

Julien Carton

Intégration et exploitation de traitements transauraux pour la production au format multicanal 5.1

Mémoire de fin d'études
de la Formation Supérieure aux Métiers du Son
du Conservatoire National Supérieur de Musique et de Danse de Paris

réalisé à l'Institut de Recherche et Coordination Acoustique/Musique (Ircam)
dans l'équipe Espaces Cognitifs et Acoustiques (EAC)
et au Conservatoire National Supérieur de Musique et de Danse de Paris (CNSMDP)
Paris – France

Encadrants :
Jean-Marc Lyzwa, ingénieur du son au Service Audiovisuel du CNSMDP
Olivier Warusfel, responsable de l'équipe Espaces Acoustiques et Cognitifs à l'Ircam
avec la collaboration de Thibaut Carpentier, Alexis Baskind et Markus Noisternig

Mars 2012

Abstract

As of 2012, the most widely-used technique for sound spatialization remains the 5.1 format (ITU-R BS.775), which presents serious limitations regarding the precision of the localization of lateral sources.

To achieve a more realistic surround image, a specific model has been conceived, with the purpose of putting the listener at the conductor's place. This model is based on a multi-layer approach involving traditional intensity panning as well as transaural techniques. Thanks to crosstalk-cancellation filters optimized for timbre developed at Ircam by Clément Cornuau, the use of transaural audio was made possible without sensitively damaging the spectrum, even from a musical perspective.

After having defined the key parameters of this system and their interdependance, the model has been integrated into a multichannel environment, thus enabling its use in a professional post-production situation ; additional features include stereophonic downmixing through virtualized loudspeakers.

The perceptive quality evaluation we performed show that our method provides significant improvements on the spectrum over standard transaural audio and enhanced lateralisation when applied to 5.1 post-production.

Keywords : 5.1 Multichannel, Transaural Audio, 3D Sound, Localisation, Perception, Evaluation.

Résumé

Aujourd'hui le format le plus répandu de diffusion audio multicanale, le 5.1 (recommandation ITU-R BS.775) comporte des lacunes dans la précision de la localisation des sources latérales.

Afin de parvenir à une image plus réaliste, un modèle spécifique utilisant les techniques transaurales a été proposé, visant à mettre l'auditeur au centre de la scène sonore. Ce modèle est basé sur une superposition de couches utilisant le panoramique d'intensité habituel ainsi qu'un panoramique transaural. Grâce à des filtres optimisés pour le timbre, développés à l'Ircam par Clément Cornuau, les techniques transaurales sont applicables à la musique sans dégradation notable de la qualité sonore.

Une fois définis ses principaux paramètres et leurs inter-dépendances, le modèle a été intégré à un environnement professionnel de post-production multicanale ; d'autres fonctionnalités ont été développées, comme la réduction à deux canaux au moyen de haut-parleurs virtuels.

Les tests d'écoute menés pour évaluer la qualité perceptive du prototype montrent une amélioration sensible du spectre par rapport aux techniques transaurales conventionnelles, et une meilleure latéralisation des sources pour le 5.1.

Mots-clés : Multicanal 5.1, Transaural, Son 3D, Localisation, Perception, Tests d'écoute.

Remerciements

Merci à Jean-Marc Lyzwa, Thibaut Carpentier, Olivier Warusfel, Clément Cornuau,
Hélène Diakhaté, Alexis Baskind et Markus Noisternig pour leur aide précieuse,

Merci à tous ceux qui ont participé aux tests perceptifs,

Merci à tous mes professeurs,
en particulier à Thierry Garacino, Alban Sautour, Pascal Besnard, Patrick Sigwalt,
Jean Taxis, Marie-Jeanne Séréro, Jean-Pierre Halbwachs, Klaus Blasquiz,
Jean-Christophe Messonnier, François Eckert, Isabelle Duha et Jean-François Boukobza.

Merci aux FSMS pour ces quatre années au Conservatoire,

Merci à Ariane,
à ma famille, à mes amis.

Avant-propos

Ce mémoire de fin d'études s'inscrit dans le cadre d'un stage réalisé entre septembre 2011 et janvier 2012 à l'Institut de Recherche et Coordination Acoustique/Musique (Ircam) dans l'équipe Espaces Acoustiques et Cognitifs, en collaboration avec le centre audiovisuel du Conservatoire de Paris, sous la direction de Jean-Marc Lyzwa, ingénieur du son et chargé de cours au Conservatoire de Paris.

L'implémentation des méthodes détaillées au long de ce compte-rendu et le développement de leur interface dans l'environnement logiciel Max/MSP accompagnent ce travail écrit.

Les références sont indiquées entre crochets, et invitent le lecteur à consulter la bibliographie située en fin de dossier.

Table des matières

I.Le modèle.....	9
I.1.Le multicanal 5.1.....	9
I.1.a.Présentation du système de restitution 5.1.....	9
I.1.b.Présentation des lois de panoramiques utilisées.....	10
I.1.c.Les sources latérales.....	11
I.1.d.La gestion des appoints microphoniques.....	12
I.2.Le modèle binaural.....	13
I.3.Application aux traitements transauraux.....	14
I.4.Limites et optimisations de la synthèse transaurale.....	15
I.4.a.Régularisation des HRTF.....	15
I.4.b.Conditionnement des filtres transauraux.....	16
I.4.c.Filtrage fréquentiel du CTC.....	17
I.4.d.Réduction de l'amplitude.....	17
I.4.e.Recours au double transaural.....	18
I.5.Conclusion.....	19
II.Le prototype.....	20
II.1.Conventions.....	20
II.2.Gestion des filtres.....	20
II.2.a.Conversion des types des filtres.....	20
II.2.b.Choix de la fréquence d'échantillonnage.....	21
II.3.Diagramme de fonctionnement.....	21
II.3.a.Cas d'une source monophonique.....	21
II.3.b.Passage au transaural.....	24
II.3.c.Schéma de principe.....	25
II.3.d.La gestion du système principal.....	26
II.3.e.Les possibilités de <i>downmixes</i>	26
II.4.L'interface.....	27
II.5.Interdépendance de paramètres.....	29
II.5.a.Cas de l' <i>auto front/back</i>	29
II.5.b.Cas de l' <i>auto dry/wet</i>	30
III.Tests perceptifs.....	34
III.1.Présentation.....	34

III.1.a.Objectifs.....	34
III.1.b.Contexte.....	34
III.2.Test de timbre.....	35
III.2.a.Protocole.....	35
III.2.b.Résultats.....	37
III.3.Test de localisation.....	38
III.3.a.Objectifs.....	38
III.3.b.Description de l'Ambisonics.....	39
III.3.c.Protocole.....	39
III.3.d.Résultats.....	40
III.4.Test de balance.....	42
III.4.a.Protocole.....	42
III.4.b.Résultats.....	43
IV. Conclusions.....	45
IV.1.Résultats.....	45
IV.2.Perspectives.....	46
Annexe.....	47
Bibliographie.....	48

Introduction

Alors que les outils technologiques se démocratisent, une réflexion importante est actuellement menée dans le cadre des nouveaux médias : en témoignent l'expansion de l'audiovisuel à domicile avec le DVD, le Blu-Ray, la télévision HD et les applications de diffusion par Internet, mais aussi l'arrivée prochaine du binaural dans le domaine de l'écoute mobile. Ces évolutions s'accompagnent de nouveautés, mais s'appuient aussi sur des normes déjà existantes ; c'est le cas du 5.1, le format multicanal le mieux implanté aujourd'hui. Sa compatibilité avec la stéréophonie à deux canaux et la sensation d'immersion qu'il apporte, mais aussi sa notoriété et son succès commercial lui confèrent une place importante dans la production audiovisuelle actuelle.

Cependant, ce format souffre de limitations importantes, notamment au niveau de la latéralisation des sources. De nombreux travaux se sont penchés sur les inconvénients liés à la norme 5.1, et avancent des solutions pour pallier ces problèmes bien identifiés. Dans cette étude, nous nous intéressons au modèle combinatoire proposé par Jean-Marc Lyzwa et l'Ircam, modèle qui répond aux faiblesses de localisation et de stabilité des sources latérales par le recours à une couche de panoramiques par synthèse transaurale.

Lors du stage mené à l'Ircam, nous avons développé une implémentation fonctionnelle de ce modèle. Le prototype ainsi réalisé s'appuie sur les filtres transauraux, développés par Clément Cornuau dans son étude, qui permettent une sensible amélioration du timbre de la synthèse transaurale, rendant ainsi son utilisation possible en post-production musicale professionnelle.

Après avoir présenté le format multicanal 5.1 puis les techniques binaurales et transaurales, nous reprendrons dans un premier temps les modifications des filtres transauraux effectuées par Clément Cornuau. Puis, dans un second temps, nous détaillerons leur intégration au modèle combinatoire et les enjeux qu'elle soulève : quels sont les paramètres clés et leurs plages de variations ? Comment proposer de manière simple une réduction du nombre de paramètres, dans un souci de simplification ? Enfin, dans un troisième temps, nous présenterons les premiers tests perceptifs menés à l'Ircam permettant d'évaluer les performances des nouveaux filtres transauraux et de notre système, notamment sur le plan du timbre et de la localisation.

I. Le modèle

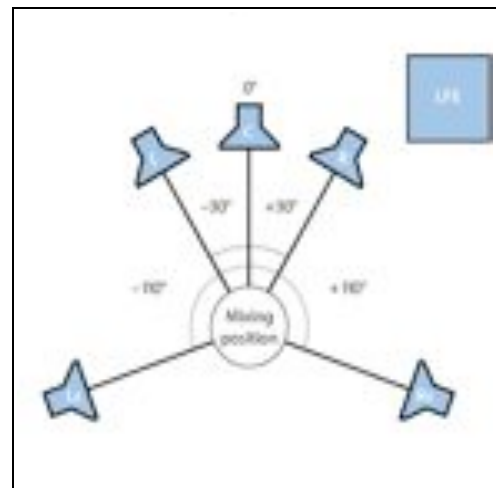
I.1. Le multicanal 5.1

I.1.a. Présentation du système de restitution 5.1

On peut regrouper les techniques de spatialisation sonore en trois familles : les modèles reposant sur la reconstitution d'un champ physique, tels que la WFS (*Wave Field Synthesis*, ou Synthèse par front d'onde) et l'Ambisonics ; le modèle « signal », basé sur un échantillonnage de la fonction de directivité de la tête et la reproduction de la localisation binaurale naturelle, et enfin, les modèles dits « perceptifs », dont sont issus la stéréophonie à deux canaux et ses extensions (2.0, 4.0, 5.0, 5.1, 6.1, 7.1, etc.).

De par la complexité de leur mise en œuvre pratique, l'Ambisonics et la WFS sont à l'heure actuelle beaucoup moins répandus que le 5.1, qui est le format multicanal le mieux implanté dans l'industrie audiovisuelle : Home Cinema, DVD (Dolby Digital AC-3 / DTS), SACD et DVD-A, Blu-ray.

La norme ITU (recommandation ITU-R BS.775) liée au système 5.1 spécifie l'emplacement des points de diffusion et leur calibration : les cinq haut-parleurs large bande L, C, R, L_s et R_s doivent être situés sur un cercle, aux azimuts respectifs -30°, 0°, +30°, -110° et +110° (avec, pour L_s et R_s, une marge de tolérance de 10° autour des positions prescrites) [Signoret 2011]. Le LFE (Low-Frequency Effects) désigne le canal réservé aux basses fréquences. On peut remarquer que les enceintes L et R sont placées à -30° et +30°, assurant une compatibilité directe avec la stéréophonie à deux canaux.



1. Norme ITU sur la diffusion multicanale 5.1

Le multicanal 5.1 a d'abord été développé pour le cinéma ; le canal central est utilisé pour la parole, les canaux frontaux L et R contiennent l'essentiel des bruitages et ambiances, tandis que les canaux arrières L_s et R_s sont le plus souvent dédiés à des effets d'immersion, et impliquent rarement des sources dont la localisation est précise [Courribet 2005].

I.1.b. Présentation des lois de panoramiques utilisées

Du recours à la norme de diffusion du multicanal 5.1 découle le choix de la loi de panoramique utilisée. Les outils de post-production dont on dispose habituellement (stations de travail audionumériques, consoles analogiques ou numériques pour le multicanal) sont équipés d'un panoramique multicanal, sur lequel on peut agir en deux dimensions : un panoramique gauche/droite, et un panoramique avant/arrière, tous deux fonctionnant en différence d'intensité, se ramenant ainsi au cas connu du panoramique d'intensité utilisé couramment en stéréophonie à deux canaux. Il existe d'autres paramètres : la *divergence* permet de doser la quantité de signal envoyée dans l'enceinte centrale (source réelle) par rapport à celle envoyée dans les enceintes frontales L et R (image fantôme). Enfin, il est aussi possible de contrôler l'action du canal de basses fréquences pour chacune des sources au moyen d'un niveau d'envoi. Cependant, nous nous concentrerons dans cette étude sur les canaux L, R, L_s et R_s.

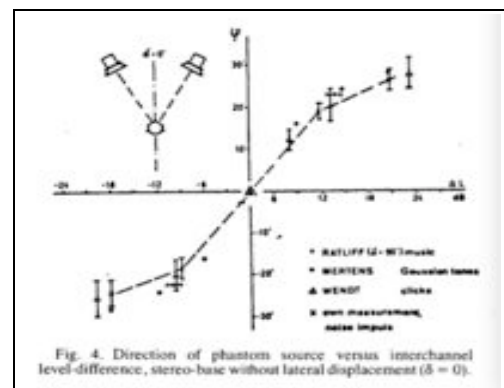
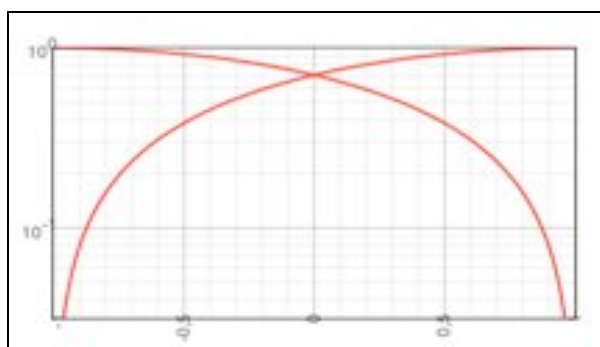
Pour rappel, la loi suivante est couramment utilisée en stéréophonie à deux canaux.

$$gauche(x) = \cos\left(\frac{\pi}{4}(x+1)\right) \quad droite(x) = \sin\left(\frac{\pi}{4}(x+1)\right)$$

où x est la valeur, comprise entre -1 et 1, sur laquelle l'ingénieur du son peut intervenir directement par l'intermédiaire d'un potentiomètre.

(1) $gauche(0) = droite(0) = \frac{\sqrt{2}}{2}$ soit 3 dB de moins que le niveau obtenu à gauche pour $x = -1$.

(2) $gauche^2(x) + droite^2(x) = 1$ c'est-à-dire l'énergie globale est conservée.



2. A gauche : loi de panoramique d'intensité à -3 dB en stéréophonie à deux canaux : atténuation des canaux gauche et droit en fonction de x .
3. A droite : direction perçue de l'image fantôme en fonction de la différence d'intensité entre les deux canaux [Theile 1977].

Il existe d'autres types de panoramiques, comme par exemple le panoramique par différence de temps, qui consiste en un décalage temporel entre les signaux issus de chacun des deux canaux. A

nouveau, on a formation d'une image fantôme qui se déplace vers le côté du canal qui est en avance par rapport à l'autre, à mesure que la différence de temps augmente.

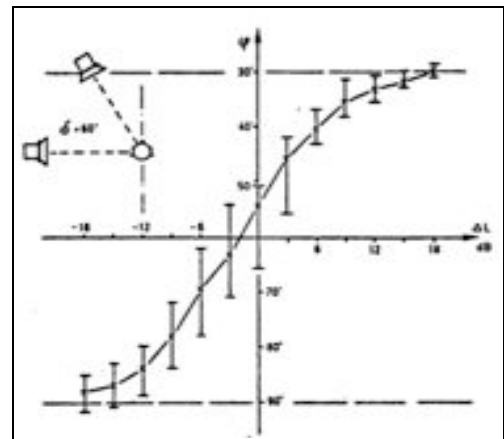
Les mécanismes perceptifs de ce type de panoramique sont d'ailleurs mis à profit dans les systèmes microphoniques dits non-coïncidents : puisque les temps d'arrivée des signaux aux capsules sont légèrement différents, leur restitution sur une base stéréophonique provoque la formation d'une image fantôme pour chaque incidence des signaux présents dans l'espace de prise de son.

Cependant, les panoramiques par différence d'intensité ont longtemps été les seuls mis en oeuvre, car leur implémentation est relativement simple sur les consoles analogiques, contrairement aux panoramiques en différences de temps qui nécessitent de recourir à des lignes à retard.

I.1.c. Les sources latérales

Les paires d'enceintes L et L_s d'une part, et R et R_s d'autre part, peuvent être considérées comme des systèmes stéréophoniques avec un écartement de 80° (au lieu de 60° pour la stéréophonie à deux canaux), avec un décalage angulaire de 70°. Le type de panoramique habituellement employé en post-production est aussi un panoramique d'intensité.

Or, des études ont été menées sur la formation des images fantômes par différence d'intensité pour des couples de haut-parleurs non frontaux. On note δ le décalage angulaire dans le plan horizontal par rapport à une base frontale symétrique. La figure ci-contre illustre les directions de l'image fantôme relevées en fonction de la différence d'intensité pour $\delta = 60^\circ$ (c'est à dire des enceintes placées respectivement à $\pm 30^\circ$ et 90°). Cette figure est à mettre en relation avec la figure n°3 de la courbe obtenue pour $\delta = 0^\circ$, soit la base stéréophonique habituelle.



4. Direction perçue de l'image fantôme en fonction de la différence d'intensité entre les deux canaux, pour $\delta=60^\circ$ [Theile 1977]

Pour $\delta = 60^\circ$, on relève des écarts de 5 à 10° autour de la valeur moyenne. Pour $\delta = 80^\circ$ (c'est à dire des enceintes placées respectivement à $\pm 50^\circ$ et 110°), on relève des écarts allant jusqu'à 30° autour de la valeur moyenne.

En fait, la localisation par image fantôme en différence d'intensité fonctionne de moins en moins bien à mesure que la base se symétrise. Un décalage angulaire de la base stéréophonique conduit aux

phénomènes suivants :

- 1) Une différence d'intensité nulle ne produit pas nécessairement une source fantôme située au centre de la base.
- 2) Une variation d'intensité plus faible produit une variation d'azimut plus importante de la source fantôme, à mesure que δ augmente ; la précision de localisation diminue.
- 3) Les variations inter-individus augmentent avec δ .

De par son irrégularité, le système multicanal 5.1 présente donc des inhomogénéités de localisation et des imperfections quant à la précision et à la stabilité des sources latérales. A cela s'ajoutent les mouvements avant-arrière de l'auditeur, qui modifient beaucoup les différences de temps et d'intensité relatives aux haut-parleurs avant et arrière. Il s'ensuit que de nombreux mixages multicanal sont élaborés comme un mixage stéréophonique frontal, avec des réverbérations ou des microphones d'ambiances envoyés aux canaux arrières.

Il est à noter que plusieurs travaux vont dans le sens d'expertises et d'améliorations de la spatialisation offerte par le 5.1 [Lund 2000] [Lyzwa 2000] [Deschamps 2003] [Craven 2003].

I.1.d. La gestion des appoints microphoniques

Aussi bien dans les productions stéréophoniques que dans les productions au format multicanal, le rôle de l'appoint microphonique est extrêmement important pour l'ingénieur du son. Dans le cas d'une prise de son dite acoustique, on choisit le plus souvent un système principal (appelé *couple microphonique* dans le cadre de la stéréophonie à deux canaux), qui donnera une image satisfaisante de la scène sonore, et le plus souvent représentative de ce qu'on peut entendre dans l'espace de prise de son, mais que l'on peut avoir besoin de compléter à l'aide de microphones d'appoint, par exemple si l'on veut rapprocher, préciser ou stabiliser une source sonore, ou encore offrir à l'auditeur un relief et un détail plus intéressants. Par ailleurs, dans les musiques amplifiées, les appoints représentent le plus souvent la majorité des signaux dont on dispose au mixage.

Or, de nombreux chercheurs se sont penchés sur la question du système principal de prise de son, sans aborder la question des appoints microphoniques, éléments incontournables lors de la post-production. Sur la base de ces constatations a été émise l'idée d'une technique qui puisse répondre aux problèmes de stabilité des sources latérales, en passant par une gestion nouvelle des appoints microphoniques dans le mixage qui favorise un étalement spatial des sources plus important que si elles se limitaient au seul secteur frontal [Lyzwa 2006].

L'objectif de cette démarche est de placer l'auditeur au milieu de la scène sonore, soit, en musique classique, à la place du chef d'orchestre, en visant ainsi à s'approcher au maximum de l'image mentale

conçue par le compositeur lors de son travail d'écriture. Cette technique s'appuie sur le modèle, dit *combinatoire*, de superposition de couches [Lyzwa 2009] :

- (1) la première des couches est formée par le système principal de prise de son, par exemple, un quintuplé microphonique, un système Ambisonics, une tête artificielle, un double MS, un Holophone H3D, etc.
- (2) la deuxième couche est formée par l'ensemble des appoints microphoniques de proximité, positionnés avec une technique traditionnelle de panoramique par différence d'intensité.
- (3) la troisième et dernière couche est formée par un panoramique transaural de ces mêmes appoints microphoniques, cette doublure visant un meilleur ancrage de l'image sonore.

Même si cette démarche est élaborée dans le contexte de la post-production de musique classique, les outils de spatialisation qui en découlent, développés dans la suite de cette étude, peuvent se généraliser pour tous les types de mixages : jazz, prises de son *live*, productions électro-acoustiques variées, mais aussi mixage film.

Nous allons maintenant expliquer les principes de la synthèse transaurale, qui conditionne la formation de la troisième couche du modèle combinatoire.

I.2. Le modèle binaural

La synthèse binaurale est un autre modèle de spatialisation sonore qui s'appuie sur le procédé naturel humain de la perception de la localisation des sources. Celui-ci utilise des indices de plusieurs types pour déterminer la position dans l'espace des sources environnantes : différences interaurales d'intensité et de temps (respectivement appelées *Interaural Level Difference* ou ILD, et *Interaural Time Difference* ou ITD), ainsi que des indices spectraux (phase et amplitude) qui sont spécifiques à l'angle d'incidence de la source et à l'auditeur, causées par des phénomènes de réflexion, d'absorption et de diffraction liées à la morphologie du sujet.

Ces données sont équivalentes à des jeux de filtres nommés HRTF (Head-Related Transfer Functions), propres à chaque individu, qui sont interprétés par le cerveau humain non pas comme des modifications de timbre, mais comme des informations de localisation [Blauert 1983].

Les HRTF peuvent être obtenues par des mesures au cours desquelles on enregistre au niveau des oreilles du sujet les deux réponses impulsionnelles correspondant aux trajets acoustiques du haut-

parleur vers chaque oreille, calculées à partir de signaux de mesure, et ce pour chacune des positions.

Il existe d'autres méthodes permettant de générer des HRTF, comme, par exemple, la simulation à partir de modélisations 3D ou de clichés 2D ou 3D, celle-ci restant toutefois au stade expérimental à l'heure actuelle.

Ainsi, pour simuler une source avec une incidence choisie, il suffit de convoluer le signal audio issu de cette source par les réponses impulsionnelles correspondant à cette incidence pour chacune des oreilles, et d'écouter le produit de convolution sur écouteurs ou casque.

Différentes limites apparaissent lors de la mise en pratique de ces techniques :

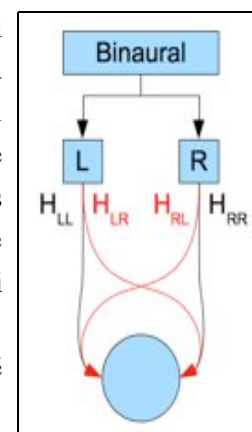
- les mesures caractérisent la perception auditive spécifique d'un individu, à la manière des empreintes digitales
- les mesures doivent être réalisées en chambre anéchoïque afin de s'abstraire de la contribution de l'acoustique de la pièce
- les mesures s'effectuent le plus souvent uniquement sur une sphère, l'effet de proximité est obtenu par des modèles additionnels (modèle géométrique par exemple)

Enfin, la reproduction de signaux binauraux est limitée à l'écoute au casque (dite *dichotique*, impliquant la séparation acoustique des canaux gauche et droit lors de la restitution).

I.3. Application aux traitements transauraux

Il existe des procédés permettant d'étendre le modèle binaural à l'écoute sur enceintes, appelés techniques transaurales.

Lors de la restitution d'un enregistrement binaural sur des haut-parleurs, si l'on ne prend pas de précautions particulières, les trajets croisés perturbent la séparation des canaux, empêchant ainsi l'écoute dichotique. En effet, le signal issu du canal gauche parvient à l'oreille gauche (H_{LL}), mais aussi à l'oreille droite (H_{LR}) ; de même, le signal issu du canal droit parvient à l'oreille droite (H_{RR}), mais aussi à l'oreille gauche (H_{RL}). On modifie donc les signaux d'origine pour que chaque oreille ne perçoive que le signal ipsilatéral en amont du transaural, et ainsi reconstituer le champ sonore d'origine au voisinage des oreilles de l'auditeur. Les traitements ainsi appliqués visent donc à supprimer les trajets croisés ; ce procédé est appelé *Crosstalk Cancellation* (CTC).



5. Les trajets croisés

Les filtres transauraux se composent des informations suivantes : H_{LL} et H_{RR} dits *trajets directs* ou *ipsilatéraux*, et sur H_{LR} et H_{RL} dits *trajets croisés*.

Depuis l'introduction de la première méthode de CTC [Schroeder 1963], plusieurs variantes ont été implémentées, correspondant à différentes structures, dont *Feed Forward asymétrique* [Gardner 1997], *Feed Forward symétrique*, *Shuffler symétrique* [Bauck 1989], *Shuffler asymétrique*. Nous ne détaillerons pas dans cette étude l'implémentation des filtres, nous conformant au choix de Clément Cornuau de travailler avec le *Feed Forward asymétrique*.

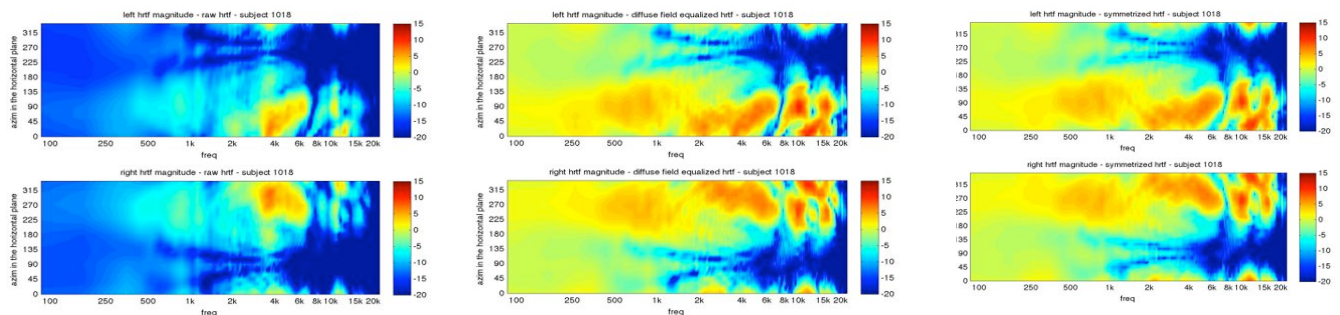
I.4. Limites et optimisations de la synthèse transaurale.

La synthèse transaurale a fait l'objet de nombreuses recherches, mais les effets de détimbrage qu'elle engendre ainsi que les contraintes importantes liées à la taille de la zone d'écoute l'ont souvent laissée à l'écart de l'audio professionnel.

Le projet présenté ici cherche d'abord à limiter les effets indésirables sur le timbre des techniques transaurales [Cornuau 2011], mais aussi à les relayer par d'autres méthodes de spatialisation dans les cas où ces dégradations apparaissent.

I.4.a. Régularisation des HRTF

Le premier des traitements est couramment appliqué aux filtres binauraux, il s'agit d'égaliser ces mesures par rapport au champ diffus (l'ensemble des mesures sur la sphère), puis de symétriser les deux filtres HRTF.



6. A gauche, mesures brutes pour les oreilles gauche et droite ; au centre, mesures égalisées par rapport au champ diffus ; à droite, mesures égalisées puis symétrisées.

Ces premiers traitements peuvent s'apparenter à un lissage des filtres, permettant de minimiser d'éventuelles erreurs de mesure ; la symétrisation est appliquée dans la perspective d'une écoute non individualisée, partant du principe que les données « asymétriques » sont probablement propres au sujet que l'on mesure.

I.4.b. Conditionnement des filtres transauraux.

Une des étapes nécessaires à l'établissement des filtres utilisés pour la synthèse transaurale est l'inversion des trajets croisés ; or dans le cas général, les HRTF ne sont pas inversibles, d'inverse stable [Cornuau 2011]. Il s'agit donc de trouver une approximation de ces trajets croisés qui soit satisfaisante pour établir les filtres transauraux.

Il existe un critère de l'inversibilité d'un système, nommé *conditionnement*, donnant une bonne indication de la qualité des filtres transauraux générés du point de vue fréquentiel. En effet, des valeurs élevées du conditionnement conduisent à des irrégularités dans la reconstruction du spectre de magnitudes, que l'on appelle *ringing frequencies*. Pour plus de détails sur le conditionnement de systèmes d'annulation de trajets croisés, le lecteur pourra se reporter aux travaux de Clément Cornuau.

Une première piste d'investigation pourrait être d'essayer de réduire l'écart angulaire entre les haut-parleurs (appelé *span*), et d'essayer des installations plus resserrées que dans la stéréophonie à deux canaux, dont le span est de 60°. Des haut-parleurs plus resserrés permettraient par ailleurs d'avoir une influence similaire de l'acoustique sur les signaux provenant des deux canaux, d'où une résistance accrue à des conditions réverbérantes [Ward 1999].

Il existe d'ailleurs une méthode de spatialisation qui s'appuie sur ces observations : il s'agit de l'*Ambiophonics* (à ne pas confondre avec l'*Ambisonics*), qui utilise des paires de haut-parleurs appelées *Ambiopoles*, dont le span est compris entre 20° et 30° [Glasgal 2003], à l'instar des systèmes stéréophoniques formés par les points de diffusion L et C d'une part, et C et R d'autre part, qui présentent un span de 30°. A ce propos, l'*Ambiophonics* préconise l'utilisation de systèmes microphoniques tels que la sphère Schoeps KFM-6, mais nous n'avons pas trouvé de références à une démarche concernant les appoints. Il se positionne toutefois comme une amélioration de la restitution pour la stéréophonie à deux canaux et pour le multicanal 5.1.

Les études menées par Clément Cornuau montrent qu'en réalité, la réduction du span ne résout pas les problèmes de *ringing frequencies* observés, tandis que des positions plus écartées de haut-parleurs permettraient une meilleure robustesse vis-à-vis des mouvements de l'auditeur.

I.4.c. Filtrage fréquentiel du CTC

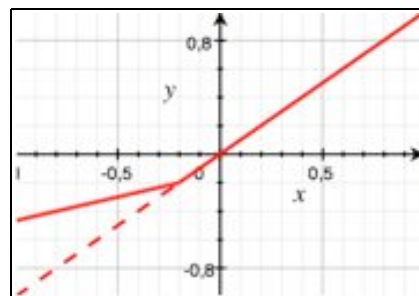
Afin de corriger les artefacts de timbre causés par l'annulation des trajets croisés, on se propose de limiter en fréquence la partie du CTC intervenant sur les trajets croisés, H_{LR} et H_{RL} [Bauck 1989]. A l'issue des travaux de Clément Cornuau, la pente du filtre passe-bas a été fixée à 30 dB par octave, et la fréquence de coupure à $f_c = 6$ kHz, fréquence au-delà de laquelle les haut-parleurs diffusent effectivement le signal binaural non traité en CTC.

En fait, on peut considérer qu'en raison de la diffraction de la tête de l'auditeur, les trajets croisés deviennent négligeables par rapport aux trajets ipsilatéraux pour les hautes fréquences. A titre indicatif, la longueur d'onde d'un signal sinusoïdal de fréquence 6 kHz est de 5,5 cm, à comparer au diamètre moyen de la tête, 17 cm.

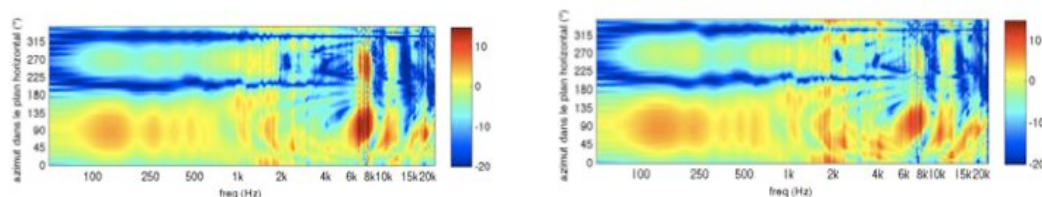
I.4.d. Réduction de l'amplitude

Lors des premières exploitations en post-production de la synthèse transaurale implémentée par Alexis Baskind, Jean-Marc Lyzwa remarque qu'une compression de la dynamique de la couche transaurale permet d'améliorer sensiblement le timbre des traitements appliqués. Il utilise pour cela les compresseurs présents dans les tranches de la console SSL Axiom MT de la régie Schaeffer.

Sur la base de ces constatations, et en s'inspirant du travail de Jean-Marc Lyzwa, Clément Cornuau propose d'intégrer cette réduction d'amplitude à la partie des filtres transauraux concernant les trajets directs, en appliquant dans le domaine spectral une compression avec une pente de 3:1 pour les magnitudes de ces filtres inférieures à -2 dB.



7. Courbe de compression utilisée pour les trajets directs [Cornuau 2011]
(x signal d'entrée, y signal de sortie)



8. Signaux envoyés au haut-parleur gauche, sujet 1087, span 60° ;
à gauche, sans compression ; à droite, avec compression 3:1. [Cornuau 2011].

Des écoutes informelles montrent que cette technique est assez efficace pour améliorer le timbre de la synthèse transaurale, en limitant les colorations marquées et les phénomènes de discontinuités spectrales, sans toutefois altérer outre mesure les performances en localisation.

Ce traitement intégré dans les filtres HRTF est linéaire, contrairement aux compresseurs « hardware », qui eux travaillent dans le domaine temporel, avec des constantes d'attaque et de retour. Ainsi, il relève davantage de la réduction de l'amplitude que de la compression de dynamique au sens où l'entendent les ingénieurs du son ; toutefois nous choisissons de nous conformer à la terminologie associée aux compresseurs traditionnels pour la pente appliquée, appelée *ratio*.

Dans son étude centrée principalement autour de la tête IRC_1087, Clément Cornuau préconise les valeurs 3:1 et 10:1 pour le *ratio*, précisant qu'un rapport de 10:1 permet de rééquilibrer le spectre pour les positions latérales de la source virtuelle, mais crée toutefois une carence en aigus à l'avant et à l'arrière.

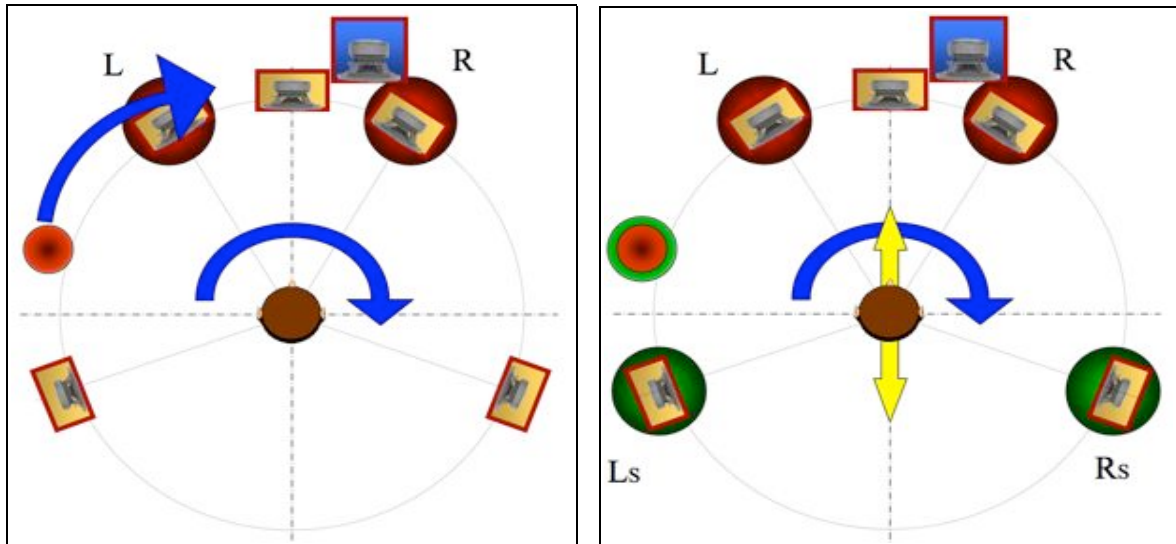
I.4.e. Recours au double transaural

A l'instar des techniques binaurales, les techniques transaurales s'appuient sur les HRTF de l'auditeur, mais dépendent aussi de la position des haut-parleurs, de leur directivité et de l'acoustique de la pièce, ainsi que de la position de l'auditeur.

Comme on cherche à recréer le champ acoustique au voisinage des oreilles de l'auditeur, l'un des points faibles bien connus de la synthèse transaurale à deux canaux est inhérent au principe de cette technique de spatialisation ; il s'agit du repliement inévitable de la scène sonore vers les haut-parleurs réels dès lors que l'auditeur modifie sa position d'écoute ou effectue une rotation de la tête, causant ainsi la perte immédiate de l'immersion et de l'extra-largeur apportée par le transaural, ou de la précision de la localisation de la source sonore latérale.

Jean-Marc Lyzwa et Alexis Baskind ont proposé le recours aux haut-parleurs arrières et à une seconde synthèse transaurale avec le *span* 240° pour relayer le couple transaural avant.

On peut inférer que le repliement qui est alors occasionné devient négligeable. Il faut trouver la balance adéquate entre les deux couples transauraux pour que la perception par images fantômes prenne le relai sur le transaural, lors d'éventuels déplacements de l'auditeur.



9. A gauche : repliement de l'image occasionné par une rotation de la tête, système transaural frontal.
A droite : le double transaural stabilise les sources latérales. [Lyzwa 2009]

I.5. Conclusion

Au cours de cette première partie, nous avons eu l'occasion de présenter le multicanal 5.1, les techniques transaurales, et le modèle combinatoire basé sur la superposition de couches permettant d'améliorer la spatialisation des sources latérales pour le 5.1.

Dans le chapitre suivant, nous allons nous intéresser à l'intégration de ce modèle à un environnement de post-production multicanale, ce qui implique d'en définir les paramètres principaux et de délimiter leurs plages de variations, mais aussi d'évaluer leurs éventuelles inter-dépendances, dans un souci de simplification de l'utilisation.

Nous avons choisi d'implémenter ce modèle à l'aide du logiciel Max/MSP en utilisant des bibliothèques du *Spatialisateur*, un programme développé à l'Ircam par l'équipe Espaces Acoustiques et Cognitifs ; le terme *prototype* renvoie au résultat de cette implémentation.

II. Le prototype

II.1. Conventions

On définit l'espace comme un modèle sphérique, dont les coordonnées sont l'azimut, l'élévation et la distance. Si un sujet est placé au centre de l'espace ainsi défini, une incidence de 0° d'azimut et 0° d'élévation correspond à une incidence frontale dans le plan horizontal de l'auditeur.

Les incidences comprises entre 0° et 180° d'azimut sont situées dans le demi-espace à droite du sujet, celles comprises entre -180° et 0° (ou, indifféremment, entre 180° et 360°) dans le demi-espace à sa gauche. Les incidences comprises entre 0° et 90° d'élévation sont situées dans l'hémisphère supérieur de la sphère, celles comprises entre 0° et -90° dans l'hémisphère inférieur.

II.2. Gestion des filtres

II.2.a. Conversion des types des filtres

A l'issue des travaux de Clément Cornuau, nous disposons de plusieurs filtres implémentés sous forme de filtres à réponse impulsionnelle finie (*Finite Impulse Response* ou FIR), filtres ayant comme inconvénient majeur un coût de calcul important. Pour la synthèse binaurale comme pour la synthèse transaurale, il est possible de modéliser les HRTF sous forme d'un retard pur (ITD) et d'un filtre à phase minimale, à réponse impulsionnelle infinie (*Infinite Impulse Response* ou IIR, appelé filtre auto-régressif), permettant ainsi une réduction de la charge en DSP [Steiglitz 1965]. La conversion des filtres FIR en IIR a été faite sous Matlab.

A titre d'exemple, sur un MacBook Pro Intel Core 2 Duo 2.5 GHz, dans l'environnement Max/MSP, à une fréquence d'échantillonnage de 48 kHz, la synthèse transaurale d'un seul flux binaural sur une seule paire de haut-parleurs utilise près de 40% des ressources processeur pour un filtre à réponse impulsionnelle finie à l'ordre 1024, là où son équivalent IIR biquadratique au quarante-huitième ordre utilise de 3 à 4% de ressources. Pour plus de détails, il existe des travaux qui se penchent sur l'analyse objective de l'impact de la modélisation IIR [Larcher 2001].

La conversion des filtres vers une structure auto-régressive, même d'ordre élevé, a ainsi permis de réduire considérablement la charge DSP, et ainsi d'implémenter plus facilement ces techniques transaurales dans le cadre de la post-production en multicanal.

II.2.b. Choix de la fréquence d'échantillonnage

La majorité des média concernés par la production audio en format multicanal fonctionne avec une fréquence d'échantillonnage de 48 kHz, ou un multiple de 48 kHz (le plus souvent, 96 kHz ou 192 kHz). Il nous a donc fallu recalculer les filtres utilisés par le prototype pour une exploitation à une fréquence d'échantillonnage de 48 kHz, car la campagne de mesures HRTF menée par l'Ircam a été faite avec une fréquence d'échantillonnage de 44.1 kHz.

Cette conversion de fréquence d'échantillonnage a été effectuée sous Matlab. Les précautions nécessaires ont été prises pour éviter les phénomènes de repliement spectral.

II.3. Diagramme de fonctionnement

II.3.a. Cas d'une source monophonique.

Présentation

Conformément à la description du modèle combinatoire, les appoints microphoniques figurent dans les couches de panoramique par différence d'intensité et de panoramique transaural.

Il y a au moins deux méthodes pour combiner ces deux techniques de positionnement [Lyzwa 2004]. La première consiste à travailler par bandes de fréquences, en privilégiant la synthèse transaurale dans les moyennes fréquences, et le panoramique par différence d'intensité dans les basses et hautes fréquences, ce afin d'éviter les colorations spectrales apportées par le transaural tout en conservant les indices de localisation se trouvant dans la bande médium. Cette méthode est assez délicate à mettre en oeuvre à cause des problèmes de détimbrage rencontrés autour des points de recouvrement.

L'autre méthode propose de combiner les deux techniques de positionnement en travaillant sur l'ensemble de la bande de fréquences. Nous avons choisi de privilégier cette méthode ; là où la première méthode demande de définir les caractéristiques d'un filtre comportant deux fréquences de coupure et deux pentes, la seconde méthode ne fait appel qu'à deux paramètres, qui sont la balance d'intensité et le délai (appelé aussi *calage temporel* ou *balance en temps*) entre les deux types de panoramiques.

Balance d'intensité

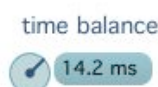


Par analogie avec la terminologie habituelle des effets audio, nous allons appeler *dry/wet* cette balance entre le panoramique par différence d'intensité et le panoramique transaural. Ce *dry/wet*, que l'on notera τ , est un nombre réel compris dans $[0;1]$, à partir duquel on obtient les deux gains respectifs

des deux techniques de positionnement. Par convention, pour $\tau=0$, le gain du panoramique par différence d'intensité est unitaire et celui du panoramique transaural est à $-\infty$ (*dry*) ; pour $\tau=1$, c'est l'inverse (*wet*), le gain du panoramique transaural est à 0 dB, et celui du panoramique par différence d'intensité est à $-\infty$.

Les lois de ce type en *dry/wet* sont exploitables en production lorsque l'énergie globale est conservée : ce réglage doit permettre de faire varier le paramètre étudié, en l'occurrence le taux de panoramique transaural utilisé dans la balance, sans faire varier l'impression de niveau global. Cette hypothèse peut s'écrire $dry^2 + wet^2 = 1$; on se propose d'utiliser une loi de la même forme que celle couramment utilisée en stéréophonie à deux canaux pour le panoramique par différence d'intensité (cf. I.1.b.).

Balance en temps



La balance en temps entre les deux couches correspondant aux deux techniques de positionnement sera notée δt . Par convention, pour les valeurs positives de δt , on introduit un délai dans le signal positionné par synthèse transaurale ; pour les valeurs négatives de δt , c'est le signal du panoramique d'intensité qui est retardé.

La loi du premier front d'onde – parfois désignée sous les noms *effet Haas*, ou *effet de précedence* – indique que sous certaines conditions, deux stimuli parvenant aux oreilles dans un intervalle de temps assez court sont perçus comme un seul dont la localisation est donnée par le premier, le second étant interprété par le système perceptif comme des modifications du timbre du premier. [Blauert 1983]

Au-delà d'une certaine durée, le cerveau humain perçoit deux sons distincts, cette durée variant de quelques millisecondes pour les impulsions à quelques dizaines de millisecondes (de 40 à 50 ms) pour des sons plus complexes. Les valeurs maximales que l'utilisateur donne à δt sont donc supposées ne pas dépasser 50 ms, d'où la plage de variations choisie pour δt de -50 ms à 50 ms ; à l'intérieur de cette plage, l'utilisateur peut jouer sur le calage temporel des couches transaurale et en panoramique par différence d'intensité pour affiner la combinaison des deux techniques.

L'une des possibilités qu'offre cette technique mixte de balance entre les deux couches de spatialisation est qu'elle permet d'utiliser le panoramique transaural comme premier front d'onde pour donner l'indice de localisation, tout en intégrant les informations de timbre de la couche de panoramique par différence d'intensité retardée. On peut imaginer que de nombreuses autres possibilités sont à explorer par les utilisateurs...

Azimuts

vbap azimuth
☐ follow main pan

On note que les positions respectives des sources correspondant à chacune des deux techniques de positionnement ne sont pas nécessairement les mêmes ; par défaut, l'azimut de la source en panoramique par différence d'intensité est identique à celui de la source en panoramique transaural, mais selon les sujets et les stimuli, les directions apparentes des deux sources ne coïncident pas toujours. C'est la raison pour laquelle nous choisissons de laisser à l'utilisateur la possibilité d'ajuster séparément les azimuts associés à chacune des deux techniques de positionnement, par l'intermédiaire d'un paramètre supplémentaire contrôlant l'azimut de la source en panoramique par différence d'intensité, pouvant éventuellement suivre celui de la source transaurale, ou encore par l'introduction d'un paramètre appelé *offset* qui correspond à l'écart angulaire entre les deux sources, réglé à 0° par défaut.

Panoramique par différence d'intensité

La méthode retenue pour la couche de panoramiques par différence d'intensité est la VBAP (*Vector-Based Amplitude Panning*) ; dans le cas du 5.1, si on fait abstraction de la notion de divergence, elle est similaire à la loi de panoramique adoptée par les consoles numériques, à ceci près que les consoles ne manipulent généralement pas des azimuts en degrés, mais plutôt des valeurs allant de -100 à 100 pour la latéralisation et le contrôle avant/arrière.

La bibliothèque du *Spatialisateur* intègre une implémentation de la VBAP, il suffit d'indiquer les positions des haut-parleurs.

Synthèse binaurale

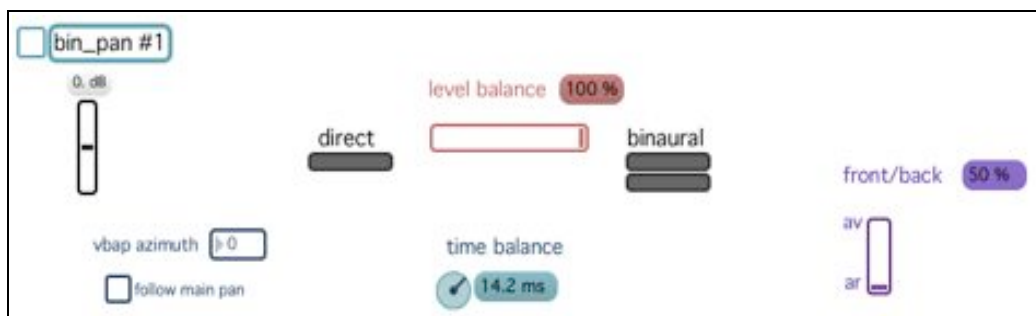
Tandis que la partie *dry* du signal, éventuellement retardée et atténuée, est envoyée vers le panoramique par différence d'intensité puis routée vers les sorties du prototype, la partie *wet* du signal passe dans un bloc appelé `spat.bin` recevant les informations de positions de la source, et réalisant le panoramique binaural associé à cette source. Nous avons choisi d'implémenter la distance apparente des sources au moyen des modèles présents dans l'objet.

Enfin, la partie *wet* du signal, désormais en format binaural, doit être répartie entre les deux couples transauraux. Pour ce faire, nous utiliserons le même type d'objet que pour un panoramique par différence d'intensité, afin de déterminer les deux gains respectifs des transauraux avant et arrière ; le paramètre contrôlant cette balance transaurale avant/arrière est appelé *front/back*, et sera noté ϕ . Le

lecteur pourra se reporter au paragraphe I.1.b sur la loi de panoramique par différence d'intensité, en remplaçant simplement les occurrences de x par ϕ .

Implémentation

En termes de programmation Max/MSP, on envoie chacune des sources monophoniques vers des blocs appelés `source_control`. Pour chacune de ces sources, la position (azimut, élévation et distance pour la partie binaurale ; azimut seul pour la partie en panoramique d'intensité) et les valeurs de τ , de δt et de ϕ sont des données réglables par l'utilisateur au moyen de l'interface proposée ; ce bloc `source_control` réalise la double balance en niveau et en temps entre les deux techniques de positionnement, puis génère la version binaurale *wet* au moyen de l'objet `spat.bin`, auquel il faut indiquer les filtres binauraux à utiliser, et enfin lui applique respectivement les atténuations correspondant à chacun des paires transaurales avant et arrière.



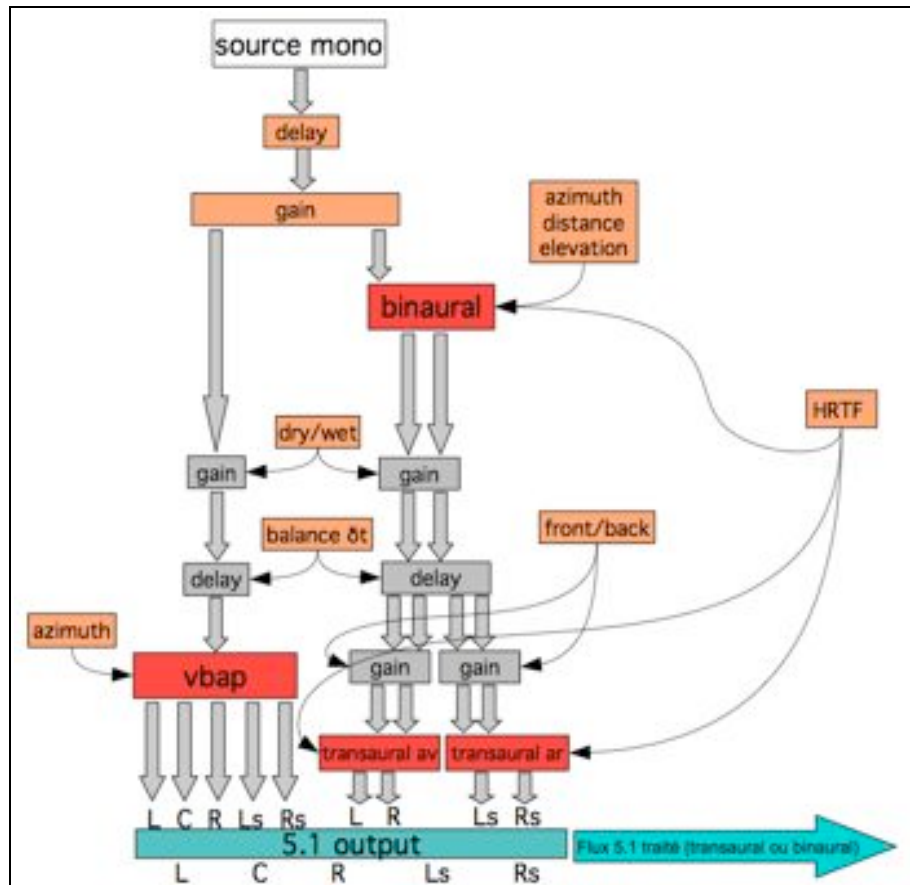
10. Interface de l'objet `source_control`

II.3.b. Passage au transaural

Les deux blocs suivants sont ceux qui vont transformer les signaux binauraux de l'étape précédente en signaux transauraux, on les appelle respectivement `trans_av` et `trans_ar`. Leur implémentation est directe, encore une fois grâce aux bibliothèques du *Spatialisateur*, qui comprend des objets nommés `spat.bin` (`bin2trans`) capables d'implémenter les nouveaux filtres optimisés pour le prototype et auxquels il faut fournir les filtres transauraux adéquats, pour chacun des spans implémentés, en l'occurrence 60° pour L et R, et 240° pour L_s et R_s .

II.3.c. Schéma de principe

Voici le diagramme reprenant le fonctionnement du prototype pour une source monophonique. Les blocs représentés en orange sont les paramètres que l'utilisateur peut ajuster ; les blocs rouges sont les blocs faisant appel aux bibliothèques du *Spatialisateur*.



11. Schéma de fonctionnement du prototype pour une source monophonique.

Le fonctionnement détaillé ci-dessus pour une source monophonique peut-être facilement étendu aux sources stéréophoniques, auquel cas le *dry/wet*, la balance en temps et le *front/back* seront des paramètres communs aux deux canaux.

II.3.d. La gestion du système principal.

L'implémentation directe du quintuplé microphonique de Jean-Marc Lyzwa (pour plus de détails, voir [Lyzwa 2009]), ainsi que sa compatibilité directe avec les mixages multicanal 5.1 nous a tourné vers le choix exclusif de ce système principal, dans le cadre de cette étude. Néanmoins, il est évidemment possible de travailler avec n'importe quel type de système de prise de son multicanal,

comme les systèmes Ambisonics DPA ou Soundfield, le double MS, l'Holophone H3D, les têtes artificielles, pour peu que l'on sache les décoder et les intégrer à un environnement de post-production multicanal 5.1.

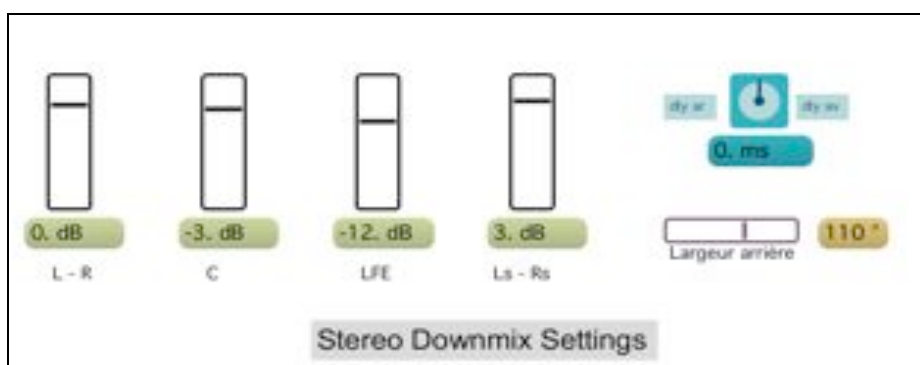
II.3.e. Les possibilités de downmixes

On appelle *downmix* une réduction d'un mixage vers un nombre moins important de canaux. Par exemple, dans le cadre d'une post-production 5.1 pour la bande son d'un film, on réalise un mixage 5.1 pour le cinéma (voire 5.1 EX, 6.1, 7.1), mais on doit également fournir une version stéréophonique à deux canaux pour la télévision, ou le DVD par exemple. Cette version peut être générée automatiquement à partir du mixage 5.1, ce qui est généralement proposé par les consoles multicanal ; ainsi, la console SSL C300 HD propose une matrice de réduction de la forme suivante.

$$L' = L + (C - 3\text{dB}) + L_s + (LFE - 12\text{dB}), R' = R + (C - 3\text{dB}) + R_s + (LFE - 12\text{dB})$$

Cependant, ce type de réduction pour la stéréophonie à deux canaux replie l'intégralité du mixage multicanal dans la base frontale, ce qui entraîne une perte de la sensation d'enveloppement et de la localisation des sources latérales et arrière, mais aussi des effets de masquage, voire des détimbrages importants. Nous nous proposons d'apporter une réponse à cette limite de la réduction stéréophonique par le recours à la synthèse transaurale, et aux possibilités de virtualisation de sources qu'elle offre. En intégrant une version transaurale des canaux arrière L_s et R_s dans la paire de haut-parleurs frontaux, on peut ainsi conserver les sensations d'immersion et de l'extra-largeur apportée par les sources latérales.

Trois paramètres entrent en jeu lors de ce processus de réduction stéréophonique : le niveau des canaux arrière, les positions des sources virtuelles, qui jouent sur la sensation de largeur arrière, et le délai entre l'avant et l'arrière, permettant de réduire les effets de masquage éventuellement induits lors de la réduction. Le canal central et le canal de basse fréquence sont, eux, toujours réinjectés dans la base stéréophonique frontale ; on dispose pour cela de deux réglages pour les niveaux d'envoi respectifs de ces deux canaux.

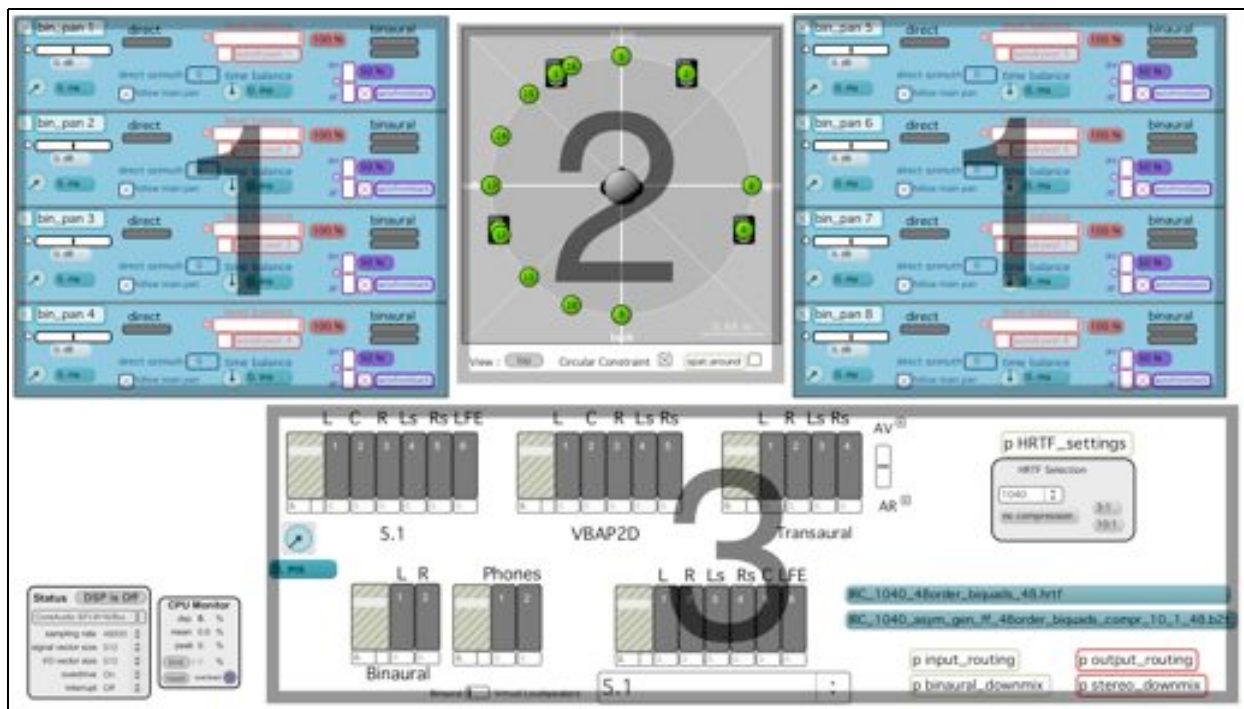


12. Interface de contrôle des paramètres du downmix stéréo.

II.4. L'interface

La première interface que nous proposons pour ce prototype comporte trois zones principales :

- (1) les contrôles séparés de chacune des sources traitées, impliquant le nombre correspondant d'instances de l'objet `source_control`,
- (2) la fenêtre de visualisation des sources issue du *Spatialisateur*, rassemblant les positions de chacune des sources traitées,
- (3) le contrôle plus global du fonctionnement du prototype : choix des filtres binauraux et transauraux utilisés, visualisation et réglage des niveaux de chacune des couches du modèle combinatoire, paramétrage des sorties du prototype.



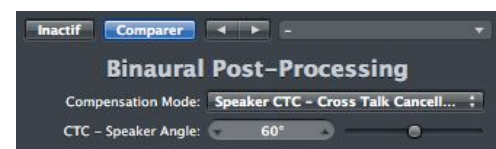
13. La fenêtre principale de la version Max/MSP du prototype.

Le développement du prototype en un seul bloc est direct dans Max/MSP, mais peu adapté au contexte de la post-production audio intégrée à un DAW (*Digital Audio Workstation*, ou station de travail audionumérique) comme Pyramix, ProTools, Logic ou encore d'autres systèmes. En effet, le traitement de n sources simultanées suppose d'écrire une version du prototype comportant n entrées, alors que l'ingénieur du son est plus habitué à ouvrir un *plug-in* mono ou stéréo en insert pour chacune des tranches audio du DAW dédiées à un appoint à traiter.

Cependant, cette architecture supposerait autant de modules réalisant la CTC que de canaux traités, là où le prototype développé sous Max/MSP ne fait appel qu'à un module transaural pour l'avant et un autre pour l'arrière, chacun agissant sur la somme des binauraux. C'est pourquoi nous proposons la structure de fonctionnement hybride suivante : une interface « mère », et des interfaces satellites. Les satellites permettent de contrôler les paramètres de la zone 1 concernant chacune des sources – position, *dry/wet*, *front/back* et balance en temps – tandis que l'interface mère offre une fenêtre regroupant les positions de toutes les sources (zone 2), les contrôles en niveau de chacune des trois couches, les réglages des sorties et le choix des filtres binauraux utilisés (zone 3).

Ce fonctionnement est d'ailleurs semblable à celui retenu par le module de synthèse transaurale proposé dans la station de travail Logic Pro : on ouvre une fenêtre *binaural panner* sur chacune des tranches déclarées binaurales par l'utilisateur, et un insert *binaural post-processing* sur le bus dans lequel sont sommés les signaux binauraux.

Le *Binaural Post-Processing* proposé par Logic Pro permet d'ailleurs d'appliquer au choix un traitement transaural CTC pour diffusion sur haut-parleurs (il est d'ailleurs possible d'ajuster le span), ou des variantes d'égalisation pour le casque : FF, pour une optimisation des incidences frontales ; HB, optimisation des incidences horizontales ; DF, égalisation en champ diffus (soit l'ensemble des directions).



14. A gauche, le Binaural Panner de Logic Pro ; à droite, le Binaural Post-Processing.

II.5. Interdépendance de paramètres

Afin de simplifier l'utilisation du prototype, nous proposons dans ce paragraphe des investigations visant une réduction du nombre de paramètres en anticipant leurs dépendances mutuelles. Deux exemples sont présentés : l'automatisation du paramètre *front/back*, puis celle du paramètre *dry/wet* en fonction de l'azimut de la source virtuelle.

II.5.a. Cas de l'*auto front/back*

Pour commencer, nous faisons l'hypothèse d'un lien existant entre l'azimut d'une part, et la balance des deux couples transauraux avant et arrière d'autre part. En termes d'énergie, si l'on veut placer une source entre les azimuts -90° et $+90^\circ$, il semble intéressant de favoriser la paire transaurale avant plutôt que l'arrière ; cela pourrait aller dans le sens de réduire les erreurs d'inversions avant/arrière, qui peuvent parfois être gênantes avec la synthèse binaurale.

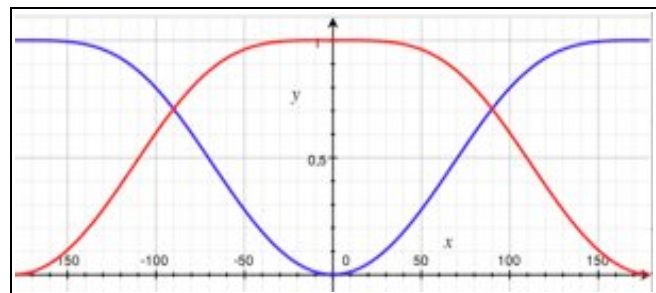
On choisit de fixer les conditions suivantes pour cette loi :

- (1) aux azimuts -90° et 90° , les deux systèmes transauraux fournissent des niveaux équivalents,
- (2) on impose la conservation de l'énergie globale de la couche par panoramique transaural,
- (3) le niveau du système transaural avant est une fonction de l'azimut croissante sur $[-180^\circ ; 0^\circ]$ et décroissante sur $[0^\circ ; 180^\circ]$

En notant α l'azimut de la source, on pose $\phi(\alpha) = \cos(\alpha)$; on remarque que $\phi(\alpha) \in [-1 ; 1]$ indique le caractère « avant / arrière » de la position de la source. Il ne reste plus qu'à obtenir les coefficients respectifs donnés pour chacun des couples transauraux. En s'inspirant de la loi de panoramique d'intensité de la stéréophonie à deux canaux, on peut formuler l'asservissement de la balance front/back à la position de la façon suivante :

$$front(\alpha) = \sin\left(\frac{\pi}{4}(\phi(\alpha) + 1)\right)$$

$$back(\alpha) = \cos\left(\frac{\pi}{4}(\phi(\alpha) + 1)\right)$$



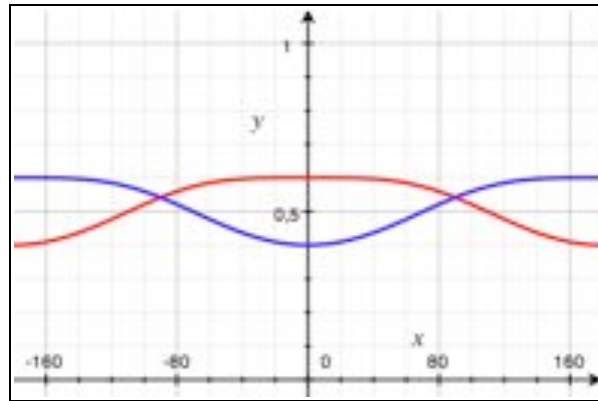
15. Courbe du paramètre *front/back* en fonction de l'azimut de la source ; en rouge, niveau de L et R ; en bleu, niveau de L_s et R_s .

Cependant, on veut aussi pouvoir pondérer l'influence de cette loi autour de la position moyenne, à l'aide d'une constante $k \in [0; 1]$. On aboutit finalement au couple d'équations suivantes :

$$front(\alpha) = k \left(\sin\left(\frac{\pi}{4}(\phi(\alpha) + 1)\right) - 0.5 \right) + 0.5$$

$$back(\alpha) = k \left(\cos\left(\frac{\pi}{4}(\phi(\alpha) + 1)\right) - 0.5 \right) + 0.5$$

Au fur et à mesure de l'utilisation de cette loi en post-production, nous sommes parvenus à la valeur de $k = 0,2$, qui donne une répartition aux extrêmes de 40% / 60%, introduisant ainsi une substantielle différence d'intensité entre les deux couples transauraux, sans trop déséquilibrer la troisième couche.



16. Courbe du paramètre *front/back* en fonction de l'azimut de la source ($k = 0,2$) en rouge, niveau de L et R ; en bleu, niveau de L_s et R_s .

II.5.b. Cas de l'*auto dry/wet*

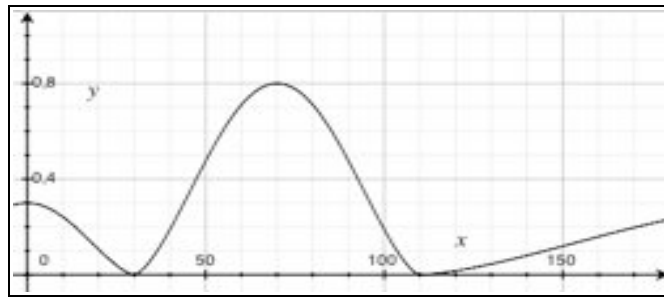
Toujours dans le but de simplifier l'utilisation du prototype, nous nous proposons à présent d'asservir le *dry/wet* à la position d'une source. Le *dry/wet* est un paramètre très important du plug-in, il y a donc quelques précautions à prendre si l'on veut essayer d'automatiser son fonctionnement. Malgré l'optimisation des filtres transauraux, le recours à des valeurs trop élevées de ce paramètre est susceptible d'entraîner des modifications importantes et non désirées du timbre de la source traitée. Il semble donc important de caractériser les portions de l'espace dans lesquelles la synthèse transaurale est à la fois la plus intéressante par rapport à l'objectif que nous visons, et la moins destructive par rapport au timbre des sources.

Le phénomène des *ringing frequencies* identifié lors de l'étude théorique de Clément Cornuau est aisément audible lorsque l'on place les sources virtuelles au voisinage des haut-parleurs. Dans tous les cas, il semble peu intéressant d'utiliser la synthèse transaurale pour positionner une source à l'endroit où

se trouve déjà un haut-parleur réel. De façon plus générale, le recours à des techniques transaurales pour positionner des sources dans l'espace frontal $[-30^\circ ; +30^\circ]$ semble peu intéressant, étant donné, d'une part, que les techniques de spatialisation de la stéréophonie à deux canaux peuvent d'ores et déjà répondre à la question de façon satisfaisante [Theile 1977], et d'autre part, que le canal central ajouté par le multicanal 5.1 contribue lui aussi à un meilleur ancrage de l'image frontale.

En revanche, c'est sur les portions latérales $[-110^\circ ; -30^\circ]$ et $[+30^\circ ; +110^\circ]$ que la synthèse transaurale peut être la plus intéressante pour consolider la position d'une source, puisque, comme nous l'avons vu dans la première partie, les techniques de panoramiques par différences d'intensité ne sont pas suffisantes dans ce cas.

Il semble donc opportun de ne pas utiliser de forts taux au voisinage des enceintes afin d'éviter les artefacts décrits par Clément Cornuau, mais de conserver des taux plus élevés autour des positions médianes. Voici l'allure de la courbe qui serait issue de cette réflexion.



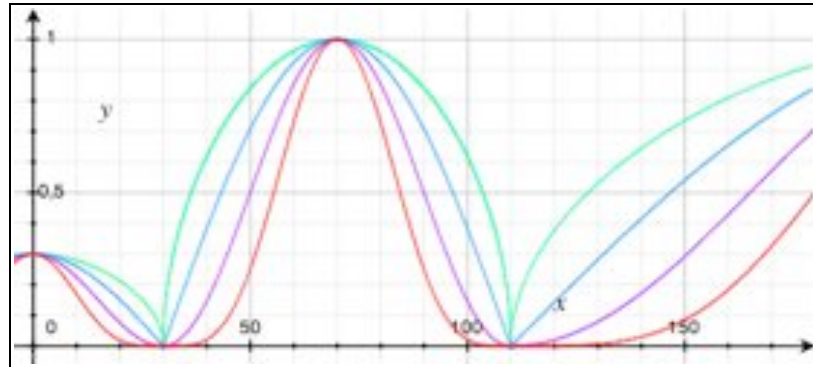
17. Allure de la courbe donnant le *dry/wet* en fonction de l'azimut, pour les azimuts positifs.

Si on fixe les racines du *dry/wet* aux azimuts 30° et 110° , emplacements des haut-parleurs, on peut maintenant établir des formules séparées suivant la portion de l'espace azimutal dans laquelle on travaille. Une fois définis $\mu_i(\alpha)$ le taux de *dry/wet* obtenu pour la portion angulaire d'index i , α l'azimut de la source, et γ une constante positive, on peut définir des coefficients multiplicatifs k_i à appliquer à chacune des formules indépendamment, de façon à pondérer cette formule en fonction de la portion angulaire délimitée par les enceintes dans laquelle on place la source.

$$\mu_{\text{frontal}}(\alpha) = k_{\text{frontal}} \left| \cos\left(\frac{\alpha}{2 \times 30} \cdot \pi\right) \right|^\gamma, \alpha \in [0 ; 30]$$

$$\mu_{\text{latéral}}(\alpha) = k_{\text{latéral}} \left| \cos\left(\left(\frac{\alpha - 30}{110 - 30} - 0.5\right) \cdot \pi\right) \right|^\gamma, \alpha \in [30 ; 110]$$

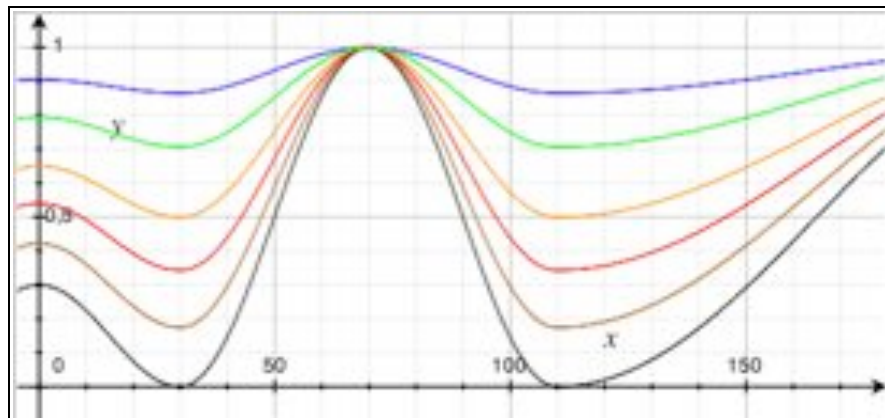
$$\mu_{\text{arrière}}(\alpha) = k_{\text{arrière}} \left| \cos\left(\frac{\alpha}{2 \times 110} \cdot \pi\right) \right|^\gamma, \alpha \in [110 ; 180]$$



18. Dry/wet obtenu en fonction de l'azimut, pour $\gamma = 0,5$ (vert), $\gamma = 1$ (bleu), $\gamma = 2$ (violet), $\gamma = 4$ (rouge)

Pour asservir le taux de *dry/wet* à la position de la source, on propose d'ajouter au taux de *dry/wet* $\mu_i(\alpha)$ obtenu avec le seul azimut une quantité qui dépende de l'élévation notée ϵ , afin d'introduire une quantité non nulle de synthèse transaurale aux azimuts $\pm 30^\circ$ et $\pm 110^\circ$ pour les élévations non nulles.

$$v_i(\alpha, \epsilon) = \mu_i(\alpha) + \sin(\epsilon)(1 - \mu_i(\alpha)) = \sin(\epsilon) + (1 - \sin(\epsilon))\mu_i(\alpha)$$



19. Dry/wet obtenu pour $\epsilon = 0^\circ$ (noir), $\epsilon = 10^\circ$ (marron), $\epsilon = 20^\circ$ (rouge), $\epsilon = 30^\circ$ (orange), $\epsilon = 45^\circ$ (vert) et $\epsilon = 60^\circ$ (bleu).

Enfin, de la même façon que pour l'*auto front/back*, on veut pouvoir pondérer l'influence de cette formule, et lui donner des valeurs minimale et maximale m et M à respecter, où $m, M \in [0; 1]$.

Finalement, le *dry/wet* peut s'exprimer ainsi :

$$\tau_i(\alpha, \epsilon) = m + \frac{1-m}{M}(\sin(\epsilon) + (1 - \sin(\epsilon))\mu_i(\alpha))$$

L'asservissement du taux de *dry/wet* en fonction de la position de la source nécessite donc d'autres paramètres, qui sont au nombre de six :

- les trois constantes k_{frontal} , $k_{\text{latéral}}$ et $k_{\text{arrière}}$
- la constante γ permettant de gérer la pente du taux au voisinage des enceintes
- les bornes inférieure m et supérieure M de la plage de variations du *dry/wet*.

On peut imaginer que ces coefficients soient fixés pour l'ensemble de l'instance du plug-in, et éventuellement accessibles à l'utilisateur chevronné, ceci permettant d'affiner le comportement du *dry/wet*, par exemple lors d'automations de la position.

Voici l'interface contrôlant les principaux paramètres de cette loi d'*auto-dry/wet*. Les paramètres front, side, rear correspondent aux trois constantes k_{frontal} , $k_{\text{latéral}}$ et $k_{\text{arrière}}$; le curseur nommé *slope* correspond à la constante γ .

Enfin, les valeurs du minimum et du maximum permettent de pondérer l'action de l'asservissement de la position sur le taux généré, taux qui apparaît sur le curseur en rouge.

20. Interface contrôlant les paramètres introduits pour l'*auto dry/wet*

III. Tests perceptifs

Cette troisième partie présente la campagne de tests perceptifs menée à l'issue du stage à l'Ircam, en février 2012. Ces tests constituent avant tout une première mise à l'épreuve du prototype développé, et pourront servir de base aux prochains travaux d'expertise le concernant.

III.1. Présentation

III.1.a. Objectifs

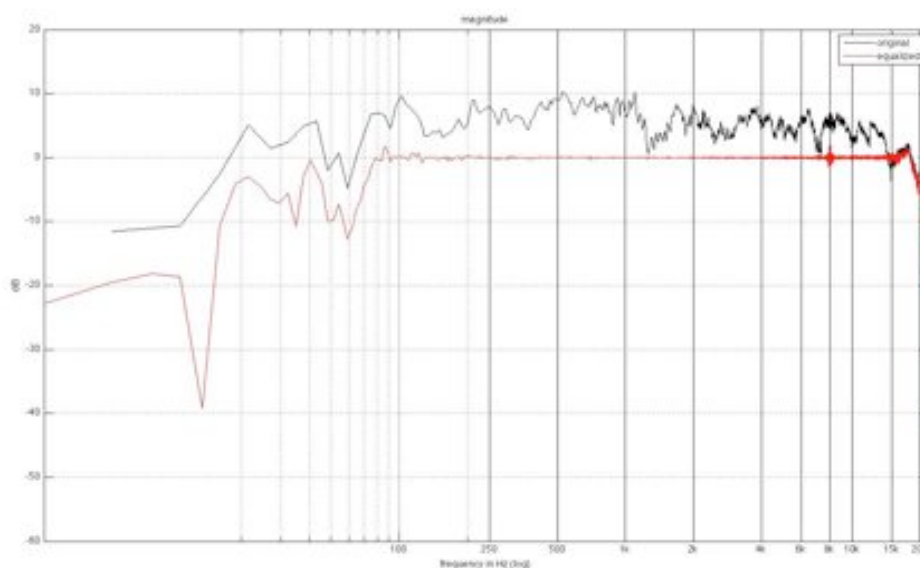
On souhaite déterminer les réglages de base du prototype pour la post-production multicanal au format 5.1, et évaluer la pertinence des hypothèses formulées dans les inter-dépendances des paramètres.

L'un des paramètres les plus importants est le choix des filtres utilisés, qui diffèrent entre eux notamment par leur couleur générale et par leurs performances en localisation.

Nous nous sommes penchés sur trois têtes HRTF issues des campagnes de mesures menées à l'Ircam : IRC_1040, IRC_1052 et IRC_1089. Le panel de têtes HRTF que nous proposons est le résultat d'un choix préliminaire parmi la base de données de l'Ircam, après écoutes informelles en binaural et en transaural, tout en essayant de conserver un panel d'ITD suffisamment large. Les maxima des valeurs de l'ITD dans le plan horizontal sont respectivement de 0,66 ms, 0,72 ms et 0,77 ms pour les têtes IRC_1052, IRC_1040 et IRC_1089.

III.1.b. Contexte

Les tests perceptifs se sont déroulés en février 2012, dans le studio 4 de l'Ircam. Nous avons utilisé pour L, R, L_s et R_s quatre enceintes Tannoy 600 passives coaxiales, alimentées par un amplificateur Yamaha, positionnées respectivement aux azimuts -30°, +30°, -110°, +110° dans le plan de l'auditeur, avec un rayon d'écoute de 1,56 m. Les réponses impulsionnelles de ces enceintes une fois placées dans le studio 4 ont été mesurées à l'aide d'un microphone de mesure au *sweet spot* puis inversées afin d'assurer les meilleures conditions d'écoute possible.



21. Spectre de magnitudes de l'enceinte Ls, avant égalisation (en noir), et après égalisation (en rouge)

Nous avons interrogé 18 participants bénévoles (12 hommes et 6 femmes), dont 15 sont ingénieurs du son (5 professionnels et 10 étudiants de la Formation Supérieure aux Métiers du Son).

III.2. Test de timbre

III.2.a. Protocole

Le but de ce premier test est d'observer le comportement spectral de plusieurs jeux de filtres, en suivant deux axes d'investigation :

- d'une part, en conservant la même tête pour les parties binaurale et transaurale du prototype (en l'occurrence, IRC_1040) et en faisant varier le type de traitements transauraux mis en jeu : sans compression, compression 3:1, compression 10:1, avec *bandlimiting* (filtrage passe-bas des trajets croisés) ; compression 3:1 *fullrange* (c'est-à-dire sans bandlimiting)
- d'autre part, en gardant le même type de traitements transauraux (bandlimiting, compression 3:1) et en faisant varier le choix de la tête HRTF (IRC_1040, IRC_1052 et IRC_1089)

Par ailleurs, nous voulions aussi savoir si la préférence pour un jeu de filtres était liée à la nature du stimulus, à son contenu spectral, et à sa localisation. Le contenu audio est choisi de sorte à pouvoir mettre en lumière les éventuels défauts des HRTF faisant l'objet du test, en particulier des colorations marquées dans les hautes fréquences [Cornuau 2011].

Le protocole de test est le suivant : en utilisant une échelle de 0 (très mauvais) à 100 (excellent), le participant doit noter sept extraits sonores, nommés de A à G, en comparaison avec une référence donnée R. Plus un extrait est jugé proche de la référence, plus le sujet doit donner une note proche de 100.

Les tests MUSHRA (*MUltiple Stimulus with Hidden Reference and Anchor*, ITU-R BS.1534) sont des tests couramment utilisés pour l'évaluation subjective du niveau de qualité intermédiaire des systèmes de codage. L'interface de ce premier test est adaptée d'une interface pour Max/MSP dédiée aux tests MUSHRA développée par Chris Hummersone (University Of Surrey). Elle comporte 6 pages, chacune d'entre elles portant sur les 7 occurrences nommées de A à G de l'extrait sonore. A l'issue des 6 pages de tests, le participant a donc donné 42 notes de 0 à 100.

Evaluez la qualité de chacun des extraits sonores suivants en utilisant les échelles.

	R	A	B	C	D	E	F	G
Très bon								
Bon								
OK								
Mauvais								
Très mauvais								
	0	0	0	0	0	0	0	0

Page 1 of 6

22. Une page de l'interface du test de timbre

La référence utilisée est un stimulus positionné au moyen d'un panoramique par différence d'intensité ; chacun des extraits est une version transaurale du stimulus de base, positionnée au même azimut que la référence, en utilisant les différents jeux de filtres que l'on veut mettre à l'épreuve.

A la différence des tests MUSHRA, nous n'avons pas choisi d'utiliser une ancre basse, notion délicate à définir dans le cadre de notre étude.

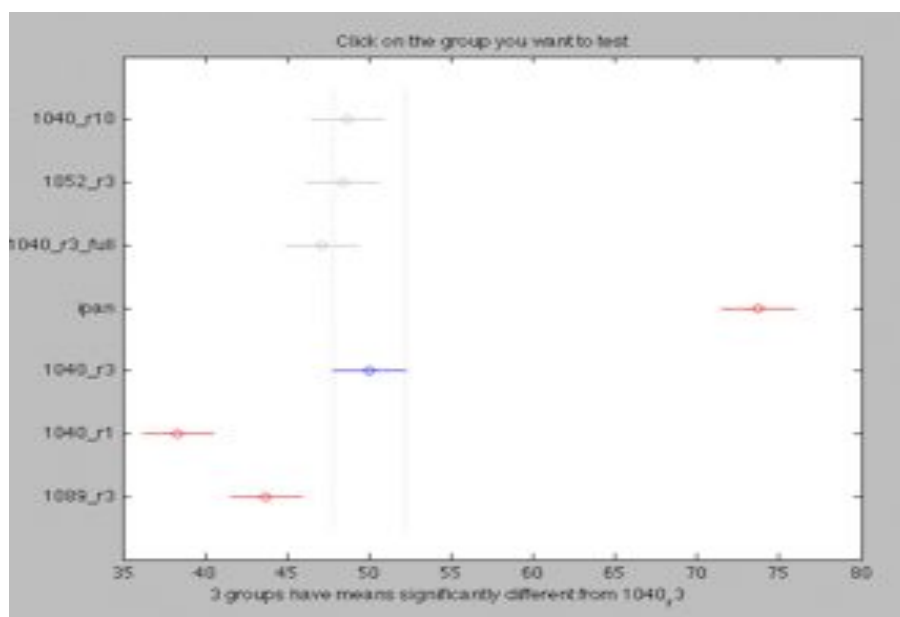
Les six extraits choisis sont les suivants :

- v1 : violon solo, -40°
- v2 : violon solo, +90°
- a1 : guitare acoustique 70° / shaker 70°
- a2 : guitare acoustique 40° / shaker 40°
- ta : voix chantée +70°
- ga : mixage multicanal de musique électronique (musique d'attente pour la Géode)

III.2.b. Résultats

Nous avons normalisé les notes pour avoir une plage de variations de 0 à 100 pour chacun des participants, ce afin de donner des poids équivalents aux réponses de tous les sujets. Les 42 stimuli notés par les 18 participants comportaient 6 extraits soumis à 7 types de traitements différents : panoramique par différence d'intensité, noté *ipan* ; tête IRC_1040 avec des ratios de 1:1, 3:1, 10:1, tête IRC_1052 avec un ratio de 3:1, tête IRC_1089 avec un ratio de 3:1, avec *bandlimiting*, notées respectivement **1040_r1**, **1040_r3**, **1040_r10**, **1052_r3**, **1089_r3** ; et enfin, tête IRC_1040 avec un ratio de 3:1, *fullrange*, notée **1040_r3_full**.

Une analyse de la variance (ANOVA pour *Analysis Of VAriance*) sur le type de traitements transauraux opérés permet de déterminer quatre groupes de filtres distincts à l'issue de ce premier test, présentés dans le tableau ci-dessous. Cette analyse statistique propose un indice de la significativité de ses résultats, appelé *p*. On considère que le paramètre observé a une influence significative sur les résultats obtenus lorsque $p < 0,05$.



23. Graphique présentant les résultats de l'ANOVA effectuée sur la note en fonction des filtres utilisés.

Les résultats de cette analyse de la variance sont repris dans le tableau ci-dessous.

Groupes de filtres	Note moyenne
1040_r1	38
1089_r3	44
1040_r3, 1040_r3_full, 1040_r10, 1052_r3	50
ipan	75

Les traitements transauraux ne mettant pas de réduction d'amplitude en jeu (1040_r1) obtiennent une note significativement plus mauvaise que les autres ($p < 10^{-6}$). L'observation de ces résultats montre que le choix du type de traitements a bien une incidence sur le timbre, ce qui semble valider les modifications proposées par Clément Cornuau dans son étude : la réduction des magnitudes des filtres permet effectivement une amélioration significative de la synthèse transaurale dans son comportement spectral.

En revanche, ni le choix de l'extrait audio traité, ni son positionnement ne semblent être des facteurs significatifs dans le cadre de ce premier test ($p = 0,39$), ni même le fait que le corpus d'extraits comprenne un mixage multicanal qui subit une réduction stéréophonique avec les différents traitements transauraux considérés. Ceci tend à montrer que les regroupements des filtres présentés dans le tableau ci-dessus semblent être valables dans un cas assez général.

Nous soulignons toutefois que 7 participants sur les 18 ont noté des difficultés dans ce premier test, difficultés portant principalement sur les critères de l'évaluation des extraits. En effet, le paramètre que l'on demandait de noter était le timbre ; cependant, d'autres différences pouvaient être perçues, notamment de niveau et/ou de proximité, mais aussi de spatialisation, par nature de la synthèse transaurale. On remarque ainsi une grande disparité de l'évaluation entre les extraits correspondant aux filtres transauraux d'une part, et à ceux correspondant à la référence en panoramique par différence d'intensité d'autre part.

III.3. Test de localisation

III.3.a. Objectifs

Nous avons envisagé de mettre à l'épreuve les performances en localisation du modèle transaural proposé. Pour ne pas allonger la durée de passage pour chaque participant, le test de timbre durant en moyenne 30 à 40 minutes, nous avons choisi de ne tester qu'un nombre limité de conditions : **1052_r3**,

1040_r3 et **ipan**, avec un seul type de stimulus (harpe enregistrée avec un microphone à ruban placé en proximité dans l'espace Maurice Fleuret du CNSMDP), positionné dans les secteurs latéraux, en nous restreignant aux élévations nulles et aux distances unitaires.

III.3.b. Description de l'Ambisonics

Pour ce second test, on utilisera un pointeur audio que le sujet doit placer à l'endroit où il entend le stimulus (appelé consigne, ou cible). Cette méthode a l'avantage de ne faire appel qu'à l'audition, par rapport à d'autres méthodes qui feraient intervenir des erreurs liées à une traduction des résultats (report de la localisation sur un graphique, ou pointage de la source avec un laser).

La diffusion sonore Ambisonics repose sur l'idée de Michael Gerzon de décomposer le champ acoustique en harmoniques sphériques. L'ordre du dispositif est déterminé par le nombre de haut-parleurs, et caractérise la précision du système.

En l'occurrence, le studio 4 de l'Ircam est équipé d'un dispositif Ambisonics d'ordre 4, dont les haut-parleurs sont équi-répartis sur une sphère de rayon sensiblement égal à celui du système d'écoute utilisé par le prototype.



24. Le studio 4 de l'Ircam, le système Tannoy et le dispositif Ambisonics

III.3.c. Protocole.

Nous utilisons le système Ambisonics pour diffuser le pointeur acoustique. Cette méthode permettra ainsi de donner directement une indication des performances en localisation du prototype par rapport à un système connu, en l'occurrence la demi-sphère Ambisonics du studio 4.

L'interface permet de passer à volonté de l'extrait proposé au pointeur Ambisonics, pour chacune des étapes du test. On met à disposition du participant une souris 3D *SpaceNavigator* pour placer ce pointeur Ambisonics dans la direction de l'extrait proposé. Nous avons eu recours à la bibliothèque `aka.spacenavigator` pour l'intégration de ce périphérique de pointage à l'interface Max/MSP utilisée pour nos tests. Une interpolation des données de contrôle du pointeur s'est avérée nécessaire, afin de permettre un contrôle plus aisé des mouvements du pointeur Ambisonics.



25. Le dispositif de pointage



26. Interface Max/MSP du test de localisation

III.3.d. Résultats

Pour chacune des mesures effectuées, nous appelons *écart azimuthal* la valeur absolue de l'écart angulaire des azimuts réponse et consigne. Une analyse de la variance de cette grandeur en fonction du mode de diffusion ($p = 0,013$) montre que la tête IRC_1052 a des performances en localisation significativement plus mauvaises (écart azimuthal moyen de 20°) que la tête IRC_1040 ou le panoramique par différence d'intensité, qui présentent tous deux des écarts azimuthaux de 15° en moyenne.

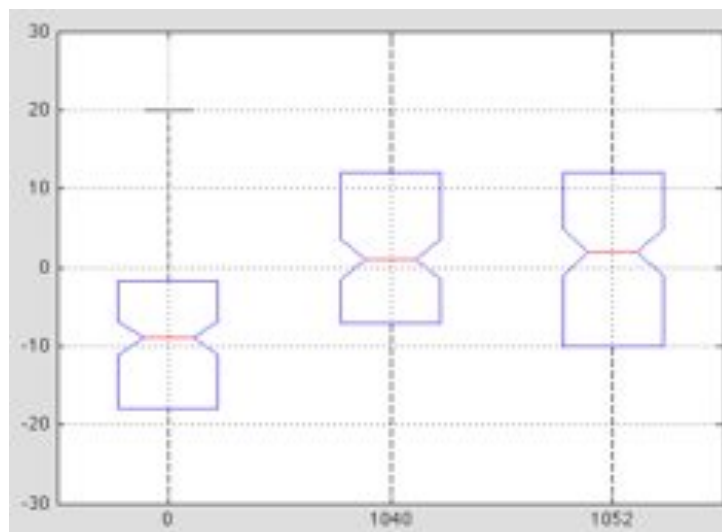
On appelle maintenant *écart en élévation* la valeur absolue de l'écart angulaire des élévations réponse et consigne (l'élévation de l'extrait consigne étant toujours nulle dans notre test). L'analyse de la variance de l'écart en élévation en fonction du mode de diffusion donne $p = 0,0004$, et révèle une erreur en élévation moyenne de 10° pour les panoramiques transauraux issus des têtes IRC_1040 et IRC_1052, valeur supérieure aux 7° obtenus avec le panoramique par différence d'intensité.

On se propose à présent d'introduire la *latéralisation*, notée λ , définie en fonction de l'azimut α par la formule suivante :

$$\lambda = \arcsin(\sin|\alpha|)$$

Ceci revient à replier la gauche sur la droite, puis l'arrière sur l'avant, faisant ainsi abstraction d'éventuelles inversions avant-arrière ; ceci peut causer des erreurs puisque l'angle 105° est replié sur 75° , entraînant alors une erreur sur l'azimut de 30° alors que la latéralisation est invariante.

L'analyse de la variance de la latéralisation en fonction du mode de diffusion donne $p = 10^{-9}$ et révèle une sous-latéralisation de l'ordre de 10° pour le panoramique par différence d'intensité, tandis que l'erreur de latéralisation liée aux panoramiques transauraux est proche de 0° .



27. Graphique présentant l'erreur de latéralisation mesurée en degrés (médiane en rouge, quartiles en bleu).
À gauche : panoramique par différence d'intensité ; au milieu : IRC_1040 ; à droite : IRC_1052.

Ce second test nous montre donc que l'apport en localisation du panoramique transaural dans les secteurs latéraux est lié à une latéralisation plus fiable que celle obtenue pour le panoramique par différence d'intensité couramment utilisé en post-production au format multicanal 5.1.

Nous présentons ci-dessous un tableau récapitulant les résultats de ces tests de localisation.

Mode de diffusion	Ecart azimuthal moyen	Ecart en élévation moyen	Latéralisation moyenne
Panoramique d'intensité	15°	7°	-9°
IRC_1040, ratio 3:1	15°	10°	+1°
IRC_1052, ratio 3:1	20°	10°	+2°

Nous attirons l'attention du lecteur sur le fait que ces résultats doivent tout de même être pris avec quelques précautions. En effet, même si les haut-parleurs du système de diffusion Ambisonics du studio 4 ont été calibrés, ses performances en localisation n'ont pas fait l'objet d'une évaluation préalable. Les résultats présentés pour l'écart azimuthal permettent de montrer qu'il est bien possible de discerner deux têtes HRTF, mais on ne peut pas savoir si les valeurs comparables que l'on trouve pour le panoramique par différence d'intensité et la tête IRC_1040 sont vraiment liées au mode de diffusion, ou si elles sont imputables à un défaut de précision du pointeur.

Toutefois, des études ont été menées sur des systèmes ambisoniques similaires à celui utilisé dans le studio 4, indiquant que des systèmes Ambisonics d'ordre 4 fournissent des performances en localisation voisines de celle de la perception naturelle [Bertet 2009] ; en particulier, le système utilisé lors de ce test devrait donc être assez précis pour notre étude, mises à part les éventuelles difficultés de manipulation du pointeur 3D. A ce propos, 7 participants sur les 18 ont indiqué que les déplacements du pointeur audio n'étaient pas toujours assez précis.

Un des manques de ce test est la limitation de la couverture des azimuts sur les secteurs latéraux (à l'arrière, 105° au maximum). Il faudrait également prévoir une phase préalable de familiarisation, afin de minimiser l'influence du dispositif de pointage. Enfin, il serait intéressant de tester d'autres types de stimuli que la harpe, d'évaluer les performances en localisation de traitements transauraux bruts par rapport à des traitements impliquant une réduction d'amplitude, et enfin de mesurer la perception de l'élévation et de la proximité obtenue par le prototype. Ces conclusions appellent la réalisation d'une nouvelle campagne de tests lors de prochains travaux sur le sujet.

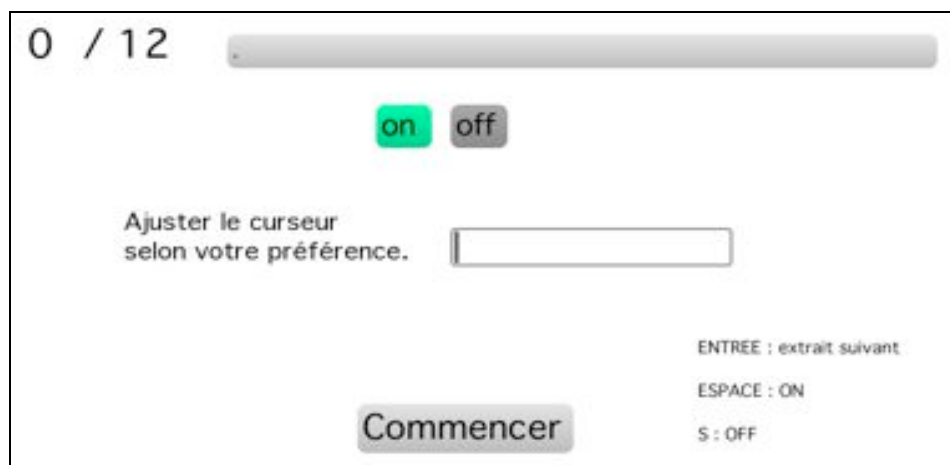
III.4. Test de balance

III.4.a. Protocole

On cherche maintenant à mettre un ingénieur du son en situation d'utilisation de l'interface. Nous voulions voir si les hypothèses formulées en II.5.b. sur l'asservissement du taux de *dry/wet* sont intéressantes.

Pour cela, nous avons confronté les utilisateurs du test à un ensemble de stimuli, et pour chacun d'entre eux, nous leur avons demandé de régler un curseur selon leur préférence subjective, en leur indiquant simplement que l'action du curseur était inconnue et susceptible de varier d'un stimulus au suivant. Pour faire abstraction d'effets indésirables liés à la visualisation du curseur, nous avons ajouté des plages minimales et maximales dans lesquelles les variations de position du curseur n'avaient pas d'incidence sur le son.

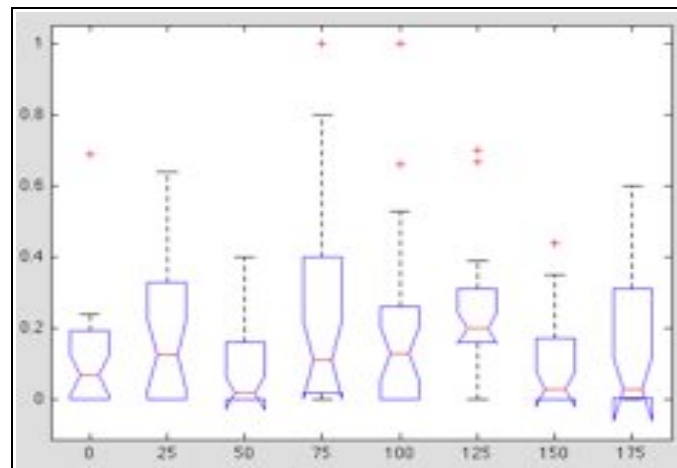
Les stimuli étaient spatialisés à différents azimuts (élévation nulle), et utilisaient uniquement les filtres associés à la tête IRC_1052 avec un ratio de 3:1, dans un souci de limitation de la durée du test.



28. Interface du test de balance

III.4.b. Résultats

Après un sous-échantillonnage des azimuts testés par plages de 25°, une analyse de la variance du *dry/wet* donné par les participants, menée sur les azimuts ($p = 0,04$) nous donne les résultats suivants :



29. Réponses *dry/wet* données par les participants au test 3, en fonction de l'azimut.

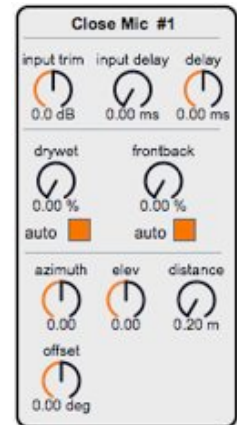
Cette courbe ne permet pas de confirmer l'hypothèse émise en II.5.b sur l'*auto dry/wet*. Ce résultat doit tout de même être nuancé ; en effet, les participants n'ont pas suivi de phase d'entraînement préliminaire aux tests. Il faut prendre en compte leur inexpérience relative en matière de techniques transaurales et en ce qui concerne le modèle combinatoire, mais aussi le fait que 4 participants sur les 18 ont indiqué avoir eu des difficultés avec les effets de phase générés par la synthèse transaurale.

Enfin, ce test doit être reproduit en utilisant d'autres têtes HRTF, la tête IRC_1052 évaluée ici ayant reçu de moins bons résultats que IRC_1040 dans le deuxième test.

IV. Conclusions

IV.1. Résultats

Durant ce stage, l'implémentation fonctionnelle du modèle tel qu'il a été décrit par Jean-Marc Lyzwa a été menée à terme. L'équipe de Flux::, qui propose des plug-ins de traitements audionumériques en collaboration avec l'Ircam, s'est montrée très intéressée par le prototype que nous avons développé au cours de ce travail. A l'heure actuelle, le prototype Max/MSP a déjà donné lieu à une intégration dans les bibliothèques du *Spatialisateur* de l'Ircam, dans lesquelles on peut retrouver l'interface présentée ci-contre.



30. L'interface de l'objet `spat.transpan` porté dans le *Spatialisateur*

L'utilisation du prototype en situation de production montre que le recours à la synthèse transaurale permet une amélioration nette des performances du 5.1, en particulier en ce qui concerne la localisation des sources latérales. Il serait intéressant de parvenir à une première explication objective des mécanismes perceptifs mis en jeu dans le modèle combinatoire utilisé, afin de mieux comprendre l'enjeu des différents paramètres qui sont apparus dans la conception du prototype ; ceci conditionnerait en particulier le choix de k pour l'*auto dry/wet* et pour l'*auto front/back*.

Les premiers résultats des tests réalisés au cours du stage tendent à confirmer les hypothèses formulées par Clément Cornuau dans son étude théorique ; les traitements appliqués à la couche de spatialisation par panoramique transaural permettraient bien une amélioration significative du timbre de cette dernière.

Toutefois, de nombreuses possibilités du prototype sont encore à évaluer, notamment en ce qui concerne l'élévation et la proximité, et la localisation en général. Il serait également intéressant de mettre à l'épreuve la robustesse de notre système vis à vis de systèmes de compression avec perte. D'autre part, une nouvelle campagne de tests plus complète permettrait de caractériser l'influence de la balance en temps entre les deux couches de spatialisation, ainsi que celle du *ratio* de la réduction d'amplitude.

IV.2. Perspectives

Dans le contexte de l'adaptation du prototype par Flux::, une étude sur l'interface de post-production définitive doit encore être menée, notamment en phase de bêta-test, afin de valider les paramètres retenus et leurs plages de variations, mais aussi de déterminer quels paramètres doivent être effectivement réglables par l'utilisateur.

De nouvelles réflexions sur l'interface nous ont poussé à réduire encore davantage le nombre de paramètres accessibles à l'utilisateur, ce afin d'améliorer la lisibilité du produit dans la gamme déjà existante de Flux::.

Le prototype pourrait aussi être complété dans deux directions :

- (1) la possibilité d'ajuster les positions des enceintes pour élargir la compatibilité du prototype avec davantage de systèmes d'écoute.
- (2) la possibilité d'inclure ses propres HRTF et de générer les filtres transauraux.

Ces deux options supposent l'intégration d'une bibliothèque à même de générer les filtres transauraux, travail qui était jusqu'alors effectué sous Matlab, et qui nécessite un suivi concernant l'adaptation du format de fichier des HRTF. Il faut aussi souligner que le fait de disposer de ses propres HRTF n'est pas forcément un atout majeur dans l'utilisation du prototype, pour la raison suivante : ce logiciel vise à être utilisé avant tout par des professionnels de la post-production, qui cherchent à atteindre tous types de publics. C'est pourquoi le recours à des HRTF individualisées ne semble pas indispensable ; dans cette mesure, on pourrait envisager le recours à d'autres têtes HRTF que celles mesurées sur des sujets lors des campagnes menées par l'Ircam, notamment la tête artificielle KEMAR.

D'autre part, il serait intéressant de mener une réflexion sur l'intégration d'un algorithme de calage en temps et en espace des appoints microphoniques par rapport au système principal. Cet aspect a été abordé par Alexis Baskind, mais reste à l'heure actuelle au stade expérimental.

Le prototype a déjà été utilisé dans un contexte de post-production professionnelle par Jean-Marc Lyzwa dans l'environnement Max/MSP relié à une console numérique SSL C300 HD en régie Schaeffer au Conservatoire de Paris. Des développements quant à un pilotage du plug-in à travers l'interface des consoles multicanales seraient un véritable atout dans l'ergonomie d'utilisation du plug-in. Il serait possible de piloter l'azimut et le *dry/wet* au moyen des deux encodeurs *Free Assign Knobs* de chaque tranche, afin de laisser à l'opérateur la possibilité d'écouter les modifications de spatialisation depuis la console plutôt que depuis la station de travail, qui se trouve être souvent excentrée dans les régies de post-production, et donc peu accessible depuis la zone d'écoute.

Annexe

Nous proposons ici une liste non exhaustive des productions du Service Audiovisuel du Conservatoire de Paris dans lesquelles Jean-Marc Lyzwa a utilisé en régie Schaeffer, avec une console numérique SSL C300 HD et le prototype implémenté dans le cadre de cette étude.

Suyus, Juan Arroyo (Journées de la composition 2011, CNSMDP)

Concert de la classe de direction de chœur : Oeuvres de Bach, Mendelssohn, Brahms, Reger (Notre-Dame de Paris, novembre 2011)

Oeuvres mixtes du compositeur Luis Naon : réductions stéréophoniques pour un CD
La Muse en Circuit

Le Roi David (Notre-Dame de Paris, 2012)

Dialogue de l'ombre double et *Notations*, Boulez ; *La Mer*, Debussy. (Cité de la musique, janvier 2012)

Jazz symphonique 2012 (Grand Plateau d'Orchestre, CNSMDP, mars 2012)

Bibliographie

[Bertet 2009]

S. Bertet. *Formats audio 3D hiérarchiques : Caractérisation objective et perceptive des systèmes ambisonics d'ordres supérieurs*. Thèse de doctorat. Institut National des Sciences Appliquées de Lyon. 2009.

[Blauert 1983]

J. Blauert. *Spatial Hearing : The Psychophysics of Human Sound Localization*. MIT Press, Cambridge, MA. 1983.

[Bauck 1989]

J.L. Bauck and D.H. Cooper. Prospects for transaural recording, *J. Audio Eng. Soc.*, vol. 37, no. 1/2, pp. 3-19. 1989.

[Cornuau 2011]

C. Cornuau. *Etude et optimisation de la synthèse transaurale à deux canaux*. Mémoire de fin d'études. Formation supérieure aux métiers du son, Paris. 2011.

[Courribet 2005]

B. Courribet and R. Béguin. *Outils de panoramisation pour l'environnement 5.1 dans Max/MSP*. Journées de l'informatique musicale. 2005.

[Craven 2003]

P. G. Craven. *Continuous Surround Panning for 5-speaker reproduction*. AES 24th International Conference on Multichannel Audio. 2003.

[Deschamps 2003]

M. Deschamps. *Restitution de l'enveloppement en 5.1 : Optimisation du système « Hamasaki Square »*. Mémoire de fin d'études. Formation supérieure aux métiers du son, Paris. 2003.

[Gardner 1997]

W. Gardner. *3-D Audio Using Loudspeakers*. Thèse de doctorat. Massachusetts Institute of Technology. 1997.

[Gerzon 1977]

M. Gerzon. Criteria for evaluating surround-sound systems. *J. Audio Eng. Soc.*, 25(6) : 400-409. 1977.

[Glasgal 2003]

R. Glasgal. *Surround Ambiphonic Recording and Reproduction*. 24th International Conference on Multichannel Audio. 2003.

[Larcher 2001]

V. Larcher. *Techniques de spatialisation des sons pour la réalité virtuelle*. Thèse de doctorat. Université Paris 6. 2001.

[Lund 2000]

T. Lund. *Enhanced Localization in 5.1 Production*. AES 109th Convention, Los Angeles. 2000.

[Lyzwa 2000]

J. M. Lyzwa and J. C. Messonnier. *Utilisation d'une rampe microphonique circulaire pour la prise de son et la post-production de sources sonores réparties sur 360° et destinées à un système de restitution multicanal 5.1*. Conservatoire de Paris. 2000.

[Lyzwa 2003]

J. M. Lyzwa. *Prise de son et restitution en multicanal 5.1. Problématique d'une oeuvre spatialisée : Répons*, Pierre Boulez. Conservatoire de Paris. 2003.

[Lyzwa 2004]

J. M. Lyzwa, A. Baskind and X. Meunier. *Utilisation de techniques binaurales et transaurales en production multicanal 5.1*. Conservatoire de Paris, Ircam, Ecole Nationale Supérieure Louis-Lumière. 2004.

[Lyzwa 2006]

J. M. Lyzwa. *Réflexion sur la prise de son et la post-production en multicanal 5.1*. Conservatoire de Paris. 2006.

[Lyzwa 2009]

J. M. Lyzwa and A. Baskind, *Utilisation de techniques binaurales et transaurales en prises de son et en post-productions multicanales 5.1*. AES 3rd Latin American Conference, Sao Paulo, Brazil. 2009.

[Pulkki 2005]

V. Pulkki and T. Hirvonen. Localization of virtual sources in multichannel audio reproduction. *IEEE Transactions on Speech and Audio Processing*, 13(1) : 105–119. 2005.

[Schroeder 1963]

M. R. Schroeder and B.S. Atal. Computer simulation of sound transmission in rooms. *Proceedings of the IEEE*, 51(3) : 536–537. 1963.

[Signoret 2011]

P. A. Signoret. *Le son multicanal*. Conservatoire de Paris. Formation Supérieure aux Métiers du Son. 2011.

[Steiglitz 1965]

K. Steiglitz and L. E. McBride, A technique for the identification of linear systems. *IEEE Trans. Automat. Contr.*, AC-10 : 461–464. 1965.

[Theile 1977]

G. Theile and G. Plenge. Localization of lateral phantom sources. *J. Audio Eng. Soc.*, 25(4) : 196-200. 1977.

[Ward 1999]

D. P. Ward and G. W. Elko, Effect of Loudspeaker Position on the Robustness of Acoustic Crosstalk Cancellation. *IEEE Signal Processing Letters*, Vol. 6, No. 5. 1999.