DASP3rd Chapter 9 - Exercises

1 Psychoacoustics

1. Human hearing

   (a) In which frequency range is a human able to perceive sounds?

   Solution:

   - 20 Hz -20 kHz

   (b) What is the frequency range of speech?

   Solution:

   - 125 Hz-8 kHz

   (c) In the above specified range, where is human hearing most sensitive?

   Solution:

   - 2 kHz-4 kHz

   (d) How is the absolute threshold of hearing determined?

   Solution:

   - The absolute threshold can be given by

$$\frac{L_{T_q}}{\text{dB}} = 3.64\left(\frac{f}{\text{kHz}}\right)^{-0.8} - 6.5\text{exp}\left(-0.6(\frac{f}{\text{kHz}} - 3.3)^2\right) + 10^{-3}\left(\frac{f}{\text{kHz}}\right)^4$$

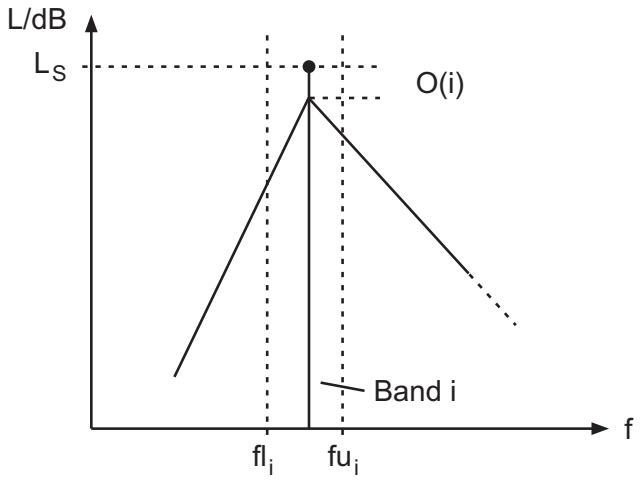
   - Absolute threshold of hearing:
     * Play for many subjects a sinusoidal tone at very low level of power.
     * Increase slowly the power until the subjects can hear the tone.
     * Repeat the process for all frequencies in the human auditory range.
   - Refer to the plot of absolute threshold in Fig 9.5.

2. Masking

   (a) What is frequency-domain masking?

   Solution:

   - Frequency masking is a phenomenon by which a masking signal (the masker) with a certain frequency range of very high amplitude can make the neighboring frequencies inaudible if they are of lower amplitude than the threshold of the masking signal (see Fig. 9.15 shown below).



(b) What is a critical band and why is it needed for frequency masking phenomena?

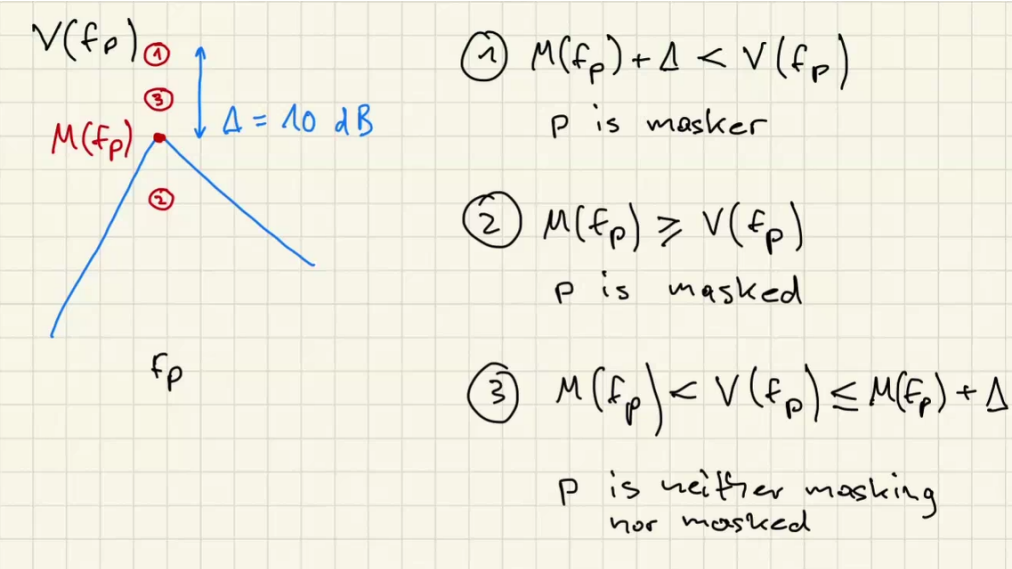
Solution:

- Critical bands are those frequency groups obtained due the capability of human hearing to subdivide the sound events in frequency groups. It refers to the smallest band of frequencies around a given frequency, which activate the same part of the basilar membrane. One Bark is equal to a width of one critical band. A transformation of the linear frequency scale into a hearing adapted scale is given by

$$\frac{z}{\text{Bark}} = 13 \arctan\left(0.76\frac{f}{\text{kHz}}\right) + 3.5 \arctan\left(\frac{f}{7.5\text{kHz}}\right)^2. \tag{1}$$

(c) Consider a_p and f_p to be respectively the amplitude and the frequency of a partial at index p and $V(a_p)$ to be the corresponding sound pressure level in dB. The difference between the level of the masker and the masking threshold $M(f_p)$ is $\Delta = 10\,\text{dB}$. The masking curves toward lower and higher frequencies are described respectively by a left slope (27 dB/Bark) and a right slope (15 dB/Bark). Explain the main steps of frequency masking in this case and show with plots how this masking phenomena is achieved.

Solution:



(d) What are the psychoacoustic parameters used for lossy audio coding?

Solution:

- The important parameters are the masking threshold, critical bands, and absolute threshold.

(e) How can we explain temporal masking and what is its duration after stopping the active masker?

Solution:

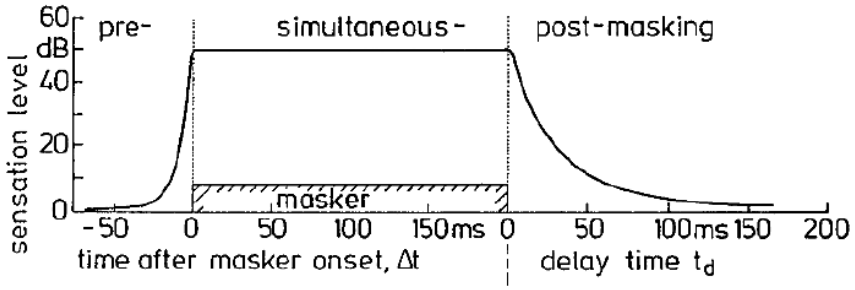


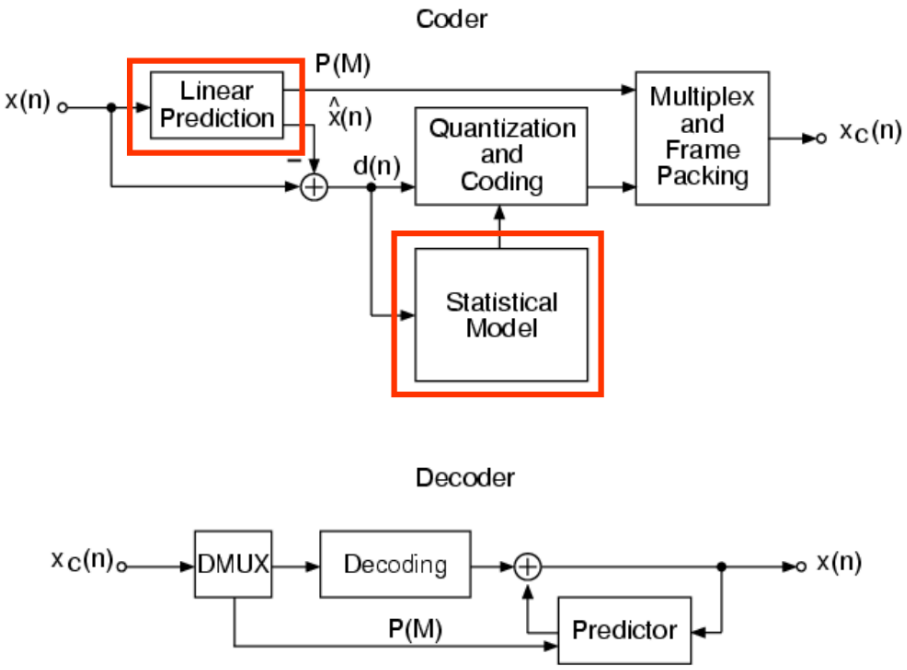
Fig. 4.17. Schematic drawing to illustrate and characterize the regions within which premasking, simultaneous masking and postmasking occur. Note that post-masking uses a different time origin than premasking and simultaneous masking [Zwi90]

- Ability of a sudden (transient) sound to make other sounds inaudible, present preceding or following the sudden sound.
  * Pre-masking duration: up to 50 ms
  * Post-masking duration: up to 200 ms

2 Audio coding

1. How do a lossless coder and decoder work.

   Solution:



2. What is the achievable compression factor for lossless coding?

   Solution:

   - The achievable compression factor depends on the statistics of the audio signal and allow a compression rate of up to two.

3. How do an MPEG-1 Layer 3 coder and decoder work.

   Solution:

   - See simplified block diagrams of ISO-MPEG1 coder and decoder in Figs. 9.13 and 9.14 on page 304.

4. How do an MPEG-2 AAC coder and decoder work.

   Solution:

   - See MPEG-2-AAC coder and decoder in Fig. 9.19 on page 311.

5. What is temporal noise shaping?

   Solution:

- Temporal noise shaping (TNS) is a method for adapting the time-frequency resolution of a filter bank, where a MDCT/ IMDCT to the signal characteristic is based on linear prediction along the spectral coefficients in frequency domain. It is a weighting of the temporal envelope of the time domain signal.

6. How do an MPEG-4 coder and decoder work.

   Solution:

   - See MPEG-4 parametric coder and decoder in Figs. 9.32 and 9.36 on pages 322 and 324, respectively.

7. What is the benefit of SBR?

   Solution:

   - Spectral band replication (SBR) is beneficial in reducing the bit rate during audio coding. In SBR, the audio signal is decomposed into lowpass (LP) and highpass (HP) components. The LP part is downsampled and coded by a standard coder. The HP part is not directly coded. Instead, it is reconstructed from the lowpass (LP) part, and the corresponding difference from the actual HP part along with the HP spectral envelope are coded.