

Le son 3D dans les futurs services de télécommunication

Determination of occupational noise exposure : a progressive approach combining three levels of analysis

R. Nicol, M. Emerit, L. Gros, G. Pallone,
J. Palacino, S. Moulin
Orange Labs
2, avenue Pierre Marzin
F-22307 Lannion CEDEX,
E-mail : rozenn.nicol@orange-ftgroup.com

Résumé

Le son 3D représente une opportunité d'enrichissement pour les télécommunications. Pour les services conversationnels, associée ou non à une image, la spatialisation de la voix des locuteurs permet de se rapprocher des conditions d'une conversation naturelle, gage d'un meilleur confort d'écoute. Pour les services concernant la diffusion de contenus (TV, VoD, musique...), le son 3D améliore l'immersion, et par suite la qualité d'expérience. Pour rendre tout ceci possible, il faut développer des technologies à la fois pour encoder, transmettre, décoder et adapter au système d'écoute (incluant la personnalisation individuelle à l'auditeur) des flux audio 3D. L'article présente un panorama des principaux enjeux et des outils pour y répondre.

Abstract

3D audio is a great opportunity to enhance telecommunication. For instance, spatialisation of the speaker's voice brings the communication closer to natural and interactive communication for a better comfort of listening. When providing audio or audiovisual content (TV, VoD, music streaming...), 3D audio improves the immersion, and consequently the Quality of Experience. For all this, technologies are needed to encode, transmit, decode and adapt to the playback system (including the individual customization to the listener) 3D audio streams. The paper will present an overview of the main challenges and the tools available to answer them.

Aujourd'hui, plus que jamais, le domaine des télécommunications est en évolution accélérée : augmentation des capacités et étendue des réseaux, éventail des services proposés par les opérateurs, type et fonctionnalités des terminaux, usage en mobilité ... Même si le multimédia tend à se généraliser, l'audio reste un moyen privilégié pour communiquer. Longtemps limitée à la bande passante [300 Hz – 3 400 Hz], la téléphonie a connu dans les années 2 000 une première révolution avec l'extension de cette bande passante à [50 Hz – 7 kHz]. Définie comme la «bande élargie», elle offre une amélioration considérable de la qualité de la restitution de la voix. La révolution suivante a été le passage du son monophonique au son multicanal permettant la spatialisation de la voix des interlocuteurs : c'est la conférence audio spatialisée qui ouvre la possibilité de converser simultanément avec plusieurs personnes. Cependant, ces services conversationnels ne représentent plus qu'une partie des activités des opérateurs de télécommunication qui s'étendent maintenant à la diffusion des contenus audiovisuels (télévision, films, musique) avec, par exemple, les services de vidéo/musique à la demande. Dans ce contexte, le son est d'emblée stéréophonique, voire multicanal.

En termes de technologies, une première question à traiter est le transport des flux audio 3D qui peuvent comporter plusieurs dizaines de canaux et qui requiert le développement de codecs¹ audio dédiés. Une question tout aussi cruciale est celle de la compatibilité des contenus audio multicanaux avec les équipements (système stéréophonique, dispositif 5.1, casque) dont dispose l'utilisateur pour les écouter. Face à l'hétérogénéité des formats et des moyens matériels, il s'avère nécessaire de proposer des outils pour adapter les contenus au dispositif de restitution, voire même à la morphologie de l'auditeur pour une écoute au casque.

Telle se définit la problématique actuelle de l'audio 3D dans le contexte des télécommunications qui va être détaillée dans le présent article, avec un focus sur la conférence audio spatialisée qui représente la première étape de la conquête des télécommunications par l'audio 3D, avant d'en présenter les prochains jalons en relation avec la téléphonie mobile et la vidéo 3D.

¹ Un codec est un dispositif capable de compresser et/ou décompresser un signal numérique

Le téléphone en 3D ?

Plus d'un siècle après la création du «Théâtrophone» (1881) de Clément Ader [1], premier service de retransmission «live» de spectacles, combinant à la fois téléphonie et son spatialisé, qui a séduit aussi bien Victor Hugo [2] que Marcel Proust, il faut paradoxalement constater que, dans notre usage quotidien, le téléphone est aujourd'hui toujours monophonique.

Pourtant l'idée d'enrichir la communication téléphonique avec la spatialisation sonore afin de se rapprocher des conditions naturelles d'écoute, n'est pas nouvelle, du moins dans les laboratoires de recherche. Dès 1925, Harvey Fletcher a déposé un brevet décrivant un système de «téléphone binaural» [3] où il proposait d'utiliser une tête artificielle pour réaliser une prise de son binaurale et ainsi restituer à l'auditeur les indices acoustiques d'une écoute naturelle (i.e. la pression induite à l'entrée des conduits auditifs). À la fin du siècle dernier, le concept de «téléprésence» est introduit : l'objectif est de donner l'illusion aux différentes personnes en communication qu'elles sont présentes dans le même espace, alors qu'elles peuvent être distantes de plusieurs dizaines de kilomètres [4]. La téléprésence est l'aboutissement ultime de la visioconférence associant l'image au son. Que reste-t-il aujourd'hui de tous ces rêves ?

Du système «Varèse» au «Mur de téléprésence»

Au sein des équipes de recherche de France Telecom (anciennement CNET, Centre National des Télécommunications), les premières expérimentations concernant le son spatialisé ont porté sur la visioconférence haut de gamme illustrée par le système «Varèse» [5]. Ce système repose sur une salle équipée d'écrans vidéo pour une restitution visuelle en grandeur réelle des participants distants, et d'une rampe de haut-parleurs pour la reproduction de leur voix. La voix de chaque personne est captée par un microphone placé devant lui, avant d'être spatialisée par synthèse pour alimenter les haut-parleurs de façon à restituer la localisation des locuteurs en cohérence avec l'image. La technique de spatialisation est simple et repose sur le contrôle d'une différence d'intensité appliquée sur la paire de haut-parleurs encadrant la position de la source virtuelle désirée (technique de panoramique d'intensité). Les spécifications des salles sont assez proches de celles d'un studio de télévision. Les deux (voire trois) salles sont mises en communication via un réseau ATM (*Asynchronous Transfer Mode*) en point à point, les flux audio et vidéo étant transmis séparément.

Le flux audio est encodé au format TDAC (codec propriétaire défini comme «*Time-Domain Aliasing Cancellation*»), tandis que le flux vidéo est transmis au format H263. Le système Varèse mise sur la qualité audiovisuelle du service : les conditions se rapprochent d'une réunion en vis-à-vis. Mais cet objectif de qualité est obtenu au prix de contraintes assez lourdes : il nécessite la construction d'une installation dédiée dont l'usage est soumis à réservation. Le «Mur de Téléprésence» (cf. Fig. 1 & 2) en propose une déclinaison plus souple, sous la forme d'une fenêtre de communication virtuelle installée dans un lieu de convivialité et ouverte en libre service pour favoriser les échanges informels, par exemple entre les salariés de deux sites d'une même entreprise.

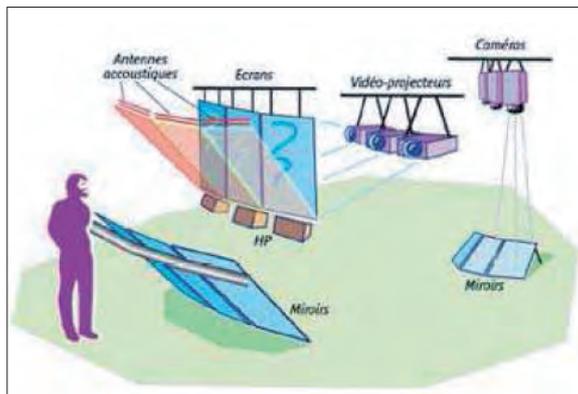


Fig. 1 : Synoptique du mur de téléprésence



Fig. 2 : Illustration d'une réunion entre 2 équipes de Lannion et Paris par le mur de téléprésence

Pour répondre à l'objectif de téléprésence, la spatialisation est enrichie pour se doter de la technique WFS (*Wave Field Synthesis*) [6] afin d'offrir une zone d'écoute étendue, dans laquelle les auditeurs sont libres d'évoluer.

La conférence audio spatialisée

De par leurs contraintes de mise en œuvre, les systèmes précédents sont d'un usage limité. *A contrario*, la conférence audio offre une solution séduisante par sa simplicité pour proposer des conversations téléphoniques à plusieurs participants. La voix étant un vecteur privilégié de communication, la conférence audio reste très attractive, vis-à-vis de la visioconférence, afin de limiter les transports et favoriser les réunions à distance, surtout si elle est couplée à des outils de travail collaboratif. Cependant, pour constituer une alternative suffisamment attractive, notamment dans le cas de réunion de longue durée, la conférence audio se doit de gagner en qualité et en confort d'écoute ; avec la spatialisation de la voix des locuteurs, la conférence audio spatialisée répond à cet enjeu, en apportant une meilleure intelligibilité (cf. effet «*cocktail party*») et en se rapprochant des conditions naturelles d'une conversation, essentiellement en termes d'immersion.

Une étude a montré que le gain qualitatif apporté par la spatialisation est du même ordre que celui perçu pour l'extension de la bande passante.

On peut aussi tirer bénéfice de l'ergonomie de la spatialisation sonore, en choisissant la position des locuteurs selon l'intérêt qu'on y porte, voire organiser ponctuellement un aparté.

Différentes configurations peuvent être envisagées : conférence entre plusieurs postes individuels, ou entre une salle dans laquelle plusieurs participants sont réunis et des postes individuels, ou encore entre plusieurs salles. Chacun de ses scénarios impliquent différents choix technologiques.

Pour un poste individuel, le dispositif le mieux adapté pour la prise et la restitution du son est un micro-casque. La spatialisation des voix est alors réalisée *a posteriori*, en synthétisant des sources virtuelles à des positions arbitraires ou selon les préférences de l'utilisateur. La restitution se faisant sur casque, la spatialisation binaurale est la technologie privilégiée. En revanche, pour un groupe de personnes, il convient de mettre en œuvre des outils de prise de son spatialisée, tels qu'un couple stéréophonique, une tête artificielle ou un microphone HOA (*Higher Order Ambisonics*). La restitution sonore se fait sur un dispositif multi haut-parleurs conforme au format de spatialisation des signaux enregistrés. Cependant, dans ce cas, il faut noter que la présence simultanée de plusieurs microphones à proximité d'un ensemble de haut-parleurs pose le problème de l'écho acoustique multivoie qui n'a pas encore trouvé de solution générale.

Aujourd'hui en phase d'évaluation en interne chez Orange, la conférence audio spatialisée sera proposée aux entreprises à brève échéance, avant d'être ouverte au grand public. La solution développée chez Orange repose sur un pont de conférence logiciel tout IP qui réalise un mixage spatialisé pour les clients connectés à l'aide d'un client VOIP permettant un rendu multicanal (rendu stéréo binaurale sur micro-casque par exemple) [7]. Cette solution est potentiellement accessible depuis un ordinateur ou un téléphone mobile. Un mixage monophonique est réalisé pour les terminaux classiques. Cette solution utilisant un pont mixeur et des codecs normalisés (G711, G722, G722.1, ...) présente l'avantage d'être compatible en mode monophonique avec le parc de terminaux existants; à la différence de solutions comme Dolby Axon ou Murmur qui utilisent des ponts répliquants intelligents avec des codecs propriétaires. Un pont répliquant ne mixe pas les flux audio : il se contente de les dupliquer et de les router vers les différents clients connectés. Dans ses futures évolutions, elle pourra alors être étendue du service conversationnel ciblé sur la voix au partage d'ambiances sonores, dans lequel, à l'image du Théâtrophone, l'utilisateur pourra faire partager à son interlocuteur un concert ou tout autre événement, en bénéficiant de l'immersion du son 3D.



Fig. 3 : Prototypage de terminal pour la conférence audio 3D

Qualité d'expérience

L'introduction de la spatialisation sonore sert à enrichir les services de télécommunication, mais elle permet aussi d'améliorer la qualité d'expérience (ou QoE pour *Quality of Experience*) de l'utilisateur, comme l'illustre la conférence audio 3D en proposant des réunions téléphoniques avec un confort d'écoute et une interactivité proches d'une conversation naturelle. Des méthodologies et des métriques, à la fois objectives ou subjectives, sont disponibles pour mesurer la qualité perçue en contexte conversationnel ou pour la diffusion de contenus (télévision, cinéma, musique).

Cependant, la majorité des outils existants restent limités à un contexte d'écoute monophonique (mesure de qualité vocale en communication [8], validation de codecs [9-10], mesure de sonie [11]...) et ont donc besoin d'être modifiés pour prendre en compte une écoute binaurale sur casque ou sur un dispositif multi haut-parleurs. Il existe néanmoins des outils standardisés pour évaluer la qualité audio spatiale (cf. notamment les normes ITU-R BS 1116, BS 1534, BS 1284) mais ils ne s'avèrent pas pleinement satisfaisants, dans la mesure où ils ont été définis spécifiquement pour le son stéréophonique ou multicanal de type 5.1, qu'ils mesurent une note de dégradation, et que la grille d'évaluation porte principalement sur la qualité audio globale, la qualité de l'image frontale et la qualité d'enveloppement sonore. Ils nécessitent *a minima* d'être revisités avant d'être appliqués à d'autres technologies de spatialisation sonore dont les propriétés de la restitution sonore, notamment en termes d'étendue de l'espace (i.e. spatialisation véritablement et totalement 3D) et de précision de la localisation des sources virtuelles se démarquent radicalement de la stéréophonie 2.0 ou 5.1 [12].

Par ailleurs, dans le contexte de systèmes de haute qualité, la mesure de dégradation héritée du contexte du codage audio s'avère inadaptée, du fait qu'en l'absence de dégradation au sens propre, les aspects à évaluer concernent plutôt la qualification de la perception (i.e. description et analyse des dimensions perceptives). En dehors du champ des télécommunications, il existe des méthodologies d'évaluation de la qualité du son spatialisé [13-14-15], mais ces dernières se focalisent sur les contenus de type musical.

Pour les contenus de type parole, la mesure d'intelligibilité est la principale métrique disponible. Or, ces contenus ne sont plus représentatifs des usages dans les contextes actuel et futur des télécommunications dont l'évolution nécessite de proposer, et de standardiser, de nouvelles métriques pour rendre compte, par exemple, de la qualité globale perçue d'une conférence audio spatialisée au-delà du seul gain en intelligibilité.

La mise en œuvre du son spatialisé en situation d'interactivité ou de réalité virtuelle requiert aussi de nouvelles méthodologies pour évaluer la qualité de spatialisation en termes de facilité pour exécuter une tâche, ou de contribution à la sensation d'immersion dans un monde virtuel.

Codecs audio 3D

Les codecs disponibles aujourd'hui pour traiter des flux audio spatialisés se limitent aux contenus stéréophoniques ou multicanaux de type 5.1. Il existe d'une part les formats propriétaires proposés par Dolby (par ex. DD pour Dolby Digital, DD+, ou Dolby Pro Logic), DTS ou Windows Media Audio, et d'autre part des solutions standardisées (MPEG Surround, HEAAC). On distingue 2 stratégies : soit les canaux sont codés séparément ou par paires (Dolby, DTS, HEAAC),

	SeTup Box Orange	Débit (kbit/s)	Mobile	Débit (kbit/s)	PC (Web)	Débit (kbit/s)
TV Live	SDTV : MPEG1L2	128	Edge : AMR-8k	6,7	SD & HD WM9.1-48k (2.0)	64
	HDTV : MPEG1L2 ou AC3 (2.0 & 5.1)	192 384	UMTS :AMR-8k	12,2		
	FTTH-SDTV : MPEG1L2	256	«HD» (HSDPA ou 3G RAB384): HE-AACv1-32k ou AMR8k ou HE-AACv2-48k	24 12,2 24		
	FTTH-HDTV : MPEG1L2 ou AC3 (5.1)	256 384	«HD» (HSDPA ou 3G RAB384): HE-AACv1-32k ou AMR8k ou HE-AACv2-48k	24 12,2 24		
	Mosaïque Orange & PiP : 17xMPEG1L2 (mono)	64				
	Premium HD : DTS-Core	768	Hotspot Wifi : WM9.1-32	48		
VoD	SD : MPEG1L2 ou AC3 2.0 ou AC3 5.1	128 128 384	Cf. TV Live		WMA MP3	
	HD : MPEG1L2 ou AC3 2.0 ou AC3 5.1	192 192/128 384				
	SDTV (à venir) : DTS neural					
Musique & Radio	Pas encore disponible		MusicMix : MP3		Mix : MP3	
			Musique Collection : MP3		Collection : MP3	
			Musique Premium Deezer : MP3	128	Musique Premium Deezer : MP3	320
			MusicStore : AAC & AAC+	64 48/32	Deezer : MP3	128
			Radio profile : AMR-NB-8k & AAC-LC-22k	7,95 32	MusicStore : MP3	256/320

Tabl. 1 : Illustration des codecs utilisés dans les différents services de diffusion de contenus d'Orange

soit un schéma spécifique audio 3D est mis en œuvre en représentant le flux par un downmix² mono ou stéréo associé à des paramètres de spatialisation correspondant à des différences inter-canaux de temps, de niveau ou de cohérence (MPEG Surround). En pratique, les codecs utilisés dépendent du terminal (Setup Box, téléphone mobile, ordinateur), du réseau et de l'application (TV, VoD ou musique), comme l'illustre le tableau 1 qui liste les codecs implémentés dans les différents services proposés par Orange. La figure 4 présente la qualité mesurée pour les principaux codecs disponibles pour les contenus 5.1 en fonction du débit. Le principal enjeu porte sur la réduction des débits : ainsi MPEG Surround rend possible la diffusion 5.1 à un débit compris entre 32 et 160 kbit/s. Au-delà on utilisera le codec HE-AAC. La qualité de MPEG Surround est acceptable pour les débits inférieurs à 64 kbit/s, bonne entre 64 et 128 kbit/s et excellente au-delà.

Cependant, le panorama des formats audio 3D s'enrichit continuellement de nouvelles propositions rivalisant pour offrir un nombre croissant de canaux (par exemple 7.1, 9.1, 10.2, 22.2 ou HOA), pour lesquelles le codage paramétrique de type MPEG Surround n'est plus forcément le mieux adapté. A minima il apparaît nécessaire

de revisiter, voire de repenser les schémas de codage audio 3D. Mais surtout, face à une diversité toujours plus grande des formats audio 3D, les représentations basées «objet» où chaque source sonore est représentée par ces coordonnées spatiales s'annoncent comme une alternative avantageuse aux représentations conventionnelles basées canal où la scène sonore est représentée par l'ensemble des signaux destinés à alimenter un dispositif de haut-parleurs.

Les représentations basées « objet » sont en effet intrinsèquement compatibles avec n'importe quel dispositif d'écoute. Une réflexion est actuellement menée sur ces questions, notamment au sein du comité de standardisation MPEG. Un autre axe d'étude concerne le développement de codecs spécifiques pour les nouveaux formats WFS et HOA qui sont basés sur une représentation radicalement différente de l'approche conventionnelle stéréophonique ou multicanal. Enfin, il faut souligner que pour l'instant, les schémas de compression n'exploitent pas les propriétés de la perception audio spatiale, où l'on peut espérer des pistes d'amélioration, par exemple en tirant profit des limites de la résolution spatiale, ou encore des phénomènes de masquage spatial [16].

² Un downmix consiste à réduire le nombre de canaux d'un flux audio multicanal.

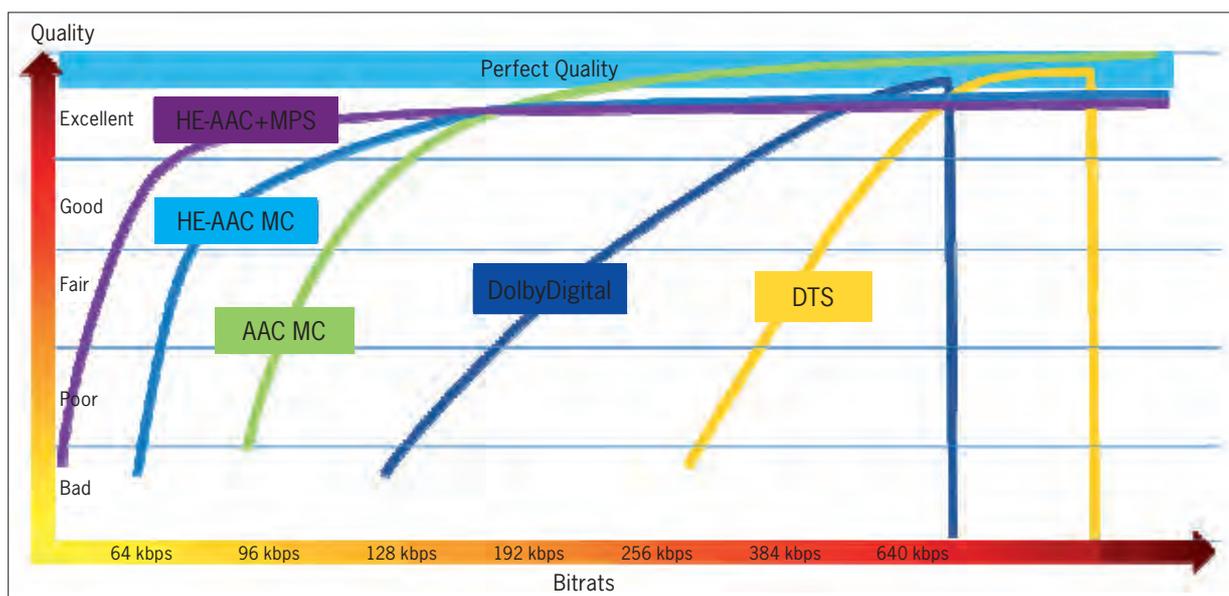


Fig. 4 : Qualité des codecs multicanal 5.1 en fonction du débit

L'audio 3D en mobilité

Une récente évolution des usages concerne l'écoute en mobilité : les terminaux mobiles, de type smartphone ou tablette tactile, deviennent aujourd'hui le principal media pour accéder aux contenus de type musique ou vidéo. Même s'ils conservent des moyens limités, quoiqu'en croissance continue, ils offrent pour la plupart une écoute stéréophonique, ouvrant la porte à la technologie binaurale.

La spatialisation binaurale constitue une réelle aubaine pour les terminaux mobiles dont elle est capable de démultiplier virtuellement les moyens de reproduction sonore. La synthèse de haut-parleurs virtuels que permet la synthèse binaurale offre en effet la possibilité d'écouter sur casque n'importe quel type de contenus audio multicanal (5.1, ou HOA). En conséquence, via l'adaptation binaurale, les terminaux mobiles sont compatibles avec tous les formats audio 3D, dès lors que les codecs pour transmettre les contenus dans ces formats sont disponibles, et intégrés dans les terminaux.

Pour la prise de son, elle reste aujourd'hui monophonique, mais des dispositifs ont été récemment proposés pour intégrer un microphone stéréophonique au terminal (Iphone par exemple). Le partage d'ambiance sonore spatialisée (concert, événement festif...) est donc aujourd'hui une réalité technique qui ne demande qu'à être enrichie avec des moyens plus performants de prise de son 3D [17].

En juin 2011, une expérience a été menée pour tester l'apport du son 3D dans le contexte des terminaux mobiles. L'opération a consisté en une restitution *live* sur Ipad d'un opéra («L'enlèvement au Sérail» de Mozart) donné à l'Opéra de Rennes (cf. Fig. 5). Pour la prise de son, 4 dispositifs ont été mis en œuvre : un couple stéréo AB, un microphone Soundfield® (ambisonique ordre 1), et deux têtes artificielles (Head Acoustics, Neumann KU100). Une écoute sur casque était proposée, moyennant des traitements pour adapter les contenus stéréophoniques ou ambisoniques.

Un test d'écoute (méthodologie MUSHRA [12]) a comparé la qualité des contenus et a montré que la prise de son binaurale naturelle (i.e. par une tête artificielle) était préférée [17]. Cependant l'enregistrement par le microphone Soundfield® s'en rapprochait sensiblement, à condition d'appliquer un décodage binaural performant.

Audio 3D et vidéo 3D

Dernière (r)évolution, l'essor de la vidéo 3D (au cinéma, ou à la télévision comme pour la chaîne Orange 3D) pose de nouvelles questions au son 3D. Par défaut, le son associé à l'image 3D est certes spatialisé, avec a *minima* la stéréophonie ou le son multicanal de type 5.1, mais jusqu'à présent aucun autre format audio 3D n'a été mis en œuvre. Or, l'ajout de la profondeur visuelle appelle le rendu de la profondeur sonore, ce qui n'est possible qu'avec les technologies WFS [18] ou ambisonique [19]. Cet aspect soulève la question de l'audio 3D en contexte de multimodalité, c'est-à-dire en présence d'autres modalités sensorielles [20-21]. Concernant les interactions entre audition et vision, on sait que l'information spatiale est le privilège de la vision, ce qui conduit à l'effet «ventriloque» dans lequel la localisation des percepts auditifs est attirée par les percepts visuels associés. Cet effet a été démontré pour la localisation en azimut, mais reste à explorer pour la perception de la distance [22]. Les enjeux sont les suivants : si l'on observe que la vision domine également la perception de la profondeur, on peut relâcher les exigences sur la restitution de la spatialisation sonore. Cependant, on peut aussi observer que la qualité perçue de l'image 3D est susceptible d'être améliorée par le rendu de la profondeur sonore, permettant ainsi de renforcer l'effet audiovisuel 3D global. Ces questions font l'objet d'études en cours [23].

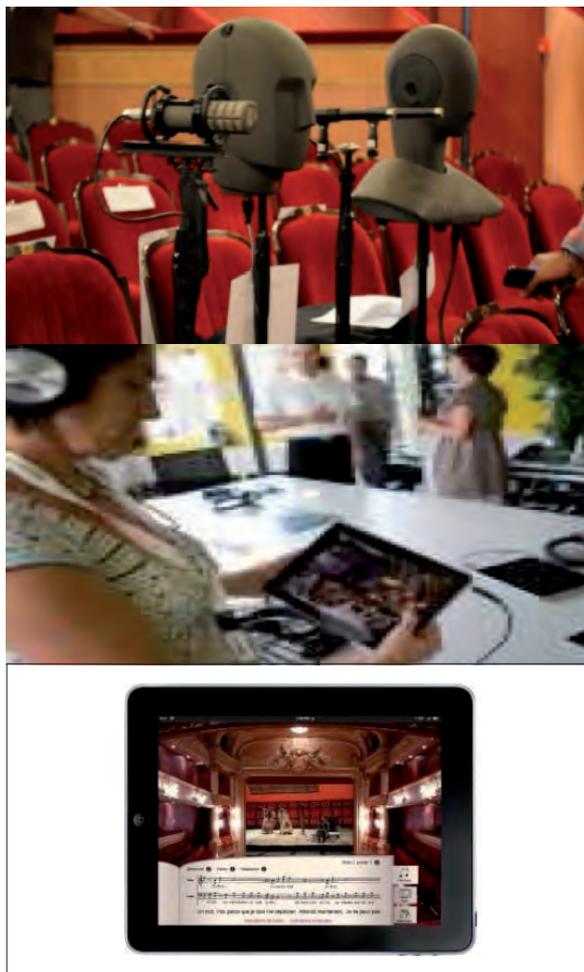


Fig. 5 : Opération de «L'enlèvement au Sérail» (Opéra de Rennes) : Dispositifs de prise son audio 3D et restitution sur iPad

En complément des enjeux scientifiques, il faut aussi résoudre un point bloquant que constitue l'équipement audio 3D du grand public : il est difficile d'envisager de tapisser de haut-parleurs les murs du salon. Il est donc nécessaire de développer des outils pour adapter les contenus audio 3D à des dispositifs non idéaux, voire sous-optimaux, de reproduction sonore. Plus généralement, pour faciliter la diffusion des contenus audio 3D, il importe de coupler le rendu audio 3D à des traitements additionnels (potentiellement implémentés dans la *setupbox*) pour :

- corriger les erreurs de positionnement des haut-parleurs, ainsi que compenser l'effet de salle du lieu d'écoute [24-25], afin d'optimiser la qualité spatiale quelles que soient les conditions d'écoute,
- adapter le format d'entrée au format du dispositif de restitution, notamment par des traitements de «*downmix*» ou «*upmix*» permettant de garantir la meilleure qualité en dépit de l'incompatibilité des 2 formats,
- adapter le rendu binaural à la morphologie de l'auditeur afin de garantir une bonne localisation des sources notamment dans la zone définie par le champ de vision,
- rendre les technologies simples à utiliser, soit par l'information du grand public, soit par des outils intégrés pour identifier la configuration matérielle en vue de fournir les traitements (codecs) les mieux adaptés, ainsi que pour diagnostiquer les pannes.

Conclusion

Le monde des télécommunications est à la fois riche d'opportunités pour le son 3D, et en continuelle (r)évolution, ce qui oblige les technologies à s'adapter en permanence aux nouveaux terminaux et aux nouveaux usages. Ce domaine constitue ainsi un champ fertile d'expérimentations et d'explorations. Certains des prochains enjeux sont déjà identifiés : d'abord la multi-modalité lorsque le son se combine aux autres sens (vision, tactile, proprioception...), mais aussi des interfaces où les transducteurs s'effacent aux yeux de l'utilisateur pour des technologies plus transparentes...

Références bibliographiques

- [1] D. Laster, « Splendeurs et misères du Théâtrophone », *Romantisme*, année 1983, vol. 13, no 41.
- [2] V. Hugo, « Choses vues », in *Oeuvres complètes*, édition du Club Français du livre, tome XVI, 1970, p. 911
- [3] H. Fletcher et al, "Binaural telephone system", Brevet US1624486, 1927.
- [4] M. Miyoshi & N. Koizumi, "NNT's research on acoustics for future telecommunication services", *Applied Acoustics*, 36. pp. 171-218.
- [5] Y. Mahieux et al, « Le traitement du son pour la téléconférence », *L'Echo des Recherches*, 1998, pp. 65-76.
- [6] R. Nicol, "Restitution sonore spatialisée sur une zone étendue: Application à la téléprésence", Thèse, Université du Maine, Le Mans, 1999.
- [7] A. Nagle, « Enrichissement de la conférence audio en voix sur IP au travers de l'amélioration de la qualité et de la spatialisation sonore », Thèse, Telecom Paris TECH, Paris, 2008.
- [8] Recommandation ITU-T P.800 : « Méthodes d'évaluation subjective de la qualité de transmission », International Telecommunications Union, Radio-communication Assembly 1996.
- [9] Recommandation ITU BS 1116-1: «Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems», International Telecommunications Union, Radio-communication Assembly 1997.
- [10] Recommandation ITU BS 1534-1: «Method for the subjective assessment of intermediate quality level of coding systems», International Telecommunications Union, Radio-communication Assembly 2003.
- [11] Recommandation ITU-T P.76 : « Détermination des équivalents pour la sonie ; Principes fondamentaux » «International Telecommunications Union, Radio-communication Assembly 1997.
- [12] S. Le Bagousse, C. Colomes, M. Paquier, and S. Moulin, "Sound quality evaluation based on attributes – application to binaural contents", Presented at the 131th AES Convention, New-York, NY, U.S.A., 2011.
- [13] J. Berg, and F. Rumsey, "Systematic evaluation of perceived spatial quality", *Proceedings of the 24th AES International Conference on Multichannel Audio*, 2003.
- [14] J. Berg, "Evaluation of perceived spatial audio quality". *Proceedings of the 9th World Multi-Conference on Systemics, Cybernetics and Informatics*, 4(2), pp. 10-14, 2005.
- [15] S. Bech and N. Zacharov, "Perceptual Audio Evaluation–Theory, Method and Application", John Wiley & Sons, 2006.
- [16] A. Daniel, "Spatial Auditory Blurring and Applications to Multichannel Audio Coding", Thèse, Univ. Pierre et Marie Curie, 2011.
- [17] J. Palacino, R. Nicol, M. Emerit & L. Gros, "Perceptual assessment of binaural decoding of first-order Ambisonics", *Acoustics 2012, Nantes*, 2012.
- [18] C. Renard, « Analyse objective et subjective d'une technique de rendu sonore 2D sur une zone d'écoute étendue, l'hologophonie, en vue de réaliser un mur de téléprésence », Université du Maine, 2000.
- [19] G. Kearney, M. Gorzel, F. Boland & H. Rice, "Depth Perception in First and Higher Order Ambisonic Soundfields", *Act Acustica United with Acustica, Special Issue on Ambisonics and Spherical Acoustics, Volume 98, Number 1*, pp. 61-71, January/February 2012.
- [20] M. Rébillat, « Vibrations de plaques multi-excitateurs de grandes dimensions pour la création d'environnements virtuels audio-visuels. Approches acoustique, mécanique et perceptive », Thèse, Ecole Polytechnique, 2012.
- [21] K.V. Nguyen, « Technologie binaurale et contexte de réalité virtuelle :Etudes perceptives et optimisation », Thèse, Univ. Pierre et Marie Curie, 2012.
- [22] N. Cote, V. Koehl & M. Paquier, « Ventriloquism effect on distance auditory cues », *Acoustics 2012, Nantes*, 2012.
- [23] S. Moulin, R. Nicol & L. Gros, "Spatial audio quality in regard to 3D video", *Acoustics 2012, Nantes*, 2012.
- [24] www.audyssey.com
- [25] www.trinnov.com/technologies/loudspeaker-room-optimization/