Module 9
AUDIO CODING

Lesson 28
Basic of Audio Coding

Instructional Objectives

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

1. Name at least three different audio signal classes.
2. Calculate the bit-rate requirements for stereo quality audio.
3. State the basic requirements of low bit-rate audio coders
4. Outline the scope of MPEG audio standards.
5. Define critical bands of auditory response system
6. Define simultaneous masking and masking threshold
7. Define signal to mask ratio (SMR) and noise to mask ratio (NMR)
8. Define temporal masking
9. State the objectives of perceptual coding
10. Present the block diagram of perception based audio coders.

28.0 Introduction

Over the past 15 years, we have witnessed a phenomenal growth in the digital audio technology, which form an essential part of multimedia standards and technology. Digital audio finds widespread application in domains such as CD/DVD storage, digital telephony, satellite broadcasting, consumer electronics etc. In the last two modules, we have extensively covered the technological aspects of still images and video sequences, along with their standards. In this module, we are going to focus upon the digital audio. The current lesson, that is the first one in this module covers the basics of audio coding. It introduces the concepts of critical bands and masking, which form the background of perceptual audio coding based on which the audio coding standards are framed.

28.1 Audio signal classes

Based on the applications, audio signals may be grouped into three major classes.

- ***Telephone speech*:** This is a low bandwidth application, covering the frequency range of 300-3400 Hz. The intelligibility and naturalness of speech, though poor, is just acceptable for telephony and some video telephony services.

- ***Wideband speech:*** This covers a bandwidth of 50Hz-7000Hz for improved speech quality.

- ***Wideband audio:*** This includes high fidelity audio (speech, as well as music) applications requiring a bandwidth of at least 20 KHz for digital audio storage and broadcast applications.

28.2 Bit-rate requirements for stereo quality audio

In the early years of digital audio technology Compact Disc (CD) quality stereo audio was used as a standard having the following specification :

- *Sampling frequency* : 44.1 KHz
- *16-bits/ sample for each of the two stereo channels*

Therefore, the net bit-rate required is 2 x 16 x 44.1 x 10 = 1.41 Mbits/sec. However, considerable extra bits are required for synchronization and error correction, resulting in 49 bits for every 16-bit audio sample. Thus, the total stereo bit-rate requirement 1.41 x 49/16 Mbit/sec = 4.32 Mbit/sec.

Although high bandwidth channels are available, there is a necessity to achieve compression for low bit rate applications in cost effective storage and transmission. In applications such as mobile radio, channels have limited capacity and efficient bandwidth compression must be employed.

28.3 Basic requirements of low bit-rate audio coders

The low bit-rate audio coders should fulfill the following requirements:

- Robustness against variations in audio levels and spectrum.
- Robustness against random and bursty channel errors and packet losses.
- Low complexity and low power consumption.
- Low encoder/ decoder delays.
- Graceful degradation of quality with increasing bit error rates in mobile radio and broadcast applications.

Over the past few years, there have been significant research contributions to fulfill these objectives. Recent results in speech and audio coding indicate that an

excellent coding quality can be achieved with bit rates of 0.5 to 1 bit/ sample for speech and wideband speech and 1 to 2 bit/sample for audio.

28.4 Scope of MPEG audio standards

The MPEG (Moving Pictures Experts Group) provided the standards for digital audio coding, as a part of multimedia standards. Three standards viz, MPEG-1 MPEG-2 and MPEG-4 catered for the following requirement :

a) **MPEG-1 audio** : In this standard, out of a total bit rate of 1.5 Mbit/sec for CD quality multimedia storage, 1.2 Mbits/sec is allocated to video and 256 Kbits/sec is allocated to audio. Up to two channels of audio are accommodated.

b) **MPEG-2 audio** : This standard fulfils the requirements of HDTV applications. In its audio part, two to five full bandwidth audio channels are accommodated. The standard also offers a collection of tools known as Advanced Audio Coding (MPEG-2 AAC)

c) **MPEG-4 audio** : The MPEG-4 standards for audiovisual coding addresses applications ranging from mobile access, low complexity multimedia terminals to high complexity multichannel sound systems. The major feature of MPEG audio is that instead of using any model for audio source (like vocal tract model used for speech signals), the coders exploit the perceptual limitations of the human auditory system. Compression is achieved by eliminating the perceptually irrelevant parts of audio, which cannot produce any audio distortion.

28.5 Human auditory perception

The human auditory system is fairly complicated, but after conducting a large number of pshychoacustic tests the human auditory response system is believed to perform short term critical band analysis and may be modeled as a bank of bandpass filters with overlapping frequencies. The power spectrum is not on linear frequency scale and the bandwidths are in the range of 50-100 Hz for signals below 500 Hz and up to 5000 Hz for higher frequencies. Such frequency bands of auditory response system are referred to as *critical band*. Up to 26 critical bands covering frequency range of 24 KHz are considered.

28.5.1 The masking phenomenon:

It is observed that a low level audio signal is rendered inaudible, if there is a *simultaneous occurrence* of a stronger audio signal, which is close in frequency to the former. This phenomenon is known as *masking*. The stronger signal that masks the weaker one is called *masker* and the one that is masked is the *maskee*. It is furthermore observed that the masking is largest in the critical band

in which the masker is located and to a lesser degree, masking is also effective in the neighboring bands.

It is possible to define a *masking threshold*, below which the presence of any audio will be rendered inaudible. The *masking threshold* will depend upon the sound pressure level (SPL), the frequency of the *masker* and the characteristics of the masker and the *maskee*, such as whether the *masker* or *maskee* is a tone or noise.

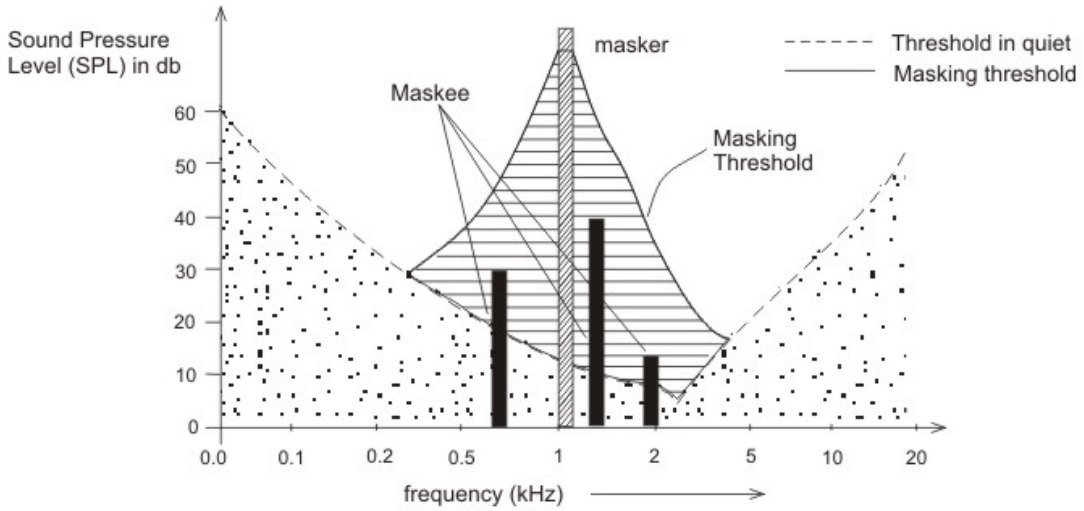

FIGURE 28.1 Effects of masking in presence of a masker at 1 kHz

Fig. 28.1 Illustrates the concepts of masker, maskees and masking threshold.

In this diagram, we have considered a strong signal at 1 kHz, acting as the masker. The masking threshold (shown in solid line) drops down considerably, as we deviate from the masker frequency. The three solid bars have their SPL s below the masking threshold and are therefore masked. The dotted curve indicates quiet threshold, that is, without the presence of any masker. It is interesting to note that the quiet threshold is of lower value at the frequency range (500 Hz-5 kHz) of the audio spectrum.

It is furthermore observed that the slope of the masking threshold is steeper towards the lower frequencies, that is lower frequencies are not masked to the extent in which the higher frequencies are masked.

The masking characteristics is measured by he following parameters :

a) ***Signal to mask ratio (SMR)*****:** The *SMR* at a given frequency is expressed as the difference (in dB) between the SPL of the masker and the masking threshold at that frequency.

b) ***Mask to noise ratio (MNR):*** The *MNR* at a given frequency is expressed as the difference (in dB) between the masking threshold at that frequency and the noise level. To make the noise inaudible, its level should be below the masking threshold i.,e the *MNR* should be positive.

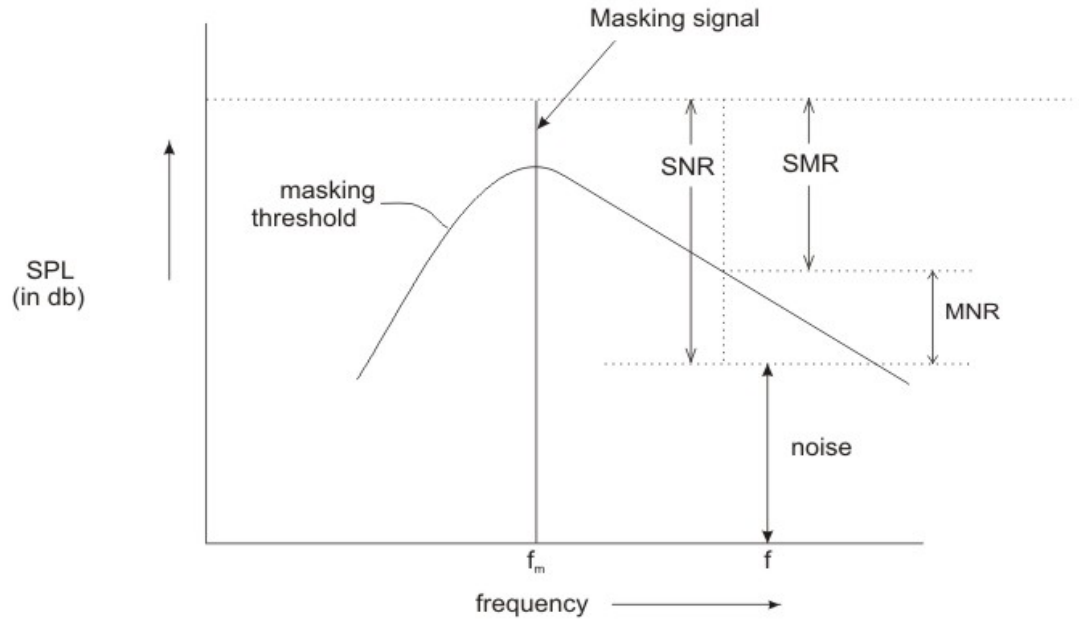

FIGURE 28.2 Masking characteristics SMR and MNR.....

As illustrated in fig.28.2, a masking signal occurs at frequency f_m, giving rise to the masking threshold curve, as shown. At a given frequency *f*, the *SMR*, the signal to noise ratio (*SNR*) and the *MNR* corresponding to the noise level is shown. It is evident that

SMR (f) = SNR (f) – NMR (f)

We have so far considered only one masker. If there are more than one maskers, then each masker contributes to its own masking threshold and a *global masking threshold* is computed that describeds*just noticeable distortion* as a function of frequency.

28.5.2. Temporal masking

The masking phenomenon described in the previous subsection is also referred to as *simultaneous masking*, where the masker and the maskee are assumed to occur at the same time instant. Masking is also observed when two sounds appear within a short time interval and the stronger one mask the weaker one. This phenomenon is known as *temporal masking*. In addition to simultaneous masking, temporal masking also plays a major role in human auditory perception.

Temporal masking is possible, even if the maskee precedes the masker by a short time gap and is associated with *premasking* and *postmasking* where the former has one-tenth duration as compared to the latter. The order of postmaking duration is 50 to 200 *ms*.

28.5.3. Perceptual coding in MPEG audio:

An efficient audio source coding algorithm should fulfill the following two basic objectives:

a) *redundancy removal* , in which the statistical redundancies between the adjacent samples are exploited, and

b) *irrelevance removal*, which is perceptually motivated since anything that our ears cannot hear, may be removed.

In irrelevance removal, simultaneous and temporal masking phenomena play dominant roles in MPEG audio coding. We have already noted that the noise level should be below the masking threshold. Since the noise due to quantization is a function of the number of bits to which the samples are quantized, the bit allocation algorithm has to take masking characteristics into consideration. Fig 28.3 shows the block diagram of a perception based coder, that exploits the masking phenomenon.

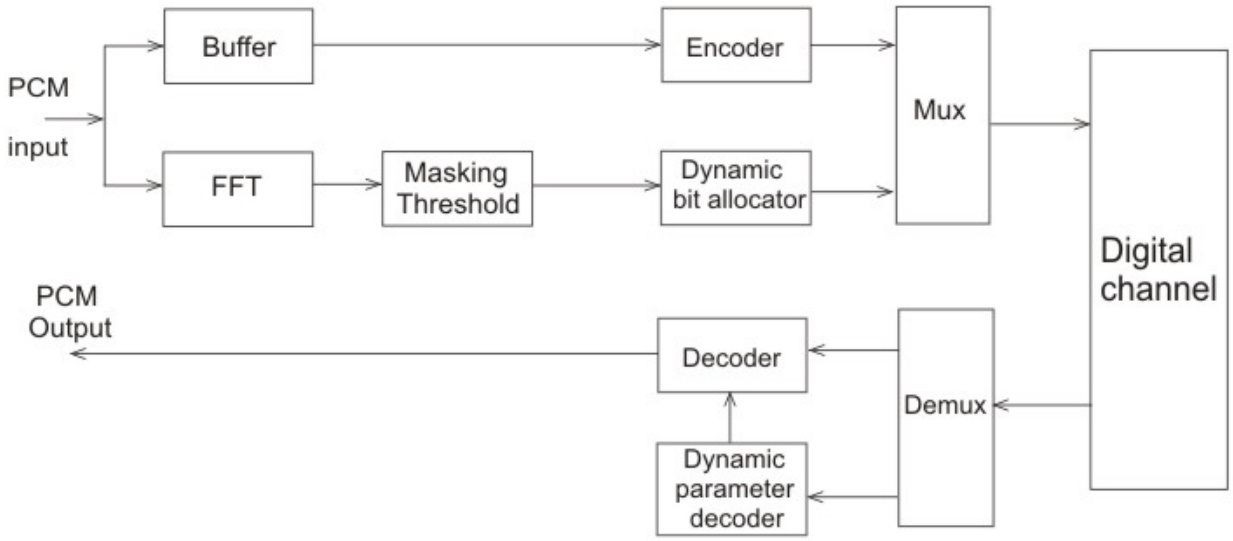

FIGURE 28.3 Block-diagram of perception-based coder

As shown, Fast Fourier Transform (FFT) is computed from the incoming PCM audio samples and the complete audio spectrum is obtained, from which the

tonal components of masking signals can be determined. Using this, a global masking threshold and also the *SMR* in the entire audio spectrum is computed. The dynamic bit allocator to encode the bit stream uses the *SMR* information. It is possible to make the coding scheme perceptually transparent by keeping the quantization noise below the global masking threshold. A perceptually transparent encoding process will make the decoded output indistinguishable from the input.

It may however be noted that our knowledge in computing the global masking threshold is limited in the sense that the perceptual model considers only simple and stationary maskers and it may fail in some practical situations. The solution to this problem is to keep sufficient safety margin.

28.6 Conclusion

In this lesson, we have learnt some of the basic ideas pertaining to audio coding. For an efficient encoding, the perceptual aspects of masking have been utilized in the MPEG standard. In subsequent lesson we are going to learn further details of the audio codecs.