

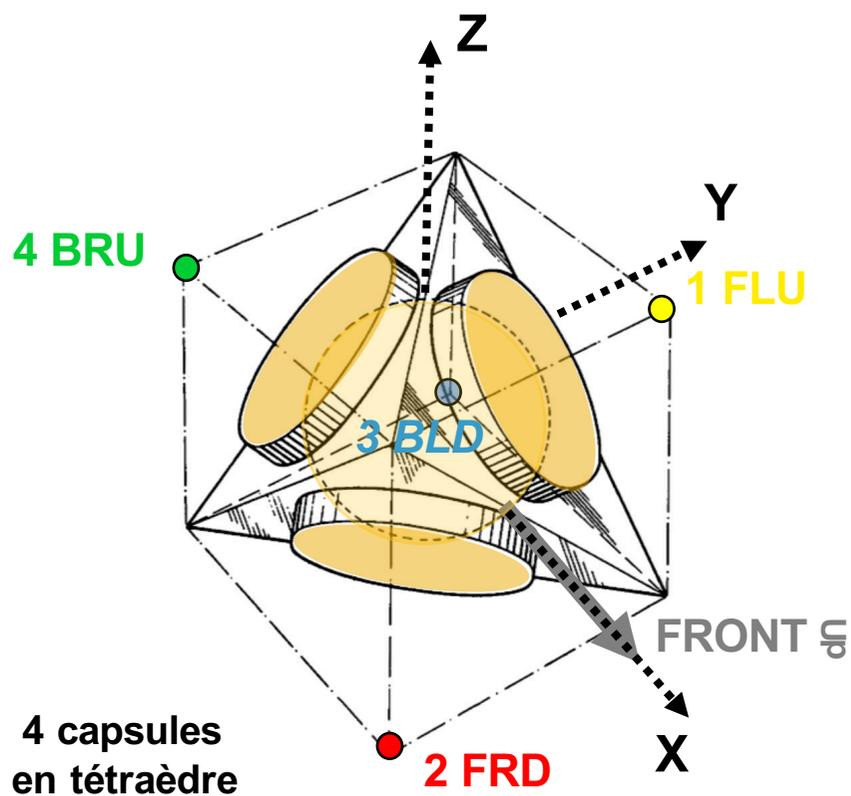


SENNHEISER



Photo SENNHEISER

# Sennheiser Ambéo<sup>®</sup> VR mic



Bernard Lagnel

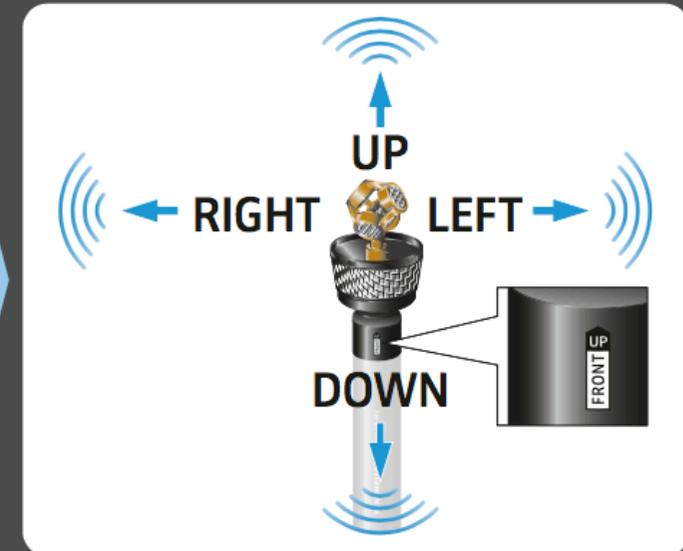
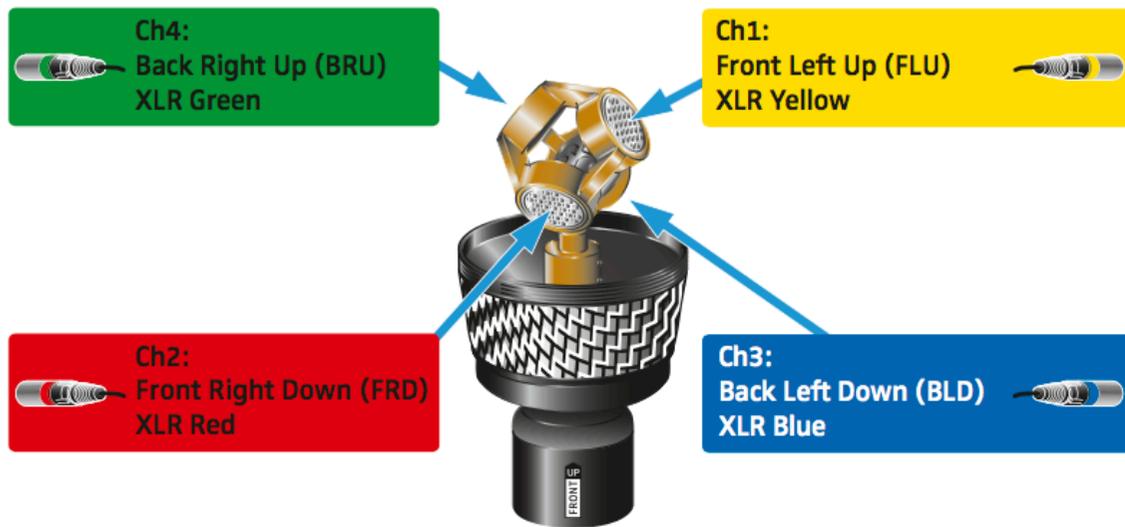
Mars 2017

# AMBEO VR MIC

Quick Guide | Kurzanleitung | Guide rapide | Guida rapida | Guía de inicio rápido | Guia rápido | Beknopte handleiding | Krótka instrukcja obsługi | Snabbguide | Kort vejledning | Pikaohje | Γρήγορος οδηγός | Краткая инструкция | 快速指南 | 快速使用指南 | クイックガイド | 요약 설명서



≈ 1500 €

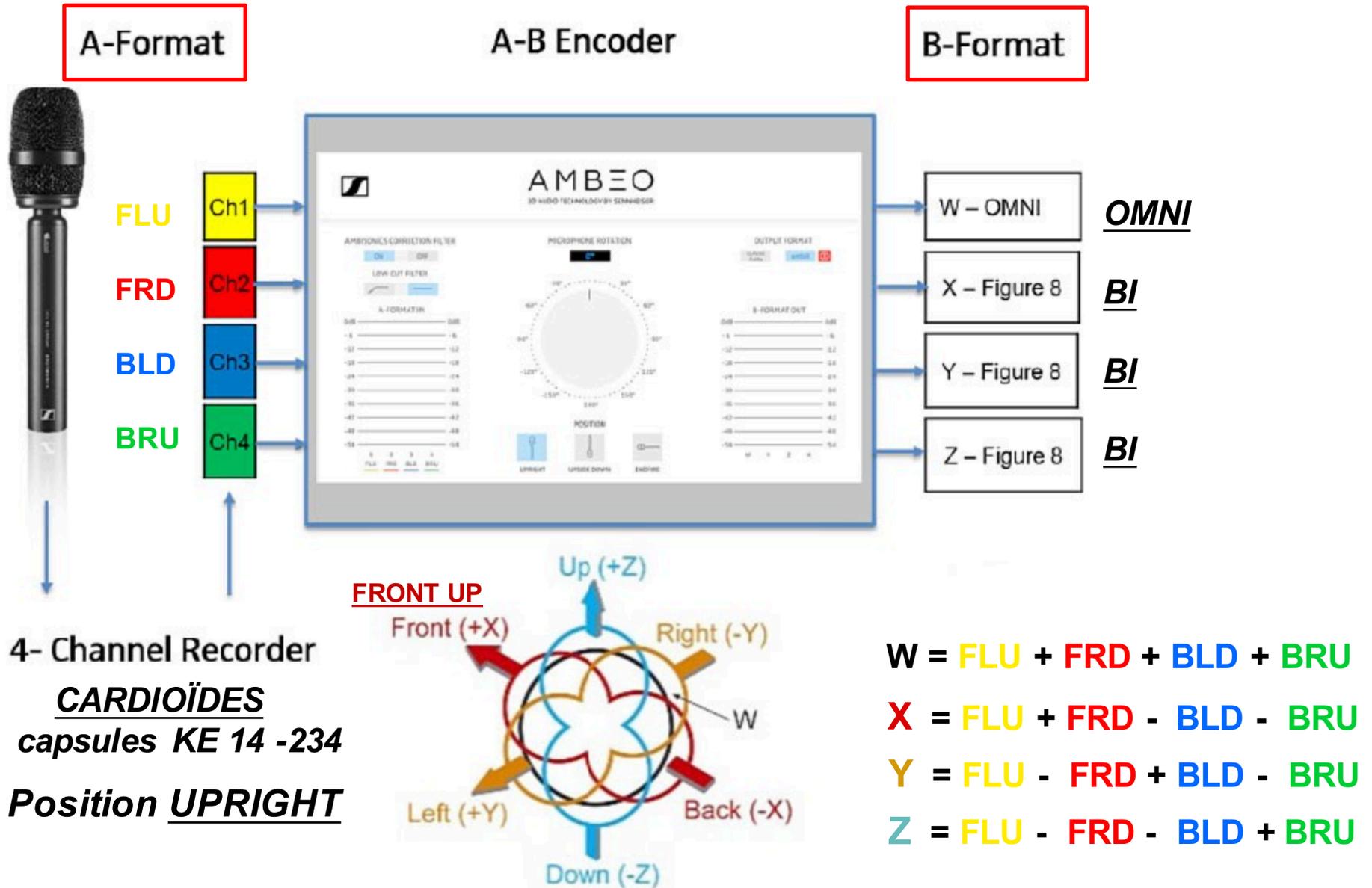


***Position UPRIGHT***

***Configuration recommandée***

# RECORDING WORKFLOW

Plug-in convertisseur de format A-B AMBEO spécialement conçu par Sennheiser, téléchargeable gratuitement en format VST, AU ou AAX.



4- Channel Recorder  
CARDIOÏDES  
 capsules KE 14 -234  
 Position UPRIGHT

FuMa : WXYZ (Furse - Malham)  
ambiX : WYZX (youtube 360)

**AMBISONICS CORRECTION FILTER**

ON  OFF

**LOW CUT FILTER**

**A-FORMAT IN**

0dB -6 -12 -18 -24 -30 -36 -42 -48 -54

1 2 3 4

FLU FRD BLD BRU

**MICROPHONE ROTATION**

0°

-30° 30°

-60° 60°

-90° 90°

-120° 120°

-150° 150°

180°

**POSITION**

UPRIGHT  UPSIDE DOWN  ENDFIRE

**OUTPUT FORMAT**

CLASSIC FuMa  ambiX

**B-FORMAT OUT**

0dB -6 -12 -18 -24 -30 -36 -42 -48 -54

W X Y Z

**Position par défaut :**

### [Ambisonics Correction Filter ▼](#)

Use this filter to optimize the recording for the 360° sound image.

This filter is activated as default.

### [Low Cut Filter ▶](#)

### [Microphone Rotation ▼](#)

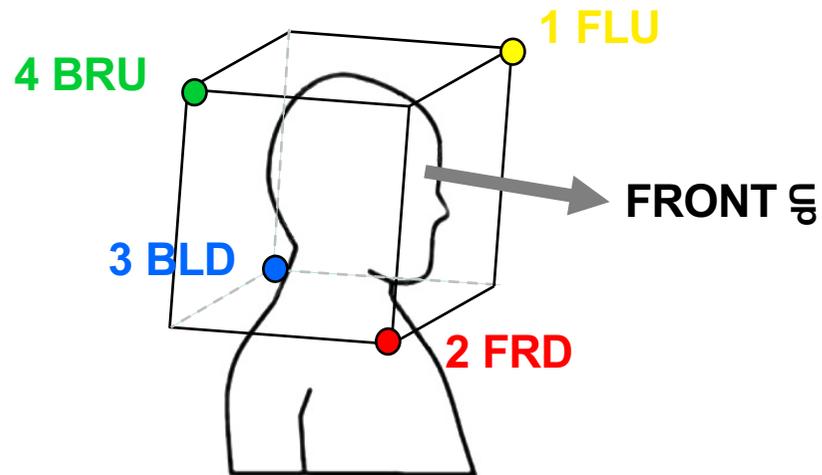
Use the rotary switch to correct the horizontal orientation of the microphone, if you e.g. have to later align the direction of sound-pick-up with the video recording.

The default setting of the rotary switch is 0°.

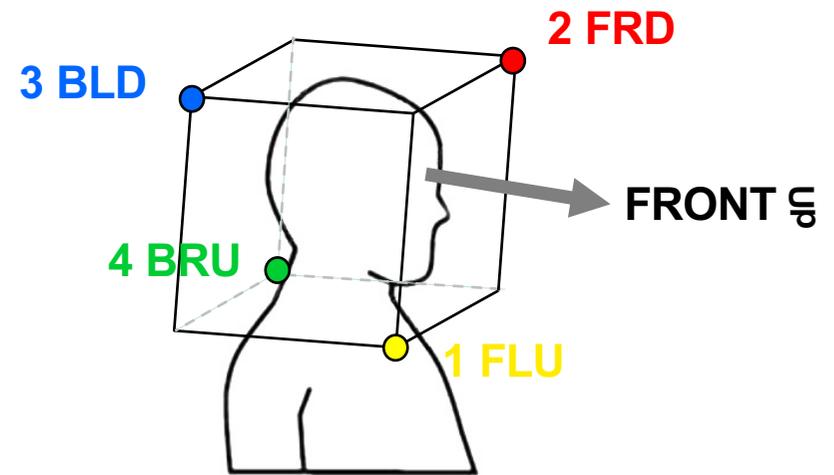
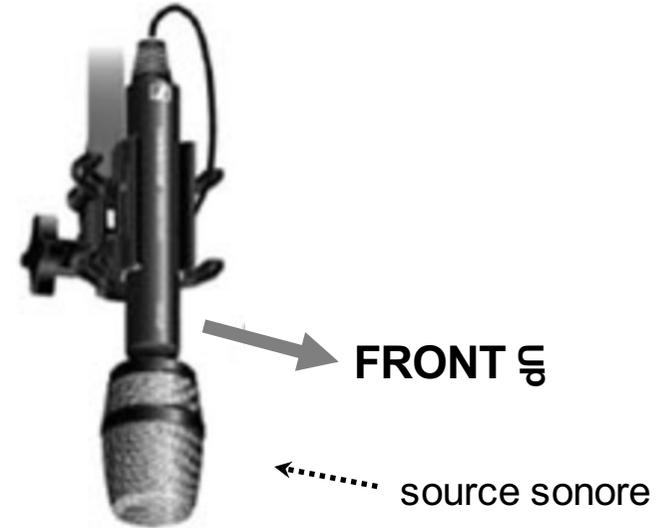
**AMBISONICS CORRECTION FILTER, propre à l'AMBÉO VR mic (correction non linéaire des capsules et correction de non coïncidence)**

# Sennheiser Ambéo® VR mic

Position UPRIGHT

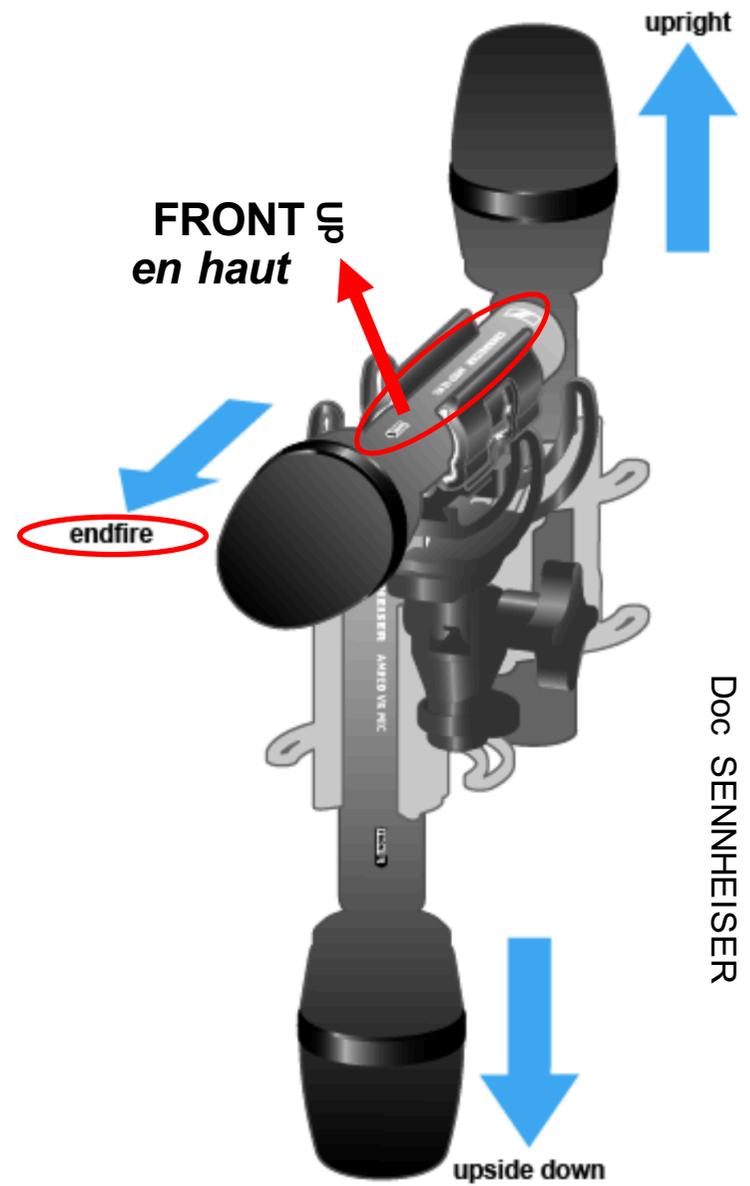
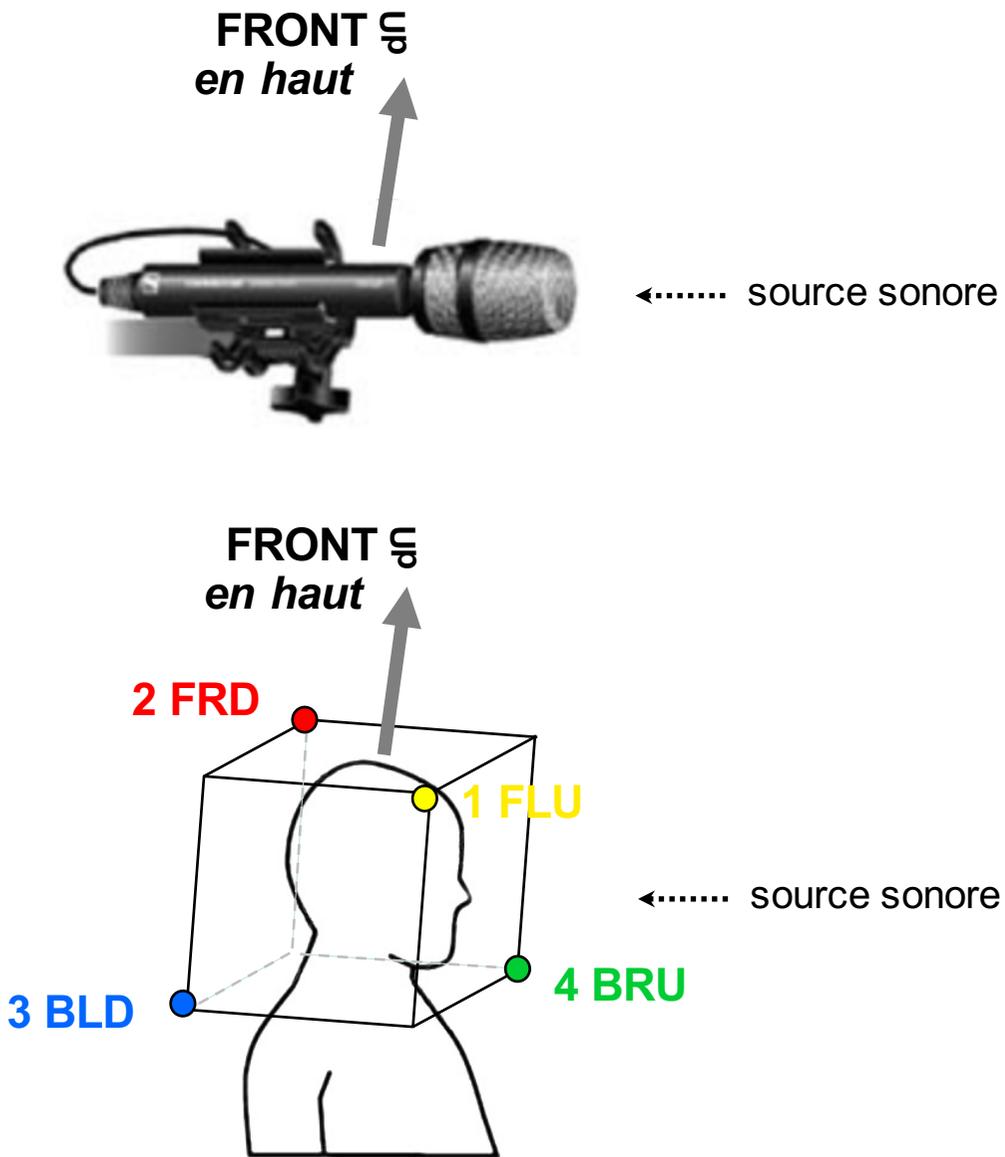


Position UPSIDE DOWN

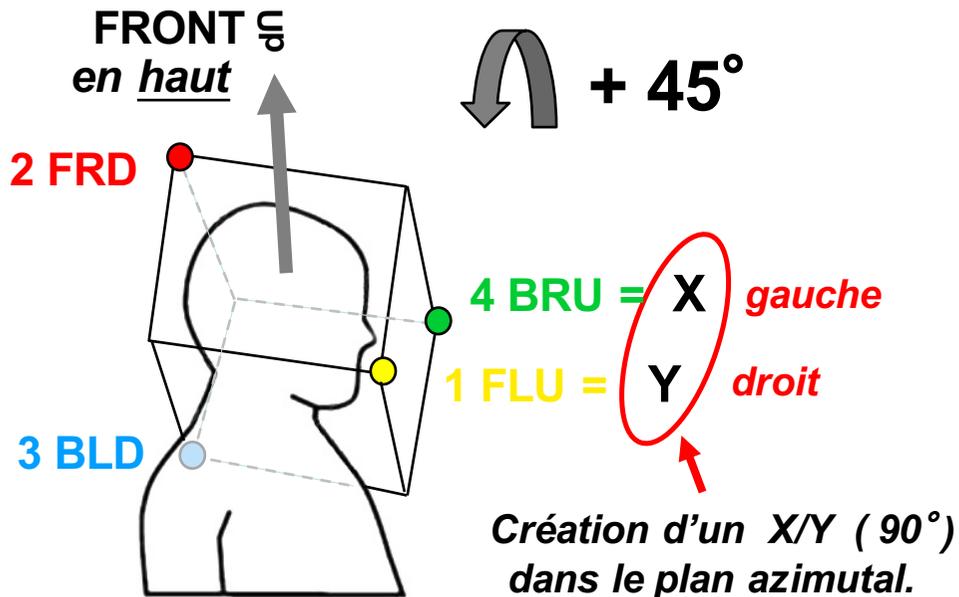
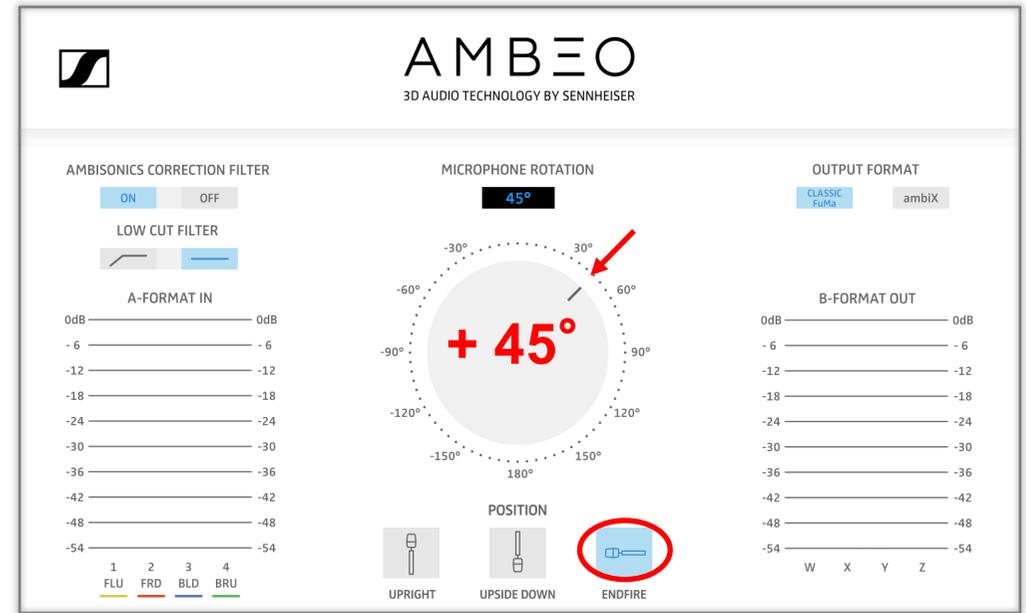
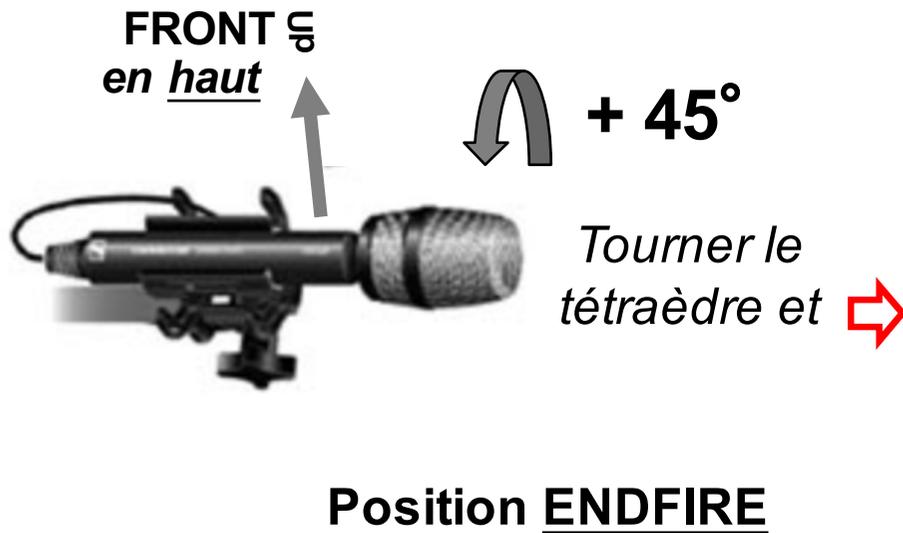


# Sennheiser Ambéo® VR mic

## Position ENDFIRE



# Sennheiser Ambéo® VR mic

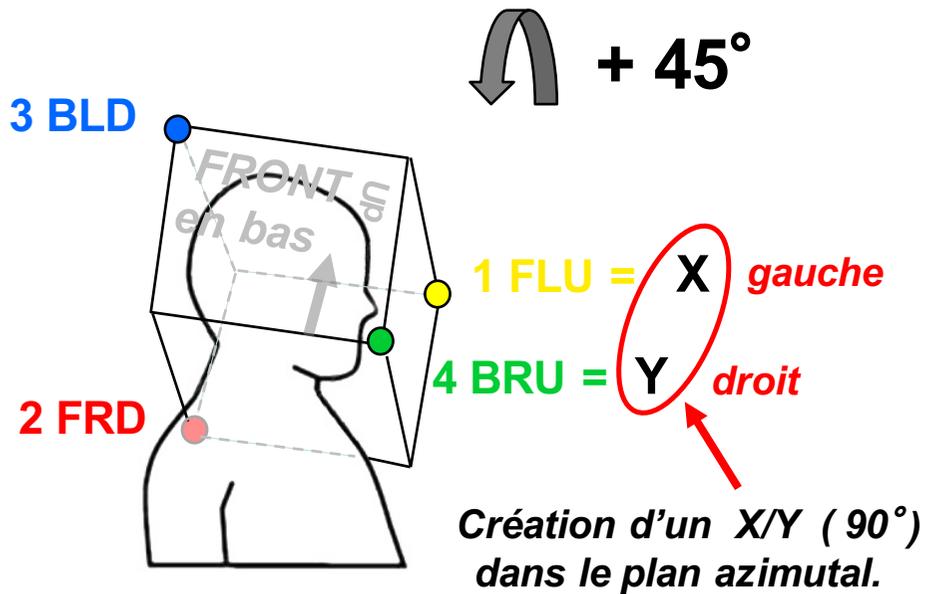
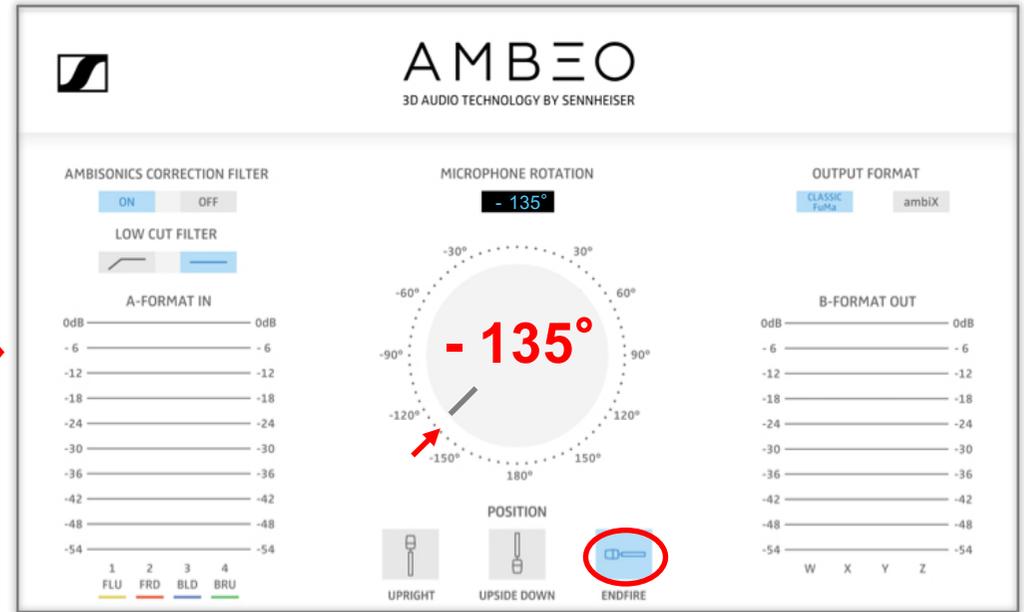


Création d'un XY à partir du VR mic afin d'écouter les pistes 4 (gauche) et 1 (droite) couplées aux DPA 4060 dans mes oreilles (reportage sur le terrain).

# Sennheiser Ambéo® VR mic

Configuration possible si on est habitué au Soundfield...

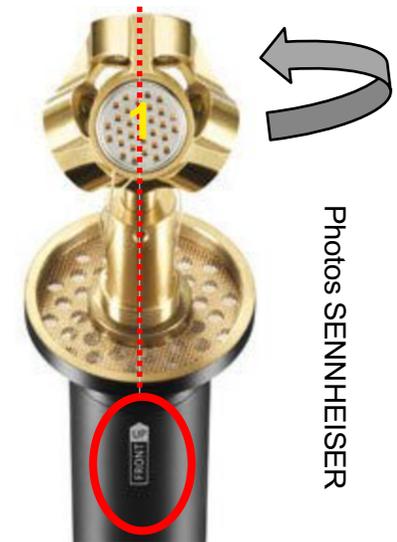
## Position ENDFIRE



Configuration normale :



Capsules tournées de -45° ?



## Caractéristiques du couple stéréophonique :

Copyright © 2008 Bernard Lagnel

\* Directivité  
des micros L et R

0,500

Angle entre  
les micros L et R

90°

Distance entre  
les micros L et R

0 cm

\* Directivité après la  
SOMMATION de L et R  
( signaux en phases )

0,586

Distance de la  
source sonore

10,0 m

Pourcentage en niveau  $\Delta L$  et en temps  $\Delta T$   
( entre les micros L et R )

$\Delta L$  dB

$\Delta T$  ms

100 %

0 %

Affaiblissements  
à l'avant 0°  
du couple

-1,4 dB

Affaiblissements  
à l'arrière 180°  
du couple

-16,7 dB

Après SOMMATION :  
coefficient de directivité  
du couple Q  
( réf du Cardio :  $Q = 3$  )

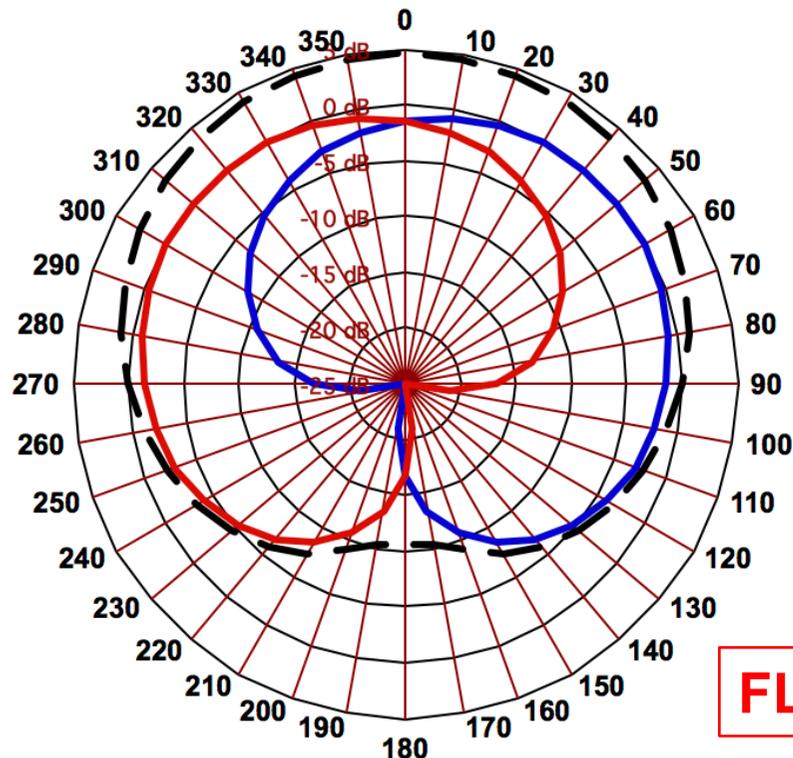
2,5

Rapport de capture  
ou Facteur de Distance =  $\sqrt{Q}$

1,6

Angle total de  
prise de son utile  
du couple

150°



### \* NOTE :

Micro OMNI = 1

Micro INFRA  $\approx 0,66$  ( -10 dB arrière )

Micro CARDIO = 0,5

Micro SUPER  $\approx 0,375$  ( -12 dB arrière )

Micro BI = 0

LES LIENS :

[SYSTÈME MMAD \( WILLIAMS -LE DÛ\)](#)

[hauptmikrofon Image Assistant 2.0](#)

**FLU + FRD**

**Matriçage**

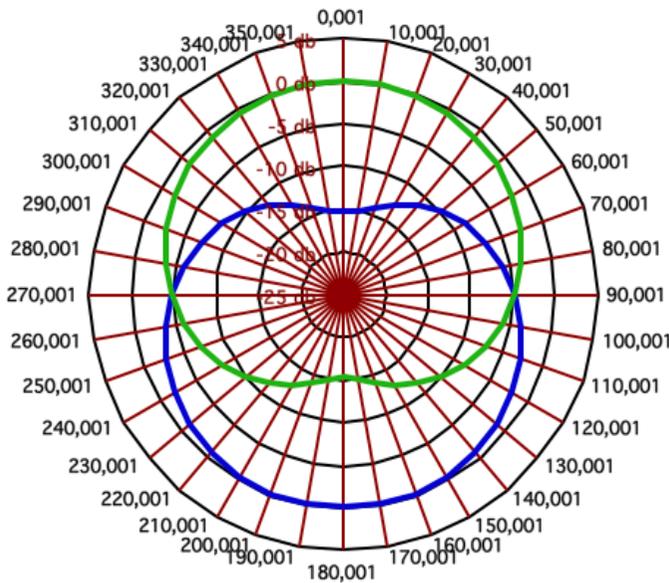
* Caractéristique du micro <b>FRONTAL</b>	<b>0,586</b>
* Caractéristique du micro <b>DORSAL</b>	<b>0,586</b>

Différence de sensibilité entre le micro <b>FRONTAL</b> et le micro <b>DORSAL</b>	
<b>0,0 dB</b>	

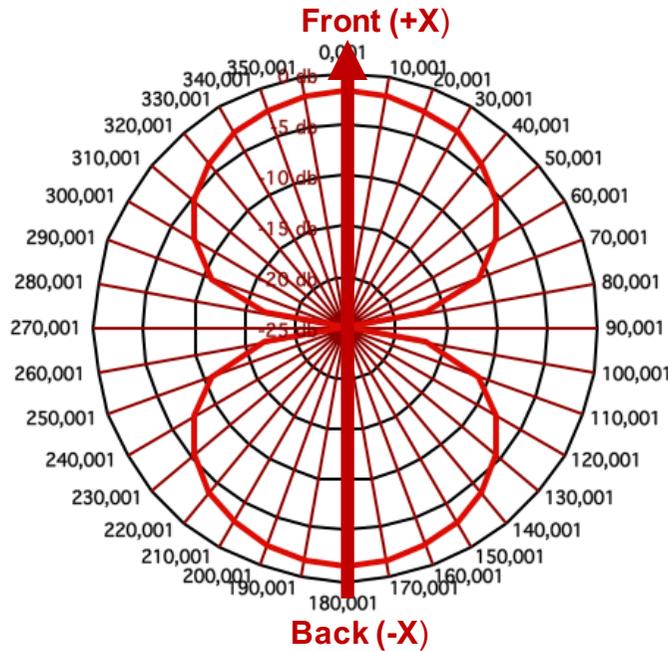
RÉSULTATS DE LA SOMMATION DES 2 MICROS VISANT DANS DES DIRECTIONS OPPOSÉES	
* Caractéristique de directivité du micro après sommation	Niveau maximum du micro après sommation
<b>0,000</b>	<b>-1,6 dB</b>

\* NOTE :  
 Micro OMNI = 1  
 Micro INFRA ≈ 0,660 ( -10 dB arrière)  
 Micro CARDIO = 0,5  
 Micro SUPER ≈ 0,375 ( -12 dB arrière)  
 Micro BI = 0

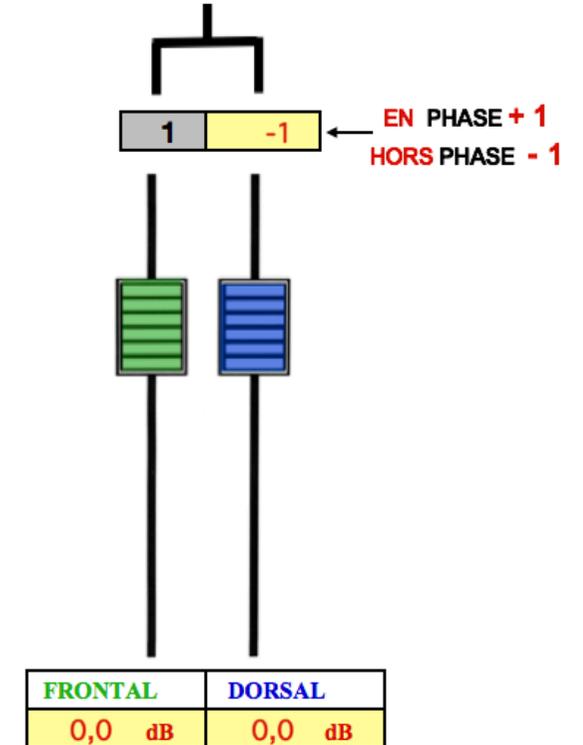
AVANT SOMMATION



APRÈS SOMMATION



SOMMATION



Copyright © 2008 Bernard Lagnel

$$X = FLU + FRD - BLD - BRU$$

*Matriçage*

## Caractéristiques du couple stéréophonique :

Copyright © 2008 Bernard Lagnel

\* Directivité  
des micros L et R

0,000

Angle entre  
les micros L et R

90 °

Distance entre  
les micros L et R

0 cm

\* Directivité après la  
SOMMATION de L et R  
( signaux en phases )

1,000

Distance de la  
source sonore

10,0 m

Pourcentage en niveau  $\Delta L$  et en temps  $\Delta T$   
( entre les micros L et R )

$\Delta L$  dB

$\Delta T$  ms

100 %

0 %

Affaiblissements  
à l' avant 0°  
du couple

Affaiblissements  
à l' arrière 180°  
du couple

-3,0 dB

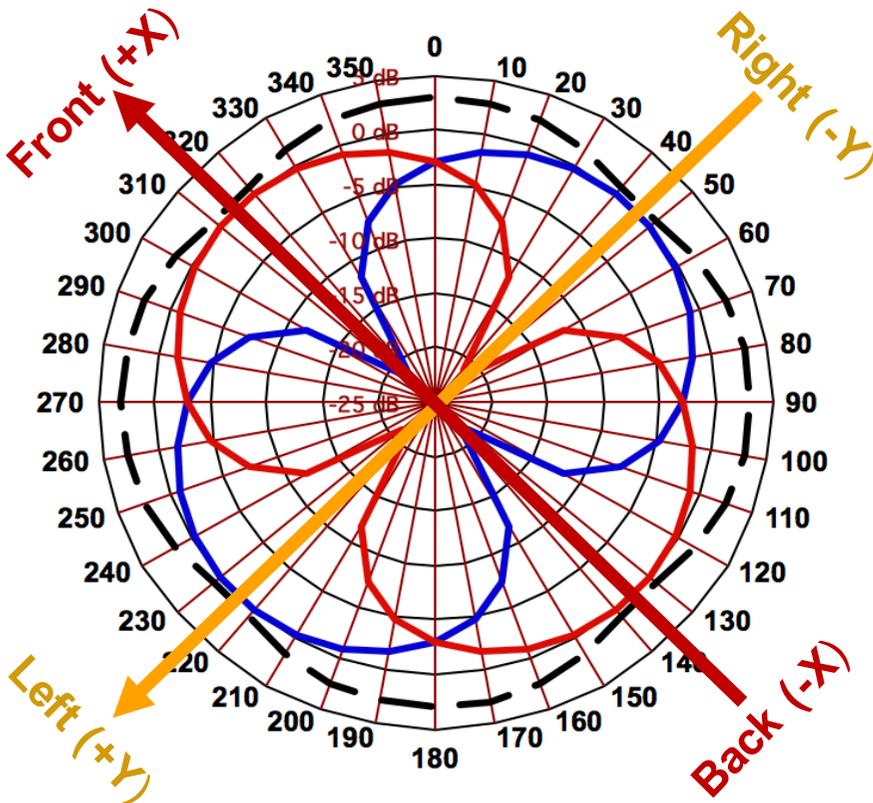
-3,0 dB

Après SOMMATION :  
coefficient de directivité  
du couple Q  
( réf du Cardio : Q = 3 )

1,2

Rapport de capture  
ou Facteur de Distance =  $\sqrt{Q}$

1,1



**FRONT UP**

Front (+X)

Up (+Z)

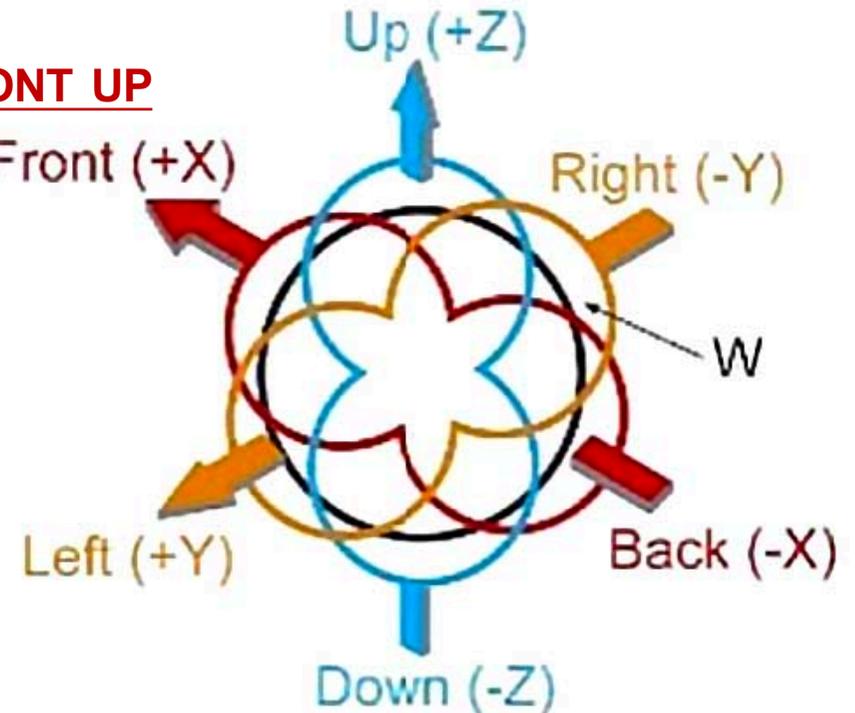
Right (-Y)

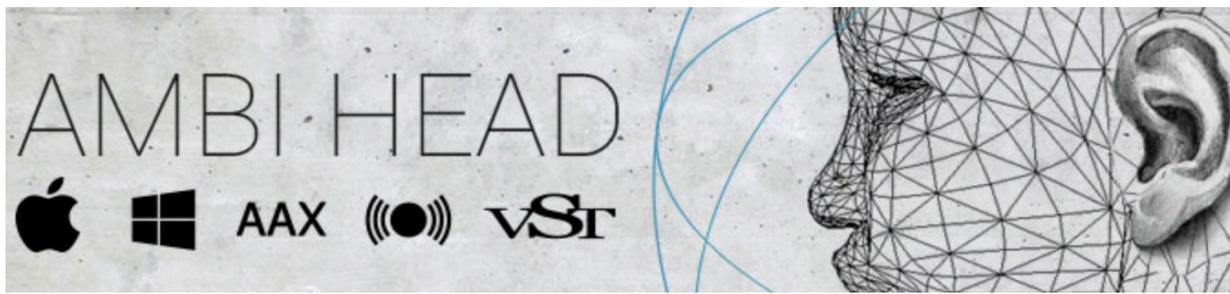
Left (+Y)

Down (-Z)

Back (-X)

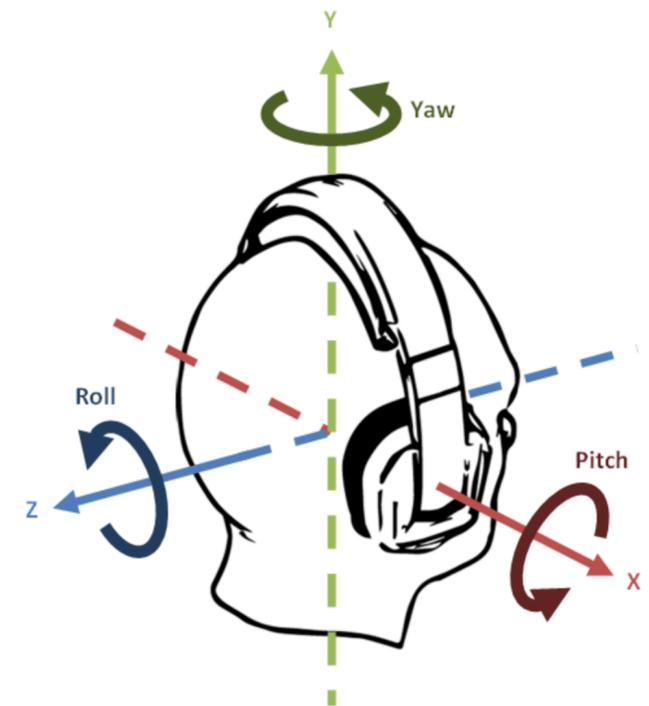
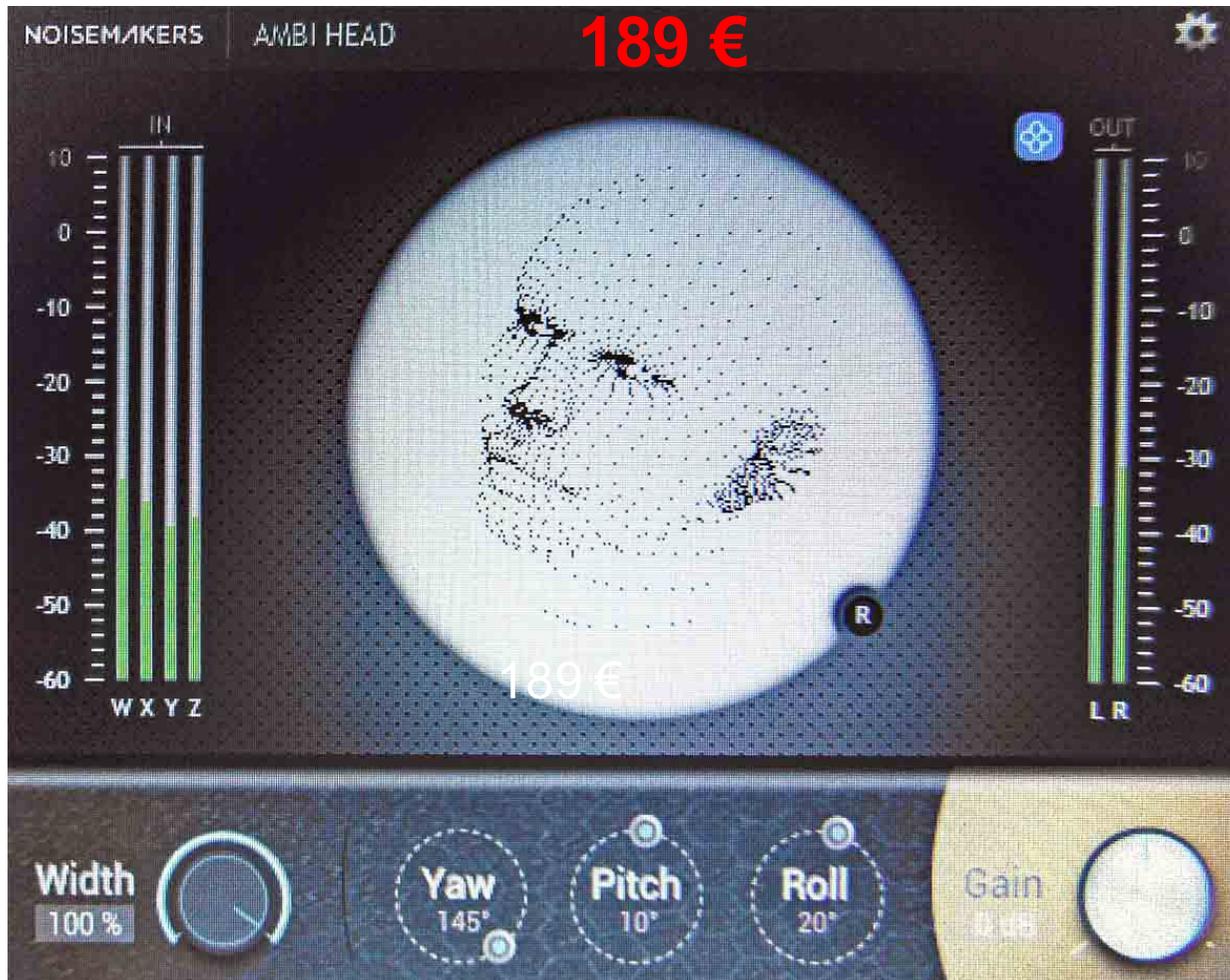
W





Ambi Head de [NOISE MAKERS](http://www.noisemakers.fr/) est un plugin professionnel pour convertir des signaux ambisoniques en audio binaural 3D. Il permet des manipulations de scènes 3D (rotations et contrôle de la largeur spatiale) et contient des filtres HRTF spécialement conçus pour une écoute immersive et une reproduction précise des signaux en format B.

Les filtres HRTF personnalisés peuvent être chargés via l'importateur [SOFA](http://www.sofaconventions.org/) intégré, permettant aux utilisateurs de choisir des "têtes personnalisées" pour un rendu binaural personnalisé.



<http://www.noisemakers.fr/>

[https:// www.sofaconventions.org/mediawiki/index.php/Files](https://www.sofaconventions.org/mediawiki/index.php/Files)

# Caractéristiques techniques du VR mic

Dimensions	215 x 49/25 mm (L x D)
Réponse en fréquence	20 Hz to 20 kHz
Niveau maximal de pression sonore (SPL)	130 dB(A) for 1 kHz
Type de transducteur	Pre-polarized condenser microphone
Poids	<b>Le VR mic est 2 fois plus lourd qu'un XY Schoeps CMXY 4V</b> <b>400 g</b>
Connecteur de microphone	DIN12M, use enclosed adapter cable to convert to 4x XLR3M Polarity: Pin1 (GND), Pin2 (+), Pin3 (-)
Directivité	4x cardioid, in A-format arrangement (1: front-left-up [FLU], 2: front-right-down [FRD], 3: back-left-down [BLD], 4: back-right-up [BRU])
Impédance nominale	Approx. 200 $\Omega$
Impédance de charge minimale	1000 $\Omega$
Température d'entreposage	-20 °C to 70 °C (-4 °F to 158 °F)
Température de service	-10 °C to 55 °C (14 °F to 131 °F)
Niveau de bruit équivalent	18 dB-A as per IEC 61672-1 / 27 dB CCIR as per CCIR 468-3
Alimentation électrique	4x phantom powering (P48) as per IEC 61938 (48 $\pm$ 4 V), 3.5 mA each
Sensibilité	<b>Le VR mic est 8 dB plus sensible qu'un XY Schoeps CMXY 4V</b> <b>31 mV/Pa (-30 dBV)</b>
Humidité relative de l'air Fonctionnement	Max 90% at 40 °C
Matrix reference	Center of the capsule holder

**Annexes :**

# Capsules KE 14-234 utilisées aussi dans le micro : e 914

## e 914 High grade pre-polarized condenser microphone for ambitious recordings and live performances

≈ 300 €

### FEATURES

- Exceptionally lively, clear sound
- Three-position sensitivity switch
- Three-position bass roll-off/cut-off switch
- Very high sound pressure handling capability
- Exchangeable capsule
- Excellent directivity across the whole frequency range
- Wide frequency response

The **e 914** is a high grade condenser microphone for very demanding applications. Its outstanding sound properties qualify the e 914 for highly sophisticated tasks. The e 914 is a perfect microphone for ambitious recordings and live performances. Its main areas of application are acoustic guitars, cymbals, percussion, overhead, orchestras, grand pianos etc.



### ARCHITECT'S SPECIFICATIONS

The microphone shall be a cardioid pre-polarized condenser designed for ambitious recordings and live performances.

It shall be fitted with a three-position sensitivity switch (0 dB, -10 dB, -20 dB) and a three-position bass roll-off/cut-off switch (roll-off: 130 Hz with 6 dB/octave, cut-off: 85 Hz with 18 dB/octave). The frequency response shall be 20 Hz-20,000 Hz and the sensitivity (free field, no load) shall be 7 mV/P at 1 kHz. Nominal impedance shall be 100 Ω, with a min. terminating impedance of 1 kΩ. Equivalent noise level shall be 19 dB(A). Maximum SPL at 1 kHz shall be 137 dB with 0 dB preattenuation, 147 dB with -10 dB preattenuation or 157 dB with -20 dB preattenuation.

The microphone shall operate on 48 V phantom power and shall provide a 3-pin XLR connector. Dimensions shall be 24 x 157 mm (0.94" x 6.18"). Weight shall be 198 grams (6.98 oz). The microphone shall be the Sennheiser e 914.

### TECHNICAL DATA

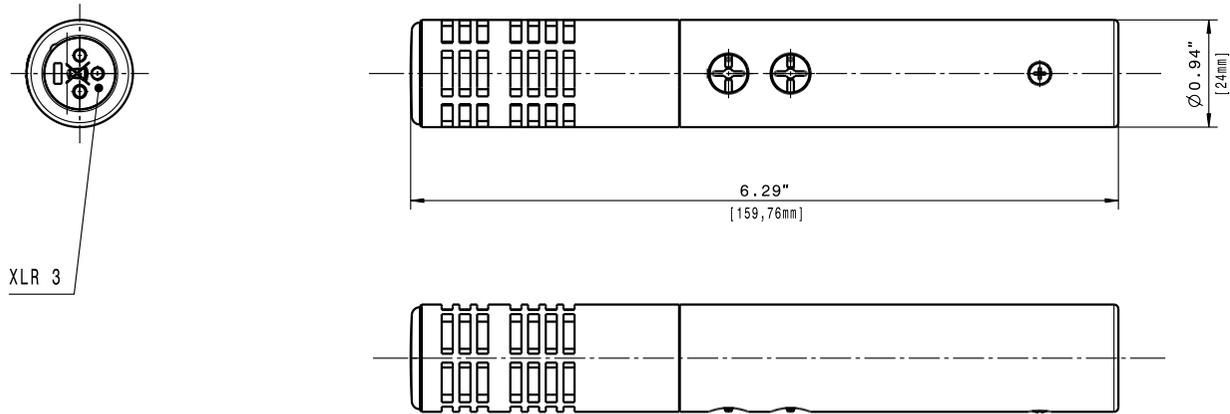
Transducer principle .....	pre-polarized condenser microphone
Pick-up pattern.....	cardioid
Frequency response .....	20.....20,000 Hz
Sensitivity (free field, no load at 1 kHz) .....	7 mV/Pa
Nominal impedance.....	100 Ohm
Min. terminating impedance.....	1000 Ohm
Equivalent noise level .....	19 dB(A)
Preattenuation .....	0, -10, -20 dB
Bass filter .....	linear
	roll-off 130 Hz, 6 dB/oct.
	cut-off 85 Hz, 18 dB/oct.
Max. sound pressure level (active).....	137, 147, 157 dB
Phantom powering.....	48 V/2.2 mA
Connector .....	XLR-3
Dimensions.....	24 x 157 mm
Weight .....	198 g

### DELIVERY INCLUDES

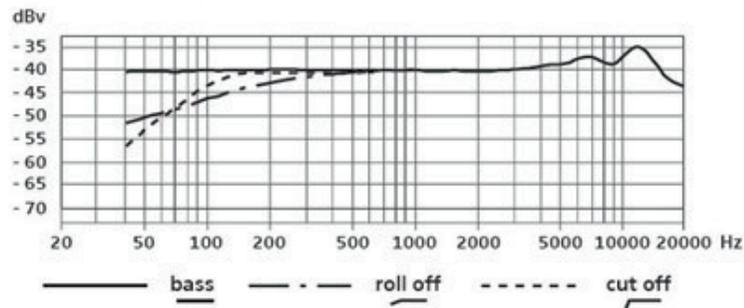
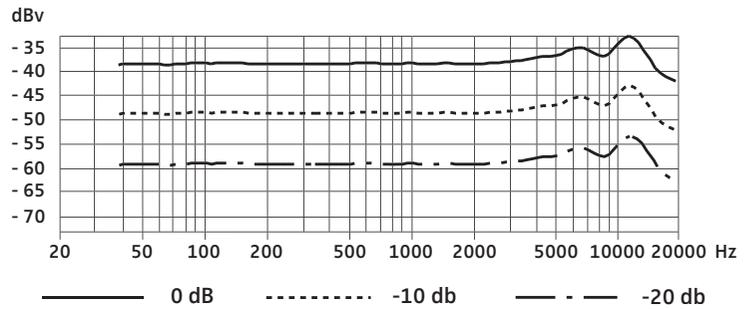
- 1 e 914
- 1 microphone clamp
- 1 pouch
- Instructions for use

# High grade pre-polarized condenser microphone e 914 for ambitious recordings and live performances

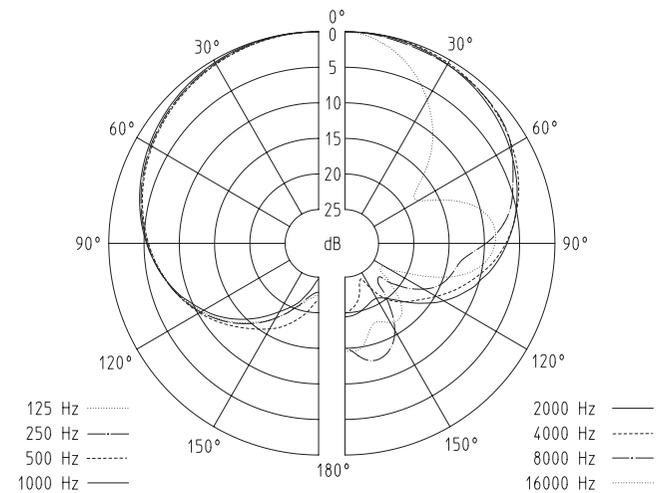
## DIMENSIONS



## FREQUENCY RESPONSE



## POLAR PATTERN



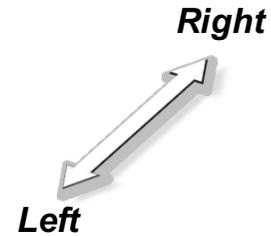
# TSL Soundfield ST450 MKII Portable Mic System

## ST450 Control Unit



End Fire OFF  
Invert OFF

Source sonore →

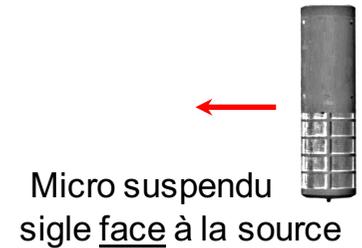
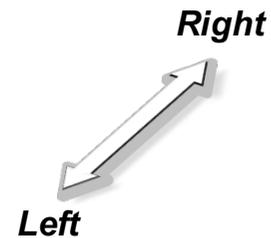


## ST450 Microphone



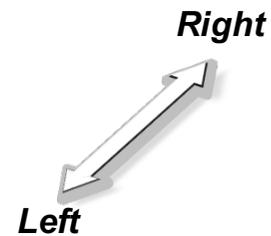
End Fire OFF  
Invert ON

Source sonore →



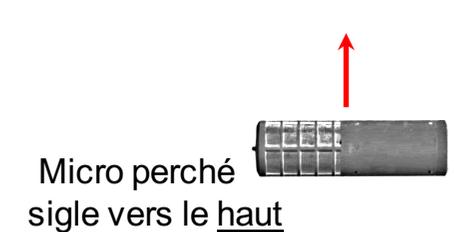
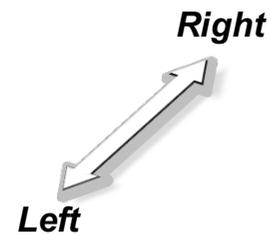
End Fire ON  
Invert OFF

Source sonore →

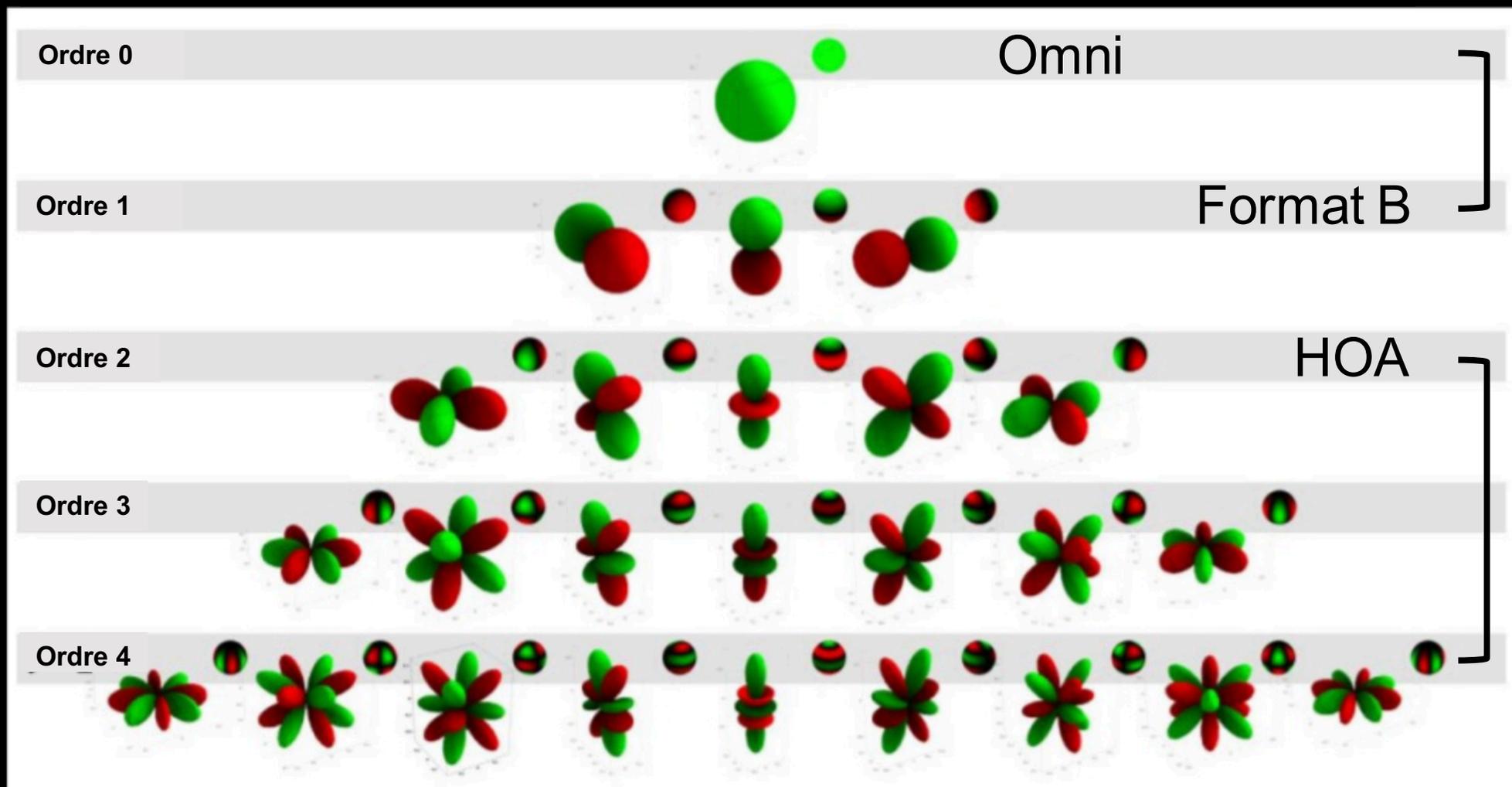


End Fire ON  
Invert ON

Source sonore →



# Spherical Harmonics Projection & Reconstruction



Benoît Courribet

Julien Bréval

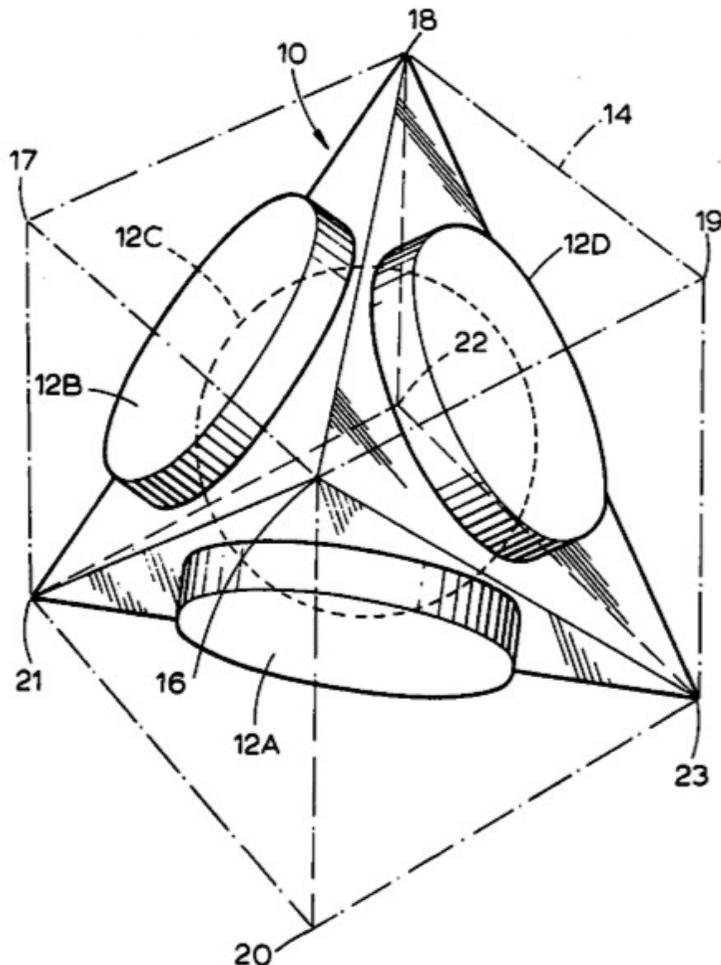
Timothée Baschet

# Son surround sur supports standard : une enquête technologique pour le musicien

**CICM**, MSH Paris Nord, Université de Paris 8

# Enregistrement en *Format A*

4 capsules cardioïdes



- Conversion en *Format B* en utilisant les filtres théoriques
- Conversion en *Format B* en utilisant des filtres issus de mesures pratiques
- Conversion en *Format B* avec l'encodeur logiciel ou matériel fourni

# Enregistrement en *Format A*

## Système SoundField ST350

Microphone

Préamplificateur, encodeur  
en *Format B* et décodeur  
pour stéréophonie 2/0



Détail des capsules

# Enregistrements en *Format A*

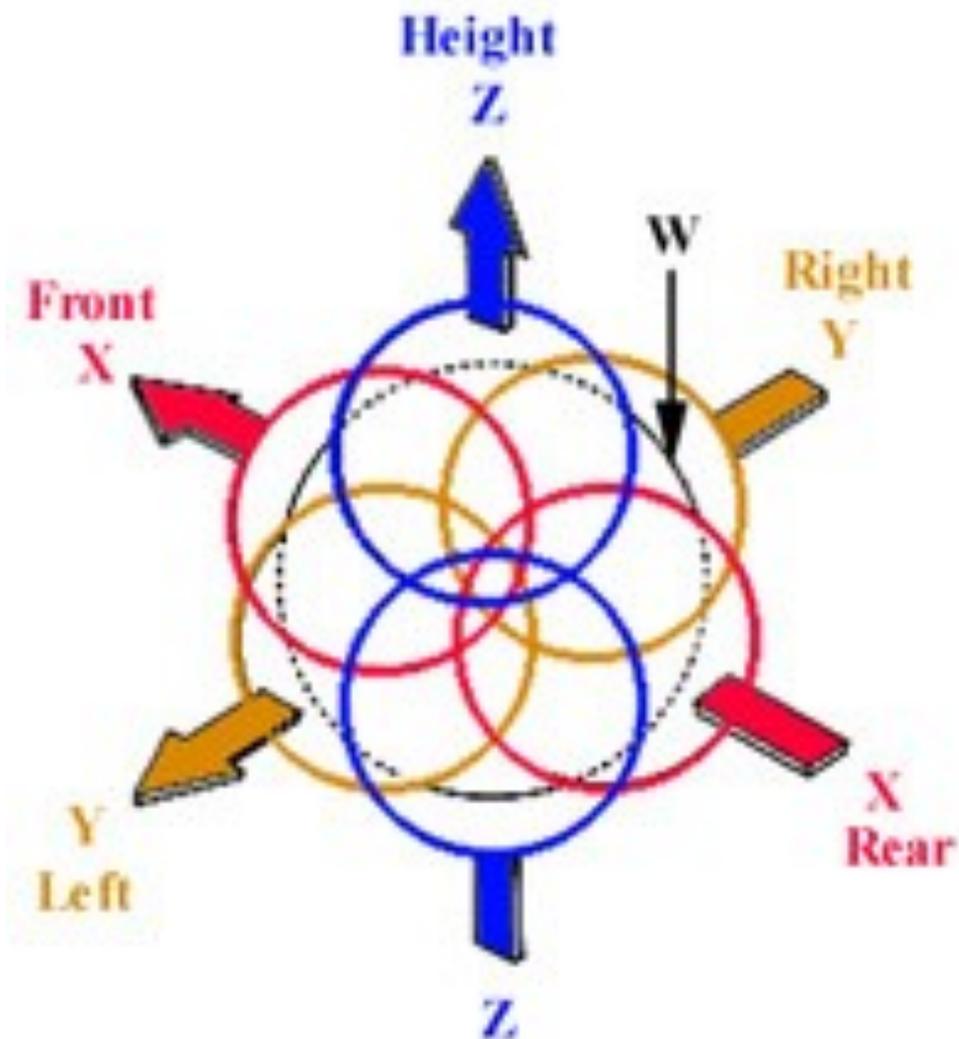


**DPA-4** (ici en cours de calibration)



**Core Sound TetraMic**

# Le modèle *Ambisonics* à l'ordre 1

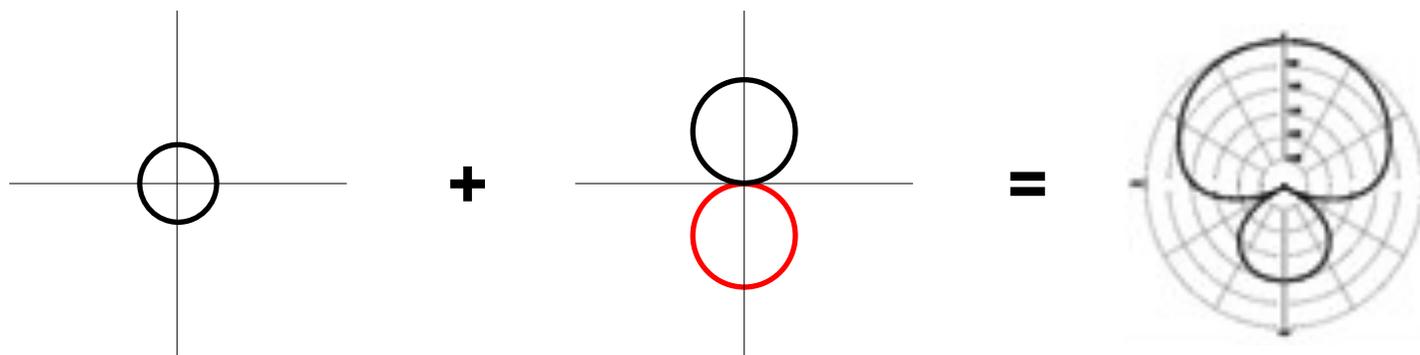
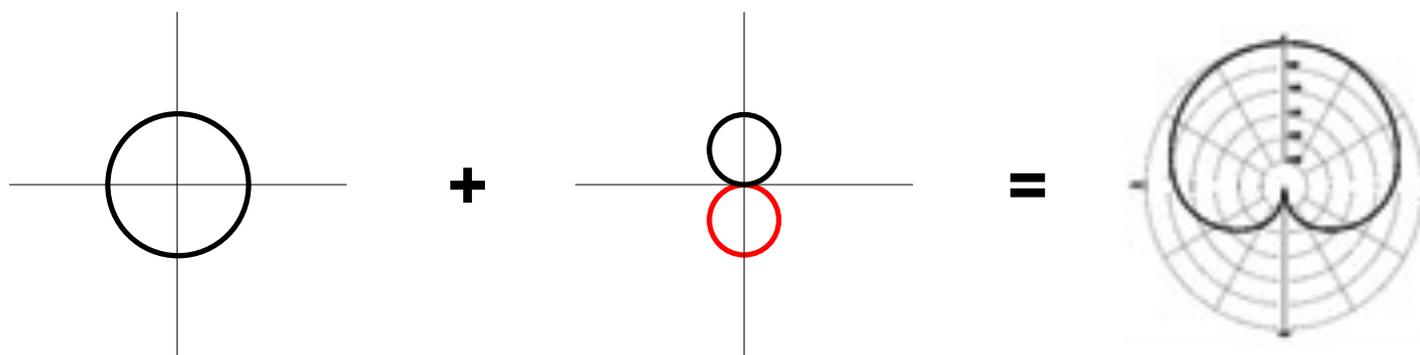
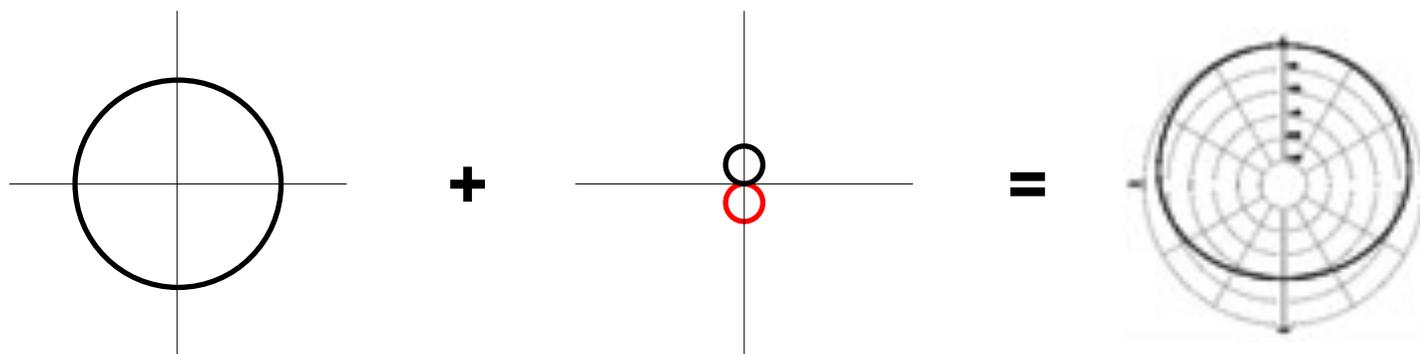


- Cet encodage du champ acoustique comprend 4 composantes : W, X, Y, Z.
- On appelle *Format B* ces composantes.
- On peut y associer un fichier audio de quatre canaux.

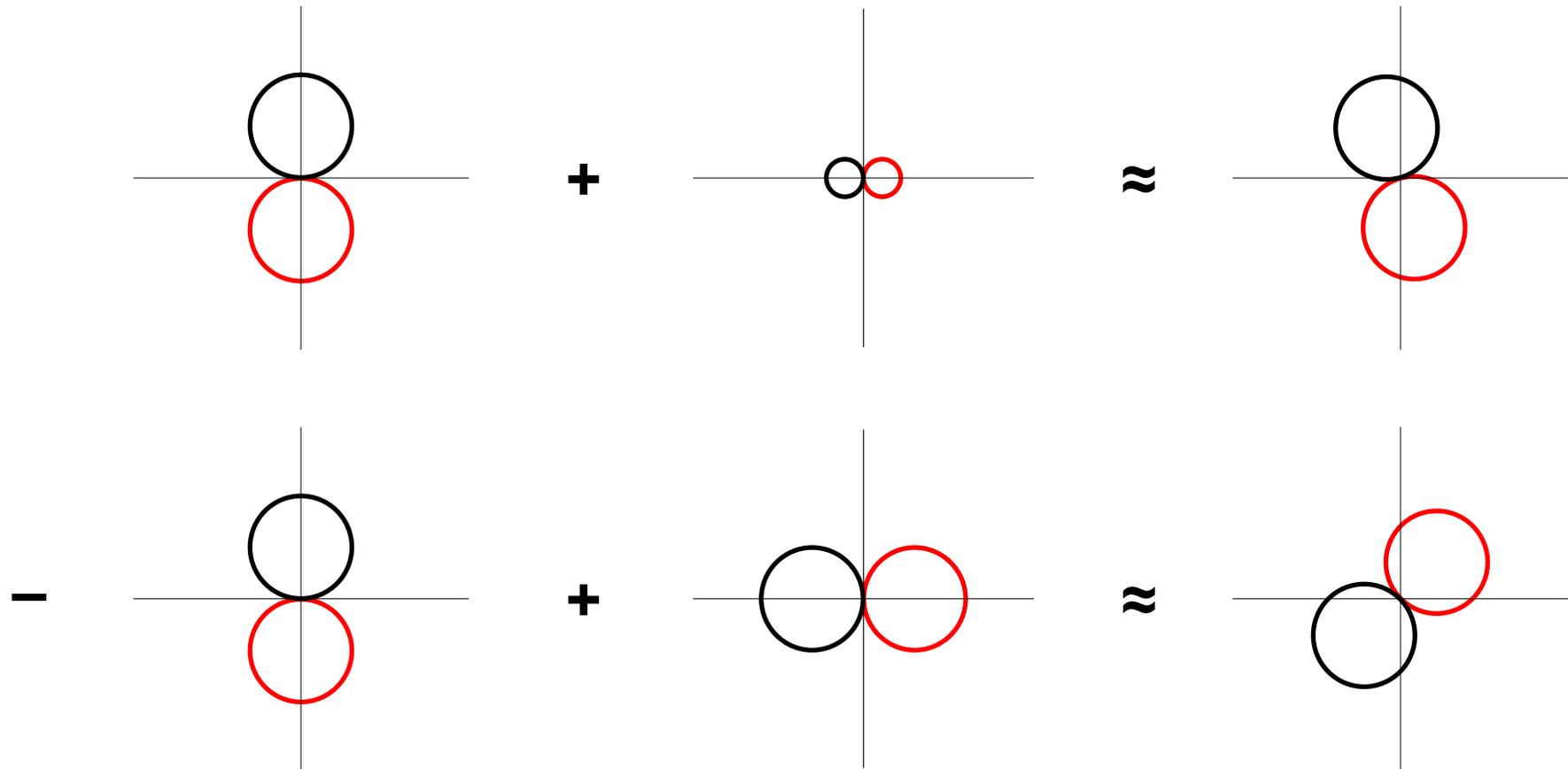
# Synthèse d'un *microphone virtuel*

- Un *microphone virtuel* est entièrement défini par sa directivité et son angle de positionnement.
- Le mixage d'un microphone omnidirectionnel et d'un microphone en figure de "8" coïncidents permet de synthétiser un microphone dont la **directivité** dépend des coefficients de mixage.
- Le mixage de deux microphones en figure de "8" orthogonaux et coïncidents permet de synthétiser un microphone en figure de "8" dont l'**angle de positionnement** dépend des coefficients de mixage.

# Exemples de synthèse de directivité



# Exemples de synthèse d'angle



# Technique de décodage

- Un *microphone virtuel*  $M$  est défini par la relation suivante :  $\mathbf{M} = a\mathbf{W} + b(c\mathbf{X} + d\mathbf{Y})$ , où  $a$ ,  $b$ ,  $c$  et  $d$  sont des coefficients de mixage compris entre -1 et 1.
- Dans chaque enceinte du système de reproduction sonore, on diffuse un seul *microphone virtuel* dont on peut choisir l'angle et la directivité.
- S'il y a un caisson de basses, on peut y diffuser le canal  $W$ .

# Prise de son en Haute Résolution Spatiale

Système d'enregistrement 5.0 Trinnov SRP



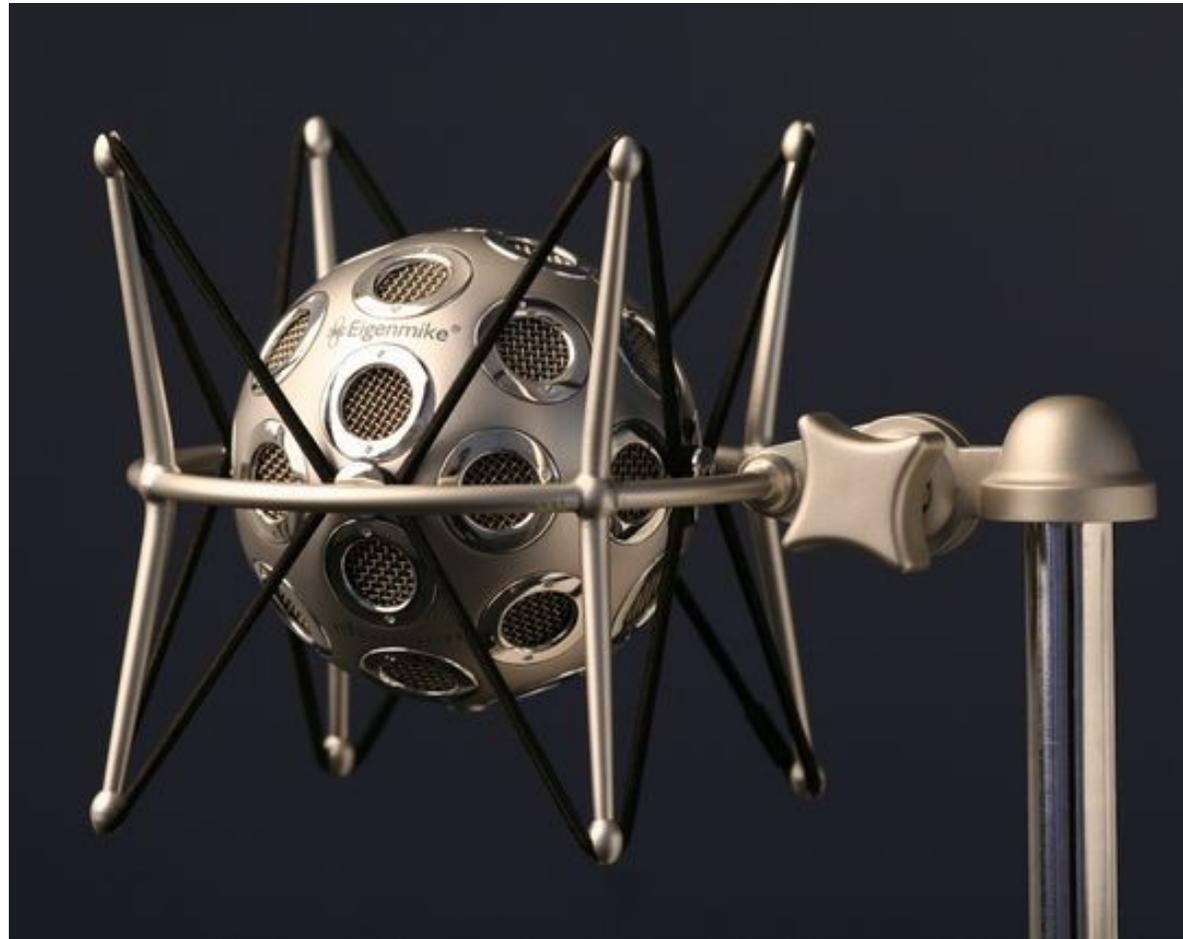
+



=



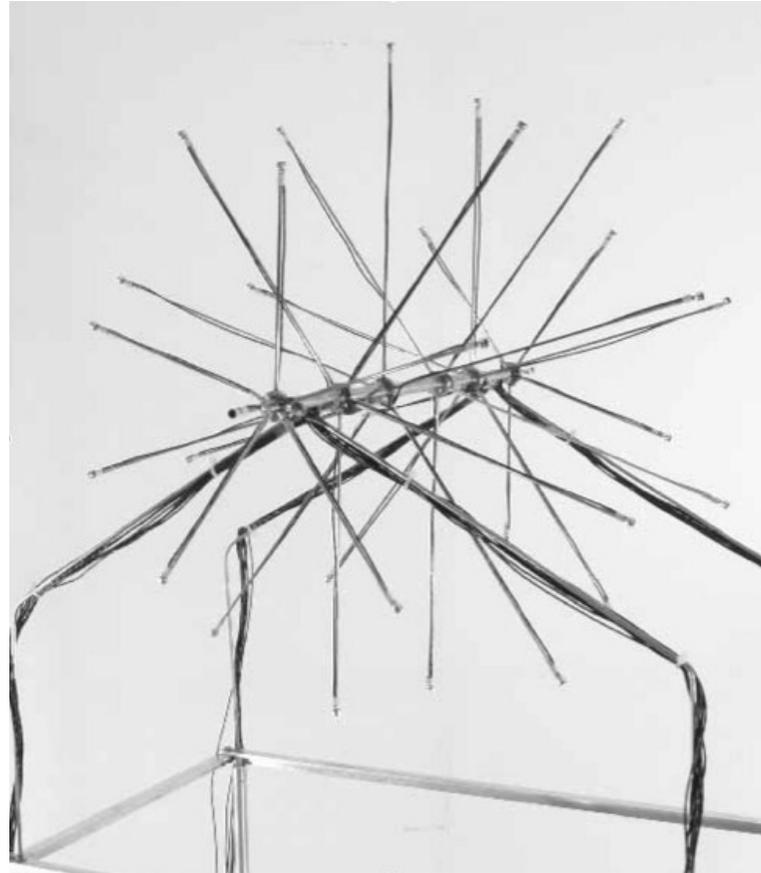
# Prise de son en Haute Résolution Spatiale



**32 capteurs  
Ordre 4**

**MH Acoustics E32 Eigenmike**

# Prise de son en Haute Résolution Spatiale



Sphère géodésique de microphones pour l'analyse  
directionnelle des champs acoustiques reverbérés  
(**NRC - Canada**)

mardi, 24 juin 2014 20:12

**Juin 2014**

## **Spheric Collection, une nouvelle sonothèque ambisonique made in France**

Écrit par Dorian Darcourt

Voilà une toute nouvelle sonothèque indépendante qui voit le jour : Spheric Collection.

En plus d'être d'origine française (on la doit au monteur son cinéma Gaël Nicolas), elle a la particularité de ne proposer que des sons en format-B (ambisonique 1er ordre, 4 pistes W/X/Y/Z), qui peuvent être décodées par plugin dans la majorité des DAWs du marché, avec Harpex-B ou Surround Zone 2, ce dernier étant complètement gratuit. On peut ainsi obtenir des ambiances allant de la mono au 7.1, et même obtenir des sons en 9.1 pour l'Auro-3D ou l'Atmos moyennant quelques manipulations. Notez aussi que les plugins permettent des manipulations du champ sonore assez intéressantes pour le son à l'image, notamment avec Haprex-B.

Spheric Collection propose dès son lancement une 20aine de collections thématiques comptant plus de 500 sons, et des démos des collections sont écoutables sur le site en binaural (ça n'a pas fonctionné ici via Firefox, mais ça passe avec Safari).

Bonne chance à ce nouveau venu, et puisse l'ambisonique se démocratiser !

Lu **5808** fois

Publié dans Sonothèque

Tags: Sonothèque Multicanal Ambisonique

Ajouter un commentaire |





### **VR Mic dans une Pianissimo de Cinela :**

- Poids de l'ensemble 750 g
- Longueur 32 cm
- Diamètre 16 cm
- Prix VR mic + Pianissimo + cables  
≈ 3000 €

## **Sennheiser Ambeo VR Mic**

*samedi 04 mars 2017*

Suivant la mode de la Virtual Reality et des caméras 360°, Sennheiser propose un nouveau microphone de technique Ambisonique particulièrement adapté à ce marché pour une prise de son multicanal autonome et compacte, avec un conception assez similaire à celle du Soundfield SPS200.

Probablement le seul microphone de Sennheiser à n'avoir pas de référence alphanumérique, l'**Ambeo VR Mic** est un microphone Ambisonique d'ordre 1 au *Format A* à quatre capsules polarisées electret KE 14 nécessitant chacune une alimentation Phantom 48V à la consommation réduite de 3,5 mA, alors que le SPS200 utilise des capsules à condensateur statiques à alimentation Phantom 48V.

Avec un bruit de fond de +18 dBA pour une dynamique de 130 dB et une réponse en fréquences de 20-20 kHz, les caractéristiques acoustiques des capsules sont relativement correctes pour des champs sonores ordinaires mais le souffle sera nettement audible dans un environnement très calme. Pas question d'aller enregistrer avec ce microphone les infimes frémissements de la banquise au Pôle Nord, encore moins les déplacements du scarabée sacré dans le désert d'Égypte.

Les quatre capsules de l'Ambeo sont disposées en tétraèdre et englobées dans une bonnette grillagée métallique (là où le SPS200 les laisse à nu). D'un poids assez conséquent de plus de 400 g et de taille plus imposante que le SPS200 (21,5 cm de long, corps de diamètre 25 mm et tête de diamètre 49 mm), il présente à sa base un connecteur DIN vissable à 12 points. Un câble Mogami multipaire de 3 mètres à connecteurs Lumberg, fabriqué par Cinela pour DC Audiovisuel, relativement souple, et un épanoui DIN 12 points vers 4 XLR-3M permettent de le raccorder à un enregistreur à quatre entrées micro (Sonosax SX-R4, Cantar Mini). Une suspension à lyres conçue par Sennheiser autorise son utilisation en intérieur.

Pour l'extérieur, Cinela a spécialement adapté pour l'Ambeo VR Mic une bonnette Pianissimo, avec un connecteur DIN mâle 12 points vissable Lumberg directement intégré dans la rotule.

Le repérage spatial des capsules ayant une importance capitale, une gravure sur le corps du microphone indique le devant (*front*) et le haut (*up*) du micro (du même côté que le logo et le nom du microphone *Sennheiser Ambeo VR Mic*). En effet les signaux des quatre capsules (1=avant-gauche-haut, 2=avant-droit-bas, 3=arrière-gauche-bas et 4=arrière-droit-haut) doivent ensuite être combinés dans un ordre défini pour générer le *Format B*.

L'usage standard est le micro vertical la tête en haut, la gravure front vers l'avant (mode *Upright* dans le plugin Sennheiser, normal dans le plugin Soundfied). Il est

possible de l'utiliser verticalement, tête en bas, gravure front vers l'avant (mode *Upside Down* chez Sennheiser, *Invert* chez Soundfield) ou encore horizontalement, tête vers l'avant, gravure front vers le haut (mode *Endfire* chez Sennheiser, *End Fire+Invert* chez Soundfield) et enfin horizontalement, tête vers l'avant, gravure front en bas (mode *End Fire* chez Soundfield uniquement, le plugin Sennheiser ne permet pas cette position).

Il est très important de communiquer à celui qui va exploiter les fichiers Ambisonique la position du microphone lors de la prise de son !

Très peu utilisé en fiction cinéma et télévisuelle, Ambisonique est le fruit des recherches dans les années 70 de deux chercheurs de la *British National Research Development Corporation* à l'université de Oxford. Ceux-ci ont par la suite créé, en 1978, la société Soundfield, récemment racheté par le fabricant de microphone RØde, pour la commercialisation de microphones Ambisonique. Les brevets concernant l'Ambisonique ayant expiré, quelques très rares autres fabricants se sont lancés dans la fabrication de microphones utilisant cette technologie. Surfant sur la vague de la réalité virtuelle, Sennheiser est donc le dernier en date.

La technique Ambisonique, fondée sur la décomposition d'un champ sonore en harmoniques sphériques, vise à une reproduction tridimensionnelle de l'espace sonore en un point donné. Le *Format B* correspond à cette décomposition limitée à l'ordre 1. Quatre canaux sont ainsi nécessaires : *W* (pression omnidirectionnelle) et *XYZ* (gradients bidirectionnels selon les trois axes).

9 canaux seraient nécessaires pour la décomposition d'un champ sonore à l'ordre 2, et 25 pour celle à l'ordre 3 ! Plus l'ordre est élevé, plus le résultat est spatialement précis. L'Ambisonique d'ordre 1 est un compromis simple et minimal qui résulte donc en une image, pouvant être considérée par certains comme un peu *floue*, du champ sonore capté.

Les dispositifs d'ordre plus élevé sont dénommés les HOA (High Order Ambisonique). Ils sont très rares, complexes et coûteux. La reproduction de tels signaux est encore un vaste champ d'investigations, mais il y a plein d'applications potentielles (hyper-directivité ou diffusion multicanal classique par plugin). Le CNAM en association avec Cinela travaille à un projet HOA à l'aide d'une sphère incluant

50 capsules, soit 200 capteurs Mems, selon une géométrie de Lebedev.

Le *Format A* reprend simplement le signal des quatre capsules plus ou moins cardioïdes, telles qu'elles sont disposées en tétraèdre par les fabricants. Il n'est pas exploitable directement. Il doit être matricé (encodé) au *Format B* à l'aide d'un plugin (principalement VST ou AAX) traitant les quatre signaux pour produire les quatre composantes WXYZ qui peuvent être enregistrés dans un fichier WAV polyphonique au format classique .amb ou au format plus spécifique Apple AmbiX .caf.

Concrètement, Sennheiser construit le *Format B* avec W (omni) égalant la somme des quatre capsules, le X égalant la somme des deux capsules avant moins la somme des deux capsules arrière, le Y égalant la somme des deux capsules gauche moins les deux capsules droit et enfin le Z égalant la somme des deux capsules haut moins les deux capsules bas.

Au cinéma comme en télévision, les composantes Ambisonique WXYZ nécessitent ensuite d'être réparties sur des systèmes de diffusion, stéréo, 5.1 ou 7.1, voire Dolby Atmos.

Sennheiser ne fournit qu'un plugin VST (pas de AAX pour Pro Tools) 32 ou 64-bits encodant le *Format A* des quatre capsules de l'Ambeo en *Format B* WXYZ. Il est donc nécessaire ensuite d'utiliser un autre plugin pour transformer ce *Format B* en un format de diffusion. Sennheiser déconseille l'utilisation d'un autre plugin que le sien pour passer du A à B, arguant d'un traitement spécifique aux caractéristiques des capsules.

Soundfield propose un plugin VST 32 ou 64-bits permettant de partir soit du *Format A*, soit du *Format B* pour arriver au *Format B* ou différents formats multicanaux du simple stéréo au 7.1 en passant par 5.0, 5.1, 6.0, 6.1 et 7.0.

De nombreux concepteurs de logiciels fournissent des plugins Ambisonique ([https://en.wikipedia.org/wiki/List\\_of\\_Ambisonic\\_Software](https://en.wikipedia.org/wiki/List_of_Ambisonic_Software)).

Le kit complet incluant le Sennheiser Ambeo VR Mic, sa bonnette Cinela Pianissimo, sa suspension, son câble multipaire et son épanoui, le tout dans une valise de transport adaptée, est d'ores et déjà disponible à la location chez DC Audiovisuel.

# HoaLibrary

## High Order Ambisonics Library

### Les formats :

Le matériel, les logiciels et la littérature liés à l'ambisonie font souvent références aux formats A, B ou encore C. Au début de l'ambisonie, sous l'influence notable de Gerzon [Gerzon, 1974][Gerzon, 1975], il parut nécessaire d'établir des normes autour des techniques ambisoniques pour répondre, en outre, à des contraintes commerciales et matérielles. Sans revenir sur les causes, l'ambisonie n'a pas eu le succès escompté sur un plan commercial mais a eu un fort impact sur le plan de la recherche notamment publique et universitaire (Université de York, Université des arts de Zurich, Université Queen Mary, IRCAM, Université Paris 8). Les formats liés aux normes ont donc perduré car loin d'être basés sur de simples contraintes commerciales, elles offrent de fortes optimisations logicielles en prédéterminant l'ordre ambisonique et les configurations du système de restitution. Bien que toujours utilisées [Wakefield, 2006][Neukom, 2009b], aujourd'hui, ces optimisations ne sont plus nécessaires et les formats peuvent restreindre les recherches; cependant il paraît nécessaire de revenir les principaux formats afin d'offrir une certaine compatibilité et pérennité aux différents projets liés à l'ambisonie.

### Le format B et les conventions d'encodage

Le format B (ou B-format), apparu au cours des années 70, est la norme originale établie pour l'encodage de champs sonores au format ambisonique. Il repose sur une décomposition en harmoniques sphériques d'ordre 1. Pour un signal  $S$ , les harmoniques sphériques sont définies tels que présenté dans le Tableau 1.

Tableau 1: Normes du format B

Canaux	Ordre	Indice	Equation 3D	Equation 2D
W	0	0	$S$	$S$
X	1	1	$S \times \cos(\delta) \times \cos(\theta)$	$S \times \cos(\theta)$
Y	1	-1	$S \times \cos(\delta) \times \sin(\theta)$	$S \times \sin(\theta)$
Z	1	1	$S \times \sin(\delta)$	$\varphi$

Aujourd'hui, le format B est augmenté jusqu'à l'ordre 3 (pour un tableau récapitulatif des normes jusqu'à l'ordre 3, se référer à Neukom [Neukom, 2009b] et Daniel [Daniel, 2001]). Malgré que cette norme possède de nombreux atouts liés à l'optimisation informatique, le format B atteint ses limites à cause notamment de la limite imposée aux ordres. Il existe de nombreuses conventions d'encodage plus ou moins liées au format B tels que Semi-Normalisation, Normalisation, Max-Normalisation et Furse-Malham set [Daniel, 2001][Malham, 2003]. Elles permettent de simplifier l'écriture de l'encodage, de faciliter le

décodage en fonction de l'ordre ambisonique et du nombre de dimensions. Ces normes possèdent en général leurs propres normalisations et peuvent limiter l'ordre donc le nombre d'harmoniques sphériques. Pour de plus amples informations se référer à Daniel [Daniel, 2001].

## Le Format A et le microphone soundfield

Le format A a été développé afin de pouvoir enregistrer des champs sonores 3D. Il est difficile d'enregistrer directement les composantes W, X, Y, Z d'un champ sonore à l'aide de microphones omnidirectionnels et en figure de 8. Grâce à un microphone SoundField, composé d'une série de capteurs cardioïdes placé sur les faces d'un tétraèdre, nous obtenons les quatre canaux LF, LB, RL, RB qui définissent le format A. Cette approche revient à discrétiser la sphère et permet par la suite de recomposer les canaux ambisoniques au format B [CG77] d'après les relations présentées dans le Tableau 2.

Tableau 2: Relations entre le format A et le format B

Canaux
$W = LF + LB + RF + RB$
$X = LF - LB + RF - RB$
$Y = LF + LB - RF + RB$
$Z = LF - LB - RF + RB$

Restreint, à l'origine, à l'ordre 1, le format A peut être envisagé aux ordres supérieurs, pour de plus amples informations se référer à Colafrancesco [Colafrancesco, 2012]. Cette norme peut être transposée en deux dimensions en décomposant un cercle en trois points ou plus (Figure 1), elle permet d'éviter l'utilisation d'un microphone ambisonique, notamment pour l'enregistrement de réponses impulsionnelles « surround » et offre de nouvelles possibilités musicales.

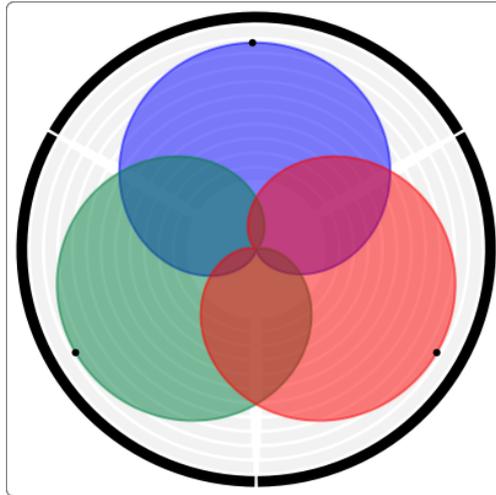


Figure 1 : Représentation des figures directives des trois microphones qui discrétise un cercle en trois points.

## Le format c et les systèmes standardisés

Le format C, aussi souvent appelée UHJ, comprend une série de normes qui permettent de rendre compatible l'ambisonie pour des systèmes de restitutions standardisés et plus répandus dans le commerce tels que la stéréo, le 5.1 et le 7.1 et des outils de stockages tels que le CD ou le DVD. Les canaux perdent inévitablement une partie de l'information liée aux champs sonores et la restitution subit une perte de qualité sur de nombreux aspects. Ces opérations relèvent souvent de « fine-tuning » et fonctionnent donc pour des ordres et des systèmes de restitution très fixes. Bien qu'envisageable, aucune généralisation de ces processus n'a encore été établie. Pour plus d'informations, se référer à Gerzon, Daniel, [Gerzon, 1975][Daniel, 2006]. Afin de répondre aux mêmes types de contraintes sont apparues d'autres formats tels que la SuperStereo et le format G [Wikipedia, 2012].

La bibliothèque Hoa:

Les principes sous-jacents aux formats qui viennent d'être présentés peuvent être retrouvés dans la bibliothèque Hoa de manière transformée afin de fonctionner aux ordres supérieurs et pour des systèmes de restitutions et d'enregistrements variés. L'encodeur ambisonique est une élévation du format B pour tous les ordres, les opérations liées au format A se retrouvent dans les outils de projection et de recombinaison et le format C possède les mêmes fonctionnalités que les décodeurs pour des systèmes standardisés comme la stéréophonie et le 5.1.

Références pour citer cet article :

“Documentation de la bibliothèque Hoa : Les formats”, Auteur : Pierre Guillot Année : 2012-2013 © CICM - Université Paris 8, Labex Arts H2H.

# Production, transmission et restitution temps réel d'une scène sonore dans un format audio 3D flexible

Sébastien Moreau<sup>1</sup> Jérôme Daniel<sup>1</sup> Adil Chraa<sup>1</sup>

<sup>1</sup> France Télécom R&D - DIH / IPS

2, avenue Pierre Marzin, 22 307 Lannion cedex – France

{sebastien.moreau, jerome.daniel, adil.chraa}@francetelecom.com

## Résumé

Cet article présente une plateforme de production, transmission et restitution temps réel d'une scène sonore 3D. Parmi les formats audio existants, le format High Order Ambisonics (HOA) a été retenu pour sa souplesse d'utilisation. Des outils d'exploitation, inexistant pour ce format encore récent, ont été élaborés pour développer la plateforme temps réel. Cette dernière est constituée d'un microphone haute résolution, d'un module de transmission/réception par réseau IP, et d'un décodeur qui permet la restitution sur divers systèmes de diffusion. L'implantation a été effectuée avec la technologie VST de Steinberg et s'adapte à beaucoup de logiciels audio grand public. La plateforme réalisée peut concerner la diffusion musicale, radiophonique, etc.

## Mots clefs

High Order Ambisonics, spatialisation sonore, format audio.

## 1 Introduction

Nous présentons ici un système temps réel de spatialisation sonore dont le but est de transmettre et de reconstruire un champ sonore 3D naturel dans un lieu distant du champ original. Un format de représentation du champ acoustique a été choisi parmi ceux existants en fonction de critères objectifs tels que la souplesse d'utilisation et la quantité d'information à transmettre.

## 2 Représentation d'une scène sonore 3D : quel format ?

Une technique de spatialisation sonore permet de représenter les informations spatiales et temporelles d'un champ acoustique dans le but de les stocker, transmettre, ou reproduire. Ceci pose le problème de la définition d'un *format* de représentation du champ acoustique.

### 2.1 Considérations préliminaires

La Figure 1 représente les principes de bases de la spatialisation sonore : l'*encodage spatial* d'une scène

sonore dans un *format* de représentation, puis le *décodage* de la scène formatée permettant sa reproduction.

Le *format* de représentation contient les informations spatiales et temporelles du champ sonore nécessaires à sa reconstruction. Il décrit donc, dans les limites de ses capacités, un champ sonore. Il est constitué d'un ensemble de signaux temporels auxquels peuvent être associés des paramètres de spatialisation.

L'*encodage spatial* désigne le processus qui permet d'acquérir, de traduire une scène sonore dans un format spécifique de représentation.

Le *décodage spatial* concerne quant à lui l'adaptation des données formatées au système de restitution (casque, 5.1, etc.). Si le format est constitué de signaux destinés à alimenter directement un système de diffusion particulier, le décodage n'est pas nécessaire (binaural, stéréophonie).

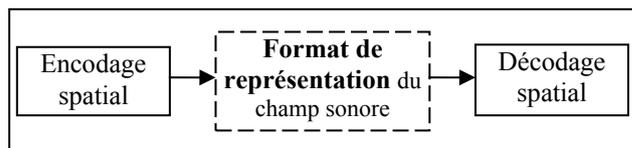


Figure 1 – Principe de base de la spatialisation sonore

### 2.2 Choix du format

Le choix du *format* de représentation d'une scène sonore est un choix décisif puisque de lui dépend la qualité potentielle de la reconstruction spatiale des sources sonores. Pour le développement de notre application temps réel, le choix s'est effectué selon trois critères :

- la possibilité de *manipuler* la scène sonore encodée ;
- la *flexibilité* par rapport au système de restitution ;
- la *concision* de la scène sonore encodée.

On entend par *manipulation* d'une scène sonore encodée le fait de modifier la position spatiale des sources sonores déjà présentes.

D'autre part, un format est dit *flexible* s'il peut s'adapter à divers systèmes de restitution pour la reconstruction de la scène sonore.

Enfin, la *concision* de la représentation est une contrainte imposée par les ressources matérielles limitées. Il est donc avantageux qu'un format de représentation contienne peu d'informations redondantes ou qu'il puisse être compressé

sans entraîner une dégradation de la qualité (du moins perceptive) de reproduction.

## 2.3 Comparaison des principaux formats existants

Le Tableau 1 permet de comparer les formats associés aux principales techniques de spatialisation existantes suivant les trois critères définis plus haut. Ces principaux formats sont : la *stéréophonie* (deux canaux ou plus) ; le *binaural* ; *High Order Ambisonics* (HOA) ; et le *tout paramétrique*.

	Manipulation de la scène sonore	Dispositif de restitution	Taille de la scène encodée
<b>Stéréophonie</b> (2.0, 5.1, etc.)	Impossible, scène sonore figée	fixe, lié au format	1 signal par HP
<b>Binaural</b>			2 signaux
<b>HOA</b> (ordre $M$ )	globale (rotation, etc.)	adaptable, indépendant du format	3D : $(M+1)^2$ 2D : $(2M+1)$ signaux
<b>tout paramétrique</b>	chaque source de façon indépendante		1 signal par source + paramètres de position

Tableau 1 - Comparaison des principaux formats de représentation de scène sonore

Le format *stéréophonique* tel que défini ici, est constitué de signaux destinés à alimenter directement les haut-parleurs d'une configuration prédéfinie (2.0, 4.0, 5.1, etc.). Le format *binaural* est composé de deux signaux, appelés *signaux binauraux*, contenant les transformations subies par le son lors de son interaction avec le corps de l'auditeur (pavillons, tête et torse) et devant être reproduit au niveau des tympanes au moyen d'un casque.

Le format associé à la technologie *High Order Ambisonics* (HOA) est constitué de signaux appelés *signaux HOA* qui résultent de la décomposition en harmoniques sphériques d'un champ sonore en un point. Le nombre de signaux dépend de l'ordre  $M$  de la décomposition :  $(M+1)^2$  en 3D et  $(2M+1)$  en 2D. De l'ordre dépend également la résolution spatiale de la description.

Enfin, un format *tout paramétrique* contient autant de signaux que de sources sonores, auxquels sont associés des paramètres de position spatiale. La reconstruction est ensuite effectuée lors du décodage selon la technique de spatialisation choisie (*binaural*, *HOA*, *Wave Field Synthesis*, etc.) et le dispositif de restitution.

## 2.4 HOA : un format avantageux

Les formats *HOA* et *tout paramétrique* sont particulièrement intéressants car ils permettent de

représenter une scène sonore 3D indépendamment du système de restitution. Bien que le second paraisse offrir plus de possibilités en terme de manipulation de scène sonore, il comporte un nombre important de signaux lorsque le champ acoustique encodé est complexe. De plus il suppose la prise de son isolée des sources sonores et s'adapte donc mal à l'enregistrement naturel en 3D.

Le format HOA semble réaliser un bon compromis entre taille et souplesse d'utilisation. Il a donc logiquement été choisi pour développer notre application temps réel. Cependant, les outils de production qui lui sont associés sont à l'heure actuelle relativement limités. Nous en avons nous-même développés sur les bases théoriques suivantes.

## 3 High Order Ambisonics

Cette section introduit les éléments théoriques qui nous permettrons d'exploiter le format HOA dans notre application temps réel. Seuls certains aspects correspondant aux fonctions requises seront abordés : l'encodage de champ sonore naturel, les rotations du champ sonore, et le décodage pour divers systèmes de restitution.

### 3.1 Série de Fourier-Bessel

La technologie *High Order Ambisonics* (HOA) se base sur une représentation harmonique spatiale du champ sonore, solution de l'équation d'onde exprimée en coordonnées sphériques (azimut  $\theta$ , élévation  $\delta$ , rayon  $r$ ). Sur une zone exempte de source sonore, cette solution est la *série de Fourier-Bessel* :

$$p(\theta, \delta, kr) = \sum_{m=0}^{\infty} j_m^m j_m(kr) \sum_{0 \leq n \leq m, \sigma = \pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \delta) \cdot (1)$$

Les fonctions radiales  $j_m(kr)$  sont les *fonctions de Bessel sphériques* de première espèce d'ordre  $m$  ( $k$  étant la longueur d'onde). Les fonctions angulaires  $Y_{mn}^{\sigma}(\theta, \delta)$  sont les *fonctions harmoniques sphériques*. Et enfin, les signaux  $B_{mn}^{\sigma}$  exprimés ici dans le domaine fréquentiel, sont les *signaux HOA* qui décrivent le champ sonore sur la zone considérée [1].

Notons que les harmoniques sphériques sont orthogonales deux à deux, ce qui se traduit mathématiquement par l'équation suivante :

$$\int_S Y_{mn}^{\sigma}(\theta, \delta) Y_{m'n'}^{\sigma'}(\theta, \delta) dS = \delta_{mm'} \delta_{nn'} \delta_{\sigma\sigma'} \quad (2)$$

où  $\delta_{mm'}$  est le symbole de Kronecker, l'intégration se faisant sur la sphère unité.

En tronquant la série (1) à un ordre fini  $M$  ( $m \leq M$ ), on obtient une approximation du champ sonore sur une zone de l'espace d'autant plus grande que  $M$  est élevé.

### 3.2 Encodage de champ sonore naturel 3D

L'encodage de champ naturel au format HOA consiste à extraire en un point d'un champ acoustique les signaux

$B_{mn}^\sigma$ . La stratégie adoptée consiste à mesurer la pression sonore à la surface d'une sphère rigide de rayon  $a$  fixé. Celle-ci peut s'exprimer de la façon suivante :

$$p_a(\theta, \delta) = \sum_{m=0}^{\infty} W_m(ka) \sum_{0 \leq n \leq m, \sigma = \pm 1} B_{mn}^\sigma Y_{nm}^\sigma(\theta_q, \delta_q) \quad (3)$$

avec 
$$W_m(ka) = \frac{i^{m-1}}{(ka)^2 h_m^-(ka)}$$

En utilisant la *propriété d'orthogonalité* des harmoniques sphériques (2), on peut calculer d'après l'équation (3) les signaux HOA (projection sur la base des harmoniques sphériques) :

$$B_{mn}^\sigma = \frac{1}{W_m(ka)} \int_S p_a(\theta_q, \delta_q) Y_{nm}^\sigma(\theta_q, \delta_q) dS \quad (4)$$

En pratique, la pression sonore n'est mesurée qu'en un nombre fini  $Q$  de positions à la surface de la sphère. On effectue alors une approximation de la décomposition harmonique jusqu'à un ordre  $M$  maximal restreint :

$$B_{mn}^\sigma = \frac{1}{W_m(ka)} \sum_{q=1}^Q p_a(\theta_q, \delta_q) Y_{nm}^\sigma(\theta_q, \delta_q), \quad m \leq M. \quad (5)$$

Le nombre de signaux HOA estimé est alors de  $K=(M+1)^2$ . Pour une bonne estimation, on choisit  $Q$  tel que  $Q \geq K$ . D'autre part, l'équation (5) suppose que la base échantillonnée des harmoniques sphériques (représentée par les positions des microphones) est orthogonale. Dans des conditions plus générales, on préférera à cette *méthode de projection*, une méthode d'estimation basée sur une *approximation aux moindres carrés* [2].

Quelque soit la méthode choisie, les signaux HOA sont obtenus par *matricage* puis *égalisation*  $EQ_m = 1/W_m(ka)$  des signaux microphoniques. Cette égalisation nécessite, notamment en basse fréquence, d'être limitée en amplitude pour sa mise en œuvre [2].

### 3.3 Rotation de la scène sonore

Lorsque la scène sonore pivote (suivant trois degrés de liberté), le groupe formé par les  $(2m+1)$  signaux HOA de même ordre  $m$  est transformé de la façon suivante :

$$b'_m = R_m \cdot b_m, \quad (6)$$

avec  $b'_m = [b'_{mm}, b'_{m,m-1}, \dots, b'_{m,m-1}, b'_{m,m-2}, \dots, b'_{m,0}]^t$  et

$$b_m = [b_{mm}, b_{m,m-1}, \dots, b_{m,m-1}, b_{m,m-2}, \dots, b_{m,0}]^t.$$

$b'_m$  est le vecteur contenant les  $(2m+1)$  signaux HOA, exprimés dans le domaine temporel, résultant de la manipulation,  $R_m$  la matrice de dimensions  $(2m+1) \times (2m+1)$  caractérisant la rotation, et  $b'_m$  le vecteur contenant les  $(2m+1)$  signaux HOA de départ, exprimés également dans le domaine temporel.

La rotation de scène sonore est ici destinée à être utilisée en association avec une *head-tracker* pour corriger les mouvements de tête en écoute binaurale (avec casque). Nous ne considérons donc ici que les rotations dans le

plan azimutal (rotation gauche-droite). Dans ce cas particulier, chaque couple de signaux  $(b_{mn}^1, b_{mn}^{-1})$ ,  $m$  et  $n$  étant fixés, s'obtiennent pour une rotation d'angle  $\theta'$  d'après la relation :

$$\begin{bmatrix} b_{mn}^1 \\ b_{mn}^{-1} \end{bmatrix} = \begin{bmatrix} \cos(n\theta') & -\sin(n\theta') \\ \sin(n\theta') & \cos(n\theta') \end{bmatrix} \begin{bmatrix} b_{mn}^1 \\ b_{mn}^{-1} \end{bmatrix}. \quad (7)$$

Nous pouvons ainsi calculer la matrice de rotation en fonction de l'ordre  $m$  considéré. Par exemple, pour l'ordre  $m = 2$ , la matrice caractérisant une rotation  $\theta'$  en azimut est la suivante :

$$R_2 = \begin{bmatrix} \cos 2\theta' & -\sin 2\theta' & 0 & 0 & 0 \\ \sin 2\theta' & \cos 2\theta' & 0 & 0 & 0 \\ 0 & 0 & \cos \theta' & -\sin \theta' & 0 \\ 0 & 0 & \sin \theta' & \cos \theta' & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (8)$$

### 3.4 Décodage

Le *décodage* consiste à adapter une scène sonore encodée au format HOA à un dispositif particulier de  $L$  haut-parleurs, que l'on choisit concentrique. Les haut-parleurs sont au moins aussi nombreux que les  $K$  signaux HOA.

On considère que les haut-parleurs émettent des ondes planes (champ lointain) qui contribuent à la reconstruction des signaux HOA suivant l'équation matricielle :

$$b = C \cdot s, \quad (9)$$

où  $b = (b_{00}^1, b_{11}^1, b_{11}^{-1}, \dots, b_{mm}^\sigma)^t$  est le vecteur constitué par les signaux HOA (domaine temporel),  $C$  est la matrice contenant les gains  $Y_{mn}^\sigma(\theta_l, \delta_l)$  associés à la direction  $l$  de chaque haut-parleur, et enfin  $s = (s_1, s_2, \dots, s_L)^t$  est le vecteur contenant les signaux temporels émis par les haut-parleurs.

Ces signaux sont obtenus en inversant l'équation (9), soit par combinaison (*matricage*) des signaux HOA :

$$s = D \cdot b \quad (10)$$

avec

$$D = C^t (C C^t)^{-1}.$$

$D$  est appelé la *matrice de décodage*, c'est la matrice pseudo-inverse de  $C$ .

Si on désire effectuer un décodage pour le casque d'écoute, il est nécessaire d'ajouter une étape de simulation binaurale des haut-parleurs : on applique aux signaux issue de (10) des filtres binauraux  $h_G(\theta, \delta)$  et  $h_D(\theta, \delta)$  caractérisant le canal acoustique entre les haut-parleurs et les oreilles gauche et droite.

Pour plus d'efficacité, on synthétise les opérations de *matricage* et de *filtrage* binaural sous la forme d'un ensemble  $f$  de filtres s'appliquant directement sur les signaux HOA. On obtient :

$$f_G = h_G \cdot D \text{ et } f_D = h_D \cdot D, \quad (11)$$

avec  $f_G = [f_{G00}^1, \dots, f_{Gmm}^\sigma]$ ,  $f_D = [f_{D00}^1, \dots, f_{Dmm}^\sigma]$ ,

$$h_G = [h_G(\theta_1, \delta_1), \dots, h_G(\theta_L, \delta_L)], \quad h_D = [h_D(\theta_1, \delta_1), \dots, h_D(\theta_L, \delta_L)].$$

Le décodage consiste alors à effectuer les opérations de filtrage suivantes (\* désigne l'opération de convolution) :

$$g = \sum_{m,n,\sigma} b_{mn}^{\sigma} * f_{Gmn}^{\sigma} \text{ et } d = \sum_{m,n,\sigma} b_{mn}^{\sigma} * f_{Dmn}^{\sigma}, \quad (12)$$

où  $g$  et  $d$  sont les signaux binauraux gauche et droit.

## 4 Plateforme temps réel

La technologie HOA a donc été choisie pour réaliser un système d'acquisition, transmission, et restitution temps réel d'une scène sonore 3D naturelle. Plusieurs applications sont visées : diffusion sur Internet d'émissions radiophoniques spatialisées, de concerts musicaux ; partage d'ambiance sonore, etc.

### 4.1 Description

La Figure 2 montre les différents éléments constituant la plateforme développée. Un *Encodeur* encode un champ sonore naturel 3D au format HOA. La scène encodée est ensuite transmise au *décodeur* qui permet de l'adapter à divers systèmes de restitution.

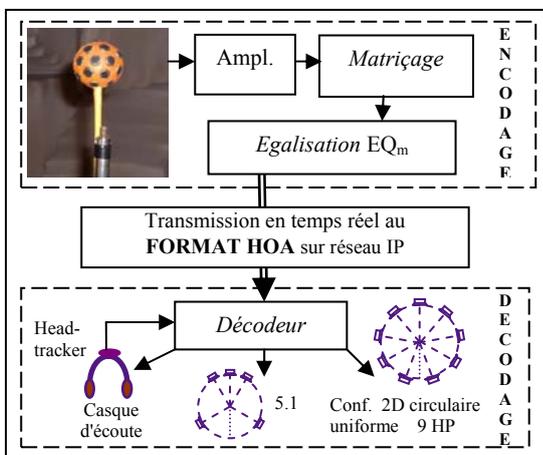


Figure 2 - Description de la plateforme temps réel

L'encodage du champ sonore 3D s'effectue grâce à un microphone pouvant estimer les signaux HOA jusqu'à l'ordre 4. Il est constitué de 32 capsules disposées à la surface d'une sphère rigide de rayon 2,6 cm. Les 32 signaux acquis sont matricés puis égalisés (section 3.2) de façon à obtenir un champ sonore encodé au format HOA indépendant du système microphonique.

Le décodage adapte la scène sonore encodée à divers systèmes de restitution : 5.1, configuration 2D circulaire uniforme de 9 HP, casque d'écoute (simulation binaurale d'une configuration 3D de 12, 20 ou 32 HP suivant l'ordre considéré) avec *head-tracker*, donnant au décodeur en temps réel les informations de position de tête.

### 4.2 Implantation

La technologie VST (*Virtual Studio Technology*) de Steinberg a été choisie pour l'implantation des différents

modules de la plateforme décrite ci-dessus. Cette technologie permet d'intégrer dans un grand nombre de logiciels audio professionnels et grand public (Cubase, Wavelab, Nuendo, etc.), sous forme de *plugins*, les traitements relatifs à la technologie HOA.

Quatre *plugins VST* ont été développés :

- Un premier concerne les traitements de matricage/égalisation des signaux microphoniques.
- Un *plugin serveur* permet ensuite l'envoi de la scène encodée sur le réseau vers une adresse IP.
- Un *plugin client* a en charge la réception, sur un ordinateur distant, de la scène transmise.
- Enfin, un dernier *plugin* effectue le décodage de la scène sonore encodée. Il prend comme paramètre la nature du dispositif de restitution choisi.

### 4.3 Transmission de la scène sonore 3D

La transmission de la scène sonore 3D encodée est une étape délicate puisqu'elle doit se faire en temps réel et concerne une quantité d'information importante : à l'ordre 4, 25 signaux HOA sont nécessaires en 3D.

Il peut donc s'avérer indispensable de réduire cette quantité d'information. Plusieurs stratégies sont étudiées. La première suppose que l'on connaisse à l'avance le dispositif de restitution et consiste à effectuer le décodage avant transmission, réduisant ainsi le nombre de canaux au nombre de haut-parleurs ou à deux pour le casque d'écoute. Il est également possible de réduire la résolution spatiale (ordre  $M < 4$ , restriction 3D → 2D). On peut enfin avoir recours à la compression audio (MP3, AAC).

Différents tests de transmission sur un LAN et sur Internet sont actuellement menés. Ils permettront d'évaluer, en fonction du débit disponible, l'impact de la réduction des informations transmises sur la qualité de restitution.

## 5 Conclusion

Un système d'acquisition, transmission et restitution en temps réel d'une scène sonore 3D a été développé et est actuellement en cours de test. Il est basé sur la technologie *High Order Ambisonics* dont l'intérêt réside dans la souplesse d'utilisation, notamment pour la restitution. Un certain nombre d'outils, encore inexistant pour cette technologie récente, ont été conçus pour mener à bien ce projet. Les applications possibles sont multiples : diffusion radiophonique, musicale, partage d'ambiance sonore, etc.

## Références

- [1] J. Daniel, R. Nicol, et S. Moreau. Further Investigations of Higher Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging, 114<sup>ème</sup> convention de l'AES, Amsterdam, 2003.
- [2] S. Moreau et J. Daniel. Study of Higher Order Ambisonic Microphone. Dans *Actes de la conférence CFA/DAGA'04*, Strasbourg, Mars 2004.

# Evaluation de systèmes d'encodage ambisonique aux ordres supérieurs.

Stéphanie Bertet<sup>1</sup>, Jérôme Daniel<sup>1</sup>, Laetitia Gros<sup>1</sup>, Etienne Parizet<sup>2</sup>, Olivier Warusfel<sup>3</sup>

<sup>1</sup> France Télécom R&D, 22300 Lannion, France, courriel : stephanie.bertet@francetelecom.com,

<sup>2</sup> LVA, INSA-Lyon, 69000 Lyon, France, courriel : parizet@lva-insa.fr

<sup>3</sup> IRCAM, 75004 Paris, France, courriel : warusfel@ircam.fr

## Introduction

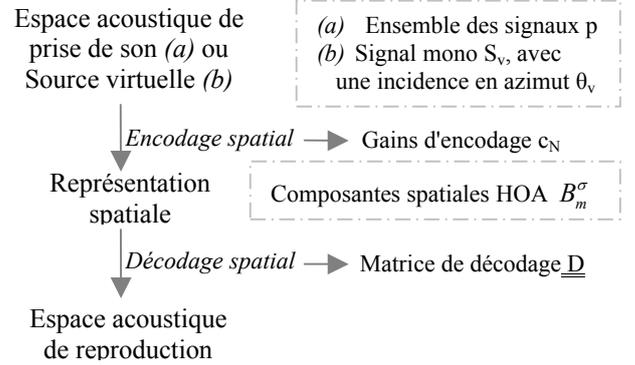
Différentes techniques de captation et reproduction d'une scène sonore 3D existent. Par exemple, pour la restitution stéréophonique frontale sur 2 haut-parleurs ou panoramique sur 5 haut-parleurs, il existe une variété de couples et arbres microphoniques qui leur sont dédiés. Nous nous intéressons ici à une approche plus « générique », dite « ambisonique », et plus généralement son extension aux ordres supérieurs, nommée « Higher Order Ambisonics » (HOA). Basée sur une décomposition du champ acoustique en « harmoniques sphériques », elle présente des propriétés intéressantes telles que: la scalabilité spatiale (la résolution spatiale de restitution peut être différente de celle d'enregistrement), la possibilité de manipuler le champ sonore avant diffusion, une flexibilité quant aux dispositifs d'écoute (restitution sur plusieurs configurations de haut-parleurs possibles, 5.1, restitution par casque). Le champ acoustique encodé peut provenir d'une synthèse de sources virtuelles ou d'un enregistrement avec un microphone HOA.

L'encodage HOA de champs sonores naturels (c'est-à-dire obtenus grâce à un enregistrement) est important pour constituer du contenu réel, réaliste, immersif. Il repose sur une prise de son par un réseau de microphones et ne peut réaliser que de façon approximative les formules d'encodage théoriques (par exemple il y a des artefacts spatiaux dû à l'espacement des capsules microphoniques). En 1993, le microphone Sound Field, ambisonique 1<sup>er</sup> ordre basé sur les théories de Gerzon, est commercialisé. Différentes études au sein de France Telecom R&D [1 2] ont abouti à 2 prototypes HOA, un prototype d'ordre 2, boule microphonique comprenant 12 capteurs et un prototype d'ordre 4, boule microphonique avec 32 capteurs.

Afin d'évaluer les effets de l'ordre ambisonique sur la résolution spatiale perçue et en particulier sur la localisation dans le plan horizontal, un test d'écoute est mis en œuvre.

## L'approche ambisonique

Le traitement du champ (ou de la source virtuelle) peut se décomposer en deux étapes: l'encodage et le décodage.



L'encodage ambisonique est basé sur une décomposition du champ acoustique en « harmoniques sphériques », ou décomposition de Fourier-Bessel. En 2D, elle peut être représentée par une décomposition en harmoniques cylindriques (1).

$$p(r, \theta_r) = J_0(kr) \cdot B_0^+ + \sum_{m=1}^M 2j^m J_m(kr) [B_m^+ \cos m\theta_r + B_m^- \sin m\theta_r] \quad (1)$$

où les  $J_m(kr)$  représentent les fonctions de Bessel.

En tronquant la décomposition à un ordre  $M$ , on retient un nombre fini de signaux qui constituent le format d'encodage spatial ambisonique. Le champ sonore ainsi encodé est décrit avec une résolution spatiale d'autant plus fine que l'ordre  $M$  est grand avec une zone de reproduction élargie en fonction de la fréquence. Il en est de même du champ sonore reproduit, moyennant un décodage spatial et une diffusion sur un dispositif de haut-parleurs appropriés.

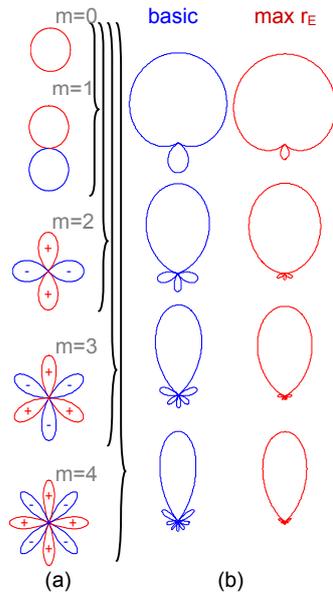
Soit une source virtuelle  $S_v$ , d'incidence  $\theta_v$ ; les composantes ambisoniques encodées jusqu'à l'ordre  $M$  sur le plan horizontal sont présentées équation 2.

$$B_m^\sigma \begin{cases} B_0^{+1} = 1 \cdot S_v \\ = \vdots \\ B_m^{+1} = \sqrt{2} \cos m\theta_v \cdot S_v \\ B_m^{-1} = \sqrt{2} \sin m\theta_v \cdot S_v \end{cases} \quad (2)$$

où  $B_0^\sigma$  représente la composante omni (ordre 0) et  $B_m^\sigma$  les composantes d'ordre supérieur  $m$ .

Pour la reproduction sur haut-parleurs, les signaux  $S_n$ , à diffuser sur les  $n$  haut-parleurs sont obtenus par application d'une matrice de décodage  $D$  sur le vecteur de composantes spatiales  $B_m^\sigma$  :  $\mathbf{S} = \mathbf{D} \cdot \mathbf{B}$  [3].

La matrice de décodage « basique » est composée de gains à appliquer aux signaux encodés. Le décodage  $\max r_E$  « concentre » les contributions énergétiques des HP dans la direction voulue. La Figure 1 (b) montre la reconstruction des lobes de directivité d'un microphone virtuel équivalent correspondant au décodage « basique » ou «  $\max r_E$  ».



**Figure 1.** (a) Directivités des composantes encodées d'ordre  $m=0$  à 4. (b) Directivités d'un microphone virtuel situé au centre du système de restitution, reconstruites après décodage basique ou  $\max r_E$  des composantes ambisoniques associées.

Pour une reconstruction par un système de restitution sur haut-parleurs, dans le plan azimutal, des composantes spatiales d'ordre  $M$ , un nombre minimum de  $N=2M+2$  haut-parleurs est recommandé.

Pour un enregistrement HOA, l'estimation des composantes spatiales repose en pratique sur une captation discrète du champ acoustique (i.e. par un réseau fini de microphones) associée à un traitement par une matrice de filtres. Cet échantillonnage spatial induit des artefacts d'encodage: d'une part l'aliasing spatial en haute fréquence, lié à l'écart entre capsules, d'autre part une directivité réduite (moindre résolution spatiale) en basse fréquence à cause des dimensions réduites de la boule microphonique [2]. Un compromis est donc à trouver sur la taille de la sphère.

## Ecoute spatialisée

### Notions de localisation

Les performances d'un système de reproduction 3D sont souvent évaluées à travers sa capacité à rendre l'effet de localisation d'une source sonore. Pour cela, on s'intéresse aux indices de localisation induits par une source sonore réelle. La position de la source est définie par son azimut, son élévation et sa distance dans un système de coordonnées sphériques centré sur la tête comprenant les plans médian, horizontal et frontal [4]. Une source placée sur le plan horizontal (hors du plan médian) sera à une distance différente entre l'oreille ipsilatérale et contralatérale provoquant des différences interaurales. Les différences

interaurales de temps (ITD), pour les basses fréquences jusqu'à 1500Hz [5], sont prépondérantes sur les différences interaurales de niveau (ILD) opérant plutôt en haute fréquence. Les sources placées sur le cône de symétrie autour de l'axe interaural, appelé cône de confusion ont un ITD identique qui peut provoquer des confusions avant-arrière [5]. Pour lever ces ambiguïtés, l'audition spatiale exploite les modifications du spectre induites par le pavillon de l'oreille, qui varient en fonction de l'incidence du son: ce sont les indices spectraux.

Le flou de localisation d'une source varie selon la nature de la source : contenu fréquentiel et durée, pour une source située à l'azimut 0 dans le plan horizontal, le flou de localisation varie approximativement de  $1^\circ$  à  $4^\circ$  [4] ; et selon sa position. Considérant les signaux large bande, plusieurs études [4 6] montrent une précision de localisation entre  $\pm 3^\circ$  et  $\pm 10^\circ$  que la source soit à  $0^\circ$  devant l'auditeur, à  $90^\circ$  ou à  $180^\circ$ .

### Reconstruction des indices de localisation avec des systèmes de haut-parleurs

La reproduction de scène sonore permet une reconstruction plus ou moins fidèle des indices de localisation. La dégradation (par rapport à une écoute naturelle) est quantifiable grâce à deux critères de qualité du système de reproduction, introduits par Gerzon, les vecteurs vitesse et énergie [6]. Notons que ces critères ne donnent pas un accès à la qualité de reconstruction des indices spectraux mais plutôt à celle des indices interauraux (ITD et ILD). Le module du vecteur énergie  $r_E$  ( $r_E \leq 1$ ) peut-être interprété comme un facteur de réduction de l'effet de latéralisation [7]. Il dépend de l'ordre  $M$ . L'angle  $\alpha_E = \arccos(r_E)$  caractérise le flou de la source virtuelle reconstruite [2].

### Le test d'évaluation

L'étude présentée participe à une investigation plus globale sur l'impact d'une résolution spatiale « objective » (caractérisée notamment par l'ordre ambisonique) sur une résolution spatiale subjective et plus généralement l'ensemble des qualités spatiales perçues.

Le présent test est élaboré pour évaluer la qualité d'encodage des systèmes de prise de son HOA.

### Les systèmes testés

Trois microphones ambisoniques ont été mesurés: le microphone Sound Field, dont les signaux sont encodés en B-format (ambisonique 1<sup>er</sup> ordre), le microphone 12 capsules (où les capsules sont positionnées en dodécaèdre sur une sphère semi-rigide de 7cm de diamètre), microphone ambisonique d'ordre 2 et le microphone 32 capsules (capsules positionnées en pentaki-dodécaèdre sur une sphère identique à celle utilisée pour le 12 capsules [2]) : microphone ambisonique d'ordre 4. Avec un traitement focalisé sur le champ horizontal (restitution 2D), la fréquence d'aliasing est autour de 7,5kHz pour le microphone 32 capsules et 4.4kHz pour le microphone 12 capsules.

Les réponses impulsionnelles des trois microphones ont été mesurées en chambre anéchoïque tous les  $5^\circ$  sur le plan azimutal [2]. Les réponses ont été interpolées tous les degrés pour avoir une précision de  $1^\circ$  pour le test, vérifiant que cette interpolation n'introduit pas d'artefacts sur la bande passante audible. Les composantes ambisoniques des systèmes HOA (ordre 2 et ordre 4) sont calculées avec les réponses des microphones mesurés. Grâce aux mesures, l'influence des capsules, les différences inter capsules et les erreurs de positionnement potentielles sur la sphère sont prises en compte pour une définition optimale dans l'encodage ambisonique [2]. Les signaux du microphone SoundField enregistrés sont encodés en B-format directement.

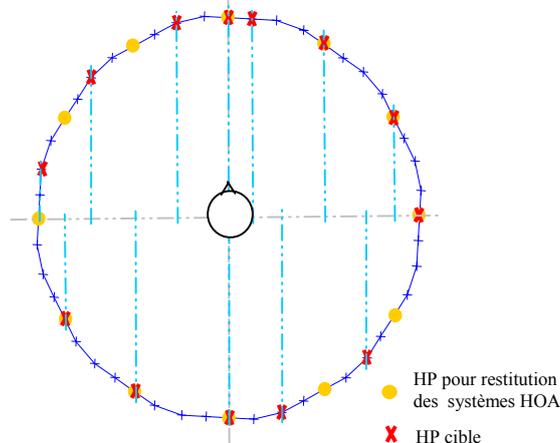
Afin de tester l'influence des capsules positionnées hors du plan horizontal pour une restitution 2D, un système d'ordre 3 a été encodé utilisant les 8 capsules du plan horizontal du microphone 32 capsules.

Un cinquième système testé, un encodage ambisonique « idéal » d'ordre 4 (simulation d'une source virtuelle) est ajouté aux systèmes de prise de son en tant que référence.

Le décodage utilisé pour tous les systèmes est une combinaison d'un décodage « basic » (jusqu'à la fréquence limite de reconstruction basse fréquence, qui croît avec l'ordre  $M$ ) et  $\max r_E$  (pour optimiser la localisation dans le domaine haute fréquence complémentaire). Dans cette étude, nous nous concentrons sur l'encodage, le décodage est le même pour tous les systèmes. Une fréquence de coupure, fixe, optimale pour l'ordre 1 est choisie.

## Le système de restitution

Un dodécagone de 1.5m de rayon est composé de 48 haut-parleurs Studer (4 haut-parleurs par côté) montré Figure 2. Chaque haut-parleur est séparé de  $7,5^\circ$ . Un rideau acoustiquement transparent cache le cercle à l'auditeur.



**Figure 2.** Système de restitution. Les 48 haut-parleurs sont représentés par les croix bleues. Les haut-parleurs "cible" sont représentés par les croix rouges et les haut-parleurs servant à la restitution ambisonique sont représentés par les ronds jaunes.

Les haut-parleurs ont été mesurés à la place de l'auditeur. Leurs réponses ont été déconvoluées aux signaux pour

compenser à la fois l'influence des haut-parleurs et la concentricité imparfaite des haut-parleurs.

12 haut-parleurs sont utilisés pour la restitution des systèmes ambisoniques afin de bénéficier d'une configuration régulière.

13 sources cibles sont choisies parmi les 48 haut-parleurs disponibles, privilégiant l'évaluation de la zone d'écoute frontale. Le placement des sources est montré Figure 2. 7 positions de cible sont situées sur des haut-parleurs utilisés pour la spatialisation ambisonique.

## Stimuli

Du bruit large bande est utilisé pour couvrir une plage fréquentielle sur laquelle l'ensemble des indices de localisation auditive sont sollicités.

Les réponses des cinq systèmes correspondant aux 360 positions possibles du pointeur sont convoluées avec un bruit uniformément masquant.

Le stimulus cible est un train d'impulsions de bruit blanc modulé en amplitude. Le stimulus cible et le stimulus du pointeur acoustique sont différents de manière à éviter l'appariement perceptif par distance spectrale et non par distance spatiale. Les deux sources sonores sont égalisées en niveau.

## La méthode de pointage

Dans la littérature, nous pouvons trouver plusieurs méthodes de report utilisées pour les tests de localisation. Wenzel et Wightman demande aux auditeurs de reporter oralement l'angle perçu de la source [8 9] ce qui nécessite un apprentissage très long. Seeber [11] utilise un pointeur laser qui nécessite l'aide de la vision, et par conséquent une évaluation uniquement frontale des systèmes. Afin d'éviter les biais causés par les méthodes « classiques » de report de localisation, une méthode d'appariement avec pointeur acoustique [3 10] est utilisée. Cependant, le pointeur n'est pas un haut-parleur physique à ajuster à la position de la source virtuelle [10]. Le pointeur est le système ambisonique à tester. Il s'ajuste à la position de la cible située sur un des haut-parleurs physiques du cercle grâce à un bouton rotatif. L'utilisation d'un encodage ambisonique comme pointeur permet d'avoir une résolution d'ajustement proche de l'audition ( $1^\circ$ ) et non une résolution physique limitée au placement des haut-parleurs.

## Protocole

Le sujet, placé au centre, déplace une source, spatialisée suivant l'un des modes d'encodage testés, (le pointeur) pour ajuster la direction perçue à celle d'une source réelle (la cible). L'ajustement se fait à l'aide d'un bouton rotatif que le sujet tient dans la main. Les signaux de rotation sont transmis numériquement puis interprétés depuis une interface développée sous Max MSP.

Les fichiers étant prétraités, le pointeur ne peut pas être déplacé pendant la durée du stimulus. La rotation du bouton rotatif est effective au prochain son joué. Le pointeur a donc une durée relativement courte, 150 ms, pour ne pas gêner

l'auditeur par l'impossibilité de bouger le son « en continu ». Les deux signaux (cible et pointeur) sont présentés consécutivement, après un silence de 50 ms. L'auditeur a 25 répétitions de la séquence cible – pointeur pour positionner le pointeur sur la cible. Si l'auditeur ajuste les deux sources avant la fin des 25 répétitions, il peut passer à la position suivante en appuyant sur un bouton déclenchant la séquence suivante. Les positions du son cible sont tirées aléatoirement parmi les 13 positions choisies. La position initiale du pointeur est choisie aléatoirement entre  $\pm 20^\circ$  et  $\pm 60^\circ$  autour de la cible. Chaque système est présenté 3 fois par position afin de valider les réponses de l'auditeur.

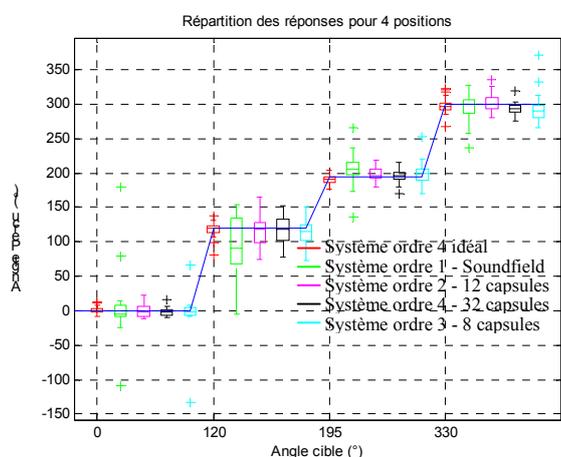
Le sujet est au milieu du cercle. La tête de l'auditeur n'est pas fixée mais on lui demande de s'orienter vers une marque indiquant le  $0^\circ$ . Il lui est demandé de ne pas détourner la tête en direction de la source. Une séquence d'apprentissage est effectuée présentant à l'auditeur un échantillon des systèmes testés.

## Déroulement du test

Les tests se déroulent à France Télécom R&D. A ce jour, 8 auditeurs, 7 hommes et 1 femme, âgés de 23 à 46 ans, ont passé le test.

## Analyse

Les premiers résultats sont en cours d'analyse. Leur tendance est conforme aux résultats attendus : le système d'ordre 4 « idéal » est le meilleur en terme de localisation, le système ordre 4 mettant en œuvre le microphone 32 capsules a de bons résultats avec une plus grande variabilité. Les angles perçus sont identiques aux positions de la cible à  $5^\circ$  près pour le système idéal d'ordre 4, pour les systèmes d'ordre 4 et d'ordre 2 (mettant en œuvre le microphone 12 capsules) à  $6^\circ$  près.



**Figure 2.** Résultats représentés par la médiane et les interquartiles à 25% et 75% des 5 systèmes respectivement : ordre 4 idéal, microphone SoundField (ordre 1), microphone 12 capsules (ordre 2), microphone 32 capsules (ordre 4) et 8 capsules (ordre 3), pour 4 positions de la cible :  $0^\circ$ ,  $120^\circ$ ,  $195^\circ$  et  $330^\circ$ .

Les résultats peuvent varier jusqu'à  $\pm 30^\circ$  par rapport à la position cible pour le système « idéal »,  $\pm 40^\circ$  pour le système d'ordre 4 « 32 capsules » et  $\pm 50^\circ$  pour le système

d'ordre 2. Les plus grandes erreurs d'angle ont été perçues avec le système d'ordre 1, mettant en œuvre le microphone SoundField (un angle moyen perçu de  $90^\circ$  pour une position de la cible à  $120^\circ$ ), avec une grande variabilité dans les résultats). Les résultats de 4 positions de la cible sont montrés Figure 2. Pour tous les systèmes un flou de localisation est visible pour les sources latérales. Quelques confusions avant arrière sont constatées.

Une analyse plus approfondie sur le temps de réponse des auditeurs, l'influence inter individu reste à finaliser. L'effet d'apprentissage est à évaluer.

## Conclusion

Le test met en œuvre le SoundField, les 2 prototypes HOA, d'ordre 2 et 4 et un système « idéal » avec une restitution par 12 haut-parleurs. Une tendance se dessine après les premiers résultats. On tend à observer que les performances de pointage (donc de localisation) s'améliorent avec la résolution d'encodage et lorsque l'encodage est dénué d'artefact (ordre 4 idéal).

On espère qu'une analyse plus complète permettra de corrélérer ces résultats à une caractérisation objective des systèmes [2].

## Références

- [1] Study of Higher Order Ambisonic Microphone. Moreau S. et Daniel J., CFA/DAGA, Strasbourg, 2004.
- [2] 3D Sound Field Recording with Higher Order Ambisonics – Objective Measurements and validation of a 4<sup>th</sup> order Spherical Microphone. Moreau S., Daniel J. et Bertet S., convention AES, Paris, 2006.
- [3] Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions. Daniel, J., Rault, J.B., and Polack, J.D. AES 105th Conv. 1998.
- [4] Spatial Hearing. Blauert J., 1999.
- [5] An introduction to the psychology of hearing, 4<sup>th</sup> edition. Moore B., 1997.
- [6] Two-dimensional sound localization by human listeners. Makous J. J. Acoust. Soc. Am. 1990, 2188-2200.
- [7] Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia. Daniel, J. Paris 6, 2001.
- [8] Headphone simulation of free field listening II : Psychophysical validation. Whithman. F. J. Acoust. Soc. Am. 1989, 868-878.
- [9] Localization using nonindividualized head-related transfer functions. Wenzel E. J. Acoust. Soc. Am. 1993, 111-123.
- [10] Localization of virtual Sources in Multichannel Audio Reproduction. Pulkki V. IEEE. 2005. 105-119.
- [11] A new method for localization studies. Seeber B. Acta Acustica. 1997.

Lidwine Hô

France télévisions - innovations & développements



francetélévisions

Hervé Dejardin

Radio France Innovation



<http://www.bili-project.org/tutoriel-production-audio-pour-diffuser-sur-youtube-360/>

# TUTORIEL

Production audio pour  
diffusion Youtube 360°



Novembre 2016