Course: Digital Audio Signal Processing

The Listener’s Model 2

Positional Models

Aurelio Uncini

Dept. of Information Engineering, Electronincs and Telecommunications

SAPIENZA
UNIVERSITÀ DI ROMA

intelligent signal processing
and multimedia lab

Premise

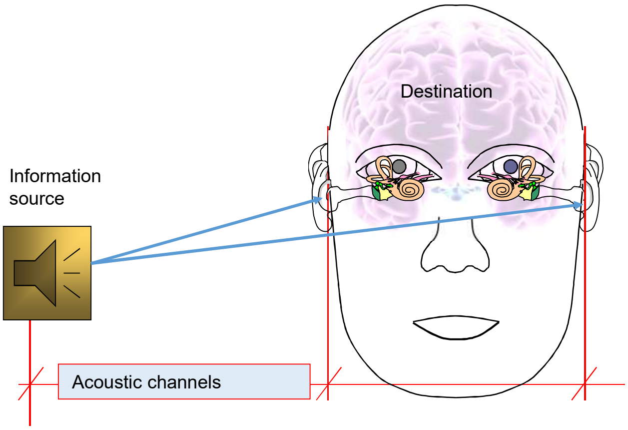
- The human ears perceive the sound waves, i.e. the vibrations that change the air pressure.
- The shell shape of the outer ear (auricle or pinna) performs the task of collecting sound waves from different directions and directing them according to various paths to the ear canal.
- The pinna, and other parts of the body such as the head, shoulders and torso, actually modify the sound waves, helping the brain to determine where the sound comes from: front, back, top or bottom.
- Human ability to locate a sound source in space is still an extremely interesting field of research.

Premise

The waves reflected from the various points of the pinna and towards the mouth of the ear canal, are each other not in phase due to the different paths.

From these phase displacements and the consequent spectral variations, even if minimal, the **nervous system is able to obtain directional information of the incoming sound**.

Although experimental evidence has shown that with only one ear there is some sense of the direction of arrival, **the precise localization of the sound source requires the presence of two ears**.

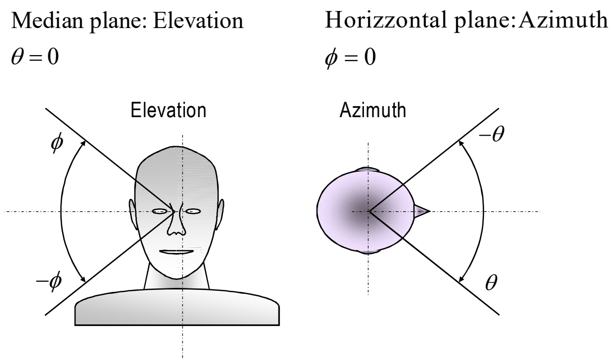


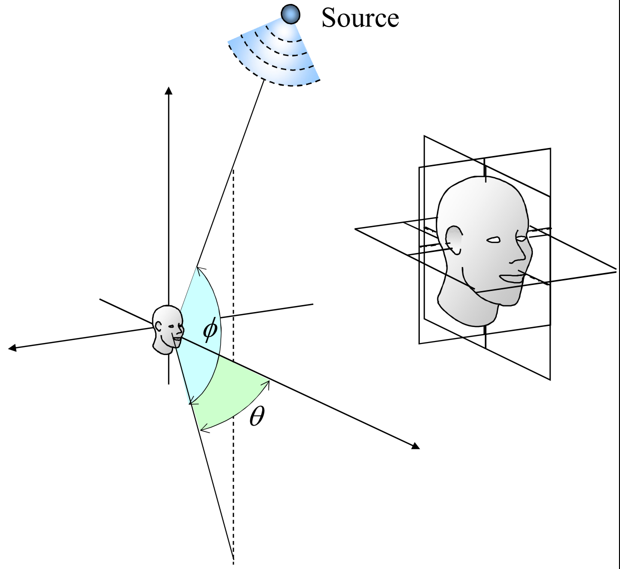
The paths between the source and the auditory channels are modeled with transfer functions indicated as head related transfer functions (HRTF).

Positional Models for 2D Audio

The Duplex Theory

Reference system for the relative position of the source with respect to the listener.





The Duplex Theory[1]

Approximate 2D model for the estimation of the spatial localization phenomenon by a pair of human ears. The theory is based mainly on the following two quantities:

- the **interaural time difference** (ITD): the temporal difference with which the sound waveform reaches the two ears;
- the **interaural level difference** (ILD): difference in intensity, measured in decibels, perceived by the two ears.

Both these quantities are important for the localization of a sound in the **azimuthal plane**, that is in the discrimination of the provenance between right and left.

[1] A. D. Blumlein, (1931), U.K. Patent 394,325. Reprinted in Stereophonic Techniques, Audio Eng. Soc., NY, 1986.

Geometry of Duplex model.

For low frequency <1500 Hz

$$\theta = 0,\ d = 0.1m$$

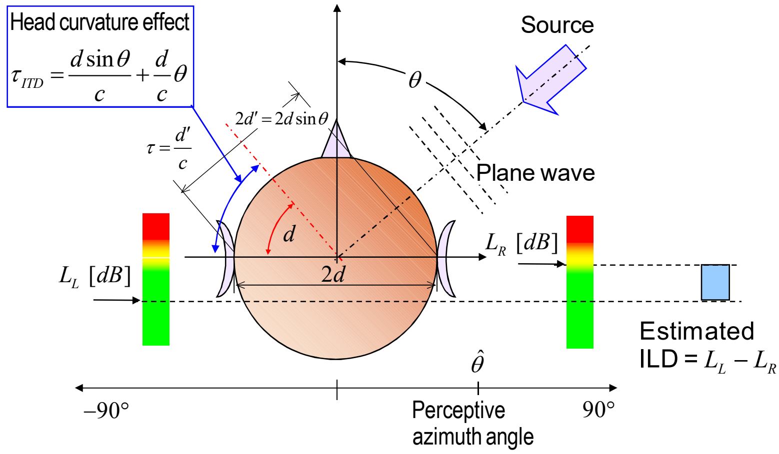
$$\tau_{ITD} = \frac{d\sin\theta}{c} + \frac{d}{c}\theta = \frac{d(\sin\theta + \theta)}{c}$$

$$= \frac{0.1(0 + 1.5707)}{344} = 0.46ms$$

$$\theta = \tfrac{\pi}{2},\ d = 0.1m$$

$$\tau_{ITD} = \frac{d\sin\theta}{c} + \frac{d}{c}\theta = \frac{d(\sin\theta + \theta)}{c}$$

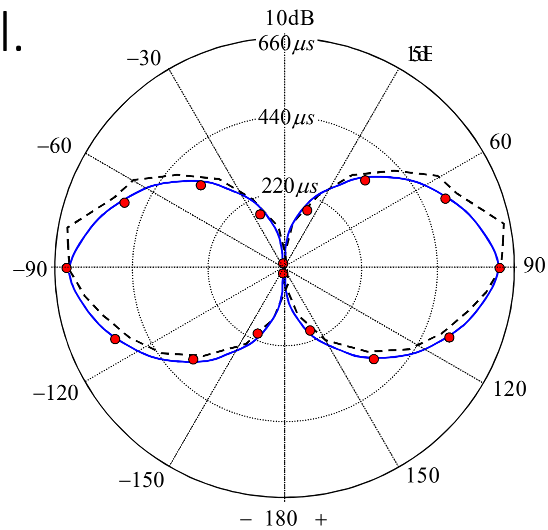
$$= \frac{0.1(0 + 1.5707)}{344} = 0.75ms$$



Interaural level difference (ILD), and **interaural time difference** (ITD) with the Woodworth/Schlosberg's formula. In the Duplex theory model the direction from the azimuthal angle is evaluated by means of these quantities.

Geometry of Duplex model.

For low frequency <1500 Hz



ITD on the horizontal plane (in μs).
The continuous segment is related to the Eqn.
.

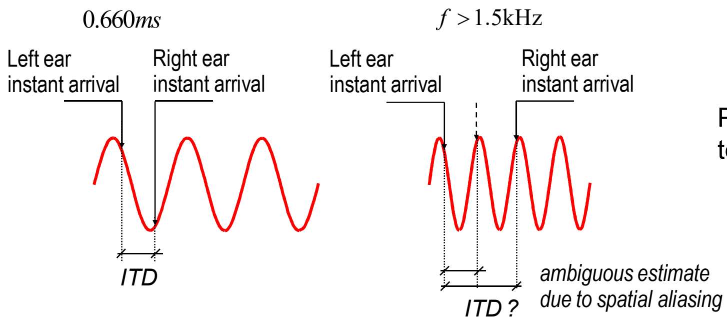
$$\tau_{ITD} = \frac{d}{c}\left(d\sin\theta + \theta\right), \quad -\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$$

The points represent an average measurement evaluated on 70 subjects. The dotted line is relative to an ellipsoidal model (Courtesy of [40]).

Spatial Aliasing ao Duplex Model

The radius of a typical adult's head is about 10 - 11 cm. Based mainly on the low-frequency content of a sound signal, the ITD can be used (without phase ambiguity from space aliasing) for wavelengths up to the Nyquist spacing of the distance between the ears, corresponding to about 1700 Hz.

In other words, around 1500 - 1700 Hz, the wavelength of the sine wave becomes comparable with the distance between the ears, and for shorter wavelengths, the delay time between the two ears is significantly longer than a length wavelength.



Phase ambiguity due to **spatial aliasing.**

Interaural Level Difference

At these frequencies the level difference perceived between the two ears becomes decisive. As shown in Fig. 2.23, the shadowed ear, compared to the sound source, will receive a lower sound energy intensity than the one placed in the frontal direction.

The smallest detectable ILD is about 0.5 dB, regardless of frequency. However, the far-field ILD doesn't exceed 5-6 dB whereas the near-field ILD. For example, at 500 Hz exceeds 15 dB [56]. In fact, the interaural level difference, is more sensitive in near-field (less than 1 meter source distance) than far-field, i.e the relationship model the ILD is not linear, furthermore this dependence is strongly linked to the frequency of the waveform. So, in the case of far field at low-frequency ILD, is negligible, and becomes large at close distances [57].

For certain frequencies, the azimuth of perception varies approximately linearly with the logarithm of the inteural intensity difference, and even if Eqn. (2.10) is frequency independent, in some other binaural models it is dependent on frequency.

Interaural Level Difference

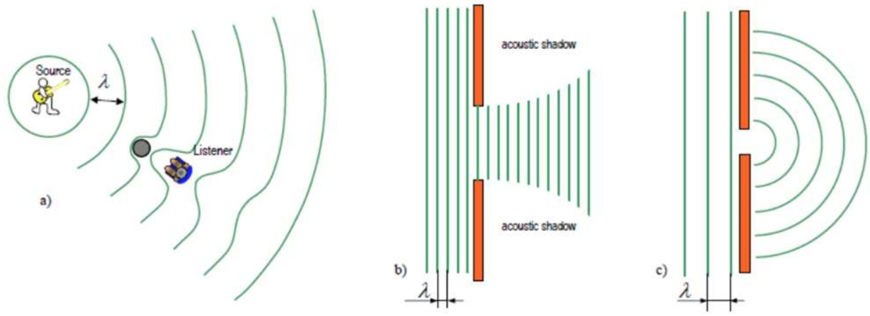
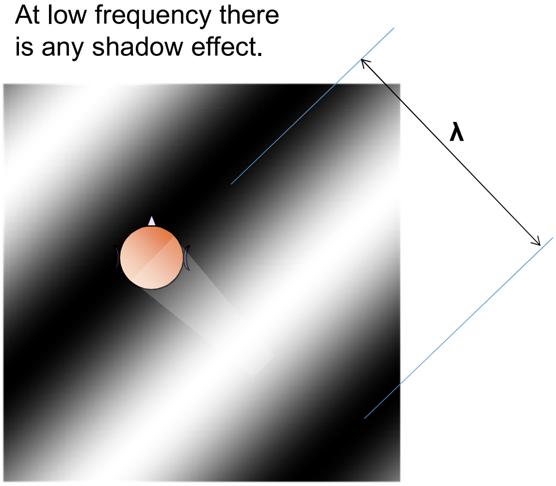


Fig. 1.55 Diffraction phenomenon. a) The wave encounters an obstacle in its path similar in size to the wavelength. b) Diffraction due to a wall with a hole larger in size than the wavelength of the acoustic source. c) Due to a wall with a hole of a size comparable to the wavelength of the acoustic source.

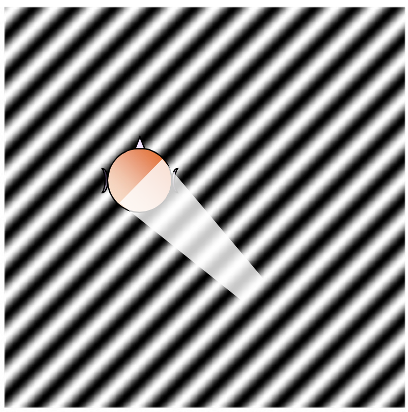
Interaural Level Difference

At low frequency there is any shadow effect.



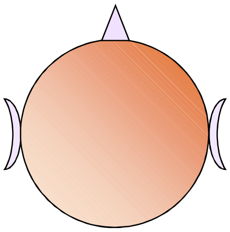
λ > **d**: The vibration is propagated, almost without attenuation.

At high frequency there is shadow effect.



λ < **d**: the acoustic wave is strongly attenuated by the shadow effect of the head.

The shadow effect of the head is the cause of the **Interaural Level Difference (ILD)**



Wavelength in air sound propagation

f **= 20 Hz** → **λ = 17 m**

f **= 20 kHz** → **λ = 1.7 cm**

2.4.1.3 Precedence effect

The spatial localization through ITD is related to the binaural psychoacoustic phenomenon known as the *precedence effect* or *first wavefront law.*

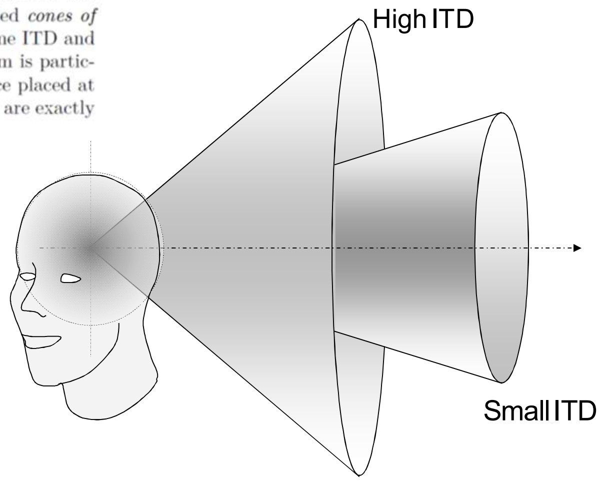
Considering two sounds, even of different levels, coming from different directions and separated by a sufficiently short time, the listener perceives a single auditory event; the direction of arrival perceived by the listener is that of the sound arriving first.

The delayed sound also affects the perceived position. However, its effect is attenuated by the first perceived acoustic event.

2.4.2 The Cone of Confusion

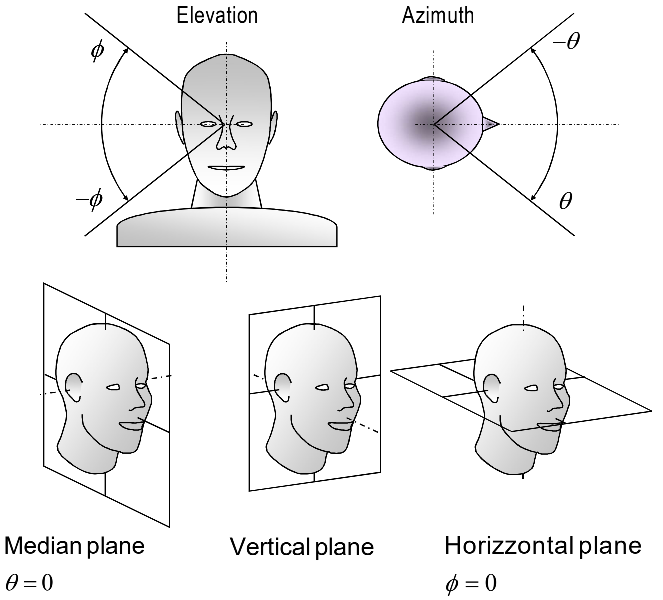
Duplex theory can correctly discriminate only the position of the source for the various azimuthal angles. However, it does not provide a satisfactory description for the complete location in the free space where the sound can also be characterized by an elevation angle and a distance. If one attempted to describe a source at any point in three-dimensional space, the Duplex theory would provide infinite points of localization along curves of equal distance from the listener's ears which associated the same ILD, ITD vlues. As shown in Fig. 2.26 these curves are cones, called *cones of confusion*. All the points in the cone of confusion have, therefore, the same ITD and ILD and are indistinguishable in accordance with the above. The problem is particularly accentuated in the median plane (see Fig. 2.27) where, for a source placed at any point of it, considering an ideal ear model, both the ITD and the ILD are exactly zero, minimizing the information.

Fig. 2.26 Cone of confusion calculated considering the head modeled as a spherical shape and in absence of pinna. The acoustic sources in the confusion cone have the identical ITD and ILD.

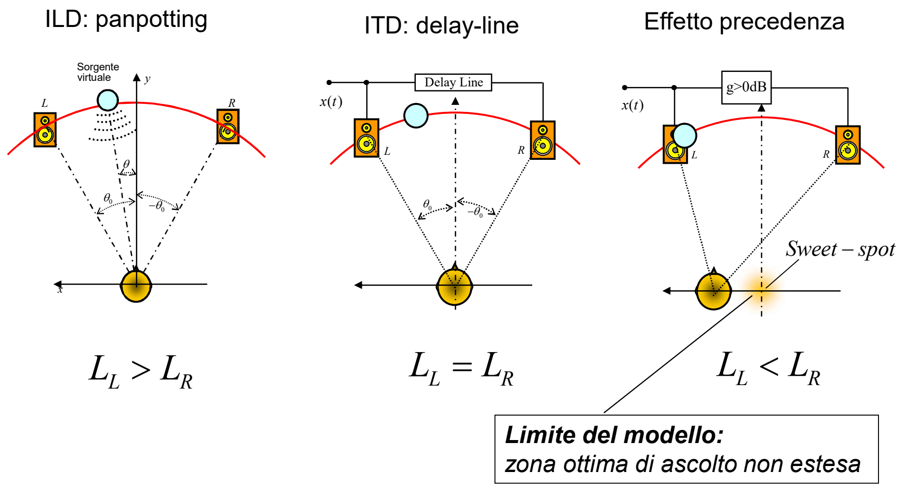


Cone of Confusion

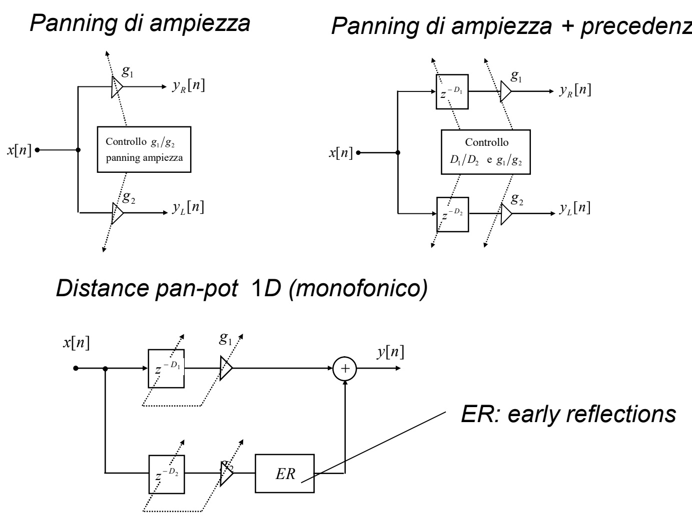
Fig. 2.27 Definition of elevation, azimuth and reference planes. It is assumed that the coordinates relative to the right ear are $(\theta, \phi) = (+90^{o}, 0^{o})$, while for the left ear $(\theta, \phi) = (-90^{o}, 0^{o})$.



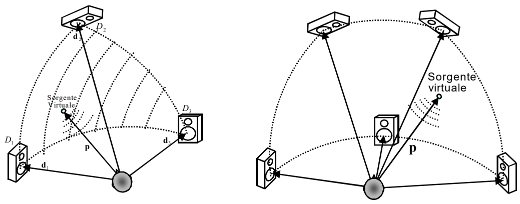
The Duplex Theory Limit



The Duplex Theory: Panning



Vector Base Amplitude Panning [1] (VBAP) NON FUNZIONA



Vantaggi:
Semplice estensione modello ILD al caso 3D
Sfera di diffusori in griglie poligonali di tre o cinque diffusori

Svantaggi:
Modello approssimato: solo per effetti. In pratica NON FUNZIONA
Non si tiene conte dell'ITD e del modello della sala di ascolto *Room Transfer Function* (RTF)

[1] V. Pulkki, "Virtual source positioning using vector base amplitude panning", J. Audio Eng. Soc., 45(6):456–466, June 1997.

Positional Models for 3D Audio

Premise

- In order to have an estimate of the position in 3D space, in addition to ITD and ILD, the entire spectral structure of the sound, due to the interference of the outer ear and the body, must also be considered.
- The physical effects related to the interaction of the acoustic waves with the pinna, the head and the torso, are mainly manifested as reflection, diffraction and linear distortion.
- By virtue of such effects, the acoustic waves which enter the ears are subject to micro spectral variations, depending on the direction of arrival of the wave [29].
- Experimental evidence (see in this regard [34]-[41]) suggests that amplitude and phase changes (of complex spectra sounds), allow the hearing organ to estimate, more or less correctly, the position of the acoustic source in all spatial dimensions in both monaural and binaural modes.

The Head Related Transfer Functions

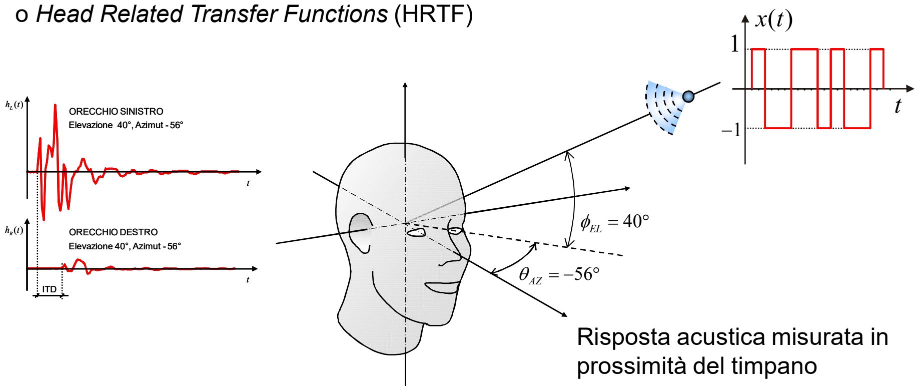
Definition 2.2. *Head related transfer functions* - For an acoustic source located in the space in the point relative to the listener of coordinates: (r, θ, ϕ) (distance r, azimuth θ, elevation ϕ), we define HRTF, indicated with $H^R_{\theta,\phi}(\omega)$ and $H^R_{\theta,\phi}(\omega)$, the ratio (complex) between the sound pressure level (SPL) in a point of the auditory canal near the eardrum and the SPL at the location of the source in free field (i.e. measured in an anechoic chamber).

Typically HRTFs are measured on a dummy-head, or on a real subject for both ears, by setting a reference angle for the position of the head. In practice, HRTFs are measured at different azimuths (Right-Left direction) and at different degrees of elevation (Low-High directions). The HRTF therefore represents a *Linear Time Invariant* (LTI) system. The impulsive response associated with HRTF is defined as *Head Related Impulse Response* (HRIR). The sampled versions of the HRIRs are commonly implemented as FIR filters, which, as we shall see, are minimum phase filter.

Head Related Impulse Response (HRIR)

Sorgente di calibrazione

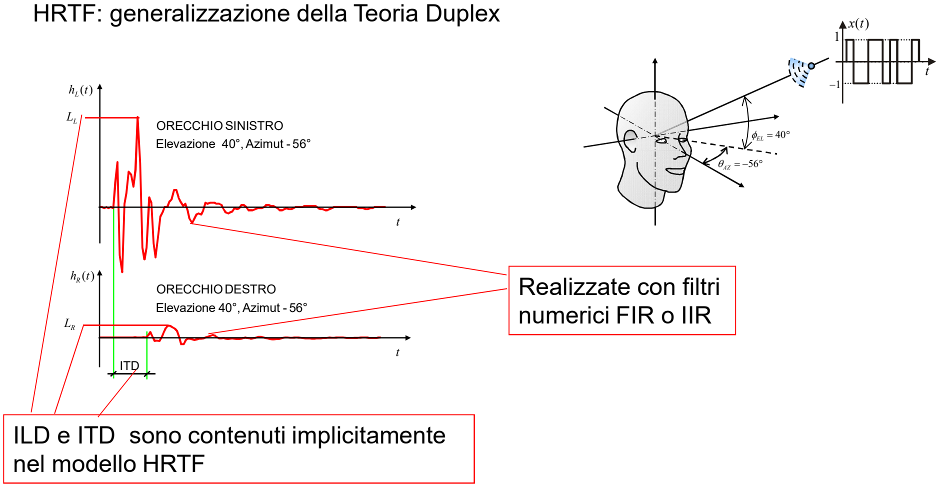
o *Head Related Transfer Functions* (HRTF)



HRTF
modello dell'ascoltatore, basato su **filtri dei percorso acustici**, che tiene conto anche dell'**elevazione**.

Generalization of Duplex Theory by HRTF

HRTF: generalizzazione della Teoria Duplex



HRIR

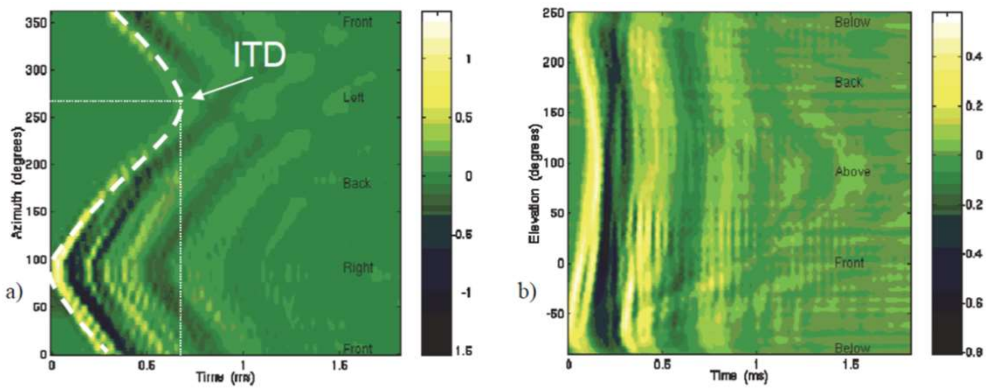


Fig. 2.29 Image representing HRIRs for the right ear. a) Evaluated on the azimuthal plane, measured for a position $\theta = (0°, 360°)$. b) Evaluated on the median plane ($\theta = 0°$), measured for a position $\phi = (-100°, 250°)$.

HRTF

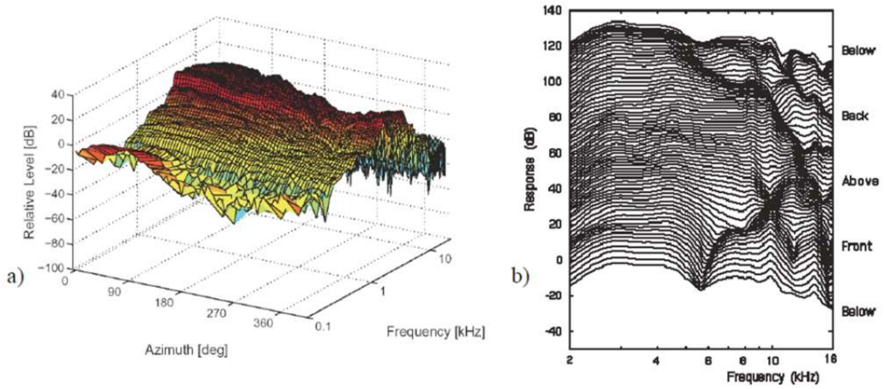
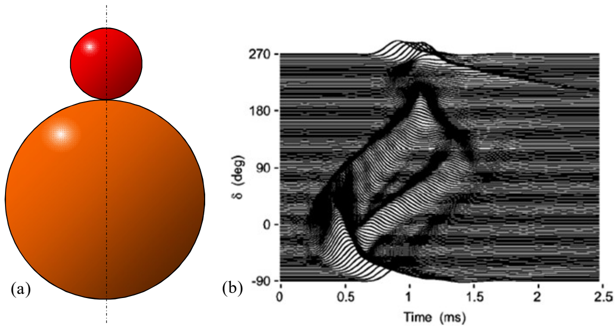


Fig. 2.30 Amplitude responses of HRTFs. a) Evaluated in the horizontal plane for various azimuth angle values [45]. b) Evaluated on the median plane for various elevation angle values [46].



The Duplex Theory

Reference system for the relative position of the source with respect to the listener