

Dispositif de spatialisation sonore 3D à l'Espace de Projection de l'IRCAM - un réseau de 345 haut-parleurs pour une restitution par WFS et HOA

**Markus Noisternig, Thibaut Carpentier,
Olivier Warusfel**
UMR 9912 STMS IRCAM-CNRS-UPMC
1 place Igor Stravinsky
75004 Paris
E-mail : markus.noisternig@ircam.fr

Résumé

L'Espace de Projection de l'IRCAM, salle de concert à acoustique variable, a récemment été équipé d'un ambitieux dispositif de synthèse de champ sonore en temps réel. Ce système de spatialisation sonore à vocations de recherche artistique et scientifique est constitué d'une ceinture de 264 haut-parleurs régulièrement répartis autour de la scène et du public pour une diffusion en Wave Field Synthesis (WFS) et d'un dôme de 75 haut-parleurs pour une diffusion tri-dimensionnelle en Ambisonics (HOA). Après un rappel des principes sur lesquels reposent ces méthodes de spatialisation et de certaines de leurs propriétés remarquables, cet article décrit les caractéristiques techniques de l'équipement et discute certaines contraintes pratiques et leurs conséquences du point de vue du traitement algorithmique.

Abstract

The variable acoustic hall of IRCAM, *Espace de Projection*, has been equipped with a high-end electro-acoustic system for real-time sound field synthesis. This 3D audio equipment is dedicated to artistic as well as scientific research. Four horizontal arrays consist of a total of 264 linearly spaced and independently driven loudspeakers that are distributed around the stage and the audience for Wave Field Synthesis (WFS). This horizontal array is complemented with a dome of 75 loudspeakers distributed on the walls and ceiling for 3D reproduction using High Order Ambisonics (HOA) format. The paper reviews the principles of these two sound field synthesis techniques and some of their remarkable properties. The main technical features of the equipment are then described together with the practical constraints of the installation and their consequences on an algorithmic standpoint.

Créé en 1977, l'Espace de Projection (ESPRO) de l'IRCAM est une salle de concert dont l'acoustique est modulable par un ensemble d'éléments mécaniques. La forme et le volume de la salle sont modifiables grâce au plafond divisé en trois parties dont les hauteurs peuvent être réglées indépendamment. Les murs et les plafonds sont subdivisés en modules prismatiques, dénommés périactes, présentant trois types de faces aux caractéristiques acoustiques différentes (absorbantes, réfléchissantes et diffusantes). Ces périactes sont motorisés et contrôlables à distance depuis un ordinateur. Par le jeu du volume et de la nature des parois, la qualité acoustique de la salle peut être grandement modifiée. À titre d'exemple, la plage de temps de réverbération accessible est comprise entre 0,6 et plus de 3 secondes. Dédicé à la création musicale contemporaine, l'Espace de Projection fait naturellement appel à des dispositifs électroacoustiques pour la diffusion des parties électroniques ou des traitements informatiques appliqués en temps réel

sur les sons produits par les instruments acoustiques sur scène. Chaque œuvre requiert un dispositif *ad hoc* mais le plus souvent, le système de reproduction est constitué de quelques enceintes réparties de manière concentrique autour du public.

Cependant, les capacités de traitement du signal des ordinateurs et l'existence de périphériques audio massivement multicanal permettent aujourd'hui de concevoir des dispositifs recréant un champ sonore physiquement cohérent sur une très grande zone d'écoute en s'appuyant sur un réseau dense de transducteurs. La cohérence physique du champ sonore ainsi recréé représente un saut qualitatif important par rapport aux formats de restitution sonore spatialisée actuellement utilisés dans l'industrie audio. Notamment elle donne accès à des effets de parallaxe sonore, de sources sonores focalisées c'est-à-dire semblant exister à l'intérieur même de la salle d'écoute, ou de sources virtuelles dont non seulement la position mais également l'orientation peuvent être contrôlées.

Dans cet esprit, l'Espace de Projection se voit aujourd'hui enrichi par l'installation d'un dispositif de spatialisation sonore combinant deux systèmes de pointe en matière de synthèse du champ sonore, la *Wave Field Synthesis* (WFS) et Ambisonics aux ordres supérieurs (*High Order Ambisonics* ou HOA). Ce nouveau dispositif est constitué d'une ceinture de haut-parleurs densément répartis autour de la scène et du public pour la diffusion en WFS et d'un dôme de haut-parleurs pour une diffusion tri-dimensionnelle en Ambisonics. Les 345 haut-parleurs constituant le dispositif sont contrôlés indépendamment par un ensemble d'ordinateurs gérant en temps réel la spatialisation des sources sonores.

Dans le cadre de son installation à l'Espace de Projection, ce système a naturellement pour vocation principale la création musicale. Sur le plan artistique et notamment musical, l'enjeu est de permettre à des compositeurs ou designers sonores de s'emparer de ces nouvelles technologies et de ces nouvelles modalités d'écoute de sorte à enrichir le vocabulaire des traitements sonores. Au-delà de la création musicale, les technologies audio basées sur des réseaux denses de haut-parleurs commencent à susciter également un intérêt dans le domaine de l'industrie cinématographique ou des télécommunications. Ce type de dispositif peut trouver des applications non seulement pour le domaine du spectacle vivant mais aussi pour le cinéma ou la réalité virtuelle et augmentée.

Cet article rappelle les principes généraux des deux méthodes de synthèse de champ sonore exploitées dans ce système, décrit leurs propriétés principales et leur implémentation.

Méthodes de Spatialisation

Les systèmes de restitution sonore spatiale visent, d'une manière générale, à reproduire aux oreilles d'un auditeur les indices acoustiques appropriés pour contrôler la localisation apparente des sources virtuelles composant la scène sonore et les caractéristiques d'un effet de salle éventuellement associé. Ces systèmes sont catégorisables en trois familles. La plus ancienne et encore la plus répandue, exploite les propriétés de la perception, et repose sur un modèle simplifié de différence de niveau sonore ou de temps d'arrivée introduit entre deux haut-parleurs voisins et qui suffit à créer l'illusion d'une source «fantôme» localisable entre les positions physiques des haut-parleurs, pourvu cependant que l'auditeur soit correctement placé. Ce principe, dit de panoramique, à la base de la restitution stéréophonique classique, peut être étendu à des ensembles de haut-parleurs plus complexes répartis autour de l'auditeur soit en couronne pour une restitution 2D (ex. : format 5.1) soit en volume pour une restitution 3D (ex. : VBAP, 22.2, ...). Ces systèmes souffrent toutefois de la contrainte du placement de l'auditeur. La restitution n'est valable qu'en une zone très réduite. Les techniques binaurales, réservées à la reproduction sur casque, font quant à elles abstraction de tout modèle perceptif ou physique et reconstruisent directement l'ensemble des indices de localisation à partir de l'échantillonnage des fonctions de directivité relevées sur une tête humaine ou artificielle et dénommées HRTF (*Head Related Transfer Functions*). Ces

techniques binaurales permettent donc de reproduire les indices acoustiques de manière exacte, mais s'adressent par définition à un auditeur unique.

Des dispositifs de reproduction sonore plus récents visent à reconstruire les propriétés physiques du champ acoustique dans un voisinage plus ou moins large autour de l'auditeur. Ces dispositifs se fondent sur des formalismes mathématiques exprimant une solution à l'équation d'onde dans la zone de restitution. Parmi ces techniques, on peut distinguer en pratique les approches intégrales comme la *Wave Field Synthesis* et les approches modales comme les techniques *Ambisonics*. Leurs principes respectifs sont brièvement rappelés ci-dessous avant d'en décrire la mise en œuvre pratique dans l'équipement installé dans l'Espace de Projection.

Wave Field Synthesis (WFS)

Aperçu théorique

Le principe de la *Wave Field Synthesis* (WFS) dérive du formalisme intégral de Kirchhoff-Helmholtz qui relie le champ de pression résidant à l'intérieur d'un volume exempt de toute source acoustique au champ induit par les sources extérieures au domaine sur une surface fermée quelconque entourant ce volume (cf. Figure 1).

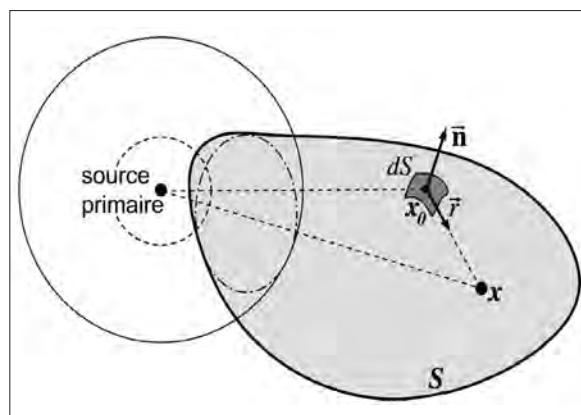


Fig. 1 : Notation pour l'intégrale de Kirchhoff-Helmholtz

Cette relation est exprimée par l'équation 2.1 où P est la pression acoustique résultante en un point x du domaine de restitution, x_0 est la variable d'intégration sur la surface entourant la zone de restitution et sur laquelle réside le champ de pression P_0 induit par les sources primaires, \vec{n} est la normale en x_0 , $k = \omega/c$ est le nombre d'onde, $r = \|\vec{r}\|$ et $j^2 = -1$.

$$P(x, \omega) = \iint_S [\vec{\nabla} P_0(x_0, \omega) \cdot \vec{n} - \vec{r} \cdot \vec{n} (1 + jkr) P_0(x_0, \omega) / r] \frac{e^{-jkr}}{4\pi r} dS \quad (2.1)$$

L'équation 2.1 peut s'interpréter comme la somme du rayonnement d'un double réseau de sources secondaires (sources en pression et sources en vitesse) distribuées sur la surface entourant la zone de restitution et alimentées respectivement par les contributions de vitesse et de pression induites par les sources primaires que l'on veut simuler ou restituer. Dans le cadre d'une reproduction de champ sonore, le principe consisterait à enregistrer la scène sonore par deux

ensembles de microphones de pression et de gradient de pression et à rediffuser respectivement ces signaux par des haut-parleurs dipolaires et des haut-parleurs monopolaires situés aux mêmes endroits que les microphones. Dans le cadre d'une synthèse, les haut-parleurs sont alimentés par le signal de la source à spatialiser et subissent chacun un filtrage simulant le rayonnement de la source et la propagation acoustique de cette source jusqu'au microphone virtuel. Ce traitement doit être répété pour chaque source composant la scène sonore.

Réalisation pratique

L'implémentation pratique de la WFS, que l'on doit aux travaux de A.J. Berkhouit menés dans les années 80, résulte d'un ensemble d'approximations et de termes correctifs associés [1],[2]. La première approximation consiste à limiter la reconstruction à un seul réseau de sources secondaires, typiquement le réseau de sources en pression. La conséquence essentielle est que le système n'est plus isolé de son milieu. En effet, dans la formulation théorique, les champs reproduits par les deux réseaux de sources secondaires s'annulent à l'extérieur du domaine de restitution et tout se passe comme si le lieu d'écoute était anéchoïque [3]. La restriction à un seul réseau, fait apparaître l'interaction du système avec la salle. Selon l'objectif du système, on peut tenter de compenser les réflexions ainsi engendrées [4], ou au contraire tirer parti de la présence de cet effet de salle qui peut contribuer au réalisme de la situation.

La seconde approximation suppose que la majorité des sources virtuelles à reproduire est située dans un plan horizontal où se situent également les auditeurs. Dans ce cas, la contribution majeure des sources secondaires est apportée par celles qui sont situées dans le plan horizontal. On peut alors restreindre la distribution surfacique de sources secondaires par une distribution linéaire ceinturant la zone d'écoute dans le plan horizontal. Enfin, inévitablement, la réalisation pratique passe par une discréétisation de la distribution sous forme d'un réseau de haut-parleurs régulièrement espacés. Cet échantillonnage spatial entraîne une fréquence d'aliasage spatial (aliasing) au-delà de laquelle l'exactitude de la reproduction n'est plus garantie. En pratique, on cherche à maintenir cette fréquence au-dessus de 1kHz de façon à assurer au moins la fidélité de reproduction de l'indice interaural de retard de phase sur l'ensemble de la zone d'écoute.

Propriétés

La figure 2 illustre certaines des propriétés remarquables du champ sonore reproduit par un banc linéaire de haut-parleurs contrôlés par *Wave Field Synthesis* pour différentes positions de la source virtuelle. Dans la situation (a) la source virtuelle est située en arrière du banc de haut-parleurs. Dans la situation (b) le front d'onde est associé à une source virtuellement placée à l'infini et parvenant selon la direction 30°. Enfin, dans la situation (c), le front d'onde constitué semble provenir d'une position interne au lieu d'écoute. Dans chacune de ces situations, on observe que le front d'onde se déploie de manière cohérente à partir de la position supposée de la source. En chaque point de l'espace, l'orientation locale du front d'onde, véhicule les indices de localisation auditive cohérents avec la position de la source virtuelle. Cette cohérence physique du champ sonore permet d'assurer la congruence des variations d'indices de la localisation auditive perçues par un sujet en fonction de ses déplacements dans la zone d'écoute ce qui renforce la sensation de présence de la source.

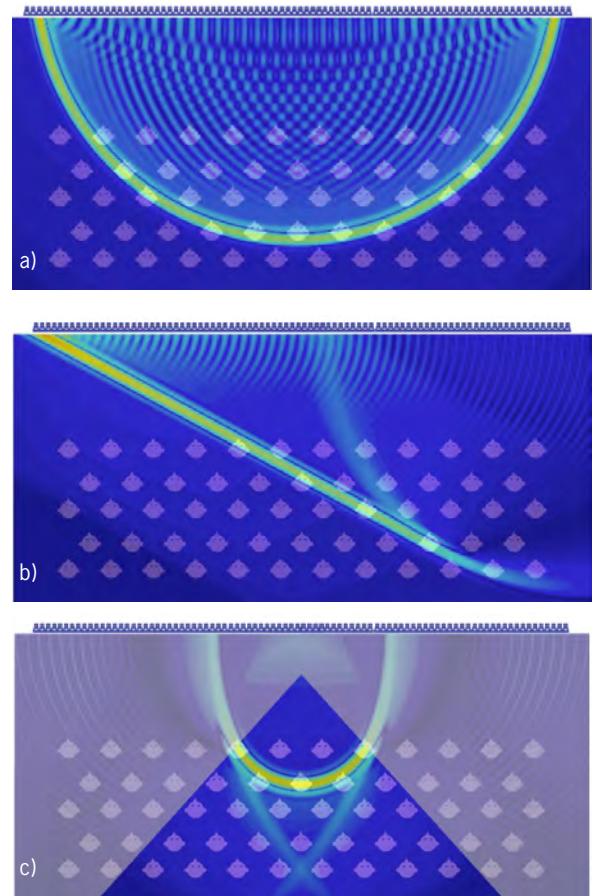


Fig. 2 : Simulation du champ acoustique synthétisé par un banc de 88 haut-parleurs espacés de 16 cm pour différentes positions de sources virtuelles. (a) source ponctuelle située 1 m derrière les haut-parleurs ; (b) onde plane de direction 30° ; (c) source ponctuelle focalisée située 3 m devant les haut-parleurs ; seule la zone non-grisée correspond au domaine de validité de la restitution.

À noter cependant que pour le cas de la source focalisée (c) la zone de validité spatiale exclut la portion de l'espace située entre la position de la source virtuelle et le banc de haut-parleurs. Différents artefacts sont également observables. Outre les franges d'interférences, on remarque sur la figure 2.b la présence d'un front d'onde secondaire lié à la troncature du réseau de transducteurs. Cependant ce front intervient après le front principal et perturbe peu la localisation de la source. En revanche, sur la figure 2.c on remarque deux fronts secondaires qui se propagent cette fois en avance par rapport au front principal. Ces deux fronts limitent la zone pour laquelle la localisation apparente de la source virtuelle sera correctement perçue.

Au cours des études qui ont précédé la réalisation de cet équipement, il était apparu intéressant de simuler non seulement la position de la source (direction et distance), mais également sa directivité associée, c'est-à-dire son angle d'ouverture et son orientation [5]. La fonction de directivité d'une source fait partie de son identité et joue un rôle important pour la perception. De sorte à rendre ce contrôle le plus souple possible, la directivité est synthétisée par combinaison de figures de directivité élémentaires, dites harmoniques circulaires et qui sont issues de la restriction planaire des harmoniques cylindriques.

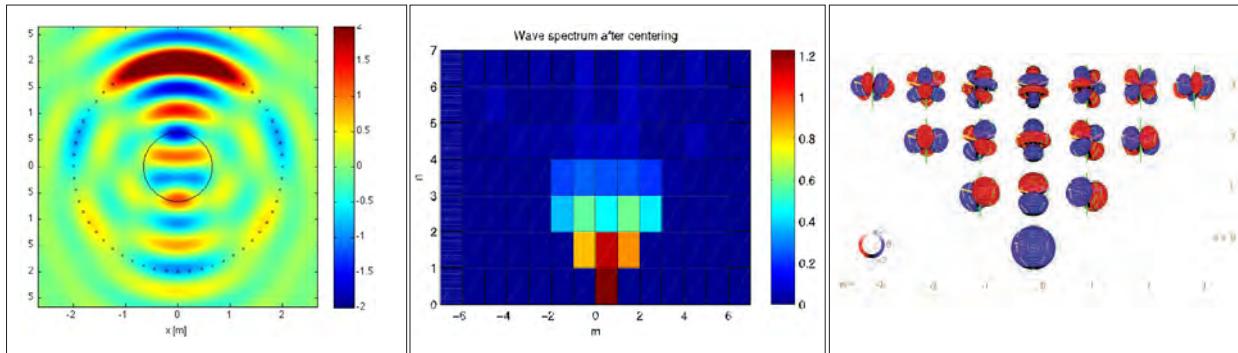


Fig. 3 : Représentations conjuguées d'un champ sonore (à gauche) et de son spectre spatial (au milieu).
À droite : motifs spatiaux des harmoniques sphériques jusqu'à l'ordre 3

Ambisonics

L'extension du principe de la WFS en trois dimensions n'est pas réaliste puisqu'elle conduirait à installer un maillage surfacique uniformément dense dans toutes les directions. Par ailleurs, notre perception ne présente pas la même résolution angulaire suivant les dimensions latérale et verticale et l'on peut admettre relâcher les contraintes de densité spatiale des transducteurs selon la dimension verticale.

Différentes techniques existent pour la reproduction de scènes sonores par un réseau périphérique tri-dimensionnel de haut-parleurs. Ces solutions se répartissent principalement suivant deux catégories selon qu'elles s'appuient sur un modèle physique ou perceptif. La technique VBAP (Vector Based Amplitude Panning), extension de la stéréophonie dans le cas d'un réseau tri-dimensionnel de haut-parleurs, repose sur la notion de source fantôme réalisée par pondération de l'amplitude du signal au sein du triplet de haut-parleurs dont les directions entourent celle de la source virtuelle. Comme la stéréophonie, cette technique souffre d'une dépendance de la direction perçue en fonction de l'emplacement de l'auditeur.

Le formalisme physique Ambisonics, retenu ici, se prête particulièrement à la description de champs sonores tridimensionnels grâce au recours à une décomposition sur la base des harmoniques sphériques.

Aperçu théorique

La famille de format Ambisonics, introduite par Gerzon dans les années 70 [6] qui a récemment fait l'objet de nombreuses recherches [7],[8],[9], repose également sur un formalisme physique en adoptant d'une part, un point de vue égocentrique, symbolisé par un repère en coordonnées sphériques, et d'autre part, une approche modale où la distribution spatiale est décomposée sur une base de fonctions spatiales. La solution de l'équation des ondes dans un volume sphérique dépourvu de sources acoustique s'exprime par l'équation 2.2.

$$P(kr, \varphi, \theta) = \sum_{n=0}^{\infty} \sum_{m=-n}^n b_{nm} j_n(kr) Y_n^m(\varphi, \theta) \quad (2.2)$$

La dépendance radiale est traduite par les fonctions de Bessel sphériques de première espèce $j_n(kr)$, tandis que la dépendance directionnelle est représentée par les harmoniques sphériques dont les motifs spatiaux des premiers ordres sont représentés Figure 3 (partie droite). Les coefficients b_{nm} sont les coefficients du spectre spatial associé à la distribution de pression. La figure 3 illustre la correspondance entre le champ sonore (à gauche) et son spectre spatial associé (figure du milieu).

Réalisation pratique

Le format ambisonique consiste à transmettre la représentation du champ sonore dans le domaine des harmoniques sphériques ce qui suppose par conséquent le développement d'une étape d'encodage et de décodage. Pour l'encodage, par exemple dans le cadre d'un enregistrement, il s'agit en quelque sorte de représenter le champ sonore au point de référence par une série de microphones virtuels dont les directivités respectives seraient conformes à celles des harmoniques sphériques (monopole, dipôle, quadripôle, etc.). En pratique, bien qu'il n'existe pas de cellules microphoniques permettant de capter directement les composantes spatiales d'ordre élevé on peut néanmoins approcher celles-ci par la combinaison d'un ensemble de cellules microphoniques simples distribuées concentriquement autour du point de référence [10] [11]. Réciproquement, les signaux de ce format d'encodage ne sont pas destinés chacun à un haut-parleur particulier. Ils doivent être décodés par matriçage sur l'ensemble des haut-parleurs répartis autour de la zone de reproduction en fonction de leur nombre et de leur direction respective [12].

Propriétés

La propriété principale de l'approche Ambisonics est de reposer sur un format hiérarchique et indépendant du dispositif de reproduction. À chaque ordre n , $2n+1$ canaux supplémentaires permettent de préciser la résolution spatiale de la scène sonore enregistrée ou synthétisée et aboutissent à un total de $(n+1)^2$ canaux. Naturellement, à l'étape de restitution, plus les haut-parleurs sont nombreux et régulièrement répartis, plus on sera en mesure de décoder correctement les ordres supérieurs et meilleure sera la résolution spatiale de la scène sonore reproduite.

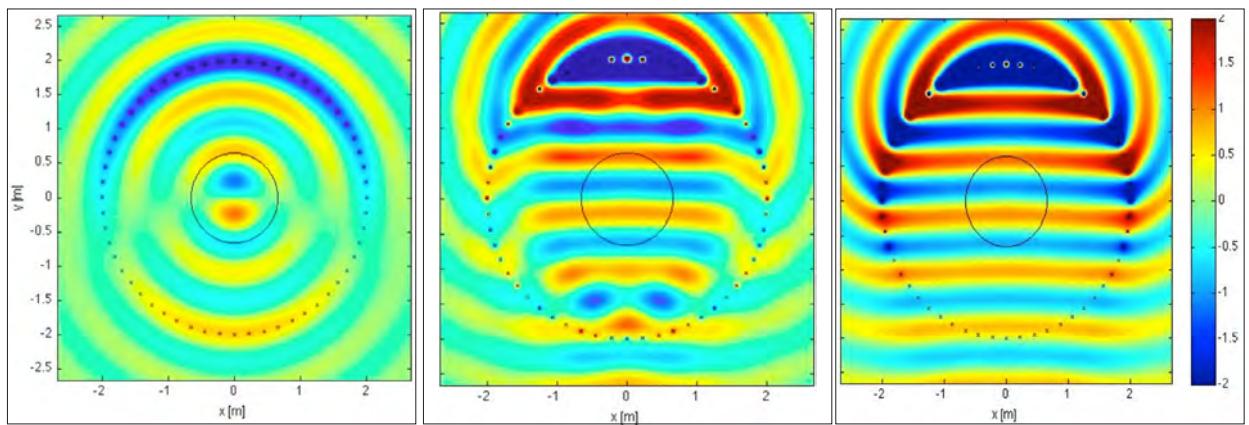


Fig. 4 : Reproduction 2D d'une onde plane par un ensemble de 56 haut-parleurs pour différents ordres de reconstruction HOA. De gauche à droite, successivement ordre 1, 10 et 20

La figure 4 illustre l'évolution du champ sonore reproduit par un ensemble de 56 haut-parleurs équi-répartis dans le plan en fonction de l'ordre maximal auquel on mène la décomposition d'une onde plane parvenant de la direction frontale. À l'ordre 1, seule la zone centrale de restitution bénéficie d'une restitution correcte. À l'exception des transducteurs latéraux, l'ensemble du dispositif de restitution est cependant sollicité contribuant pour les places non centrales à une impression d'enveloppement et de flou de la localisation. Au fur et à mesure que les composantes d'ordre supérieurs sont ajoutées, la zone de restitution s'élargit et la direction de provenance se précise.

Description du dispositif

L'installation des réseaux de haut-parleurs pour la diffusion en mode WFS et en HOA est fortement contrainte par les spécificités de l'ESPRO. D'une part, il ne faut pas entraver la rotation des périactes qui pavent l'ensemble des murs et des plafonds et par ailleurs il faut également autoriser le déplacement des ponts techniques (cf. Figure 5). Ces deux contraintes imposent à la fois l'emplacement et l'encombrement maximal des transducteurs composant ces réseaux. En particulier, la plupart des haut-parleurs ne doivent pas excéder une largeur ou une hauteur de 20 cm de sorte à pouvoir être disposés entre deux rangées de périactes. De même leur profondeur ne doit pas excéder 20cm de sorte à ménager le passage des ponts techniques.

Réseau linéaire pour la reproduction en WFS
La diffusion en WFS repose sur un réseau de haut-parleurs ceinturant la salle et divisé en quatre segments. Le segment frontal et le segment arrière sont respectivement composés de 88 et de 64 haut-parleurs espacés de 16 cm. Il s'agit de haut-parleurs actifs coaxiaux alimentés en chaîne par un flux numérique transporté par fibre Ethernet sous protocole Ethersound.

Pour des raisons de contraintes d'encombrement, les réseaux latéraux sont constitués de haut-parleurs coaxiaux passifs dont l'espacement est cette fois de 29 cm. Les modules de conversion Ethersound/analogique et d'amplification sont déportés dans le plenum situé entre les périactes et la structure primaire de l'ESPRO. Les bancs latéraux sont composés de 64 haut-parleurs, dont 9 d'entre eux sont partagés avec l'hémisphère Ambisonics décrit ci-dessous.

Au total la ceinture comprend 280 haut-parleurs (dont 264 sont dédiés à la reproduction en WFS) placés à 2m50 du sol.



Crédit photographique Sylvia Gomes



Fig. 5 : Segment frontal (scène) et segment latéral du réseau de haut-parleurs pour la WFS

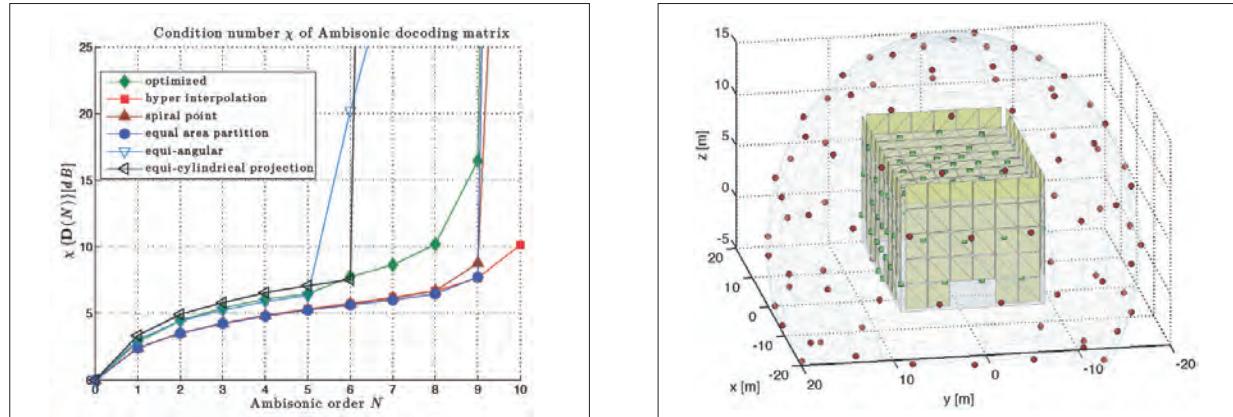


Fig. 6 : Conditionnement de la matrice de décodage pour différentes distributions spatiales théoriques des haut-parleurs. En vert : conditionnement de la matrice de décodage pour la répartition effective des haut-parleurs dans l'ESPRO. À droite : emplacement théorique (rouge) et effectif (en vert) des haut-parleurs

Réseau périphérique tri-dimensionnel pour la reproduction en Ambisonics

Au total l'hémisphère destiné à la reproduction HOA est composé de 75 haut-parleurs dont 59 sont des haut-parleurs individuels répartis en partie haute des murs et sur le plafond et les 16 autres sont intégrés aux bancs de haut-parleurs latéraux décrits précédemment.

Du nombre et de la répartition spatiale des haut-parleurs constituant le réseau vont naturellement dépendre le nombre d'harmoniques sphériques qui pourront être correctement reproduites et, par voie de conséquence, la précision angulaire de restitution. D'un point de vue général, il s'agit de trouver la répartition de haut-parleurs réalisant la discréttisation de la sphère la plus régulière possible. La logique sous-jacente d'une installation HOA est celle d'une répartition sphérique ou hémisphérique des haut-parleurs. Toutefois, pour des raisons pratiques, les haut-parleurs sont souvent reportés sur les parois de la salle, sachant que les différences de distances par rapport au point de référence peuvent être compensées par un retard et un gain de sorte à ré-aligner virtuellement tous les haut-parleurs sur une sphère.

Dans l'Espace de Projection, le positionnement des haut-parleurs est toutefois contraint par la présence des périactes dont il ne faut pas empêcher la rota-

tion. La distribution spatiale retenue résulte par conséquent d'un compromis entre une distribution idéale et la distribution obtenue après avoir déporté chaque haut-parleur vers l'emplacement disponible le plus proche de la direction souhaitée. La qualité de reproduction peut être estimée par le conditionnement de la matrice décrivant la discréttisation des harmoniques sphériques selon les directions retenues. Plus la valeur de conditionnement est basse, meilleure est la qualité de reproduction. À titre de comparaison, la figure 6 présente l'évolution de la valeur de conditionnement pour différentes répartitions spatiales théoriques d'un ensemble de 75 haut-parleurs en fonction de l'ordre de décomposition visé. La courbe en rouge montre les valeurs théoriques optimales que l'on peut obtenir avec la répartition selon le schéma «hyper-interpolation». À l'inverse, la courbe bleu clair montre le conditionnement obtenu avec une répartition équi-angulaire dont les performances se dégradent très rapidement au-delà de l'ordre 5. La répartition effective des 75 haut-parleurs dans l'ESPRO décrite par la courbe verte, permet de maintenir une valeur de conditionnement proche des valeurs optimales jusqu'à l'ordre 8 et encore acceptable à l'ordre 9, ce qui laisse augurer de très bonnes performances de reproduction spatiale.



Fig. 7 : Différentes versions des haut-parleurs PMX5-SLIM utilisés pour l'hémisphère de diffusion en mode HOA. Les versions diffèrent par l'inclinaison des pavillons de sorte à orienter le lobe principal vers le centre de la salle

Les contraintes d'encombrement liées au passage des ponts techniques limitent également la plage de rotation des haut-parleurs sur leur lyre de fixation. En conséquence, différentes variantes géométriques des haut-parleurs ont été conçues (cf. Figure 7, page précédente). En fonction de leur emplacement dans la salle, l'axe de la membrane et du moteur co-axial est incliné de 30° ou 45° de sorte à pouvoir être orienté vers le centre de référence de l'hémisphère. Comme pour les bancs de haut-parleurs latéraux, les convertisseurs et amplificateurs associés sont disposés dans le plenum situé entre les périactes et la structure primaire de l'ESPRO.

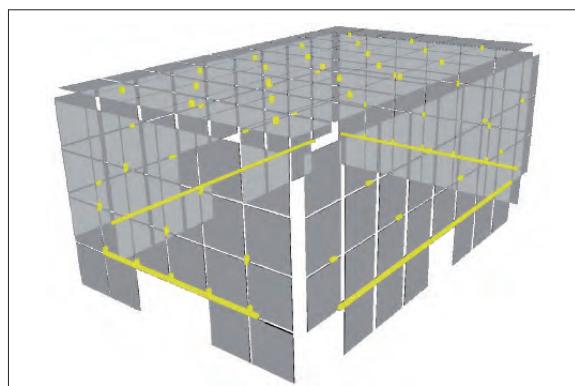


Fig. 8 : Implantation finale du réseau horizontal périphérique de haut-parleurs pour la WFS et du réseau hémisphérique pour la diffusion HOA

La figure 8 montre l'implantation finale des haut-parleurs avec, d'une part, le ceinturage des bancs de haut-parleurs pour la WFS et d'autre part, l'hémisphère de haut-parleurs pour la reproduction HOA. Au total, les bancs de haut-parleurs horizontaux comprennent 280 haut-parleurs répartis comme suit :

- 88 en fond de scène [16 cm d'espacement]

- 64 à l'arrière [16 cm d'espacement]
- 64 sur chacun des murs latéraux Nord et Sud [29 cm d'espacement], dont 56 servent à la restitution WFS, tandis que les 8 autres servent à la restitution en HOA.

L'installation HOA est composée de 75 haut-parleurs, dont :

- 23 en plafond,
- 16 sur chacun des murs Sud et Nord (8 d'entre eux étant intégrés dans les bancs WFS),
- 10 sur chacun des murs Est et Ouest.

L'installation est complétée par 6 haut-parleurs de très basses fréquences.

Organigramme de traitement de l'installation finale

La figure 9 représente l'organigramme de traitement de l'installation finale.

Étage d'entrée

Les signaux d'entrée sont véhiculés sur fibre MADI, permettant à terme l'exploitation de 64 sources en parallèle. Les signaux d'entrée sont répliqués sur l'ensemble des unités de traitement par un routeur MADI.

Unités de traitement

Le traitement audio-numérique lié à la WFS est pris en charge par quatre machines, chacune étant responsable d'un segment linéaire du réseau. Une cinquième machine prend en charge le traitement de l'ensemble de l'hémisphère HOA.

Étage de sortie

En sortie, les signaux sont envoyés à une matrice MADI qui gère le routage de l'ensemble des signaux fournis par les différentes unités de traitement vers l'ensemble des transducteurs. Cette étape est importante car elle permet d'assurer l'indépendance entre les unités de traitement et les transducteurs. Même si au moment de l'achèvement de l'installation, les unités de traitement sont spécialisées en fonction de leur mode de reproduction (WFS ou HOA), cette distinction sera petit à petit levée de sorte que chaque machine pourra s'adresser à n'importe quel transducteur.

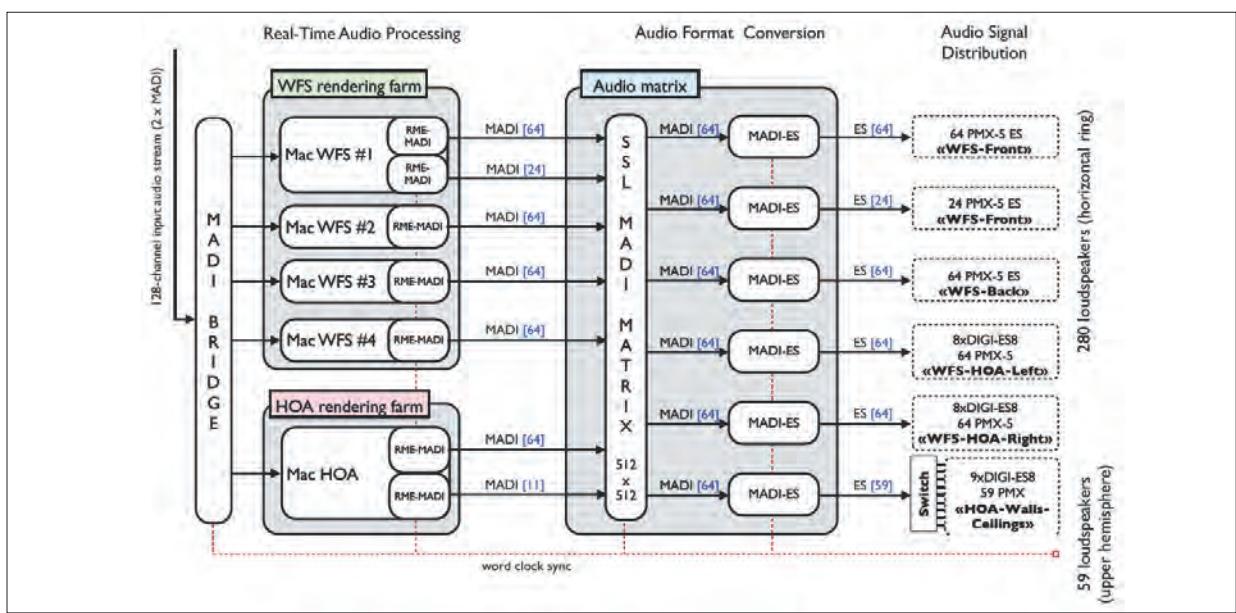


Fig. 9 : Organigramme de l'installation WFS et HOA finale

Cette possibilité est d'ores et déjà exploitée au niveau des haut-parleurs des bancs latéraux. Certains d'entre eux peuvent être affectés indifféremment soit à la reproduction WFS soit à la reproduction HOA. De même, pour la reproduction HOA, l'accent a été mis au départ sur une reproduction 3D exploitant l'hémisphère de 75 haut-parleurs et autorisant une reproduction correcte jusqu'à l'ordre 9. La présence de cette matrice permet d'envisager également une diffusion Ambisonics en 2D exploitant l'ensemble de la couronne horizontale de 280 haut-parleurs avec un ordre de reproduction par conséquent supérieur à 100.

Transduction

L'ensemble des unités de traitement et de matriçage est situé en régie. En sortie de matrice MADI, les signaux sont transmis dans l'ESPRO par flux Ethersound qui sont décodés et amplifiés soit directement par les enceintes actives des bancs WFS avant et arrière, soit par des convertisseurs et amplificateurs répartis dans le plenum de l'ESPRO ou sur le plafond pour la diffusion sur les bancs de WFS latéraux et sur l'hémisphère de diffusion HOA.



Crédit photographique Markus Noisternig



Crédit photographique Olivier Warusfel

Fig. 10 : Vues de l'installation en cours de montage et achevée.

En haut : Passage des câbles dans le plenum de l'ESPRO et installation des haut-parleurs du plafond. En bas : Murs Est et Sud montrant les réseaux linéaires pour la WFS et quelques haut-parleurs HOA

Algorithmes de traitements

L'ensemble des traitements de synthèse sonore est réalisé en temps réel dans l'environnement Max/MSP.

Wave Field Synthesis

Chaque unité prend en charge une portion du réseau de haut-parleurs, typiquement 64 sorties. Pour chaque

source en entrée, chaque unité de traitement charge un jeu de filtres correspondant au traitement à réaliser l'ensemble des haut-parleurs dont elle est responsable. Ces filtres dépendent tout d'abord de la position de la source virtuelle que l'on veut synthétiser.

La figure 11 représente l'ensemble des positions de sources qui peuvent être synthétisées pour un réseau de haut-parleurs frontal (symbolisé par la ligne noire). Les positions sont distribuées à la fois devant (sources focalisées) et derrière le réseau de haut-parleurs. Compte tenu des propriétés de la perception auditive, la résolution angulaire est maintenue constante sur l'ensemble de l'ouverture de la fenêtre spatiale représentée par le réseau de haut-parleurs tandis que la résolution en distance se relâche au fur et à mesure que les sources s'éloignent. Pour des distances virtuelles supérieures à une quinzaine de mètres, on assimile le front d'onde à une onde plane. L'ensemble des jeux de filtres correspondant à cette distribution de positions de sources virtuelles est pré-calculé (dans Matlab) et chargé dans chacune des machines. Différentes bases de données de filtres peuvent être chargées selon la configuration géométrique finale des haut-parleurs et la distribution plus ou moins dense de sources virtuelles que l'on veut synthétiser.

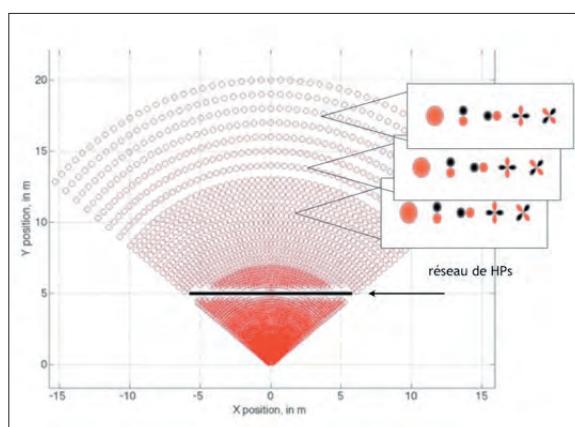


Fig. 11 : Principe de traitement audio pour la WFS. Pour chaque source, l'unité de traitement charge le jeu de filtres associés à la position et à la directivité de la source virtuelle. Le réseau de haut-parleurs est symbolisé par la ligne noire, tandis que les ronds rouges représentent l'ensemble des possibilités de positionnement des sources virtuelles

Comme symbolisé sur la Figure 11, pour chaque position de source virtuelle, plusieurs jeux de filtres sont en fait calculés et correspondent aux harmoniques circulaires des premiers ordres utilisés pour synthétiser la directivité de la source virtuelle.

Au final le traitement effectué par les unités de traitement se décompose pour chaque source en étapes successives : un étage assure la synthèse de la directivité de la source par combinaison linéaire d'un ensemble de filtres synthétisant les directivités élémentaires, un second étage assure la formation du front d'onde par ajustement de retards et gains en fonction de la position désirée de la source. Des traitements complémentaires réalisent l'égalisation des transducteurs composant la portion du réseau de haut-parleurs dont l'unité de traitement à la charge.

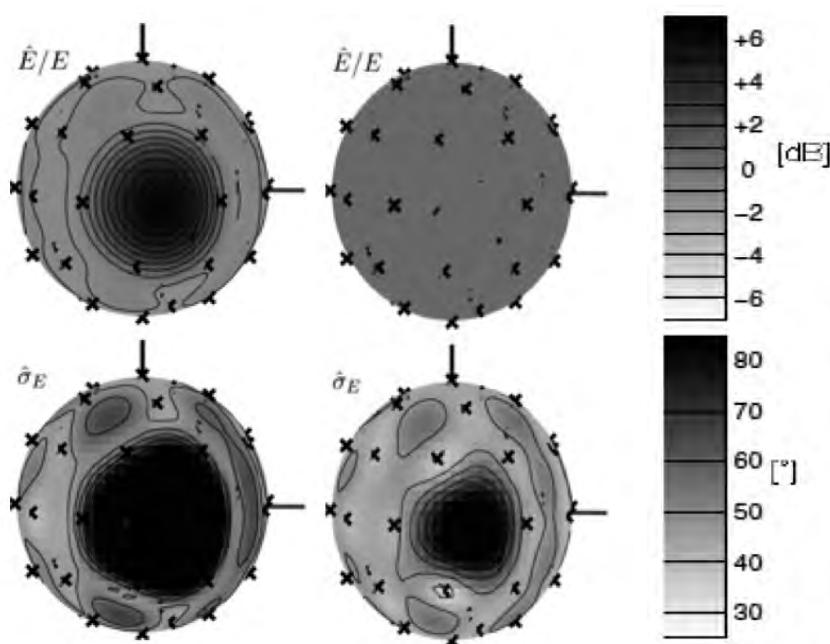


Fig. 12 : Comparaison des performances théoriques de différents décodeurs Ambisonics à l'ordre 5 en présence d'une hétérogénéité de distribution des haut-parleurs. Sont représentées d'une part, la différence relative de niveau \hat{E}/E (dB) et la largeur de source $\hat{\sigma}_E$ ($^{\circ}$) en fonction de la direction de la source virtuelle pour une distribution uniforme de 41 haut-parleurs (marqués par des croix) dont on ôte le haut-parleur central. À gauche : décodeur par identification modale. A droite : décodeur « à préservation d'énergie ». On observe pour ce dernier une restitution plus homogène du niveau sonore et un moindre étalement de la source dans la direction du haut-parleur manquant.

High Order Ambisonics

La reproduction d'une scène sonore encodée en format Ambisonics devrait s'appuyer en théorie sur un continuum sphérique de sources. En pratique, cette reproduction est approximée par le décodage des canaux Ambisonics sur un nombre limité de haut-parleurs.

Les approches de décodage les plus courantes reposent sur l'échantillonnage des harmoniques sphériques selon les directions des haut-parleurs. D'autres procèdent par identification des harmoniques sphériques et des modes d'excitation du champ sonore par les haut-parleurs [8]. Lorsqu'elles sont appliquées à des distributions régulières de haut-parleurs, ces deux approches fournissent de bonnes performances de localisation sonore. En revanche, leurs performances se dégradent sensiblement pour des distributions de haut-parleurs non uniformes, en particulier lorsque le réseau de haut-parleurs ne couvre qu'une portion de l'espace. Dans le cas particulier de l'Espace de Projection, le système de reproduction est limité à l'hémisphère supérieur. Par ailleurs, compte tenu des contraintes d'emplacement liées aux périactes, nous n'avons pas une entière liberté de positionnement des haut-parleurs.

La distribution des haut-parleurs s'éloigne donc sensiblement d'une répartition régulière et on peut noter des hétérogénéités de la restitution du niveau en fonction de la direction de la source virtuelle. Une étude a été consacrée à l'amélioration des techniques de décodage

tenant en compte le cas des réseaux de haut-parleurs non réguliers. La méthode s'inspire des travaux de M.Gerzon qui a introduit le principe de décodeurs préservant un critère d'énergie [12]. Cette méthode réduit les variations de niveau perçu dans les directions éloignées des haut-parleurs. Dans le cas de systèmes présentant de larges zones dépourvues de haut-parleurs, comme le cas d'un réseau réduit à un hémisphère supérieur, ces décodeurs doivent également s'appuyer sur de nouvelles bases de fonctions spatiales orthonormales (fonctions régularisées). L'observation de critères d'homogénéité d'énergie et de largeur apparente de la source synthétisée montre le bénéfice apporté par cette approche (Figure 12).

Le décodeur Ambisonics dit « à préservation d'énergie » [13] a été implémenté dans le Spatialisateur (librairie de traitement de spatialisation temps réel développée à l'IRCAM). Ce décodeur qui vient s'ajouter aux deux autres types de décodage déjà développés (décodage « direct-sampling » par échantillonnage de l'excitation des harmoniques sphériques sur les positions des haut-parleurs et « mode-matching » par appariement des modes du champ sonore au dispositif de haut-parleurs) fera donc l'objet d'une validation expérimentale lors de la mise en œuvre de l'installation finale du dispositif. D'autres types de filtrage visant la synthèse d'effet de proximité des sources (« Nearfield Compensated Higher Order Ambisonic ») [14][15] ont été intégrés et feront également l'objet d'études expérimentales avec le nouveau dispositif.

Utilisation artistique

Au cours de son développement, le système a été exploité par différents artistes compositeurs ou metteurs en scène désirant explorer les possibilités particulières de diffusion sonore offertes par la WFS et Ambisonics. Grâce à la mobilité d'une partie du dispositif (réseau WFS frontal) réalisée lors de la première phase de l'installation en 2008, certaines productions ont eu lieu hors les murs. Dans la pièce *Le Père* de Michael Jarrell, représentée au théâtre de Schwetzingen (Allemagne) puis au théâtre de l'Athénée à Paris, le réseau WFS était installé en nez de scène et était en particulier utilisé pour créer des voix chuchotées virtuellement situées au sein même du parterre de spectateurs. Ce même réseau, étendu à 128 haut-parleurs a également été utilisé lors des représentations théâtrales de la *Tragédie du Roi Richard II* de Shakespeare, mise en scène par Jean-Baptiste Sastre dans la Cour d'Honneur du Palais des Papes à Avignon. Le système y était principalement utilisé comme renfort électroacoustique en exploitant la possibilité de co-localiser les voix amplifiées avec les acteurs y compris au cours de leur déplacement.

En 2012, deux compositeurs ont été accueillis en résidence pour expérimenter l'ensemble du dispositif dont l'installation s'est achevée fin 2011. Les travaux issus de ces résidences prennent à la fois la forme d'esquisses musicales créées en novembre 2012 et de matériaux sonores pouvant servir à organiser des tests sur la perception auditive spatiale.

Dans son projet, la compositrice Natasha Barrett, experte de l'utilisation du format Ambisonics, s'est consacrée principalement à la manipulation et la restitution de scènes sonores au format Ambisonics aux ordres supérieurs. Grâce aux performances du système, elle a notamment pu comparer différents ordres de synthèse et de restitution de scènes sonores jusqu'à l'ordre 9. Dans sa pièce *Hidden, Values* elle superpose différents ordres de reproduction sonore ce qui permet de varier les impressions d'espace en privilégiant soit l'enveloppement et l'immersion pour les ordres faibles ou la précision de localisation et la transparence de la scène sonore pour les ordres élevés.

Dans son projet, le compositeur Rama Gottfried s'est concentré sur l'utilisation temps réel de l'ensemble du système dans une pièce pour violoncelle et électronique qui explore les relations entre l'espace des timbres associés à différents modes de jeu du violoncelle («jeté», «écrasé», «archet circulant», «un seul crin») et les différentes figures spatiales autorisées par le système. Il a notamment développé des algorithmes simulant des mouvements collectifs («boids») appliqués à des foules de grains sonores issus d'une captation en temps réel des sons produits par le violoncelle sur scène. Grâce aux propriétés particulières de la WFS et aux capacités de restitution tri-dimensionnelle d'Ambisonics ces nuées de sons peuvent traverser l'espace des spectateurs et peupler l'ensemble du volume de la salle.

Nous espérons qu'à travers cet équipement, l'Espace de Projection puisse à la fois garder son caractère exceptionnel de salle de concert expérimentale, jouer un rôle moteur pour l'exploration des technologies de spatialisation appliquées à la musique et susciter de nouvelles collaborations autour de ces questions scientifiques.

Remerciements

Cet équipement a bénéficié du soutien financier de la Région Ile-de-France (programme d'équipement Sesame), du CNRS et de l'UPMC.

Références bibliographiques

- [1] Berkhout A.J. A holographic approach to acoustic control. *J. Audio Eng. Soc.*, 36(12), pp.977-995, 1988.
- [2] Berkhout A.J., de Vries D. and Vogel P., Acoustic control by wave field synthesis. *J. Acoust. Soc. Am.*, 93(5), pp. 2764-2778, 1993.
- [3] Nicol R., Restitution sonore spatialisée sur une zone étendue : application à la téléprésence. Thèse Université du Maine, 1999.
- [4] Spors S., Buchner H., Rabenstein R. and Herbordt W., Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering, *J. Acoust. Soc. Am.* 122 (354)
- [5] Corteel E., Synthesis of directional sources using Wave Field Synthesis, possibilities and limitations. *EURASIP Journal on Advances in Signal Processing*, special issue on Spatial Sound and Virtual Acoustics. Volume 2007, Article ID 90509, doi:10.1155/2007/90509, 2007.
- [6] Gerzon M.A., Maximum Directivity Factor of n'th- Order Transducers, *J. Acoust. Soc. Am.*, 1976. 60 p. 278-280, 1976.
- [7] Daniel J., *Représentation de Champs Acoustiques, Application à la Transmission et à la Reproduction de Scènes Sonores Complexes dans un Contexte Multimédia*. University of Paris 6: Paris, 2000.
- [8] Poletti M. A., A Unified Theory of Horizontal Holographic Sound Systems. *J. Audio Eng. Soc.* 48 pp., 1155-1182, 2000.
- [9] Fazi F. and P. Nelson, A theoretical study of sound field reconstruction techniques. in 19th International Congres on Acoustics, Madrid, (2007).
- [10] Daniel, J., Evolving views on HOA : from technological to pragmatic concerns. 1st Ambisonic symposium, 2009.
- [11] Rafaely B., Analysis and Design of Spherical Microphone Arrays. *IEEE Transaction on Speech and Audio Processing*, 2005.
- [12] Gerzon M.A., Psychoacoustic Decoders for Multispeaker Stereo. in AES 93rd Conv. 1992. San Francisco, 1992.
- [13] Zotter F., Pomberger H., Noisternig M. Energy-preserving Ambisonic decoding. *Acta Acustica united with Acustica* vol. 98 pp37-47, 2012.
- [14] Sontacchi A. and R. Höldrich. Further investigations on 3D sound fields using distance coding. In DAFX-01, 2001.
- [15] Daniel J., Spatial Sound Encoding Including Near Field Effect : Introducing Distance Coding Filters and a Viable, New Ambisonic Format. in AES 23rd International Conference, 2003.

AMÉLIORER LES PERFORMANCES ACOUSTIQUES ET VIBRATOIRES

Les sous-couches DAMTEC garantissent une excellente désolidarisation des bruits de structure et sont également parfaitement adaptées à l'amortissement des vibrations en bâtiments commerciaux et industriels.



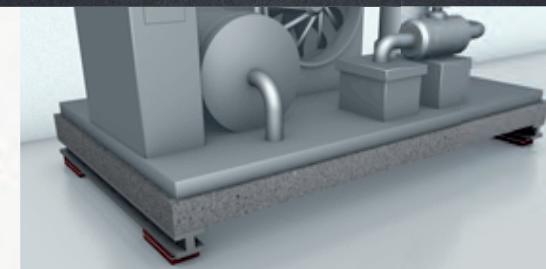
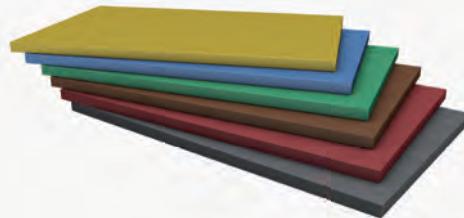
Pour en savoir plus:
Sébastien Bisteul . Tél. 06 07 89 38 51
sebastien.bisteul@kraiburg-relastec.com
www.kraiburg-relastec.com/damtec



Spécialiste des matériaux flexibles en Caoutchouc Composite

Construction, Bâtiment et Industrie

Isolation acoustique et vibratoire pour réduire les nuisances sonores et les niveaux de bruits d'impact dans les habitations, bureaux et les nuisances vibratoires dans l'industrie



RubberGreen Vibra et Vibrafoam

L'innovation en isolation sonore et vibratoire pour le bâtiment et l'industrie

Supports anti-vibratoires

