



***Master MVA***

**Spatialisation des sons**

**Gaël RICHARD**

**Février 2016**



« Licence de droits d'usage"»

[http://formation.enst.fr/licences/pedago\\_sans.html](http://formation.enst.fr/licences/pedago_sans.html)





# Spatialisation des sons

- Introduction
- Perception Spatiale
- Systèmes multicanaux
  - La stéréophonie.
  - Les systèmes 'surround', ou 'multiphonie'.
  - Ambisonics.
  - Wave-field synthesis.
- Techniques binaurales
- Applications





# Introduction

- **Temps – Fréquence** : **Son**
- **Espace/Temps – Fréquence** : **Champ Sonore**
  
- **Spatialisation** : **définitions.**
  - Science
  - Ingénierie du son
  - Art
  
- **Spatialisation en pratique** : **enregistrement**  
: **synthèse**
  
- **Approches** : **perceptives**  
: **physiques**



# Bref historique de l'utilisation de la spatialisation sonore dans la musique

(d'après A. Dutto: <http://www.nikkojazz.fr/2009/05/la-mise-en-espace-de-la-musique.html>)

## ■ Les prémices:

- 4<sup>ème</sup> siècle: l'Antiphonie, e.g. alternance entre 2 chœurs
- 1573: motet à 40 voix de Thomas Tallis
- Double chœur ensuite utilisé par Vivaldi, Bach,..

## ■ Au 19<sup>ème</sup> siècle:

- Requiem de Berlioz – Tuba Mirum (1837):
  - orchestre monumental + imposante masse chorale + 4 groupes de cuivres aux 4 points cardinaux
  - Question ?: où doit se positionner l'auditeur ?

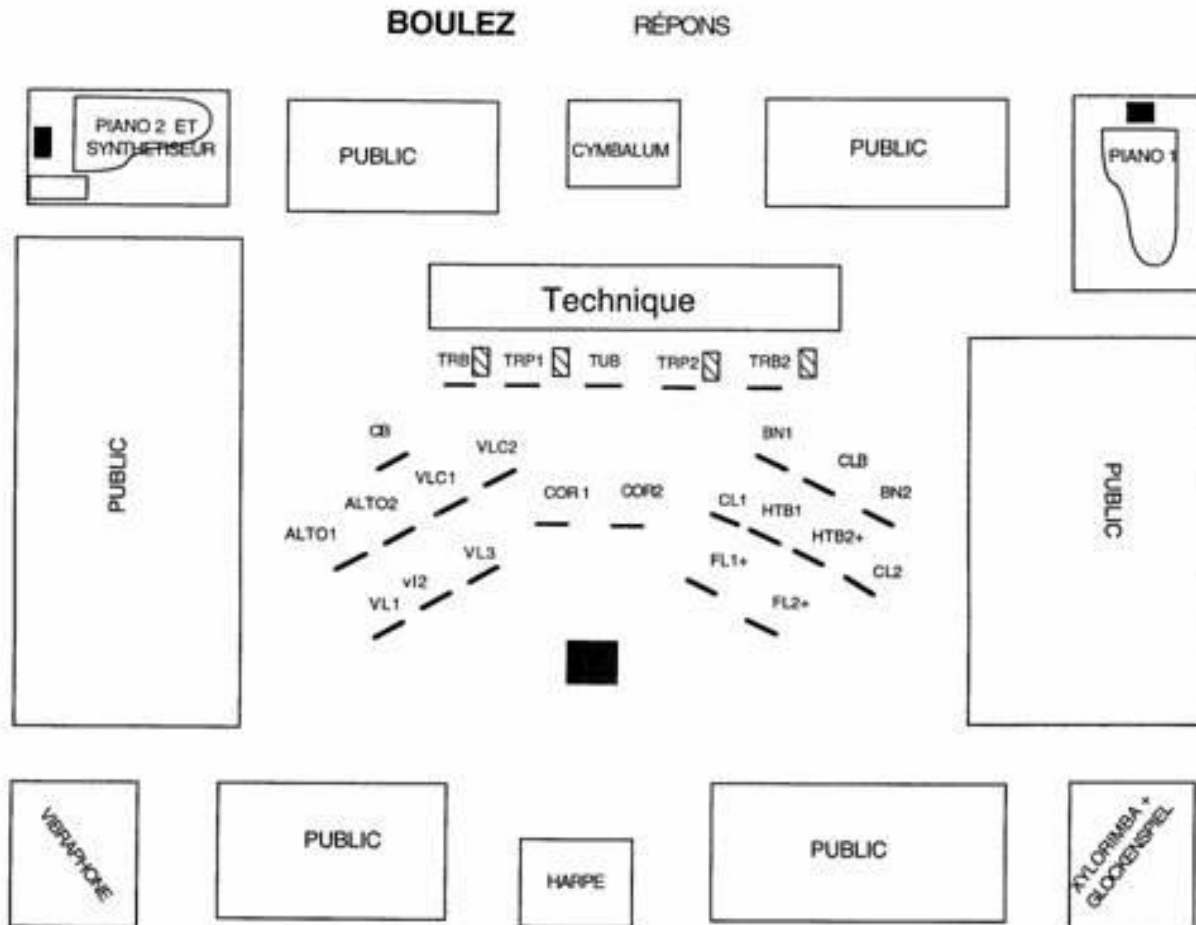
## ■ Le 20<sup>ème</sup> siècle: exploitation de l'espace

- Maderna, Carter, Stockhausen, Xenakis...Boulez



# Bref historique...

## ■ Disposition pour « Repons » de P. Boulez...

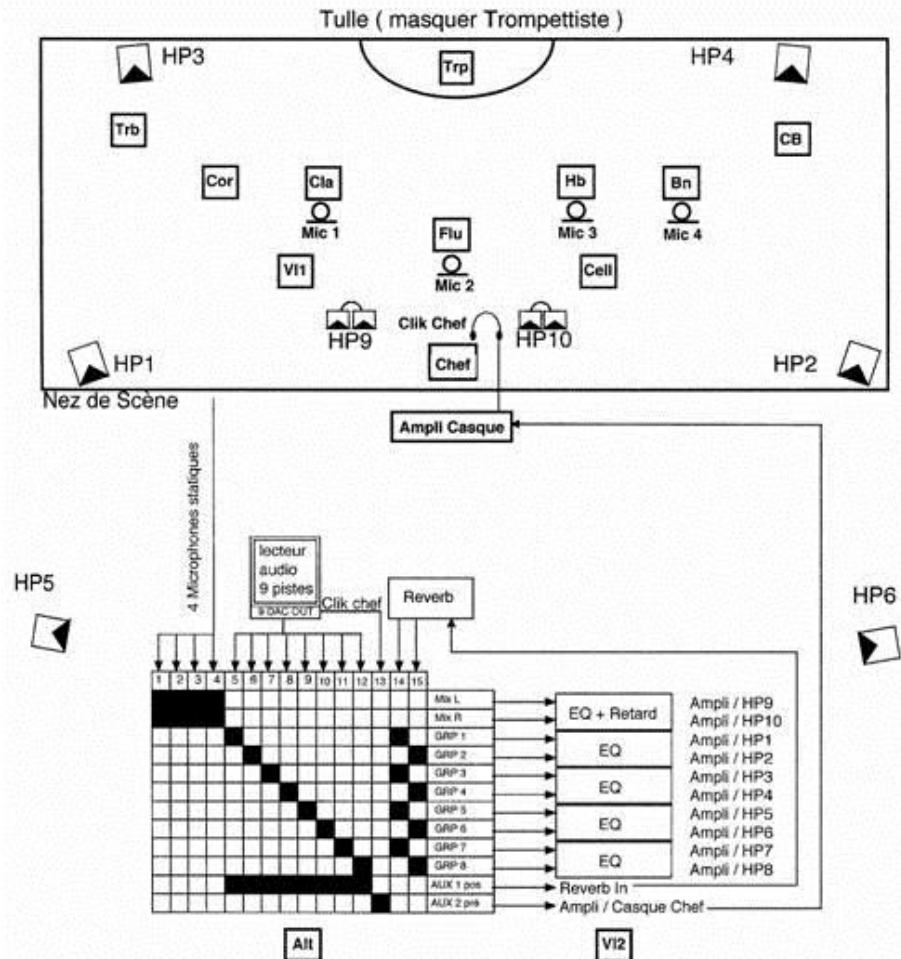


# Bref historique...

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

## ■ Un autre exemple: *Le plein du vide*, Xu Yi (1997)

- 8 pistes d'ordinateur diffusées sur 8 HP disposés sur scène et dans le public,
- Le public est au centre de l'oeuvre



G. Richard

# Techniques de spatialisation

- **Physiques** : champ sonores en tant que phénomènes physiques
- **Perceptives** : champ sonores en tant que phénomènes physiques perçus
  - Paramètres pertinents: IID, ITD, IPD, HRTFs..
- **Multicanale** : écoute collectif
  - Notion de sweet spot
- **Binaurale** : écoute individuel
  - Optimale pour l'écoute en casque





# Spatialisation des sons

- Introduction
- Perception Spatiale
- Systèmes multicanaux
  - La stéréophonie.
  - Les systèmes 'surround', ou 'multiphonie'.
  - Ambisonics.
  - Wave-field synthesis.
- Techniques binaurales
- Applications





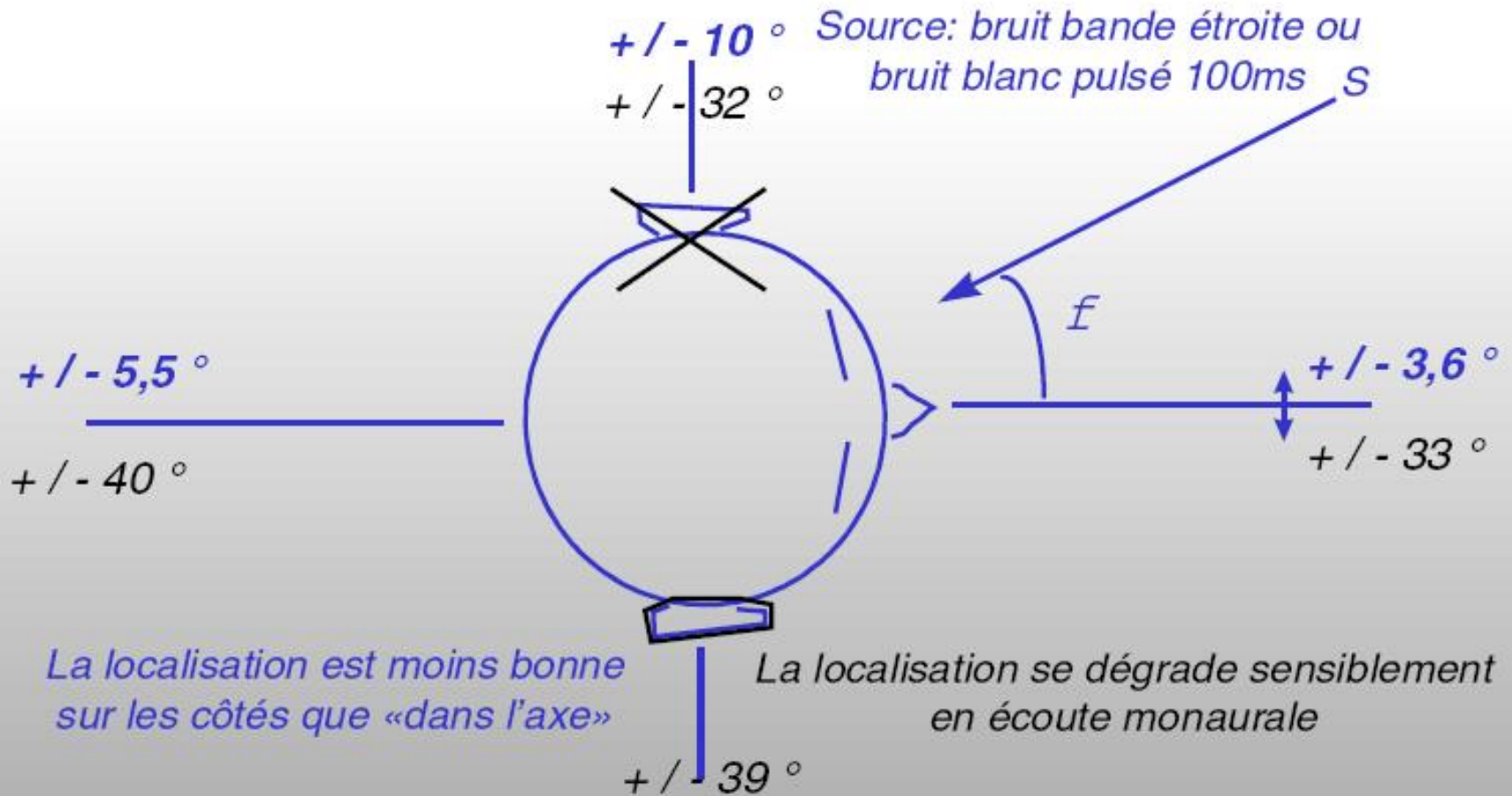


# Introduction

- **2 aspects principaux de la perception de l'espace:**
  - ***la localisation des sources***
    - Cette capacité est d'une précision remarquable, en particulier pour les sons brefs
    - Elle utilise des indices binauraux essentiellement, mais pas uniquement
    - On distingue *localisation* et *latéralisation* (au casque, des sons « dans » la tête)
  - ***l'étendue subjective des sources***



# Localisation des sons en azimut

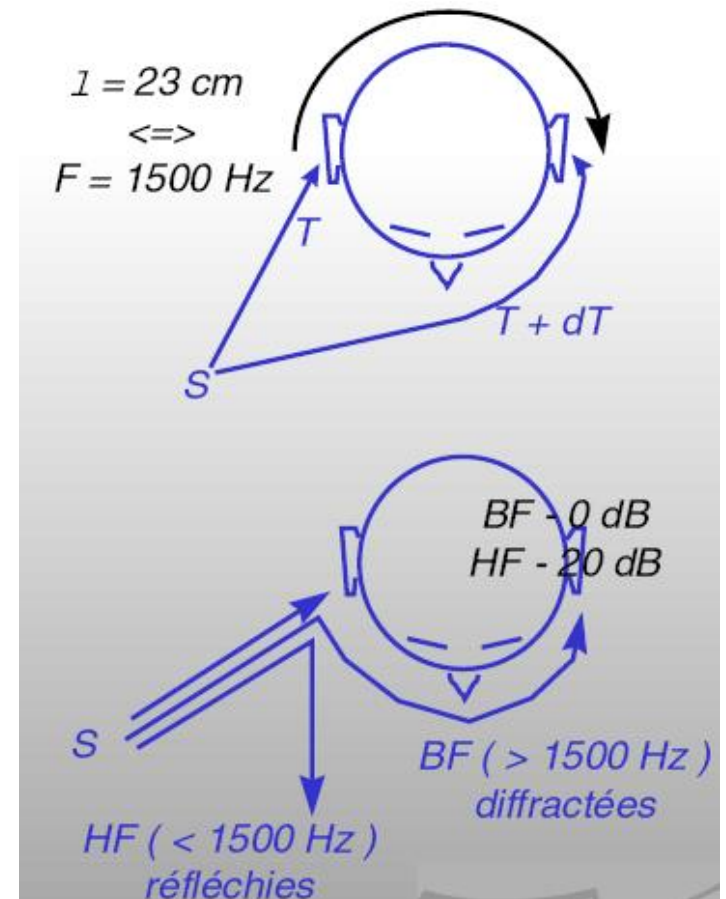


# Localisation des sons en azimut

## ■ 2 types d'indices (Steven et Newman, 1936):

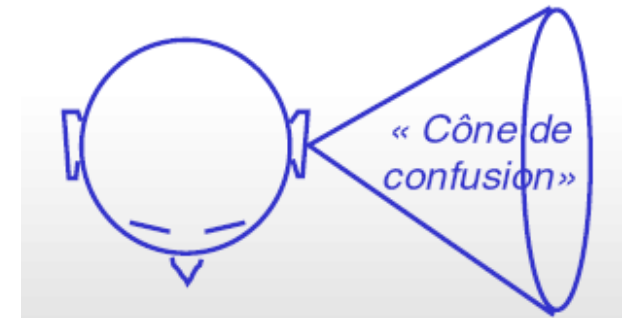
- la diff. interaurale en temps (ITD)  
... surtout en Basses Fréq.
- la diff. interaurale en intensité (IID)  
... surtout en Hautes Fréq.

## ■ ambiguïté maxi. vers 1500Hz



# Capacités monaurales: cône de confusion

- Toutes les sources sur le cône ont la même ITD
- Les ambiguïtés sont résolues par des mouvements de la tête
- Si les mouvements de tête sont permis, la localisation monaurale est presque aussi bonne
- ainsi d'autres indices que les ITD et IID sont utilisés: en particulier, le changement du pattern spectral suivant la direction d'incidence

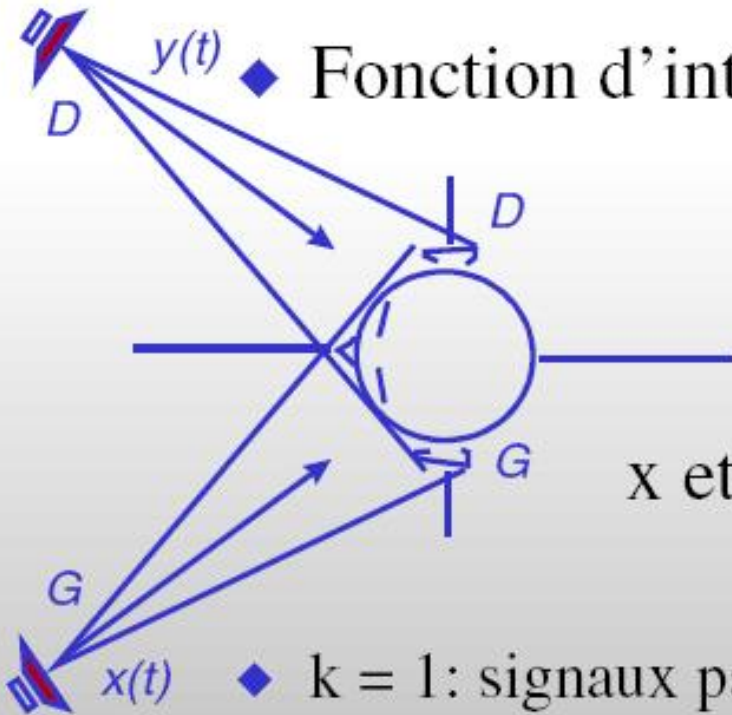


# Rôle de la tête et du pavillon

- *ITDs+IIDs+mouvements de tête ne suffisent pas à expliquer nos capacités de localisation (en élévation)*
- **tête + pavillon : filtrage complexe, effet notable de 500 à 16 000 Hz**
  - pavillon: essentiellement pour les HF (> 6 kHz)
  - baisse des capacités en HF si on obstrue les reliefs
- **la localisation exploite des connaissances sur la source et sur le local, dont l'acquisition est très rapide**



# Etendue subjective des sources



◆ Fonction d'intercorrélation normalisée :

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

x et y sont dits cohérents si:

$$(k =) \max |F_{xy}(t)| = 1$$

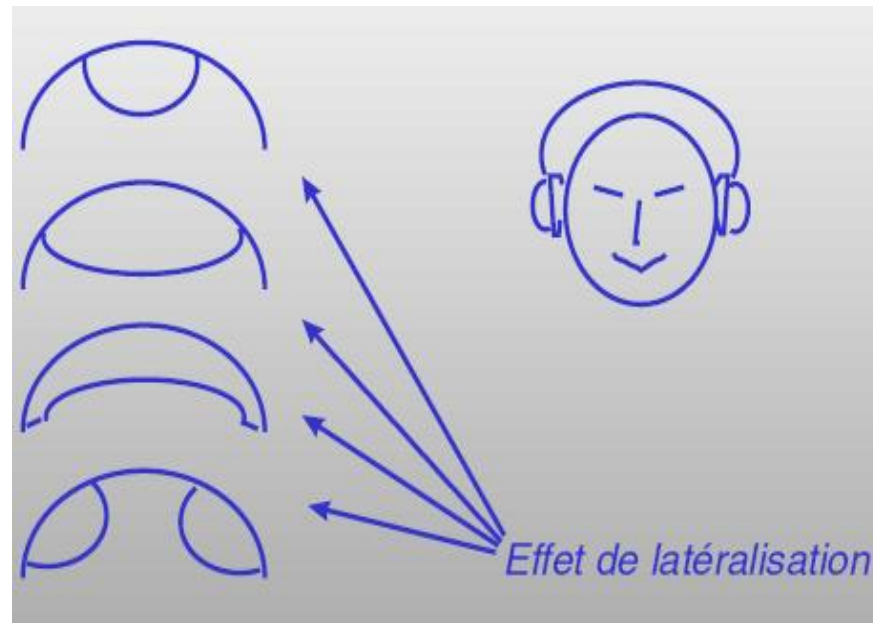
- ◆  $k = 1$ : signaux parfaitement corrélés, 1 seul événement auditif de localisation variable
- ◆  $k = 0$ : signaux décorrélés, 2 événements auditifs localisés chacun sur un des HP
- ◆  $k \approx 0,4$ : sensation équilibrée d'étendue



# Application à la latéralisation

- En utilisant des filtres (passe-bas, passe haut) on peut créer la valeur désirée de  $k$

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



$k=0$





# Conclusion sur la perception de l'espace

- **Elle fait appel à de nombreux indices, notamment:**
  - les ITD
  - les IID
  - le filtrage du pavillon, dépendant de la direction
- **la redondance est très utile en situation difficile, elle est exploitée au mieux par le traitement binaural**
- **notre performance est**
  - bonne pour la localisation des sources en azimut, plus mauvaise en élévation







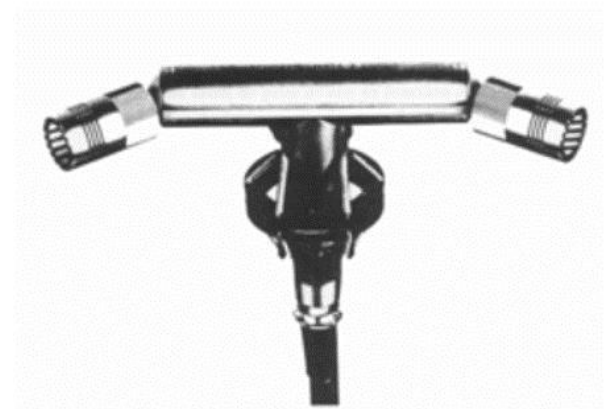
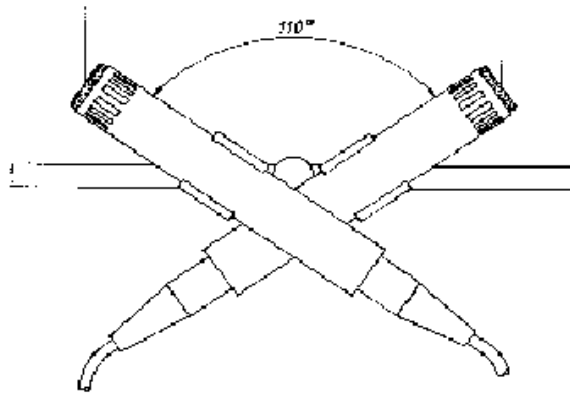
# Spatialisation des sons

- Introduction
- Perception Spatiale
- **Systèmes multicanaux**
  - La stéréophonie.
  - Les systèmes 'surround', ou 'multiphonie'.
  - Ambisonics.
  - Wave-field synthesis.
- Techniques binaurales
- Applications

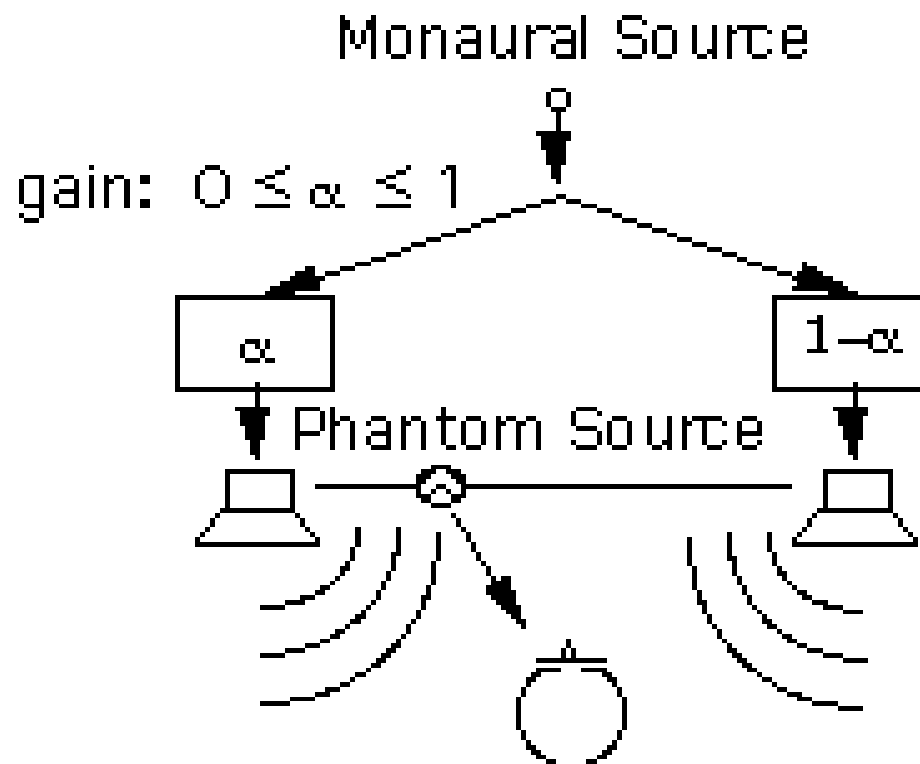


# Enregistrement Stéréophonique

- Utilise les différences d'intensité (IID) et les décalages temporels (ITD) et de phase (IPD).
- Par exemple, enregistrement à l'aide de 2 microphones de directivité différente (XY) ou placés à des positions différentes (AB)



# Synthèse stéréophonique



## ■ Enceintes: deux enceintes avec ouverture de 60 degrés.

- Images fantômes sur la ligne entre les deux haut-parleurs.
- Sweet spot

## ■ Casque

- Localisation à l'intérieur de la tête



# Surround: Systèmes multicanaux

## ■ Système à 2 canaux:

- Limitation de la localisation dans une zone restreinte (frontale en stéréo)
- Contrainte sur la position de l'auditeur

⇒ **augmenter le nombre de canaux**

## ■ Approches « surround »

- 1940: Fantasia (« Fantasound »)
  - 3 canaux (mur gauche, centre, mur droit)
- Années 50: format 4 canaux.
  - Problèmes de compatibilité avec le monophonique
  - Pas de réel succès commercial



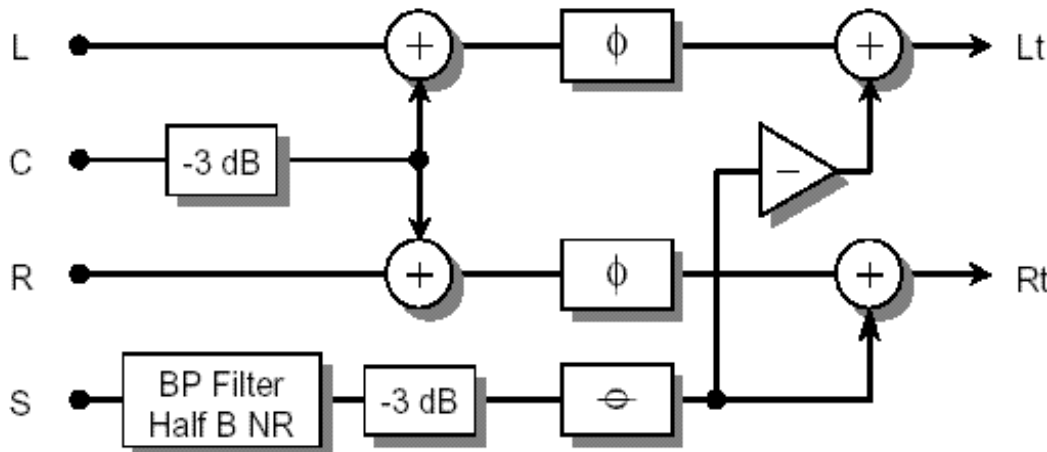


# Histoire

- 1977 : **Dolby Stéréo pour le cinéma (*Starwars*)**
- 1982 : **Dolby Surround pour la maison**
- **Systemes à 4 signaux (G,D,C,S), mais 2 canaux.**
- 1992: **Dolby AC-3 coding, système numérique à 5.1 canaux (qui est devenu le *Dolby Digital*)**
- **Dolby...mais aussi DTS,THX...**



# Dolby surround: codage



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

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

$$L\_d\acute{e}cod\acute{e} = L_t$$

$$R\_d\acute{e}cod\acute{e} = R_t$$

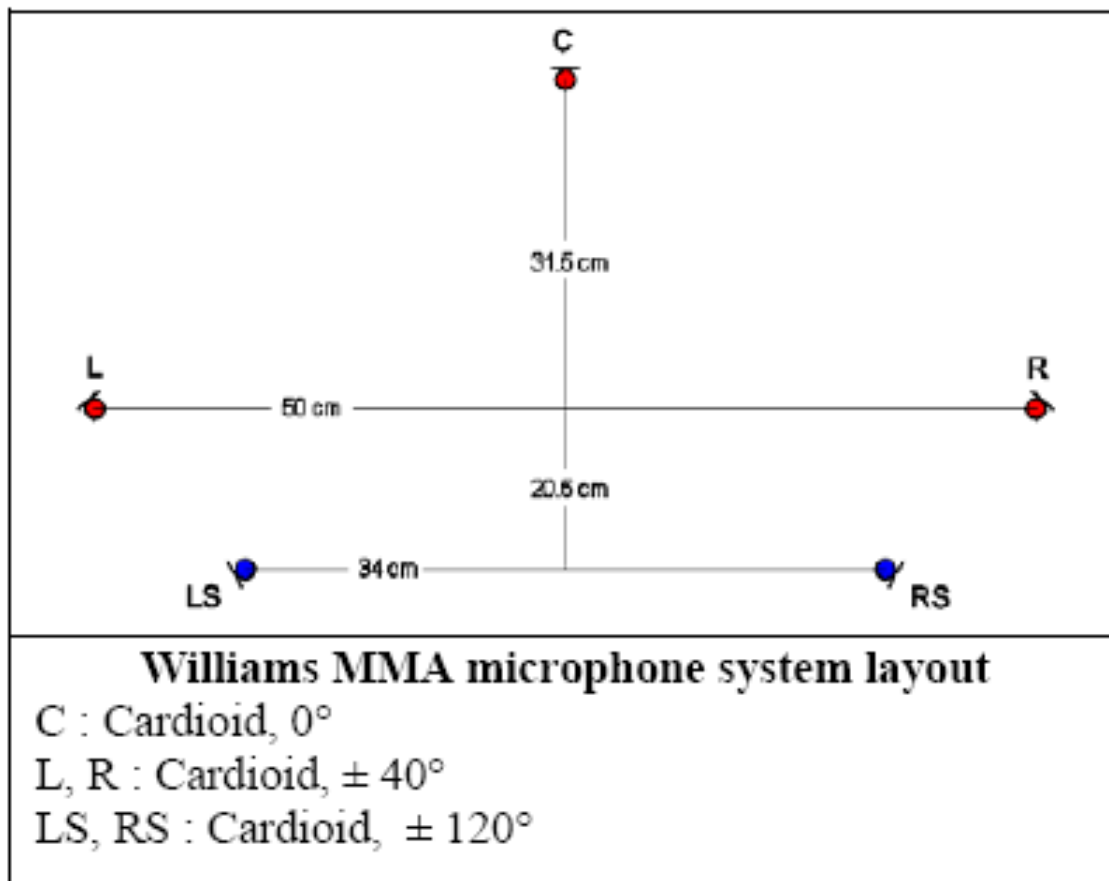
$$C\_d\acute{e}cod\acute{e} = \frac{1}{\sqrt{2}} (L_t + R_t) = C + \frac{1}{\sqrt{2}} (L + R)$$

$$S\_d\acute{e}cod\acute{e} = \frac{1}{\sqrt{2}} (L_t - R_t) = i * S + \frac{1}{\sqrt{2}} (L - R)$$

**(Compatibilit\acute{e} st\acute{e}r\acute{e}o)**

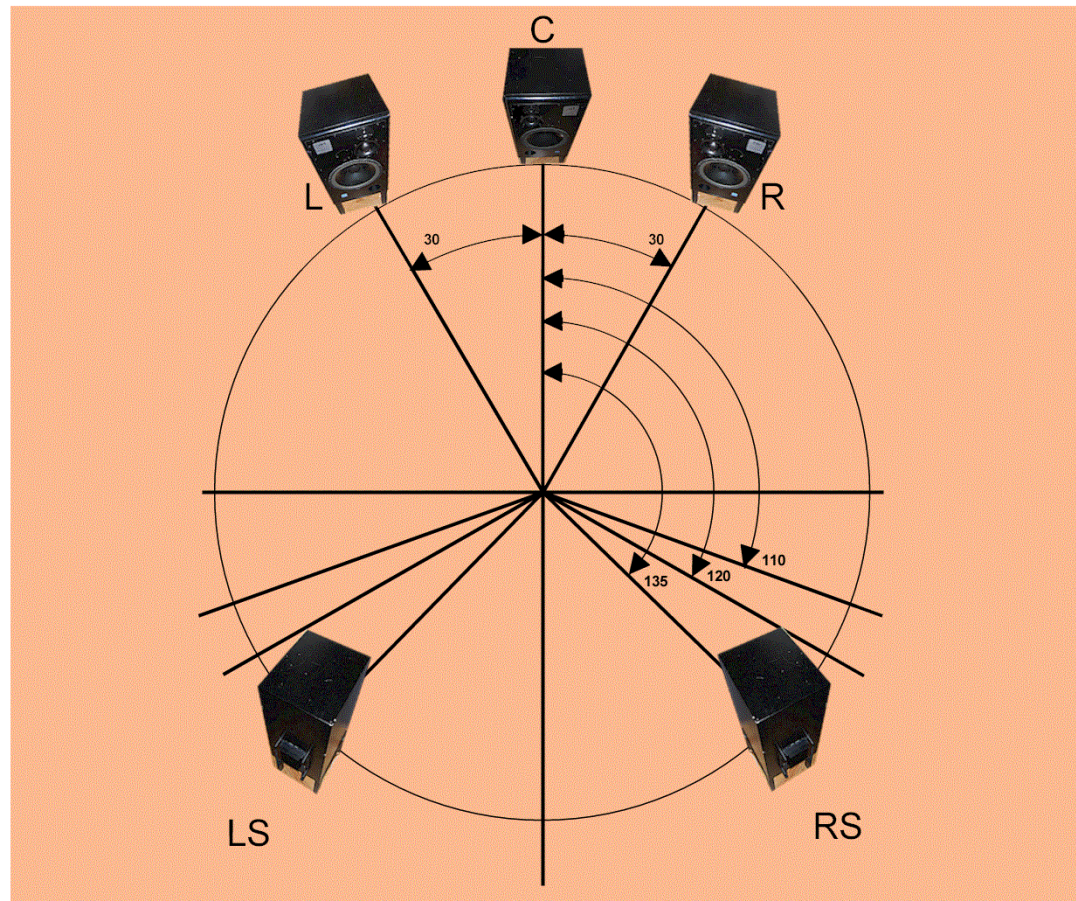


# Dolby 5.1: enregistrement





# Reproduction multicanale (5.1)



# Spatialisation des sons

## ➤ Introduction

## ➤ Systèmes multicanaux

➤ La stéréophonie.

➤ Les systèmes 'surround', ou 'multiphonie'.

➤ Ambisonics (zone étendue, physique).

➤ Wave-field synthesis (zone étendue, physique).

## ➤ Techniques binaurales

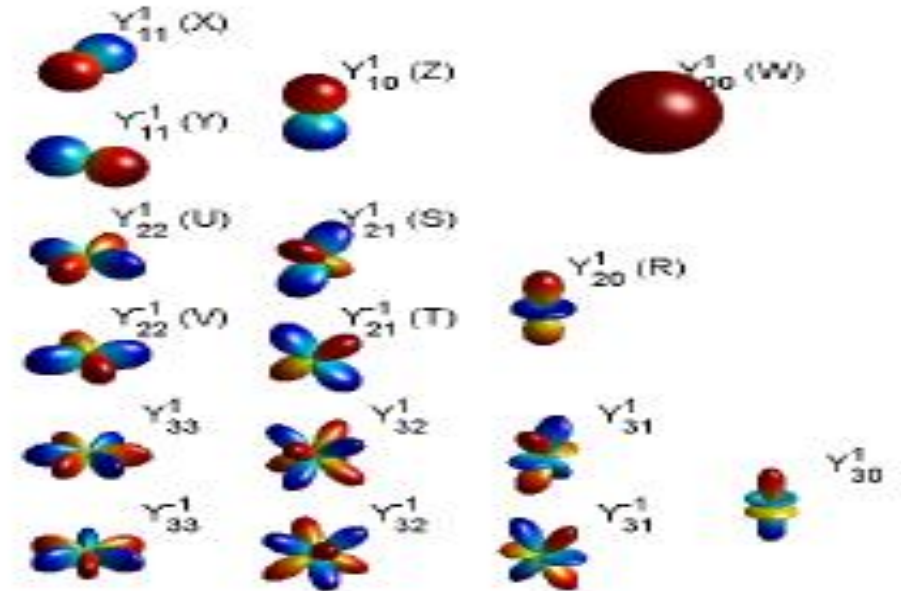
## ➤ Applications



# Ambisonics

- Système à référence centrale.
- Basé sur la décomposition du champ sonore en harmoniques sphériques .

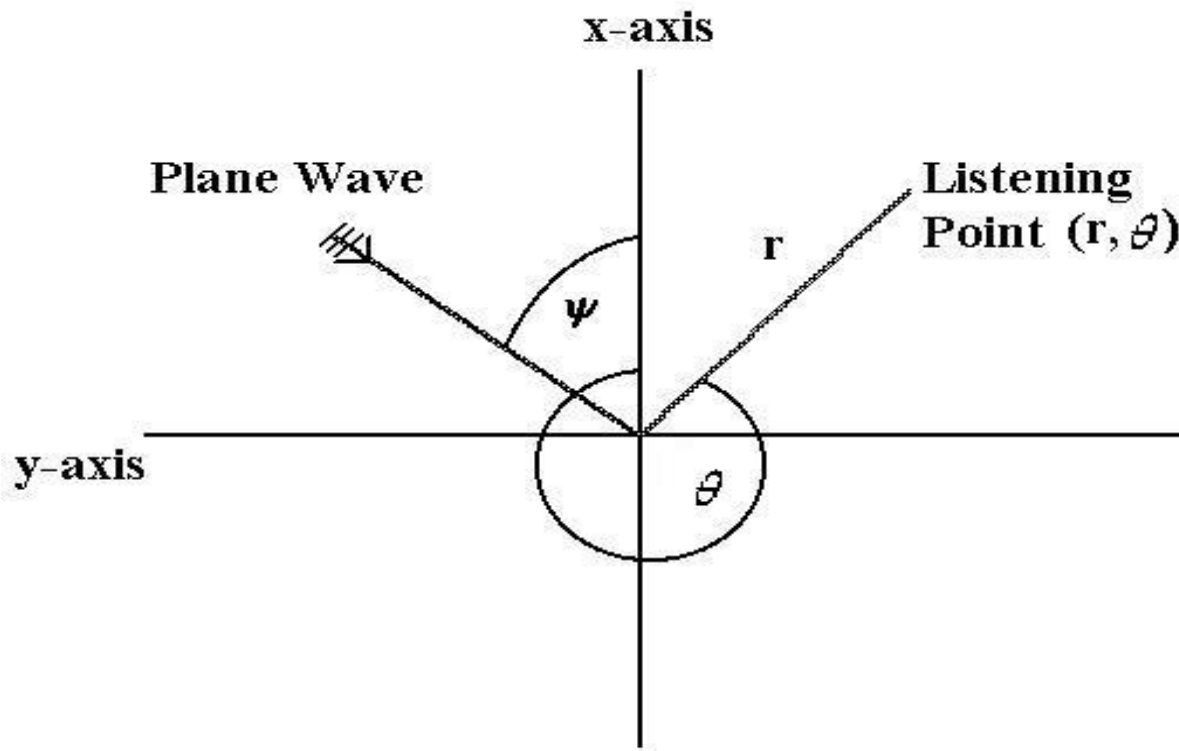
$$p(r, \theta, \varphi) = \sum_{m, n, \sigma} \tilde{A}_{mn}^{\sigma} \tilde{Y}_{mn}^{\sigma}(\theta, \varphi) j^m j_m(kr)$$



# Ambisonics: synthèse d'une onde plane

## ■ Onde plane

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



# Ambisonics: synthèse d'une onde plane

- Le plan d'onde s'écrit en terme d'harmoniques sphériques

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

- En supposant que l'onde provenant de chaque haut parleur est approximativement une onde plane à la position d'écoute et en considérant que le haut parleur est à la position  $\phi_n$

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

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





# Ambisonics: synthèse d'une onde plane

- On peut alors obtenir les équations suivantes en identifiant terme à terme les équations précédentes.

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

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

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



# Ambisonics: synthèse d'une onde plane

## ■ B-Format : Restriction au premier

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

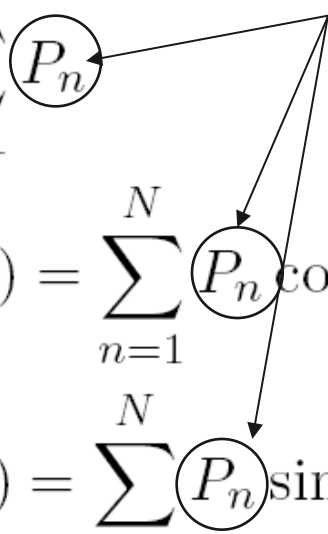
*Composante omnidirectionnelle*

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

*Composante selon l'axe X*

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

*Composante selon l'axe Y*



- On peut montrer que (coefficients de restitution pour une onde plane)

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



# Ambisonics: décodage et reproduction

- Système ambisonics: écriture matricielle:  $\mathbf{b} = \mathbf{A} \cdot \mathbf{x}$

$$\mathbf{b} = [W, X, Y, U, V, \dots]^T$$
$$\mathbf{x} = [P_1, P_2, P_3, \dots, P_N]^T$$
$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \cos(\varphi_1) & \cos(\varphi_2) & & \cos(\varphi_N) \\ \sin(\varphi_1) & \sin(\varphi_2) & & \sin(\varphi_N) \\ \vdots & \vdots & & \vdots \\ \sin(m\varphi_1) & \sin(m\varphi_2) & \dots & \sin(m\varphi_N) \end{bmatrix}$$

- La solution est donnée par:

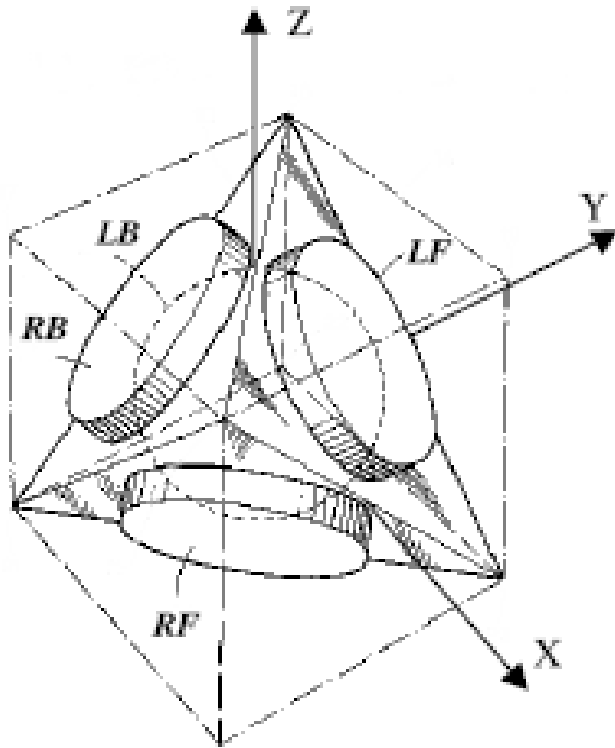
$$\mathbf{x} = \mathbf{A}^T \cdot (\mathbf{A} \cdot \mathbf{A}^T)^{-1} \cdot \mathbf{b}$$





# Ambisonics: enregistrement

## ■ SoundField microphone



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

*Format-B*

*Format-A*



# Microphone HOA ou « soundfield »



Experimental:  
*France Telecom, JES 2006)*



Tetramic (Coresound)

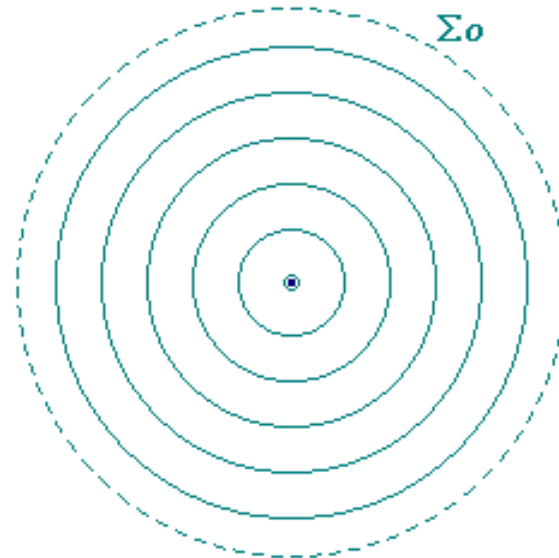


Microphone NAB



## Approches en synthèse des fronts d'onde

- Principe de Huyghens: « ***Les vibrations qui se propagent à l'extérieur d'une surface fermée  $\Sigma_0$  contenant la source sont identiques à celles qu'on obtiendrait en supprimant cette source et en la remplaçant par des sources convenablement réparties sur la surface  $\Sigma_0$***  »



# Wave Field Synthesis

(DELFT, IRCAM, France TELECOM, University of Erlangen Nuremeberg)

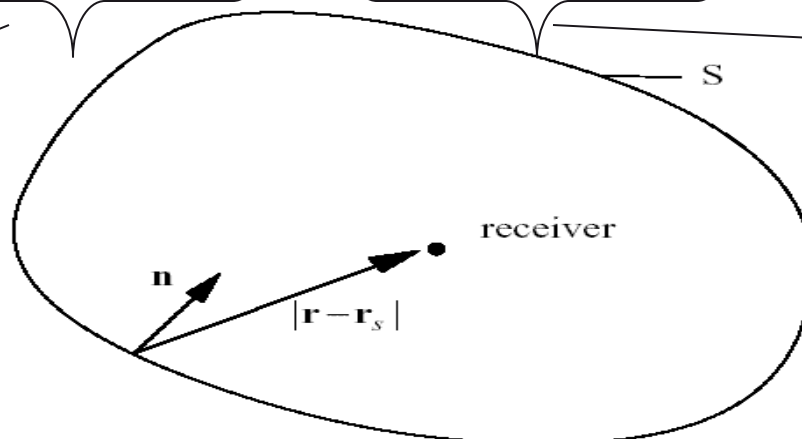
- Théorie de l'acoustique linéaire: Il est possible de recréer n'importe quel champs sonore à l'intérieur du volume délimité par une surface en plaçant des sources (monopoles et dipôles) sur cette surface (à condition qu'il n'y ait pas d'autres sources sonores à l'intérieur).

- Intégrale de Kirchoff-Helmholtz

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

**Dipôles:**  
sources reliées  
à la pression

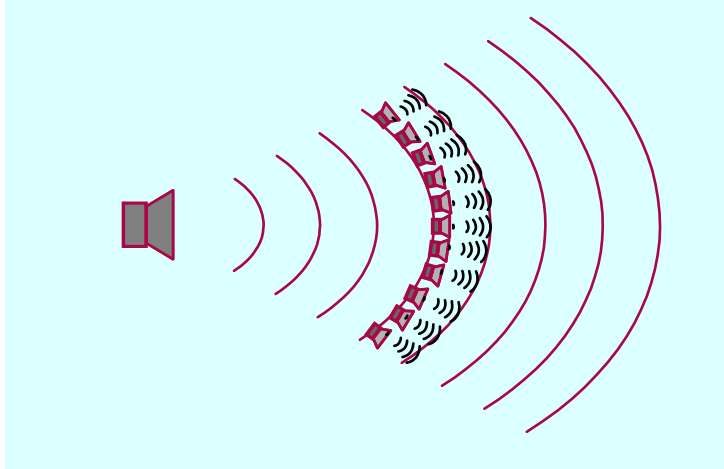
**Monopoles:**  
sources reliées à  
la vitesse du  
champs sonore



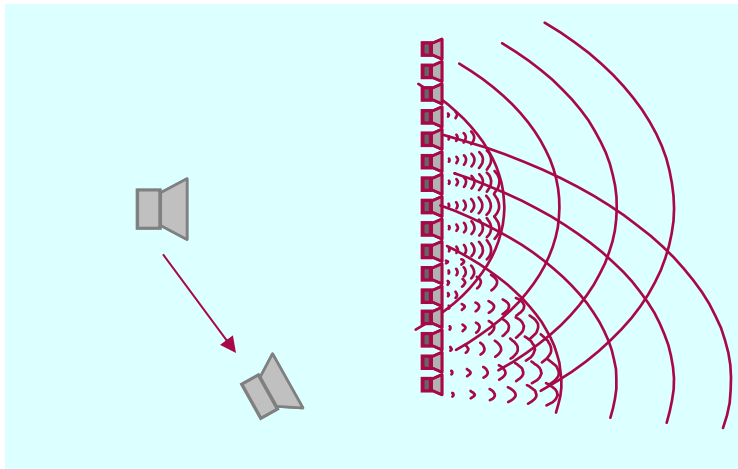
G. Richard



# Wave Field Synthesis (WFS) – Concept



- **Reproduction physique du champ sonore émis par une source primaire avec un ensemble de sources secondaires: référence périphérique**



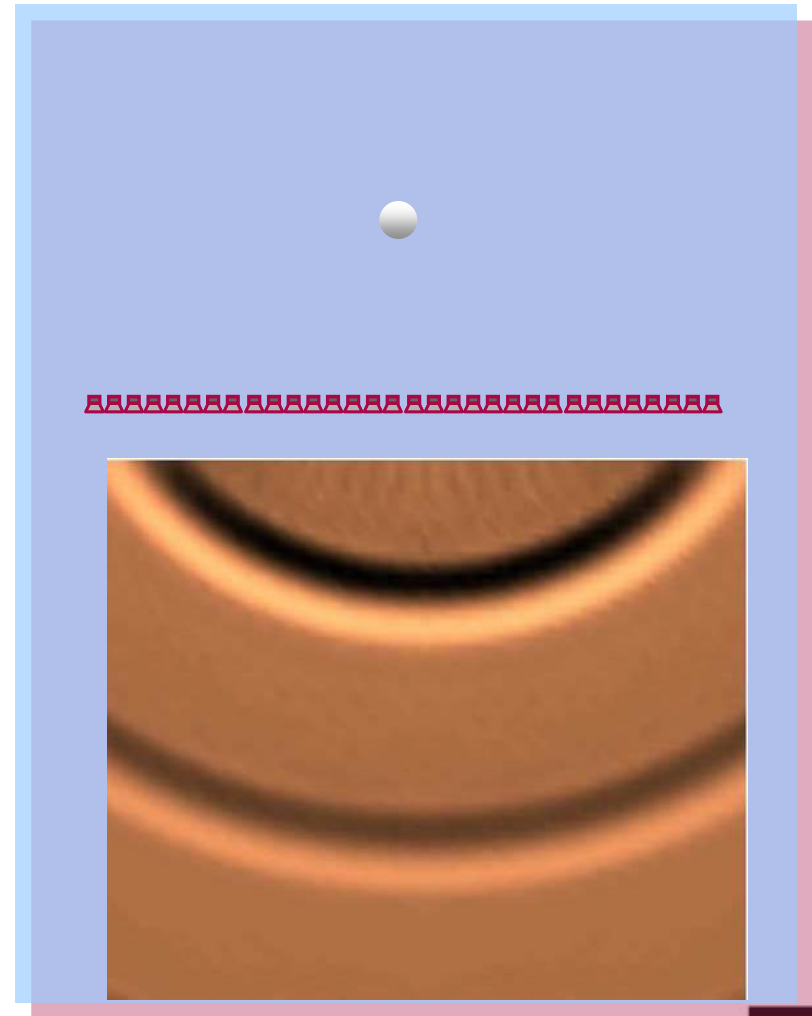
- **Relation d'impédance: la connaissance de la pression suffit.**

- **Adaptation à un réseau fini et avec une géométrie quelconque**



# Enregistrement et Reproduction en WFS

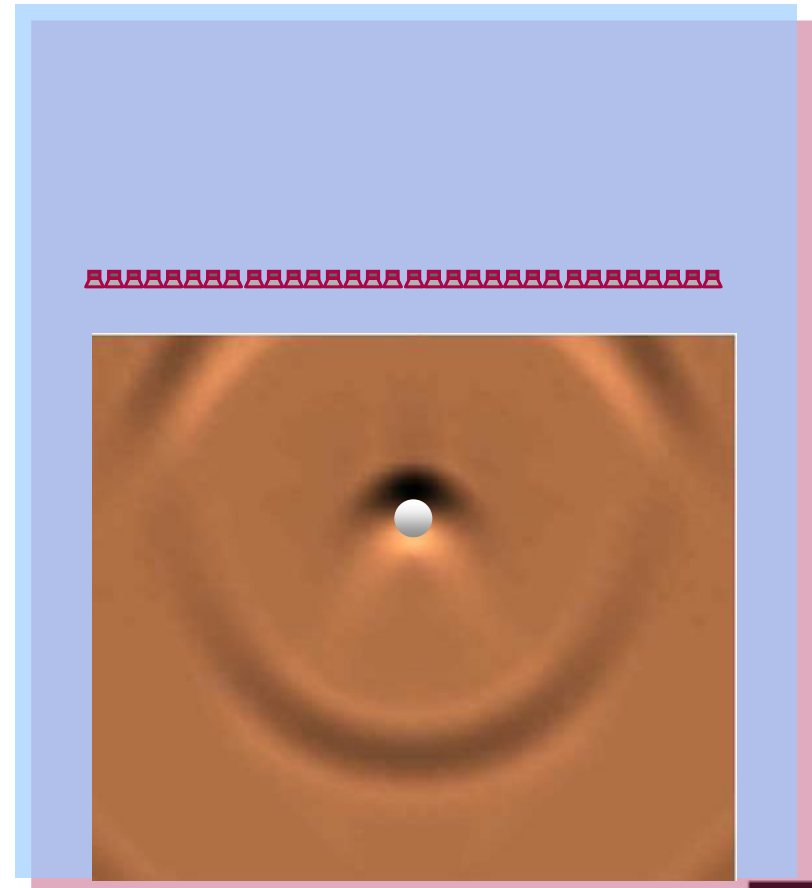
- Enregistrement
- Reproduction
- Synthèse
  - Source ponctuelle “virtuelle”.
  - Reproduction sur zone étendue.



# Reproduction en WFS

(d'après E. Corteels, cours option ATIAM)

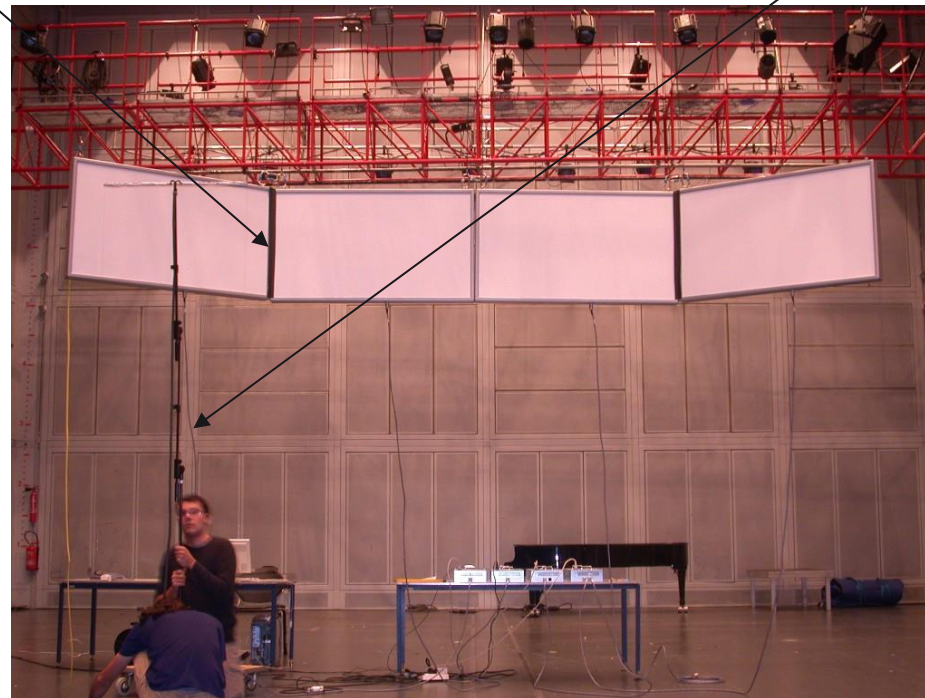
- Source ponctuelle “virtuelle”
- Eventuellement devant les haut-parleurs!
- Champ reproduit non valable entre les haut-parleurs et la source virtuelle...



# Installations: IRCAM

**Haut-Parleurs MAP « Multi  
Actuator Panel »  
(8 excitateurs par panneau)**

**Banc de  
Microphones pour  
correction de la  
salle**





# Spatialisation des sons

- Introduction
- Techniques multicanale
  - La stéréophonie.
  - Les systèmes ‘surround’, ou “multiphonie”.
  - Ambisonics (zone étendue, physique).
  - Wave-field synthesis (zone étendue, physique).

## ■ Techniques binaurales

## ■ Applications



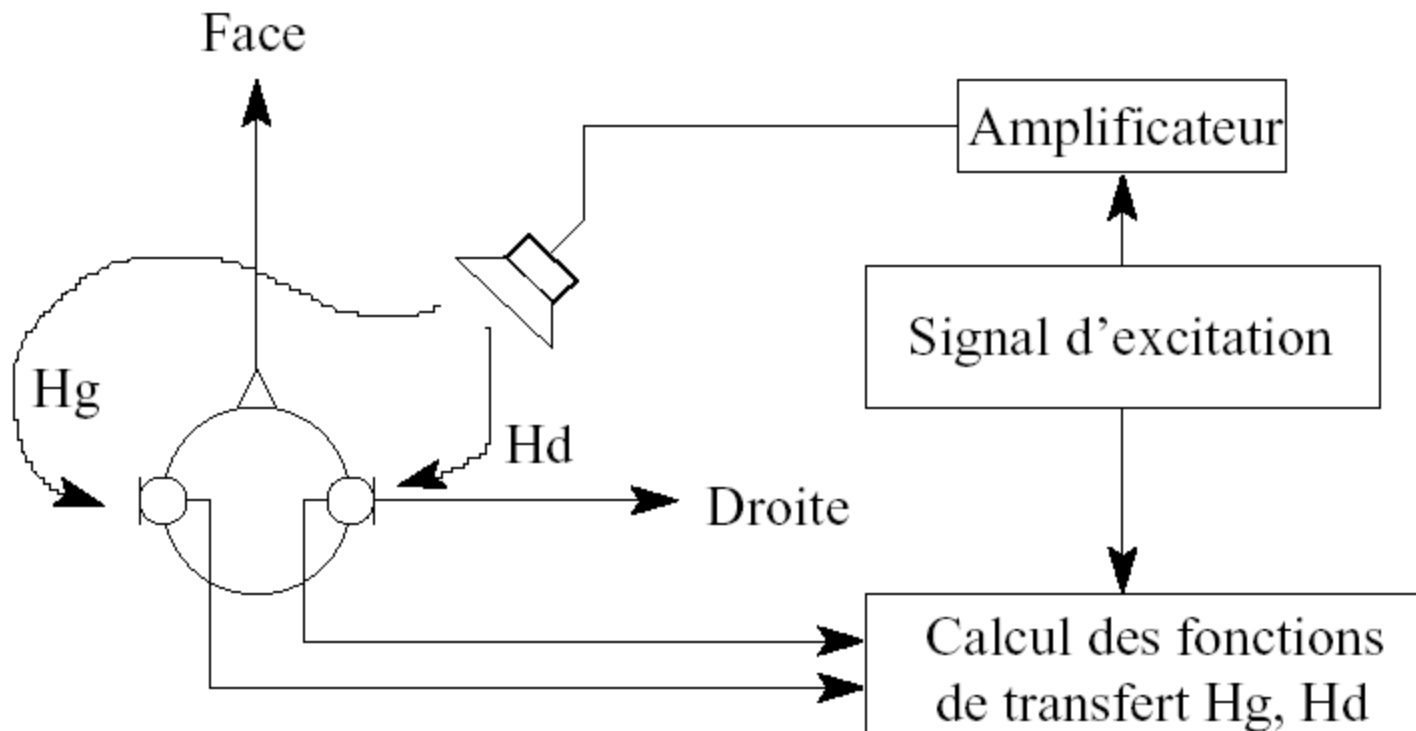
# Techniques binaurales

- **Enregistrement: enregistrer exactement le signal au tympan de l'auditeur (tête artificielle, microphones binaurales).**
- **Synthèse: à partir d'un enregistrement monophonique d'une source (anéchoïque), en utilisant des filtres binauraux et des réponses impulsionnelles, synthétiser une écoute binaurale.**
  - Pour une écoute au casque: localisation « hors de la tête »
  - Pour une écoute sur haut-parleurs (transaural): localisation en 3D (et pas seulement dans un plan situé entre les deux haut-parleurs!!)



# Synthèses binaurales

- Mesure des filtres binauraux (Head Related Transfer Functions, HRTFs)



# Mesures de réponses impulsionnelles

## ■ Les Séquences de Longueur Maximale (LMS)

- Signaux périodiques de période  $L = 2^k - 1$
- Signaux ne prenant que des valeurs instantanées  $+1$  ou  $-1$
- Leur fonction d'autocorrélation est quasiment un dirac:

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

- On vérifie alors que:

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

## ■ Sweeps Logarithmiques



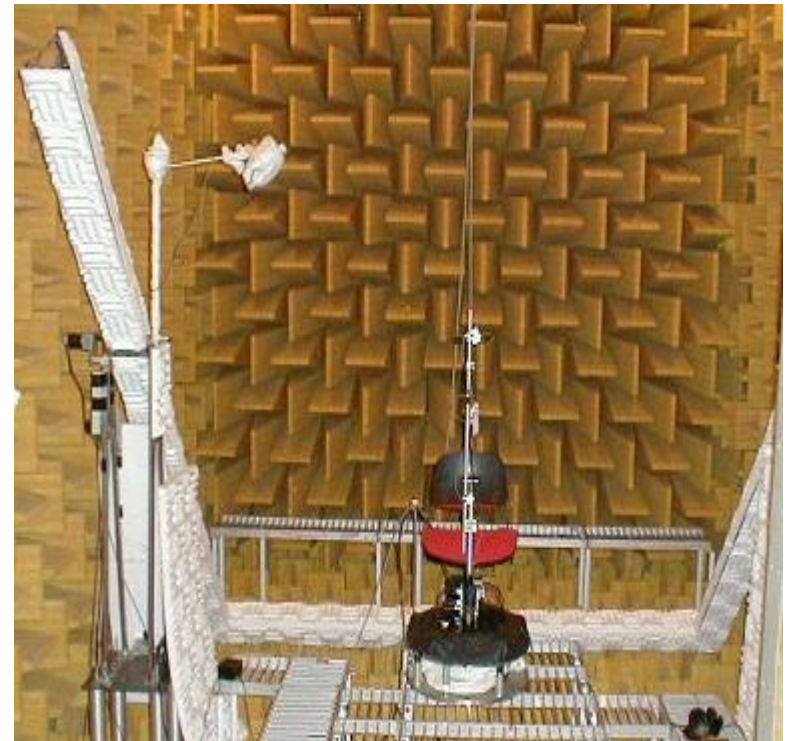
# Mesures des filtres binauraux

## ■ Mesures des HRTF à l'aide d'une tête KEMAR



# Mesures des filtres binauraux

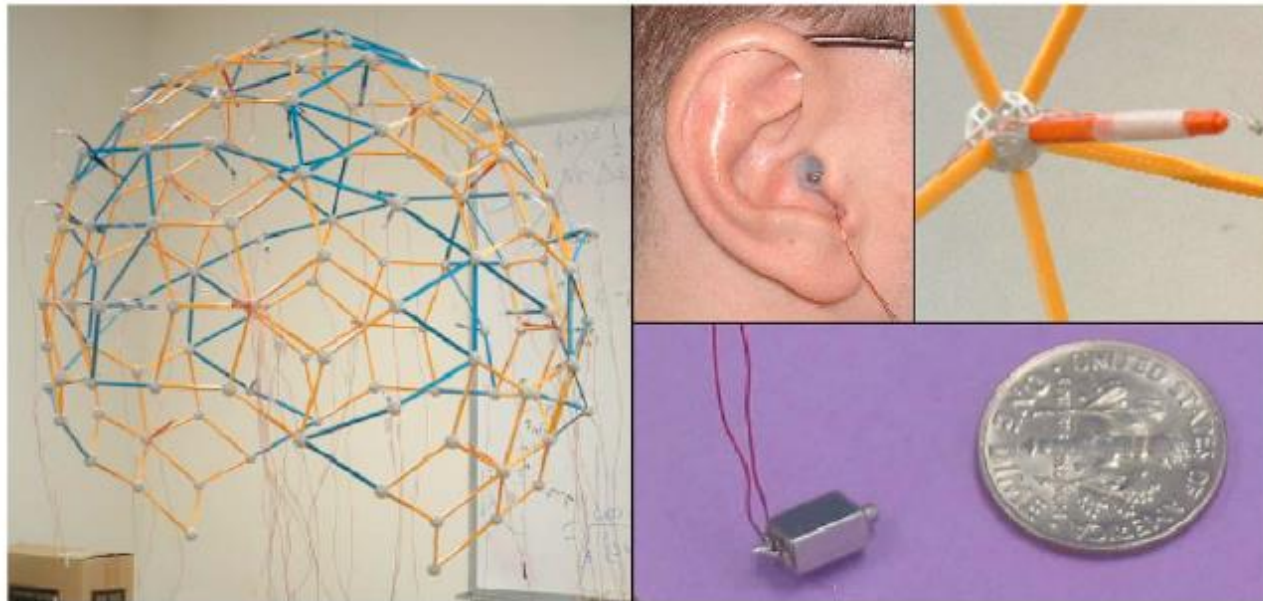
## ■ Mesures des HRTF pour des individus (ENST, IRCAM)



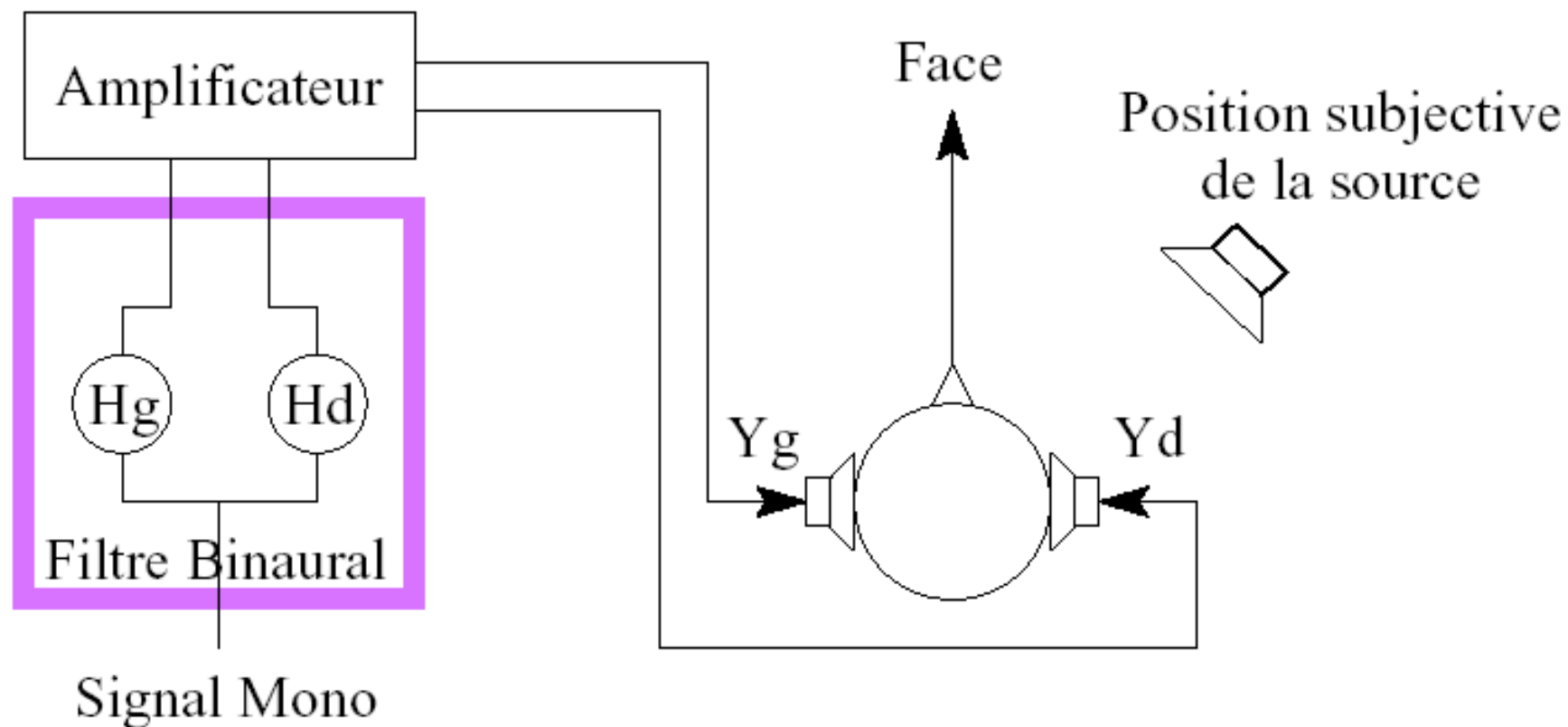


# Mesures des filtres binauraux

- Mesures des HRTF pour des individus (Maryland University)



# Synthèse binaurale







# Problématiques liées aux HRTFs

- **Un problème majeur: personnalisation des HRTFs**
- **Systemes dynamiques: interpolation des HRTfs et suivi de mouvement**
- **Auralisation binaurale**
- **Rendu de la distance**
- **Modélisation des HRTFs...**



## ■ IRCAM: LISTEN

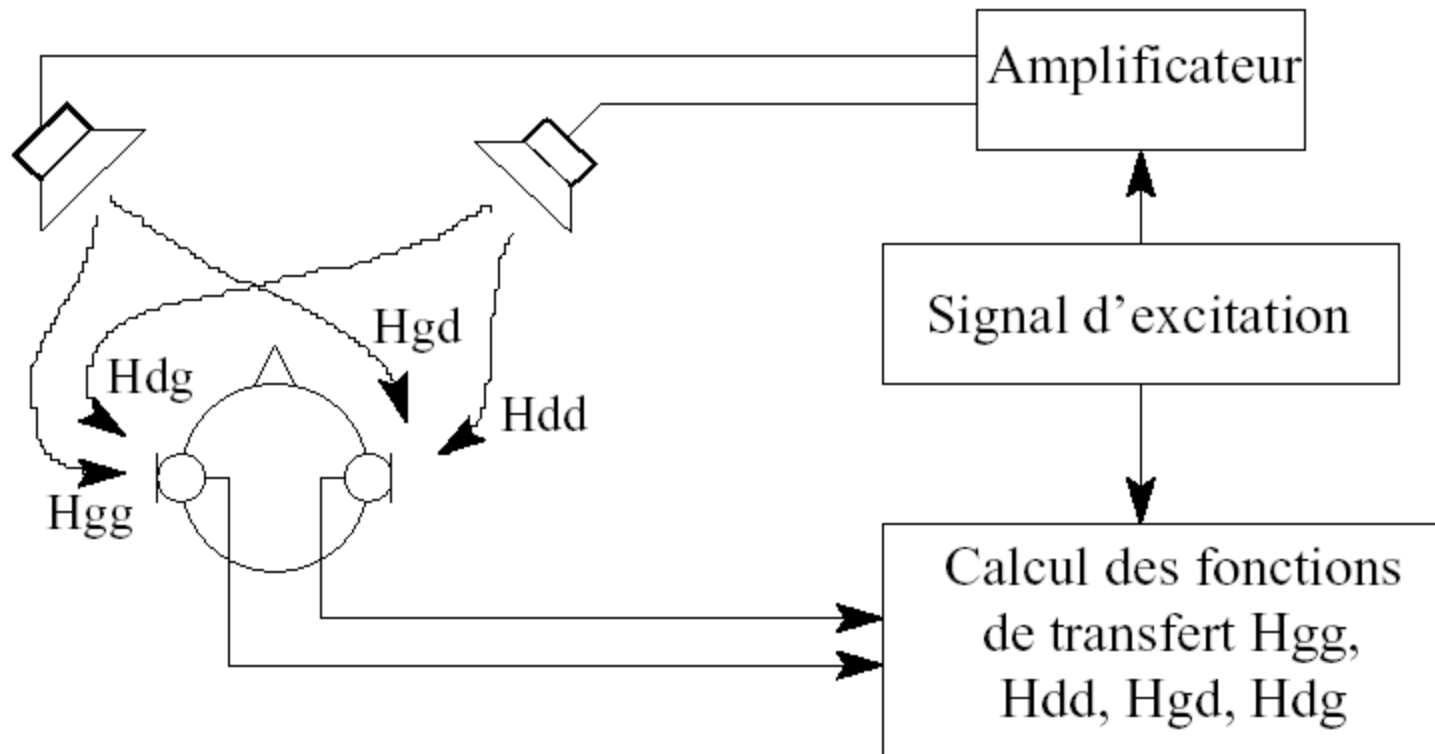
## ■ CIPIC

- HRTFs mesurées pour des valeurs discrètes de l'azimuth, de l'élévation et du temps
  - Échantillonnées à 44.1 kHz (200 échantillons)
  - 25 azimuths (compris entre  $-90^\circ$  et  $+90^\circ$ )
  - 50 élévations (compris entre  $-45^\circ$  et  $230^\circ$  par pas de  $5.625^\circ$ )
- HRTFs mesurées pour plus de 90 personnes (l'ensemble des HRTFs de 45 personnes constitue une base publique!)



# Conversion binaural $\Leftrightarrow$ transaural

- Nécessite la mesure des 2 fonctions de transfert binaurales correspondant aux positions des haut-parleurs et de l'auditeur dans le lieu d'écoute



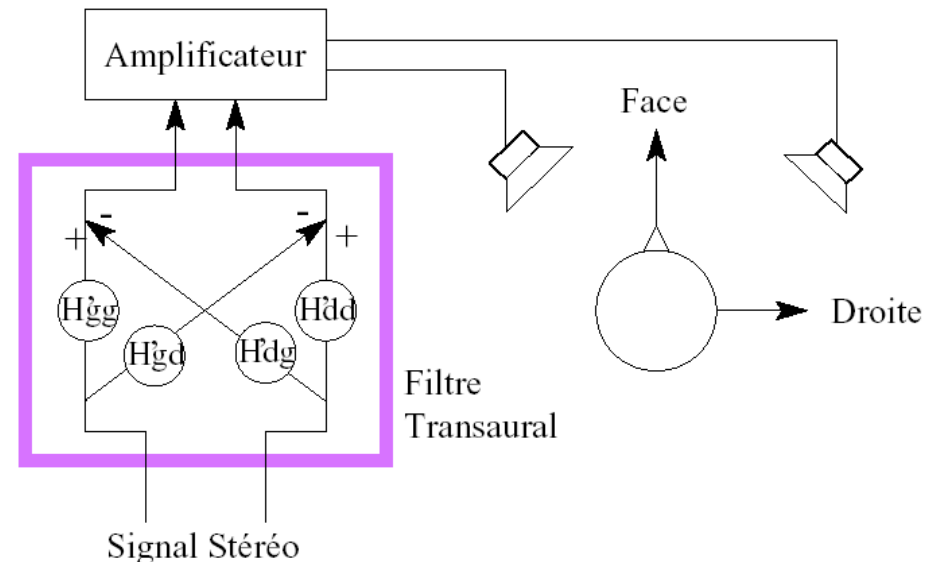
# Conversion binaural $\Leftrightarrow$ transaural

- Matrice de transfert caractérisant la transformation des signaux émis par les hauts parleurs vers les signaux reçus par les conduits auditifs ( $Y_g$  et  $Y_d$ )

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

- La conversion binaural  $\Rightarrow$  transaural est la matrice de transfert inverse

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



# Conversion binaural $\Leftrightarrow$ transaural

## ■ Hypothèses de symétrie du dispositif de reproduction et des conduits auditifs

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

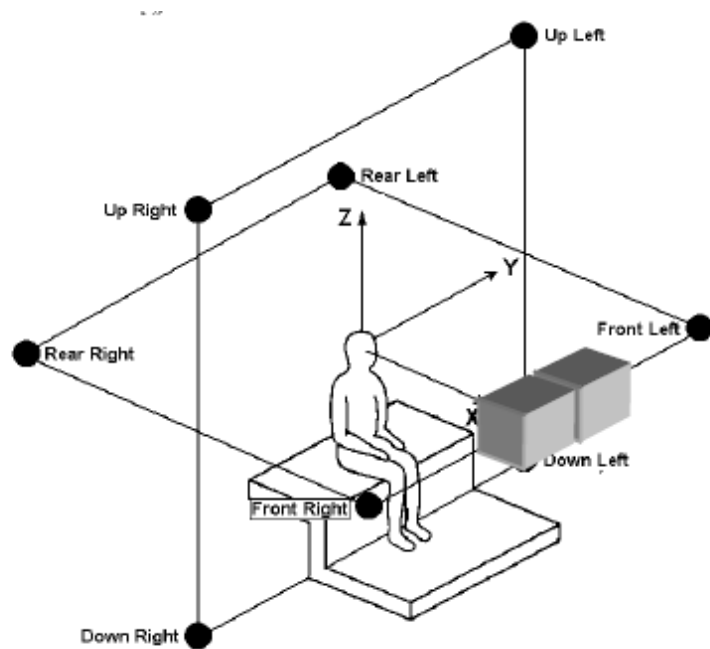
- ✓ Fonctions de transfert binaurales
  - ✓ ***G et D*** : pour une position quelconque de la source
  - ✓ ***G<sub>0</sub> et D<sub>0</sub>*** : pour la position du Haut-Parleur gauche

### ✓ On obtient alors:

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



# Un système transaural: Ambiophonics



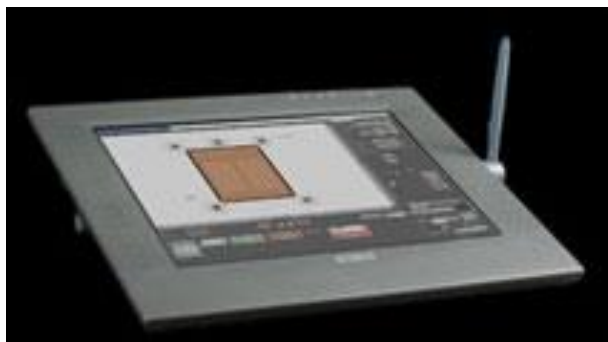
# Spatialisation des sons

- Introduction
- Techniques multicanale
  - La stéréophonie.
  - Les systèmes ‘surround’, ou “multiphonie”.
  - Ambisonics (zone étendue, physique).
  - Wave-field synthesis (zone étendue, physique).
- Techniques binaurales

## ■ Applications



# Entreprises: IOSONO





# Applications existantes...

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



# Applications existantes...

- Ambisonics (<http://www.ambisonic.net/>): utilisé en ingénierie du son...micro Soundfield

- Dolby Headphones

..ou presque

- MPEG Surround

- Binaural Mobile

- Binaural iPod

- Spatial music production



# Organismes impliqués ..en France

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

