

Introduction

A la fin du XIX^e siècle, notre perception du monde sonore a été bouleversée par deux inventions: le téléphone (G. Bell, 1876) et le phonographe (Ch. Cros / T. Edison, 1877). Avec la première, il devenait désormais possible de transmettre les sons dans l'espace, et avec la seconde, de les transmettre dans le temps. Alors qu'auparavant, un son était indissociable de l'instant et du lieu où il était émis, il pouvait dès lors être *reproduit* à des *instants ultérieurs* et en des *lieux différents*, ce qui ouvrait la porte à une nouvelle expérience du monde sonore¹...

Cependant, qu'il s'agisse de les transmettre dans l'espace ou dans le temps, ces découvertes posent d'emblée le *problème de la reproduction des sons*. Reproduire un son signifie reproduire toutes ses propriétés telles qu'elles existent dans une situation d'écoute réelle. Or, un son est un *phénomène à quatre dimensions*: il évolue en fonction du *temps* et des trois coordonnées d'*espace*. Il est frappant de constater qu'un peu plus d'un siècle après l'invention du téléphone et du phonographe, le problème de la reproduction d'un son est loin d'être maîtrisé dans son ensemble. Même si la reproduction des phénomènes temporels peut être considérée comme un problème résolu, la reproduction des aspects spatiaux soulève encore de nombreuses difficultés et le domaine du son spatialisé est un secteur de recherche en pleine expansion en cette fin du XX^e siècle. L'intérêt pour le son spatialisé est motivé par ailleurs par le nombre croissant d'applications qui y font appel, qu'il s'agisse du cinéma, du monde de la réalité virtuelle en général, des applications Multimédia ou des systèmes de Télécommunication...etc...

Le travail de thèse qui est rapporté dans le présent document, s'inscrit au cœur de ce problème, puisqu'il a trait à la fois au *domaine des Télécommunications* et à la *spatialisation sonore*. Il concerne la reproduction spatialisée du son dans les systèmes de communication de groupe que sont les visioconférences. Avec ces nouveaux systèmes de télécommunication, le dialogue téléphonique se conjugue au pluriel, entre plusieurs groupes d'individus répartis en différents lieux, et l'image s'associe au son pour constituer une interface de communication audiovisuelle qui combinent des caméras, des écrans vidéo, des microphones et des haut-parleurs. Dans ce contexte, une reproduction sonore spatialisée intervient d'abord pour renforcer l'illusion d'une situation réelle et rendre ainsi l'interface de communication la plus transparente possible du point de vue des utilisateurs. De cette façon, on cherche à imiter les conditions naturelles d'une conversation, dans le souci d'une interactivité optimale. A terme, on veut donner l'impression aux différents interlocuteurs qu'ils partagent le même espace, alors qu'ils peuvent être distants de plusieurs milliers de kilomètres. Cette idée définit le concept du *mur de téléprésence*. Cependant la restitution sonore spatialisée présente d'autres intérêts à plus court terme: elle contribue à améliorer l'*intelligibilité de la parole*, de même qu'elle réduit l'effort d'*attention* dans le suivi d'une conversation.

1. Les idées contenues dans cette introduction sont largement inspirées d'une conférence présentée par D. Malham dans le cadre des journées d'études DAFX98 [Malham, 1998].

La mise œuvre d'un système de spatialisation sonore dans une application de visioconférence était donc le problème posé à l'origine de cette thèse. Les solutions envisagées pour intégrer une restitution sonore spatialisée aux systèmes de visioconférence sont le plus souvent basées sur le principe stéréophonique ou les technologies binaurales [Mahieux *et al.*, 1998] [Miyoshi & Koizumi, 1992] [Rimell, 1999], mais ces approches souffrent de plusieurs limitations que nous avons cherché à dépasser en abordant le problème de la spatialisation sonore du point de vue le plus général. Le contexte de visioconférence pose en effet des contraintes spécifiques auxquelles les techniques classiques de spatialisation sonore ne permettent pas de répondre de façon satisfaisante et robuste. La zone d'écoute doit d'abord être suffisamment grande pour englober l'ensemble des participants. Ensuite, ces derniers doivent pouvoir se déplacer au sein de la zone d'écoute en conservant une restitution correcte et stable. Enfin, des systèmes de restitution contraignants, tels qu'un casque (par exemple pour une reproduction binaurale), sont exclus.

Fondamentalement, pour reproduire un champ sonore spatialisé, il suffit de le recréer à l'identique sur une zone donnée. Ainsi, un auditeur placé à l'intérieur de cette zone, dont les dimensions devront nécessairement englober au minimum celles de la tête de l'individu, percevra un champ en tout point équivalent à celui qu'il aurait perçu dans la situation réelle, de telle sorte qu'il soit parfaitement capable de localiser les positions et les mouvements des sources sonores, comme si elles étaient réellement présentes. Cette approche de reproduction du champ physique à l'identique correspond au procédé défini par M. Jessel comme l'*holophonie* et qui s'inspire du Principe de Huygens [Jessel, 1973]. Elle consiste à enregistrer le champ acoustique sur une surface pour le reproduire à l'intérieur d'un volume et s'identifie donc à l'équivalent acoustique de l'holographie.

L'holophonie constitue l'approche la plus générale de reproduction sonore spatialisée: en d'autres termes, l'ensemble des autres techniques de spatialisation sonore n'en sont que des cas particuliers dans lesquels la reproduction sonore est simplifiée en limitant la taille de la zone de restitution ou la précision de la reproduction du champ sonore. Avec l'holophonie, la reproduction des informations spatiales est basée sur une analyse physique des phénomènes (Principe de Huygens), par laquelle les performances de reconstruction du champ sonore, notamment en terme de taille de zone d'écoute et de qualité de reproduction, sont directement reliées au nombre de microphones et de haut-parleurs utilisés à la prise et à la restitution du son, ce qui démontre l'intérêt de cette approche pour une appréhension globale du problème de la spatialisation sonore.

Bien que, dans son principe, l'holophonie soit connue depuis longtemps [Jessel, 1973], elle n'a été mise en œuvre que récemment, avec les travaux menés au Laboratoire d'Acoustique de l'U.T.D. (Université Technologique de Delft) qui, depuis une dizaine d'années, développe des systèmes de spatialisation sonore sous le concept de *Wave Field Synthesis* qui n'est autre qu'un exemple particulier d'implémentation d'un système holophonique. L'apport fondamental des recherches de l'U.T.D. a consisté à prouver qu'il était possible, certes moyennant certaines approximations, de réaliser un système holophonique opérationnel avec une complexité raisonnable.

La question de la faisabilité de l'holophonie étant ainsi résolue, la première étape du travail de thèse a visé à approfondir la connaissance du procédé holophonique, en étudiant à la fois la théorie générale contenue dans le Principe de Huygens et l'Intégrale de Kirchhoff-Helmholtz [Jessel, 1973] [Bruneau, 1983] et les résultats des travaux de l'U.T.D. [Start, 1997] [Verheijen, 1997]. Dans une seconde étape, un prototype de système holophonique adapté au contexte de visioconférence a été réalisé, puis monté dans la chambre anéchoïque du C.N.E.T.. Une fois son fonctionnement validé, ce prototype permettra d'évaluer l'apport d'un tel système dans une application de visioconférence, en particulier en le comparant à d'autres techniques de spatialisation sonore.

Le présent document se veut le reflet de cette démarche. Le premier chapitre pose le problème de la reproduction sonore spatialisée dans le contexte de la visioconférence, en soulignant les contraintes spécifiques à ce type d'application. Dans le second chapitre, la reproduction d'un champ sonore spatialisé est abordée d'un point de vue général: les différentes méthodes de spatialisation sonore (stéréophonie, technologies binaurales, système ambisonique) sont présentées en faisant ressortir leurs avantages et leurs inconvénients du point de vue du contexte de la visioconférence. La conclusion de cette étude met en évidence l'intérêt de l'holophonie par rapport au problème posé au premier chapitre.

Au troisième chapitre, les fondements théoriques de l'holophonie sont exposés à partir du Principe de Huygens et de l'Intégrale de Kirchhoff-Helmholtz. Il est également montré comment cette approche peut être simplifiée, principalement en ce qui concerne le nombre de transducteurs requis, moyennant des hypothèses compatibles avec notre contexte applicatif. Ces résultats proviennent en grande partie des travaux de l'U.T.D.

Dans la perspective de la réalisation d'un système holophonique, les deux chapitres suivants s'intéressent à deux problèmes spécifiques: l'échantillonnage spatial et la troncature du réseau de transducteurs utilisés pour la prise et la restitution du champ sonore. Le second point est davantage détaillé, car il a fait l'objet d'une étude originale, dans laquelle le phénomène de troncature, qui peut être identifié à un phénomène de diffraction, est modélisé grâce à l'intégrale de Rubinowicz. Il s'en déduit une nouvelle méthode de correction des effets de troncature, qui consiste à annuler le champ diffracté résultant de la troncature spatiale au moyen de sources additionnelles, dites sources de Rubinowicz. Cette méthode est comparée à d'autres approches, notamment celles qui sont utilisées à l'U.T.D..

Le prototype de système holophonique mis au point à l'issue de cette étude théorique est décrit au sixième chapitre. Sa conception s'inspire largement du système de *Wave Field Synthesis* de l'U.T.D. Il a fait l'objet d'une première phase de validation objective, menée à la fois en chambre anéchoïque et dans un studio expérimental de visioconférence, et dont les résultats sont présentés.

La démarche qui a été retenue tout au long de cette thèse, en privilégiant une approche générale du problème de la spatialisation sonore, a conduit à une formulation générale du principe de reconstruction de champ acoustique, au sein duquel ont pu être fondus deux approches de spatialisation sonore, considérées jusqu'alors comme parfaitement distinctes, à savoir: l'holophonie et le système ambisonique. En particulier, il a été établi que le système ambisonique n'était qu'un cas particulier de l'holophonie. Mais surtout, en les intégrant dans une formulation commune de reproduction d'un champ sonore spatialisé, on a pu effectué une comparaison directe de leurs performances d'encodage (prise de son) et de décodage (restitution) spatial, en termes de taille de zone d'écoute et de fidélité de reproduction du champ sonore, sans perdre de vue les aspects pratiques de mise en œuvre. Ces résultats incitent à reconsidérer d'autres techniques de spatialisation sonore à la lumière de cette approche holophonique généralisée.

Avant de conclure, les perspectives pour la suite de l'étude seront proposées.

