Module 9

AUDIO CODING

Lesson 32

Psychoacoustic Models

Instructional Objectives

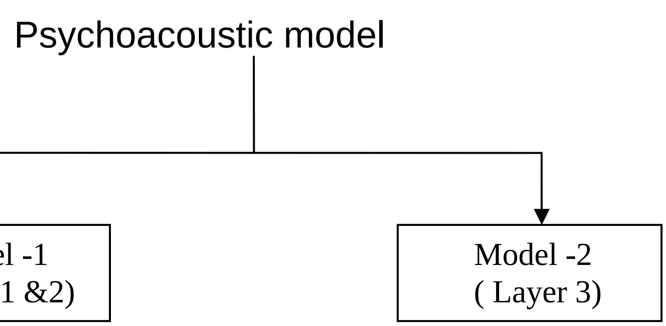
At the end of this lesson, the students should be able to

1. State the basic objectives of both the psychoacoustic models.
2. Identify the tonal components from an auditory spectrum.
3. Identify the non-tonal components from an auditory spectrum.
4. Prune the list of tonal and non-tonal components using a sliding window.
5. Define masking index, masking function and masking threshold for both tonal and non-tonal components.
6. Calculate the global masking thresholds.
7. Explain the partition-domain transformation for psychoacoustic model-II.

32.0 Introduction

In the last few lessons, we have discussed the basic philosophy of audio encoding and the bit allocation policy followed in MPEG-1 audio standard. The MPEG-1 audio standard follows two psychoacoustic models. In this lesson, we are going to cover the two psychoacoustic models in details.

32.1 Psychoacoustic model classification



Model – 1:
- is computationally simple.
- has high accuracy at high bit rate.

Model – 2:
- is computationally complex.
- has high accuracy at low bit rate.

Essential philosophies of both the models:

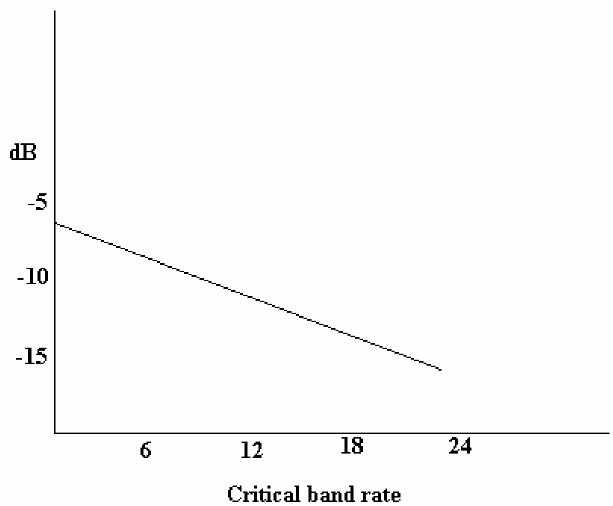
- Compute Fourier power spectrum of the signal. (512 point FFT for layer 1 & 2 / 1024 point FFT for layer 3).
- Map the spectrum into critical band domain.
- Distinguish between the tonal and non-tonal components.
- Calculate the masking function.
- Map these functions back to the sub-band domain.

32.2 Psychoacoustic model – I

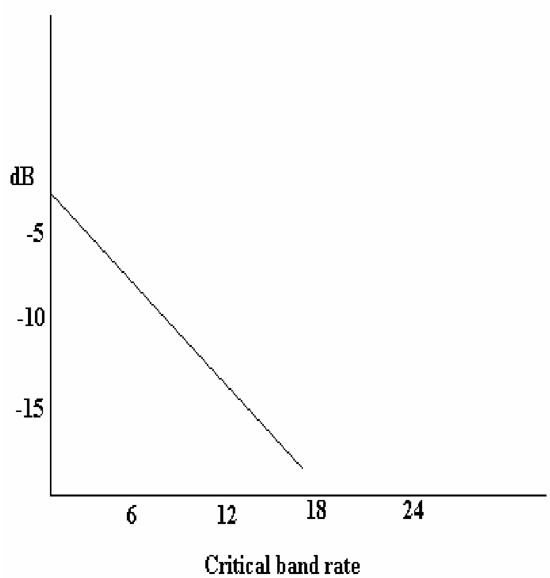
- The auditory spectrum is approximated by a list of tonal and non-tonal components.
- Tonal components are selected by identifying the maxima in the spectrum whose height is greatest in the neighbourhood.
- All the remaining spectral lines are used for calculating the non-tonal components. They are grouped into critical bands. Within each critical band, a non-tonal component is represented.
- Then, the list of tonal and noise components are decimated by eliminating those components which are below the auditory threshold or are less than one half of a critical band width from a neighbouring component.
- To compute the masking effect of a tonal or non-tonal component on the neighbouring spectral frequencies, the strength of the component is summed with two terms called the masking index and the masking function.
    - Masking index: An attenuation term which depends on the critical band rate of the component and whether it is tonal or non-tonal.
    - Masking function: An attenuation factor which depends on –

- Displacement of the component from neighbouring frequency.

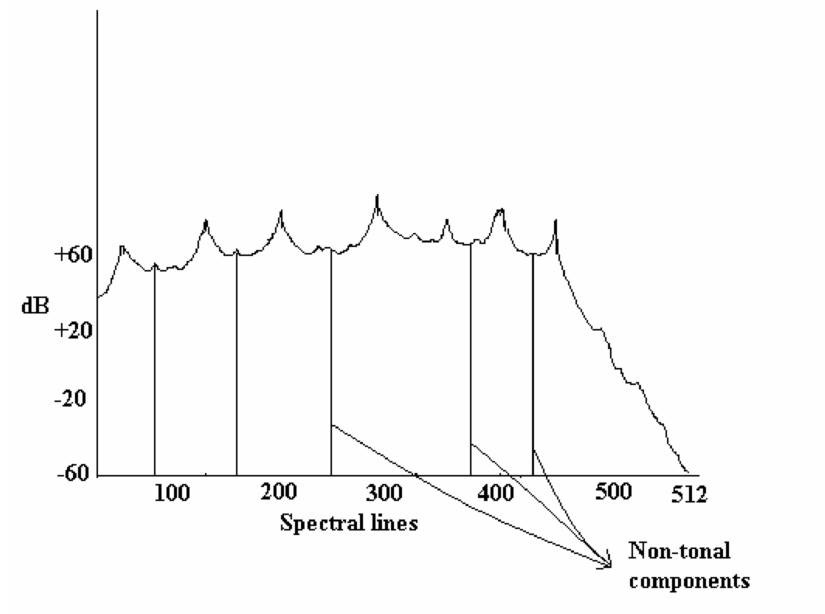
- The component signal strength**.**

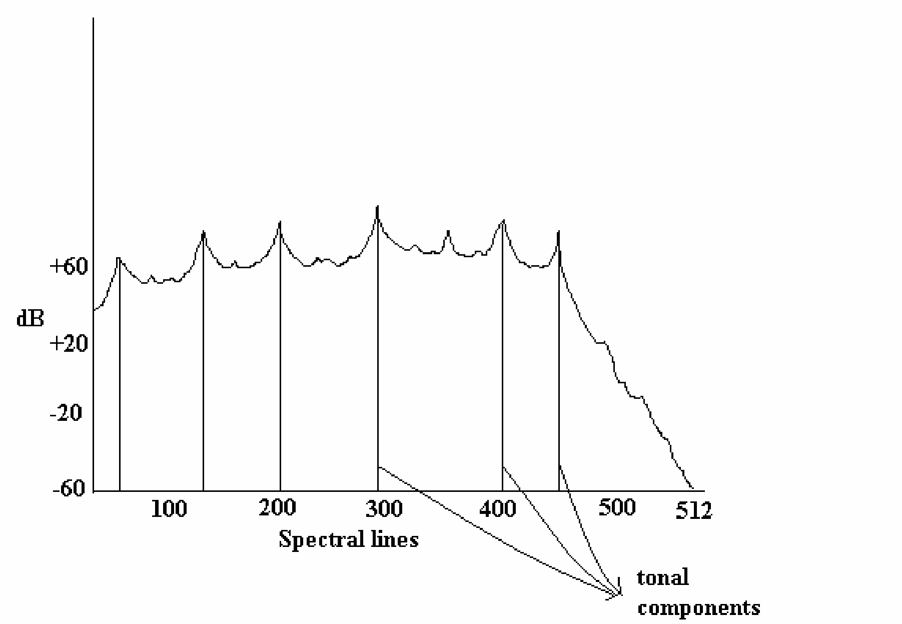


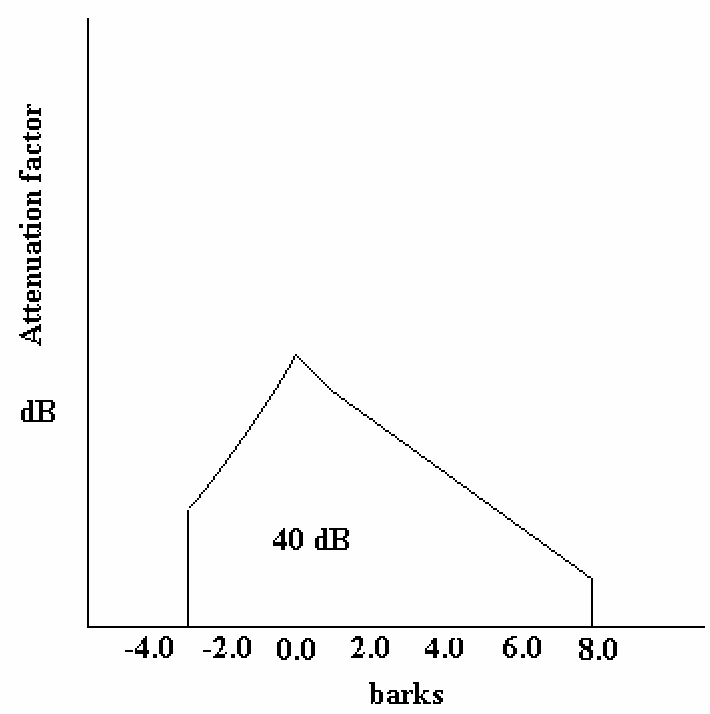
Tonal masking index

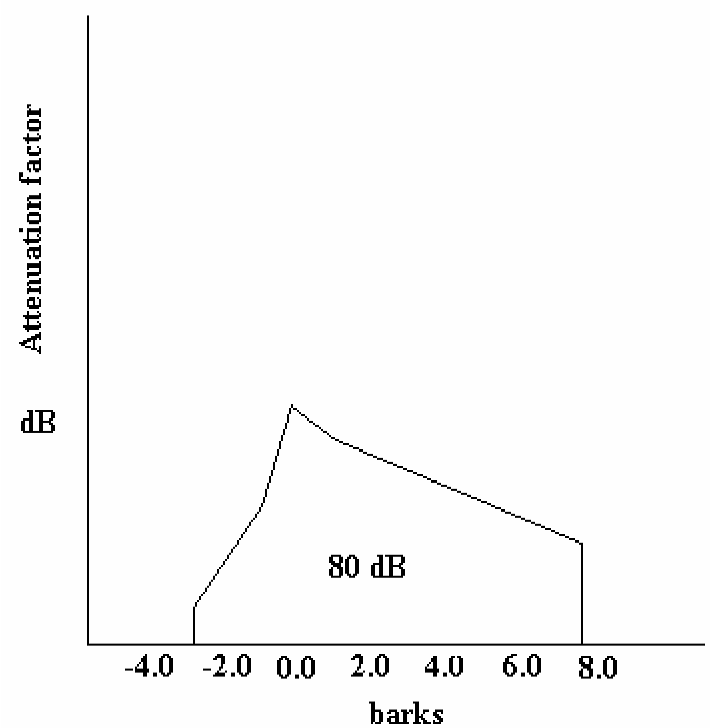


Non-tonal masking Index


+60
dB
+20
-20
-60
100
200
300
400
500
512
Spectral lines
Non-tonal
components


+60
dB
+20
-20
-60
100
200
300
400
500
512
Spectral lines
tonal
components


Attenuation factor
dB
40 dB
-4.0 -2.0 0.0 2.0 4.0 6.0 8.0
barks


Attenuation factor
dB
80 dB
-4.0 -2.0 0.0 2.0 4.0 6.0 8.0
barks

For a tonal component j, at critical band rate z(j), the masking threshold $LT_{tm}(j,i)$ at critical band rate z(i) is given by

$$LT_{tm}(j,i) = X_{tm}(j) + av_{tm}(z(j)) + vf[z(i) - z(j), X_{tm}(j)]$$

where,

X_{tm} (j) is the strength of tonal component at frequency j,

av_{tm} (j) is the tonal masking index at the critical band rate z(j),

vf (i,j) is the masking function with i representing displacement

and j representing signal strength.

For non-tonal components the masking index can be calculated as:

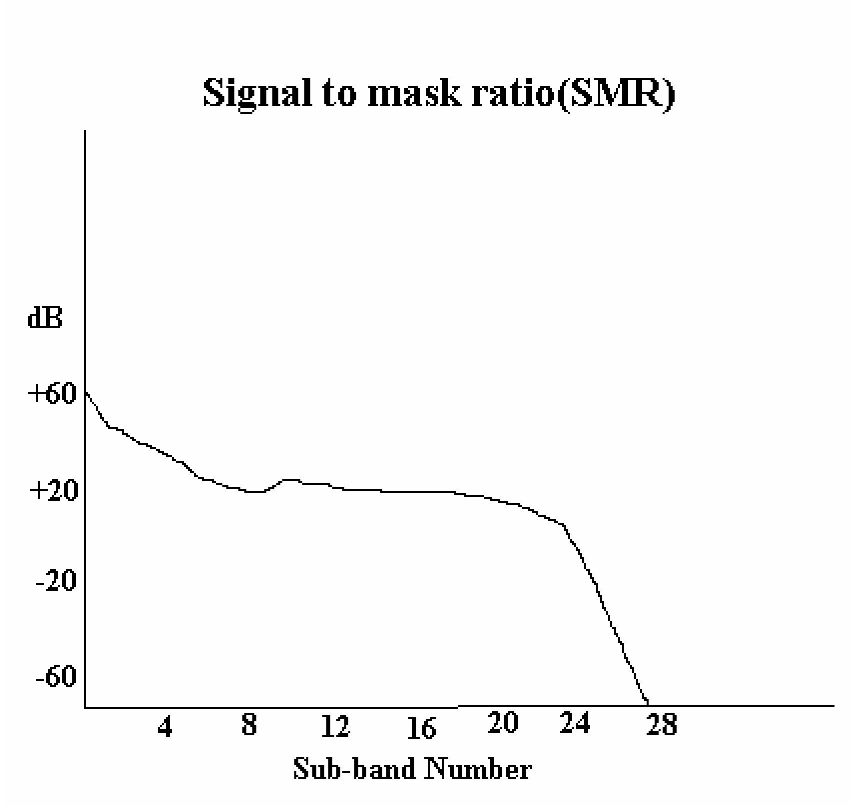
$$LT_{nm} (j, i) = X_{nm} (j) + av_{nm}(j) + vf[z(i) - z(j), X_{nm}(j)]$$

.

- The global masking thresholds are computed for all spectral frequencies by adding the masking thresholds computed above, for all the neighbouring tonal & non-tonal components, with the threshold of hearing.

- Then, the minimum masking threshold function is determined for each sub-band from the minimum of all the global masking thresholds contributing to that sub-band.

- Finally, signal to mask ratio (SMR) is computed.



32.3 Psychoacoustic model – II

- It does not make a distinction between the tonal and non-tonal components.
- Spectral data is transformed into a partition domain.
- 1024 point FFT computation is used.
- Tonality is decided by the unpredictability of the spectrum with time.

32.3.1 <u>Layer –III encoding</u>

There are several new features in the Layer-III encoding scheme:

- Each of the 32 sub-bands in now split up into 18 spectral lines using the Modified Discrete Cosine Transform(MDCT) with window length 36 followed by a sequence of alias reduction butterflies.
- Variable frequency/time resolution is used to control transient effects.

- Variable bit rate coding using a choice of 32 Huffman tables is used to encode the data.