Human Auditory System

Audio Processing Guide

ELECTGON
www.electgon.com
ma_ext@gmx.net

03.06.2019

Contents

1 Introduction 3
2 Properties of Sound 3
3 Human Ear Interaction 4
4 Critical Bands 6
5 Auditory Model 7
6 Masking 7
6.1 Simultaneous Masking 8
6.2 Temporal Masking 10

1 Introduction

Acoustics is the study of sound and is concerned with the generation, transmission and reception of sound waves. Generation of sound starts when energy makes disturbance in a medium, molecules of this medium starts to vibrate back and forth. For example ringing bell generates a sound because an incident energy hit the bell causing disturbance in the bell body, this means that bell body surface starts to vibrate. In consequence, air molecules surrounding this bell will be affected by this vibration and its molecules start to vibrate. Thus bell body surface is causing disturbance in air by pushing on air molecules forth and back. This disturbance consists of regions of pressure above and below equilibrium pressure. This pressure is causing displacement in the air molecules. The local displacement of air molecules occurs in the direction in which the disturbance is traveling. Thus sound undergoes to longitudinal form of transmission. A receptor placed in the sound field similarly moves according to the pressure acting on it. The denser the medium, the easier is the task of propagation, for example sound travels more easily in water than in air.

2 Properties of Sound

Sound can be defined as mechanical oscillation or vibration of an elastic medium that potentially can be heard. This oscillation is referred to a wave. We have then a mechanical energy has been transduced by sound source to acoustical energy which is contained in a propagating sound wave. Dealing with sound as a propagating wave, enables us defining its properties. As a general properties of any wave, sound wave undergoes to constructive or destructive interference with other sound waves. It undergoes to diffraction, in which it bends through opening or around obstacles. It can refract, in which it bends because of its velocity changes.

What is important for research scope here is to shed light on other attributes of the sound wave, namely amplitude and frequency. Amplitude of the sound wave describes the sound pressure displacement above and below equilibrium pressure. In absolute pressure terms, sound pressure is very small. If atmospheric pressure is 101325 Pa, for human ear, a loud sound might cause a 20 Pa (this is actually highest pressure that human ear can receive safely) [1]. However, sound wave is not measured based on this absolute pressure value but is measured by Sound Pressure Level (SPL) which defines the sound wave pressure relative to minimum pressure that human ear can sense. According to ISO 1683-2 standard, minimum pressure detected by healthy human ear is 20 micro Pa [1]. Remembering that maximum pressure is 20 Pa, linear scale of between these minimum and maximum values is not appropriate to deal with, instead, logarithmic scale is used to represent sound pressure then which is defined as follows $20\,log_{10}\frac{p_1}{p_0}$ with P_0 as the reference value which is minimum detectable pressure 20 µPa. P_1 is the pressure that we need to identify its worthy level. According to that, maximum sound pressure level for human 140 dB, minimum pressure is 0 dB. Table 1shows different sound pressure level cause by some daily activities.

Source / observing situation	Typical sound pressure level (db SPL)
Hearing threshold	0 dB
Leaves fluttering	20 dB
Whisper in an ear	30 dB
Normal speech conversation for a participant	60 dB
Cars/vehicles for a close observer	60-100 dB
Airplane taking-off for a close observer	120 dB
Pain threshold	120-140 dB

Table 1: Sound Pressure Levels by Daily Activities [1]

Same values can be obtained if we expressed the sound wave in terms of its power (not pressure). Sound power is energy rate, i.e. Energy of sound per unit time from a sound source measured in watt. Sound intensity is the unit used to express the sound wave and is defined as sound power per unit area $I = \frac{P}{A} W/m^2$ We can with the previous approach define sound intensity level as

$$L_I = 10\, log \frac{I}{I_{ref}} \tag{1}$$

with I_{ref} is the minimum intensity that human ear can accept which is defined also by ISO 1683-2 as:

P_{ref} = 10^{-12} W

I_{ref} = 10^{-12} W/m^2

Another aspect of the sound wave also is the frequency of this wave. Human ear is able to operate with sound wave within specific frequency range from 20 Hz to 20 KHz. It means no thing to use sound signal out of this range for audio applications but there are many other medical and industrial applications in which other frequency ranges are used. Sound waves with frequencies higher than 20 KHz are called Ultrasonic waves. On the other side, sound wave with frequencies lower than 20 Hz are called infrasonic waves.

3 Human Ear Interaction

After we have seen properties of sound waves, it is important then to describe how human ear interact with these sound waves. It was mentioned before that sound as a wave with acoustical energy travels in the medium, assuming now this medium is air, this sound wave when received by human ear (received by the outer ear) it will be converted to mechanical energy (by middle ear). Inner ear will then convert it to electrical impulses which are sent to brain. Figure 1 shows human ear system and its components.

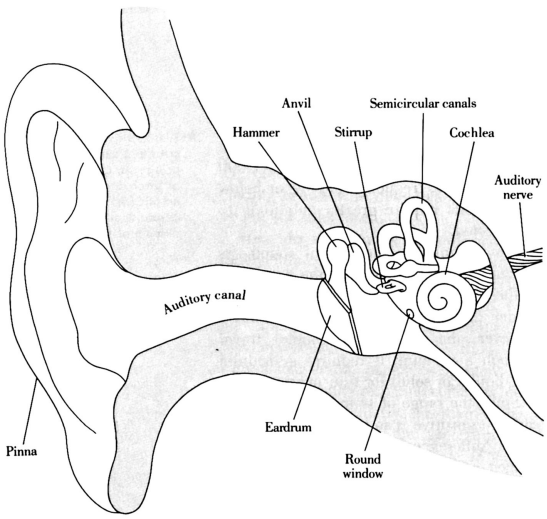

Figure 1: Cross Section of Human Ear[2]

What happens then is the outer ear collects sound waves which in turn pass through ear canal till it hit the eardrum. At this point acoustic wave will be converted to mechanical wave as the Hummer, Anvil and stirrup working as impedance matching to make efficient transition of this acoustic wave from air medium to fluid-filled medium of the inner ear. The inner ear has the Basilar membrane which deals with the incident wave in frequency domain. According to amplitude and frequency of the incoming wave, it generates electrical impulses to the brain. Basilar membrane has in its top Vestibular which is responsible for motion detection i.e. by this human brain can understand spatial dimension of the sound. Basilar membrane has about 30000 hair cells arranged in multiple rows. These hair cells detect membrane vibration and produce corresponding electrical impulses. Figure 2 depicts the Basilar membrane and its response to incident waves in terms of its frequency.

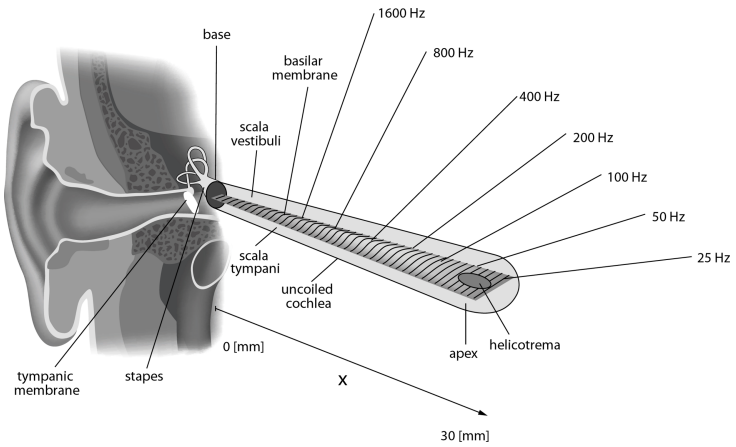

Figure 2: Basilar Membrane Response for Frequencies[3]

4 Critical Bands

Human ear then interacts with sound wave according to its frequency components. This frequency decomposition process is called tonotopicity. Experiments showed that at low frequencies tones of few Hertz apart can be distinguished but at high frequencies tones must differ by hundreds of Hertz. Now what if human ear received two tones at the same frequency but with different amplitude. In this case, human ear will respond to the strongest. Actually according to experiment done by Harvey Fetcher in 1940 this strongest component will cover tones of its frequency and any tone in near frequency. So it will have effect in its local region. This effect defined what is called Critical Bands of the human ear. Critical Band is range of frequencies in which the human ear responds to the strongest tone in that band. It was found that human ear has 24 critical bands. You can think of it as human ear has 24 frequency band pass filters, each filter passes the strongest tone. To define these critical bands, it means that we need to find Bandwidth of each band. This can be done experimentally by triggering sound tones at different frequencies with different intensity and observe response of human hearing. Table 2 was concluded as definition for human critical bands. These critical bands width has a measurement unit called Bark which is a non-linear frequency scale modeling the resolution of the human hearing system. One Bark distance on the Bark-scale equals to critical bandwidth.

Band no.	Start Frequency (Hz)	End Frequency (Hz)	Bandwidth (in Hz)	Bandwidth (in Bark)
0	0	100	100	0.5
1	100	200	100	1.5
2	200	300	100	2.5
3	300	400	100	3.5
4	400	510	110	4.5
5	510	630	120	5.5
6	630	770	140	6.5
7	770	920	150	7.5
8	920	1080	160	8.5
9	1080	1270	190	9.5
10	1270	1480	210	10.5
11	1480	1720	240	11.5
12	1720	2000	280	12.5
13	2000	2320	320	13.5
14	2320	2700	380	14.5
15	2700	3150	450	15.5
16	3150	3700	550	16.5
17	3700	4400	700	17.5
18	4400	5300	900	18.5
19	5300	6400	1100	19.5
20	6400	7700	1300	20.5
21	7700	9500	1800	21.5
22	9500	12000	2500	22.5
23	12000	15500	3500	23.5
24	15500	End of hearing		

Table 2: Critical Bands of Human Ear [7]

5 Auditory Model

Human ear has special sensitivity to sound waves. It responds to sound waves with special parameters. First parameter is the frequency of the sound wave. Human ear is able to hear only sound waves in range of 20 to 20 KHz. Second parameter is the sound level. Figure 3 shows region of audibility of human ear in terms of frequency and sound level. Third parameter is the arrival time of the sound. This arrival time is important for the case when more than one sound wave are received by the ear. For instance, and since human has two ears, the difference between arrival time of sound at each of the two ears together with the difference in intensity of the sound that reaches each ear is used by the auditory nervous system to determine the location of the sound source. Stereo or multichannel sound systems are built based on this property. Fourth parameter is discussed previously in so what is called critical bands which makes sound components of the same critical band is inaudible if they differ in sound level.

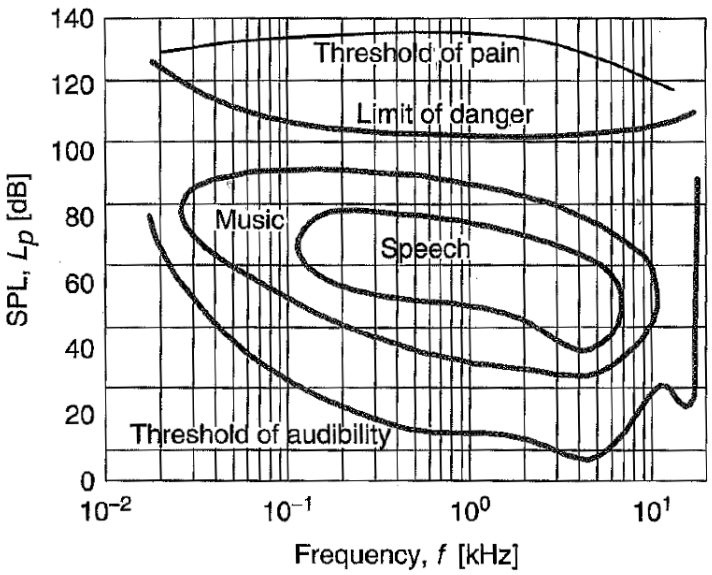

Figure 3: Audible Regions[4]

6 Masking

Human ear has special treatment for sound waves according to what is discussed previously. What is important about this nature is the ability to differentiate between sounds of different critical bands only. Sounds in the same critical band can't be distinguished. This leads to the masking property. Masking is the effect that happens when a strong sound tone covers adjacent less tone in the same critical band. What is meant by 'covers' here is the stronger tone makes the weak tone inaudible by human ear. Masking is a kind of interference with the audibility of a sound (called maskee) caused by the presence of another sound (called masker), if both of these sounds are close enough to each other in frequency and occur simultaneously or close to each other in time. Figure 4 discusses the idea of Masking in which three frequency components are placed in the same critical band and with center frequency 1.2 KHz but they

have different sound levels. Signal C will make masking for the other two signals

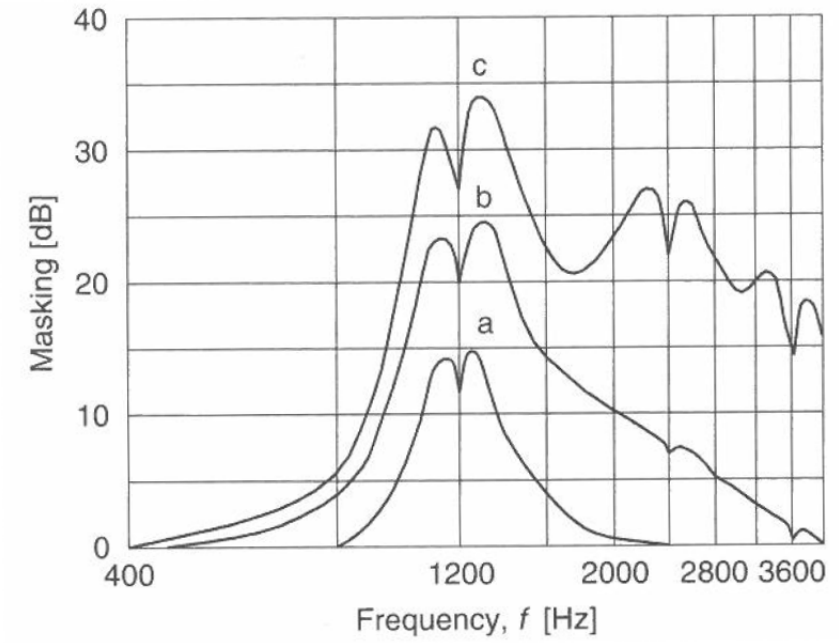

Figure 4: Masking between Signals[4]

If a lower level maskee is inaudible, because of a simultaneous existence of a higher level masker, this effect is referred to as the simultaneous masking. If an inaudible maskee precedes the masker or follows the masker, this phenomenon is called temporal masking. It worth mentioning here also that according to that concept of masking; four different cases can be distinguished: tone-masking-tone, noise-masking-tone, tone-masking-noise, noise-masking-noise. These four cases are just discussing what the masker and maskee signals, for example in case of tone-masking-noise we have an audio tone is a masker and associated noise is the maskee. Importance of first two cases comes when designing audio coders that is able to code audio signal without distortion. The latter two cases are important also for making quantization noise of these decoders are inaudible [4].

6.1 Simultaneous Masking

Threshold of audibility represented in figure 3 can be simplified as in figure 5. In which the minimum level of audible sound is plotted. Human ear is able to hear sound levels only above this plotted contour.

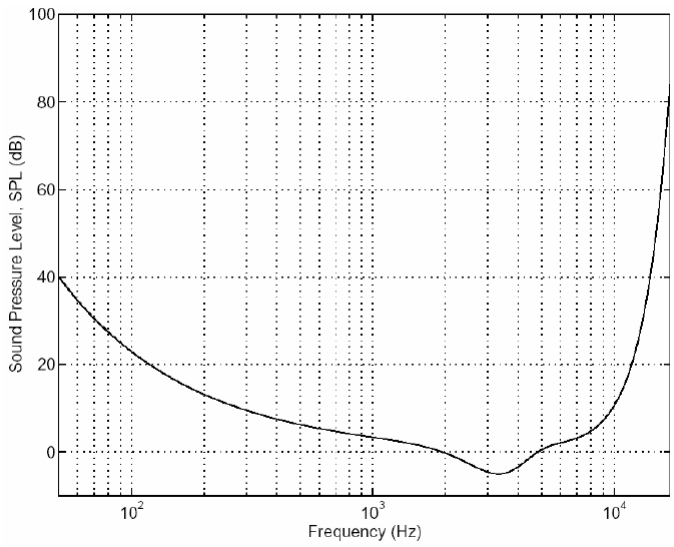

Figure 5: Absolute Threshold of Human Hearing[5]

When masking effect is happening, the behaviour of this threshold is changing around the band that has this masking. This can be explained as follows. An excitation takes place at the basilar membrane because of incident sound signals, the strong noise or masker blocks detection of the weaker signals within the same critical band. However, Inter-band masking has also been observed, i.e., a masker centered within one critical band changes detection thresholds in other critical bands so that it raise this threshold. This effect, also known as the spread of masking. Some other context approximating this change in threshold by a triangular spreading function. Figure 6 shows example of effect of spread masking on auditory threshold.

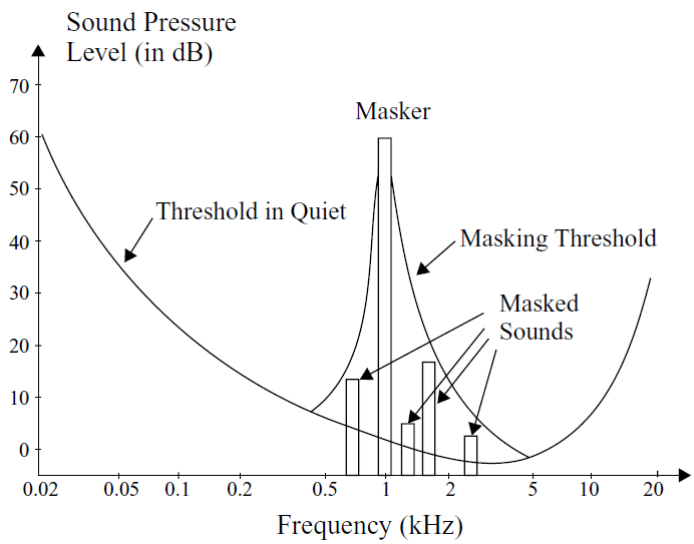

Figure 6: Spread Masking[6]

6.2 Temporal Masking

Masking effect can happen also in time domain. It is called then Temporal masking, in which a Masker can mask a preceding or following maskee signal. Figure 7 shows Temporal masking which is composite of premasking+simultaneous masking+postmasking. Premasking effect appears in 10-20 ms before the masker. Whereas the postmasking effect last for 50-200 ms after the masker.

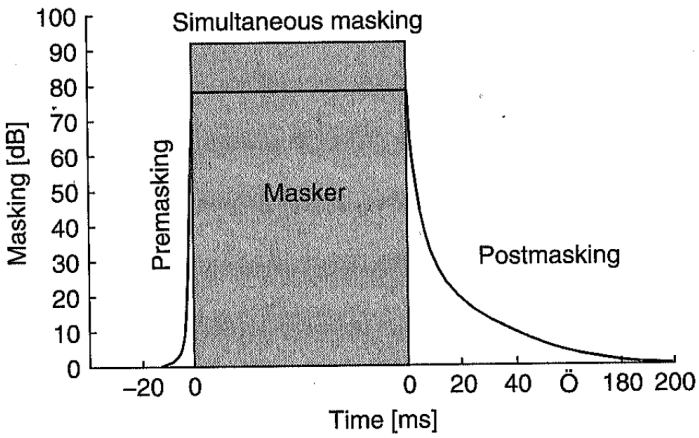

Figure 7: Temporal Masking[4]

Figure 8 shows the effect of masking in both frequency and time domain to summarize the difference between both types.

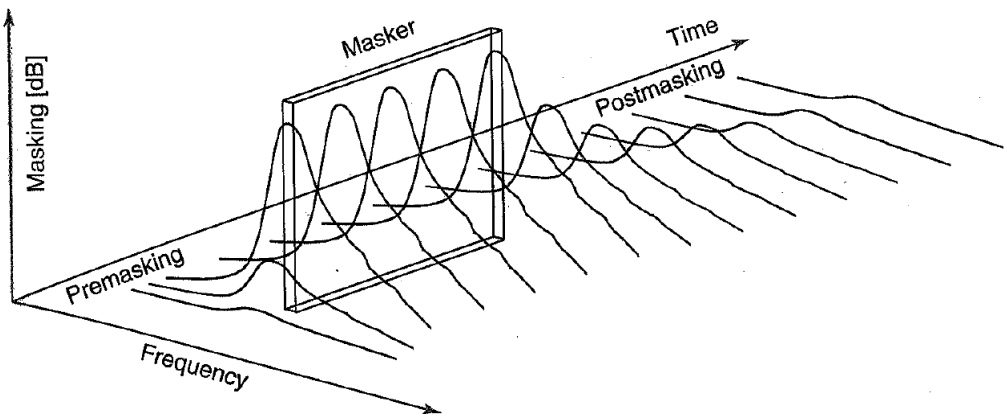

Figure 8: Masking in Time and Frequency Domain[4]

Bibliography

[1] S. Committe, Health Risk from Exposure to Noise from Personal Music Players, sep 2008.

[2] http://www.intropsych.com, June 2015.

[3] https://www.wikipedia.org/, June 2015.

[4] V. G. Oklobdzija, Ed., Digital Systems and Applications. CRC Press, 2008.

[5] R. Raissi, "The theory behind mp3," Tech. Rep., December 2002.

[6] K. Lagerström, "Design and implementation of an mpeg-1 layer iii audio decoder," Master's thesis, Chalmers University Of Technology, Sweden 2001.

[7] E. Z. Hugo Fastl, Psychoacoustics: Facts and Models. Springer, 2007.