UNIVERSITÀ DEGLI STUDI
DI MODENA E REGGIO EMILIA

Lecture notes for
Multimedia Data Processing

Sound Perception

Last updated on: 20/05/2020

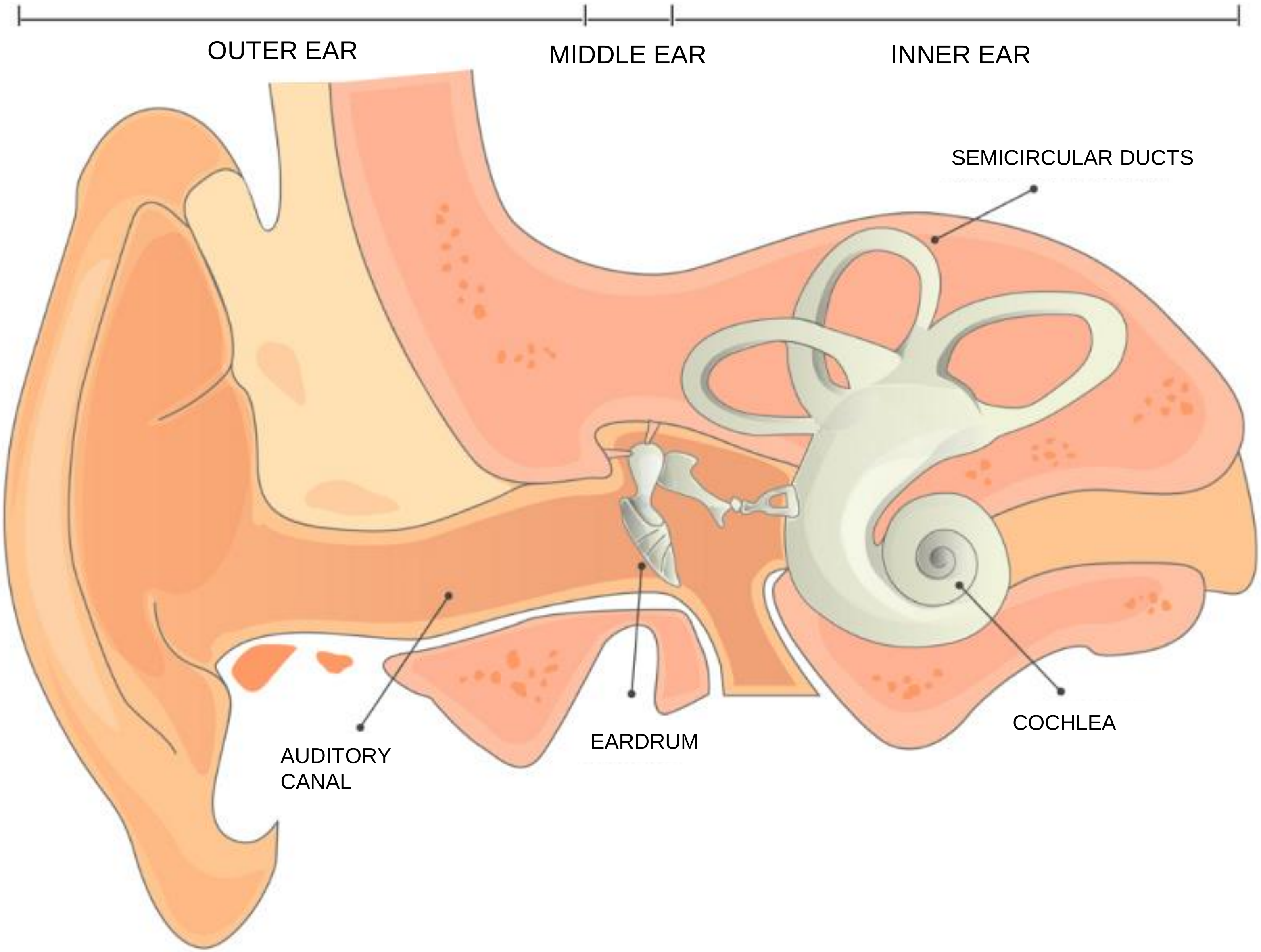
The Human Ear

- The human ear acts as a **transducer in transforming** acoustic energy, first into mechanical energy and then into electrical energy.

- Once the energy has been converted **from mechanical** to **electrical form by the ear**, the electrical impulses come to the brain through nerve endings.

- The auditory system is made up of three sections: the **outer ear**, the **middle ear** and the **inner ear**.

The Human Ear



Outer Ear

- The first organ that sound encounters when it reaches the ear is the **auricle**.

- The sound is reflected by the auricle and concentrated towards the **ear canal** whose length is on average equal to **3 cm.**

- **Resonant frequency of the ear canal** is on average 3KHz.
- This means that when a group of frequencies of value around 3KHz reach the ear, the ear canal resonates and therefore those frequencies undergo a natural amplification.

Middle Ear

- The ear canal ends on a membrane, the **eardrum** (also called **tympanic membrane** or **myringa**), which vibrates in accordance with the sound that has reached the ear.
- On the opposite side of the **eardrum** there are three small bones called: **hammer**, **anvil** and **stirrup**.
- These have the function of amplifying the vibration of the eardrum and retransmitting it to the **cochlea**.
- This **amplification** is **necessary** because while the eardrum is a very **light membrane** suspended in the air, the cochlea is filled with a **dense fluid** and therefore much more difficult to put into vibration.
- **An opening inside the middle ear** leads to the so-called **Eustachian tube** which consists of a canal that leads to the oral cavity (allows to balance the atmospheric pressure on the two sides of the eardrum).

Internal Ear

- The **stirrup** is in contact with the **cochlea** through a membrane which is called **oval window**.
- The **cochlea** is a snail-shaped bone containing fluid.
- The fluid receives the vibration from the stirrup through the oval window and transports it inside where the real organ responsible **for converting mechanical energy into electrical energy is present: the Corti organ**.
- Inside the organ of Corti we find the **basic membrane** which houses a population of eyelashes, about 4000, which vibrate according to the vibration of the fluid.
- Each **group of eyelashes** is connected to a **nerve termination** capable of converting the vibration received by the fluid into electrical impulses to be sent to the brain and to be processed and perceived as sounds.
- **The extension of the eyelashes excited** by the single frequency is called the **critical band** and is the basis of many psychoacoustic phenomena.
- The extension of the critical band increases with increasing frequency.

Critical Band

- The phenomenon of critical bands is at the origin of the **masking phenomenon**, used in many audio data compression algorithms.
- Information relating to **frequencies that fall in the same critical band** is **replaced** with a **single frequency** representative of all.
- The sound perceived in this way will not be degraded much, while the sound information to be stored will be decreased, thus creating **data compression**.
- The concept of critical band is also the basis of another psychoacoustic phenomenon: **beats** and occurs when the **two frequencies that originate the beat fall into the same critical band**.

Anechoic Chamber

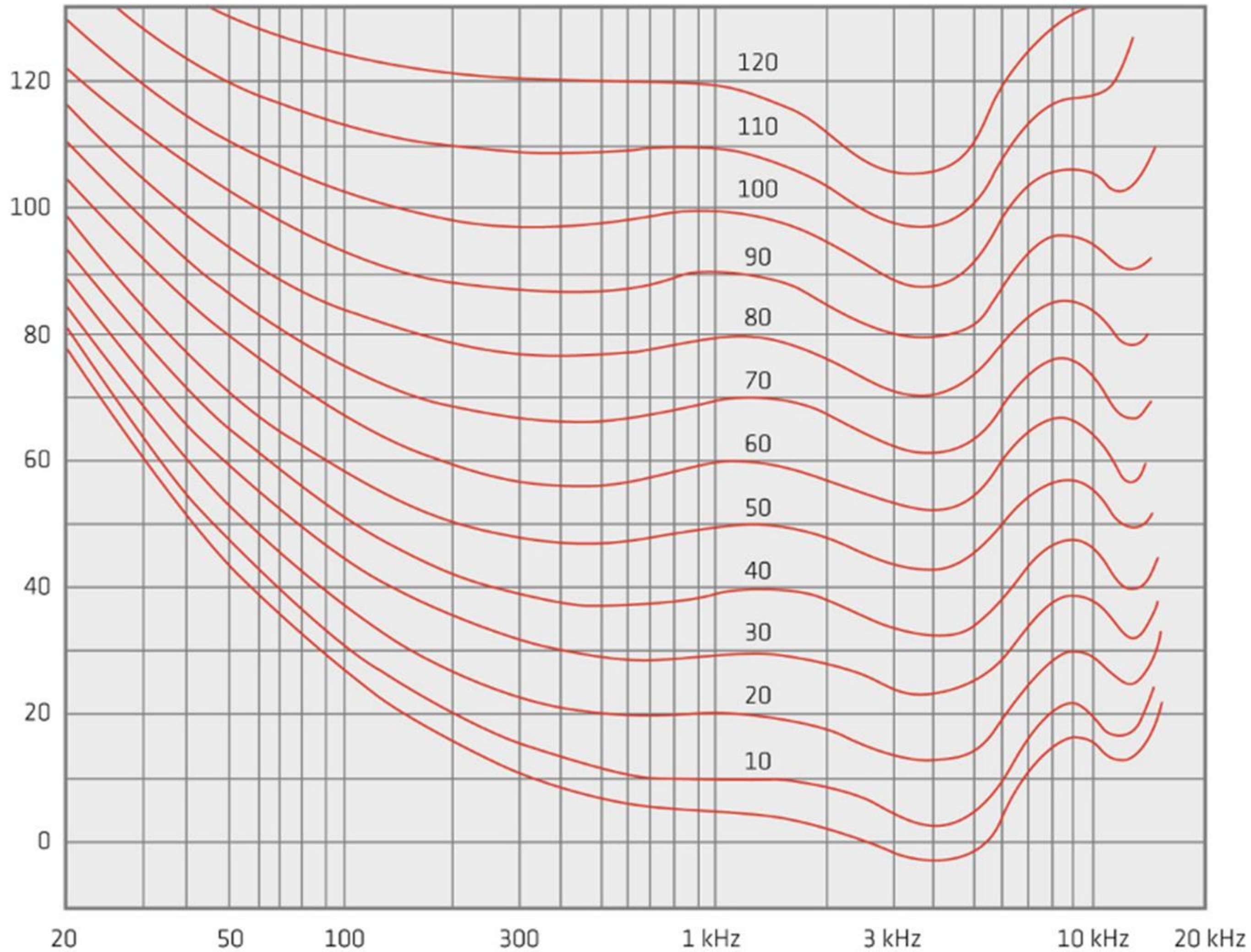
- The isophonic curves were obtained by processing the data on a statistical sample subjected to a series of sounds produced in an **anechoic chamber** (designed to minimize reflections on the walls).

Isophonic Curves

- Suppose we have a **sound source** capable of generating **sine waves with variable frequency and constant amplitude**.
- Setting the **amplitude** for example to **80 dB$_{spl}$** we would notice that a listener perceives the low frequencies as having a very low volume and **as the frequency is increased he would have the perception that the volume also increases** (while the sound pressure actually generated is always **80 dB$_{spl}$**).
- This behavior is explained by the fact that **the human ear has a different perception of sound intensity** as the frequency changes.
- The **isophonic curves** are called such as they indicate the dB$_{spl}$ value necessary to perceive a sound always **at the same volume along each curve**.
- The **reference frequency** for each curve is **1KHz** and at this frequency, the dB$_{spl}$ value is equal to the value that identifies a particular curve and which is called **phon**. For example, the isophonic curve **at 40 phon is the one that has an amplitude of 40 dB$_{spl}$ at 1 KHz**.

Isophonic Curves

- They are very important **charts** that allow you to have a reference on **how the human ear reacts to different frequencies.**



Audibility Threshold

- The **lowest isophonic curve** of all is called the **audibility threshold** and indicates the smallest pressure change that the ear is able to detect at different frequencies:

Frequency Zone	**Hz**	**dB_{spl}**
Reference	1000	5
Low Frequencies	50	42
High Frequencies	10000	15

- The table above shows some reference values relating to this curve (which represents the limit below which no sound is perceived on statistical samples)
- **Pain threshold (120 phons)** the ear begins to perceive physical pain and for prolonged exposures non-reversible damage can be generated.

Psychoacoustics

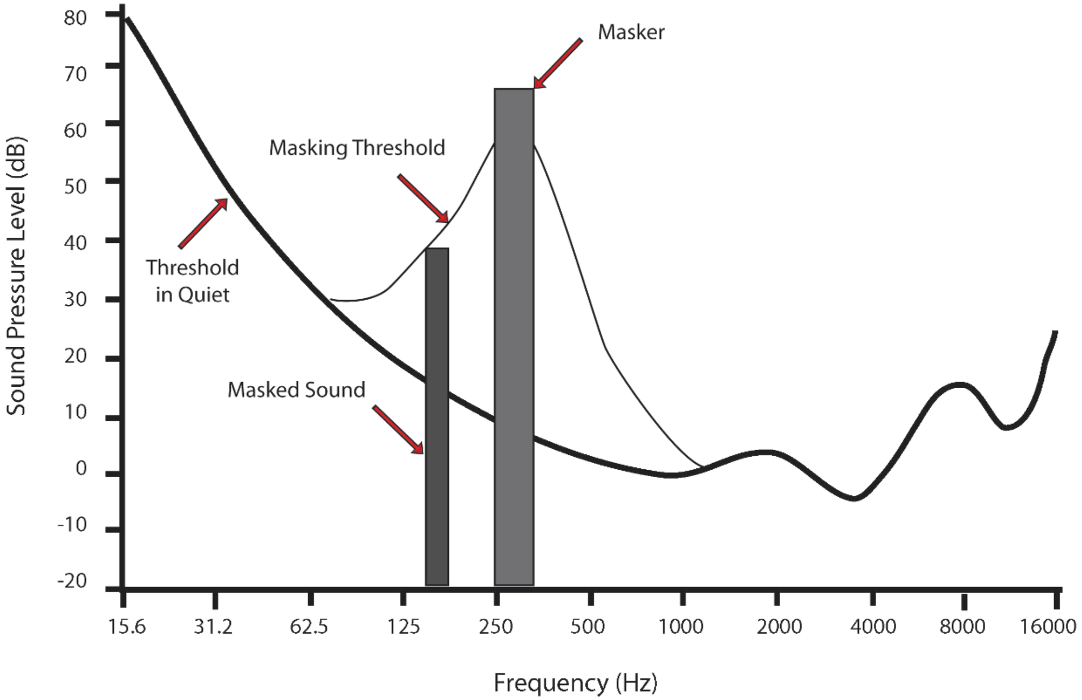
- **The hearing**, like the view that interprets the light, **is capable of perceiving only a part of the acoustic waves** that surround us and therefore returns a partial picture.

- The perceived waves are processed by the brain which thus interprets the sounds it has to process.

- **Psychoacoustics** studies the brain's sound processing mechanisms.

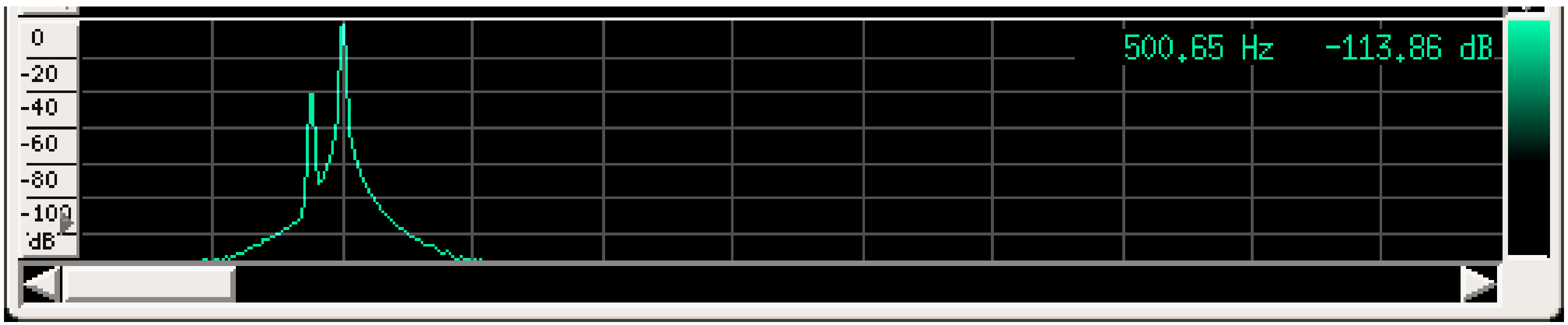
Simultaneous Masking

- A frequency component with **high amplitude can mask components at nearby frequencies** with lower amplitudes since nearby frequencies are decoded by cilia belonging to the same critical band.
- A sound of a certain frequency and intensity (masker) can temporarily change the frequency response of the ear bringing other sounds (masked) below the audibility threshold.



Simultaneous Masking

- The bandwidth of the masking threshold depends on the intensity and frequency of the masker and the intensity of the masking effect decreases moving away from the frequency of the masker sound.
- Example: a louder pure sound masks another weaker sound within the same critical band (between 400 and 510 Hz)



- This property is massively exploited to create algorithms for compressing audio data in digital format such as MP3 or AAC.
- These algorithms allow compressions in the order of 5: 1

Temporal Masking

- The phenomenon occurs when a weak sound **follows**, or, incredibly, **precedes a more intense sound**.
- **The weak sound is not audible**, if the time interval between the two is less than a certain threshold.
- For forward masking, i.e. when the loud **sound precedes the weak sound**, **the threshold is about 50 ms**. For **masking backwards about 10 ms.**
- In the following example a **glissando** from 200 to 3200 Hz is performed three times.
- In all cases, as can be seen by looking at the sonogram, **the sound is interrupted for 150 ms**, but the interruption **is noticeable only in the third repetition**. In the first repetition the short silence is masked by white noise, while in the second by white noise from which the band from **900 Hz to 2000 Hz** has been subtracted (which would contain the frequency of the glissando in the 150 ms in which it instead ceases).

Temporal Masking

- Temporal masking that gives the illusion of continuity of sound.