

Ecoute au casque multicanal (1^{ère} partie) Facteurs perceptifs et principe de fonctionnement

L'écoute au casque touche aujourd'hui beaucoup de monde et a un bel avenir devant elle, face aux besoins et aux modes de vie de plus en plus mobiles, à travers l'ordinateur portable, le smartphone, l'ipod, et la tablette numérique, et leurs nombreuses applications.

Nous avons, depuis l'arrivée du casque stéréo dans les années 1950, développé une culture d'écoute définie par la description d'une image sonore interne, c'est à dire que nous localisons les sons dans notre tête. Ainsi, sans procédé de virtualisation, un son monophonique écouté au casque positionne la source au centre de la tête, tandis qu'un signal stéréo positionne les sons sur une ligne de gauche à droite, en passant par le centre de la tête, mais jamais en dehors, comme dans la perception naturelle. Aujourd'hui, les nouvelles technologies multicanal s'intègrent progressivement dans les casques et les lecteurs audio mobiles, et nous proposent de sortir de la perception interne et de s'initier sérieusement à la spatialisation du son, comme nous le percevons naturellement. Mais comment spatialise-t-on un son pour le casque stéréo ? Définissons tout d'abord les facteurs perceptifs sur lesquels le principe du casque multicanal repose.

Localisation

La localisation du son au casque stéréo repose sur les mêmes indices que pour l'écoute sur une installation d'enceintes stéréophonique : les paramètres ITD (différences interaurales de temps) et ILD (différences interaurales de niveau). On définit l'oreille ipsilatérale comme étant l'oreille éclairée par l'onde sonore, et l'oreille contralatérale, celle dans l'ombre. L'ITD traduit, par une valeur de temps, la différence de marche que subit l'onde sonore pour

atteindre l'oreille opposée, contralatérale et l'ILD traduit par une différence de niveau, la diffraction et l'absorption que subit cette même onde.

L'ITD peut être ou non significatif par rapport aux longueurs d'ondes des signaux sonores. Cela se traduit par deux types de valeurs :

- les différences temporelles de la structure fine, dites aussi IPD (différences de phase), valables pour les fréquences basses, inférieures à 1500 Hz
- les différences de temps d'enveloppe, valables pour les fréquences aiguës, à partir de 200 Hz.

Pour une écoute au casque stéréo traditionnel, seuls les paramètres ITD et ILD contrôlent la localisation latérale des sources, entre les oreilles.

Tout d'abord, les écoutes stéréo au casque et sur enceintes ne sont pas tout à fait compatibles. En effet, la largeur de l'image sonore sera toujours un peu plus importante au casque que pour l'écoute sur enceintes. La raison, ce sont les valeurs ITD et ILD, qui pour le casque sont plus petites (0,7 ms / 10 dB environ pour latéraliser complètement une source sur une oreille, contre 1,1 ms / 15 dB environ sur une enceinte).

Par ailleurs, la localisation de l'image sonore pour un casque stéréo est interne alors que la perception pour une écoute sur enceintes est externe. Comment externaliser alors l'image sonore dans un casque stéréo ? En écoute naturelle spatialisée, nous utilisons les indices ITD et ILD pour localiser les sources, mais le filtrage de notre tête, de nos épaules, des pavillons de l'oreille joue aussi un rôle très important. Nous parlons alors d'indices spectraux ou courbes de transfert relatives à la tête HRTF (Head Related Transfer Function). On définit plus précisément par indices spectraux, les courbes de transfert HRTF qui présentent une certaine saillance pouvant être détectée, analysée et interprétée par

l'oreille afin de localiser le son.

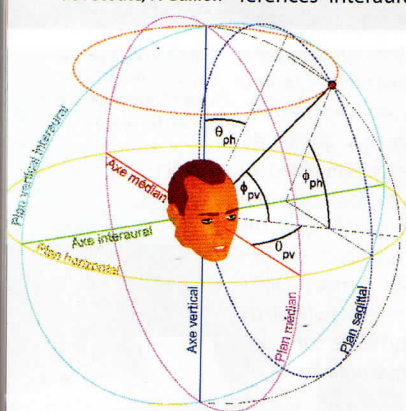
Une étude particulièrement intéressante dans ce domaine a été menée récemment par Rozenne Nicol, au sein d'Orange Labs. Les courbes HRTF sont la conséquence de la diffraction due à la tête, des réflexions sur les épaules, sur le torse, et des réflexions et des diffractions issues des pavillons de l'oreille. Elles se traduisent par des courbes de réponses en fréquence spécifiques en fonction de l'angle de provenance de la source.

Ces phénomènes sont considérés comme l'empreinte acoustique de l'auditeur et celle-ci diffère d'un individu à l'autre. Le pavillon de l'oreille fonctionne comme une cavité résonnante, il ajoute des colorations spectrales, variables en fonction de la direction de la source. Il semble être, d'après R. Nicol, l'élément clef dans la caractérisation de la fonction HRTF. Chaque direction du son a une courbe de transfert spécifique et l'oreille est capable d'extraire cette caractéristique fréquentielle pour retrouver la provenance angulaire de la source. La fonction de transfert HRTF définit la morphologie de la tête, et elle est représentée par une courbe de réponse en fréquence (HRTF), ou par une réponse impulsionnelle (Head related Impulse Response HRIR), son équivalent dans le domaine temporel. Elle dépend de la fréquence ou du temps, de la localisation de la source sonore et de la morphologie de l'individu.

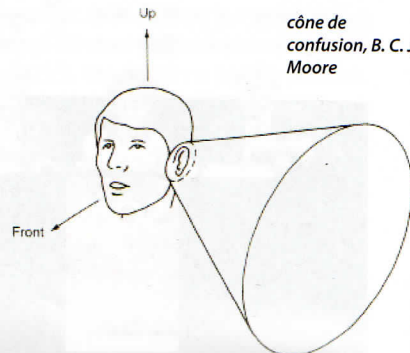
L'ITD et l'ILD sont des critères à valeur constante en fonction de la position de la tête, ils permettent de localiser le son dans ce que l'on appelle « le cône de confusion », dont l'axe de symétrie est l'axe interaural. En effet, pour un même angle d'incidence, la différence de marche est exactement la même si le son vient de l'avant ou de l'arrière. Les indices spectraux vont alors lever cette indétermination frontal / arrière et permettre aussi la localisation en hauteur.

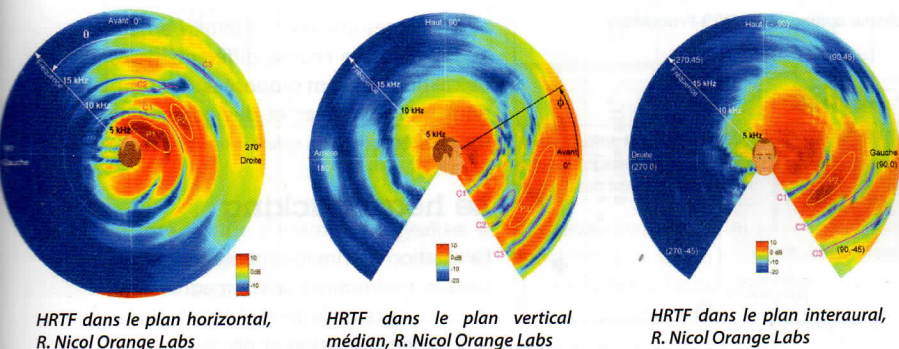
Les indices spectraux, qui contribuent le plus à la localisation, sont ceux issus des effets des cavités des pavillons de l'oreille.

Représentation des différents plans et axes de l'écoute, P. Guillon



cône de confusion, B. C. J. Moore





Ceux, issus de la tête et des épaules, sont d'un intérêt moindre et correspondent à des fréquences en dessous de 4 kHz. Les fréquences qui jouent un rôle important dans les indices spectraux des pavillons sont contenues dans la bande 4 kHz -16 kHz. Les résonances induites par la forme du pavillon, qui est propre à chacun, n'affectent pas les basses fréquences des sons en dessous de 4 kHz, puisque sa taille est trop petite en comparaison avec la forme d'onde. En revanche, pour les hautes fréquences, l'effet du pavillon est décrit par des pics et des trous dont les fréquences dépendent de la direction du son.

Voici une représentation des courbes HRTF suivant trois plans, avec la présence des pics P et des creux C :

Pour localiser le son, le cerveau utilise simultanément deux méthodes. La première consiste à utiliser l'indice spectral porté par le spectre global de la source, en comparant le spectre perçu et les spectres des HRTF enregistrés dans la mémoire depuis le plus jeune âge. Ainsi, les mouvements de la tête, faisant appel à différentes courbes HRTF, jouent un rôle important pour préciser la localisation, on parle d'indices spectraux dynamiques. La seconde méthode fait référence aux irrégularités créées par les pics et les creux dans le spectre perçu, représentés sur les figures des HRTF.

Dans le traitement des indices spectraux, l'oreille utilise concomitamment là aussi deux méthodes : elle traite de façon commune les indices recueillis par l'oreille gauche et droite, cela s'appelle le traitement binaural, et elle les traite aussi de façon indépendante, ce qui correspond à exploiter les HRTF monaurales. L'ensemble des courbes HRTF, dont le principe de fonctionnement vient d'être décrit, sont totalement absentes pour une écoute au casque stéréo traditionnel. Le casque multicanal intègre donc ces courbes, et tient compte aussi des mouvements de la tête de l'auditeur, en faisant varier les HRTF afin de reproduire les indices spectraux dynamiques.

Perception de la distance

La perception de la distance fonctionne à partir de plusieurs indices, notamment pour les sons familiers, le niveau perçu est un critère important. Cet indice s'avère être encore plus efficace lorsque plusieurs sources sonores sont présentes, de telle sorte que l'oreille peut comparer les différents niveaux perçus. C'est un indice efficace dans le jugement des distances absolues.

L'indice spectral est aussi utilisé pour des distances modérées, mettant en œuvre les propriétés d'absorption de l'air. Les hautes fréquences sont plus rapidement atténuées que les basses fréquences, en fonction aussi de la présence de vent et du taux d'humidité de l'air. Cet indice est dépendant de la familiarité avec les sources, il est efficace dans le jugement des distances relatives mais pas dans l'évaluation des distances absolues.

Pour les sons proches de la tête, ce sont les valeurs d'ILD plus importantes, qui permettent d'évaluer leur distance.

Dans une salle, l'évaluation de la distance en absolu et en relatif se fait grâce au rapport champ direct sur champ réverbéré CD/CR et au temps initial, délai entre le son direct et les premières réflexions. Le spectre du son réfléchi, altéré par les premières réflexions, sert aussi d'indice de jugement de la distance.

Pour conclure, les informations obtenues des différents indices sont combinées pour évaluer la distance, les deux principaux critères restant le niveau et le rapport CD/CR. Ces deux indices donnent des résultats dépendants du type de source, et de sa position angulaire (azimuth).

Impression spatiale

Abordons maintenant la notion d'impression spatiale qui caractérise fortement l'image sonore multicanal.

L'étude de l'acoustique des salles nous apporte des informations pertinentes

pour analyser l'espace que l'on ressent à l'écoute d'une œuvre en multicanal. Voici la définition de l'impression spatiale que J. Blauert donne : c'est l'image conceptuelle du type, de la taille et des propriétés d'une salle réelle ou virtuelle. C'est la sensation d'espace, l'impression d'être entouré par les sons.

Cette impression repose sur différents attributs perceptifs (cf J.M. Potter, M. Deschamps) :

- la taille apparente de la salle
- la spatialisation ou « Spaciousness », qui comporte :

- la largeur apparente des sources

- l'enveloppement de l'auditeur

- la réverbérance : rémanence temporelle des événements sonores résultant des réflexions tardives et de la réverbération

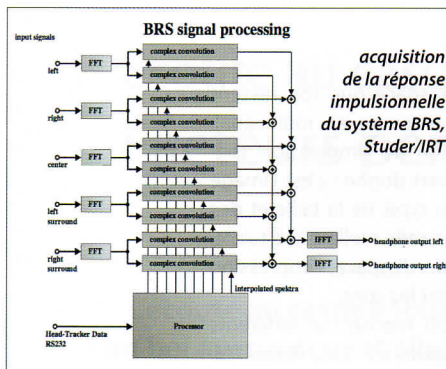
Le terme spaciousness pourrait être traduit en français par spatialisation mais la signification n'est pas exactement la même. En effet, la notion de spatialisation induit une localisation des sources dans une salle, plutôt qu'une quantité d'espace occupée par celles-ci. L'enveloppement est un critère que les acousticiens définissent comme reposant essentiellement sur la décorrélation interaurale. Le contraire d'une image sonore multicanal présentant une grande impression spatiale est une image sonore fermée, très corrélée, pouvant être perçue de façon interne, proche de la tête.

En conclusion, on demande au casque multicanal un bon rendu en localisation avec une perception externe des sources, la possibilité de choisir la distance de notre écoute multicanal virtuelle et les caractéristiques de l'acoustique du lieu ainsi qu'une bonne restitution de l'impression d'espace.

Voyons maintenant le principe de fonctionnement du casque multicanal.

Prise de son binaurale ou synthèse binaurale ?

Il existe aujourd'hui diverses solutions de reproduction virtuelle de l'espace pour le casque. La première consiste à reproduire au casque stéréo une prise de son binaurale, faite à partir d'une tête artificielle ou personnalisée, en intégrant des microphones miniatures dans les conduits auditifs de nos oreilles. Les résultats sont très réalistes mais cette étude ne fait pas l'objet de ce premier article. L'autre solution consiste à synthétiser un format binaural à partir d'un mixage multicanal 5.1 ou 7.1.

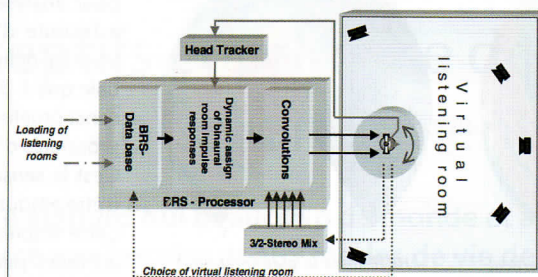


Réponse impulsionnelle binaurale de la salle

La synthèse binaurale reproduit virtuellement une écoute multicanal et fait appel à l'information sur les enceintes de diffusion, l'environnement et l'auditeur. Tout signal audio interagit avec ces trois facteurs et est au final écouté, il peut être étudié grâce à la réponse impulsionnelle de chacun d'entre eux. Lorsqu'un signal impulsionnel est généré par des enceintes, il est réfléchi, diffracté et absorbé par toutes les surfaces de la pièce et crée ainsi un champ réverbéré. Ce champ sonore définit et associe les réponses impulsionnelles des enceintes, des amplificateurs et de la pièce. On le définit par la courbe de réponse impulsionnelle de la pièce (Room Impulse Response RIR). De façon similaire, lorsque le son parvient au niveau du conduit auditif, il est filtré comme défini précédemment par nos pavillons, nos épaules, notre tête, suivant les courbes HRIR ou HRTF. Ces deux étapes peuvent être regroupées afin d'étudier le signal émis par un haut parleur et reçu au niveau du conduit auditif, elles sont définies alors par la réponse impulsionnelle binaurale de la pièce (Binaural Room Impulse Response BRIR). Durant la reproduction du son au casque, la réponse impulsionnelle binaurale de la pièce, pour chaque source associée à une enceinte, est convoluée avec le signal audio entrant. Les différents signaux convolués pour

Binaural Room Scanning (BRS)

Functional outline of the BRS-Processors



lecture par convolution de la réponse impulsionnelle du système BRS, Studer/IRT

les différentes enceintes (L, R, C, Sub, Ls, Rs) sont sommés et reproduits au format stéréo pour le casque.

Voici les différents procédés de convolution des casques BRS et Realiser 8 :

Le cerveau est alors capable, à partir de l'analyse fréquentielle du contenu, d'identifier les directions sonores associées. Deux possibilités s'offrent alors : des solutions intégrant des HRTF personnalisées et d'autres, comportant des HRTF moyennées, proposées sous formes de différents presets issus de bibliothèques.

La calibration du casque

Le système intègre une procédure de calibration du casque qui consiste à compenser le filtrage HRTF du pavillon de l'oreille pendant l'écoute, ce réglage dépend du type de casque utilisé. En effet, le filtrage est plus important pour les casques dont les oreillettes recouvrent l'oreille que pour les casques intra-auriculaires. La réponse impulsionnelle du casque / pavillon de l'oreille est alors mesurée, et un filtrage de casque individuel est généré et enregistré. La calibration prend en compte le couplage acoustique casque / oreille ainsi que la courbe de réponse en fréquence des haut-parleurs du casque. Sans cette procédure de calibration, la localisation en hauteur est

altérée et les inversions frontal / arrière augmentent. La encore, différents types de calibrations sont proposées, des solutions individualisées et des corrections moyennées.

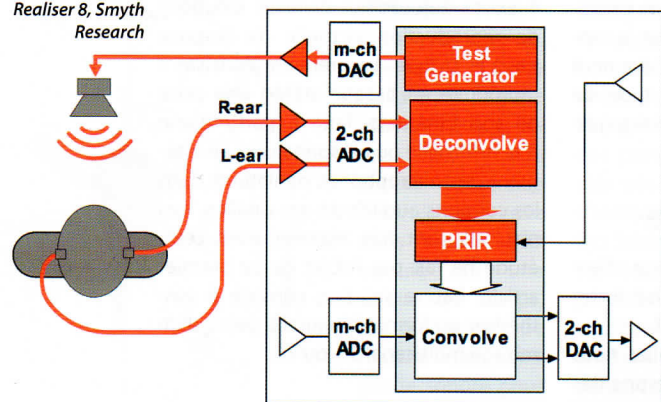
Le head tracking

La rotation de l'image sonore perçue au casque traditionnel, accompagnant les mouvements de la tête, réduit l'impression d'externalisation et de réalisme du champ sonore. Une étude menée au sein de l'IRT par G. Theile démontre la nécessité d'intégrer un head tracking pour l'écoute binaurale multicanal, car, ce dernier solutionne les nombreuses erreurs d'inversions frontal / arrière, ou de localisations internes.

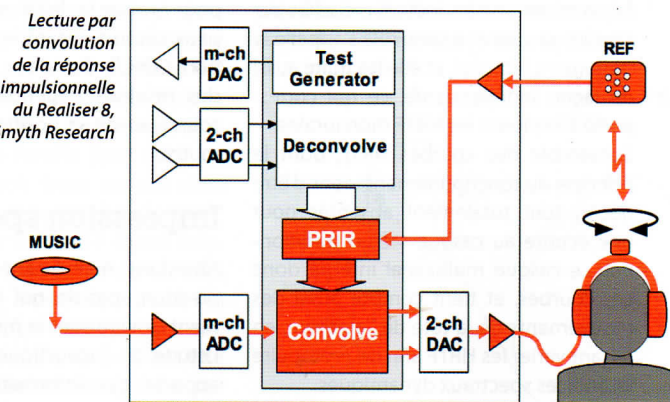
Le head tracking permet, grâce à l'utilisation d'un récepteur placé sur le casque de l'auditeur, d'avoir un suivi de l'orientation horizontale de sa tête, dans ses mouvements de rotation, sur un espace de 60 à 80 degrés, de part et d'autre du centre. Le procédé, d'une part, permet de fixer l'image sonore perçue lors des mouvements de tête, et recalcule les HRTF en intégrant la valeur ITD correspondant à l'angle de rotation de la tête. Il solutionne ainsi les inversions frontal / arrière souvent présentes et améliore le réalisme et l'externalisation du champ sonore. Le head tracking suit les mouvements de tête avec un taux de rafraîchissement d'une centaine de herz et le procédé calcule ainsi les HRTF à partir de plusieurs positions, généralement tous les 15 degrés. Le procédé demande beaucoup de puissance de calcul, c'est pourquoi des modèles de simplification sont généralement utilisés comme l'interpolation des valeurs obtenues tous les 15 degrés, pour une résolution d'environ 5°.

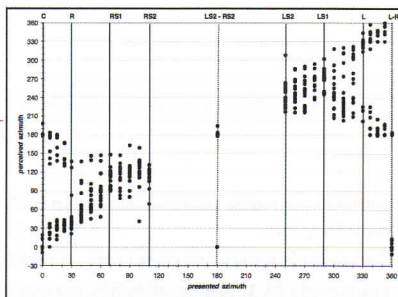
La latence, d'environ 10 ms pour les différentes solutions de head tracking, est bien sûr un critère de réactivité du casque important afin d'avoir un rendu

Acquisition de la réponse impulsionnelle individualisée du Realiser 8, Smyth Research

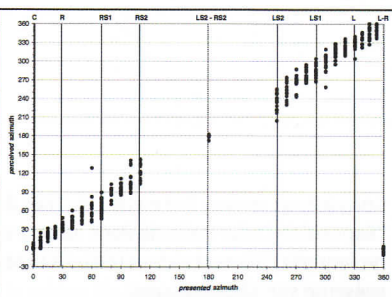


Lecture par convolution de la réponse impulsionnelle du Realiser 8, Smyth Research

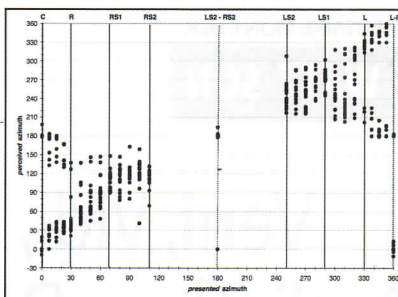




Configuration de l'écoute multicanal pour l'étude IRT, IRT



Localisation multicanal au casque avec réponses impulsionnelles binaurales d'une tête artificielle sans head tracking, IRT



Localisation multicanal au casque avec réponses impulsionnelles binaurales d'une tête artificielle utilisant un head tracking, IRT

réaliste. A titre d'exemple, le Realiser 8 permet de générer une rotation de 360° sur une seconde sans introduire d'artefact audible.

L'ensemble des données varient d'un système à l'autre, ce qui justifie des performances variables mais l'utilisation du head tracking améliore sans aucun doute le réalisme, la perception externe de l'image ainsi que la localisation circulaire des sources.

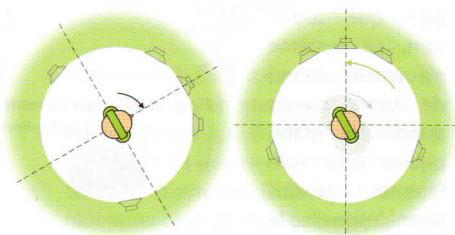


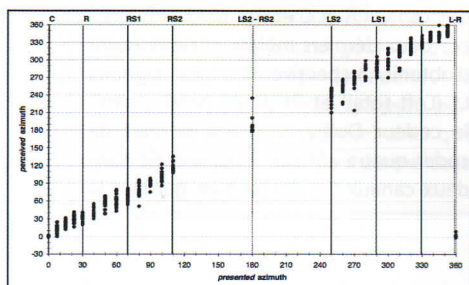
image sonore sans et avec head tracking, Beyerdynamic

Personnalisation des HRTF

La même étude menée par l'IRT compare la localisation en utilisant des réponses impulsionnelles binaurales issues d'une tête artificielle et celles obtenues pour un auditeur. Les résultats n'exposent pas de réelles améliorations pour la perception. L'étude du head tracking vertical est aussi abordée mais ce dernier n'améliore pas la perception de hauteur.

Cependant, d'autres études démontrent que l'utilisation d'HRTF non individualisées induit des artefacts tels qu'une diminution de la perception externe, avec des localisations internes, et des erreurs d'inversions frontal / arrière et haut / bas. Les indices spectraux constituent finalement le paramètre le plus affecté par la non-individualisation des HRTF dans le casque.

Pour conclure, certaines solutions exploitent l'étude IRT en proposant des réponses impulsionnelles binaurales obtenues avec une tête artificielle et d'autres appliquent des HRTF individualisées.



Localisation multicanal au casque avec réponses impulsionnelles binaurales d'un individu utilisant un head tracking, IRT

En revanche, toutes intègrent un head tracking pour suivre les mouvements de rotation horizontaux de la tête de l'auditeur, autour de son axe vertical.

La perception de salle

La réponse impulsionnelle binaurale donne l'empreinte acoustique du lieu dans lequel se fait la mesure. On peut donc imaginer plusieurs types de pièces sous forme de presets et choisir le programme qui nous convienne en fonction des caractéristiques acoustiques, de l'identité de l'écoute multicanal virtuelle. Une fois le programme chargé, ces caractéristiques ne changent plus. Il existe autrement une méthode de synthèse de réponses impulsionnelles qui permet à l'auditeur de modifier de façon continue la taille de la pièce, la distance de l'écoute virtuelle et la quantité d'effet de salle, cette méthode est par exemple utilisée par le casque Headzone.

Résultats perceptifs

Les résultats perceptifs de ces casques sont assez variables. La localisation des canaux arrière fonctionne souvent très bien. En revanche, la localisation frontale est plus limitée, avec notamment un canal central difficile à « visualiser » devant soi. Ces résultats perceptifs dépendent du contenu, surtout de la décorrélation du signal multicanal entrant et de la présence ou non de plusieurs sources en même temps. L'identification par exemple d'une voix en solo, qui passe de canal en canal, fonctionne très bien, comme l'écoute en solo des canaux d'un mixage 5.1. En revanche, le passage à un mixage complet, où l'ensemble des canaux du multicanal sont restitués, recentre généralement la perception vers l'interne, comme si l'oreille avait du mal à trier les différents filtrages HRTF associés aux canaux du multicanal. Il y a donc sans doute encore des améliorations à apporter sur l'externalisation.

Dans tous les cas, aucun de ces procédés ne peut prétendre reproduire aujourd'hui parfaitement la localisation obtenue à partir de cinq ou d'avantage d'enceintes, réparties autour de l'auditeur, mais des avancées très intéressantes se sont faites connaître dernièrement, avec notamment le Smyth

Research Realiser 8, qui offre la personnalisation des HRTF.

Les Solutions

Parmi les différentes solutions proposées, trois procédés professionnels ont vu le jour, le casque BRS développé dans les années 1990 par l'IRT et Studer, le casque Headzone développé par Beyerdynamic et Sonic Emotion en 2007 (un modèle grand public et un modèle professionnel) et le Realiser 8 développé par Smyth Research en 2009 (une version grand public et une version pro en cours de développement). Seul le projet BRS Studer n'a finalement pas été commercialisé.

Ces premières solutions se présentent sous forme de boîtiers avec des entrées audio 5.1, discrètes (analogiques ou numériques PCM) ou encodées DTS, Dolby, et une sortie casque stéréo, le procédé de spatialisation étant contrôlé par head tracking. D'autres solutions comme Dolby Headphone, Arkamys, Fraunhofer (à travers la technologie MPEG Surround et mp3D) sont des technologies de spatialisation sans head tracking. Ces dernières solutions proposent un format binaural spatialisé pour une position de l'auditeur centrée, intégrant des HRTF moyennées, souvent proposées suivant trois presets. Ces dernières solutions sont embarquées dans certains téléphones mobiles, ipods (Mpeg Surround, Arkamys), dans des casques pour le jeu vidéo et on les trouve aussi sous formes d'applications compatibles PC, Mac.

D'autres solutions très différentes, fonctionnent avec plusieurs haut-parleurs, citons par exemple le casque Tube Surround présenté à l'AES de San Francisco 2008.

Lors du prochain numéro, nous présenterons et comparerons ces différentes solutions d'écoute au casque multicanal.

Un grand merci à Rozenne Nicol pour son excellent monographe sur les technologies binaurales, édité par l'AES.



Le système BRS, Studer/IRT



Le système Headzone Pro, Beyerdynamic/Sonic Emotion



Le système Realiser 8, Smyth Research

Casque Tube Surround intégrant 6 HP

