Master MVA

Analyse des signaux Audiofréquences

Audio Signal Analysis, Indexing and Transformation

Lecture on 3D sound rendering

Gaël RICHARD

February 2023

Sound rendering

- Introduction
- Some elements on spatial perception

- Multi-channel systems
  - **Stereophony**
  - **'Surround' systems**
  - **Ambisonics.**
  - **Wave-field synthesis.**

- Binaural techniques

- Applications

Introduction

- **Time-frequency : Sound**
  Space/time-frequency : Sound field

- Rendering : **definitions.**
  - Science
  - Sound engineering
  - Art

- **Rendering in practice : Recording**
  : Synthesis

- **Approaches : perceptual**
  : physics based

Brief historical overview of 3D sound rendering in music

(from A. Dutto: http://www.nikkojazz.fr/2009/05/la-mise-en-espace-de-la-musique.html)

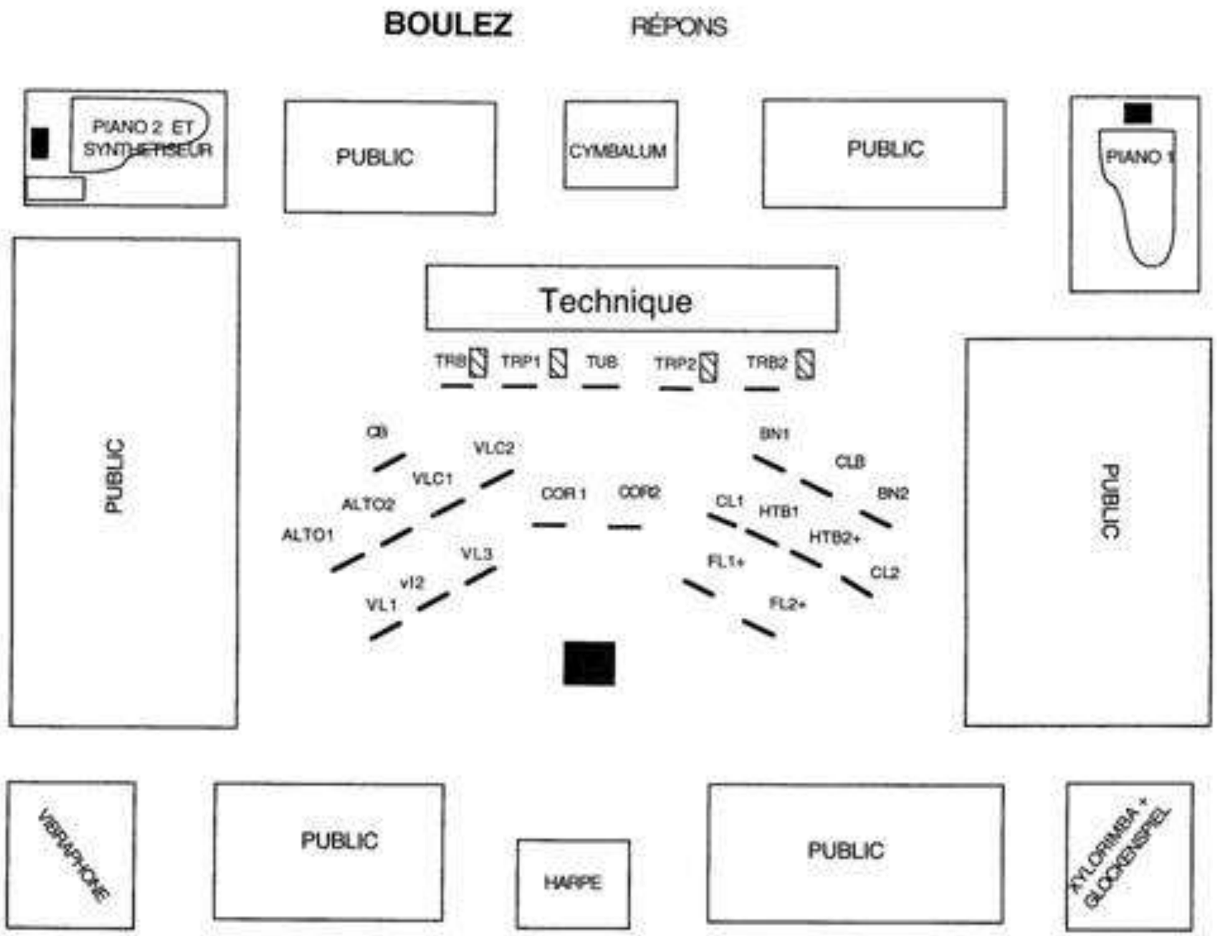
- **Premices:**
  - **4th century: Antiphonia, e.g. alternating between 2 choirs**
  - **1573: motet with 40 voices from Thomas Tallis**
  - **Double choirs then used by Vivaldi, Bach,.**

- **19th century:**
  - **Berlioz's Requiem – Tuba Mirum (1837):**
    - Monumental orchestra + large choir + 4 groups of brass instruments at the four corner of the hall
    - Question ?: where should be positioned the listener ?

- **The 20th century: space exploitation**
  - **Maderna, Carter, Stockhausen, Xenakis…Boulez**

Brief historical overview …

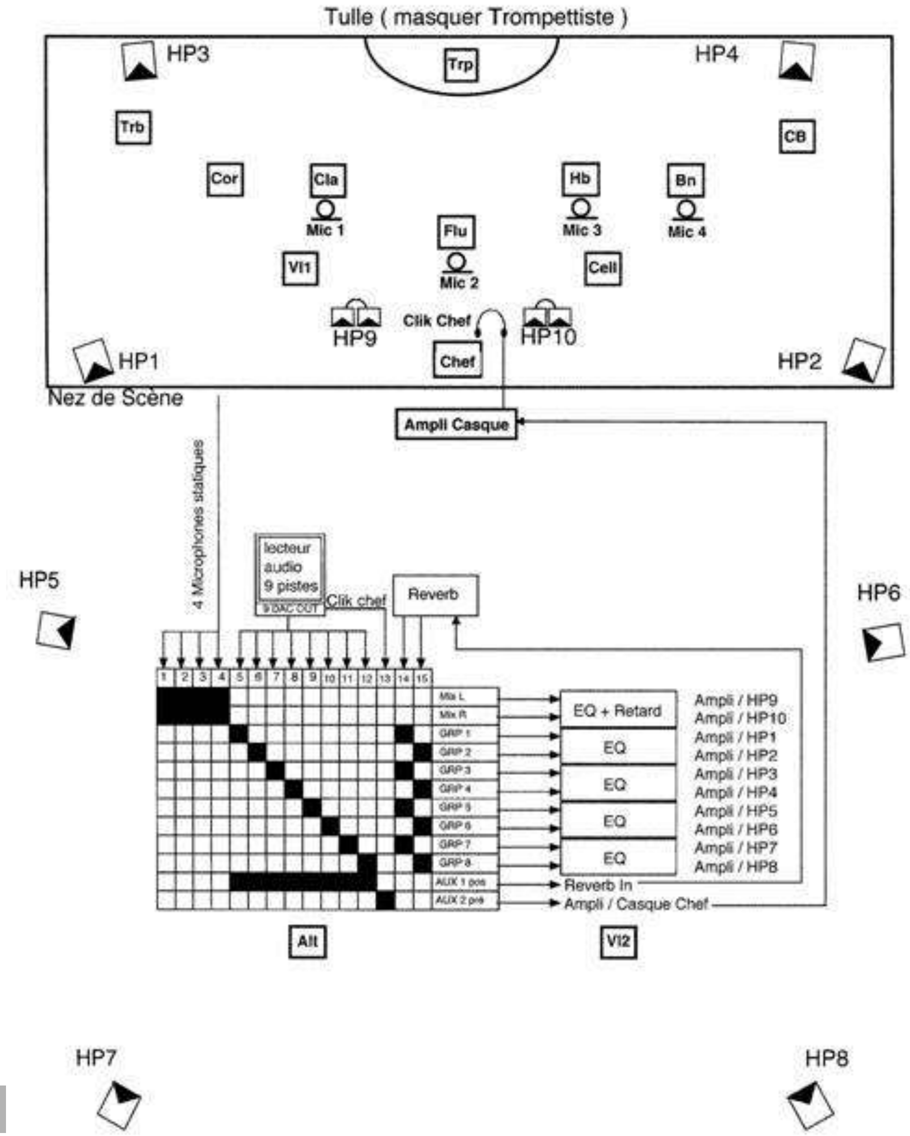
- « Repons » from P. Boulez…



Brief historical overview …

http://www.nikkojazz.fr/2009/05/la-mise-en-espace-de-la-musique.html

- Another example : *Le plein du vide,* Xu Yi (1997)
  - 8 computer tracks played on 8 loudspeakers placed on stage and in the public,
  - Public is at the center



Rendering techniques

- **Physics based:** Sound field as a physical phenomenon
- **Perceptual :**Sound field as a perceived phenomenon

  – Important parameters: IID, ITD, IPD, HRTFs..

- **Multichannel :** group listening
  – Notion of sweet spot

- **Binaural :** individual listening
  – Optimal for headphone listening

Sound rendering

- Introduction
- **Some elements on spatial perception**

- Multi-channel systems
  - **Stereophony**
  - **‘Surround’ systems**
  - **Ambisonics.**
  - **Wave-field synthesis.**

- Binaural techniques

- Applications

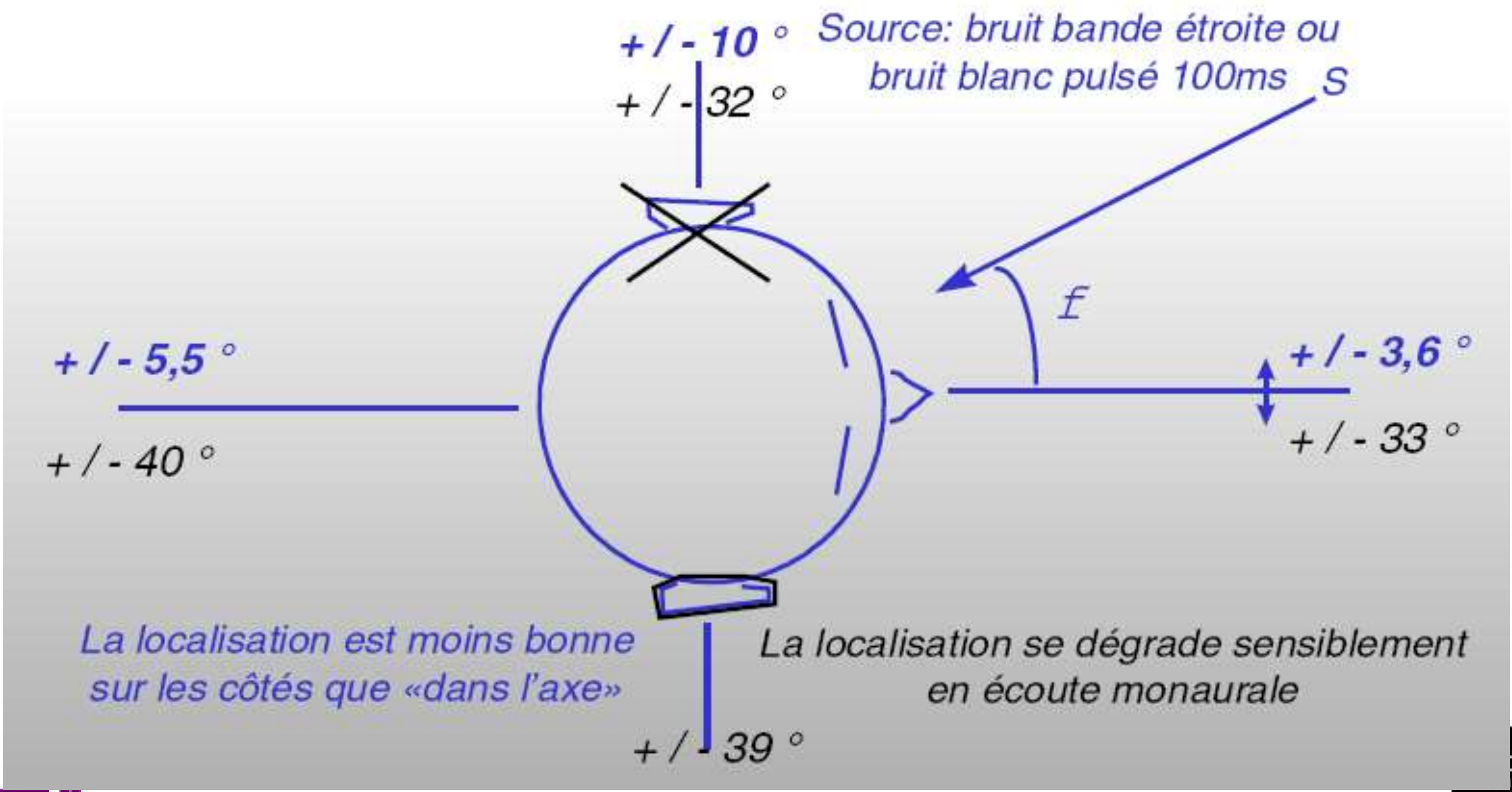
Some elements on spatial perception

- **2 main aspects of perception of space:**

- ***Sound sources localisation***
  - Remarkable human capacity for localising sounds (especially short impulsive sounds).
  - We essentially use binaural indices but not only.
  - Note that *localisation is not lateralisation* (with headphones, in this case sounds are “in” the head)

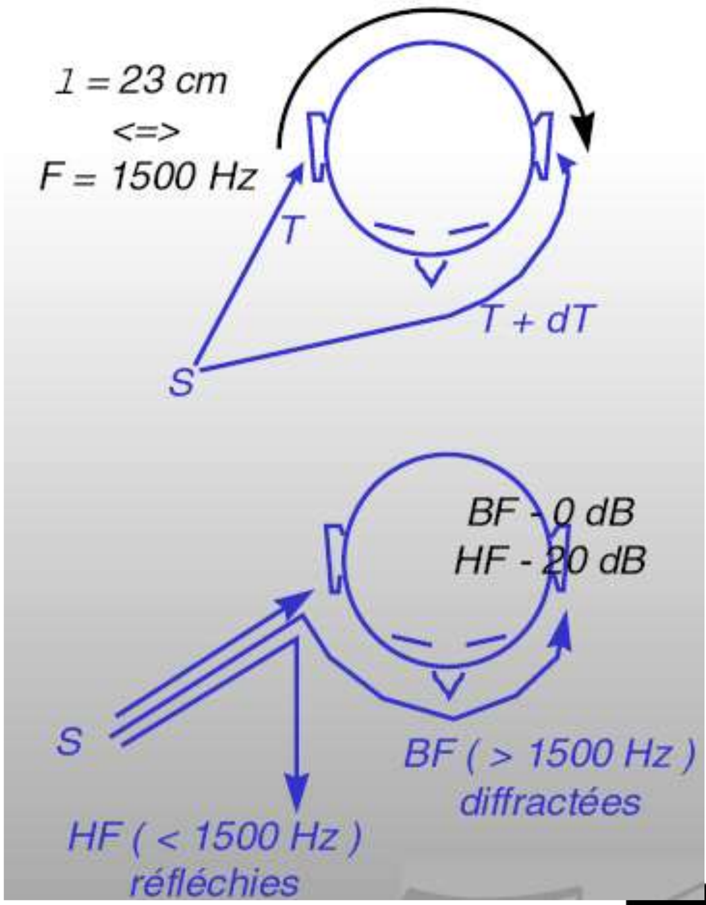
- ***Source subjective width***

Sound localisation in azimut



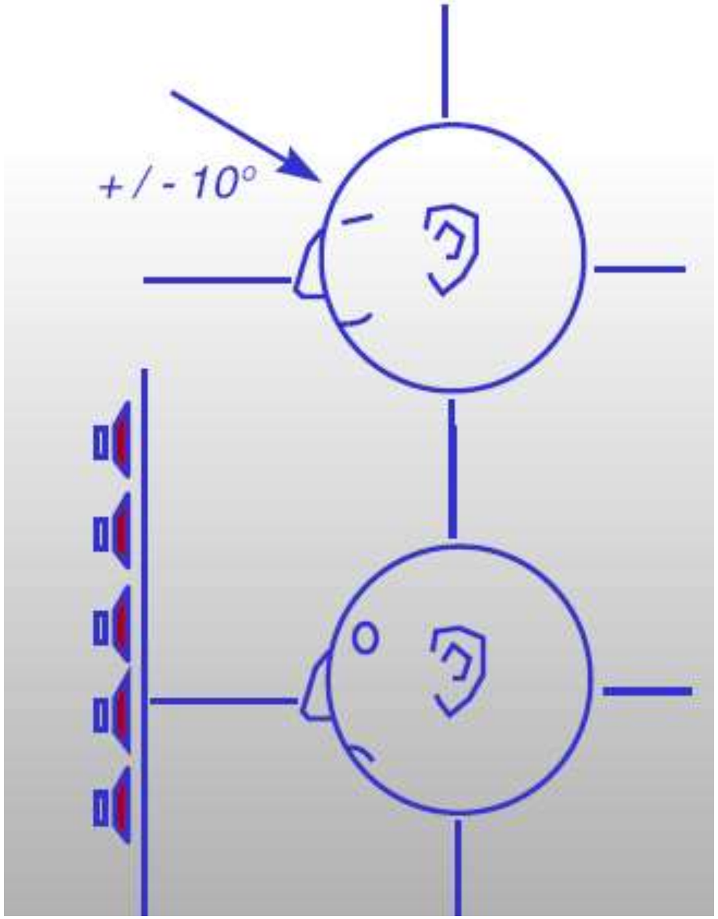
Sound localisation in azimut

- **2 types of indices (Steven and Newman, 1936):**
  - Interaural Time Difference (ITD)
    ... Especially in low frequency
  - Interaural Intensity Difference (IID)
    ... Especially in high frequencies.
- **max. ambiguity ambiguïté around 1500Hz**



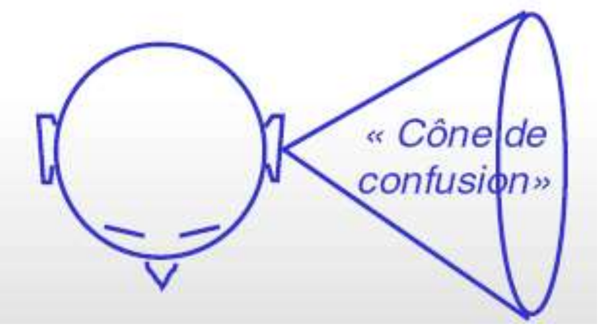
Sound localisation in elevation

- **Speech sounds localisation:**
  - + or - 10° / sound in front,
  - ~ less good / sound behind
- **Pur sounds : « localisation » is only depending on the frequency !**
  - Low frequency => source « low »,
  - High frequency => source « high »
- **One needs at least 2/3 of frequency range for a sound to be localised**



Monaural capacities: confusion cône

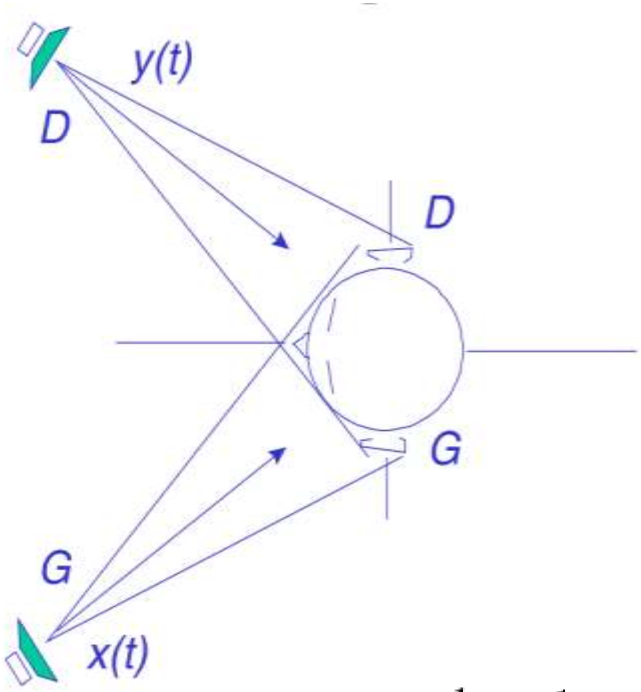
- All sources on the cone have the same ITD
- Ambiguities are solved by head movements
- If head movements are authorized, monaural localization is almost as good as binaural localization
- Other indices than ITD and IID are used: in particular change of the spectral patterns which is function of the sound direction of arrival



Impact of head and auricle (or pinna)

- ***ITDs+IIDs+head movements are not enough to explain our capacity in localization (incl. elevation)***
- **Head+ pinna : complex filtering with a clear effect between 500 and 16 000 Hz**
  - **pinna: essential for high frequencies (> 6 kHz)**
  - **Lowering of HF capacities if pinna relief are filled**
- **Localization exploits knowledge of the source and listening room (with very fast acquisition)**

Subjective width of sound sources



- **Normalised intercorrelation:**

$$F_{xy}(\tau) = \frac{\int_{-T}^{T} x(t)y(t+\tau)dt}{\int_{-T}^{T} x^2(t)dt \int_{-T}^{T} y^2(t)dt}$$

- **Let** $k = \max |F_{xy}(\tau)|$

$k = 1$ signals are coherent (perfectly correlated) and a single event is heard

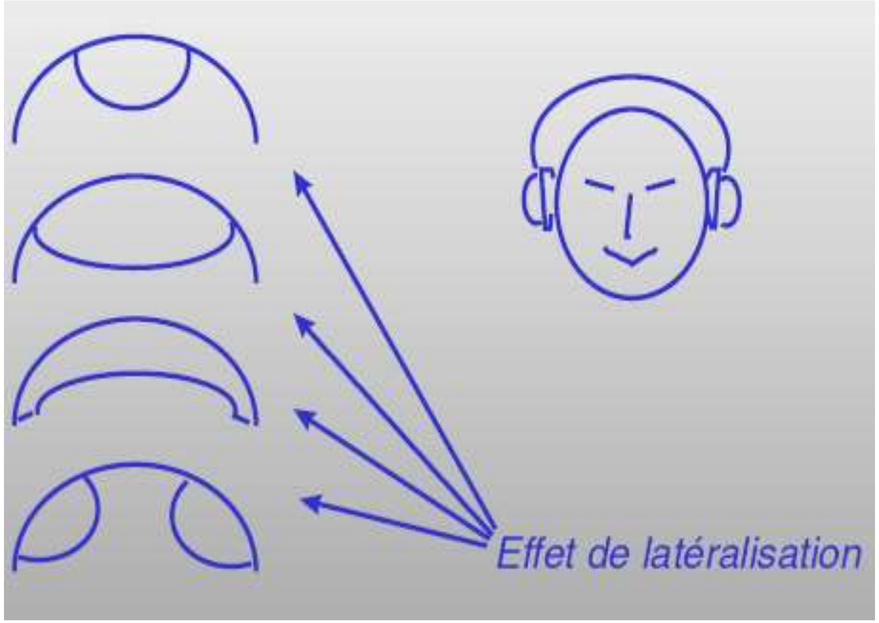
$k = 0$ Signals are decorrelated (2 events are heard in each loudspeaker

$k = 0.4$ Well balanced listening

Application to lateralisation

- **Using filters (low pass, high pass) one may impose a desired k value**

  - **k=1**
  - **k=0,85**
  - **k=0,4**
  - **k=0**



Conclusion about perception of space

- **Numerous indices, and in particular :**
  - ITD
  - IID
  - Pinna filtering, depending on sound direction of arrival

- **Redundancy is very useful in adverse situations**

- **Our localisation capacity is**
  - Very good in azimuth, much worse in elevation

Sound rendering

- Introduction
- Some elements on spatial perception

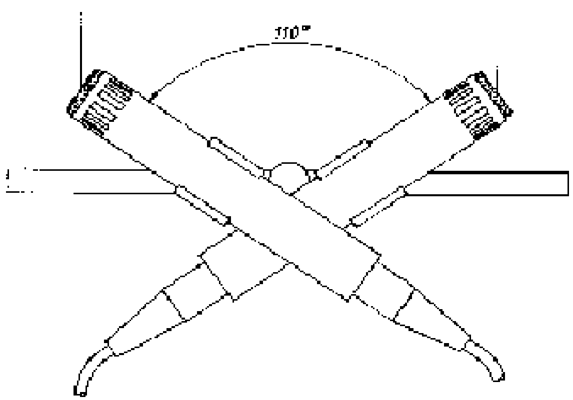
- **Multi-channel systems**
  - **Stereophony**
  - **'Surround' systems**
  - **Ambisonics.**
  - **Wave-field synthesis.**

- Binaural techniques

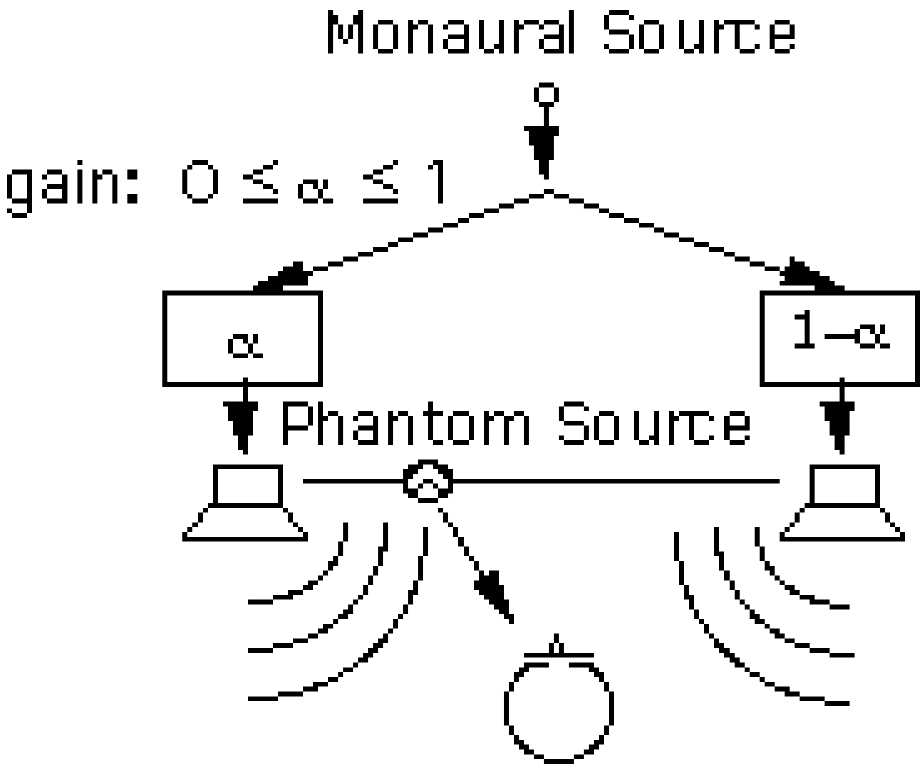
- Applications

Stereo recordings

- **Use the intensity difference (IID) and time differences (ITD) or Phase differences (IPD).**
- **For instance, recording using 2 microphones of different directivity (XY) or placed at different positions (AB)**



Stereo synthesis



Stereo rendering

- **On Loudspeakers: two loudspeakers with an aperture of 60 degrees**
  - Phantom images on a straight line between loudspeakers
  - Sweet spot
- **Headphones**
  - Localization inside the head

Surround: multi-channel systems

- 2 channels systems:
  - **Limitation of the localization in a specific zone (in front in stereo)**
  - **Constraint on the listener position**

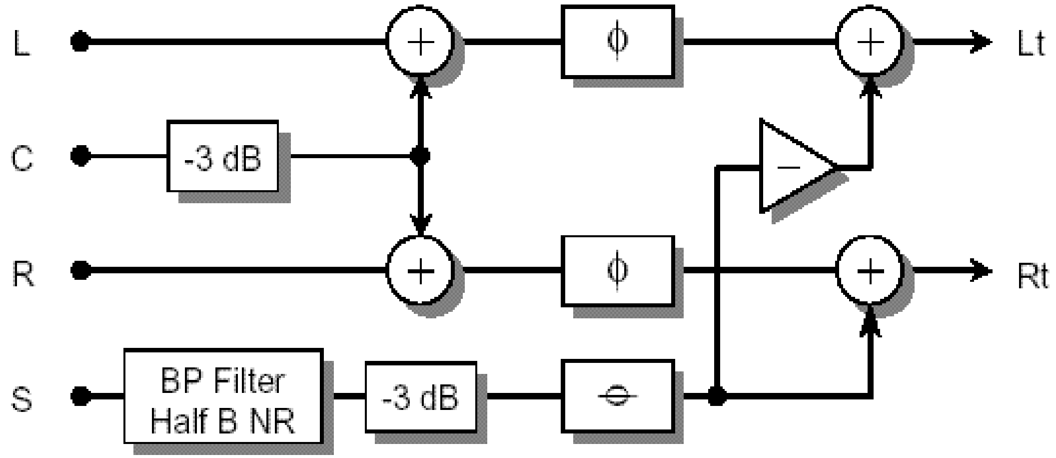
⇨**Solution: increase the number of channels**

- « surround » approaches
  - **1940: Fantasia (« Fantasound »)**
    - 3 channels (left wall, center, right wall)
  - **In the fifties: 4 channels format 4**
    - Compatibily problem with mono
    - No commercial success

Brief history

- 1977 **: Dolby Stereo for movie (*Starwars*)**
- 1982 **: Dolby Surround for home**
- **4 channels system (G,D,C,S) but on on 2 channels.**
- 1992**: Dolby AC-3 coding, digital system with 5.1 channels (which became *Dolby Digital*)**
- **2012 : Dolby Atmos (up to 128 different audio elements rendered to up to 64 individual speakers ….)**
- **Dolby but also DTS, Auro, THX,..: Auro 3D, DTS:X**

Dolby surround: coding



$$Lt = L + \frac{C}{\sqrt{2}} + i\frac{S}{\sqrt{2}}$$

$$Rt = R + \frac{C}{\sqrt{2}} - i\frac{S}{\sqrt{2}}$$

$$L\_décodé = Lt$$

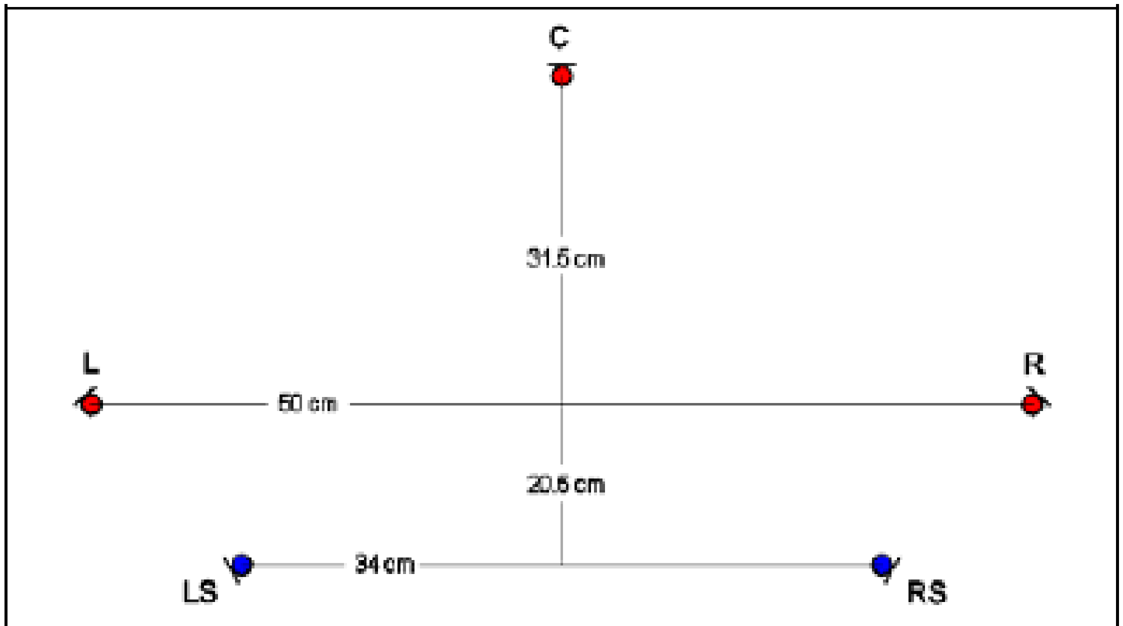
(Stereo compatibility)

$$R\_décodé = Rt$$

$$C\_décodé = \frac{1}{\sqrt{2}}(Lt + Rt) = C + \frac{1}{\sqrt{2}}(L + R)$$

$$S\_décodé = \frac{1}{\sqrt{2}}(Lt - Rt) = i * S + \frac{1}{\sqrt{2}}(L - R)$$

Dolby 5.1: recording



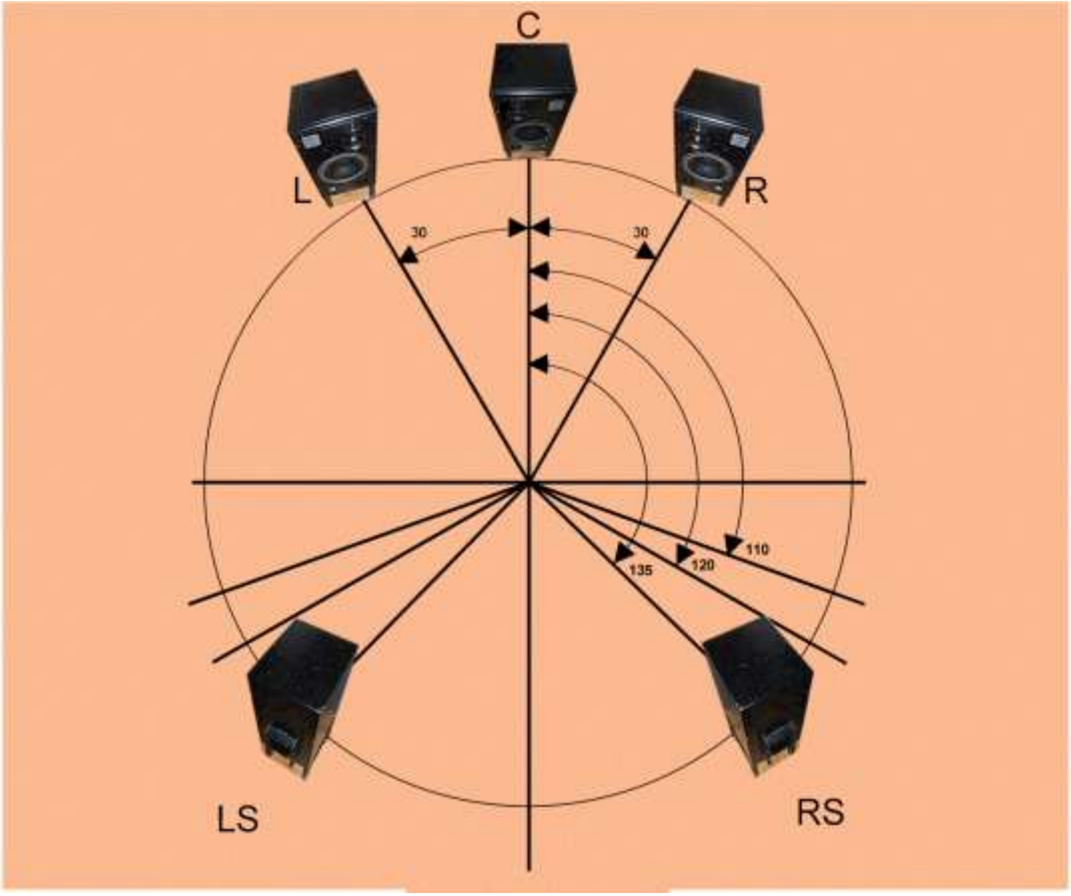
Williams MMA microphone system layout

C : Cardioid, 0°

L, R : Cardioid, ± 40°

LS, RS : Cardioid, ± 120°

Reproduction (5.1)



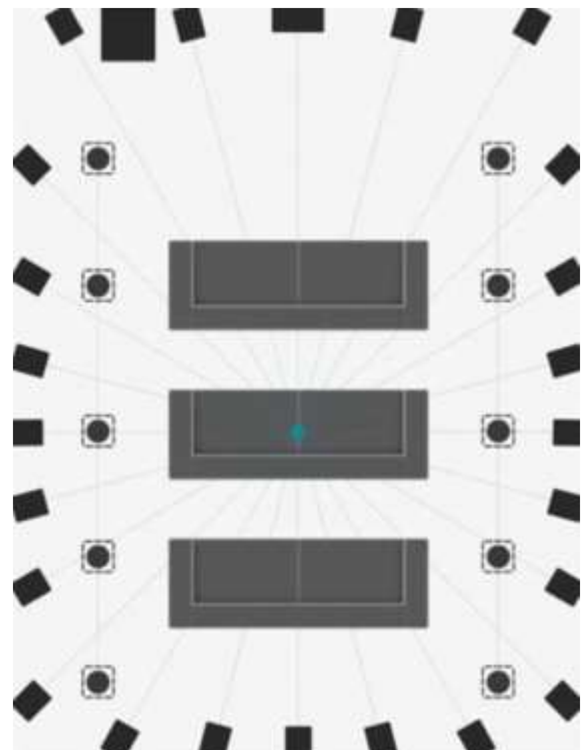
Dolby Atmos
(> 1000 movies in this format (first is « Rebelle in 2012) around 20 theater equipped in France in 2019

- **example of configurations for reproduction**

4 specific speakers + bass

• 4 x R50 et 1 caisson de basses

24 speakers
+ 8 overhead speakers



Sound rendering

- Introduction
- Some elements on spatial perception

- **Multi-channel systems**
  - **Stereophony**
  - **'Surround' systems**
  - **Ambisonics.**
  - **Wave-field synthesis.**

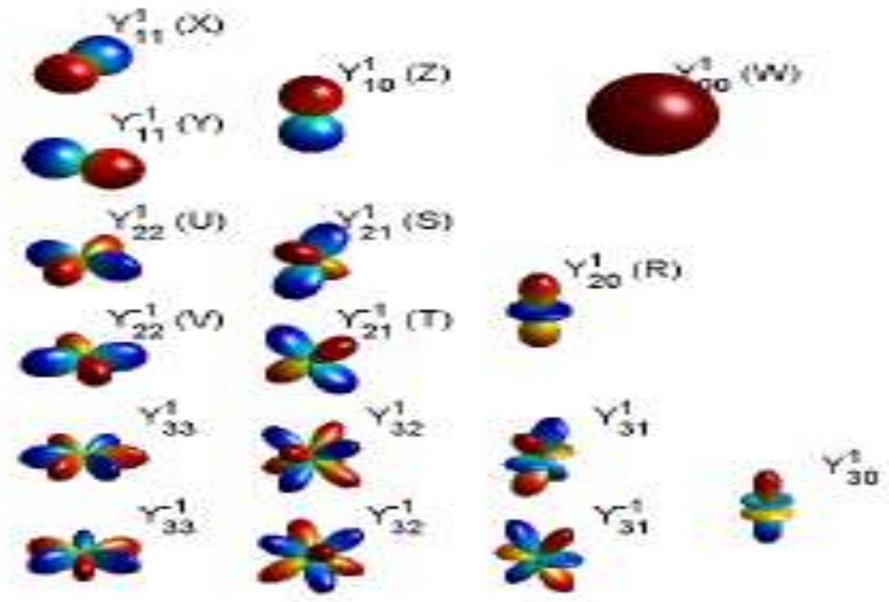
- Binaural techniques

- Applications

Ambisonics

- **System with a central reference**
- **Based on the decomposition of the sound field in spherical harmonics**

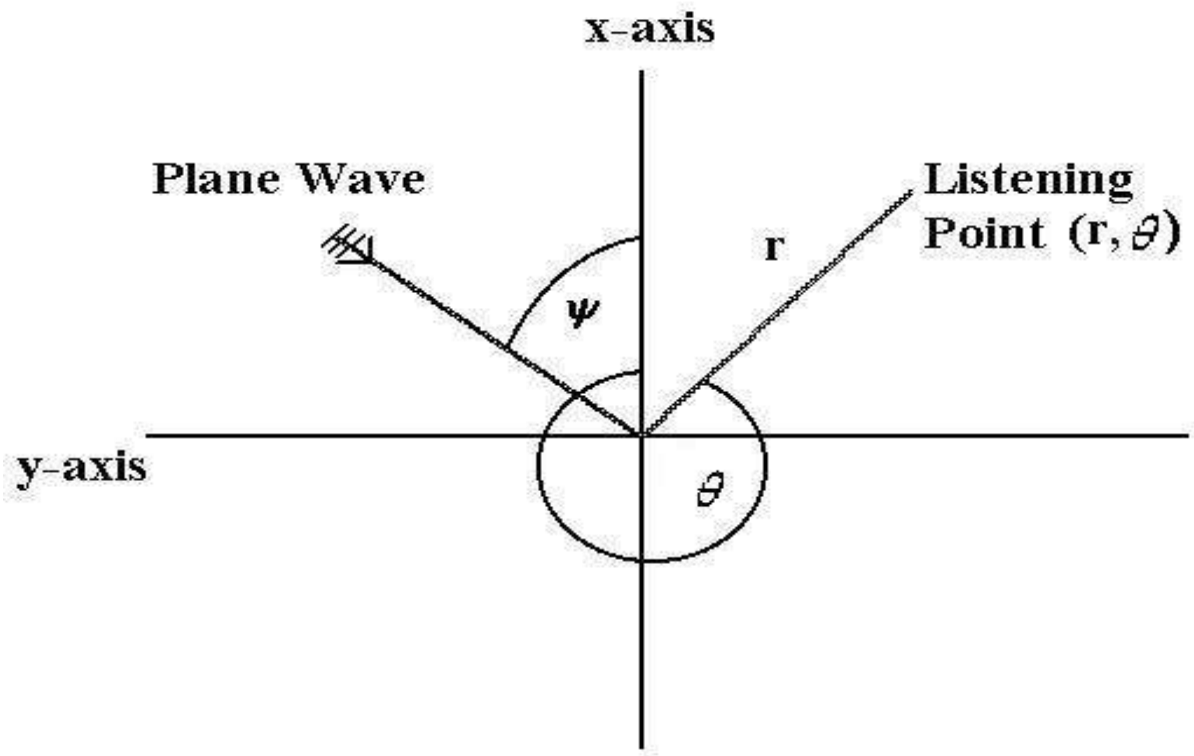
$$p(r,\theta,\varphi) = \sum_{m,n,\sigma} \tilde{A}^{\sigma}_{mn} \tilde{Y}^{\sigma}_{mn}(\theta,\varphi)\, j^{m}\, j_{m}(kr)$$



Ambisonics: synthesis of a plane wave

- **Plane wave**

$$S_\psi = P_\psi e^{ikr\cos(\theta-\psi)}$$



Ambisonics: synthesis of a plane wave

- **Plan wave in terms of spherical harmonics is given by:**

$$S_\psi = P_\psi J_0(kr) + P_\psi \left(\sum_{m=1}^{\infty} 2i^m J_m(kr)[\cos(m\psi)cos(m\theta) + sin(m\psi)sin(m\theta)] \right)$$

- **If we suppose that the wave that comes from each loudspeaker is approximately a plane wave at the listening point and if the loudspeaker is at position ϕ_n**

$$S_n = P_n J_0(kr) + P_n \left(\sum_{m=1}^{\infty} 2i^m J_m(kr)[\cos(m\phi_n)cos(m\theta) + sin(m\phi_n)sin(m\theta)] \right)$$

$$S = \sum_{n=1}^{N} P_n J_0(kr) + \sum_{m=1}^{\infty} 2i^m J_m(kr) \left(\sum_{n=1}^{N} P_n \cos(m\phi_n)cos(m\theta) + \sum_{n=1}^{N} P_n sin(m\phi_n)sin(m\theta) \right)$$

Ambisonics: synthesis of a plane wave

- **We can then obtain the following equations (term by term identification):**

$$
\begin{aligned}
P_\psi &= \sum_{n=1}^{N} P_n \\
P_\psi \cos(m\psi) &= \sum_{n=1}^{N} P_n \cos(m\phi_n) \\
P_\psi \sin(m\psi) &= \sum^{N} P_n \sin(m\phi_n)
\end{aligned}
$$

Ambisonics: synthesis of a plane wave

- **B-Format : only 1st order**

$$W = P_\psi = \sum_{n=1}^{N} P_n$$

omnidirectionnal component

$$X = P_\psi \cos(\psi) = \sum_{n=1}^{N} P_n \cos(\phi_n)$$

X axis component

$$Y = P_\psi \sin(\psi) = \sum_{n=1}^{N} P_n \sin(\phi_n)$$

Y axis component

- We can show that (*for a plane wave*)

$$P_n = \frac{1}{N}\left(W + 2X\cos\phi_n + 2Y\sin\phi_n\right)$$

Ambisonics: decoding and rendering

- **Ambisonics system: in matrice form** $\mathbf{b} = \mathbf{A} \cdot \mathbf{x}$

$$\mathbf{b} = [W, X, Y, U, V, \ldots]^{\mathbf{T}}$$

$$\mathbf{x} = \left[P_1, P_2, P_3, \ldots P_N\right]^{\mathbf{T}}$$

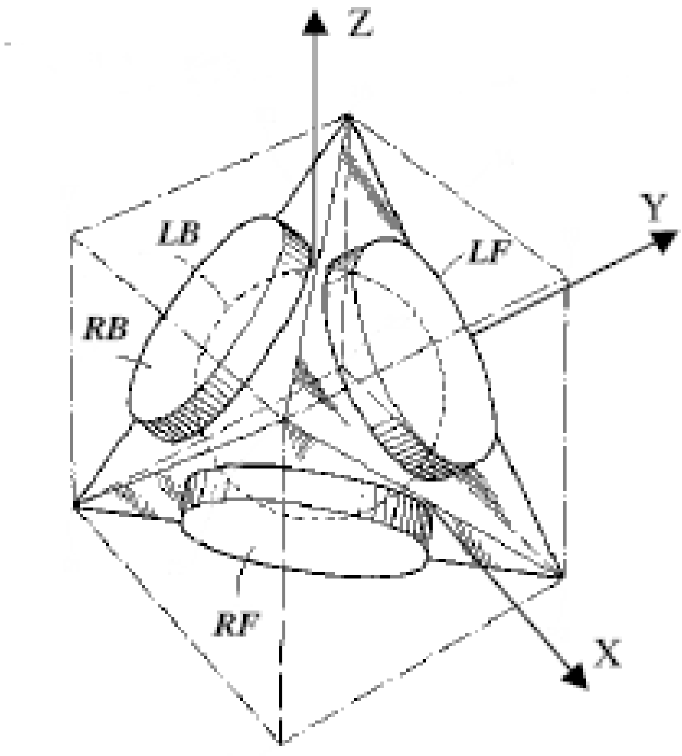
$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \cos(\varphi_1) & \cos(\varphi_2) & & \cos(\varphi_N) \\ \sin(\varphi_1) & \sin(\varphi_2) & & \sin(\varphi_N) \\ \vdots & \vdots & & \vdots \\ \sin(m\varphi_1) & \sin(m\varphi_2) & \cdots & \sin(m\varphi_N) \end{bmatrix}$$

- **Solution given by :**

$$\mathbf{x} = \mathbf{A}^{\mathbf{T}} \cdot (\mathbf{A} \cdot \mathbf{A}^{\mathbf{T}})^{-1} \cdot \mathbf{b}$$

Ambisonics: recording

- **SoundField microphone**



$$\left\{\begin{array}{l} W = LF + LB + RF + RB \\ X = LF - LB + RF - RB \\ Y = LF + LB - RF - RB \\ Z = LF - LB - RF + RB \end{array}\right.,$$

Format-B *Format-A*

Microphone HOA or « soundfield »

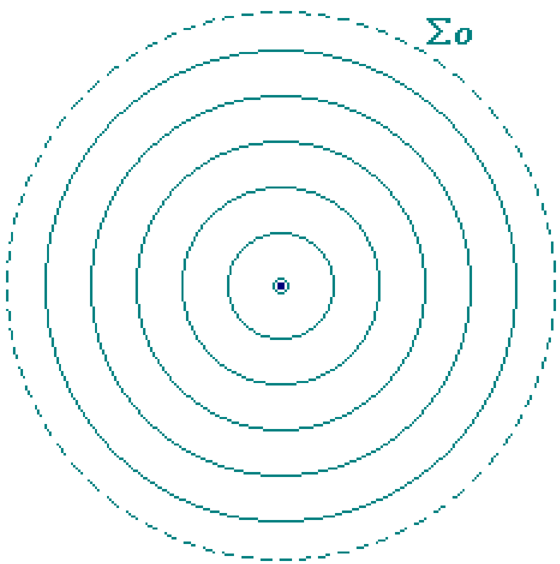
Experimental:
France Telecom, JES 2006)

Tetramic (Coresound)

Microphone NAB

Wave field synthesis

- Huyghens principle : **« Vibrations which propagates inside a closed surface *Σo* *containing the source are identical to those that we would obtain by replacing this source by new sources adequately placed on the surface Σo* »**



Wave Field Synthesis

(DELFT, IRCAM, ORANGE, University of Erlangen Nuremeberg)

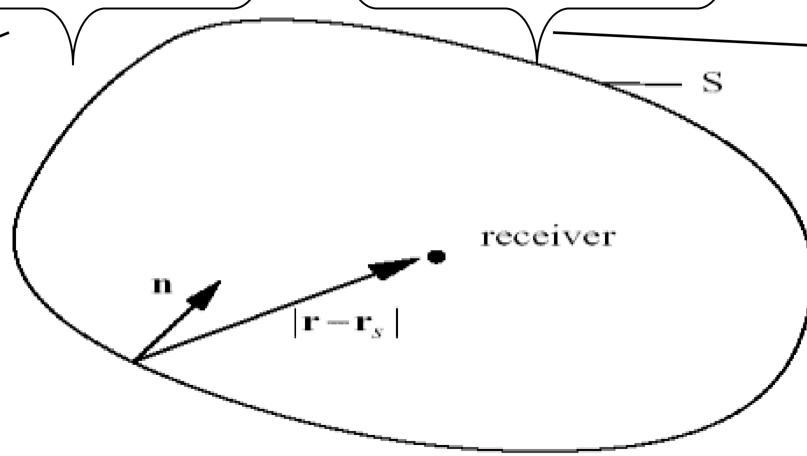
- Linear acoustic theory: **It is possible to reproduce any sound field inside a volume enclosed by a surface by placing sources (monopoles and dipoles) on this surface (considering that there are no other sources inside this volume).**

- Kirchoff-Helmholtz integral

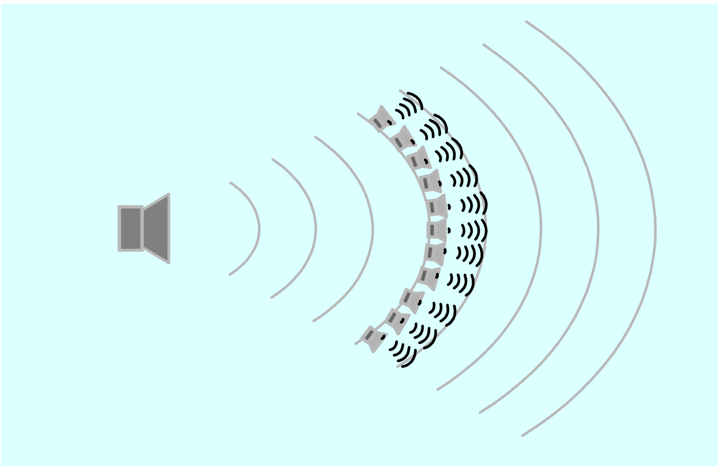
$$P(\mathbf{r},\omega)=\frac{1}{4\pi}\iint_S\left[P(\mathbf{r}_s,\omega)\frac{\partial}{\partial n}\left(\frac{e^{-jk|\mathbf{r}-\mathbf{r}_s|}}{|\mathbf{r}-\mathbf{r}_s|}\right)-\frac{\partial P(\mathbf{r}_s,\omega)}{\partial n}\frac{e^{-jk|\mathbf{r}-\mathbf{r}_s|}}{|\mathbf{r}-\mathbf{r}_s|}\right]dS$$

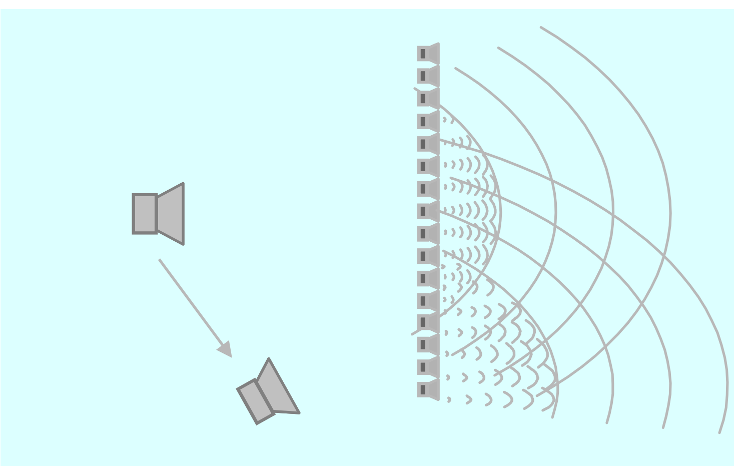
Dipoles: sources linked to pressure

Monopoles: sources linked to sound field velocity



Wave Field Synthesis (WFS) – Concept

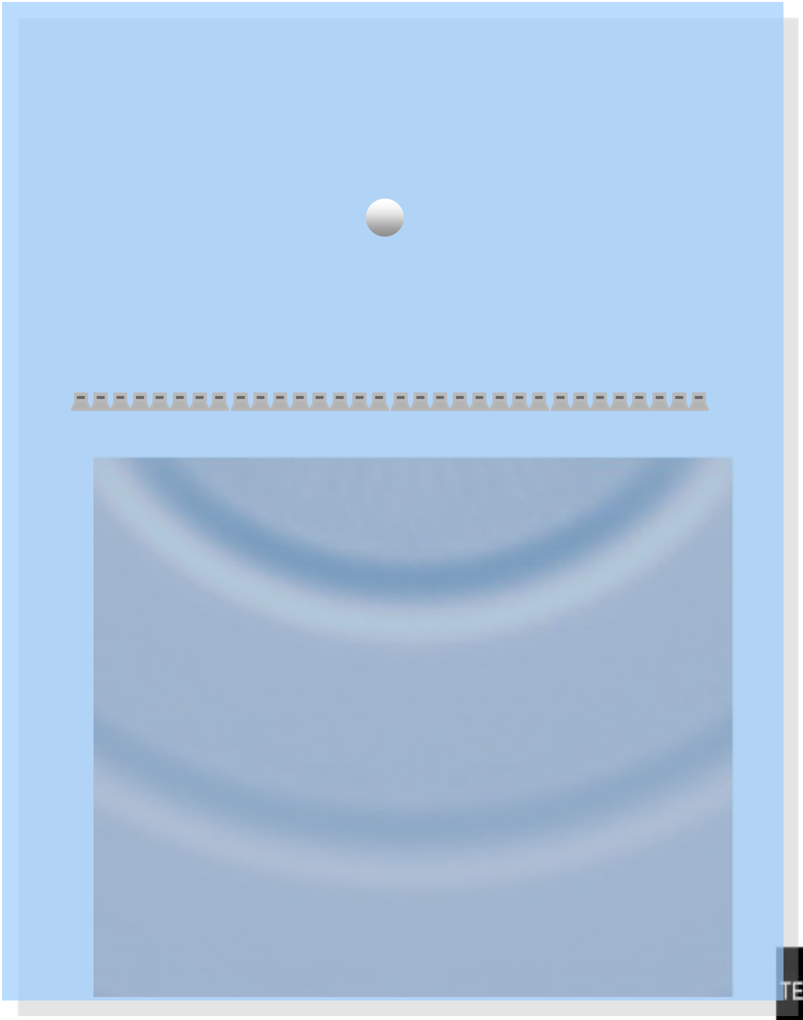




- **Physical reproduction of the sound field produced by a primary source with an ensemble of secondary sources.**
- **Impedance relation: the knowledge of pressure is enough.**
- **Possible to adapt to any geometry of loudspeaker.**

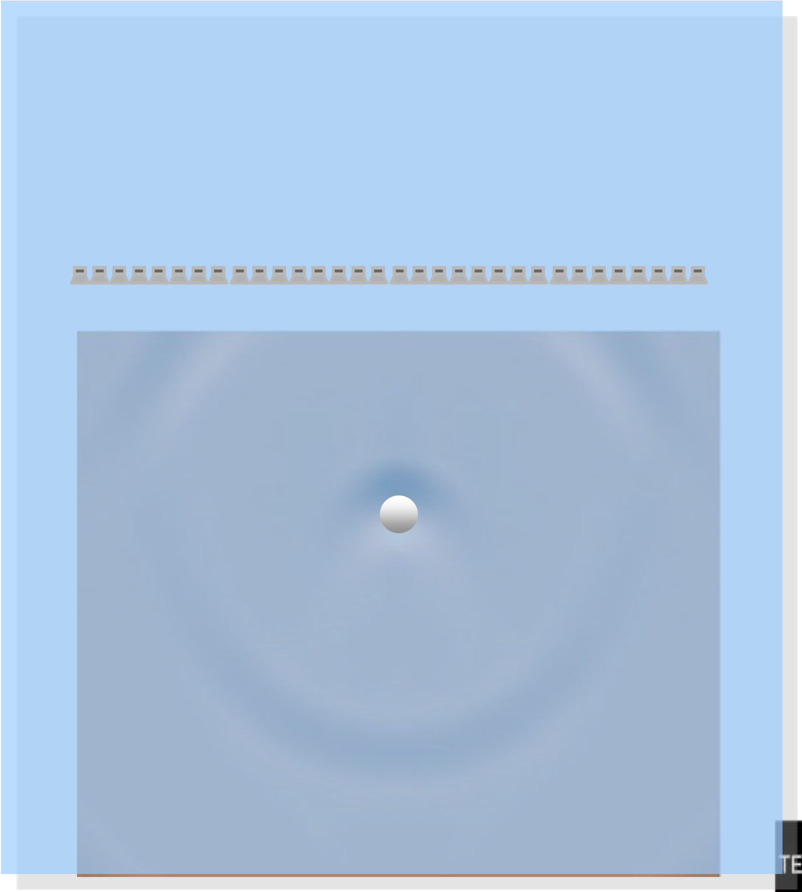
Recording and reproduction in WFS
(from E. Corteels)

- Recording
- Reproduction
- Synthesis
  - **Virtual punctual source**
  - **Large zone reproduction.**



Reproduction in WFS
(from E. Corteels)

- **Potentially in front of the loudspeakers**
- **But not realistic between loudspeakers and virtual source …**



Examples at IRCAM

Loudspeakers MAP « Multi Actuator Panel »

(8 actuators per panel)

Microphone array for room corrections

Sound rendering

- Introduction
- Some elements on spatial perception

- Multi-channel systems
  - **Stereophony**
  - **‘Surround’ systems**
  - **Ambisonics.**
  - **Wave-field synthesis.**

- **Binaural techniques**

- Applications

Binaural approaches

- **Recording:** to record exactly the signal received at the listener ears (artificial head, binaural microphones)

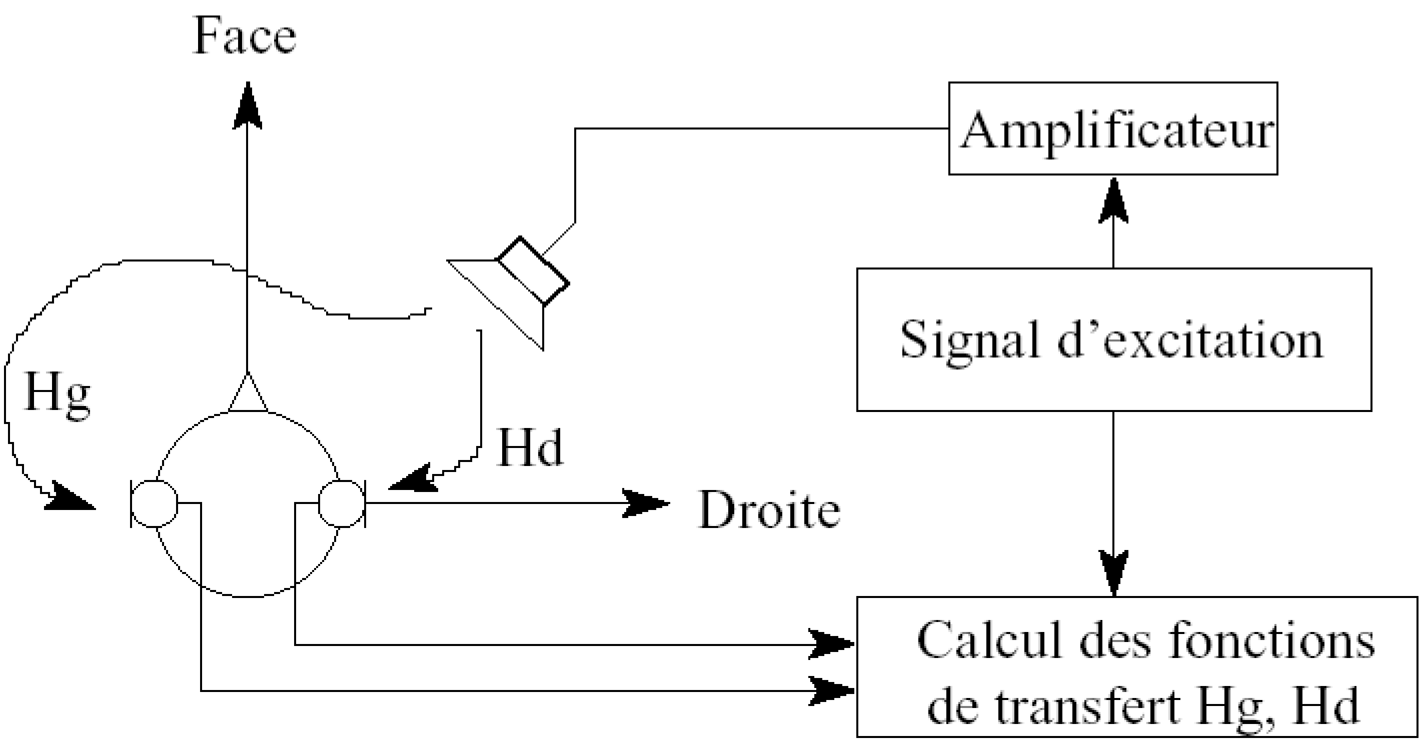
- **Synthesis:** from a monophonic recording (anechoic) using binaural filters to synthesize a binaural listening

  - **For headphone listening: localisation « outside the head »**

  - **For loudspeaker listening (transaural): 3D localisation with 2 loudspeakers (and not only in a plane between loudspeakers)**

Binaural synthesis

- **Measuring binaural filters (Head Related Transfer Functions, HRTFs)**



Measuring filter's impulse response

- **Using Maximum Length sequences (LMS)**
  - **Periodic signals of period L= 2^{k}-1**
  - **Signals take only +1 ou –1 values**
  - **Their autocorrelation function is almost a dirac :**

$$C_{xx}^{L}(n) = \frac{1}{L}\sum_{m=0}^{L-1} x_{m-n}x_m \; = (1+\frac{1}{L})\delta_{[L]}(n) - \frac{1}{L}$$

– We then have:

$$C_{xy}^{L}(n) = \frac{1}{L}\sum_{m=0}^{L-1} x_{m-n}y_m = (1+\frac{1}{L})\delta_{[L]}(n) \star h_n - \frac{1}{L}\bar{h}$$

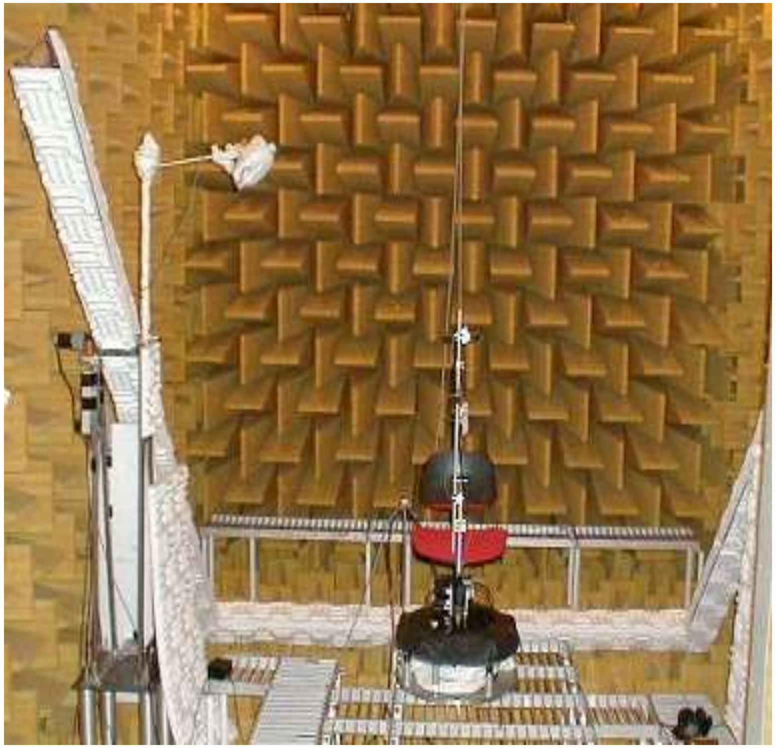
- **Sweep**

Measuring binaural filters (Head Related Transfer Functions, HRTFs)

- **Using a KEMAR head** *(example from CIPIC)*

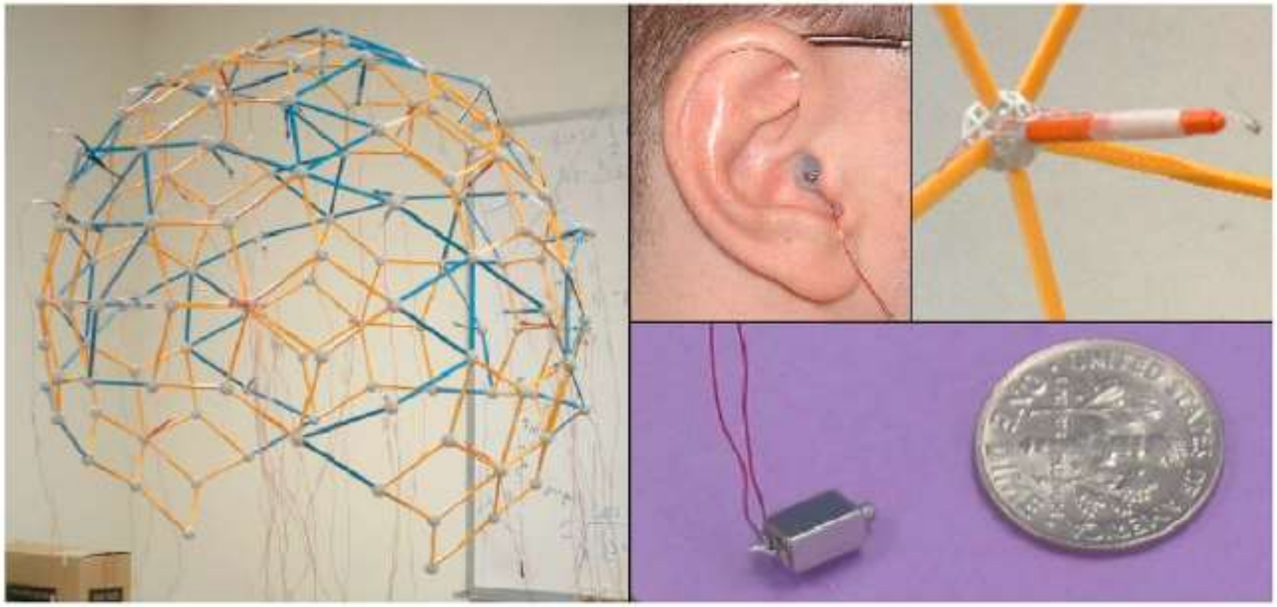
Measuring binaural filters (Head Related Transfer Functions, HRTFs)

- Examples from (Telecom ParisTech, IRCAM)

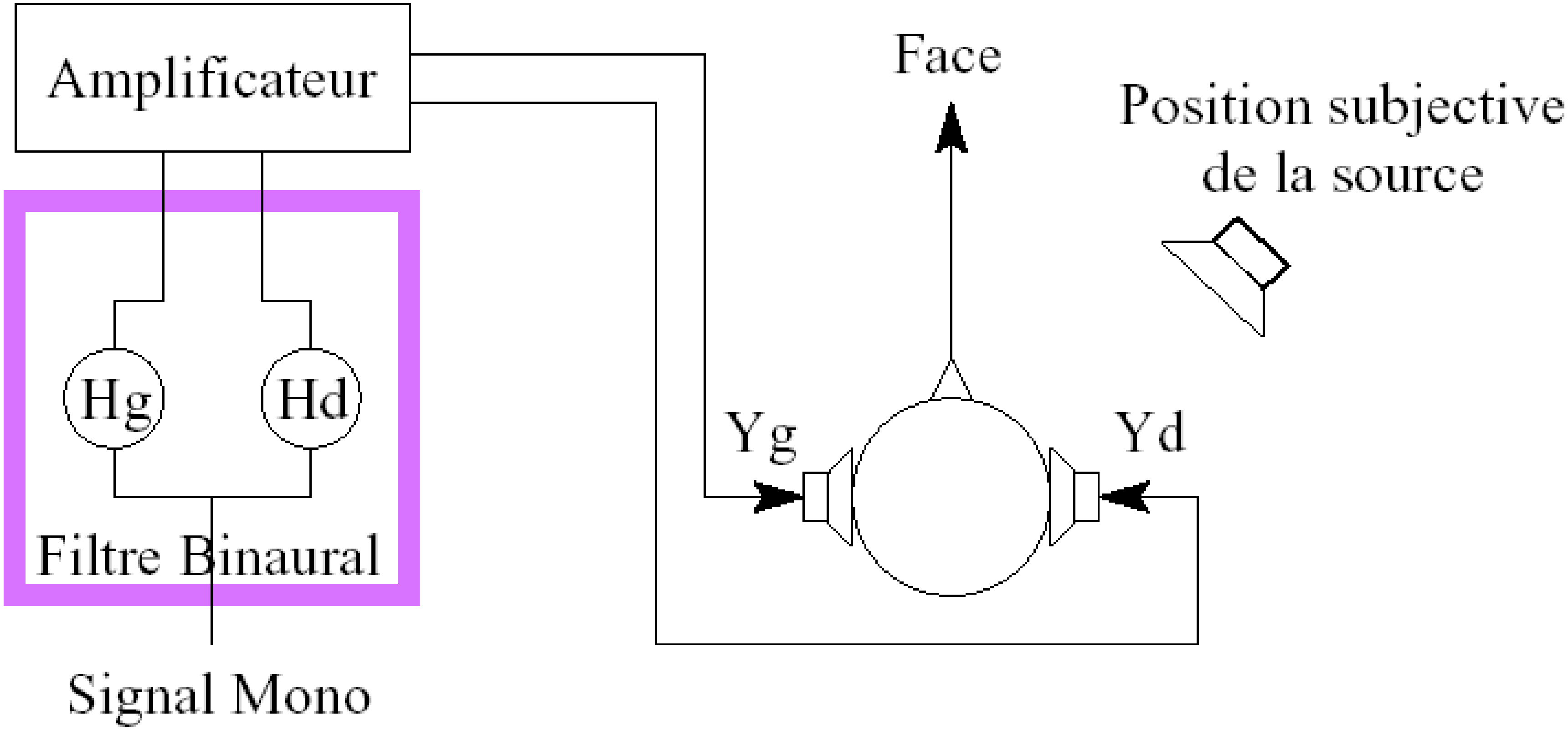


Measuring binaural filters (Head Related Transfer Functions, HRTFs)

- **For individuals** *(from Maryland University)*



Binaural synthesis



Problems with HRTFs

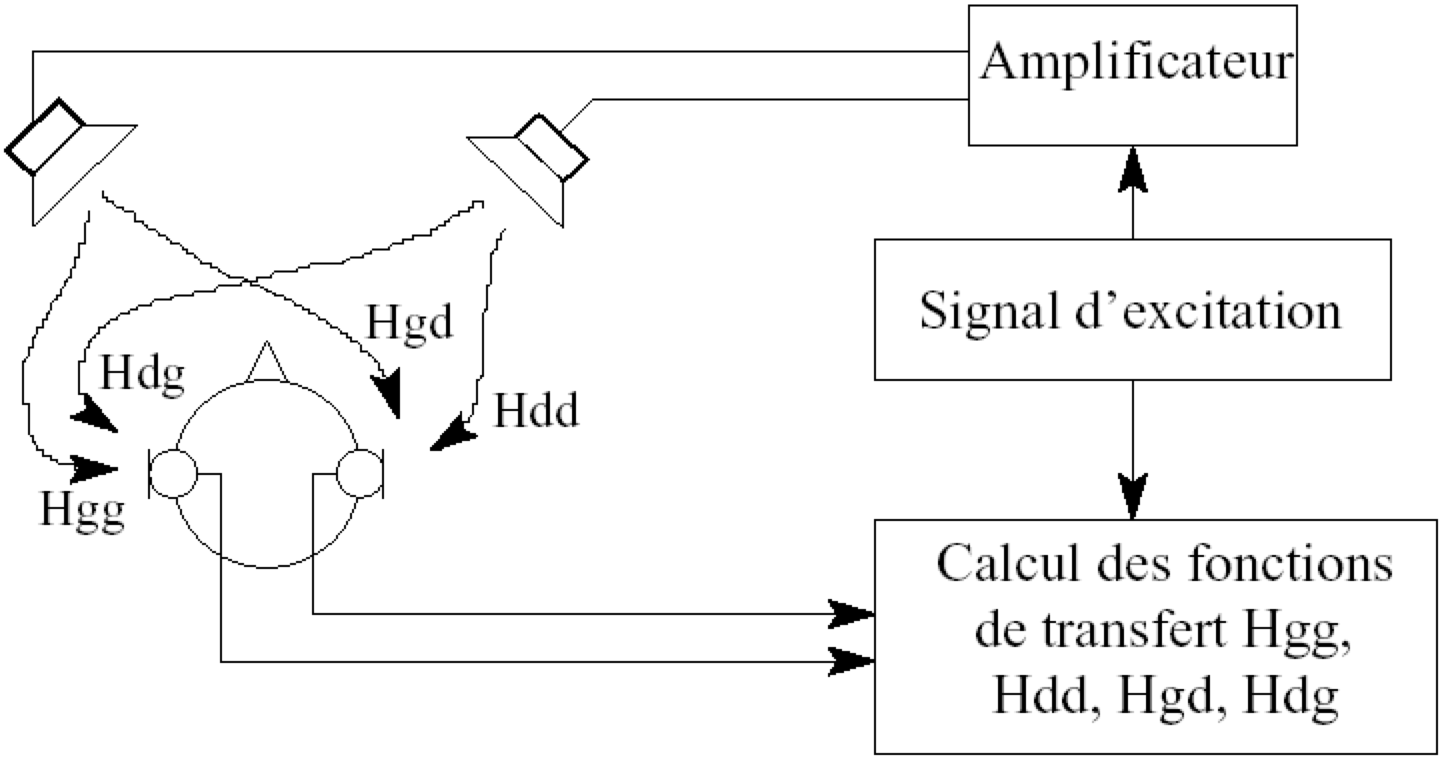
- A major problem: individualization of HRTFs
- Dynamic systems: interpolation of HRTFs and head tracking
- Distance rendering
- HRTFs modelling …

Public datasets

- **IRCAM: LISTEN**
- **CIPIC**
  - **HRTFs measured for discrete values of azimuth, elevation and time**
    - Sampled at 44.1 kHz (200 samples)
    - 25 azimuths (between –90° and +90°)
    - 50 elevations (between – 45° and 230° by steps of 5.625°)
  - **HRTFs measured for over 90 persons (the public dataset comprises 45 persons!)**

Conversion binaural ⬄ transaural

- **Needs the measurements of the 2 transfer functions from the loudspeakers to the listener.**



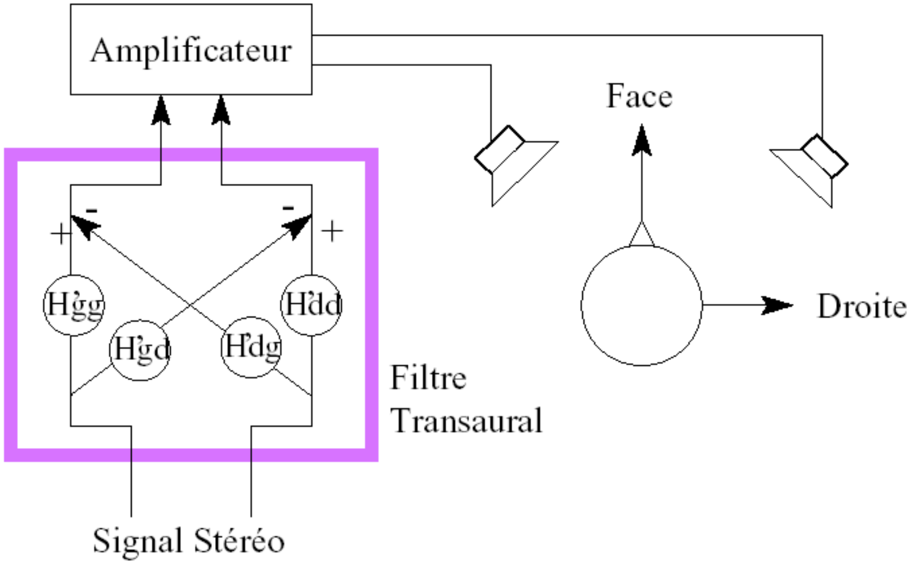
Conversion binaural ⬄ transaural

- **Transfer matrix characterizing the transformations of the signals played at the loudspeakers to the signals received at listener's ears** Y_g et Y_d

$$\begin{bmatrix} Y_g \\ Y_d \end{bmatrix} = \begin{bmatrix} H_{gg} & H_{dg} \\ H_{gd} & H_{dd} \end{bmatrix} \begin{bmatrix} Z_g \\ Z_d \end{bmatrix}$$

- The conversion binaural ⇨ transaural is the inverse matrice

$$\begin{bmatrix} Z_g \\ Z_d \end{bmatrix} = \frac{\begin{bmatrix} H_{dd} & -H_{dg} \\ -H_{gd} & H_{gg} \end{bmatrix}}{H_{gg}H_{dd} - H_{dg}H_{gd}} \begin{bmatrix} Y_g \\ Y_d \end{bmatrix}$$



Conversion binaural ⇔ transaural

- **Possible Symetry hypotheses (reproduction symetry and hearing property symetry)**

$$H_g = G \quad ; \quad H_d = D \quad ; \quad H_{gg} = H_{dd} = G_0 \quad ; \quad H_{gd} = H_{dg} = D_0$$

✓ Binaural transfer functions

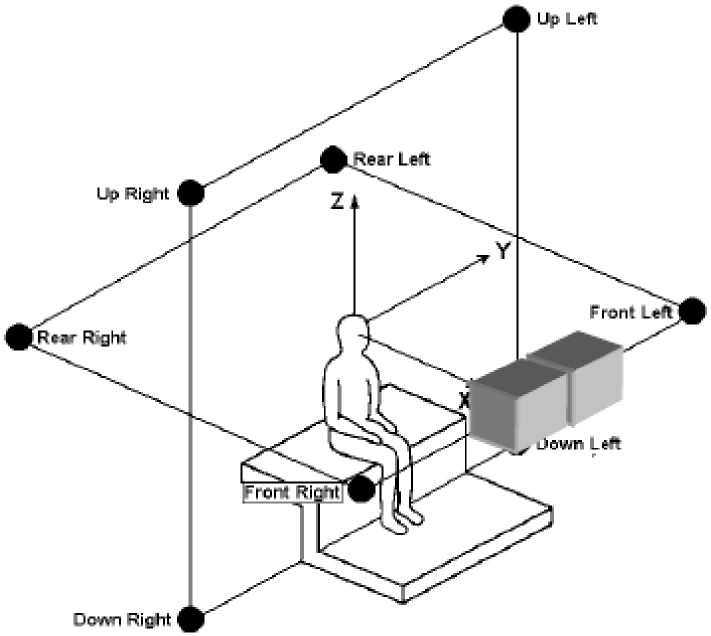
✓ ***G et D*** **:** for any source position

✓ ***G_0 et D_0*** **:** for the position of the left loudspeaker

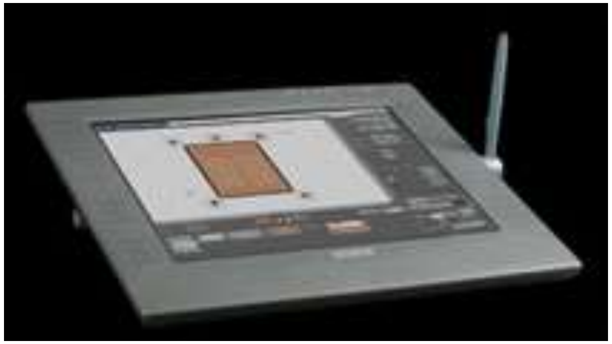
✓ **We then obtain:**

$$\begin{bmatrix} Z_g \\ Z_d \end{bmatrix} = \frac{\begin{bmatrix} G_0 & -D_0 \\ -D_0 & G_0 \end{bmatrix}}{G_0^2 - D_0^2} \begin{bmatrix} Y_g \\ Y_d \end{bmatrix} = \frac{1}{2} \left(\frac{\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}}{G_0 + D_0} + \frac{\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}}{G_0 - D_0} \right) \begin{bmatrix} Y_g \\ Y_d \end{bmatrix}$$

A transaural system : Ambiophonics



Some rendering systems examples: IOSONO



Existing applications…

- Ambiophonics Playstation
- Home Theathers (5.1 et Virtual Speaker Technology)
- WFS: cinéma
- *3D Sound* by Arkamys sur Renault Mégane (2009)

Existing applications…

- **Ambisonics (http://www.ambisonic.net/): used in sound engineering…micro Soundfield**

- **Dolby Headphones**

..or almost

- **MPEG Surround**
- **Binaural Mobile**
- **Binaural iPod**
- **Spatial music production**

A few organisation active in 3D rendering

- IRCAM
- Orange Labs
- LMA Marseille
- Genesis acoustics
- Trinnov Audio
- Télécom ParisTech…
- A-Volute
- LIMSI
- Euphonia
- ARKAMYS
- Sonic emotion
- Immersion

A few links

- **An overview on spatial reproduction**
  - **http://www.deutsche-telekom-laboratories.de/~sporssas/publications/talks/AES128_Tutorial_Spatial_Audio_Reproduction.pdf**