

■ LES AUTEURS



Matthieu AUSSAL est ingénieur de recherche au Centre de mathématiques appliquées (CMAP) de l'École polytechnique, à Palaiseau.

François ALOUGES est professeur de mathématiques appliquées au CMAP.

Matthieu Aussal a reçu en 2015 le prix AMIES (Agence pour les mathématiques en interaction avec l'entreprise et la société) pour sa thèse de doctorat.

ces systèmes laisse peu de doutes quant à leur avenir.

Ces casques de visualisation immersive commencent à livrer des informations acceptables et persuasives pour notre système visuel. Mais il n'en est pas de même pour ce qui est des informations fournies à notre système auditif.

De quoi s'agit-il ? Au même titre que nos yeux nous permettent de voir en trois dimensions, nos oreilles nous fournissent, inconsciemment et en continu, des informations spatiales contenues dans le son, notamment sur la position et le mouvement des sources sonores. Lorsque ces informations auditives sont absentes, voire contradictoires avec les informations visuelles, notre système cognitif le supporte mal. Parfois, il supprime même l'illusion offerte par le dispositif de réalité virtuelle, ce qui rend désagréable l'expérience vécue par l'utilisateur et peut même provoquer des nausées.

La compréhension des mécanismes physiques et physiologiques de la spatialisation sonore est donc indispensable pour offrir une expérience immersive, complète et persuasive de réalité virtuelle : s'il est facile de comprendre que nos deux yeux voient la même scène d'un point de vue légèrement différent, il est beaucoup plus difficile de savoir ce qu'entend exactement chacune de nos oreilles lorsque nous sommes plongés dans une scène sonore.

Parmi les nombreuses recherches consacrées à ce sujet, l'une des démarches consiste à simuler par ordinateur, à l'aide de modèles mathématiques, la propagation des ondes sonores, de la source du son jusqu'au

conduit auditif de l'oreille. Ces simulations numériques, qui sont l'objet du présent article, peuvent en effet livrer des résultats comparables à ceux des mesures effectuées en laboratoire, mais avec davantage de souplesse et de répétabilité, pour un coût modique. Ces « expériences numériques » autorisent la confrontation d'hypothèses, la validation ou l'invalidation de modèles, et fournissent des données utiles pour les applications destinées au grand public.

Pour mieux comprendre l'intérêt des recherches sur le son 3D et l'approche par la simulation numérique, rappelons qu'il existe déjà sur le marché des dispositifs réalisant une spatialisation du son, au moyen d'un ensemble de haut-parleurs diffusant chacun un canal sonore approprié. Les chaînes *home cinema* 5.1 ou 7.1, voire les 64 canaux proposés par le système Atmos de Dolby, en sont des exemples. Ces dispositifs permettent de positionner virtuellement des objets sonores autour de l'auditeur, en jouant par exemple sur des différences d'intensité entre les haut-parleurs, des effets de retard ou de réverbération. C'est une sorte de généralisation du système de stéréophonie bien connu.

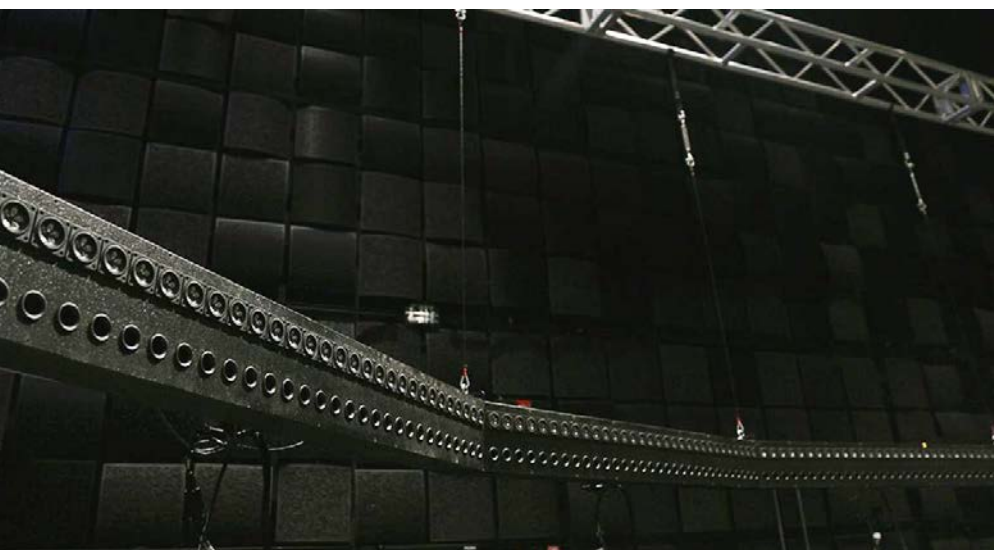
Un rendu spatial avec juste un casque

D'autres techniques plus récentes, telles que la HOA (*High Order Ambisonics*, « ambisonie d'ordre élevé ») ou la WFS (*Wave Field Synthesis*, ou « synthèse de front d'ondes »), fonctionnent sur des principes équivalents, quoique plus complexes dans leur formalisme mathématique.

Cependant, tous ces systèmes sont confrontés à la même limitation : la précision du rendu spatial est directement proportionnelle au nombre d'enceintes. Ils restent donc encombrants, malgré les efforts de miniaturisation. Aussi, bien que ces dispositifs soient tout indiqués pour équiper de grands volumes (salles de cinéma, auditoriums, salles de spectacle, etc.), ils sont incompatibles avec la portabilité et la mobilité des casques de réalité virtuelle. Il faut donc se tourner vers la technique de spatialisation sonore adaptée, à savoir la « synthèse binaurale » sur casque audio.

L'écoute binaurale désigne les perceptions auditives engendrées par une stimulation sonore des deux oreilles. L'idée de la synthèse binaurale est donc de synthétiser le son adéquat destiné à chaque oreille.

LE SYSTÈME WAVE FIELD SYNTHESIS (WFS), ou Synthèse de front d'ondes, ici celui de l'Ircam à Paris, utilise un grand nombre de haut-parleurs pour créer dans l'espace des « hologrammes sonores », c'est-à-dire des sources virtuelles donnant l'impression que c'est d'elles, et non des haut-parleurs, qu'émane le son.



Lorsqu'un son est émis dans l'espace, il arrive généralement un peu plus tôt et avec une intensité un peu plus élevée à l'une des deux oreilles (voir la figure ci-contre). Cela fournit à notre système cognitif deux indices sur la direction de la source émettrice. On les nomme ITD (*Interaural Time Delay*, « différence interaurale de temps ») et ILD (*Interaural Level Difference*, « différence interaurale d'intensité »).

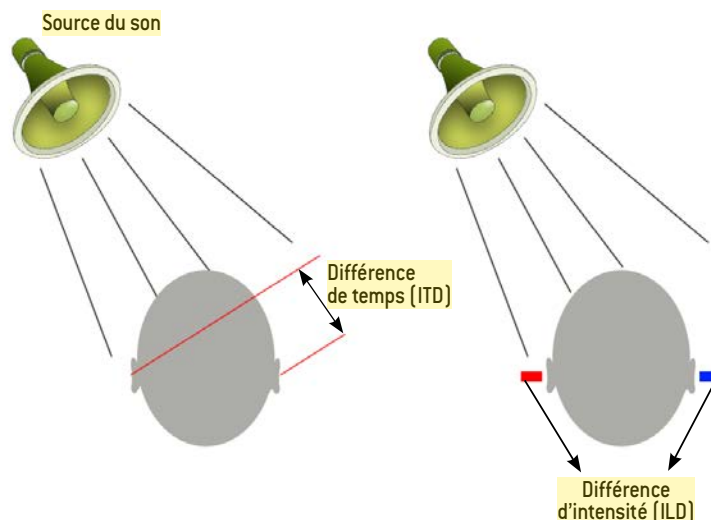
Mais ces indices de localisation sont insuffisants. Par exemple, ils ne permettent pas de différencier la provenance de sources sonores situées à la même distance des deux oreilles, c'est-à-dire dans le plan sagittal qui passe entre les oreilles. Pourtant, nous sommes capables de discriminer efficacement de telles sources sonores, provenant du dessus, de devant, de derrière, etc. Comment fait notre cerveau ?

Pour l'expliquer, il faut d'abord savoir que tout son peut être vu comme une superposition d'ondes acoustiques sinusoïdales, et que la propagation d'une telle onde dans un milieu est perturbée par les détails géométriques des obstacles qu'elle rencontre.

Ce phénomène de diffraction dépend de la longueur d'onde (distance entre deux pics successifs de l'onde sinusoïdale considérée). Des détails géométriques petits par rapport à la longueur d'onde modifient peu la propagation : l'onde correspondante n'y est pas sensible. Inversement, des détails géométriques de taille comparable ou supérieure à la longueur d'onde modifient notablement la propagation et perturbent l'acoustique. C'est d'autant plus le cas quand la forme de l'objet est complexe, comme pour le pavillon de l'oreille.

La forme de la tête et des oreilles joue

Lorsqu'un son émis par une source se propage jusqu'au tympan, la forme de la tête et des oreilles de l'auditeur diffractent toutes ses composantes dont les longueurs d'onde sont, pour l'essentiel, inférieures au mètre. En termes de fréquences, cela se traduit par une modification du spectre pour toutes les fréquences supérieures à quelques centaines de hertz – c'est-à-dire une modification des amplitudes et phases des ondes correspondant à ces fréquences. Il s'ensuit que les différents tons qui composent le signal sonore initial



LA DIRECTION DE LA SOURCE SONORE se traduit dans divers indices physiques dont se sert le cerveau pour la déterminer. Deux d'entre eux sont la différence interaurale de temps (ITD) et la différence interaurale d'intensité (ILD), qui mesurent les différences, entre les deux oreilles, du temps d'arrivée du son et du niveau perçu.

seront perçus avec des intensités relatives différentes de celles de l'émission. Or deux sons provenant d'endroits différents ne rencontrent pas les mêmes obstacles et les mêmes détails. Ils sont donc modifiés différemment. Ce troisième indice, judicieusement exploité par notre cerveau, vient compléter la panoplie perceptive et enrichit profondément notre capacité à localiser les sons dans tout l'espace.

Par extension, deux personnes ayant des morphologies – en particulier d'oreilles – différentes n'entendront pas de la même façon, surtout dans les hautes fréquences. Ainsi, chacun a son propre univers sonore !

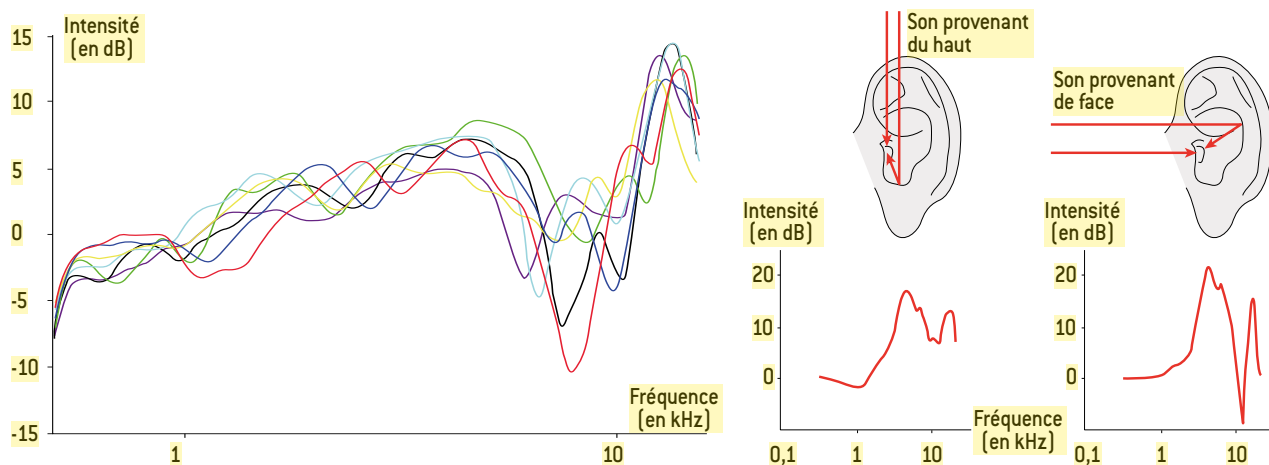
Du théâtrophone à la synthèse binaurale

En synthèse binaurale, on cherche à exploiter les indices de localisation portés par l'ITD, l'ILD et les déformations fréquentielles que nous venons de mentionner. Ces trois indices sont codés dans un ensemble de fonctions nommées HRTF (*Head-Related Transfer Functions*, c'est-à-dire « fonctions de transfert relatives à la tête »).

Plus précisément, les fonctions HRTF relient, pour chaque fréquence sonore et chaque direction d'incidence, la pression acoustique à l'entrée du canal auditif à la pression acoustique que l'on mesurerait en l'absence de l'auditeur, au milieu du volume qu'occuperait la tête de ce dernier. Ces fonctions constituent des « filtres », qui caractérisent mathématiquement les transformations acoustiques dues à la morphologie de la tête, et sur la base desquelles on réalise la synthèse binaurale.

Fréquence et longueur d'onde

La fréquence f est liée à la longueur d'onde λ par la relation $f = c/\lambda$, où c est la vitesse de propagation. Pour le son, cette vitesse est égale, dans l'air, à 340 mètres par seconde.



POUR UNE DIRECTION D'INCIDENCE donnée, le son qui parvient au pavillon de l'oreille dépend de la morphologie de la tête de l'auditeur et de son pavillon (à gauche, différentes courbes correspondant à différents

auditeurs). Les variations sont plus importantes pour les sons aigus (plus de 1 kHz). Pour un même individu, on a aussi des variations selon la direction d'incidence (graphiques de droite).

Avant d'expliquer plus précisément en quoi consiste la synthèse binaurale, on peut retracer brièvement sa préhistoire. La première expérience de spatialisation sonore remonte à la fin du XIX^e siècle, avec le théâtrophone (1881) de Clément Ader. Il s'agissait du premier service de retransmission en direct de spectacles, combinant à la fois téléphonie et stéréophonie (voir l'illustration ci-dessous).

Dans les années 1930, l'ingénieur britannique Alan Blumlein introduisit l'enregistrement stéréophonique, tandis que le physicien américain Harvey Fletcher déposa un brevet décrivant un système de téléphone binaural.

L'idée du système de Fletcher était d'utiliser une tête artificielle pour effectuer une prise de son binaurale, à l'aide d'un microphone placé au creux du pavillon de chaque oreille du mannequin, ce dernier devant imiter au mieux la morphologie humaine. Les ondes sonores sont alors diffractées par les détails morphologiques du mannequin, et le son ainsi transformé est enregistré par les microphones. Cet enregistrement stéréophonique particulier peut ensuite être écouté sur n'importe quel casque audio, ce qui donne un résultat souvent spectaculaire (écouter par exemple *Virtual Barber Shop*, sur www.youtube.com/watch?v=IUDTlvagjJA).

LE THÉÂTROPHONE conçu par Clément Ader retransmettait le son capté par deux microphones placés sur la scène de l'opéra ou du théâtre : une sorte de stéréophonie en direct, par téléphone.

Il faut souligner que la sensation d'espace procurée par le dispositif de Fletcher est uniquement perceptive : il n'y a pas de restitution physique d'un quelconque champ sonore, contrairement aux systèmes d'enceintes évoqués précédemment. Toutefois, même si l'approche de Fletcher reste aujourd'hui inégalée en termes de fidélité, tant pour le respect des timbres que pour le rendu spatial, elle nécessite *de facto* un enregistrement préalable à la restitution. Elle ne peut donc pas être interactive. Ainsi, ce procédé n'est pas utilisable dans des conditions de réalité virtuelle, où le système doit être dynamique, avec une adaptation en temps réel aux positions des sources et de l'auditeur.

Aussi, depuis les années 2000, avec l'essor de la puissance des ordinateurs, une

seconde méthode – la synthèse binaurale – a vu le jour dans de nombreux laboratoires, en France (l'Ircam, le Laboratoire d'acoustique musicale, France Telecom, consortium Bili, etc.) et à l'étranger (institut Fraunhofer en Allemagne, université de Californie à Davis, projet Orpheus, etc.). Afin de la mettre en œuvre, il faut d'abord déterminer les filtres HRTF d'un sujet (ou d'un mannequin de morphologie suffisamment proche). Pour ce faire, on place ce dernier au centre d'une sphère de plusieurs mètres de diamètre sur laquelle sont répartis des haut-parleurs. Pour chacun de ces émetteurs sonores, on mesure le son capté par des microphones placés dans les oreilles du sujet. De l'ensemble de ces mesures on déduit les fonctions HRTF.

Vient ensuite la synthèse binaurale proprement dite : si l'on veut donner à l'auditeur la sensation qu'un son provient d'une direction donnée, il suffit de traiter le son en question avec le filtre HRTF correspondant à cette direction. Pour un son numérisé, cela revient à effectuer un calcul sur ce son et à transmettre le résultat au casque de l'auditeur.

De longues mesures expérimentales...

Ainsi, par cette méthode, tout contenu sonore peut être spatialisé, au prix d'un effort de calcul. Mais la tâche reste raisonnable, car le traitement numérique est réalisable sur un ordinateur standard, voire un téléphone mobile. De plus, lorsqu'on effectue ledit calcul en temps



© DR