Module 9

AUDIO CODING

Lesson 29

Transform and Filter banks

Instructional Objectives

At the end of this lesson, the students should be able to:

1. Define the three layers of MPEG-1 audio coding.
2. Define the four modes of MPEG-1 audio coding.
3. Present the structure of MPEG-1 audio codec layer –1 and II.
4. State the basic objectives of two psychoacoustics models.
5. Explain the realization of analysis and synthesis filters through filter banks.
6. Explain the realization of analysis and synthesis filters through time domain windowing and transforms.
7. Define critically sampled analysis-synthesis system.
8. Define time domain aliasing.
9. Explain the mechanism of time domain alias cancellation.

29.0 Introduction

In lesson-28, we have learnt the basic requirements of efficient audio codec design, utilizing the perceptual aspects of audio. The MPEG-1 audio codec is the first standard to adopt this concept. The standard offers a choice of three independent layers of compression with increasing order of complexity and four possible modes to support up to two audio channels. For perceptual bit allocation, the input audio signal is analyzed through a bank of bandpass filters and based on the masking threshold, bits are allocated to each of the filtered channels.

In this lesson, we are first going to present the basic structure of MPEG-1 audio codec, followed by the design considerations of analysis filter bank. The analysis filter bank design is based on time domain alias cancellation (TDAC), whose concepts we are going to introduce too.

29.1 Layers of audio compression:

The MPEG-1 audio coding standard supports the following three coding layers with increasing complexity, delay and subjective performance –

(a) Layer –1: This is the simplest layer and best suits bit rates above 128 Kbps per channel. Relatively lower compression ratio as compared to the higher layers makes this layer simple. Layer-1 encoding is used in digital audiocassettes that use compressed bit rates of 192 Kbps per channel.

(b) Layer – II: This layer is of intermediate complexity and supports compressed bit stream up to 128 Kbps per channel. Applications of this layer include digital audio in CDs, VCDs.

(c) Layer – III: This is the most complicated of the three layers. The target bit rate is around 64 Kbps per channel. This layer is suitable for ISDN.

29.2 Modes of MPEG-1 audio

MPEG-1 supports upto two channels of audio with four modes as follows.

(a) *Monophonic mode* for single channels audio

(b) *Dual monophonic mode* for two independent channels of audio. This is useful for bilingual presentations

(c) *Stereo mode* for stereo channels that share bits but do not exploit correlations between the stereo channels.

(d) *Joint stereo mode* which exploits correlations between the stereo channels and uses an irrevancy reduction technique called *intensity stereo*.

29.3 Structure of MPEG-1 audio codec

Fig. 29.1 shows the structure of MPEG-1 audio codec for layer-I and II. It is very similar to the perceptual audio codec presented in lesson–28. To allocate bits to the audio samples, the samples are passed through an analysis filter bank, which filters the signal into 32 subbands. To derive the audio spectrum, performing FFT on the samples carries out a time to frequency conversion. The psychoacoustic models identify the tonal and the non-tonal components to derive the masking thresholds and hence the Signal to Mask Ratio (SMR). Bits are allocated on a block by block basis for each subband. Each block has 12 samples per subband for Layer-I of audio i.e., total 384 samples considering 32 sub bands and 36 subband samples (total of 1152 samples) for Layer-II and III. MPEG-1 audio codec supports two psychoacoustics' models. The psychoacoustics model I is used in Layer I. and II. and the model-II is used for Layer-III. The subband codewords, the scale factor and the bit allocation information are multiplexed into one bitstream, together with a header and optional ancillary data. At the decoder, the synthesis filter bank reconstructs the audio output from the demultiplexed bitstream.

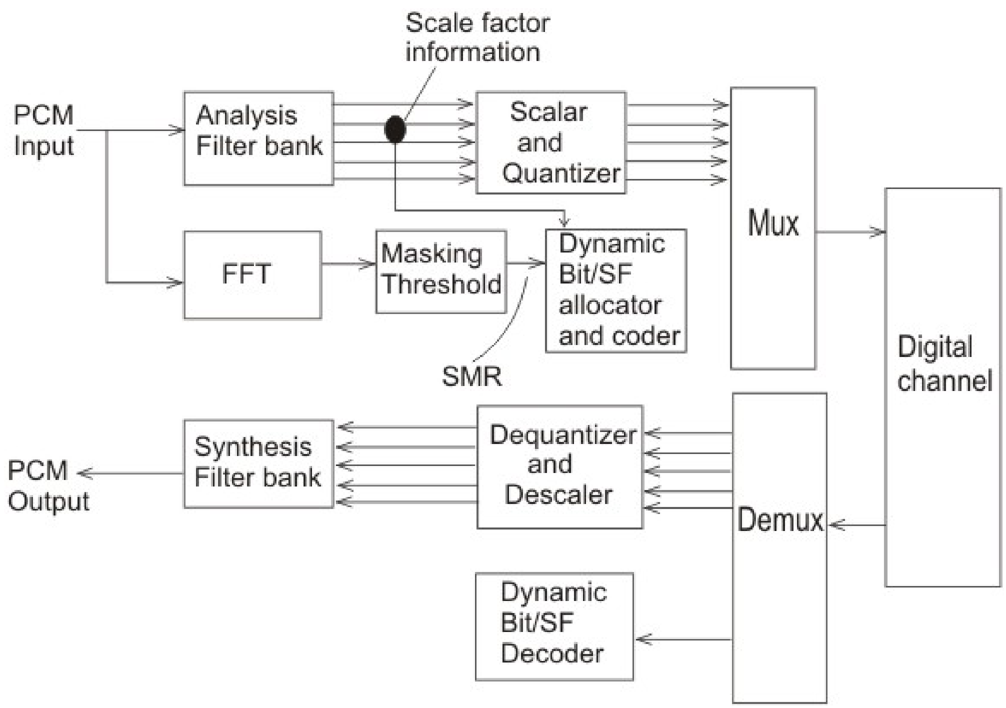

Figure 29.1 Structure of MPEG.1 audio encoder and decoder (Layers I and II)

29.4 Psychoacoustics models

As already mentioned, MPEG-1 audio codec design is based on two psychoacoustic models. This model computes the *SMR* s, considering the short term spectrum of the audio block to be coded. Only the encoder requires this model. Both model-I and II determine the masking characteristics for either "tone masks noise" or "noise masks tone" types and calculate the global masking thresholds. In both the models, the audio samples are Hann weighted.

The Psychoacoustics model II is more complex than the model- I. It takes into consideration the properties of the inner ear and the effects of pre-echoes. We are going to discuss about the psychoacoustics models with further details in the subsequent lessons.

29.5 Design issues of analysis and synthesis filters

The basic framework for an analysis / synthesis system is shown in Fig 29.2. The analysis filer bank segments the input samples $x(n)$ into a number of contiguous filter banks, whose outputs are shown as $X_k(m)$, where k=0,1, ---, K-1 are the subband numbers and m=0,1… are the sample indices. The synthesis filterbanks receive the subband samples $\hat{X}_k(m)$ from the digital channel. For a lossless channel, $\hat{X}_k(m) = X_k(m)$, but this is not true for a noisy channel. The synthesis filter bank reconstructs the signal as $\hat{x}(n)$, which should be ideally an exact replica of the input signal.

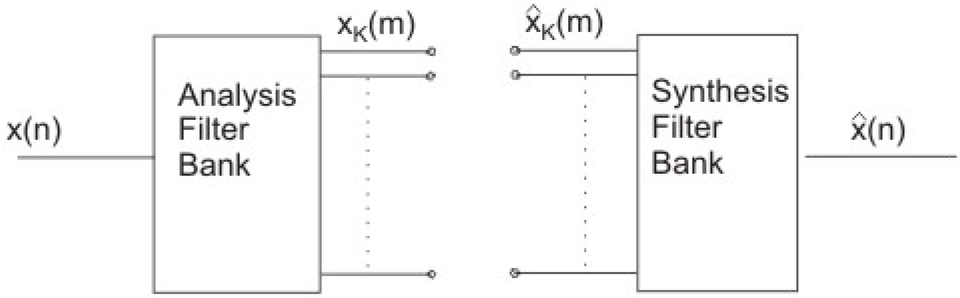

Figure 29.2 Analysis / Synthesis Filtering Framework

There are two different approaches to the design of such analysis and synthesis filters. The first is based on a bank of lowpass and bandpass filters. We conveniently express these filters in frequency domain. Exact reconstruction can be achieved only if the composite analysis-synthesis filter channel responses overlap and add, such that their sum indicates a flat response Fig. 29.3 indicates this concept. All the filters are non-ideal and have definite roll-offs, which must match.

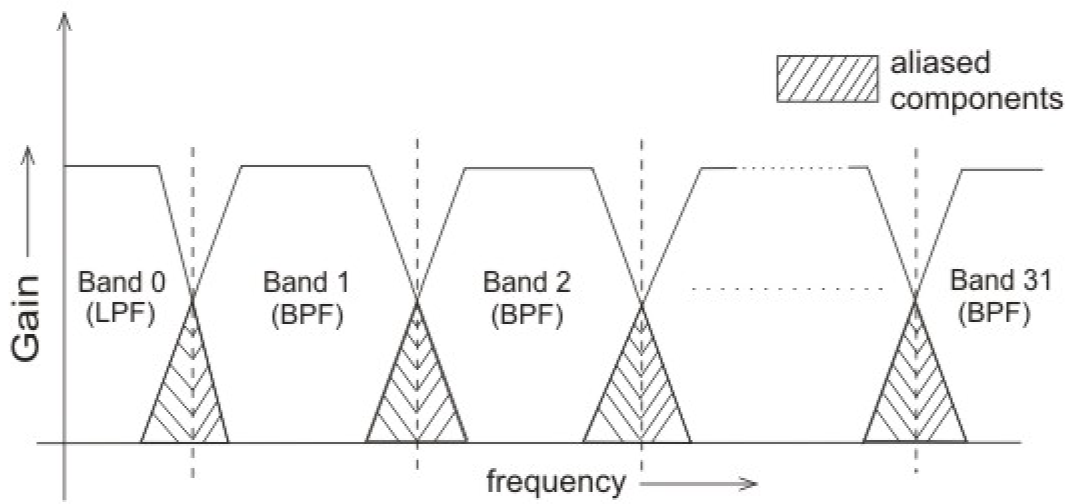

Figure 29.3 Characteristics of analysis synthesis filter.

In this scheme, it is necessary that any frequency domain aliasing introduced by representing the narrow band analysis signals at a reduced sampling rate must be removed at the synthesis filter.

A second approach towards the realization of analysis / synthesis system is to use frequency domain transform. The analysis filter first windows the samples before applying transforms like DFT, DCT etc. The transform domain samples form the channel signals. To synthesize, the channel signals are inverse transformed and the resulting time sequence is multiplied by a synthesis window, overlapped and added to generate the reconstructed signal. By a similar argument we followed for a first approach, we can say that reconstruction of $x(n)$ will be perfect if the composite analysis-synthesis window responses overlap and add in the time domain is flat and any time domain aliasing introduced by the frequency domain representation is removed in the synthesis process.

Such Time Domain Alias Cancellation (TDAC) may be seen as the dual of Frequency Domain Alias Cancellation (FDAC). The analysis-synthesis systems for coding applications are so designed that the overall sampling rate at the output is same as that of the input systems satisfying such conditions are referred to as *critically sampled system.* In a K-band filtering approach, this can be achieved by decimating each channel signal by a factor of K. For the frequency domain, i.,e subband filtering approach critical sampling introduces frequency domain aliasing except in the case where the analysis and synthesis filters have rectangular response equal to the channel width. In time domain approach, the dual of subband filtering is time domain windowing and there, critical sampling introduces time domain aliasing, except if the analysis and synthesis windows are rectangular and equal to the decimation factor.

One popular techniques of critically sampled analysis-synthesis system having overlapped channel frequency responses and reconstruction is the Quandrature mirror filtering (QMF). Overlap exists only between adjacent channel filters. For these systems, frequency domain aliasing is introduced by the analysis filters, but these are corrected by the synthesis filters. The techniques can be best described by single sideband modulation (SSB), since the channel signals are real.

A corresponding time domain description of SSB results in a block transform implementation of filter banks, using DCT and Discrete Sine Transforms (DST). An efficient weighted overlap-add (OLA) analysis and synthesis can be used to reconstruct the samples through time-domain alias cancellation. A detailed mathematical analysis of such system, along with the necessary conditions for perfect reconstruction are described by Princen *et al* [1] and is not repeated here. A qualitative description is however presented in the following section.

29.6 Mechanism of time-domain aliasing cancellation

Fig 29.3 illustrates the mechanism for time-domain alias cancellation. The input signal is first windowed with overlap between the adjacent windows. Let us first consider that the block time m is even. The recovered sequence after the forward and the inverse transform (in this case, the DCT) contains time reversed aliasing distortion, shown by the dashed curve. Although, for illustration purposes we have shown the original signal and the aliased signal separate, in practice the sum of these two would be observed. The sequence can be interpreted as periodic with period K. The synthesis window extracts a portion of the sequence of length K. In the next block time that is with m odd, the window is shifted by K/2 and a forward and an inverse DST are performed on the windowed sequence. The output contains time-reversed aliasing distortion, as shown by the dashed curve. It may be noted that the aliased samples in odd block time is opposite in sign as compared to those in even block time. The aliasing terms from the upper edge of segment from block time *m*, and from the lower edge of segment from block time *m+1* are equal and opposite and hence cancelled when these are overlapped and added.

29.7 Implementation and Window Design for TDAC

Implementation of the TDAC approach follows the block diagram shown in fig 29.4. The input signal is multiplied by an analysis window and then a transform (DCT/DST) is applied. The syntheses involve a corresponding inverse transform (IDCT/IDST), multiplication by a synthesis window, followed by overlap and add to obtain alias-free time samples. Time domain designs are more efficient for a given number of bands as compared to the frequency-based designs. For example, a critically sampled design for a 32-band system using frequency domain techniques, having stop-band attenuation greater than 40 dB and passband ripple less than 0.2 dB would require an 80-sample window. A critically sampled time domain design, in contrast requires just a window of 32 samples to implement filters with the same characteristics.

Two possible window design using TDAD are shown in fig 29.5 and the corresponding coefficient values are shown in table 29.1. The Modified Discrete Cosine Transform (MDCT) approach, adopted in MPEG-1 audio coding standard is a variant of the TDAC approach developed by Princen and Bradley [1]. In the text lesson we are going to present the design of analysis and synthesis filters.

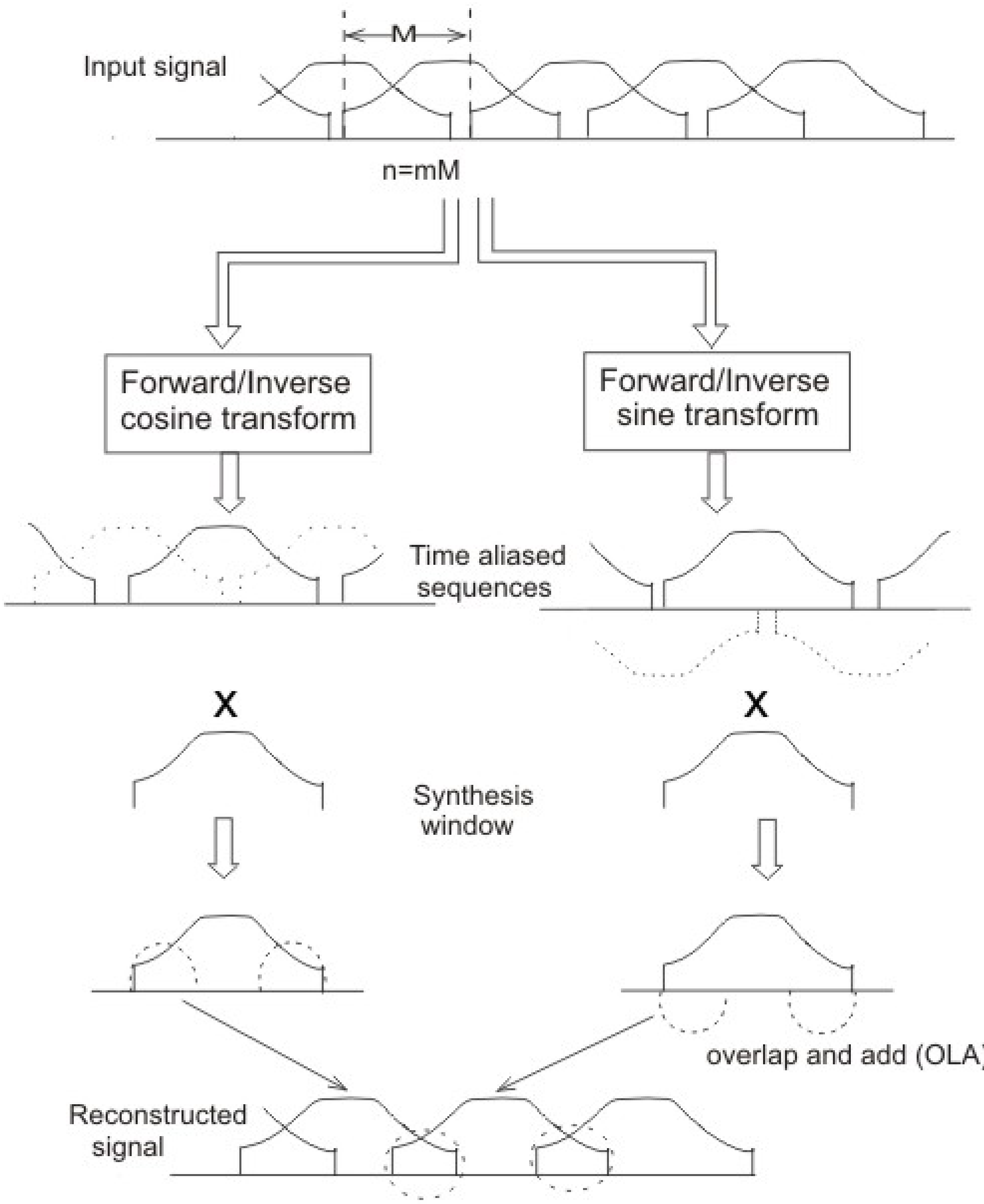

FIGURE 29.4 The mechanism of aliasing cancellation through DCT/DST and OLA

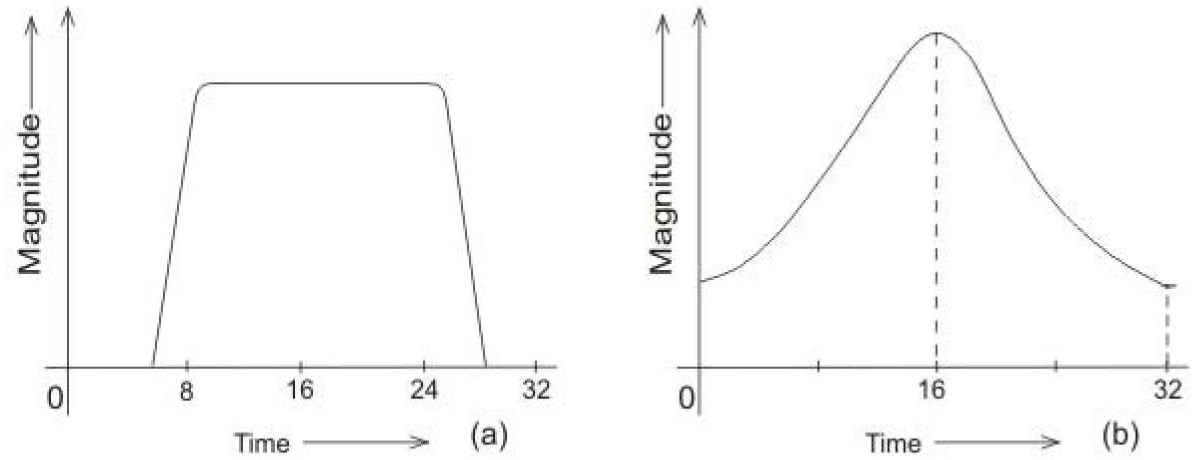

FIGURE 29.5 (a) A 20-point window for 32-band system;
(b) A 32-point window for a 32-band system.

Table 29.1 Window coefficient for two possible design in a 32 band system. (Courtesey Princen and Bradley in reference [3]. Note that the windows are symmetric, i.,e *h(31-r) = h (r))*

Coefficient no (r)	Design –1 (20 point window)	Design –2 (32-point windows)
0	0	0.18332
1	0	0.26722
2	0	0.36390
3	0	0.47162
4	0	0.58798
5.	0	0.70847
6.	0.5000	0.82932
7.	0.86603	0.94553
8.	1.11803	1.05202
9.	1.32287	1.14558
10.	1.41421	1.22396
11.	1.41421	1.28610
12.	1.41421	1.33307
13.	1.41421	1.36655
14.	1.41421	1.38866
15.	1.41421	1.40212

References

1. Peter Noll, “ MPEG Digital Audio Coding”, *IEEE signal Processing Magazine*, September 1997, pp-59-81.

2. Davis Pan, “ A tutorial on MPEG/Audio Compression*”. IEEE Multimedia*, Vol.2, No.2, 1995, pp 60-74.

3. John P. Princen and A. B. Bradley, “ Analysis/ Synthesis Filter Bank Design based on time domain Aliasing Cancellation”. *IEEE Transaction on Acoustics, Speech and Signal Processing*, Vol. ASSP-34, No. 5, October 1986, pp. 1153-1161.