

# MEMOIRE 3EME ANNEE

Année 2003-2004

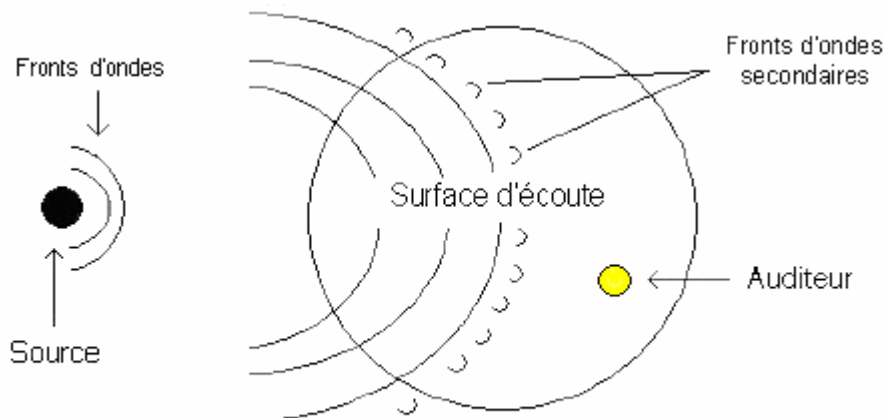
## LE SYSTEME DE DIFFUSION HOLOPHONIQUE Technique Wave Field Synthesis

Matthieu Berjon

berjonmatthieu@hotmail.com

ESRA Bretagne Section ISTS

Directeur du mémoire : Gwendal Poiron



Rayonnement d'une source en système holophonique

## Remerciements

Je tenais à remercier tout particulièrement **Gwendal Poiron** pour m'avoir orienté et motivé dans mon projet et m'a permis de le mener jusqu'au bout, ainsi que ses cours qui m'ont donné une nouvelle ouverture sur la diffusion sonore.

Je tenais également à remercier **Terence Caulkins** et **Etienne Corteel** du bureau 105 pour avoir pris le temps de me fournir de la documentation et répondu à mes questions avec sympathie.

A toute **l'équipe Production** de l'IRCAM, pour le stage que j'ai pu suivre, leur bonne humeur, et l'attention portée à mon travail.

Je voulais également remercier **Alexis Baskind** pour son éclaircissement sur l'EDR et sa démonstration de Linux.

Et enfin à mes parents et ma sœur pour leur soutien ainsi que leur aide dans ce projet qui espérons-le se poursuivra...

# **Table des matières**

## **1 Introduction**

## **2 Historique de la diffusion sonore**

## **3 Principe de fonctionnement**

- 3.1 Définition
- 3.2 Principe de Huygens, formulation de Kirchoff-Helmholtz
- 3.3 La Wave Field Synthesis
- 3.4 Les types de sources

## **4 Projet CARROUSO**

- 4.1 Description
- 4.2 Objectifs
- 4.3 Mise en œuvre
- 4.4 Conclusion

## **5 Mise en œuvre**

- 5.1 Les transducteurs et la multi égalisation
- 5.2 Les signaux, comment, pourquoi.
- 5.3 Conclusion

## **6 Phénomènes Indésirables**

- 6.1 La phase stationnaire
- 6.2 Fenêtrage et diffraction
- 6.3 L'aliasing spatial

## **7 Prise de son/Mixage d'une scène sonore**

- 7.1 La technique
- 7.2 Notion de distance holophonique
- 7.3 Prise de son
- 7.4 L'artistique

## **8 Avenir de l'holophonie**

- 9.1 Développements
- 9.2 Mutations

**Annexe A** Le MPEG-4

**Annexe B** La réverbération EDR

**Annexe C** Le projet LISTEN

**Sources/Glossaire**

**Bibliographie**

# **1 Introduction**

Depuis quelques temps nous observons l'émergence de différents formats de multi diffusion surtout pour le cinéma. Leurs apparitions ont suscité également un engouement grandissant chez les particuliers. Pourtant la multi diffusion n'en est qu'à ses prémices et évolue constamment. Les laboratoires travaillant toujours au développement de nouveaux produits. Les techniques de diffusion comme le binaural, l'Ambisonic ou même le transaural pourront voir le jour dans peu de temps sur le marché grand public. Nous nous concentrerons ici sur une technique particulière dont les possibilités sont énormes : l'holophonie. Les formats actuels de diffusion multi haut-parleurs ne donnent qu'une sensation d'immersion comme la diffusion 5.1. Ce système est très adapté au son à l'image puisque l'image n'est diffusée que sur un plan, le spectateur ne devant donc pas se détourner de celui ci.

Le principe de l'holophonie repose sur l'idée d'une reconstruction totale du champ sonore dans une zone de diffusion. Même sur une zone d'écoute étendue, on garde cette cohérence, ce qui permet d'avoir une reproduction correcte en tout point de cette zone. Ceci est un avantage majeur par rapport aux systèmes « classiques » stéréo ou à multi haut-parleurs ne fonctionnant pas sur un principe de reconstruction du champ sonore.

Depuis sa découverte de nombreuses études ont été menées par l'université de Delft ou encore par l'**IRCAM** (Institut de Recherche et Coordination Acoustique-Musique) et l'**IST** (Information Society Technologies) qui a lancé le projet **CARROUSO** (Creating, Assessing and Rendering in Real-time Of High-quality aUdio-viSual envirOnements). France Télécom mène également des recherches sur l'installation de systèmes de visioconférences.

Dans ces pages, nous nous tenterons d'expliquer le principe même du fonctionnement de l'holophonie au travers d'une technique de diffusion dite WFS (Wave Field Synthesis). De décrire les organismes qui travaillent dessus, leurs avancées technologiques sur ce système et d'expliquer sa mise en œuvre dans sa chaîne de fabrication.

Dans une deuxième partie, nous tâcherons de voir les inconvénients même du système, de voir les différentes phases du mixage tant du point de vue technique (logiciels, hardwares...), que d'un point de vue artistique (détournement du système, intention dans le résultat...).

Dans une troisième et dernière partie, nous verrons sa possible adaptation dans un environnement autre qu'un laboratoire, comme dans un espace de spectacle.

Enfin nous ferons un bilan sur les avancées technologiques, le développement du système, et les mutations qui peuvent survenir dans les années à venir.

## **2 Historique de la diffusion sonore**

Le son est présent depuis que l'humanité existe. D'abord utilisé surtout comme moyen de communication, il se diffusait par moyen mécanique comme les cordes vocales ou tout objet émettant des ondes acoustiques. C'est Pythagore, en 500 avant Jésus-Christ qui a réalisé les premières expériences sur le son, en mettant en relation la longueur d'une corde, la tension appliquée sur celle-ci et la note qu'elle produit de ce fait. Puis Vers 330 avant JC, Aristote s'intéressa aux phénomènes d'échos qui ont permis de dégager les principes de l'acoustique architecturale.

En 1630, Marin Mersenne est le premier à vouloir déterminer la vitesse du son. Il utilise pour cela un canon comme source sonore et lumineuse et détermine approximativement la vitesse à 450 m/s. En 1710, un trompettiste du nom de John Shore invente le diapason accordé à 440 Hz qui deviendra le standard de normalisation des fréquences. En 1852, **Corti** étudie la structure interne de l'oreille ainsi que son fonctionnement. En 1866, Kundt étudie les résonances des sons dans les tuyaux et en sort une formule pour le calcul des ondes stationnaires. 1877, **Thomas Edison** dépose le brevet du phonographe.

C'est en 1900 que l'acoustique architecturale se développe grâce aux études menées par un certain **Wallace Sabine**. En 1918, **Langevin** utilise les ultrasons pour la mesure de distance et fait naître en outre le sonar. En 1922, on voit l'arrivée du premier haut parleur à rendement élevé grâce à Maurice Guéritot. Puis en 1927 c'est le premier film parlant diffusé en mono appelé « The jazz Singer ». En 1930, l'enregistrement optique arrive et prend essor dans les salles de cinéma. En 1941, c'est l'arrivée de la stéréo dans les salles obscures avec Fantasia des studios Disney. 1950, c'est le début des bandes son magnétiques qui permettent d'utiliser de 4 à 6 canaux.

En 1958, le disque 33 tours se lance aussi dans la stéréo, puis en 1960 c'est la radio qui passe à la stéréo et à la diffusion *FM* (Fréquence Modulation). En 87, on voit apparaître dans les foyers le Dolby Surround Prologic utilisé déjà dix ans auparavant dans les salles de cinéma. Puis en 1992 on passe au tout numérique avec le Dolby Digital au format 5.1.

Depuis plusieurs années, on s'intéresse à la diffusion multi canal avec une prépondérance dans le marché grand public pour les systèmes 5.1, chacun essayant d'imposer sa technique de diffusion comme standard. Des systèmes comme le binaural, le transaural, l'Ambisonic ou même l'holophonie sont des techniques basées sur la reproduction des champs acoustique permettant un réalisme et des capacités bien plus grandes que le 5.1, mais qui demandent encore des améliorations, des recherches pour pouvoir sortir un produit de qualité et répondant aux besoins du consommateur. On peut donc imaginer dans plusieurs années ces techniques émerger dans le broadcast et même sur le marché grand public, compléter, voire même remplacer nos systèmes actuels stéréo ou multi canaux.

De nos jours, le grand public reste sur les standards qui lui sont offert, avec le développement du 5.1. Pourtant les nouveaux types de formats qui arrivent peuvent vite détrôner ces systèmes surtout basés sur le marché audiovisuel. Le transaural par exemple ne nécessite aucun décodage spécial et peut être lu sur une chaîne hifi classique permettant une retranscription de la scène sur deux haut-parleurs (stéréophonie). Cela entraîne un coût moins élevé permettant alors d'investir dans une écoute de meilleure qualité plutôt que de développer un système 5.1 coûteux n'apportant pas une reproduction satisfaisante. L'Ambisonic quant à lui permet une souplesse extraordinaire puisque l'on peut diffuser la scène sonore sur n'importe quel type de système d'où une adaptabilité et une économie sur le matériel (réactualisation du système en fonction du format de diffusion). L'holophonie, par la technique de la Wave Field Synthesis (développer dans ce mémoire) est un système trop important et coûteux pour le marché grand public mais à une bonne base pour être développé dans le broadcast et surtout pour le spectacle vivant.

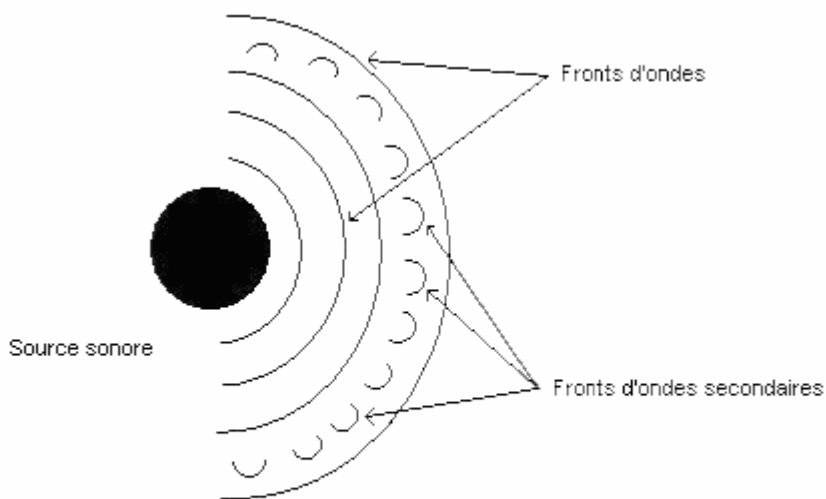
### **3 Principe de fonctionnement**

## 1. Définition

L'holophonie est un principe de diffusion émis au 17<sup>ème</sup> siècle par le biais de la géométrie. Il s'agit de reconstruire un champ acoustique en tout point d'une zone d'écoute. Pour expliquer le résultat de l'holophonie, on pourrait l'assimiler à un hologramme (reproduction visuelle en trois dimensions d'un objet quelconque), l'effet final étant sonore au lieu d'être visuel. Le but de l'holophonie est donc de recréer une scène sonore en trois dimensions dans un espace « objet », ce dernier étant délimité par le système de diffusion lui-même. L'intérêt de ce type de diffusion est que l'auditeur peut évoluer dans l'espace d'écoute tout en localisant toujours parfaitement les sources sonores qu'elles soient fixes ou en mouvements.

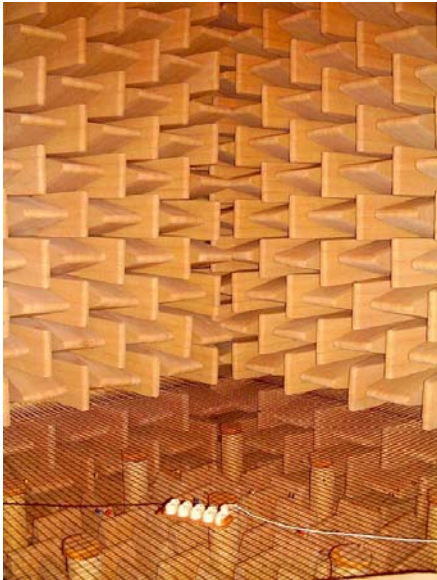
## 2. Principe de Huygens, formulation de Kirchoff-Helmholtz

Pour arriver à ce résultat **Huygens** a émis le principe qu'une onde se propageant dans une certaine direction peut être émise par une multitude de sources secondaires réparties sur cette onde.



### PRINCIPE DE HUYGENS

**Kirchoff** et **Helmholtz** ont tiré de ce principe une intégrale, disant que pour calculer un point dans la zone d'écoute, il faut connaître la pression en tout point de la surface délimitant cette zone d'écoute (système de diffusion). Sur le papier la reconstruction est idéale voire parfaite, mais sa mise en œuvre pratique comporte des problèmes non négligeables. La taille et le nombre de transducteurs, ainsi que l'importance de l'acoustique de la salle à cause des réflexions induites par les parois de celle-ci qui viendront perturber la reconstruction de ce champ acoustique. Cela suppose donc de travailler en *chambre anéchoïque*.

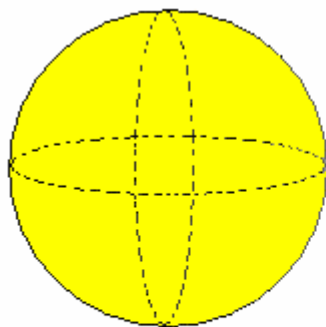


## Chambre Anéchoïque TU de Delft

### 3. La Wave Field Synthesis

On a pu voir naître, grâce à **Huygens** et Fresnel, puis grâce à **Kirchoff** et **Helmholtz**, une théorie sur la diffusion sonore à trois dimensions. Ce système de diffusion dit « holophonique » travaille donc sur une diffusion de type sphérique. De ce fait les calculs et le matériel à mettre en œuvre deviennent vite impressionnants et surtout très lourds. La première application du système holophonique a été réalisée par **Berkhout** et Al en 1988 avec la technique de la *WFS* (Wave Field Synthesis). On peut traduire Wave Field Synthesis par synthèse d'un champ d'onde puisque le principe même de cette technique est de recréer un champ acoustique par la décomposition des ondes primaires. **Berkhout** et Al ont donc améliorés le principe holophonique de départ en prenant en compte les fronts d'ondes des premières réflexions (early reflexions). Par cette technique, on peut déplacer la source, modifier les caractéristiques du lieu d'écoute virtuel, le tout indépendamment les uns des autres. De plus la restitution est valable dans tout l'espace d'écoute et non pour un seul point ou du moins une zone d'écoute restreinte comme dans la plupart des autres systèmes (mono, stéréo, 5.1....).

Ce principe de diffusion implique une simplification du système de base. Il faut donc déjà ramener le système dans un plan et non plus travailler dans l'espace.



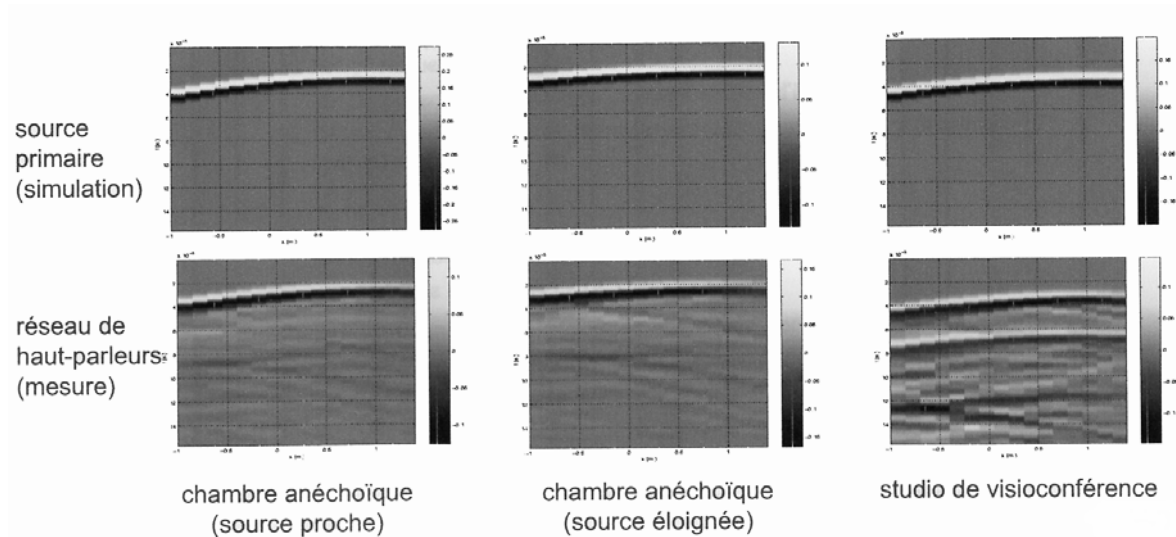
holophonie  
diffusion sur sphère



Wave Field Synthesis  
diffusion sur un plan

Principe de diffusion holophonique et WFS

On retrouve dans ce système des similitudes avec le binaural, puisqu'il faut injecter au signal enregistré une égalisation pour les haut-parleurs et une correction de la phase stationnaire pour corriger l'acoustique de la salle, le tout, sur chaque haut-parleur (chapitre 5). Pour effectuer ce type d'opération en temps réel on utilise des unités élémentaires de traitement du signal (DSP). Dans ses travaux sur la WFS, Rozenn Nicol a mis en évidence l'importance des réflexions de la salle qui viennent perturber le champ sonore généré rendant difficile sa mise en oeuvre hors d'une *chambre anéchoïque*.



**Simulation et mesures de la WFS à la fréquence de coupure  
FC=500Hz**

#### 4. Les types de sources

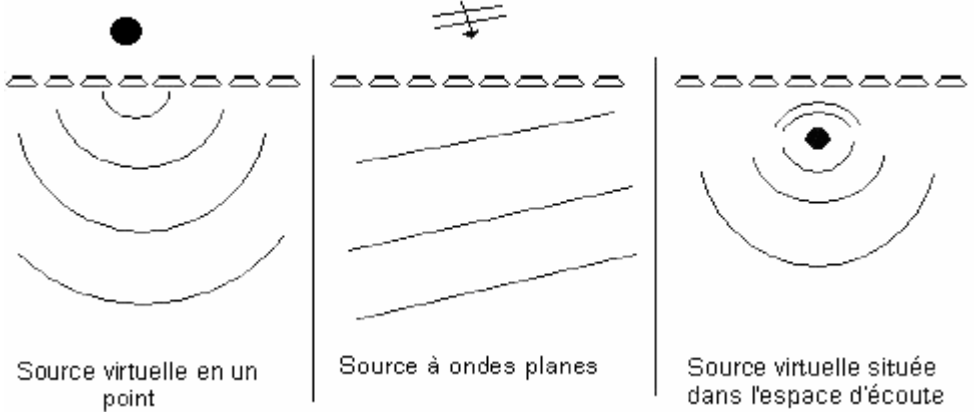
En menant plus loin les recherches sur la technique de la Wave Field Synthesis, on a dégagé plusieurs types de sources distincts. Elles sont au nombre de trois, leurs différences majeures étant la propagation de leur signal, plus précisément la modification de leurs fronts d'ondes.

Le premier est la source virtuelle dite « source non focalisée ». Elle est synthétisée en dehors de l'espace d'écoute et est perçue par l'auditeur avec une assez grande précision ainsi qu'une définition correcte du volume de la source, ceux-ci étant valable n'importe où l'auditeur évolue dans l'espace d'écoute.

Il y a également la « source à ondes planes ». Elle se situe également derrière le banc de haut-parleurs mais avec des valeurs de distances beaucoup plus élevées (voire une distance apparemment infinie). Par rapport à la pièce d'écoute, elle se trouve à l'extérieur et donc n'a pas d'équivalent acoustique dans le « monde réel ». De ce fait on a l'impression d'entendre la source avec la même intensité. La source nous suit en permanence dans l'espace d'écoute tandis que les autres sources, quant à elles, vont se déplacer. Ce type de source garde toujours sa direction angulaire quelque soit le mouvement que l'on effectue en tant qu'auditeur. L'exemple pris du soleil dans [4] reflète très bien l'idée d'une source à onde plane.

Le troisième type de source est celle simulée à l'intérieur de l'espace d'écoute dite « source focalisée ». Elle est réalisée grâce à une extension du principe de la WFS. On arrive à synthétiser une source dans l'espace d'écoute sans qu'il n'y ait pour cela une source physique. Pour réussir à recréer cette source, on fait converger plusieurs fronts d'ondes par le

biais du banc de *monopôles*. De ce fait le front d'onde se retrouve réémis dans le sens opposé et recrée la source sonore en forme et en taille. Les études montrent que le champ acoustique entre le banc de monopôles et la source cible reste imprécis mais pas au niveau de la source cible.



## **4 Projet CARROUSO**

## 1. Description

Le projet CARROUSO (Creating, Assessing and Rendering in Real-time Of high-quality aUdio-viSual enviroNements in *MPEG-4* context) a été initié par l'IST, piloté par le **Fraunhofer Institut** et financé par la communauté européenne.

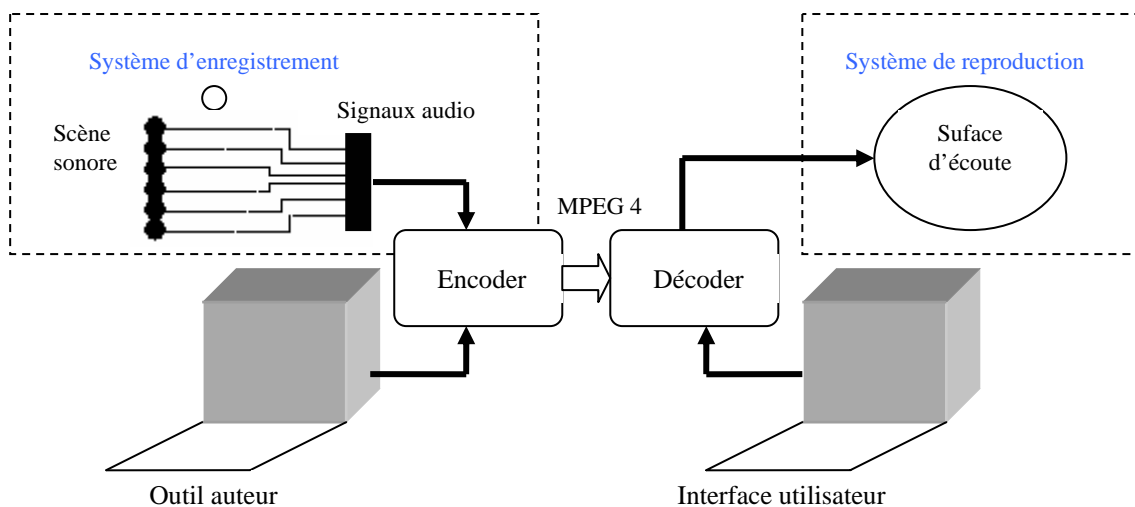
Le standard MPEG-4 (annexe I) a été redéfini plusieurs fois depuis son lancement en 1995. Une première version a vu le jour en 1999 puis une seconde mouture a été sortie en 2000.

Ce projet vise à développer un système de transmission sur réseaux et de restitution de scène holophonique (au format WFS) le tout encodé au format MPEG-4.

Le projet était de spécifier, développer, implémenter et valider une nouvelle technologie. Elle permet de transférer un champ sonore, et de générer un certain espace réel ou virtuel vers un autre espace, le tout en ayant un contrôle total et interactif de ses propriétés temporelles et spatiales. Ceci peut être associé à de la transmission de données visuelles (vidéo).

Il fallait également spécifier, développer, implémenter et valider une nouvelle modélisation, un nouveau procédé d'enregistrement, ainsi qu'une nouvelle technique d'encodage/décodage et interprétation du signal audio.

Et enfin en construire une démonstration en temps réel. Et enfin définir et contrôler des niveaux de performances de la technologie par des expériences perceptuelles.



**Projet CARROUSO et MPEG 4**

## 2. Objectifs

La partie principale du travail était de spécifier et développer un système qui peut encoder, transmettre, décoder et interpréter un champ sonore d'une manière efficace, avec une haute qualité perceptive. Le composant clé pour l'enregistrement du champ sonore se situe au niveau de la technologie utilisée pour le microphone. Le traitement de signal calcule la position des sources sonores, fixes ou en mouvements. L'ensemble des microphones ainsi que les caméras rassemblent des informations appropriées sur les conditions acoustiques de la salle d'enregistrement. Des modèles acoustiques ont été développés pour permettre de paramétrer l'ensemble des données acoustiques, les rendant ainsi aptes à la transmission.

L'encodage et le décodage utilisent un sous ensemble de composants dérivés du standard MPEG-4. Cette sélection de sous ensemble est nécessaire en fonction du manque de profil prédéfini pour les applications données dans le standard. Pour la diffusion, la transmission des données a adopté les courants de diffusion vidéo numérique.

Le procédé d'interprétation permet la reproduction d'un champ sonore enregistré ou simulé via un nombre variable d'enceintes.

Le système de pilotage des enceintes a été choisi de sorte que l'acoustique originale de la salle de reproduction, qui peut être influencera le résultat obtenu, soit corrigé de façon optimale.

Tous ces composants sont combinés pour un testeur, base pour la validation de la totalité du système. Des expériences perceptuelles et des champs de tests ont servis à la validation.

Chaque source encodée prend en compte plusieurs paramètres, comme les paramètres géométriques de la salle d'enregistrement et des murs. L'effet de salle est associé à une position dans cette dernière. Enfin l'encodage de descriptions des scènes est faite de façon perceptive et psycho acoustique.

Durant la mise en place du projet CARROUSO plusieurs alternatives ont été proposées. La première étant une description physique des paramètres basés sur les mesures des réponses impulsionnelles des salles par rapport à des positions de sources. Ces réponses de salles étant mesurées par une ligne circulaire de microphones cardioïdes. Le décodage devant se faire en recombinaison des signaux enregistrés (en une forme de microphone hyper cardioïdes), permettant l'extraction des réponses impulsionnelles en une profusion de directions élémentaires.

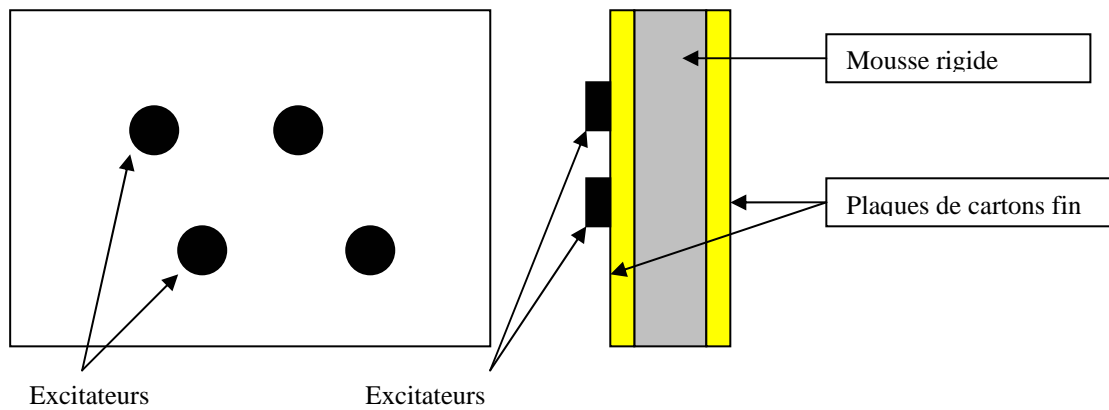
### 3. Mise en œuvre

La mise en œuvre du projet CARROUSO s'est déroulée en plusieurs étapes au sein de l'IRCAM. Ces étapes sont au nombre de 5 :

- L'égalisation des haut-parleurs
- La validation de la WFS dans les outils de spatialisation
- L'intégration de l'holophonie dans le Spatialisateur
- La définition d'un outil et d'une interface utilisateur
- L'extraction des paramètres perceptifs à partir de la réponse d'une salle

#### **Egalisation des haut-parleurs**

Dans le cadre de l'utilisation de l'holophonie dans le projet CARROUSO, ce sont des *haut-parleurs M.A.P* (Multi-Actuator Panel) qui ont été utilisés. Ce système utilise en général 4 ou 8 excitateurs, tous alimentés indépendamment, répartis suivant une ligne horizontale et collés sur une plaque servant de membrane.



### Haut-parleurs Multi Actuator Panels

La procédure mise en place pour égaliser ce type de système assure la reproduction du front d'onde d'une source virtuelle et optimise la qualité du champ acoustique reproduit (en phase et en fréquence) dans toute la zone d'écoute. Le mode de reproduction implique la constitution d'une base de données de différents filtres pour telle et telle position de source virtuelle ou de direction d'onde plane. Ces filtres seront implantés en amont de chaque exciteur et seront effectués en temps réel grâce à des processeurs de convolutions temps réel. Il faut savoir que la bande passante de chaque exciteur est de 150 Hz à 16000 Hz. Cela est suffisant puisqu'une source descendant en dessous de 120/150 Hz est quasi omnidirectionnelle et de ce fait donc quasi impossible à replacer dans un espace quel qu'il soit.

### L'intégration de la Wave Field Synthesis dans les outils de spatialisation

Pour la reproduction d'une scène sonore il faut pouvoir synthétiser un front d'onde pour une position donnée de la source virtuelle. L'effet de salle quant à lui est reproduit par des ondes planes suivant un nombre de directions angulaires limitées.

Au niveau du Spatialisateur, on utilise des *encodeurs directionnels* existants. Les différents signaux ainsi que le son direct sont transmis après traitement à des processeurs dédiés. Tout un patch *DSP* et une interface de contrôle ont été développés pour commander des modules à distances (via des ordinateurs Macintosh ou PC).

### Définition de l'outil auteur

Dans le projet CARROUSO, la définition de l'outil auteur doit se faire de façon descriptive pour permettre une interaction de l'auditeur dans la construction de la scène sonore. Le standard MPEG-4 permet de séparer complètement le codage de chaque signal et la description paramétrique de la scène. Pour créer une scène sonore, le besoin du développement d'un outil auteur a donné naissance au projet LISTEN (annexe D). Pour le mettre en place il a fallu appliquer des fonctionnalités compatibles avec le format MPEG-4. Cela veut dire que dans son développement, à chaque source virtuelle, on associe un flux audio, qu'il soit issu d'un fichier son ou d'une prise de son effectuée en temps réel, ainsi qu'un ensemble de renseignements sur la qualité acoustique de la salle. Cette scène est alors transformée en format binaire et multiplexée avec les flux audio compressés (communication UDP en format textuel à un encodeur MPEG-4). L'avantage de ce type d'encodage est que le format de restitution peut être choisi par l'utilisateur puisque la Wave Field Synthesis ne fait pas partie de la norme (rendu *Ambisonic*, binaural...).

## Extraction des paramètres perceptifs

Enfin il reste l'extraction des paramètres perceptifs. L'IRCAM se sert d'un outil de description de la réverbération nommé *EDR* qui est intégré au Spatialisateur. Cette librairie est le point de départ pour la retranscription des effets de salles dans le projet CARROUSO. Le but est de reprendre cette librairie, de la réactualiser et de faire une nouvelle interface pouvant être intégrée à l'interface utilisateur. L'avantage de cette librairie est d'être totalement modulaire permettant ainsi de rajouter d'autres outils de mesure. Le tout doit être couplé à la librairie de *MAX/MSP* fournissant un outil de mesure complet.

## 4. Conclusion

Le projet CARROUSO avait pour but le développement d'un équipement de prise de son dédié à l'holophonie pour le format WFS, d'où construction du système de prise de son et des algorithmes de calcul. Il fallait également développer un outil permettant d'éditer les paramètres pour modifier la scène sonore encodée. Le tout développé dans des formats audio et vidéo (lorsqu'il y a association à de l'image) standards étant sur le marché (en l'occurrence la norme MPEG-4). Le projet a été mené à bien avec l'apparition de ListenSpace et des algorithmes de calcul poussés et performants. Depuis l'université de Delft a développé un système de prise de son performant fonctionnant sur la décomposition du champ acoustique (chapitre 7). CARROUSO permet désormais (dans le cadre encore de l'étude en laboratoire) de travailler la matière sonore en 3 dimensions sur haut-parleurs et de poursuivre les recherches dans la réalité augmentée.

## **5 Mise en œuvre**

Dans le cadre de cette étude de la mise en œuvre d'un système basé sur la Wave Field Synthesis, nous resterons dans le cadre de la diffusion et non de l'encodage ou de la prise de son de la scène sonore elle-même. Cette étude sera bien entendu dans le cadre du projet CARROUSO et du MPEG-4 puisqu'il représente le standard d'encodage et de diffusion pour ce type de technique.

Dans un premier temps il faut étudier les différentes parties matérielles qui constitue la chaîne du signal, puis dans le chapitre suivant, de voir dans son ensemble l'installation par un exemple concret.

## 1. Les transducteurs et la multi égalisation.

Nous avons vu plus haut que pour un système holophonique il fallait que nos haut-parleurs représentent un point dans l'espace et qu'ils soient une infinité. Dans la phase pratique du système on voit bien l'irréalité de la chose. Nous avons également vu la mise au point de « plaques » de haut-parleurs nommés « M.A.P » qui permettent une cohérence plus grande dans la reconstitution du champ acoustique. On a vu dans le chapitre 4 que du fait que les haut-parleurs ont une dimension finie, cela crée un effet d'aliasing (chapitre 6). Il faut savoir que tout système de diffusion, pour pouvoir fonctionner de façon optimale, a besoin d'être égalisé. Le souci est que chaque transducteur à une directivité variant avec la fréquence. On doit avoir bien sûr des transducteurs théoriquement identiques, de ce fait on peut appliquer un filtre individuel en amont de chaque haut-parleur sachant que chaque filtre à appliquer va dépendre de la position de la source dans l'espace. Etienne Corteel a proposé une méthode d'égalisation globale du système de diffusion. Le gros avantage de cette égalisation est le fait que l'on ne tienne plus compte de la directivité des haut-parleurs ainsi que de la forme globale du système d'écoute. A contrario le phénomène d'aliasing va venir perturber la multi égalisation dans les hautes fréquences d'où le besoin de faire intervenir un égalisation individuelle.



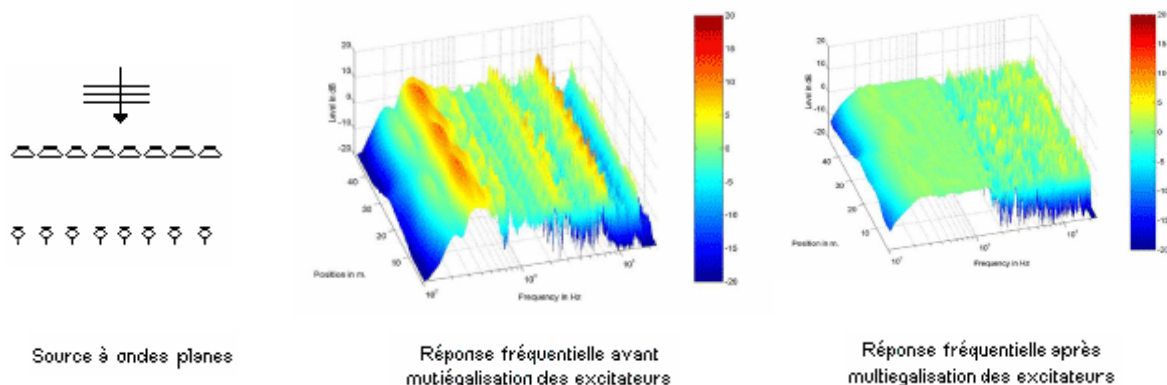
Vue de Face



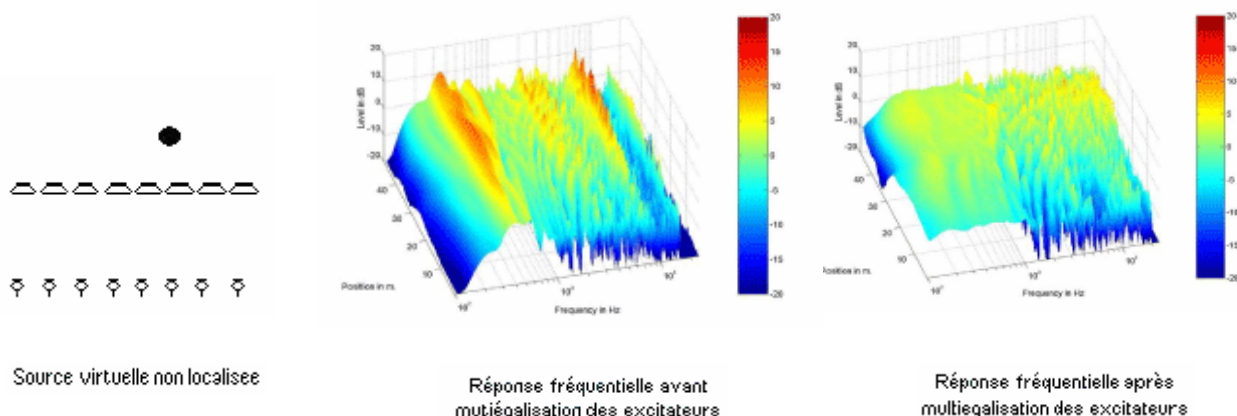
Vue de dos

### **Haut parleur MAP fabriqué à la TU de Delft**

Pour réaliser cette égalisation, on mesure le champ sonore grâce à un banc de microphones. On mesure la réponse impulsionnelle de chaque haut-parleur. Puis on mesure les coefficients à appliquer pour une position de source. En fait, on cherche à obtenir le coefficient des filtres à appliquer au banc de haut-parleurs en tenant compte de la réponse impulsionnelle obtenue par mesure et celle que l'on doit obtenir idéalement. Il n'est pas possible à l'heure actuelle d'obtenir un résultat « parfait » de ce fait on applique un algorithme de façon itérative qui permet de diminuer l'erreur en fonction du niveau d'acceptation défini en amont.



Multiégalisation d'une source à ondes planes sur Haut-parleurs M.A.P



Multiégalisation d'une source non focalisée sur Haut-parleurs M.A.P

Bien sûr ce type d'égalisation possède un certain nombre de limitations. Le système de la WFS est conçu dans un plan et non dans l'espace, de ce fait on ne peut appliquer le traitement d'égalisation que dans ce plan. En dehors de ce dernier, l'égalisation est complètement aléatoire voire même inutile. Enfin du fait du nombre fini de transducteurs, il se crée le phénomène d'aliasing (chapitre 6). A partir de la fréquence d'aliasing, ce système d'égalisation ne crée que des solutions locales, au niveau de la position des microphones de mesures, en dehors de ces points de mesure le champ sonore devient imprévisible.

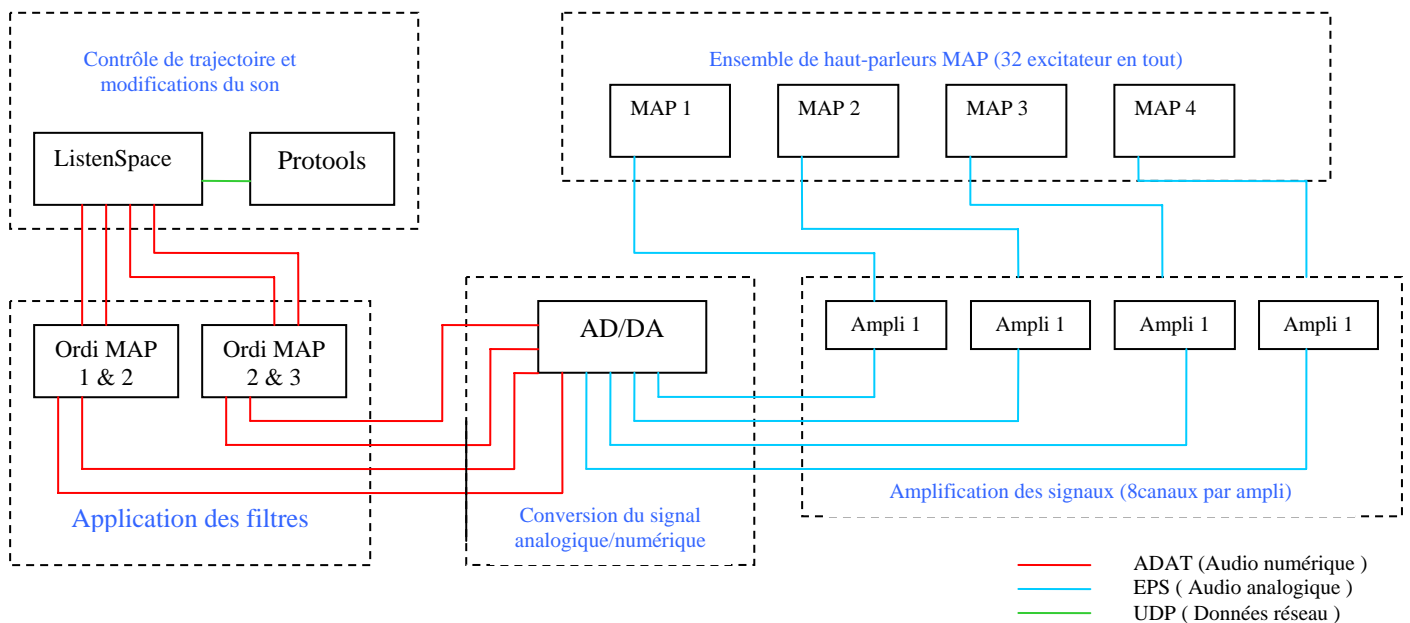
Enfin au niveau du traitement dans les hautes fréquences on effectue une égalisation individuelle. On a vu qu'à partir de la fréquence d'aliasing, le champ sonore devient non prédictible, par conséquent on effectue une égalisation par bande de fréquence en fonction de la position de la source. Pour réussir complètement l'égalisation à partir de la bande de fréquence on effectue également un contrôle énergétique du champ sonore qui va être

reproduit. Ceci permettra de garder une cohérence en terme de perception des sources et non en terme d'exactitude physique.

## 2. Les signaux, comment, pourquoi.

On a montré que la technique de la WFS impliquait la prise en compte d'un grand nombre de transducteurs (signaux) mais pas nécessairement d'un grand nombre de sources. Il faut bien distinguer ces notions de source et de signal. Dans l'exemple qui suit nous allons prendre exemple d'un système mis en place à l'IRCAM, permettant de bien voir les différents signaux transitant entre les différentes machines, ainsi que le trajet qu'ils effectuent.

Tout d'abord on voit grâce au schéma ci-dessous les différents éléments entrant en jeu. Au travers de ces machines naviguent trois types de signaux principaux servant à la fabrication d'une scène sonore.



### MISE EN ŒUVRE D'UN SYSTEME WFS (diffusion à 32 haut-parleurs/4MAP)

Le système mis en place a comme matériel un macintosh pour le contrôle d'automatisation sur protocols, un PC pour le contrôle du ListenSpace, deux ordinateurs macintosh pour l'application des filtres, un convertisseur numérique/analogique (minimum obligatoire), 4 amplificateurs 8 voies et 4 haut-parleurs M.A.P. Entre le protocols et le ListenSpace ce sont des données qui transitent par le réseau en UDP. Ce sont des messages OSC (Open Sound Control) qui s'échangent entre les machines permettant ainsi de pouvoir faire des automatisations sur Protocols et qu'elles agissent sur les fichiers son contenus dans le PC. Ensuite les flux audio sont envoyés en numérique (connectique ADAT) sur les 2 autres macintosh qui se chargent d'appliquer en temps réel les filtres en fonction du type de source et de sa position dans la scène. Les signaux filtrés circulent toujours en ADAT vers le convertisseur permettant le passage des signaux en analogique. Nos signaux une fois dans le convertisseur, se trouvent dispatcher sur des multi paires analogiques (type EP5) vers les amplis 8 voies. Les signaux une fois amplifiés sont dirigés toujours avec le même type de connectique sur les haut-parleurs M.A.P.

Ceci est une solution « simple », on pourrait imaginer réduire le nombre de canaux ADAT par une interface audionumérique type MADI, permettant la transmission sur un seul câble avec un splitter numérique au niveau des ordinateurs appliquant les filtres permettant de

dispatcher les signaux. Il faut savoir que si il y a deux ordinateurs à appliquer les filtres sur les signaux, c'est qu'il faut énormément de ressource pour gérer le nombre de filtres en temps réel (1 macintosh 2,5 GHz pour 2 M.A.P soit 16 haut-parleurs).

### 3. Conclusion.

On peut voir que pour un système holophonique de petite taille, sa mise en œuvre ne nécessite pas de grosse contraintes matériels mais demande déjà de bonne ressources informatique pour le calcul et l'application des filtres en temps réel. On imagine tout de suite la difficulté tant au niveau matériel qu'au niveau DSP pour un système équipant une salle de spectacle d'un volume important.

## **6 Phénomènes indésirables**

La mise en œuvre d'un système basé sur le principe de la Wave Field Synthesis conduit à prendre en compte un certain nombre de paramètres intrinsèques au système. Le nombre de transducteurs ainsi que leurs caractéristiques respectives et de plus l'acoustique de la pièce amène trois principaux phénomènes causant une dégradation et empêchant une restitution complète des signaux.

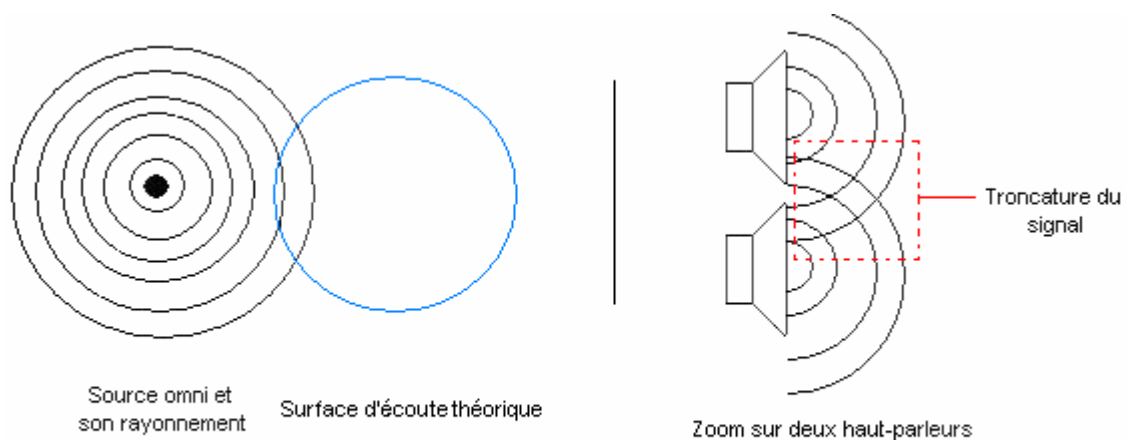
## 1. La phase stationnaire

Puisque nous travaillons sur le principe de la WFS, nous ramenons notre dispositif à un plan, avec une distribution linéaire des *transducteurs*. De ce fait, nous savons donc que les sources primaires seront distribuées par ces mêmes transducteurs, le but étant de déterminer la quantification à donner aux sources secondaires qui vont être elles aussi diffusées dans le plan des transducteurs. Il faut savoir que c'est grâce à ces dernières que l'on peut reproduire le champ acoustique (ou tout du moins une partie). Du fait que les sources secondaires sont limitées à une diffusion dans un plan, il nous est impossible de les synthétiser sur toute la surface théorique d'écoute (à savoir une sphère).

## 2. Fenêtrage et diffraction

Le but de la WFS est d'avoir un système de diffusion (théorique) cohérent et continu en tout point de la zone d'écoute. Les problèmes matériels engendrent un système cohérent mais discontinu du fait que chaque haut-parleur projette une onde primaire pouvant être découpé en une multitude d'ondes secondaires ayant une surface « finie » déterminée par la directivité des haut-parleurs.

Cet effet de diffraction se produit aux abords des haut-parleurs. Lorsque deux ondes se croisent, il se crée un phénomène de troncature perturbant le signal que l'on doit recevoir. Pour pallier un minimum à ce phénomène, il faut apporter une pondération aux signaux envoyés au banc de haut-parleurs permettant une distribution équilibrée du signal aux haut-parleurs.



Fenêtrage et diffraction

### 3. L'aliasing spatial

En plus du phénomène de fenêtrage vu plus haut, il existe un autre phénomène appelé aliasing spatial. Nous avons vu que le fait d'avoir une série de monopôles impliquait une diffusion discontinue de la scène sonore. De ce fait on crée un échantillonnage du banc de transducteurs et un donc un effet d'aliasing sur le signal. Puisque les haut-parleurs ne représentent pas un point mais une surface finie, on est obligé de n'utiliser qu'un certain nombre de haut-parleurs dont on sous échantillonne le signal et de ce fait on crée le phénomène d'aliasing. Il faut savoir que ce phénomène se rencontre qu'à partir d'une certaine fréquence appelée fréquence d'aliasing et auquel le ou les signaux sont sujet (essentiellement dans les hautes fréquences). Au niveau de la technologie M.A.P, par les calculs, on s'aperçoit que l'aliasing spatial s'opère à partir de 2000 Hz environ.

## **7 Prise de son/Mixage d'une scène sonore**

Dans le domaine de la prise de son ou du mixage multi canal, il n'y a pas vraiment de standard en la matière. On prend malgré tout exemple sur la stéréo, mais le fait que l'on évolue dans un espace en « trois dimensions » (plutôt deux dimensions puisque l'élévation n'est pas complètement prise en compte) rend les possibilités quasi infinies. Dans le système holophonique, on a également l'avantage de pouvoir donner une « dimension » à nos différents objets sonores ce qui n'est pas le cas de la stéréo classique ou des systèmes 5.1.

Il faut étudier le plan technique pour pouvoir mettre à jour des idées et des concepts de mixage sur le plan artistique. Mieux connaître les possibilités techniques permet de pouvoir mieux se servir des outils mais également de pouvoir complètement en détourner l'usage à des fins créatives.

## 1. La technique

Pour expliquer techniquement les possibilités de mixage par rapport au système adapté à la WFS (Wave Field Synthesis), il est nécessaire d'expliquer l'outil auteur développé par l'IRCAM pour le projet CARROUSO : LISTEN.

A la base LISTEN a été développé pour le système binaural puis une plate forme a été développée pour la WFS appelée ListenSpace.

On sait que la Wave Field Synthesis permet la reproduction de sources sonores à des positions virtuelles dans l'espace. En fonction de cette position dans l'espace, la courbure du front d'onde va se modifier. C'est cette courbure qui va permettre de localiser la source pendant un mouvement. On peut donc manipuler une source à une certaine distance de l'auditeur grâce à un paramètre appelé distance holophonique, celui-ci étant complètement indépendant de la notion de distance subjective qui est entre autre lié à l'équilibre entre l'intensité de la source et son niveau de réverbération (valables surtout pour les réflexions primaires). On trouve cette notion de distance holophonique en tant que paramètre sous le nom de « présence » dans le standard MPEG-4.

## 2. Notion de distance holophonique

Pour expliciter ce paramètre, nous allons reprendre l'exemple de Terence Caulkins et Etienne Corteel dans [4] qui explique les fonctionnalités de celui-ci dans une scène holophonique. On prend trois guitares et une voix dans une configuration plutôt « classique » avec les trois guitares sur un plan sonore et la voix devant ce plan.

On constitue trois scènes musicales, la première met la voix et les trois guitares à une « distance holophonique » courte. L'espace d'écoute est fermé. L'auditeur se déplace dans cet espace et peut effectuer une approche physique de chacune des sources.

La seconde situation place la voix à une courte « distance holophonique » mais les guitares à une « distance holophonique » infinie (ondes planes). Les trois guitares vont donc être perçues avec la même intensité et directivité angulaire dans tout l'espace d'écoute tandis que la voix au contraire va se trouver à une certaine place (fixe dans notre exemple) ce qui permet au spectateur de choisir son point de vue dans l'espace d'écoute.

La troisième et dernière situation donne toutes les sources en ondes planes donc à une très grande distance « holophonique ». De ce fait on a la même sensation dans tout l'espace d'écoute et la scène entière suit le spectateur dans son déplacement.

Quand on pousse le raisonnement, on s'aperçoit que du point de vue d'un spectateur statique et placé au milieu du système d'écoute, les trois scènes sonores vont lui paraître identiques. Grâce à cette observation on comprend que la notion de « distance holophonique » n'intervient que sur les effets de parallaxe (lié à une vraie position de sources dans une scène sonore) lorsque l'on se déplace dans un environnement naturel.

### 3. Prise de son

Au fur et à mesure de notre étude, nous avons observé les multiples contraintes qu'entraîne la technique de la Wave Field Synthesis dans sa mise en œuvre. Et pourtant au niveau de la prise de son, la TU de Delft a mis au point un système de prise assez impressionnant dont nous allons détailler le fonctionnement ici. Dans le cadre de la WFS, Delft a proposé une démonstration lors de la convention de l'AES à Munich en mai 2002, un système de prise de son dédié à cette technique de diffusion. On sait que dans une salle, une infinité de réflexions et de diffraction se créent et parviennent à l'auditeur par différentes directions. Pour pouvoir enregistrer le champ acoustique généré dans sa globalité, il faudrait un nombre de position de microphones supérieurs à 200 pour obtenir une bonne précision et enregistrer les données. Pour permettre l'enregistrement des données en diminuant le nombre de contraintes, on procède d'abord à l'auralisation du système dans sa globalité (enregistrement, processing, reproduction).

On définit l'auralisation par la procédure d'enregistrement des réponses impulsionnelles mesurées par les micros dans une salle et de reproduire les fronts d'ondes dans une autre salle sur un système WFS. Cette auralisation contient trois parties, l'enregistrement, le processing et la reproduction (diffusion). Lors de l'enregistrement, on mesure les réponses impulsionnelles de la salle, dans la partie processing, on calcule les filtres à appliquer en fonction des haut-parleurs et un peu comme le principe de la multi égalisation (chapitre 5), on calcule les filtres en fonction des réponses impulsionnelles mesurées dans la salle, et celle mesurées en chambre anéchoïque. Le tout étant enfin reproduit sur le système de diffusion.

Le processing est nécessaire dans le cas où les positions de microphone ne correspondent pas aux positions de haut-parleurs. De ce fait le champ sonore est extrapolé de la ligne de microphone vers la ligne de haut-parleurs. Pour réussir ce processus d'extrapolation, on se sert de l'inverse de l'intégrale de Kirchoff-Helmholtz vu plus haut. L'extrapolation du champ sonore dépend de la taille (surface) de la ligne de microphone.

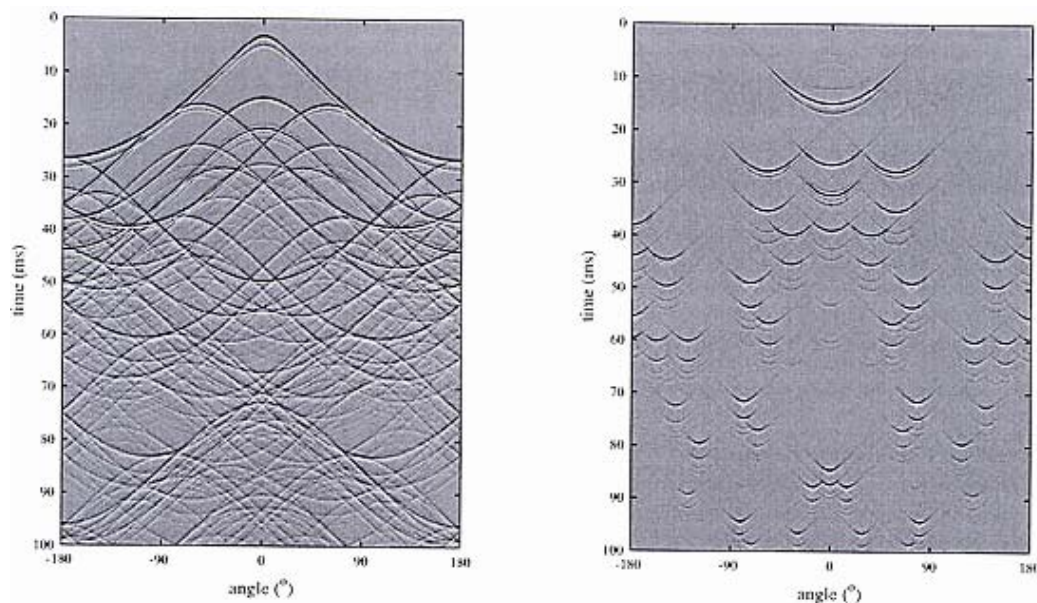
Puisque l'on se sert de l'intégrale de Kirchoff-Helmholtz on utilise donc la décomposition d'onde. Cette approche (extrapolation et auralisation) demande une grande ressource informatique car le nombre de données devient vite lourd à gérer. Il faut donc faire une réduction spatiale et une paramétrisation temporelle des données en temps réel. Nous allons voir tout d'abord la décomposition du champ sonore avant d'explicitier les paramètres de réduction des données.

#### **Décomposition du champ sonore**

D'un point de vue pratique et théorique il a été démontré qu'un système de prise de son en forme circulaire était la plus facile à utiliser, c'est donc celle-ci qui servira d'exemple dans ce paragraphe.

Pour réussir la décomposition du champ sonore dans un système holophonique, il faut utiliser une décomposition en harmonique cylindrique (procédé utilisé déjà dans les systèmes

Ambisonic). Pour réussir cette extraction des harmoniques, on tient compte de la pression, du vecteur vitesse (vitesse), du rayon du cercle de captation et de l'angle à auquel on capte le signal. On sait que lorsque un signal est émis, il est omni directionnel et a une infinité de réflexions en fonction des parois sur lesquelles il tape. Dans notre cas, seuls les signaux entrants dans notre cercle de captation nous intéressent. Le fait de décomposer nos signaux en ondes planes nous permet d'avoir une simplification du schéma et de voir les signaux utiles à la reconstruction plus tard. Ensuite on exerce une focalisation à l'infini à partir du cercle afin de pouvoir avoir la décomposition en onde plane la plus juste possible (comme en photographie, le fait d'ouvrir une focale sur l'infini permet d'avoir une profondeur du champ énorme). On ne fait que capter les signaux par la technique WFS dans une pièce rectangulaire. Puis on réalise la décomposition en ondes planes pour simplifier le système. A partir de cette simplification, on peut extraire les différents paramètres de la salle, à savoir le son direct, les



premières réflexions, les réflexions tardives et le champ diffu. Le fait que les ondes soient représentées avec une certaines courbures sont dues au fait des rapports de distances. Il est donc possible de n'obtenir que des points en décomposant les ondes en fonction du temps de trajet ainsi que de la distance focalisée.

Captation Wave Field Synthesis

Décomposition en ondes planes

### Captation et décomposition du champ sonore dans une pièce rectangulaire

Le but est donc de décomposer le champ acoustique entrant, afin de pouvoir le dispatcher par la suite sur tous les haut-parleurs en leur appliquant des filtres individuels.

#### La réduction spatiale de données

La réduction spatiale de données a pour but de diminuer le nombre de données informatiques enregistrées. Il est vrai qu'au vu du nombre de canaux enregistrés, le poids des fichiers devient vite lourd à gérer. Pour réussir cette diminution des données, on donne le nombre de canaux de diffusion dit technique adaptative. On donne donc un certain nombre de canaux et l'on analyse les signaux de façon à ne garder que les éléments dominants dans les signaux. C'est-à-dire qu'on garde par ordre de priorité le son direct, les réflexions précoces, le champs diffus dans une certaine proportion. L'utilisation de cette technique permet d'avoir une restitution correcte du signal qui, dans le cas où les signaux seraient distribués

uniformément sur le nombre de canaux disponible, il y aurait création de signaux parasites perturbant de ce fait la restitution correcte et totale du signal.

### **Paramétrisation temporelle**

La paramétrisation temporelle est l'autre moyen d'alléger les données. Pour pouvoir se faire, on utilise les données de l'oreille. On mesure une réponse impulsionnelle simple d'une salle (dans [8] il s'agit de la salle Concertgebouw d'Amsterdam.). A partir de celle-ci, on va chercher à extraire les paramètres perceptifs de la réponse mesurée. Pour pouvoir ce faire on va tenir compte de l'audition humaine (le temps et l'analyse de fréquence réalisée par l'oreille humaine au niveau de la membrane basilaire). Le fait d'avoir ces données permet de voir que l'oreille agit en tant que filtre passe bande. Pour comprendre le résultat d'une façon simple, on peut dire que le but est de décomposer les signaux par bande de fréquences faisant référence aux bandes de fréquences critiques de l'audition humaine. La réponse impulsionnelle est avant dédoublée. On calcule l'enveloppe temporelle en fonction de la puissance. Seuls les fonctions d'amplitude positives sont retenues (ainsi seuls les signaux perceptible par l'oreille humaine sont pris en compte) et donc allégeant par la même occasion la masse de données.

## **4. L'artistique**

### **L'écoute**

L'écoute d'un système basé sur la WFS est très déstabilisante. Le fait de pouvoir se mouvoir dans l'espace d'écoute est déjà étonnant mais de plus, on perd totalement ses repères face à un système non stéréophonique. La première écoute amène donc une phase d'adaptation, prendre des nouveaux repères face à un système sortant des « normes » actuelles. Lorsqu'une source ponctuelle non focalisée est synthétisée et que l'on se déplace dans cet espace, on découvre une nouvelle forme d'écoute. Il est possible également en se déplaçant qu'une source focalisée ou non (La source étant **fixe**) ait un déplacement assez court. Cela est dû au cerveau, qui lorsque l'on effectue un déplacement dépassant une certaine vitesse, met un certain temps (très court, certes) pour redéterminer l'angle de la source. Cela ne s'effectuera pas pour une source à onde plane pour les raisons suscitées dans le chapitre 3. L'avantage majeur de ce système réside dans la position d'écoute que l'on a, puisque par le principe même on peut se placer où l'on veut dans l'espace d'écoute on aura toujours une écoute précise et correcte mais avec à chaque fois un point de vue de l'écoute différente. On peut penser la composition par rapport à un emplacement donnée dans la salle d'écoute. On peut également associer le son à des objets visuels pour créer des illusions sonores. L'avantage réside donc dans le fait que tous les spectateurs, malgré l'espace qu'ils occuperont, auront une écoute différente sans avoir à se soucier de l'endroit et du « *sweet spot* ».

### **La composition**

L'art de la composition et de l'orchestration étant très subjectives et demandant une expérience longue et n'ayant pu rencontrer les personnes adéquates à l'heure où j'écris ces lignes, je vous présenterai exhaustivement les différentes possibilités que j'ai pu déduire de mes écoutes en fonction des possibilités offertes par le système de diffusion lui même.

L'idée de ce système est de pouvoir fonctionner dans le monde du spectacle. On a vu que le gros avantage était de pouvoir se mouvoir dans le système d'écoute sans perdre les informations de localisation. Lorsque l'on se retrouve en situation de concert, on imagine mal le public se déplacer pour pouvoir écouter le concert à des endroits différents. On peut donc inverser la situation et faire mouvoir les sources sonores. Dans une idée de synthèse totale de la scène sonore, pouvoir faire évoluer les sources dans l'espace d'écoute permet de jouer totalement avec l'acoustique de la pièce et la disposition du public. Dans un autre contexte, on peut mélanger de vrais musiciens en acoustique (un orchestre par exemple) et plonger des sources sonores dans celui-ci (un soliste par exemple) les laissant fixe ou en les faisant se mouvoir parmi les musiciens.

L'avantage aussi de ce type de sonorisation, est que si on utilise un système de diffusion fermé, on a une localisation plus précise des objets sonores que les systèmes plus traditionnels. A contrario, le phénomène de « *sweet spot* » est également plus grand mais si le système de diffusion est trop près des auditeurs alors on a le même phénomène que pour les autres système, toute la spatialisation est annihilée et on entend seulement ce qui sort du haut-parleur vers lequel on se trouve.

On remarque aussi des possibilités dans le cinéma, en appliquant à un objet visuel (à l'écran) sortant de cet espace visuel, une signature sonore permettant de continuer à l'entendre et de l'imaginer hors de l'écran.

## **8 Avenir de l'holophonie**

## 1. Développements

Les différentes possibilités offertes aujourd'hui par la diffusion en Wave Field Synthesis amène des idées dans de nouveaux développements et améliorations du système.

Les développements sur le système en lui-même, à l'heure actuelle, sont la synthèse de dipôles. Jusqu'à aujourd'hui toutes les sources émises par le système sont des sources monopôlaires donc ayant un rayonnement ponctuel et omnidirectionnel. Le fait de pouvoir créer une source dipolaire permettra de pouvoir synthétiser une source en fonction de son rayonnement et ayant un rendu plus fidèle lors de la retranscription d'événements sonores réels.

Les autres développements amenés sur le système seront vus plus d'un point de vue matériel et dans une utilisation artistique. Déjà l'exploitation dans les musées a été étudiée puisque le projet LISTEN découle de ce besoin avec l'adaptation du ListenSpace. Le but du ListenSpace était de pouvoir donner naissance à système de diffusion 3D permettant une distribution à un public entier. Au niveau du spectacle vivant, on voit de belle promesse dans la recréation d'ambiance ou le mélange acoustique et synthèse. Le cinéma permet également un espace de « jeu » sonore assez intéressant. La réalité virtuelle pourrait aussi être un secteur intéressant, mais semble quand même plus adapté à un système binaural, mais si le système évolue au niveau visuel, l'holophonie serait bien le prochain grand système de l'évolution sonore. Les développements quand à eux pourraient se proposer plus sur la mixité des dispositifs de composition avec le système de diffusion à proprement parler. La création de spectacle dans l'idée d'une diffusion en 3 dimensions pourrait se révéler très intéressante avec une idée de captation simultanée des mouvements, dans des pièces de théâtre ou la création d'ambiance réelle/virtuelle dans le cadre d'un vernissage ou d'une exposition (rejoignant ainsi l'idée de la diffusion dans les musées).

## 2. Mutations

On peut dire que la plupart des inventions, tout au long de leur utilisation subissent des mutations dues aux besoins différents de chacun. L'holophonie n'en n'étant qu'à ces balbutiements d'un point de vue de la mise en œuvre, que l'on travaille au fur et à mesure sans penser réellement à des utilisations hors laboratoire.

Les recherches actuelles permettent de développer la notion de directivité dans les sources synthétisée. Depuis le début des essais, on synthétisait un monopôle étant ponctuel et omnidirectionnel, on recherche donc globalement à synthétiser des dipôles dans un premier temps pour induire l'idée de directivité au sein de la source elle-même permettant de pouvoir reproduire une source plus fidèlement. Ceci sous-entend de pouvoir prendre en compte les idées de filtrage des sources face à un obstacle, l'absorption de l'air, ainsi que l'effet doppler étant déjà présent dans le Spat. Le but étant de pouvoir vraiment recréer une source fidèlement dans l'idée de se mouvoir dans un espace (réaliste ou non) mais ayant des propriétés acoustiques propres.

Cette idée amène donc le complément visuel de l'holophonie, à savoir : l'holographie. Dans cette optique, on pourrait recréer et synchroniser le son et l'image (réalité virtuelle augmentée). Le fait de pouvoir associer une image en 3 dimensions et le son adapté à l'image recréée (avec la directivité des sources ainsi que des phénomènes de filtrage).

Enfin, la technique de la Wave Field Synthesis, ne prend en compte que 2 dimensions réelles. La prochaine étape est l'intégration de l'élévation permettant ainsi une reconstruction totale du champ sonore. La possibilité d'arriver à ce système n'est pas évident tant au niveau de la puissance de calcul temps réel (c.f chapitre 5) ainsi que les problèmes purement matériel

(souci des haut-parleurs au sol). Ces soucis non négligeable, une fois résolus, rendraient le système holophonique, l'un des systèmes 3D les plus performants de l'époque.

# **Annexe A**

Le MPEG-4

**D**ans cette annexe nous allons expliquer succinctement le principe du format et de la norme MPEG-4 au niveau du son, puisque c'est l'intérêt principal de notre sujet. La norme créée avec le MPEG-4 permet de tenir compte des nouvelles technologies mise au point tant au point de vue de l'image que du son. Cette norme découle directement du MPEG-1 et 2 mis en place par l'Organisme International de Normalisation (ISO) et la Commission Internationale d'Electronique (IEC). La nouveauté intervenant dans cette norme est une philosophie orientée objet, d'où lors de l'encodage, on parlera d'une scène composée d'objet visuel et sonore. La première norme fût mise en place en janvier 1999 puis un an plus tard la deuxième et définitive version sera mise en place.

### **La norme MPEG-4 du son**

On a donc vu qu'on avait orienté le format vers un principe d'objet. A chaque objet on attribue un canal. Ce canal son aura une compression adaptée suivant le débit désiré ainsi que le type de son qu'il intègre (voix, musique, mixte ...). Le codage se fait avec un protocole propriétaire AAC (Advanced Audio Coding) mis au point par Dolby. Avec ce protocole de compression d'autres outils de codage sont venus se greffer pour compléter le format et constituer la norme. Ces outils sont choisis en fonction du débit que l'on souhaite obtenir, à savoir le T/F pour les forts et moyens débits, le codage HVXC (Harmonic Vector eXcitation Coding – codage par excitation de vecteur d'harmonique) pour le très faible débit et le CELP (Code Excited Linear Prediction - Prédiction linéaire excitée par code) pour le faible débit. Nous allons maintenant expliciter ces différents types de codage (surtout le faible débit qui nous intéresse plus particulièrement). Le codage très faible débit CELP fonctionne sur la principe de la prédiction. On génère un signal à partir de plusieurs tables de code. Puis ce signal est comparé au signal d'origine, et le résultat de la comparaison génère un signal d'erreur. Ce signal d'erreur est ensuite utilisé pour contrôler le gain des tables de code et les paramètres du filtre de synthèse ainsi que pour la création d'une table de code adaptative qui sera transmise avec les différents paramètres.

Quand au codage HVXC, il permet de coder la voix à un débit situé entre 2 et 4 Kb/s, le codeur HILN lui permet de coder les autres signaux n'étant pas de la parole à un débit supérieur à 4 Kb/s.

### **Conclusion**

Au niveau du son le MPEG-4 reprend les techniques de ses aînés (MPEG-2 et 1) déjà éprouvées mais du fait qu'elle décompose une scène en objet sonore et vidéo on obtient une flexibilité beaucoup plus grande pour l'utilisateur (opérateur pour le codage de la scène). De plus, le codage paramétrique pour les faibles débits a apporté des résultats plus que satisfaisant pour le codage de la voix. Son application pour la technique de la Wave Field Synthesis est parfaitement appropriée puisque ListenSpace intègre lui aussi un codage visuel de la scène orienté en objet sonore.

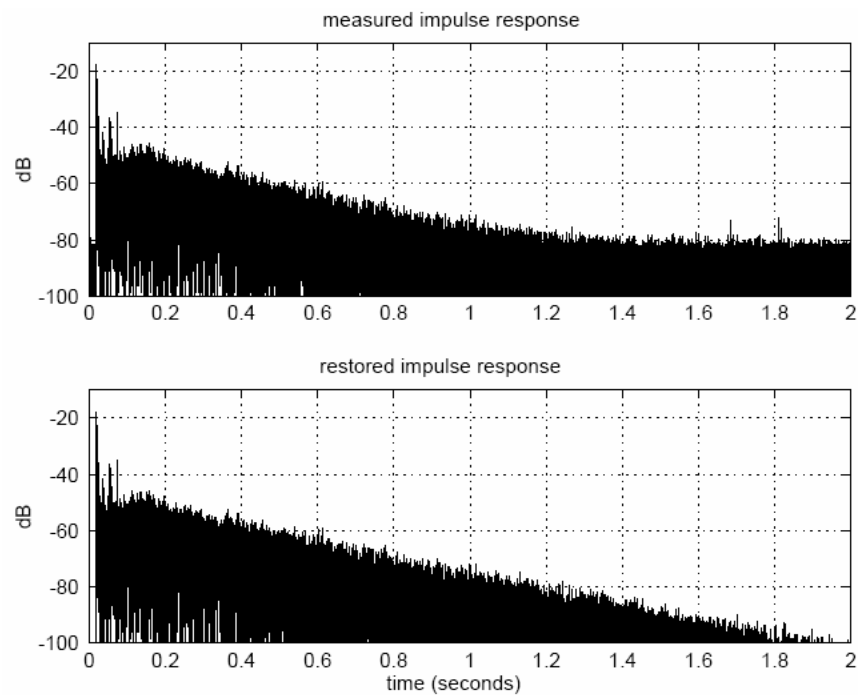
# **Annexe B**

La réverbération EDR

**E**DR est à la base un moteur d'analyse de réponse impulsionnelle basé sous l'environnement MatLab. EDR signifie Energy Decay Relief, d'où l'idée de l'analyse d'un temps de réverbération d'une salle. Ce type de moteur a pour but de pouvoir représenter la qualité acoustique d'une salle, de permettre l'égalisation de transducteurs ou bien la correction acoustique d'une salle par rapport à sa réponse impulsionnelle. A la base l'EDR permet de faire une représentation temps-fréquence du signal.

Pour définition l'EDR est la moyenne de l'ensemble de la représentation temps-fréquence de l'affaiblissement de la réverbération après l'extinction du signal engendré, si ce signal est un bruit blanc stationnaire.

Son but premier est de pouvoir prendre une réponse impulsionnelle mesurée dans une salle et de la reconstituer par la suite. La démarche est la suivante, après analyse de la réponse impulsionnelle de la salle par un bruit (impulsionnel) et l'analyse d'un bruit blanc dans cette même salle. EDR fait une analyse de Fourier sur les deux signaux, il analyse les réponses dans le temps et leurs enveloppes spectrales respectives. Il estime où est le son direct, les réflexions précoces et le champ diffu. Il reconnaît à partir de quel temps, le signal est trop perturbé (trop grand nombre de réflexions) et calcule l'affaiblissement normal à chaque fréquence, reconstituant de ce fait la totalité du temps de réverbération.



De plus, EDR servant aussi au calcul de réverbération dans le Spatialisateur, les informations analysées par ce moteur ont été divisées en deux types. Les informations de bas niveaux vont représenter les critères objectifs de la salle (informations sur le temps de réverbération, modes propres, calcul du Tr, analyse de Fourier...) et les informations de haut niveaux quant à elles vont représenter les critères perceptifs de la salle permettant de pouvoir modifier les informations selon des paramètres adaptés à un public plus large que celui des scientifiques de l'IRCAM.

# **Annexe C**

Le projet LISTEN

Le projet LISTEN (Augmenting everyday environments through interactive soundscapes) à été mené en collaboration entre l'IRCAM, le Fraunhofer Institut, la société AKG et l'université de Vienne. Le but étant de pousser la recherche dans la réalité augmentée dans le domaine sonore. Le principe de LISTEN était de pouvoir enrichir un environnement réel par des évènements sonores de manière immersive et surtout intuitive. La première application que l'on avait envisagé, était d'en faire un guide audio pour des expositions artistiques ou commerciales et de faire un rendu sonore de façon individualisé sur casque. Les principaux axes de recherches étant la modélisation de la scène sonore interactive, le moteur de rendu sonore et le développement du dispositif technique. Pour ce type de technologie, les chercheurs se sont orientés vers les techniques binaurales permettant une restitution au casque à 360°.

### **Le format multi canal pour la restitution binaurale**

Les techniques binaurales à proprement parler ne pouvant être utilisées complètement pour des raisons de puissance de calcul et surtout de coût, on va utilisé un format intermédiaire permettant de réduire les calculs et donc de dimensionner les coûts du dispositif. Le processus de spatialisation nécessite donc deux étapes importantes. La première est l'étape d'encodage où chaque source est distribuée sur un nombre déterminé de canaux, selon des coefficients dépendant de sa position dans l'espace. Par la suite ce flux peut être soit stocké soit transmis directement pour la seconde étape, le décodage de la scène. Pour cela les flux audio vont passer dans des filtres dont leurs caractéristiques sont spécifiques à l'auditeur (HRTF). Le but étant de faire une « grande distribution » du produit, on s'attellera à une étude statistique des fonctions de transfert de la tête permettant un fonctionnement maximal sur une grande partie des auditeurs utilisant le produit.

### **Mesures**

Le but des mesures étant d'améliorer la qualité et la diffusion des données sur casque, il est nécessaire d'optimiser le processus. Pour le moment chaque personne essayant un système binaural a besoin de passer par une séance de mesure pour pouvoir avoir un rendu optimal lors de l'écoute. L'idée de l'IRCAM était de définir des méthodes permettant de générer ces HRTF individuelles sans passer justement par une séance de mesures. De plus il fallait trouver une façon d'encoder et de décoder les signaux audio qui seront diffusés sur casque. Ils ont tout d'abord modifié leur système de mesures qui leur permettaient jusqu'à présent de ne capter qu'un seul canal acoustique (gênant pour le projet LISTEN qui nécessitait 2 canaux de mesures). Pour cela ils ont réécrit un logiciel dans l'environnement *Max/MSP* sous *MacOS*. Le logiciel a pris donc la forme d'une boîte à outils permettant de construire sa propre application avec deux types de mesures possible, par les codes de Golay ou les balayages fréquentiels. L'avantage de ce logiciel est également de pouvoir fonctionner soit en temps réel soit en temps différé. Lorsque l'on travaille en temps réel, alors les mesures sont enregistrées sous forme d'un fichier son, tandis que lorsque l'on travaille en temps différé il faut déconvoluer le signal enregistré. Au niveau matériel, il s'agit d'une table pouvant tourner en azimut sur laquelle une chaise avec le sujet est installée et d'un portique articulé ayant le haut-parleur diffusant le signal et mobile en élévation. Le contrôle de tout ce dispositif se fait dans l'environnement de *Max/MSP*. Seule la commande de la chaise tournante était commandée par un programme externe déjà réalisé sous l'environnement de *Windows*. Pour améliorer le dispositif de prise, l'équipe a mis au point des moules du conduit auditif permettant d'insérer les micros directement dedans de façon analogue aux mesures faites directement sur un sujet.

## **Adaptation individuelle**

Le but est de pouvoir créer des HRTF générique permettant une adaptation correcte sur chaque individu utilisant le système. Pour réussir l'implémentation d'HRTF générique il faut réussir à corrélérer plusieurs domaines qui sont les domaines morphologiques et acoustiques. Pour ce faire, plusieurs sujets ont été choisis. On a défini différents critères morphologiques. On a fait un moule de leurs oreilles externes permettant ainsi les mesures sur un mannequin. Le but des mesures étant de comparer les résultats et d'analyser les changements morphologiques et acoustiques engendrés. Les différences interindividuelles trouvées sont analysées dans le domaine morphologique, dans le domaine du signal et dans le domaine perceptif qu'ils ont cherchés à relier. Pour accélérer la phase de calcul, les ingénieurs ont planchés sur la modélisation 3D du visage de chaque individu à partir d'un scanner (IRM), de la reproduction des oreilles ou la construction 3D à partir d'un ensemble de photos. Une autre idée est de passer par des techniques numériques pour l'analyse des paramètres morphologiques. Le but de ce système est de créer des modèles simplifiés du comportement acoustique des oreilles et de la tête, ces mesures étant extraites également par des scanners. Par ce biais, il est possible de générer un ensemble de HRTF individuelles permettant de reproduire un espace immersif avec une assez grande précision.

## **ListenSpace**

ListenSpace est un outil auteur né de la nécessité de pouvoir contrôler graphiquement une scène sonore bidimensionnelle. La scène est donc décrite par des objets graphiques virtuels et physiques à la manière d'une vue d'oiseau (comme le module SpatCirc~ dans le Spatialisateur).

## **Communications**

Le besoin au niveau des communications est le fait que chaque changement d'un élément sonore dans un logiciel devait être répercuté dans les autres composants logiciels. Tout d'abord la communication se fait au niveau du fichier avec l'établissement d'une extension commune basée sur le standard XML, et ensuite au niveau du réseau lui-même pour donc partager en temps réel les différentes modifications apportées à la scène sonore dans l'un ou l'autres des composants logiciels. Le souci était le nombre d'*instances* à traiter en même temps. Lors de l'ouverture d'un fichier, chaque programme charge la scène en mémoire à partir du fichier commun et ensuite chaque programme se connecte aux autres afin de pouvoir transmettre tout modification effectuée lors de la lecture de la scène sonore. L'amélioration apportée au programme ListenSpace est qu'il peut communiquer avec lui-même, l'avantage étant de pouvoir travailler sur plusieurs *instances* en même temps, produisant donc le même scène sonore mais avec des vues différentes, ouvrant la voie ainsi aux applications multi-utilisateurs.

## **Elargissement au projet CARROUSO**

ListenSpace a donc été adapté et intégré au projet CARROUSO qui manifestait des besoins similaires en terme d'outil auteur pour le contrôle de scène sonore. Tout d'abord ListenSpace a été modifié pour la mise aux normes que nécessite le format MPEG-4. Puis les modifications se sont faites de telles sortes que l'on puisse intervenir au niveau de l'encodage de la scène sonore. D'autre part, on peut convertir la représentation graphique de la scène

sonore par le biais d'un ensemble de routines compatibles au format MPEG-4 de sorte que l'on puisse encore modifier la scène au moment du décodage et du rendu sonore.

## **Les réflexions**

Pour pouvoir améliorer la cohérence du rendu sonore, on étudie des modèles statistiques sur les réflexions précoces. Pour ce faire, l'IRCAM a mis au point un algorithme de calcul de sources images, ayant été incorporé au ListenSpace pour permettre un effet d'immersion plus précis. Il est ainsi possible de calculer les réflexions précoces d'ordre 1, 2 et 3 de plusieurs sources sonores en même temps lorsque l'architecture utilisée n'est pas trop complexe. Pour garder une souplesse dans les concepts de contrôle perceptif qui font la richesse du Spat, on filtre une partie des réflexions pour n'en garder qu'un faible nombre qui viendront affiner par un jeu de paramètres nouveaux le contrôle du module « early » du Spatialisateur.

## **Conclusion**

Le projet LISTEN a permis une avancée dans le domaine de la réalité augmentée. De plus, les développements menés ont permis d'accélérer et d'améliorer les systèmes de mesures actuels. Maintenant on peut connecter un joystick en commande MIDI permettant de contrôler les positions des sources directement. Cela donne une grosse possibilité de spatialisation au niveau d'une utilisation dans les infrastructures culturelles en rajoutant en plus sur les casques des systèmes permettant de prendre en compte les mouvements de la tête.

## Sources

**Alfonso Corti** : (1822-1876) Anatomiste Italien ayant été le premier à pouvoir décrire précisément l'intérieur de l'organe portant son nom.

**Berkhout** : première personne à avoir imaginé un système de restitution sonore sur une zone étendue.

**Edison Thomas** : (1847-1931), Inventeur américain qui tout au long de sa vie aura déposé des brevets pour un bon nombre d'inventions toutes aussi impressionnantes les unes que les autres (électricité, téléphone, cinéma, ....)

**Fraunhofer Institut** : institut de recherche créé par Joseph von Fraunhofer (1774-1826). Cet institut effectue des recherches dans tous les domaines comme les communications, l'énergie, la microtechnique...

**Fresnel Augustin** : physicien français (1788-1827). Il inventa des dispositifs donnant des interférences lumineuses dont il fera l'étude théorique. Il inventa aussi les lentilles pour phares.

**Helmholtz Hermann** : (1821-1894), physicien et physiologiste allemand

**Huygens Christian** : (1629-1695) Astronome et physicien, il a beaucoup travaillé sur les phénomènes optiques comme la nature ondulatoire de la lumière en 1690.

**IRT** : Institute für Rundfunktechnik

**IRCAM** : Institut de Recherche Acoustique/Musique

**IST** : Information Society Technologies

**Kirchoff Gustav** : physicien allemand (1824-1887). Il inventa le spectroscope à l'aide duquel il développa, avec Bunsen, l'analyse spectrale.

**Rayleigh John** : (1842-1919) Etudiant en physique et prix Nobel de Physique en 1901 il porta ces recherches essentiellement sur la lumière.

**Rozenn Nicol** : chercheur chez France Télécom R&D à Lannion. Il a basé ses recherches sur la Wave Field synthesis surtout dans le domaine de la visioconférence.

**Sabine Wallace** : (1868-1919) pionnier de l'acoustique architecturale, c'est à Harvard qu'il travailla sur l'impact de l'absorption sur le temps de réverbération.

## Glossaire

ADAT : Alesis Digital Audio Tape. Fibre optique ayant un protocole propriétaire (Alesis) pouvant faire transiter jusqu'à 8 canaux 24 bits/48 kHz en simultané.

Ambisonic : technique de diffusion basée sur la géométrie des ondes sphériques. On l'encode sur 4 canaux et permet une diffusion sur n haut-parleurs d'où une souplesse de choix du système de diffusion très grande.

Binaural : Technique existant depuis les années 70, elle a pour but la restitution d'une scène sonore à 360° selon avec une sphère avec deux transducteurs. Développée au casque, elle consiste à reconstruire au tympan d'un auditeur les signaux qui auraient été produit en situation réelle d'écoute par implémentation.

Bruit blanc : Signal reproduisant toutes les fréquences audibles à proportion égale (de 20 Hz à 20 khz).

Chambre anéchoïque : pièce étant dénuée de toutes réflexions.

Communication UDP : protocole de communication entre deux machines hardware (réseaux) dans notre exemple il s'agit de faire communiquer ListenSpace (windows) et Max/MSP (macintosh)

DSP : Digital Signal Processor. Processeur numérique permettant d'en faire une unité de calcul pour n'importe quel type d'opération.

Encodeurs directionnels : dans le Spatialisateur, il existe un module permettant de spécifier le nombre de haut-parleurs utilisé pour la diffusion et celui-ci se charge de répartir les signaux sur le nombre de directions disponibles d'où l'appellation d'encodeurs directionnels.

FM : Frequency Modulation, modulation de fréquence (voir principe de modulation)

Haut-parleurs MAP : Multi Actuator Panels. Une plaque de mousse enserrée dans deux plaques de cartons fins auxquels sont accrochés 4 ou 8 excitateurs, le panneau faisant office de membrane.

HRTF : Head Related Transfert Function. Mesures permettant de calculer les filtres à implémenter pour un système binaural (casque).

LISTEN : Augmenting everyday environments through interactive soundscapes. Projet mené par le Fraunhofer Institut pour développer dans le domaine de la réalité augmentée dans sa dimension sonore. En outre ils ont développé l'outil ListenSpace pour la technique de la Wave Field Synthesis.

MacOS : MacOS est le système d'exploitation servant dans les ordinateurs de type Macintosh, concurrent direct de Windows chez Microsoft.

Max/MSP : logiciel développé par l'IRCAM. Logiciel modulaire permettant de programmer des fonctions diverses telles que du midi, de la synthèse ....

Monopôle : source ponctuelle qui exhibe un rayonnement omnidirectionnel quelque soit sa fréquence. Ce type de source est créée selon les harmoniques sphérique d'ordre 0 (source donc omnidirectionnelle). Etant ponctuelle, elle n'a donc pas de surface ni de volume et peut être créé entre 250 Hz et la fréquence d'aliasing (environ 2000 Hz).

MPEG-4 : Moving Picture Expert Group. Format de codage multimédia ayant pour but l'encodage (description/reproduction) de scène audio/vidéo. Format initié en 1995 et a été terminé en 2000.

Principe de modulation : il existe deux types de modulation, la modulation de fréquence et la modulation d'amplitude. Leurs buts respectifs est de permettre d'envoyer des signaux grâce à un autre signal (porteuse) servant à transmettre les modulations du premier signal.

Réverbération EDR : Early Decay Reflexions. Permet dans le Spatialisateur de contrôler le niveau de premières réflexions avec un paramètre de cluster (taux de premières réflexions).

Sweet spot : Surface d'écoute théorique idéale dans un système de diffusion directif. Zone la plus appropriée pour une écoute correcte.

Transducteur : système permettant de transformer une énergie en une autre forme d'énergie (dans notre sujet, la référence est le microphone en captation et le haut-parleur en restitution)

WFS : Wave Field Synthesis. Technique dérivée de l'holophonie travaillant dans un plan et non dans l'espace. Ce système a pour but de reconstruire un champ acoustique dans une zone d'écoute sur le principe de la décomposition des fronts d'ondes primaires.

## Bibliographies

- [1] Terence Caulkins. Restitution d'un effet de salle cohérent avec le lieu d'écoute dans le cadre de Wave Field Synthesis : rapport de stage de DEA ATIAM. IRCAM, 2003
- [2] Sylvain Yon. Contrôle du champ acoustique en milieu réverbérant et applications à la communication (extrait « spatialisation sonore »)
- [3] Rozenn Nicol. Spatialisation sonore par holophonie. France Télécom, 2001
- [4] Etienne Corteel, Terence Caulkins. Sound Scene Creation and Manipulation using Wave Field Synthesis. IRCAM, 2004
- [5] Histoire du son. <http://b.schmerber.9online.fr/theorie2/histoire.htm>
- [6] Edo Hulsebos, Diemer de Vries. Parametrization and reproduction of concert hall acoustics measured with a circular microphone array. AES Convention Paper 5579 ,Munich, Mai 2002.
- [7] AES staff writer. Wave Field Synthesis : Evolution from stereophony and some practical challenges. AES, 117 th convention, May 2004.
- [8] Jean-Marc Jot, Laurent Cerveau, Olivier Warusfel : Analysis and synthesis of room reverberation based on a statistical time-frequency model. IRCAM,
- [9] Etienne Corteel, Egalisation de bancs de haut parleurs par techniques adaptatives multivoies, application à la reproduction de sources virtuelles en zone étendue. STUDER Professional Audio A. G. Juillet 2001