

Entretien avec Pascal Rueff, preneur de son 3D Le Binaural : un artisanat de la sensation ?



L'intérêt actuel des acteurs audiovisuels pour les médias tridimensionnels (image et son 3D) résulte au moins d'une ambition simple : immerger davantage le public dans l'univers de narration. Notre espace naturel de perception est en effet déployé dans quatre dimensions objectives (trois pour la géométrie, une pour le temps) et les solutions technologiques évoluent vers une prise en compte toujours plus complète de la bulle sensitive du spectateur. Les approches dites « 3D » ajoutent une dimension spatiale aux standards : l'image acquiert une profondeur et la diffusion sonore enveloppe l'auditeur comme dans la vie réelle. Cet entretien est l'occasion de faire le point sur l'expérience de preneur de son et de réalisateur de Pascal Rueff qui a peu à peu été séduit par un son 3D destiné au casque, le son binaural.

Acoustique & Techniques : Bonjour Pascal Rueff. Pouvez-vous nous décrire ce qu'est une prise de son binaurale ?

Pascal Rueff : Parmi les différents moyens de rendre compte d'une réalité sonore (et d'en capter quelque chose) la prise de son binaurale est une approche « biomimétique » dirait-on aujourd'hui.

Sur le plan technique, elle se range dans la famille des couples stéréophoniques. Dans cette catégorie, deux microphones travaillent en tandem pour trianguler les sources (les configurations varient) et reproduire un espace à l'intérieur duquel l'auditeur peut localiser les sons (il peut les pointer du doigt). Les signaux des deux micros sont associés par des relations de phase, de temps et d'intensité, qui doivent être préservées tout au long de la chaîne de réalisation. Ils sont corrélés.

Par rapport à une prise de son à un seul micro (monophonique), la prise de son stéréophonique est davantage capable de reproduire l'espace naturel d'écoute, en cela qu'elle peut commencer à rendre compte des relations spatiales (distances, ampleur) entre les sources sonores et leur environnement et, bien entendu, entre les sources elles-mêmes. Mais l'écoute sur deux haut-parleurs d'un enregistrement réalisé au couple reste cantonnée à la portion de l'espace déployée entre les deux haut-parleurs. La scène sonore n'englobe pas l'auditeur, l'expérience est conçue pour être frontale.

Des systèmes corrélés à plusieurs microphones sont aujourd'hui testés et exploités. On parle d'arbres microphoniques. Leur pendant géométrique chez l'auditeur est une configuration de haut-parleurs plus ou moins équivalente. Ainsi, tel arbre à cinq branches/cinq micros alimente les six haut-parleurs d'un système 5.1. De manière intuitive, on imagine bien qu'un nombre élevé de micros doit mieux décrire l'espace.

Un système *High Order Ambisonic* (HOA), par exemple, peut mettre en œuvre trente-deux capsules. En réalité, le succès des différentes approches qu'explore l'ingénieur dépend de la maîtrise des paramètres certes physiques, mais aussi, comme toujours, technologiques, commerciaux, voire conceptuels. La notion de fidélité par exemple, presque un mot d'ordre dans le monde de l'audio, reste un leitmotiv flottant. C'est l'exemple du déploiement des compressions psychoacoustiques - MP3 en tête - pour se conformer aux limitations de débits d'une décennie ; des algorithmes enlèvent au signal le moins utile, en visant de conserver une certaine intégrité subjective, une fidélité donc.

La mise en œuvre des systèmes, leur robustesse, leurs coûts, leur « souplesse » en postproduction, l'habitude qu'en ont les professionnels, et bien entendu l'équipement des auditeurs déterminent aussi sûrement des préférences, des modes, des esthétiques de la fidélité.

Au final, il faut bien reconnaître que le rendu de n'importe quel système de prise de son passe toujours seulement par deux oreilles !

A&T : Comment êtes vous arrivé à la prise de son binaurale ?

PR : Comme preneur de son, j'ai migré de la captation stéréophonique à la prise de son 3D par la voie la plus simple. Je suis parti du bon principe que puisque nos deux oreilles suffisent tous les jours à localiser en 3D, deux micros doivent suffire à produire le même résultat. La prise de son binaurale n'en requiert pas plus, à condition que les deux capteurs soient séparés par une tête ; notre propre configuration naturelle d'écoute (une oreille, une tête, une oreille) définit directement le système de captation équivalent ; ces mécanismes avaient été compris par l'anglais Fletcher dès 1930.

Vous savez, je suis autodidacte et je n'avais jamais entendu parler du binaural auparavant.

C'est sa simplicité qui m'a enthousiasmé.

Pour résumer froidement, c'était imparable : la fonction binaurale met en œuvre une tête et deux micros, le flux d'information est bicanal¹, et l'immersion 3D, délivrée sous casque, est complète.

Toutefois, passés les premiers émois de cette découverte, je me suis demandé pourquoi un système d'une telle évidence fonctionnelle et d'une telle efficacité n'entraîne pas dans ma boîte à outils. Soixante-dix après sa description, le procédé restait essentiellement connu des laboratoires (tester un téléphone) et des concepteurs automobiles ; quelques producteurs de musique² s'y étaient intéressés, guère plus.

Quelque chose clochait. Plusieurs choses clochaient.

1- Le terme stéréo convient (deux canaux corrélés), mais j'opte ici pour bicanal qui fait le distinguo avec les techniques stéréophoniques traditionnelles
2- Tchad Blake, notamment

A&T : Qu'entendez-vous par là ?

PR : En pratique, le signal binaural doit être «injecté» sans mélange, signal gauche dans l'oreille gauche, signal droit dans l'oreille droite ; ce qui rend l'écoute au casque obligatoire. Souvenez-vous que si l'usage d'écouteurs nous est aujourd'hui familier, la chose n'allait pas de soi voilà seulement quinze ans. L'expérience individuelle de la musique n'était pas valorisée à ce point ; c'est la technologie qui a rendu possible de promener sa discothèque.

Les mœurs ayant bougé, une écoute exclusive au casque ne paraît plus aujourd'hui comme un obstacle.

Une autre évidence suscitait des questions : l'objet sonore, outre ses qualités habituelles de timbre, de hauteur tonale, d'enveloppe, de volume, etc. s'est vu doté d'un nouvel attribut phénoménal : l'espace entier.

Pour le preneur de son que je suis, l'objet sonore binaural apparaît alors comme une globalité, le «signal utile» est inséparable de l'espace dans lequel il se développe. La prise de son stéréophonique y préparait, mais l'ouverture azimutale devient ici complète, il n'y a plus de cadre, tout l'audible est entendu (le haut et le bas apparaissent).

Cela confère au procédé une liberté maximale pour placer les sources dans l'espace.

Seulement, il faut savoir que puisqu'il s'agit de prise de son, toutes les décisions de positions et de déplacements doivent se prendre au moment de la capture (du tournage) et la voiture de passage à un kilomètre n'a peut-être pas sa place dans le scénario.

La source sonore est indémontable de sa globalité spatiale, sa position ne se manipulera pas en postproduction, de nombreuses sources secondaires, parasites ne pourront être gommées. Il faut donc penser mixage à la prise, et en 3D, et travailler dans des conditions acoustiques compatibles avec les besoins de la narration.

Par ailleurs, les premiers essais montraient que les têtes artificielles n'étaient pas très neutres sur le plan des timbres, avec un côté nasal haut perché qui semblait dénoncer un système de piètre qualité, selon nos critères de preneurs de son.

Ces considérations laissaient voir qu'il faudrait expérimenter beaucoup. Mais le potentiel d'amélioration n'était pas mince, tant en traitement du signal qu'à l'usage. Un savoir-faire, une nouvelle forme d'écriture étaient à inventer.

A&T : Il semblerait qu'un problème plus sérieux empêcherait le développement de la prise de son binaurale. Pouvez-vous nous en dire plus ?

PR : Il s'agit de la morphologie du mannequin de prise de son qui ne peut exactement coïncider avec celle de chaque auditeur. Ce problème est malheureusement incontournable à l'heure actuelle.

Si la fonction binaurale émane de la forme générale typique de la tête d'un être humain, la taille de son nez, la largeur de son visage, sa carnation, l'orientation de ses oreilles ainsi que la forme des pavillons l'individualisent absolument, la rendent unique. Les paramètres de notre morphologie singularisent notre grille de lecture de la distribution du son dans l'espace.

Vous savez que l'onde sonore voyage suffisamment vite (340 mètres/seconde) pour que la différence maximale de temps d'arrivée entre chaque oreille ne dépasse guère une milliseconde (fig.1). Cela donne une idée de la rapidité de

discrimination de notre système d'analyse psychoacoustique. Cette interaction physique de la tête et du flux sonore, précisément déconvoluée par le cerveau nous permet de détailler la provenance des sources sonores.

Si je simplifie : vous savez qu'une porte claque dans votre dos parce que les petits indices inconscients associés à la position arrière ont été identifiés. Votre cerveau a comparé les signaux de chaque oreille et déterminé que leur différentiel signalait une position arrière (ce que vos yeux ont confirmé). La base de données personnelle des différences interaurales que vous avez constituée est modélisable et on la nomme HRTF (*Head Related Transfer Function*).

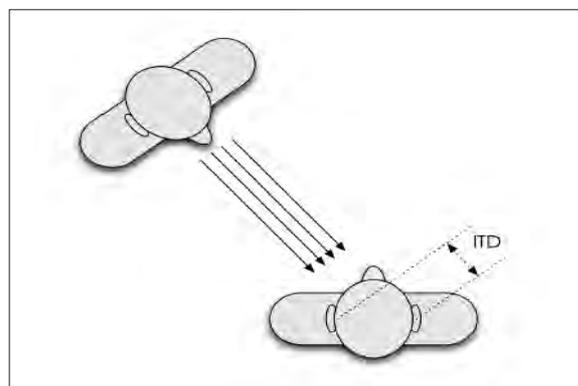


Fig 1 : Différence interaurale de temps (ITD)

Or, ce mécanisme subtil est perturbé par le signal d'une tête artificielle : les approximations consécutives à l'emploi du mannequin, forcément générique, créent des flous dans le rendu spatial, des distorsions ; la bulle 3D se déploie, tient le coup, mais se déforme. Notre grille de lecture est perturbée et déclenche des projections fausses. Différemment fausses selon les auditeurs : aucun de nous ne verra la même scène (mais nous verrons tous une scène déformée). Comment produire avec une technique qui prive le réalisateur d'une bonne part de la prédiction nécessaire ? C'est toujours une question.

L'écoute humaine est individualisée et l'empreinte captée par un mannequin ne peut exactement coïncider avec l'immense diversité des têtes d'utilisateurs, si l'on peut dire.

A&T : Quelles seraient alors les solutions à apporter ?

PR : Des solutions techniques tentent de remédier à l'absence d'individualisation de la prise de son. Le flux binaural n'émane plus, dans ce cas, d'une réalité captée par une tête, mais d'un logiciel qui manipule à la demande telle ou telle HRTF et l'applique à une séquence audio numérique captée sans espace pour la positionner dans la bulle de perception de l'auditeur.

Cette approche par synthèse binaurale fournit déjà des solutions professionnelles (pour l'écoute d'un mixage 5.1 au casque par exemple), mais ne se développera réellement auprès du public qu'à partir du moment où il sera possible de mesurer (ou d'approximer) l'HRTF de l'utilisateur et d'en renseigner le moteur de synthèse. Les questions du temps de calcul ou de la nécessaire standardisation des équipements ne résisteraient pas, probablement, à un sérieux engouement du public.

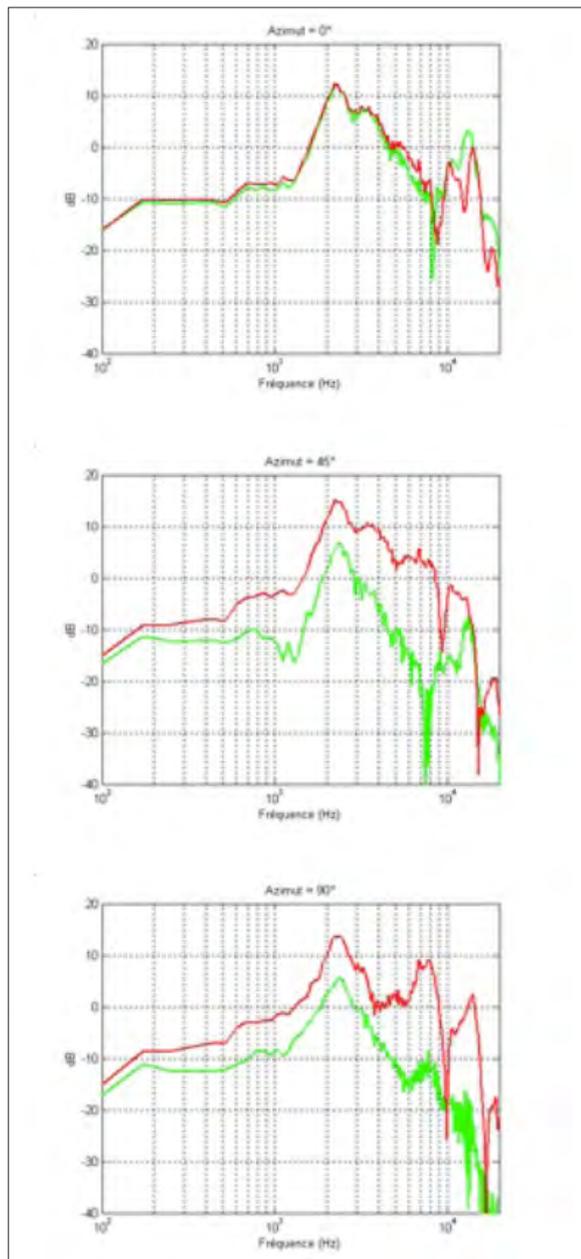


Fig. 2 : Différences interaurales de réponse en fréquence pour les azimuts 0°, 45° et 90°, oreille gauche en vert, oreille droite en rouge, mesurées sur le mannequin Kemar. Mesurées avec une autre morphologie, ces réponses varieraient. ill. Rozenn Nicol

Vous savez, j'ai commencé il y a dix ans, grâce à la conviction de Rozenn Nicol, scientifique rattachée au service R&D d'Orange, dans le cadre universitaire où nous enseignions, à Brest. Son cours de psychoacoustique expliquait les mécanismes de la perception spatiale et permettait d'écouter de courtes séquences réalisées avec des têtes artificielles. Des étudiants m'en ont parlé, nous avons fait quelques tests. Je me souviens très bien du choc de cette première expérience. Nous avons capté les préparatifs d'un concert et à la réécoute, tard dans la nuit, une fois rentré chez moi, je voyais que mon esprit tentait d'assembler des signaux visuels et sonores pourtant manifestement décorrélés : l'ambiance du hall de concert et les formes familières de mon bureau. Je me voyais composer une réalité hybride, instable, à partir de deux réalités différées. J'étais souffié que mon esprit fasse cet effort : manifestement, un certain jeu d'automatismes ne se rendait pas compte qu'il était renseigné par un casque. La distanciation ne marchait pas.

Ce point de bascule, ce point critique ne m'a jamais quitté. Produire avec un tel procédé semblait capable de susciter chez l'auditeur des modalités cognitives assez nouvelles.

Après dix ans de pratique, la prise de son binaurale et l'usage du casque me sont devenus à ce point familiers que la stéréophonie conventionnelle a perdu son caractère de standard ; j'aurais probablement du mal aujourd'hui à mixer quelque chose avec une paire d'enceintes dans l'espace réel d'un studio. En contrepartie, j'ai gagné de pouvoir écrire avec des espaces complexes et des stimuli très efficaces.

A&T : Pouvez-vous nous parler maintenant de vos expérimentations ou mieux de vos productions en binaural ?

PR : Tout d'abord, à partir de 2002, une phase d'expérimentation et de validation technique, menée avec le soutien de France Télécom R&D par le collectif OmniHead (association de preneurs de son créée pour l'occasion) avait montré qu'il était possible d'exploiter la prise de son binaurale. Mettre en œuvre un mannequin ne posait pas de problème inédit (sinon qu'il pouvait faire quelque peu sensation, installé au milieu du public!). Au prix d'une certaine approximation dans la localisation et d'un traitement de la réponse en fréquence, l'externalisation des sources sonores (le fait de les croire en dehors du casque) s'avérait robuste et facilement fonctionnelle.

Mais apprendre à travailler avec cet outil impliquait d'en observer les effets sur le public ; il fallait engager une véritable production et proposer dans un cadre réel un objet culturel abouti.

C'est pourquoi, après une première tentative autour d'un travail de collectage à Douarnenez en Bretagne, et une expérience de diffusion collective assez réjouissante au Quartz de Brest (22), l'opportunité vint en 2007 avec une commande du festival «Nuit Bleue» pour le très beau site de la Saline Royale d'Arc et Senans (25). Une pièce de vingt minutes, construite autour du voyage et de l'ambiance des trains, pouvait s'écouter en déambulant dans les jardins du site, de nuit, avec un casque sur la tête et une lampe. Par rotation de la vingtaine de casques disponibles, l'opération permettait de toucher quelques centaines de personnes.

A&T : Et votre expérience dans le domaine ?

PR : Pour ce qui me concerne, j'ai parié sur des solutions d'usages. Ignorer le problème de l'individualisation permettait, en attendant mieux, de se focaliser sur les atouts. En effet, si je veux frapper l'attention de l'auditeur avec un claquement de porte, de quelle précision de localisation ai-je réellement besoin ? Ne me suffit-il pas que la prise de son donne l'impression qu'une porte physique vient de claquer, qu'elle est à droite, à une distance de trois ou quatre mètres ? En tant que réalisateur, une rigoureuse reconstruction de l'espace m'importe finalement moins que de toucher l'auditeur pour lui raconter l'histoire.



Photo 1 : Tête de prise de son mise en œuvre sur le plateau du spectacle « Spatih 1.0 ». Crédit photo : Pascal Rueff

Le succès de ce lancement encouragea Elektrophonie, structure organisatrice du festival, à réitérer l'expérience. Ce laboratoire grandeur réelle permet de faire écouter au grand public des œuvres réalisées en binaural, mais aussi de réfléchir aux modalités d'exploitation : choisir les casques, élaborer le discours de vulgarisation, jouer avec la réalité physique en corrélant l'expérience auditive et l'environnement de la déambulation, etc.

En parallèle, Elektrophonie déclenchait une activité de R&D : associer la géolocalisation au scénario. Un acteur culturel de diffusion travaillait au développement de nouveaux moyens de narrations. Le dispositif informatique « Chimera », développé par Christophe Baratay, intègre le suivi GPS de la position de l'utilisateur et renseigne un suiveur de scénario : en fonction du parcours réel, la narration s'adapte en temps et en contenu³. Une couche de fiction se superpose à la visite du site sans le modifier. En 2009, la Saline Royale d'Arc et Senans commandait un parcours pour trois ans en complément de son offre d'accompagnement standard (guides et audioguides à touches).

Aujourd'hui des sites historiques ou des festivals sont équipés de casques et proposent à leurs publics des narrations réalisées en binaural, à partir de résidences de tournage durant lesquels la matière sonore est produite et captée.

Mais d'autres applications sont envisageables.

Depuis sept ans, nous travaillons du côté de Tchernobyl, en Ukraine ; nos séjours construisent des témoignages, nous employons différents médias. En 2008, nous avons pu tester l'emploi du binaural dans une fiction destinée à la scène, *L'île de T.* où chaque spectateur est muni d'un casque. Le public est distribué autour d'un dispositif de prise de son binaural, actif, par lequel les acteurs chuchotent à l'oreille de chacun. Des prises de son réalisées dans les grandes forêts de Tchernobyl transportent le public par le son et lui font éprouver, par la sensation, certaines des émotions que ces territoires hors normes peuvent provoquer.

L'immersion est ici renforcée par la présence des acteurs, bien réels, et l'impact troublant des voix médiatisées par la tête artificielle en direct.

S'il reste à explorer dans les années à venir une écriture adaptée à cette approche, par le biais probablement d'un laboratoire de production, force est de constater que le public est favorable à cette approche et parle volontiers d'une expérience troublante, réellement nouvelle.



Photo 2 : Un mannequin de prise de son avec protection anti-vent à poste dans la forêt contaminée de Rudnia en Ukraine. Crédit photo : Pascal Rueff

A&T : En conclusion, quels seraient vos arguments pour rallier nos lecteurs à l'utilisation de la prise de son binaurale ?

PR : En tant que preneur de son, ce système me satisfait ; il est au plus près de ma façon d'écouter le monde. En tant que réalisateur, il me donne le moyen de composer avec l'espace. En tant qu'auditeur, j'apprécie que les événements sonores aient leurs dimensions « naturelles », puissance comprise, et d'ailleurs j'apprécie la grande dynamique que l'usage du casque autorise (en m'isolant dans une bulle de moindre bruit). J'apprécie également que la diffusion ne soit plus sujette à l'acoustique du lieu d'écoute et que soit réglée la question du point d'écoute idéal (le « spot »), les transducteurs gauche et droit du casque étant idéalement placés, sans me coincer quelque part. Cette économie de moyens me séduit.

Dix ans plus tard, je constate qu'avec un flux stéréo et un casque à cinquante euros il est possible de proposer au public une expérience sensible de très grande qualité.

A&T : Pascal RUEFF, nous vous remercions.

Pour contacter Pascal Rueff Tél : 06 70 00 46 48
E-mail : pascal.rueff@orange.fr <http://www.binaural.fr/>
www.tchernobyl.fr