engineering Society*, vol. 45, pp. 4–21, Jan/Feb 1997.
- [2] K. Brandenburg and G. Stoll, "ISO-MPEG1 Audio: A Generic Standard for Coding of High-Quality Digital Audio," *Journal of the Audio Engineering Society*, vol. 42, pp. 780–792, Oct. 1994.
- [3] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, N.J.: P T R Prentice Hall, 1993.
- [4] M. A. Broadhead and C. B. Owen, "Direct Manipulation of MPEG Compressed Digital Audio," *ACM Multimedia Conference*, pp. 499–507, 1995.
- [5] S. Levine, "Effects Processing on Audio Subband Data," *International Computer Music Conference*, 1996.
- [6] C. A. Lanciani and R. W. Schafer, "Psychoacoustically-Based Processing of MPEG-I Layer 1-2 Encoded Signals," *IEEE Workshop on Multimedia Signal Processing*, pp. 53–58, 1997.
- [7] K. Nayeibi, *A Time Domain Framework for the Analysis and Design of FIR Multirate Filter Bank Systems*. PhD thesis, Georgia Institute of Technology, 1990.
- [8] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, N.J.: Prentice Hall, 1989.
- [9] R. L. Reng and H. W. Schuessler, "Measurement of Aliasing Distortions and Quantization Noise in Multirate Systems," *IEEE International Symposium on Circuits and Systems*, pp. 2328–2331, 1992.