École Nationale Supérieure Louis Lumière
Section Son
promotion 2013

Mémoire de fin d'étude

L'insertion d'un outil de spatialisation binaurale dans le flux de post-production et de diffusion sonores

Pierre Bézard

Mémoire soutenu le 18 juin 2013,
devant un jury constitué de

Jean Rouchouse (directeur interne)
Matthieu Aussal (directeur externe)
Alan Blum (rapporteur)
Laurent Millot (rapporteur)

bezardp@free.fr

Remerciements

Comme tous les mémoires d'étudiants, celui-ci doit son existence à tant de participants que je ne pourrais les remercier tous à leur juste valeur. Toutefois, j'aimerais remercier avant tout grandement Matthieu Aussal pour son investissement constant, ses conseils, ses encouragements, ses bonnes idées, son énergie, sans lesquelles ce mémoire aurait été bien plus chétif.

Merci à Jean Rouchouse, mon directeur interne, pour ses patientes relectures et corrections et son écoute.

Merci à toute l'équipe de DMS qui a su si bien m'accueillir et me fournir un stage d'une telle qualité, et plus particulièrement, merci à Mathieu Coïc pour son aide inlassable, à Alexandre Dazzoni, Pascal Chédeville, Walid Lini, et bien entendu Hervé Roux.

Merci à Brian F.G. Katz et Tifanie Bouchara, du LIMSI-CNRS, pour leur expertise en matière de tests subjectifs et leur aide tout au long de ce travail.

Merci aux ingénieurs du son de Radio France : Hervé Déjardin, Frédéric Changenet, Edwige Roncière, pour leur assistance, leur gentillesse et leur disponibilité, et merci à Jean-Christophe Messonnier et Jean-Marc Lyzwa pour leurs précieux conseils sur le multicanal.

Merci à tous mes cobayes, dont j'espère qu'ils ont profité de leurs chocolats.

Merci à Alan Blum, mon rapporteur, et même à Gérard Pelé, deux hommes à qui, par une longue histoire, je dois ce sujet de mémoire. Merci également à Laurent Millot d'avoir été mon second rapporteur sur ce travail.

Un chaleureux merci à Baptiste, mon relecteur et ami et à son groupe Les Quenelles de Requins, pour m'avoir gracieusement prêté l'un de leurs enregistrements.

Merci enfin et surtout à ceux qui m'ont soutenu lors de cette longue aventure et grâce à qui je suis arrivé là, ma famille sur qui j'ai toujours pu compter, mes camarades de classe qui ont rendu cette année si savoureuse, mes nombreux amis, et mon koala, Cloé Chope.

Résumé

Connues depuis déjà longtemps, les technologies du binaural ont profité lors de ces dernières années d'un nouvel essor, lié notamment aux progrès ostensibles de la recherche, et à un regain d'intérêt du monde de l'audiovisuel, éveillé par les possibilités nouvelles offertes par cet outil. La mise en chantier de produits concrets comme le BP84, le développement de logiciels binauraux et la mise en ligne de fichiers binauraux dans le cadre du projet NouvOson de Radio France, témoignent de cette percée du binaural sur le marché. Or cette technologie, qui permet potentiellement de restituer un espace sonore 3D au moyen d'un simple casque, présente encore plusieurs problématiques, dues non seulement aux limites actuelles de la recherche et du développement, mais aussi aux difficultés liées à la manière dont nous localisons les sons et appréhendons l'espace sonore.

L'objectif de ce mémoire est de faire un bilan des limitations actuelles du binaural, par l'emploi du logiciel de spatialisation binaurale SpherAudio, et avec en filigrane les problèmes soulevés par l'utilisation du binaural comme outil de mixage descendant dans le projet NouvOson : on tentera ainsi de déterminer dans quelle mesure le binaural permet de retranscrire l'espace sonore d'un système multicanal, ici le système 5.0, et quels sont les dommages causés au fichier binaural par la compression de données, avec l'exemple de l'AAC 192 kbps, utilisé par NouvOson, et du MP3 192 kbps. Un descriptif de notre logiciel de travail permettra d'aborder la méthodologie liée à l'emploi du binaural dans un contexte de réduction d'un mixage 5.0. Un test subjectif sera ensuite mené pour apporter des éléments de réponse à nos problématiques de départ, et dont les résultats seront mis en relation avec les limites technologiques du binaural et les besoins de l'industrie.

Mots-Clefs :

5.1, Binaural, HRTF, Localisation, Multicanal, Spatialisation, Réduction de débit.

Abstract

Binaural technologies, which have been known for some years now, have recently undergone a transformation, due to patent development in research combined with a renewed interest from the audiovisual sector in the possibilities offered by this tool. The beginning of work on concrete products such as BP84, the development of binaural software and the on-line release of binaural audio files as part of the NouvOson project, testify to the breakthrough of binaural technologies on the market. This technology, which theoretically allows the reproduction of 3D sound by the mean of simple headphones, still faces some issues, however. These are not only due to the actual limits of research and development, but also to the difficulties linked with the way we localize sound and approach our sound environment.

The objective of this thesis, is to make an assessment of the actual limits of binaural technologies, through the use of binaural spatialisation software SpherAudio. We will keep in mind the issues that are raised by the use of binaural as a downmixing tool, in the context of the NouvOson project. As a consequence we will try to determine to what extent binaural allows restitution of the sound space produced by a multichannel system, such as 5.0, and what damages are caused to binaural sound files by the use of AAC 192 kbps (used in NouvOson) and MP3 192 kbps data compression. A description of the software we used will allow us to talk about the methodology linked with the binaural downmixing of 5.0 sound. A subjective test will then be carried out so as to suggest the answers to our original questions, and our results will be linked with the binaural technological limits of today, and the needs of the industry.

Keywords :

5.1, Binaural, HRTF, Localization, Multichannel surround sound, Spatial audio, Data reduction.

Table des matières

Chapitre 1
Partie théorique

1.1 La localisation sonore

La localisation sonore est un domaine d'étude dans lequel se révèle toute la complexité des capacités de notre audition, mêlant physique, physiologie de l'oreille et psychoacoustique. Le principe de la technologie binaurale, de même que ses limites, y sont étroitement liées. Ce mémoire commencera donc par un récapitulatif de l'essentiel des connaissances actuelles, en ce qui concerne la localisation sonore tout d'abord, et le binaural, ensuite.

1.1.1 ITD et ILD

La base de la localisation sonore chez l'être humain se fonde sur les différences interaurales de temps et de niveau, ce que les anglophones nomment respectivement « *interaural time* (**ITD**) *and level* (**ILD**) *differences* » (on rencontre parfois aussi l'IID, « *interaural intensity difference* »). Ces différences sont dues à la distance entre les deux oreilles (distance interaurale, d'une vingtaine de centimètres en moyenne chez l'être humain), et au volume de la tête, qui provoque rebonds, contournement, heurts sur le trajet du son et joue sur son retard en même temps que sur son intensité. Or par l'**effet de précédence**, le cerveau a tendance à localiser la source sonore du côté de l'oreille qui l'a perçue en premier. Par ailleurs, le son est plutôt localisé du côté de l'oreille qui l'a perçu au plus fort niveau (voir à ce sujet les courbes d'Henri Mertens, établies pour des signaux parlés, reproduites notamment dans [9], p.76). Ces ITD et ILD jouent donc un rôle important dans notre localisation du son.

Ainsi que le montre la figure 1.1 (d'après [24]), le son, arrivant à la tête depuis une incidence non-nulle (ici un son provenant de la droite), atteint de fait l'oreille droite avant l'oreille gauche : à noter que ce retard est dû non seulement à l'écart entre les oreilles, mais aussi à l'obstacle formé par la tête, qui impose un contournement supplémentaire.

De même, la présence de la tête est un obstacle qui atténue le son provenant à l'oreille la plus éloignée (Figure 1.2). Ainsi sur le dessin, l'oreille gauche percevra-t-elle un son atténué par rapport à l'oreille droite.

Ces deux notions simples d'ITD et d'ILD, constitueraient d'après les études actuelles l'essentiel de la localisation azimutale chez l'homme. Cependant, en

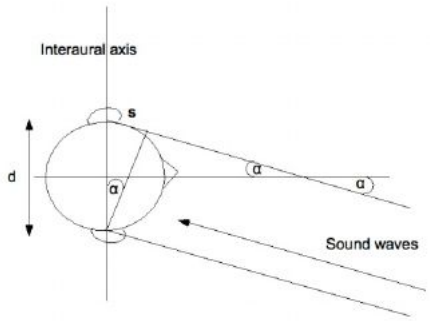


FIGURE 1.1 – Illustration de l'ITD, d'après [24]

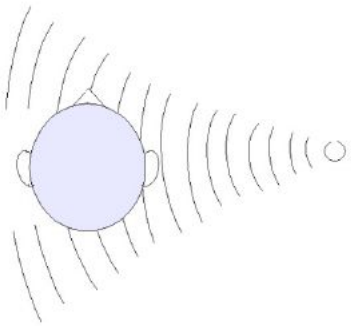


FIGURE 1.2 – Illustration de l'ILD, d'après [24]

l'état actuel des connaissances la précision de localisation azimutale n'est dans le meilleur des cas que de 3 ou 4 degrés, et varie selon la provenance du son. La figure 1.3 indique une estimation de ce que Blauert appelle **flou de localisation** (« *localisation blur* ») [3], défini comme la modification minimale des différents paramètres d'un son (ici : sa position), pour que le sujet perçoive un changement de localisation. Cette figure, obtenue pour des sujets gardant la tête fixe, apporte quelques informations supplémentaires : où l'on voit notamment que le flou de localisation est minimal à l'avant : entre 3 et 4 degrés, un peu plus important à l'arrière (entre 5 et 6 degrés), plus important encore sur les côtés (10 degrés). Ces chiffres mettent en exergue plusieurs éléments intéressants, notamment les difficultés de l'homme à localiser précisément les sons venant des côtés (90 et 270 degrés d'azimut).

1.1.2 Les HRTF

Les ITD et ILD ne sont pas suffisants pour localiser la provenance du son. Prenons le cas de deux sources sonores, S1 et S2, situées en deux endroits différents mais tels que ITD(S1) = ITD(S2) et ILD(S1) = ILD(S2). Ce cas est évoqué dans la figure 1.4. Si la localisation humaine était uniquement basée sur les différences de temps et de niveau, un auditeur ne serait pas capable, à tête fixe, d'identifier si le son vient de derrière (S1) ou devant lui (S2). Pareillement, il serait bien incapable de juger de l'élévation d'un son. C'est alors qu'entrent en jeu les **HRTF**.

Des déformations du son liées à notre tête

La Head-Related Transfer Fuction ou « fonction de transfert relative à la tête » , communément appelée HRTF, regroupe l'ensemble des déformations subies par le son sur son trajet jusqu'à notre canal auditif, et principalement dues à notre tête et notre pavillon. Elle comprend donc, outre le filtrage proprement dit, l'ITD et l'ILD. En chemin jusqu'à nos oreilles, le son issu de la source S2 dans la figure 1.4 ne va certes pas connaître de différence ILD ou ITD, mais il

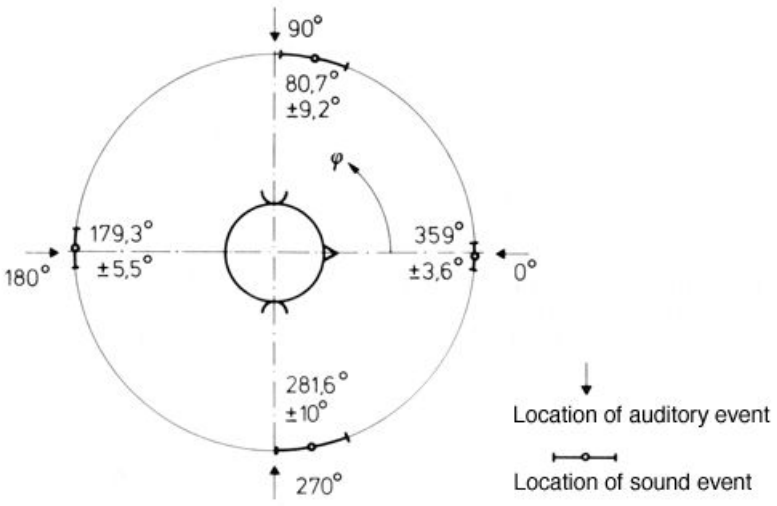


FIGURE 1.3 – Quelques mesures du flou de localisation azimutale chez l'homme. Attention au vocabulaire, que l'on doit à Jens Blauert [3] « ***Sound event*** » désigne l'origine physique du son (le chat qui miaule, devant l'auditeur) ; « ***Auditory event*** » désigne le son tel que l'auditeur le perçoit et le localise (le miaulement qui est entendu, indépendamment de la position du chat). (D'après Preibisch-Effenberger 1966 et Haustein et Schirmer 1970 ; 600-900 sujets, impulsions de bruit blanc d'une durée de 100 ms, à environ 70 phones, tête immobilisée, résultats reproduits dans [3], p.41 et [9], p.62, également sur [1]).

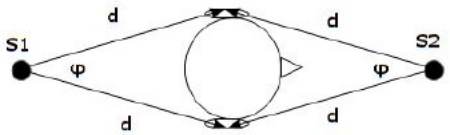


FIGURE 1.4 – Cas où ITD et ILD sont identiques (ici : égaux à 0) pour S1 et S2 : les distances des deux sources sont égales par rapport aux deux oreilles.

va se heurter à notre visage, rebondir, voir ses fréquences déphasées, en partie absorbées ou réfléchies, donc filtrées par les différentes matières, peau, cheveux, os, qu'il rencontrera, il sera dégradé par la forme de notre pavillon, avant, enfin, d'atteindre notre canal auditif. Le son issu de S1 lui, subira une déformation différente, en rencontrant non pas notre visage mais notre nuque, et en atteignant nos oreilles par l'arrière du pavillon. Ces sons ont donc été altérés suivant un filtrage dépendant de leur direction, ce qu'on appelle les **DDF** (*direction dependant filters*), et la différence de dégradation selon qu'ils sont parvenus aux conduits auditifs par l'avant ou par l'arrière, permet au cerveau de les situer. De même, c'est par les DDF que l'être humain peut percevoir l'élévation d'un son, les ITD et ILD restant théoriquement semblables en cas de faible variation d'élévation de la source. La combinaison des DDF des deux oreilles, de l'ITD et de l'ILD pour une direction donnée du son, a pour résultante la **HRTF** relative

à cette direction.

Les DDF apportent donc plusieurs éléments à la localisation d'un son :

- la discrimination avant-arrière
- la perception de l'élévation.

Individualité des HRTF et importance de l'apprentissage

La principale caractéristique de ces HRTF, c'est qu'elles sont **propres** à la morphologie de chaque être humain. C'est la forme de ma tête qui est la cause de mes HRTF, elles ne seront par conséquent pas les mêmes que celles de mon voisin. J'ai appris à les reconnaître depuis tout petit, elles sont donc le fruit d'un **apprentissage** de longue haleine, qui se poursuivra jusqu'à la fin de ma vie. C'est en accumulant les expériences et en découvrant les caractéristiques de mon corps que j'ai pu identifier, mémoriser, comprendre d'où venaient les sons selon les déformations qu'ils présentaient en entrant dans mon oreille. On peut donc supposer que la localisation chez les bébés est moins bonne que celle de l'adulte ; ils la perfectionneront en apprenant à se connaître. Des expériences consistant à court-circuiter l'oreille externe d'auditeurs en insérant des tubes directement dans leurs conduits auditifs, ont montré que ces sujets avaient beaucoup plus de mal à localiser les sons qu'auparavant : on les avait privés de leurs HRTF (Kietz 1952, Tarnoczy 1958, *in* Blauert [3], p.100).

Il est à noter que le terme de HRIR, « Head-Related Impulse Response » (« réponse impulsionnelle relative à la tête »), est parfois également employé : les HRIR sont la version temporelle des HRTF. C'est-à-dire qu'en appliquant une transformée de Fourier à une HRIR, on obtient une HRTF, et en appliquant à cette dernière une transformée de Fourier inverse on retrouve notre HRIR de départ. Les données relevées lors d'un test d'impulsion sur tête artificielle seront ainsi, en toute rigueur, ses HRIR, mais l'analyse fréquentielle se fera sur les HRTF correspondants.

L'apprentissage, le familier ont un rôle essentiel dans la localisation sonore : outre l'accoutumance qu'a une personne pour ses HRTF, le caractère familier des sons perçus aurait une incidence. Ainsi, des signaux connus par le cerveau seraient plus facilement localisés (voir Plenge et Brunschen 1971, 10-20 sujets, voix parlée, *in* [3]). C'est logique : si quelqu'un connaît par coeur le klaxon de la voiture de son voisin, et par conséquent son contenu fréquentiel, il saura jauger immédiatement de quelle manière le spectre en a été altéré par ses HRTF, et il en déduira sa provenance. S'il s'agit d'une klaxon qu'il ne connaît pas, il risque d'hésiter plus longuement.

L'individualité des HRTF est cependant pondérée par une **faculté d'adaptation** à des HRTF étrangères (voir Butler et Belenduik (1977), Morimoto et Ando (1977,1982), *in* Blauert, *op. cit.*, p.312) : l'expérience montre qu'après une période d'adaptation, des sujets sont capables de se repérer dans l'espace en utilisant des HRTF étrangères (voir aussi [17]). Une faculté d'adaptation existe donc. Celle-ci n'est pas contradictoire avec l'individualité des HRTF, au contraire : si l'homme est capable de s'adapter constamment à des HRTF qui varient avec les changements de son corps, pourquoi ne pourrait-il s'adapter à celles d'une autre

personne, tant qu'elles restent proches des siennes ? Ce résultat important aura de grandes conséquences dans l'élaboration des technologies binaurales.

La sommation des HRTF pour aiguiser la spatialisation

Une HRTF seule, une ITD et une ILD ne sont pas forcément suffisants pour localiser le son : cela signifie que, si la tête est gardée fixe, la localisation précise d'un son pourra sembler difficile. En revanche, si la tête bouge, les informations données par les HRTF sont accumulées. Le sentiment de localisation est ainsi précisé (Thurlow et al. 1967, Klensch 1948, Jongkees et van de Veer 1958, Wallach 1938, 1939, 1940, *in* Blauert, *Ibid.*, p.191). Les mouvements de tête apparaissent ainsi essentiels à la localisation des sons : ils réduisent considérablement le cône de confusion. C'est pour cette raison que, lorsque survient un son dont la provenance intrigue, le premier réflexe est de bouger la tête : le sentiment de localisation est ainsi précisé.

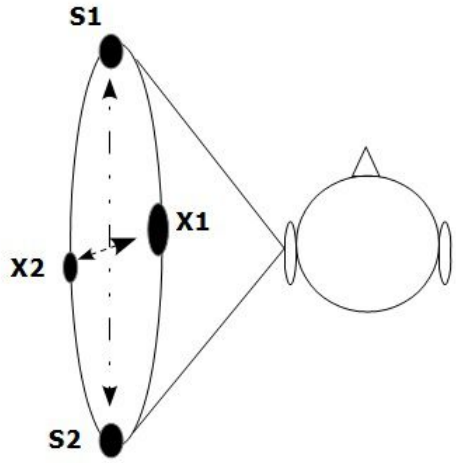


FIGURE 1.5 – Le « cône de confusion » . Pour tout évènement sonore survenant à la surface du cône, et pour un sujet gardant la tête fixe, il existera une ambigüité haut-bas (X1, X2) et/ou avant-arrière (S1, S2), sur la localisation de ce son.

Malgré l'accumulation d'informations dues aux HRTF par mouvements de tête, la précision de localisation de l'oreille humaine en élévation reste assez mauvaise. Blauert [3] évoque 10 degrés de précision au mieux dans le plan médian pour de la voix, 13 à 22 degrés au-dessus de la tête, 15 degrés derrière la tête (voir Damaske et Wagener, 1969, 7 sujets, 65 phones, tête immobilisée, *in* Blauert, *Ibid.* ; voir aussi [11]). La localisation en élévation reste donc assez imprécise au naturel, à tel point qu'elle se rattache principalement au contexte : par exemple, un son d'oiseau sera perçu par réflexe avec une élévation positive, même si l'oiseau se trouve en réalité au sol (élévation négative ou nulle), tant l'habitude de l'auditeur le pousse à penser que l'oiseau se trouve au-dessus de

lui. Cette influence du contexte, et ce flou lié à l'élévation, ne s'avéreront pas anodins lorsqu'il s'agira de restituer une impression d'espace en binaural.

1.1.3 Une localisation qui varie avec la fréquence

Le sentiment de localisation enfin est dépendant de la fréquence du son : les études montrent que selon le contenu fréquentiel de ce qui est donné à entendre, l'oreille ne localise pas de la même manière.

Structure fine et enveloppe

Blauert (*op.cit.* p. 164) et Begault ([2] p.33) font référence à une perception différente de l'ITD selon la fréquence du son : si le son est suffisamment grave, sa longueur d'onde est plus grande que la distance interaurale, et le cerveau est donc capable de comparer les fronts montants du signal qui arrivent aux deux oreilles, pour déterminer à quelle oreille il est arrivé en premier (figure 1.6, sinusoïdes A et B). Le cerveau peut donc analyser la « **structure fine** » *(fine structure)* du signal.

Mais si le son est d'une fréquence plus élevée, sa longueur d'onde finira par devenir plus petite que la distance interaurale (ce qui se produit au-delà de **1600 Hz** environ). Dès lors, le cerveau n'est plus en mesure de déterminer, entre des fronts montants très rapprochés les uns des autres, lequel arrive avant l'autre (figure 1.6, sinusoïdes D et E : D est « en avance » par rapport à E mais si l'on se concentre sur les fronts montants E et F, on pourrait croire le contraire). Le cerveau n'analyse plus la structure fine du signal, mais bien son **enveloppe** : à la moindre variation d'amplitude du signal, ce sont les fronts montants de l'enveloppe qu'il va comparer pour déterminer quel signal arrive en premier (figure1.6, enveloppes X et Y). La détermination de la différence de temps s'opère donc de façon différente selon la fréquence du signal.

Par ailleurs, à l'écoute d'un signal complexe, l'homme est capable de l'analyser par bandes de fréquences, chacune étudiée soit en structure fine soit par son enveloppe, avant de conclure quant à sa localisation. Cela peut provoquer une localisation différente des parties basse et aiguë du signal (Deatherage 1961, Sakai et Inoue 1968, Boerger 1965a, Schubert et Wernick 1969, *in* Blauert, *op. cit.*, résultats reproduits également dans [9], p.69).

Ces études expliquent aussi que l'être humain a plus de facilité à localiser les signaux comprenant beaucoup de transitoires, des impulsions (percussions, touches de piano), dans lesquelles sont présents des repères réguliers d'enveloppes et de fronts montants, plutôt que des sinusoïdes ou des notes tenues.

Feddersen *et al.* (1957) (*in* Blauert, *op. cit.*, p.156) évoque aussi une meilleure précision de localisation en ILD pure pour les fréquences autour de 2 kHz (ce qui correspond à des fréquences de la voix), que pour les fréquences supérieures et inférieures.

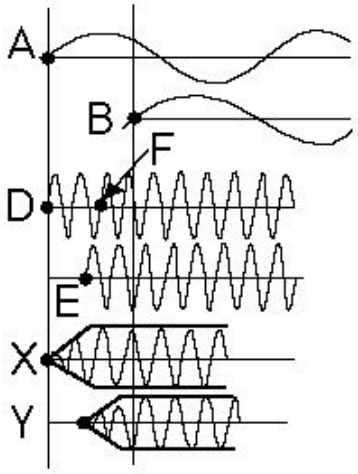


FIGURE 1.6 – Illustration de la perception de la structure fine ou de l'enveloppe du son, selon sa fréquence.

Les bandes directionnelles

Les expériences de Blauert lui-même (*op. cit.*, pp. 105-116, évoquées également dans [9], p.63), tendraient à montrer que le cerveau a tendance à associer d'office différentes bandes de fréquences à différentes localisations. C'est ce que Blauert nomme les « bandes directionnelles » . Si on expose à un sujet, tête fixe, des sons diffusés devant lui, il aura tendance à les situer devant, derrière, ou au-dessus de lui (fig.1.7). Selon Blauert les mouvements de tête ont notamment pour effet d'annuler l'incidence de ces bandes directionnelles, dans la localisation du son.

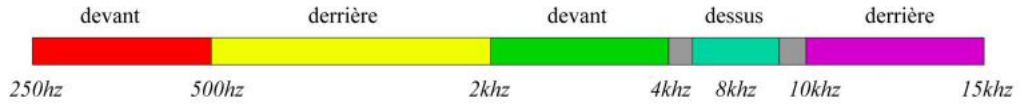


FIGURE 1.7 – Illustration des bandes directionnelles (d'après Blauert, 1968, 1969, 1969-70) (reproduits dans [14]).

1.1.4 Perception de la distance

Nous n'avons jusqu'ici pas évoqué la question de la distance. L'appréciation de cette dernière repose sur trois propriétés physiques du son :

- le niveau sonore ;
- le spectre du stimulus considéré ;
- l'effet de salle.

Ces trois éléments sont exploités au quotidien par les ingénieurs du son, qui savent que pour éloigner une source sonore dans un mixage, on peut au choix

baisser son niveau, filtrer ses aigus, augmenter la réverbération, ou combiner ces différents effets.

Les études réalisées jusqu'ici (par Pierce 1901, Arps et Klemm 1913, Schutt 1898, *in* Blauert, *op. cit.*, pp. 128-131) permettent de faire la distinction entre champ proche (distance inférieure à 1/6 de la longueur d'onde du son) et champ lointain (distance supérieure à 1/6 de la longueur d'onde). L'appréciation moyenne du champ proche et du champ lointain est sujette à débat : pour un son complexe, sont parfois évoquées comme valeurs moyennes : un champ proche à moins de 3 m, un champ intermédiaire de 3 à 15 m, un champ lointain à plus de 15 m, mais d'autres parlent d'un champ proche à moins de 1 m, et d'un champ lointain à plus de 10 m (voir Morse et Ingard 1968).

Plus précisément, en l'état actuel des connaissances :
- en champ proche, à moins de 25 cm, la perception de la distance se fait principalement par le niveau du son, et par l'effet de proximité, marqué par une bosse dans les fréquences graves (60, 80 Hz) ;
- à plus de 25 cm, elle se fait surtout par le niveau ;
- en champ lointain, la perception de la distance se fait par le niveau ;
- à plus de 10 m, elle se fait par le niveau et par la chute des hautes fréquences (au-delà de 5000, 6000 Hz).

On notera que l'influence de l'effet de salle n'est pas considérée ici.

Mais ces résultats sont à prendre avec précaution, car la perception de la distance suit également d'autres lois, et notamment, comme pour l'élévation, celles du contexte et du type de stimulus considérés. L'exemple de l'expérience de Gardner, faite en 1969, dont les résultats sont affichés dans la figure 1.8, apporte des éléments intéressants : dans le cas de la voix parlée, la perception de la distance est relativement correcte (à concurrence d'une distance comprise entre 1 et 4 m, en chambre anechoïque, donc une perception basée essentiellement sur le niveau de la source). Cependant, dès qu'il s'agit d'une voix murmurée ou criée, les résultats sont très différents : la distance de la voix murmurée est sous-estimée au-delà de 1 m, celle de la voix criée est d'emblée surestimée. La nature du stimulus joue sur la perception de la distance.

Les ingénieurs du son le savent : en champ libre on devrait normalement avoir une décroissance du niveau en 1/r lorsque la source s'éloigne. Lorsque la distance double, le gain chute donc de 6 dB. Mais pour simuler un doublement de la distance qui paraisse crédible à l'oreille, il est nécessaire de baisser le niveau de 20 dB ! (Gardner 1969, voix parlée, 5 sujets, chambre anechoïque, mais aussi von Békséy (1949) et Laws (1972). Petersen (1990) parle de 21 dB, Begault (1991) de 9 dB, *in* Blauert, *op. cit.*, p. 122). De même, pour une source dont le niveau varie mais qui reste fixe dans l'espace, il est logique qu'on ait une sensation de variation de la distance : une hausse du niveau de la source renforce les extrêmes graves et aigus perçus, donc donne la sensation d'un rapprochement ; la hausse de niveau donne la sensation d'une source plus proche ; et à l'inverse,

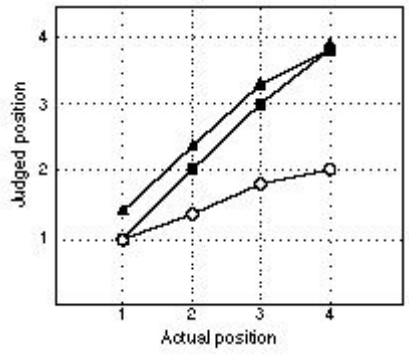


FIGURE 1.8 – Perception de la distance, en chambre anechoïque, pour des stimuli diffusés sur haut-parleur à 0 degré d'azimut. Cercles : voix murmurée ; carrés : voix parlée ; triangles : voix criée. (Gardner, 1969, *in* Begault [2], p. 77)

une source plus forte excite davantage l'acoustique de la salle, ce qui a pour effet une sensation d'éloignement. Il en résulte une perception de la distance qui peut varier constamment, alors même que la source n'a pas bougé.

Ce genre d'effet est particulièrement sensible à l'écoute d'enregistrements réalisés en stéréophonie de phase (couple AB ou ORTF) dans un lieu réverbérant, et où les sources peuvent donner la sensation de changer de plans sonores.

1.1.5 Influence de l'image

Les expériences dont les résultats sont relatés ici ont eu lieu dans des conditions où les sujets ne pouvaient distinguer l'origine des sons (sujets aveuglé par un bandeau, sources masquées par un tissu...). En effet, l'ajout de l'image apporte un vrai problème à la localisation sonore, car le visuel est dans tous les cas prioritaire. Comme la Mannschaft, l'image gagne à tous les coups : des études montrent que si la personne qui parle est à l'écran, quels que soient les effets employés pour éloigner le son (niveau, spectre, effet de salle), le spectateur situera obstinément la voix à la distance du personnage qu'il voit (Gardner 1968, Mershon, Desaulniers, Amerson, Kiefer 1980, *in* Blauert, *op. cit.*, pp. 193-196). L'effet est tout aussi prégnant sur l'azimut, et surtout l'élévation, ainsi que Michel Chion l'observait, à travers le phénomène d'**aimantation spatiale**, ou effet ventriloque ([6], pp. 221-223, 412) : si l'écran montre un chat miaulant depuis une branche d'arbre, en haut à gauche de l'écran, le miaulement sera perçu au même endroit que lui, même si le mixeur l'envoie dans un autre haut-parleur que le canal gauche, par exemple le haut-parleur central. Que le chat disparaisse de l'écran, et le miaulement basculera dans le haut-parleur central (voir Stratton 1887, Klemm 1918, Held 1955, *in* Blauert, *op. cit.*). Cet effet est même sensible dans certains cas si le miaulement est envoyé dans les enceintes arrière. On peut donc dire que, pour du son à l'image, la localisation de la source sonore par le son devient très secondaire par rapport à la localisation par le visuel. C'est la « théorie du visuel » *(visual theory)*, sur laquelle repose par ailleurs tout le succès du cinéma parlant.

1.2 Le binaural en théorie

Après ce récapitulatif des connaissances actuelles concernant la manière dont nous localisons les sons, il est temps d'aborder la théorie du binaural. Qu'entend-on par « binaural » ? Quels sont les ressorts et les effets permis par cette nouvelle technologie ?

En préambule : un peu de vocabulaire

Avant d'entamer cette seconde partie, quelques termes, pouvant porter à confusion, vont être passés en revue.

Binaural Dans notre étude, ce terme désigne la retranscription d'un espace sonore en trois dimensions (x, y, z ou distance, azimut, élévation), dans le cadre d'une écoute au casque. Ce terme s'oppose non seulement à « stéréo au casque » , aussi appelé « présentation dichotique » du son (*dichotic presentation*, cf. [3] p. 94), mais également à la retranscription d'un espace sonore en trois dimensions au moyen d'un système de hauts-parleurs : 9.1, 22.2, Dolby Atmos, etc.

Binauraliser Nous emploierons ce terme pour décrire l'étape au cours de laquelle le son est positionné dans l'espace binaural, et quitte ainsi son état d'origine : par exemple, on parlera de « binauraliser » un mixage 5.1, quand il s'agira de rentranscrire ce mixage au casque en conservant la sensation d'espace d'origine par le binaural : le canal central sera entendu devant nous, L à 30 degrés (à gauche), R à -30 degrés (à droite), Ls à 110 degrés, derrière notre nuque, et Rs à -110 degrés. Nous décrirons le plus souvent les angles utilisés dans un repère 0, 90, 180, -90 degrés, 0 étant devant, 90 à notre gauche, 180 derrière nous, -90 à notre droite. De même pour l'élévation, avec 0 devant nous, 90 au-dessus de nous, 180 derrière nous et -90 degrés en-dessous de nous.

Les HRTF Nous parlerons dans cette étude de HRTF plutôt que de HRIR. Nous avons fait le choix, dans notre rédaction, de donner à ce mot le genre féminin, décision qui peut prêter à discussion mais s'appuie sur la traduction littérale de l'acronyme HRTF : head-related transfer function (**une** fonction de transfert relative à la tête). Les textes employant cet acronyme avec le genre masculin, au demeurant nombreux, lui donnent bien entendu le même sens que nous.

Plans médian, horizontal, frontal Nous reprendrons ici, par commodité, le vocabulaire employé par Jens Blauert (voir figure 1.9) (cf. [3], p. 14) afin de découper l'environnement 3D de l'auditeur : le plan horizontal ou azimutal *(horizontal plane)* a en tout point une élévation nulle, c'est le plan (0,x,y) ; le plan médian recoupe l'axe avant-arrière (0 degré - 180 degrés) verticalement, c'est le plan (0,x,z) ; le plan frontal recoupe l'axe droite-gauche (-90 - 90 degrés) verticalement, c'est le plan (0,y,z). Ces termes nous permettront de décrire plus facilement certaines théories ou certains effets observés.

Son 3D, espace 3D Un terme souvent utilisé, notamment dans les présentations de ces nouvelles technologies, est celui de « son 3D » ou encore d'« espace 3D ». Ils désignent la capacité à positionner un son où l'on veut dans un espace en trois dimensions : (0, x, y, z), par opposition à la stéréophonie qui n'est censée permettre qu'un positionnement à une dimension (0, y). Le binaural, l'ambisonie d'ordre élevé, mais aussi les systèmes discrets type 9.1, 11.1, 22.2, et systèmes hybrides comme le Dolby Atmos, sont souvent présentés comme permettant de faire du « Son 3D ». Dans le cadre de notre étude ce terme nous paraît cependant restrictif. Le son et les ingénieurs du son n'ont pas attendu des systèmes à plus de deux canaux pour retranscrire un espace à au moins deux dimensions : l'azimut et la profondeur, par le biais des outils de mixage. Nul n'analyse le mixage d'un morceau de musique en se concentrant sur l'unique dimension (0, y) mais parle de profondeur, de plans sonores... Par ailleurs, même en monophonie le son se propage dans un espace à quatre dimensions : x, y, z, et le temps. Nous serons donc prudents avec le terme de « son 3D », que nous éviterons de trop utiliser.

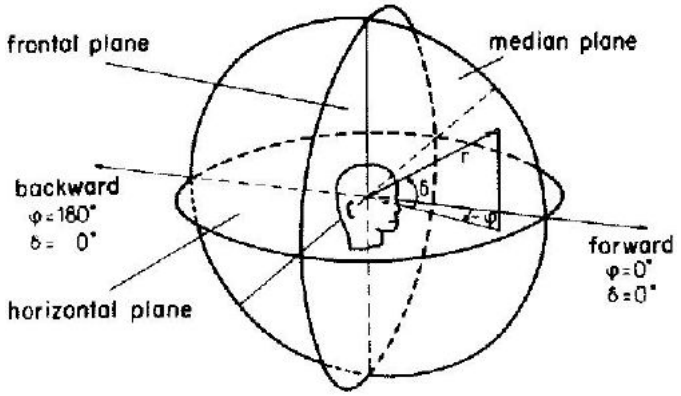


FIGURE 1.9 – Les plans médian, horizontal, frontal (Blauert) (reproduits dans [20]).

1.2.1 Principe du binaural

La convolution par des HRTF

C'est l'idée de base : puisque j'entends que le chat miaule derrière moi grâce à mes HRTF, alors si j'émule ces HRTF sur un enregistrement monophonique de miaulement de chat, et que je diffuse ce son altéré dans mon casque, alors je l'entendrai logiquement derrière moi. L'expérience a montré l'efficacité de ce principe (Blauert, *op. cit.*, p.306, voir aussi [9], pp.169-171). Chacun peut d'ailleurs de faire l'expérience de glisser des microphones-cravate omnidirectionnels dans ses oreilles, d'enregistrer la scène sonore autour de soi, en stéréophonie (une piste par microphone), et de diffuser ensuite cet enregistrement dans un casque, pour retrouver ses sensations d'espace, avant, arrière, en-haut, en-bas, de manière généralement convaincante.

La simulation des ITD et ILD n'est pas très difficile, car on l'exploite déjà en stéréophonie : monophonie dirigée, enregistrement stéréophonique d'intensité, de phase... Le plus difficile est donc l'émulation du filtrage par des HRTF. Deux solutions ont été trouvées à ce problème : la première est la solution du **binaural natif** : on place des microphones-cravates dans les conduits auditifs d'une tête (le plus souvent une tête artificielle), et on enregistre le son parvenant aux microphones, sur deux pistes séparées (voir figure 1.10). L'enregistrement qui en résulte est figé, les sons ne peuvent pas être déplacés individuellement dans l'espace, les possibilités de mixage sont réduites. Nous ne nous étendrons pas sur ce procédé, qui ne nous intéresse pas dans le cadre de notre étude.

La seconde option, consiste à enregistrer des sons en monophonie ou stéréophonie « classique » , puis de les convoluer par des HRTF (qui sont, sur le plan mathématique, des filtres comme les autres), et d'obtenir l'illusion d'un son spatialisé. Le gros avantage de cette technique, est qu'elle permet ensuite de mixer chaque son de façon indépendante, et indépendamment de la façon dont il a été enregistré (figure 1.11). Par ailleurs, le workflow traditionnel de production sonore n'est pas remis en cause, puisqu'on traite des sons enregistrés de façon traditionnelle, et qu'on obtient un simple fichier stéréo .wav ou .aiff que l'on peut écouter sur n'importe quel casque.

La question se pose alors de la manière de se procurer les HRTF dont on a besoin. Pour des contraintes matérielles évidentes, il n'est pas envisageable de relever les HRTF de chaque auditeur et de réaliser une convolution du mixage par tous les HRTF ainsi relevés. Il serait déjà très compliqué de relever les HRTF d'un seul sujet pour tous les points de l'espace tridimentionnel.

Un compromis consiste à réaliser, pour une série de positions déterminées, l'enregistrement des HRTF : on place des microphones-cravates dans les oreilles d'un sujet, ou on utilise une tête artificielle ; on diffuse un même son, pleine bande, de nature connue, en différents points d'un maillage spatial déterminé, et on analyse les HRTF obtenues via les signaux captés par les microphones. Il reste ensuite à réaliser une **interpolation** des HRTF obtenues, qui correspondent donc à quelques positions connues, afin d'obtenir une approximation des HRTF correspondant aux positions non-connues.

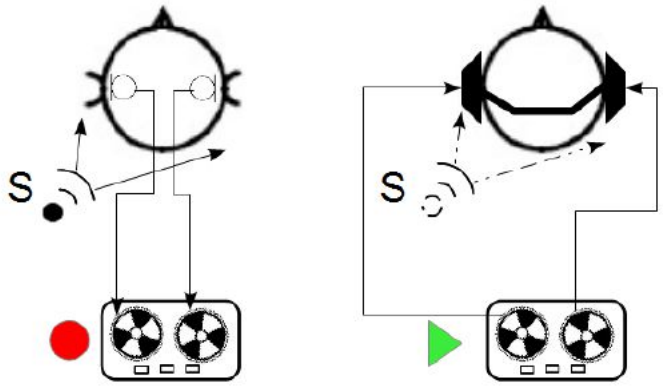


FIGURE 1.10 – Schéma de principe du binaural natif.

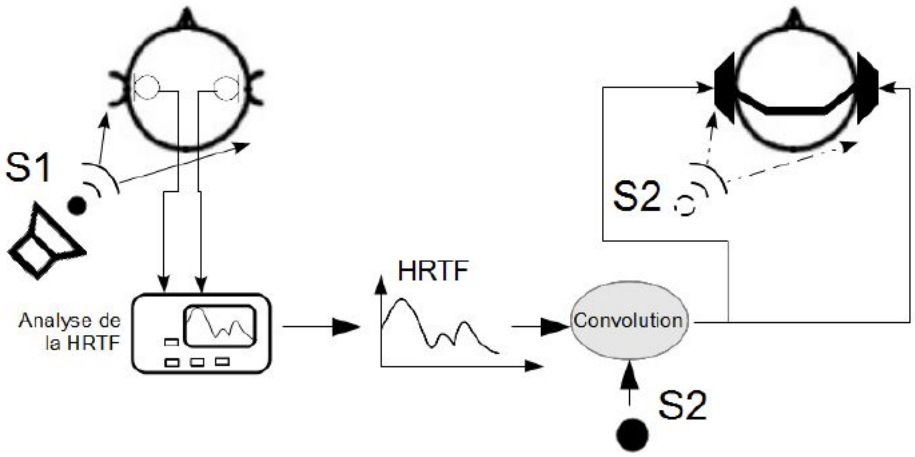


FIGURE 1.11 – Principe du binaural obtenu par traitement du signal : avec S1 on obtient les données relatives à la HRTF pour une position donnée ; en convoluant cette HRTF par un autre signal S2, on donnera l'impression à l'auditeur d'entendre S2 à cette même position.

Par exemple : on fait entendre à un auditeur équipé de microphones-cravates dans ses oreilles, un signal connu sur un cercle d'enceintes, disposées tout autour de lui, et espacées de 10 degrés d'azimut. On pourra ainsi relever les HRTF de ce sujet sur le plan horizontal, mais seulement pour 36 positions : à 0 degré, -10 degrés, -20 degrés, etc. Si l'on veut convoluer un signal audio pour donner l'illusion à l'écoute qu'il se situe à -10 degrés, en utilisant les HRTF du sujet, on dispose d'une mesure de cette HRTF. Mais si l'on veut donner l'illusion d'une position à -7 degrés d'azimut, on devra réaliser une approximation de la HRTF correspondante, à partir d'HRTF connues (figure 1.12). On pourra prendre alors, soit une combinaison des deux HRTF connues voisines, soit la HRTF la plus proche, soit encore les trois HRTF les plus proches, etc., selon des modalités dont dépendra le succès de la restitution obtenue. C'est cette interpolation qui a constitué longtemps un gros problème dans la recherche sur le binaural.

Comment choisir les HRTF

Nous l'avons dit, l'une des principales caractéristiques de nos HRTF, c'est qu'elle nous sont propres. Nous n'avons jusqu'ici mentionné que des cas où un même sujet servait à l'analyse des HRTF. Mais nous avons aussi expliqué qu'il n'était matériellement pas possible de prendre les HRTF de tous les auditeurs potentiels et de les convoluer par les sons du mixage. Alors, comment faire pour que les effets de spatialisation voulus en binaural fonctionnent pour un maximum d'auditeurs ?

Cette question est en fait celle de l'individualisation des HRTF, qui est une problématique complexe. L'optique qui a été favorisée jusqu'ici a été de rechercher un compromis, à savoir, chercher à constituer une HRTF « moyenne » qui fonctionnerait convenablement pour un maximum de personnes. Le moyennage

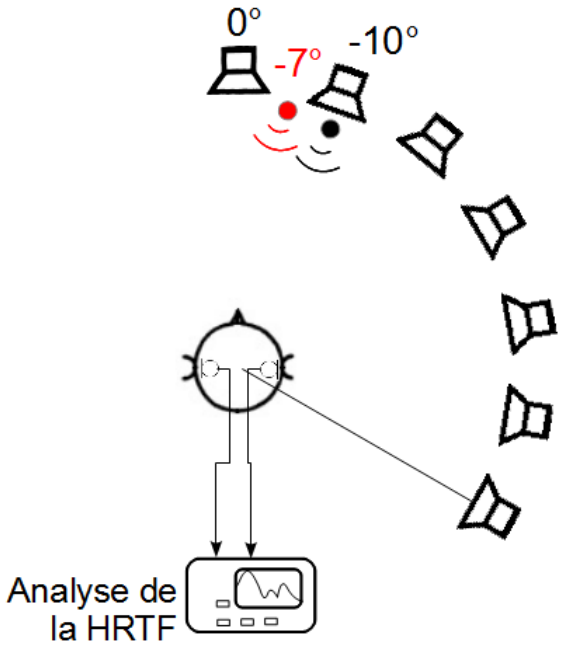


FIGURE 1.12 – Maillage d'enceintes pour le relevé de HRTF au moyen d'une tête artificielle. On voit que l'on relèvera ici les HRTF pour les positions 0 et -10 degrés, mais qu'on devra reconstituer la HRTF à -7 degrés à partir de HRTF connues pour d'autres positions (attention, les angles ne sont pas à l'échelle).

pur et simple des HRTF relevées sur un grand nombre de sujets, n'aurait pas grand sens, et conduirait à des aberrations (Mellert 1971, Mehrgardt and Mellert 1977, Platte 1979, *in* Blauert, *op. cit.*, pp. 290-295). On a parfois aussi employé le lissage d'une HRTF jugée satisfaisante pour un grand nombre de personnes. Enfin, on a parfois pensé que l'emploi des HRTF d'une tête artificielle de qualité serait suffisante. SpherAudio, le logiciel de spatialisation binaurale de Digital Media Solutions, que j'ai utilisé dans le cadre de mon mémoire, a recours de son côté à une dizaine d'HRTF, sélectionnés parmi 45 HRTF relevées en 2000 sur sujets réels dans le cadre du projet Listen, avec relevés morphologiques (cf. la page web [31]). Elles ont été choisies de telle façon que, pour chaque être humain sur Terre, au moins une d'entre elles fonctionne statistiquement de façon satisfaisante.

Malgré tout, ce procédé n'est qu'un palliatif à la problématique plus large de l'individualité des HRTF, qui reste sans doute un sujet crucial pour le développement du binaural et ses applications commerciales (voir une approche du problème dans [15]). Malgré notre faculté d'adaptation évoquée plus haut, qui nécessite une durée d'apprentissage encore peu définie (il pourrait suffire de quelques heures sur au moins trois séances d'entraînement, d'après [17]), des solutions sont activement recherchées dans l'optique d'une exploitation commerciale. Il est en effet peu probable que les consommateurs soient prêts à adopter un système qui ne fonctionne qu'après un temps d'adaptation.

La room

Techniquement, convoluer un son par le couple d'HRTF oreille gauche-oreille droite correspondant à la position qu'on veut lui donner (comprenant donc les ITD et ILD), devrait être suffisant pour le positionner correctement dans l'espace binaural. Mais c'est oublier le paramètre de l'acoustique du lieu d'écoute : dans la réalité, il est rare que l'on entende des sons en parfait champ libre. Même dans une salle de mixage, le son produit par les enceintes ne parvient à nos oreilles que réverbéré, heurté, déformé par les murs de la salle, qui n'est jamais vraiment anéchoïque. Or les HRTF du projet Listen ont été relevées en chambre anéchoïque, et ne prennent donc en compte que du son direct. Ce n'est pas anodin, car à se contenter d'une simple convolution, on s'aperçoit rapidement qu'on se heurte à deux problèmes : le premier étant l'**IHL**, « inside-the-head locatedness » (localisation intra-crânienne), soit l'impression de localiser le son à l'intérieur de sa tête, ce qui est considéré comme un effet indésirable. Le second, étant une imprécision de localisation pouvant conduire à des ambiguïtés haut-bas et avant-arrière. Ceci se produit même dans le cas de l'écoute en binaural d'un son qui a été réverbéré au mixage. L'idée est donc d'ajouter une réverbération technique, qu'on appelle la « **room** » . Il s'agit d'émuler la réverbération d'une salle d'écoute, appliquée au son.

Mais plusieurs contraintes demeurent : d'une part, sur le choix de la salle d'écoute, et donc, de l'algorithme. D'autre part, sur le rapport direct/réverbéré approprié, et sur le choix de le fixer ou de le laisser à la discrétion du mixeur. Nous reviendrons sur ces questions lorsque nous décrirons plus en détail le cas du logiciel SpherAudio.

Le head tracking

Il nous reste à exposer une dernière limite du binaural. Nous avions en effet évoqué plus haut l'importance, dans la localisation d'un son, des mouvements de tête, qui permettent de cumuler les HRTF. Or lors d'une écoute au casque, les mouvements de tête de l'auditeur n'ont évidemment pas d'influence sur le son qui arrive à ses oreilles. On perd donc potentiellement une précision importante dans la localisation des sons en binaural.

Une solution à ce problème est le *head tracking* : il s'agit de positionner un capteur sur le casque de l'auditeur, et de positionner un repère à un point de l'espace devant lui, de telle sorte que ce repère analyse en temps réel les mouvements de la tête de l'auditeur, et convolue en conséquence, en temps réel, la scène sonore que entendue par les HRTF correspondantes. Le but est de donner le sentiment que l'auditeur écoute une scène sonore fixe dont il peut préciser la disposition par des mouvements de tête libres. Cette technologie, qui sera sans doute à terme une avancée importante dans le développement du binaural, rencontre à l'heure actuelle des obstacles : les mouvements de tête de l'auditeur sont limités, le système ne prend souvent en compte que l'azimut, sur un angle restreint, cesse de fonctionner si le capteur sort de son champ de détection (l'auditeur ne peut alors baisser la tête)... Par ailleurs cette installation, bien que légère au demeurant, n'est guère compatible avec une majorité de situations

d'écoutes. Malgré tout, au stade actuel de son développement, le head tracking apporte déjà des améliorations nettes aux performances du binaural.

Le transaural

Certains lecteurs auront sans doute été surpris que ne soient mentionnées ici que les expériences de restitution 3D faites au casque : en effet, si l'on est capable de rendre un espace en trois dimensions au moyen de deux signaux sonores au casque (oreille gauche et oreille droite), pourquoi ne pourrait-on pas le faire au moyen d'une paire d'enceintes ? Cette technologie, qu'on appelle transaurale, existe effectivement. Elle ne fonctionne cependant que dans des conditions restreintes : elle consiste à envoyer sur les hauts-parleurs gauche et droit d'un système stéréophonique deux signaux convolués par des HRTF et permettant de restituer un espace en trois dimensions.

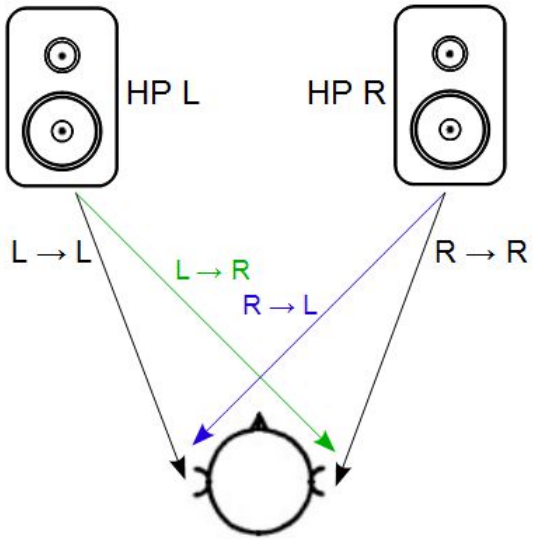


FIGURE 1.13 – Intercorrélation gauche-droite lors d'une écoute sur enceinte : l'oreille gauche entend non seulement le haut-parleur gauche (signal L vers L), mais aussi le haut-parleur droit (signal R vers L), et inversement.

Le principal problème du transaural est qu'il nécessiterait de complètement contrôler, et donc, décorreler ce que reçoit l'oreille gauche de ce que reçoit l'oreille droite. Or si une telle condition se trouve facilement validée au casque, il n'en est pas de même sur hauts-parleurs, où l'oreille gauche entend toujours le haut-parleur droit, et inversement (voir figure 1.13). On ne peut alors garantir que le son convolué par la HRTF de l'oreille gauche, ne va pas entrer en conflit, lorsqu'il sera entendu par l'oreille droite, avec le son convolué par la HRTF de l'oreille droite... La solution est donc d'introduire, dans le haut-parleur droit, le signal du haut-parleur gauche traité de manière à annuler l'influence de celui-ci sur l'oreille droite (c'est le « cross-talk cancelling », voir le système TRADIS, cf. notamment [9] pp.171-172). Ce procédé est toutefois non seulement compliqué à

maîtriser, mais restreint les mouvements de tête du sujet, et impose une position de sweet spot [1] très définie.

La technologie transaurale, outre qu'elle ne nous intéresse pas directement dans notre étude, comporte donc des inconvénients qui rendent compliquées ses applications commerciales pour le grand public, et ce, malgré les résultats étonnants qu'elle peut apporter dans de bonnes conditions d'écoute.

1.2.2 Applications

Le binaural peut se destiner à un certain nombre d'applications, qui recoupent plusieurs domaines.

Applications d'utilité publique

On a pu voir dans les technologies binaurales une amélioration possible dans la vie quotidienne, et notamment pour les handicapés : ainsi, dans l'audiodescription pour aveugles. Lorsque ceux-ci regardent un film, le binaural permettrait par exemple de restituer les sons du films dans un espace défini (stéréo, 5.1) et d'ajouter les commentaires dans un autre point de l'espace, afin que les spectateurs puissent suivre le film sans confondre commentaires et sons originaux.

De même, des projets plus ambitieux de système de guidage ont été imaginés sur les technologies binaurales : Gaëtan Parseihian [17] étudie ainsi le principe d'un système dans lequel la personne malvoyante pourrait entrer la destination qu'elle veut atteindre (une adresse, par exemple), et se laisser guider au casque, par des sons intervenant dans l'espace 3D pour lui signaler un carrefour, une boulangerie, une fontaine...

Applications techniques

Dans un cadre plus proche de l'audiovisuel, les technologies binaurales pourraient être une solution intéressante à la problématique de la prise de son multicanale, pour laquelle les ingénieurs du son ne disposent pas toujours des systèmes d'écoute appropriés. Ainsi, on pourrait imaginer une simulation de restitution d'espace quadriphonique au casque, pour un ingénieur du son effectuant des enregistrements d'ambiance en croix IRT. Ce genre d'applications est tout à fait envisageable du point de vue de la technique, et l'on pourra notamment se référer au mémoire de François Heller, étudiant à Louis Lumière (promotion 2013), et au mémoire de Pierre Bompy (promotion 2008) [4].

De la même manière, les systèmes d'intercommunication pourraient trouver des applications au binaural : jusqu'à présent les seules hiérarchies possibles entre les différents participants d'une intercommunication au casque (pour pilotes, cadreurs de télévision, militaires etc.) reposaient sur la différentiation gauche-droite (en télévision, la voix du réalisateur arrive par exemple dans l'oreille

1. Rappelons que le terme de sweet spot, qui nous vient de la statistique et du sport, définit dans notre cas l'emplacement géographique où le son est optimal, dans une salle et/ou pour un système d'écoute donné.

gauche, droite, ou les deux) et la priorité d'atténuation (la voix du réalisateur atténuant toutes les autres voix). Le binaural permettrait de diversifier les provenances des différentes voix (ou signaux d'alerte) afin de permettre plus facilement au cerveau de discriminer qui parle, et quelle est son importance. Attention toutefois, bien sûr, aux dangers de saturation par un trop-plein d'informations.

Applications artistiques ou du divertissement

Le binaural trouve des applications intéressantes dans de nombreux domaines du divertissement.

Le visionnage de films Une proportion non-négligeable d'auditeurs profitant d'oeuvres audiovisuelles dans le train, ou même chez eux, au casque, il sera sans doute intéressant de proposer sur les DVD et Blu-Ray une option d'écoute de la bande-son originale 5.1 en binaural. Il en est bien entendu de même pour les programmes télévisés qui seraient mixés en 5.1. A ce sujet, le mémoire de fin d'étude à Louis Lumière de Florent Castellani évoque des pistes de réflexion et des conclusions intéressantes [5].

La radio Outre la binauralisation des mixages 5.1 réalisés pour la radio, la fiction et le documentaire radiophoniques, notamment, pourraient tirer parti des possibilités que le binaural recèle en lui-même : il est un outil sans doute intéressant pour la création sonore, avec un espace propre, et l'intimité que confère l'écoute au casque ne serait plus ici une limite mais un atout.

A noter qu'il ne saurait sans doute s'agir ici de « radio » au sens traditionnel du terme, étant donné que les programmes, destinés à l'écoute au casque, ne pourraient pas être diffusés sur un poste (bien que Blauert, entre autres, évoque des possibilités de filtrage qui aideraient à la compatibilité des prises de son binaurales avec une diffusion sur hauts-parleurs, cf. [3], p.361, voir aussi les recherches de Günther Theile, par exemple dans [32]). Il s'agira donc plutôt de créations sonores, qui pourraient être véhiculées sur une plage de fréquences dédiée (éventualité que les applications de la radio numérique permettent), ou seraient mises en ligne.

La musique On pense en premier lieu à la binauralisation des mixages multicanaux, que peu de personnes ont les moyens d'entendre chez eux (voir le succès mitigé du SA-CD de Sony-Philips). Le binaural pourrait être une solution à ce problème. Mais on peut également penser à la création musicale : l'espace binaural ne se limitant pas au plan horizontal, il met à la portée des mixeurs un espace entier à explorer par la musique, plus accessible que les nombreux formats inventés par le cinéma. La question se pose alors toutefois du réel besoin de l'élévation en musique (qui dépend des cas : musique actuelle, musique électroacoustique... ?).

Le jeu vidéo Une application importante du binaural, surtout si l'on parvient à l'associer au head-tracking, serait très certainement le jeu vidéo. Dans ce

mode de divertissement, où l'espace sonore est à la fois une aide et un obstacle, la possibilité d'ajouter des sons à l'arrière, en haut, en bas, ajouterait sans doute une dimension appréciable à l'expérience de jeu. Par ailleurs la proportion non-négligeable de personnes jouant au casque, pourrait être un argument de plus à la propagation du binaural. Enfin, la capacité de démasquage conférée par le binaural, qui permet de clarifier une scène sonore (ou, si l'on part d'une scène sonore pauvre, qui en révèle les manques...) pourrait permettre de rendre plus lisible l'environnement sonore du jeu, entre sons environnementaux, sons interactifs, sons de personnage, sons de réponse... (d'après la taxonomie sonore tirée du modèle de Mark Grimshaw, cf. notamment [23], p.45, voir aussi [16]). A noter que des expériences de jeu en binaural sont déjà accessibles au public : citons les jeux Blindside (http ://www.blindsidegame.com/), disponible notamment pour iPhone, ou encore Papa Sangre (http ://www.papasangre.com/), tous deux sans image.

A ce stade, l'un des principaux obstacles à l'intégration d'un moteur binaural dans le jeu vidéo, réside dans les faibles espaces-mémoires disponibles, surtout ceux réservés au son, par rapport aux nécessités d'un plug-in binaural censé travailler en temps réel. On peut cependant espérer, au vu de l'optimisation constante des outils informatiques, que ce genre de limites sera finalement repoussée.

1.2.3 Concrètement, aujourd'hui en France

A l'heure actuelle, le binaural a déjà donné lieu à la réalisation d'outils de spatialisation pour le mixage, et à plusieurs applications notables.

Les outils

Pour travailler en binaural, l'ingénieur du son dispose aujourd'hui de plusieurs outils, que ce soit en binaural « natif » ou en binaural « traitement du signal » . En binaural natif, signalons la tête Neumann KU-100, qui semble donner de bons résultats tant en terme de localisation que de timbre des sources.

D'autre part, différents logiciels existent actuellement pour permettre de spatialiser des sources sonores en binaural : par exemple, l'Ircam HEar de Flux (voir le site officiel [26]), le H3D de Longcat [30], l'Auro-3D Headphones du Auro-3D Engine [25], ou encore la technologie Dolby Headphone [29] (voir aussi des recherches récentes avec [13]). Dans le cadre de ce mémoire, c'est le logiciel SpherAudio de Digital Media Solutions [27], encore en développement, qui a été utilisé.

Ces outils sont censés permettre soit la spatialisation en temps réel en binaural (par exemple le Longcat H3D), soit la binauralisation de formats discrets « figés » comme le 5.1 (Dolby, Auro-3D). A noter que le binaural est une technologie relativement récente, et les enjeux (financiers notamment) deviennent de plus en plus importants à mesure que les recherches aboutissent à des solutions logicielles concrètes. Il n'est donc pas facile de savoir précisément ce que contient chacun de ces outils : les informations précises relatives à leur fonctionnement sont pour le moment inaccessibles pour cause de secret industriel. En parallèle, la

politique commerciale de ces différentes sociétés semble les pousser à faire l'éloge de produits qui ne sont pas forcément encore au point. C'est pourquoi, alors que les annonces flatteuses sur les outils de spatialisation binaurales se multiplient, il convient sans doute de rester prudent sur leurs performances réelles.

Le projet BiLi Le logiciel SpherAudio est utilisé notamment dans le cadre du projet BiLi, qui regroupe, autour de quelques partenaires principaux (Digital Media Solutions, Orange Labs), des broadcasters et institutions liées au milieu du son (Radio France, France Télévision, CNSMDP), et des institutions et laboratoires (Ircam, LIMSI). Ce projet vise entre autres, à concevoir une solution hardware de spatialisation binaurale : le BP84 (voir la page web du constructeur [28]), dont le but est de permettre la retranscription en binaural d'un espace multicanal : il reçoit en entrée les différents canaux discrets, et renvoie en sortie les canaux gauche et droit du flux binaural correspondant. L'objectif est à terme de proposer un outil simple et polyvalent pour binauraliser les mixages multicanaux, à destination des entreprises de télévision et de post-production.

Les applications

Quelques applications du binaural existent déjà. Outre quelques (rares) expériences de mixage, fictions sonores et films, le projet le plus ambitieux actuellement réalisé est un site internet où Radio France met en ligne des contenus en version stéréo, 5.1 et/ou binaural 5.1 : le site NouvOson (http ://nouvoson.radiofrance.fr/). Les mixeurs de Radio France utilisent, depuis mars 2013, date de lancement du site, une méthodologie du type : mixage en 5.1, binauralisation, réduction de débit par le codec AAC 192, mise en ligne, comme résumé sur le schéma 1.14.

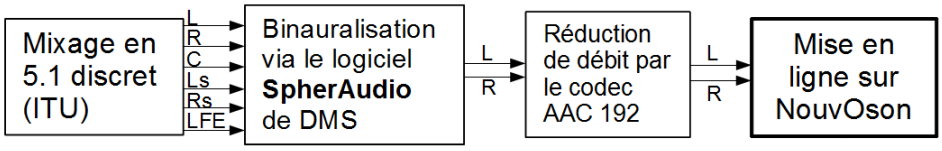


FIGURE 1.14 – Le workflow utilisé par Radio France pour les contenus binauraux de NouvOson.

C'est en grande partie sur les problématiques soulevées par cette technique de travail que les principaux aspects de ce mémoire ont été basés. En effet, si le projet NouvOson constitue l'une des applications les plus abouties du binaural à l'heure actuelle, plusieurs problématiques n'en ont pas été pleinement résolues. Par exemple, la qualité de la restitution de l'espace 5.1 en binaural, ou encore les effets du codec AAC 192 sur la spatialisation ressentie, ne semblent pas avoir été préalablement évalués. Une grande partie des choix d'expérience réalisés pour ce mémoire est donc axée vers la résolution de ces questions, de manière à pouvoir proposer des ébauches de solutions à ces problèmes concrets. La binauralisation des sons à Radio France s'effectuant avec le logiciel SpherAudio, utilisé pour

ce mémoire, l'expérience développée dans ce travail reprend donc exactement le processus de travail de Radio France pour NouvOson.

Résumé et conclusion

Le binaural repose sur des principes de localisation sonore mettant en jeu des différences d'intensité (ILD) et de temps (ITD) entre les deux oreilles, ainsi que l'emploi des DDF, filtrages propres à la morphologie de notre tête et dépendants du point d'origine des sons, qui nous permettent, en se combinant, une localisation plus ou moins précise. L'association des ITD, ILD et DDF donne ce que l'on appelle les HRIR, et leur version fréquentielle via transormée de Fourier, les HRTF. Les mouvements de tête, qui cumulent les informations données par les HRTF, ont dans ce processus une importance capitale. Les technologies binaurales cherchent donc à reproduire, en simulant ces HRTF, des espaces sonores en trois dimensions au travers d'un simple casque. Nous étudierons dans ce mémoire la spatialisation binaurale obtenue par traitement du signal (à la différence du binaural « natif »). Ces technologies ne sont pas sans poser un certain nombre de difficultés, liées tout autant aux subtilités des mécanismes physiologiques de la localisation sonore, qu'aux limites actuelles des procédés informatiques et du matériel. Cela n'empêche pas le binaural de connaître d'ores et déjà quelques applications, dont l'une d'elles, le site NouvOson de Radio France, a guidé la mise en place des expériences qui seront exposées dans la suite de ce mémoire. Ce site présente des contenus binauralisés avec le même logiciel qui sera utilisé pour cette étude : le logiciel SpherAudio.

1.3 Le logiciel SpherAudio

L'intégralité de ce mémoire est basée sur le logiciel de spatialisation binaurale développé par Digital Media Solutions : SpherAudio. Une présentation de cet outil s'impose donc, qui nous permettra d'aborder rapidement son histoire, ses performances, et son fonctionnement.

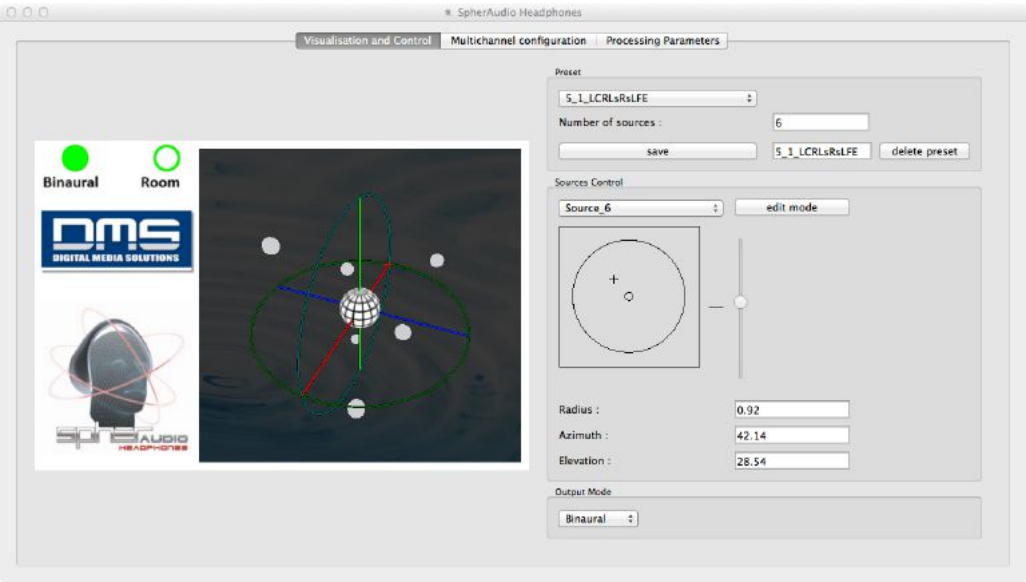


FIGURE 1.15 – La page principale du logiciel SpherAudio

1.3.1 Présentation

Digital Media Solutions

Digital Media Solutions (DMS) est une société fondée en 2009, se définissant comme « spécialisé[e] dans la conception, la fabrication et la commercialisation de solutions innovantes pour les industries du Cinéma, de la Hi-Fi, du Home Cinema, de la Télévision et de la Musique » . (http ://www.dms-cinema.com/). Elle est basée à Noisiel [2].

SpherAudio

SpherAudio est un moteur de synthèse binaurale, conçu par M. Aussal dans le cadre d'un programme de collaboration-recherche (DMS - LIMSI - CMAP), et développé sous sa forme industrielle par l'équipe Audio 3D de Digital Media Solutions.

À l'origine un code Matlab, SpherAudio est aujourd'hui un plug-in VST, fonctionnant avec le logiciel DAW [3] Reaper, permettant de spatialiser en temps

2. 45, Grande Allée du 12 Février 1934, 77186 NOISIEL, tél. +33 (0)1 80 81 52 40
3. « Digital Audio Workstation »

réel jusqu'à 64 sources sonores dans l'espace binaural 3D, en utilisant une HRTF à choisir parmi une banque de dix. SpherAudio comporte en outre plusieurs réglages intéressants : au traitement des sources par la HRTF choisie s'ajoute une « Room » optionnelle, réverbération simulant l'acoustique d'une salle d'écoute, comme vu en 1.2. SpherAudio peut également être employé en mode VBAP (*Vector Base Amplitude Panning*, voir [19] et [10]) , et non plus en binaural, afin de spatialiser des sons sur un maillage d'enceintes déterminé, option utilisée lors de ce test, et sur laquelle nous reviendrons plus loin.

1.3.2 La partie binaurale

La partie binaurale de SpherAudio comporte plusieurs options dont la théorie a déjà été abordée. A l'époque où de son utilisation dans le cadre de ce mémoire (février-mai 2013), SpherAudio comprenait un mode « binaural », un mode « stéréo » (qui avait donc l'effet d'un *downmix*) et un mode « VBAP ». En mode binaural, le plug-in convoluait en temps réel le signal entrant par la HRTF voulue. Ceci impliquait donc deux étapes : la conception d'un routing, et le choix d'une HRTF.

Les HRTF se sélectionnaient dans une banque de dix disponibles : les HRTF « Best Matching » 1, 2, 3, et les HRTF « Min Subset » 1 à 7. Les trois premières correspondaient à trois HRTF sélectionnées (sur les 45 relevées dans le cadre du projet Listen) car censées donner un rendu à peu près satisfaisant pour l'ensemble des êtres humains, toutes morphologies confondues (d'où leur nom : « best matching » soit « celles qui correspondent le mieux »). Les Min Subset 1 à 7, quant à elles, étaient censées répondre à l'axiome suivant : quelle que soit la morphologie de l'auditeur considéré, une des sept Min Subset lui correspond de façon satisfaisante.

Rappelons que cette banque d'HRTF constitue un compromis au problème de l'individualisation des HRTF : il est en effet peu pratique de devoir mixer avec une HRTF en sachant que statistiquement elle ne conviendra pas à une certaine proportion d'auditeurs. Il est probable que la recherche dans ce domaine permettra d'apporter des solutions plus globales à cette problématique.

Le choix de la HRTF s'accompagne d'une possible option « Left/Right Egalization » (voir fig.1.16). Celle-ci est destinée à compenser les déséquilibres de niveaux qui pourraient accompagner le traitement par la HRTF choisie. Ce réglage permet donc, sans dégrader les qualités de localisation de la HRTF, de ré-équilibrer les canaux gauche et droite.

La room

La room est une option de réverbération associée à l'utilisation du binaural. Comme exposé en 1.2, elle est censée favoriser l'externalisation des sons (donc réduire l'IHL), et leur bonne localisation dans l'espace 3D, en simulant l'acoustique d'une salle d'écoute.

Ce réglage pose la question de l'algorithme utilisé et de son dosage. SpherAudio permet théoriquement la modification de ces deux paramètres, mais la

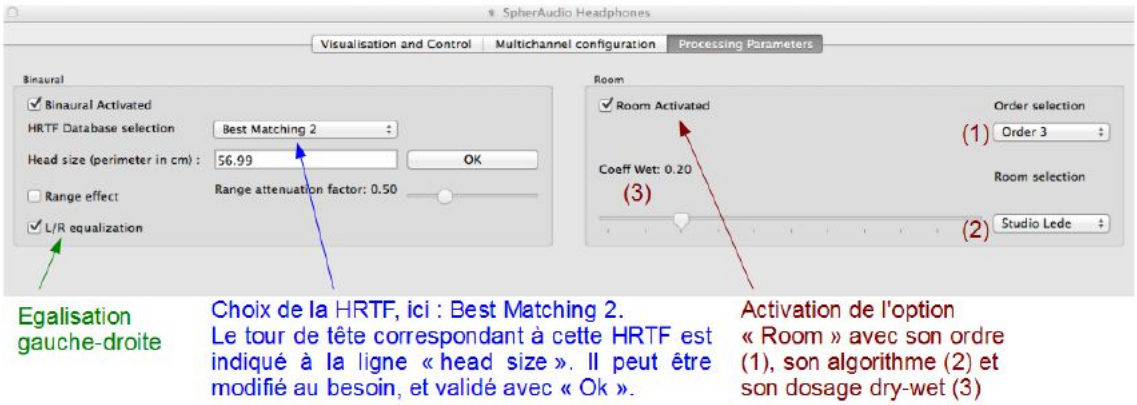


FIGURE 1.16 – La page « Processing Parameters » , où se trouvent plusieurs réglages relatifs à l'utilisation du logiciel en binaural.

version utilisée pour ce mémoire ne comportait qu'un algorithme de room : celui d'un studio LeDe (fig.1.16), dont les caractéristiques sont évoquées en figures 1.17 et 1.18. Cet algorithme simule l'acoustique d'un studio bâti selon les normes LeDe, en proposant différents ordres pour le rendu de l'espace : ordres 0 à 3. L'ordre 0 correspondrait à la captation de l'acoustique par un microphone omnidirectionnel, les ordres supérieurs simulent la captation par des capsules plus nombreuses, et donc, dans davantage de directions. L'ordre 3 rend ainsi théoriquement l'espace le mieux retranscrit, et le plus précis. Ce système présente de forts points communs avec la théorie de l'ambisonie.

Si l'ordre 3 est techniquement le plus performant, l'ajout d'une « room » est en pratique synonyme de détimbrage, même minime, et de changement du sentiment de localisation. Or à l'écoute, il peut arriver que le mixeur préfère utiliser un ordre 0, 1 ou 2, plus flous, et peut-être plus flatteurs pour certains sons. Par ailleurs, le choix de l'ordre influe sur les temps d'export. Ainsi, sous Reaper, si l'export (*bounce*) d'un son binauralisé en ordre 0 peut se faire à environ une fois la vitesse de lecture, l'export en ordre 1 se fait à 0.5 fois la vitesse, en ordre 2 : 0.3 fois et en ordre 3 : 0.2 à 0.1, en raison des temps de calcul. Pour une musique de 3 minutes, l'export peut ainsi prendre une demi-heure.

Un réglet de dosage *dry-wet* est également ajouté à la « room » : à l'époque de la réalisation de ce mémoire, ce réglet était conçu de telle sorte que, à la valeur 0, la room n'était pas appliquée, et à la valeur 1, on n'avait plus aucun son direct (comme dans une réverbération classique). Le niveau de « room » à appliquer dépend beaucoup du type de source à traiter : l'expérience montre en effet que certains sons nécessitent une dose importante de room pour faciliter l'externalisation, tandis que d'autres sons n'en ont pratiquement pas besoin. Le dosage de la room est donc laissé à l'appréciation du mixeur.

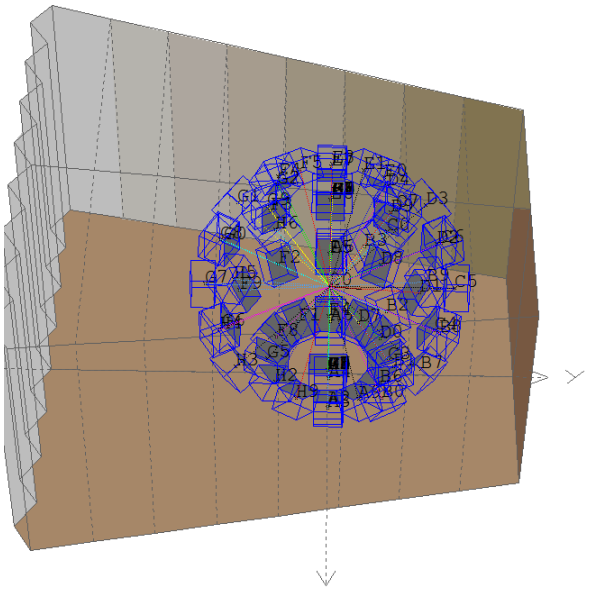


FIGURE 1.17 – La modélisation de la « room » LeDe de SpherAudio à l'ordre 3 : un studio LeDe dont l'acoustique est captée par un système ambisonique d'ordre 3.

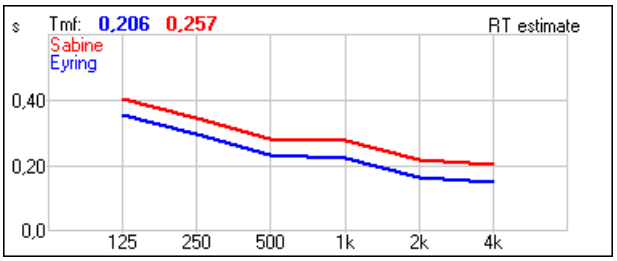


FIGURE 1.18 – Les courbes d'Eyring et de Sabine relatives au studio LeDe modélisé pour l'option de « room » .

1.3.3 Le mode « VBAP »

Bien que SpherAudio ait pour vocation première le traitement du binaural, ses concepteurs ont travaillé sur une adaptation qui permettrait d'en faire un outil de spatialisation sur système d'enceintes, en utilisant le principe du « Vector Base Amplitude Panning », ou VBAP (voir à ce sujet Pulkki, 1997 [19], et Pulkki et Lokki, 1998 [10]). Cette technologie permet la création d'un « panpot » amélioré, capable de positionner une source sonore en n'importe quel point de n'importe quel système d'enceintes, en azimut et en élévation (selon, bien entendu, les capacités du système utilisé). Il est nécessaire d'entrer préalablement les coordonnées précises des hauts-parleurs constituant le système d'écoute, et d'avoir, dans la mesure du possible, un maillage d'enceintes équidistant par rapport sweet spot (c'est-à-dire que les hauts-parleurs sont situés à la surface d'une sphère qui aurait pour centre le sweet spot du système, comme c'est le cas en 5.1 ITU) ; à la suite de quoi le logiciel pourra calculer lui-même la dose d'énergie à envoyer dans chaque enceinte pour obtenir le placement désiré.

Ainsi, si l'on change le menu déroulant « binaural » et qu'on sélectionne, par exemple, « 5.1 », alors on aura la possibilité d'envoyer nos sources où l'on veut sur un système 5.1 ITU (fig.1.19). En entrant les coordonnées adéquates des hauts-parleurs, on peut également créer des présets permettant d'employer n'importe quel type de système : ainsi le préset « Studio DMS » permettait de spatialiser les sources n'importe où sur le système 24.2 aménagé alors dans le studio de DMS (avec toutefois des distorsions de placement dues au fait que les hauts-parleurs n'étaient pas équidistants du sweet spot). On obtenait, dans un

langage d'ingénieur du son (incorrect du point de vue scientifique), un « pan-pot 26 voies » .

A noter que ce mode de fonctionnement court-circuite les réglages associés au binaural : ainsi, l'application des HRTF bien sûr, et l'option de « room » , sont alors désactivés.

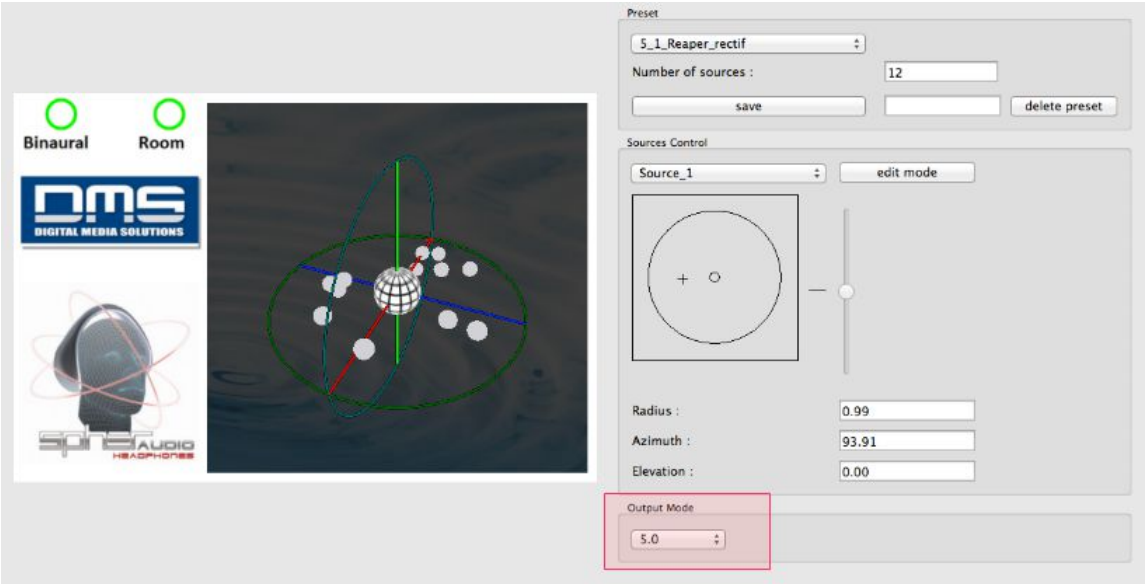


FIGURE 1.19 – La page principale de SpherAudio en mode VBAP : on a sélectionné ici un mode 5.0. Le schéma principal représente donc 5 sources en positions L, C, R, Ls, Rs, complétées ici par des sources qu'on a voulu spatialiser indépendamment de ces 5 canaux figés.

1.3.4 Comment insérer SpherAudio dans le processus de mixage ?

Du fait de sa relative lourdeur en calculs, le logiciel a été conçu pour être inséré une, à deux fois maximum dans une session de travail. La méthode de travail envisagée par les concepteurs est d'insérer le plug-in sur un auxiliaire dédié, vers lequel on envoie les sources qu'on veut binauraliser. Cela peut être comparé au mode de fonctionnement d'une réverbération, à ceci près que c'est la sortie de nos pistes, et non un « send » , qui est supposé être envoyé vers SpherAudio, puisqu'on est censé vouloir binauraliser l'intégralité de notre son (il n'y a donc normalement pas lieu de garder une possibilité de dosage entre le son non-traité et le son traité). (Voir fig.1.20)

La figure 1.21 montre le panneau de contrôle en azimut et en élévation. Si l'on veut spatialiser un son en un point de l'espace, binaural ou autre, on peut utiliser la tablette principale de SpherAudio (1) : une représentation d'un disque, au sein duquel on déplace à la souris une petite croix symbolisant la position du son (correspondant à l'une des entrées de SpherAudio, que l'on a préalablement choisie dans le menu déroulant correspondant (2)). Un réglet, sur le côté (3),

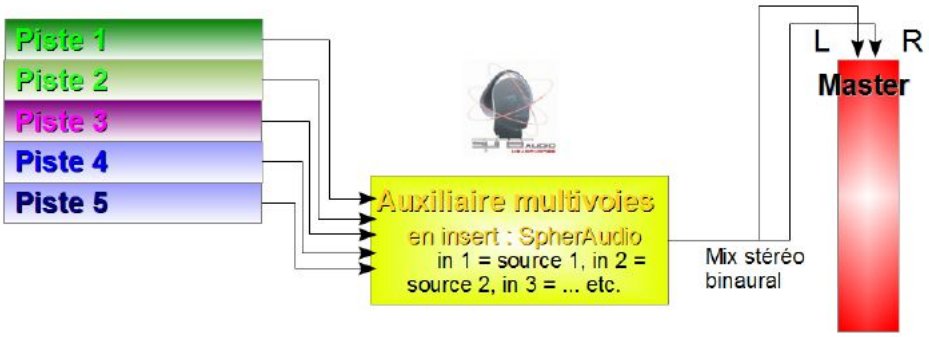


FIGURE 1.20 – Un exemple de routing pour le mixage en binaural avec SpherAudio.

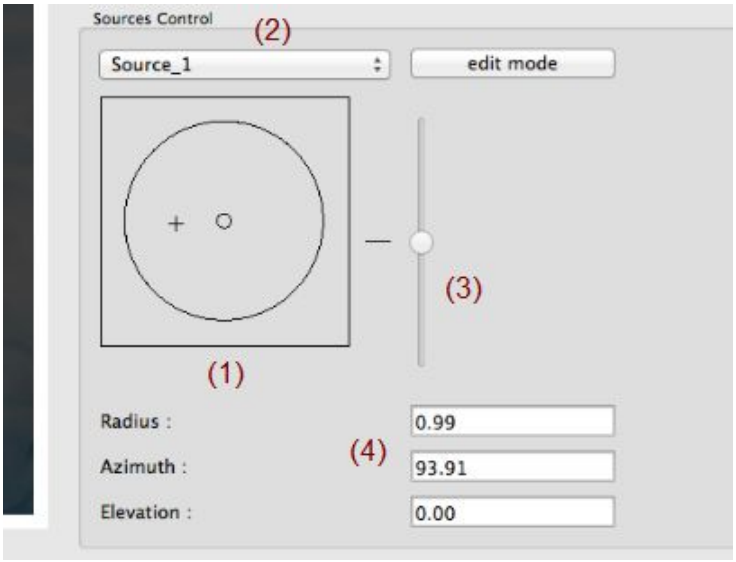


FIGURE 1.21 – Le panneau de contrôle d'azimut et d'élévation. (1) : réglage de l'azimut. (2) : choix de la source, c'est-à-dire du canal d'entrée affecté aux réglages. (3) : réglage de l'élévation. (4) affichage des valeurs de distance, azimut et élévation choisis.

permet de jouer sur l'élévation. Par exemple : si l'on a deux sources, et que l'on veut spatialiser la première : on indiquera « nombre de sources : 2 », puis on sélectionnera dans le menu déroulant (2) « Source : 1 » et la croix indiquera dès lors la position de la source 1 (correspondant normalement à l'entrée 1 de mon auxiliaire). On pourra ensuite déplacer cette croix dans le plan horizontal sur le disque (1), et en élévation avec le réglet (3). Les angles et distances correspondants s'affichent au fur et à mesure dans des fenêtres associées (4). Lorsque la localisation du son est jugée satisfaisante, on relâche la souris, et cette position restera figée jusqu'à la prochaine modification. Si l'on veut réaliser une automation de localisation, on peut armer l'écriture de l'azimut, de l'élévation et du radius (c'est-à-dire la distance), et modifier les courbes d'automation, soit en travaillant directement dessus, à la souris, soit par une « passe » de mixage

(en mode *Write*, *Touch*, ou *Latch* dans Reaper), en agissant sur le pointeur du disque et le réglet d'élévation : les réglages seront enregistrés en temps réel, et l'on pourra les modifier plus tard comme n'importe quelle automation (fig.1.22).

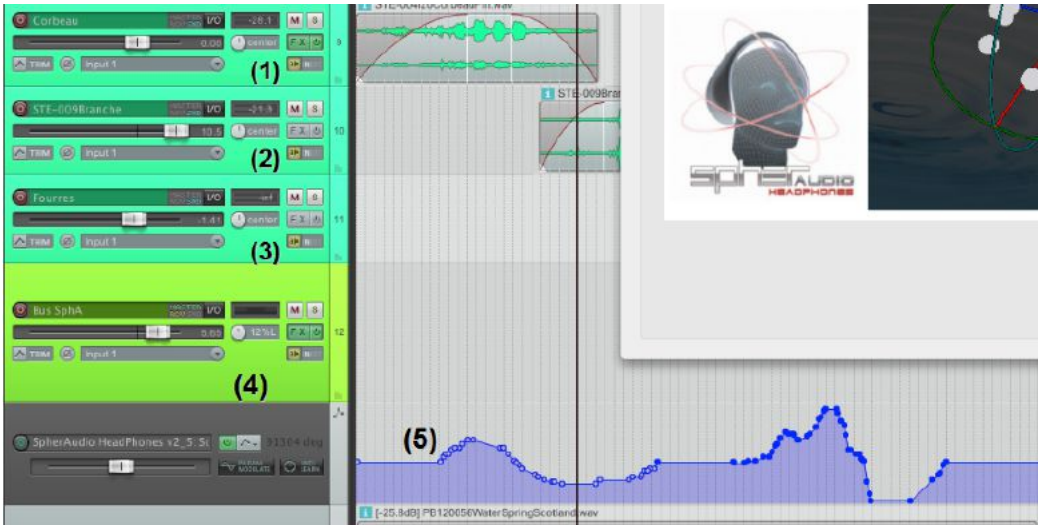


FIGURE 1.22 – Un exemple d'automation avec SpherAudio : (1), (2), (3) : pistes audio ; (4) : auxiliaire SpherAudio ; (5) : automation d'un paramètre, ici l'azimut de la source 1.

Si maintenant l'on veut mémoriser plusieurs positions de sources, et les rappeler en cours de lecture sans avoir à les replacer pointeur par pointeur, on peut mémoriser les positions en un *snapshot*, ou « Preset », qui vient se placer dans une liste. On peut ainsi changer de *snapshot* à tout moment. Ces *snapshots* peuvent également être des presets de nombres de sources et de système d'enceintes simulés en binaural : « 5.1 », « 6.1 »... Auquel cas les positions rappelées pourront figurer celles des hauts-parleurs simulés par le logiciel (fig.1.23).

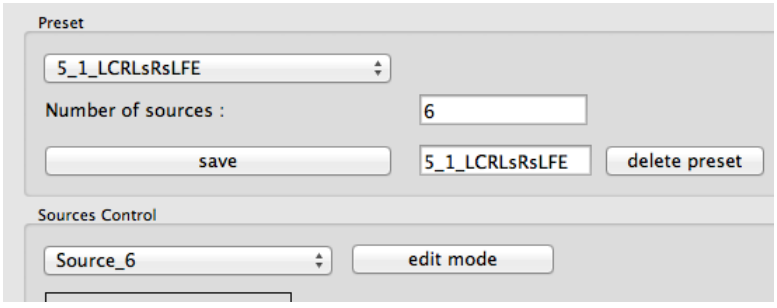


FIGURE 1.23 – La fenêtre des Presets. Ici : on a chargé un preset 5.1.

SpherAudio s'apparente donc à un plug-in assez simple d'utilisation, dès lors que l'on maîtrise les spécifités du mixage binaural. Pour terminer cette présentation, seront traités deux exemples simples, en lien avec ce mémoire, pour mieux aider à comprendre le fonctionnement du logiciel.

1.3.5 Exemples d'utilisation

En « VBAP » : on veut mixer en 5.1 sur enceintes avec SpherAudio

On a notre session de mixage complète (sous Reaper) avec toutes nos pistes. On ajoute un auxiliaire « SpherAudio » 6 voies, sur lequel on insère le plug-in. Celui-ci admettra désormais : en source 1 : l'input 1 de l'auxiliaire ; en source 2 : l'input 2, etc. On choisit l'output mode : 5.1. On envoie nos pistes audio vers notre auxiliaire.

Si l'on veut maintenant envoyer une piste vers un haut-parleur unique, il suffit de jouer sur le *routing* des pistes : en 5.1 ITU (L, C, R, Ls, Rs, Lfe), une caisse claire envoyée vers l'input 1 de l'auxiliaire, traitée comme Source 1 par SpherAudio en preset 5.1, partira directement, et uniquement, vers le haut-parleur gauche.

Si l'on veut spatialiser une source uniquement sur l'avant, on peut l'envoyer dans les entrées 1 et 2 de l'auxiliaire, et la doser avec le *pan pot* associé à sa piste. Jusqu'ici, le plug-in n'a donc pas une utilité spécifique.

Si toutefois on désire spatialiser une source de façon plus spécifique (à mi-chemin entre l'avant et l'arrière, ou en lui faisant faire un déplacement), on peut recourir à une autre méthode : il faut tout d'abord élargir l'auxiliaire, par exemple à 8 canaux. SpherAudio pourra donc traiter indépendamment 8 sources, correspondant aux 8 entrées de l'auxiliaire. Les 6 premières sources sont envoyées chacune vers une enceinte du 5.1, mais on peut spatialiser les deux dernières comme l'on veut. Ainsi, si l'on désire déplacer une guitare monophonique, on peut l'envoyer vers l'entrée 7 de l'auxiliaire. On sélectionne la source 7 dans SpherAudio, qui correspond dès lors à la guitare ; on effectue le déplacement désiré, en automatisant les paramètres correspondants. Grâce au mode « VBAP » , le déplacement sera retranscrit sur le système d'enceinte, sans modifier la position des autres sources. La seule contrainte est donc de pouvoir augmenter facilement la taille de l'auxiliaire. Cette manipulation est toutefois très simple sous Reaper. Le plug-in pouvant traiter actuellement jusqu'à 24 sources en simultané, et les auxiliaires de Reaper pouvant véhiculer jusqu'à 64 canaux, laissent donc une certaine marge de manoeuvre au mixeur.

Quand le mixage est terminé, on peut exporter les 6 sorties du master 5.1 (à noter que quel que soit le nombre de sources traitées, et donc la taille de l'auxiliaire, tant que SpherAudio reste paramétré en 5.1 il ne délivrera du son que sur les 6 premières sorties, que l'on envoie vers les 6 canaux du master 5.1). On obtient ainsi un fichier audio 5.1. Si l'on avait voulu ajouter des sons non traités par SpherAudio, il suffit d'avoir envoyé les pistes correspondantes dans les bons canaux du master, et ils seront pris en compte dans l'export et mélangés canal par canal aux sons traités par le plug-in.

En binaural : on veut binauraliser un mixage 5.1

Cette opération est relativement simple : une fois que l'on a notre mixage 5.1, on choit dans le menu déroulant du plug-in : « Output mode : binaural ».

On choisit une HRTF et un réglage de « room » (ordre et dosage), et le logiciel délivre en sortie deux canaux audio, contenant le mixage 5.1 binauralisé. On achemine ces deux canaux vers un master stéréo, à partir duquel on exporte notre mixage.

Conclusion

Les étapes décrites ci-dessus ont été réalisées dans le processus de mixage utilisé pour ce mémoire, avec quelques modifications : on a d'abord réalisé un mixage en 5.0 (le caisson de basses posant toujours problème dans la localisation des sons, et les casques retranscrivant de toute manière assez mal les fréquences graves, on avait décidé de se passer du « .1 »), puis sa binauralisation. Les réglages utilisés seront décrits avec davantage de précision dans les sections qui suivent.

Chapitre 2
Partie expérimentale

Introduction

La mise en place de la partie expérimentale de ce mémoire a suivi plusieurs guides et contraintes. La réalisation des tests subjectifs a obéi avant tout à un souci de rigueur et de pertinence, à conjuguer avec les délais impartis. Les conseils avisés de Matthieu Aussal, ainsi que de Brian Katz du LIMSI-CNRS, se sont révélés particulièrement précieux dans ce domaine. Les préoccupations formulées par les ingénieurs du son ont aussi été gardées à l'esprit, qu'ils soient de Radio France ou de DMS. Enfin, les cours dispensés par Étienne Hendrickx ont apporté une aide appréciable dans le choix des stimuli, la conception globale du test, et l'exploitation des résultats.

2.1 Expérience

Le projet initial d'expérience, très ambitieux, visait à comparer la restitution de l'espace sonore et l'appréciation du son par les auditeurs :

- sur un ou plusieurs systèmes d'enceintes ;
- sur ces mêmes systèmes émulés en binaural, sans aucun autre traitement ;
- et enfin, en binaural avec traitements.

Un test subjectif devait permettre d'évaluer ces différents éléments, dans l'optique de déterminer la fidélité du binaural par rapport à la diffusion multicanale simulée, et la robustesse du rendu binaural face à différents traitements sonores, notamment de compression de données.

Ce projet prenait directement sa source à certaines préoccupations du moment. D'une part Digital Media Solution s'interrogeait sur la transportabilité des produits réalisés avec SpherAudio : en cas de diffusion, de réduction de débit pour le stockage... D'autre part, les ingénieurs du son de Radio France, qui utilisaient déjà SpherAudio dans le cadre du projet NouvOson, mettaient en ligne des contenus 5.1 binauralisés puis traités en AAC 192kbps, sans avoir au préalable, pleinement étudié les effets d'une telle manipulation. L'expérience détaillée dans ce mémoire répondait donc à des interrogations précises et contemporaines de sa mise en place.

Enfin, il fut décidé de rajouter un dernier volet à l'expérience, consistant à évaluer quelle HRTF, parmi celles proposées par SpherAudio, était la plus adaptée pour chacun des sujets. Les sujets se verraient proposé un même déplacement de source, binauralisé via plusieurs HRTF différentes, et devraient dessiner la trajectoire entendue ; il faudrait ensuite évaluer pour quelle(s) HRTF le rendu était le plus proche du mouvement voulu. L'un des objectifs de cette expérience était de vérifier que pour chaque sujet potentiel, une des HRTF de SpherAudio offrait un rendu satisfaisant.

2.1.1 Choix des traitements

Il fut estimé que les traitements face auxquels serait évaluée la robustesse du binaural dans le cadre de l'expérience, devraient répondre aux critères suivants :

- intervenir traditionnellement en-dehors du contrôle de l'ingénieur du son, par exemple sur le master ;
- être représentatif du type de traitements qui pouvaient éventuellement être réalisés sur un master binaural, dans le cadre d'un projet concret ;
- pouvoir être appliqués dans le cadre de mes expériences.

La première condition écartait les traitements de mixage, que l'ingénieur du son est censé pouvoir rectifier si besoin, à l'écoute du rendu en binaural. De même, les traitements de mastering, étant effectués par un opérateur, sont censés respecter les spatialisations opérées par le mixeur. En conjuguant ces critères avec les interrogations de Radio France et les besoins du marché, on a finalement retenu les traitements suivants :

- la réduction de débit en mp3 192 kbps, pour le stockage de fichiers sonores ;
- la réduction de débit en AAC 192 kbps, pour la mise en ligne de contenus par Radio France ;
- la réduction en Dolby AC-3 pour les DVD
- la compression antenne, en vue d'une éventuelle diffusion radio, et comme référence, supposée extrême, des traitements que pouvait ou non supporter le binaural (et ce que les sujets en percevaient).

Le MP3 192 kbps fut choisi car jugé représentatif de la qualité audio qu'écoutait un public averti n'ayant pas l'espace pour stocker du PCM [1]. Par ailleurs Brian Katz et Fabien Prezat avaient déjà montré que des codecs à débit trop faible n'étaient pas recommandés pour du binaural [7]. Le choix de l'AAC 192 kbps venait des procédés de mise en ligne de Radio France. Le Dolby AC-3 aurait permis d'évaluer la restitution d'un 5.1 binauralisé pour du DVD, mais ce codec n'a pas été retenu faute d'avoir trouvé une implémentation fiable de cet algorithme.

Le choix d'appliquer une compression « antenne » peut être jugé contestable : la majorité des auditeurs écoutant la radio sur leur poste, et non au casque, il

1. Le PCM (*Pulse Code Modulation*, « modulation d'impulsion codée ») désigne un format audio n'ayant subi aucune réduction de débit. Les formats WAV, AIFF et BWF sont donc des formats PCM.

ne serait peut-être pas pertinent de diffuser sur les ondes un signal binaural. Cette éventualité n'était toutefois pas à écarter [2], et par ailleurs ce traitement fut jugé présenter un « maximum » des altérations que le flux de post-production et de diffusion pouvait imposer à un fichier sonore. Appliquer ce traitement à un signal binaural permettrait donc d'avoir un aperçu d'une dégradation forte du signal en binaural, et l'hypothèse fut formulée que les dégradation dues à la compression antenne (altération forte du spectre du signal, donc des indices de spatialisation ; réduction des différences d'intensité permettant la localisation gauche-droite etc.) seraient probablement plus importantes que celles causées par la réduction de débit. Cela permettait également, de ce fait, d'évaluer la compétence des sujets à repérer et estimer une dégradation audible du signal.

La compression antenne a pu se faire grâce à la collaboration d'Edwige Roncière, qui a permis d'appliquer aux stimuli la compression antenne de France Inter (donc d'une radio largement écoutée), et dotée d'un stéréo enhancer.

FIGURE 2.1 – Synoptique de l'installation de Radio France employée pour réaliser la compression antenne de mes stimuli.

La conversion des stimuli en AAC 192 s'est faite, à débit constant (constant bit rate), au moyen du logiciel Quick Time Pro 7, sur le conseil d'Hervé Déjardin de Radio France. La conversion mp3 s'est faite quant à elle au moyen du logiciel Audacity (utilisant Lame Encoder).

2.1.2 Système discret

En ce qui concernait la comparaison avec l'espace retranscrit par un système multicanal discret, plusieurs systèmes possibles avaient au départ été évoqués, incluant le 5.1, le 7.1, potentiellement le 9.1 d'Auro 3D... Tous systèmes dont le mixage serait possible dans le studio d'écoute 28.2 alors en construction à Digital Media Solutions.

Des contraintes de temps, alliées à un souci de pertinence, nous incitèrent à ne conserver que le 5.1, système employé dans le cadre de NouvOson. Par ailleurs

2. Radio France avait déjà fait des essais de diffusion hertzienne du binaural, comme me l'a confié Edwige Roncière, ingénieur du son. Les résultats avaient été jugés à l'époque, semble-t-il, peu satisfaisants.

ce système était le plus évolué qui pouvait être rencontré chez les particuliers comme dans bon nombre de studios. Le caisson de basses posant traditionnellement problème dans la binauralisation [12] (de l'avis de plusieurs ingénieurs du son de Radio France), et les casques retranscrivant de toute façon mal les graves, il fut décidé d'employer un **5.0, ITU** : des enceintes à équidistances du sweet spot permettant une meilleure utilisation du VBAP de SpherAudio.

Au final, l'expérience eut donc pour objectif de comparer les espaces sonores et l'appréciation de sources mixées en 5.0, dans le cas d'une diffusion :

- sur enceintes, en 5.0 ITU ;
- en 5.0 binauralisé, sans autre traitement ;
- en 5.0 binauralisé, avec traitements : compression antenne, réduction de débit MP3 192kbps, réduction AAC 192.

L'expérience s'articulait autour de **deux sessions de test distinctes** pour chaque sujet : l'une en 5.0 discret, et l'une au casque en binaural. Les sessions duraient **40 minutes** chacune maximum (nos cours avec Etienne Hendrickx nous ayant appris que la concentration d'un sujet baissait après 45 minutes d'expérience), et étaient séparées, au minimum, par **une journée complète** (ceci afin de permettre au sujet de se reposer, et de bien décorreler dans sa mémoire les deux parties de l'expérience). La moitié des sujets réaliserait l'expérience sur enceintes puis au casque, et l'autre moitié au casque puis sur enceintes, afin de vérifier l'influence de cet ordre sur les résultats.

Pour compléter l'expérience et éviter que l'une des sessions ne laisse au sujet la mémoire des spatialisations des stimuli (ce qui pouvait nuire à sa spontanéité lors de la session suivante), il fut décidé de réaliser deux autres mixages 5.0 de chaque stimulus, des « leurres » destinés à donner le sentiment que les spatialisations au casque et sur enceintes n'avaient rien en commun. L'un des trois mixages sur enceintes seulement serait binauralisé, les deux autres n'auraient d'autre utilité que de distraire la mémoire du sujet.

2.1.3 Choix des stimuli

Le choix des stimuli fut guidé par un souci de pertinence, d'imitation des contenus concrets auxquels se destinait le binaural, au détriment des stimuli de laboratoire (sweep, bruit blanc). Cinq type de sons furent sélectionnés :

- **un stimulus musique actuelle**, représentatif du type de contenu que Radio France peut mettre en ligne, avec une prise de son fractionnée limitant les relations de phase entre les sources ;
- **un stimulus musique classique**, représentatif là encore des contenus mis en ligne par Radio France ;
- **un stimulus voix**, potentiellement représentatif de contenus concrets (fiction radiophonique, interview radio, voire dialogues de cinéma spatialisés), et qui jouerait sur l'affinité particulière qu'a notre oreille avec la voix humaine dans la localisation des sons ;
- **un stimulus ambiance**, représentatif d'une scène de fiction radiophonique ou de film ;
- **un bruit rose**, stimulus de laboratoire de référence, au contenu parfai-

tement connu et reproductible, et demeurant malgré tout assez proche de stimuli réels (puisque fréquentiellement proche de sons naturels comme une chute d'eau).

Des stimuli statiques

Seule la localisation de stimuli statiques serait demandée aux sujets,les déplacement de sources étant donc évités. La question du déplacement de la source induit en effet d'autres problèmes : d'une part, la combinaison des HRTF convoluées pour retranscrire l'impression d'un déplacement, joue le rôle approximatif d'un mouvement de tête, et facilite la localisation, ce qui peut conduire à surestimer les capacités du binaural dans les résultats. Ensuite, l'emploi de sources statiques permettait de mettre en valeur les éventuels déplacements involontaires de sources, dus à des aberrations ou des sources fantômes. Enfin, une grande partie des contenus mis en ligne en binaural par Radio France sur NouvOson, comprend des sons « statiques » (instruments de musique, voix...).

2.1.4 Les sujets

Les recommandations ITU BS 1116 [21] et BS 1534-1 [22] relatives à l'évaluation par test subjectif, respectivement d'une dégradation faible, et du niveau de qualité intermédiaire des systèmes de codage, recommandent de recourir à une vingtaine de sujets "expérimentés", c'est-à-dire habitués à une écoute critique des sons, ou bien ayant subi un entraînement préalable à l'expérience, et disposant en tous les cas d'une audition normale. Ces éléments des recommandations ne purent cependant être respectés. Le test fut réalisé avec comme sujets les employés de DMS, par commodité autant que par leur contact avec le domaine des technologies sonores, qui n'en faisait pas tout à fait des naïfs. Quelques personnes expérimentées extérieures à DMS, et quelques sujets naïfs participèrent également. Enfin, une expérience parallèle fut menée à l'école Louis Lumière, mais les résultats ne purent en être exploités dans le cadre de ce mémoire, par suite de contraintes de temps. L'audition des sujets ne put pas davantage être vérifiée, bien qu'aucun d'entre eux ne nous ait fait part de difficultés d'audition, autres que celles liées à l'âge - or Blauert indique que la perception de la localisation sonore resterait à peu près inchangée chez un être humain jusqu'à soixante-dix ans [3]. Avec toutes les précautions nécessaires, et au regard des contraintes de mise en place, il fut donc estimé que ce choix de sujets n'invaliderait pas la pertinence des résultats, bien qu'elle puisse la nuancer.

2.2 Mise en chantier de l'expérience

2.2.1 Réalisation des stimuli

Un laps de temps important a été consacré à l'élaboration des stimuli. *In fine* ceux-ci ont dépendu des matériaux sonores et possibilités d'enregistrement

disponibles, ce qui a conduit, sur ce point comme sur d'autres, à modifier légèrement les ambitions et objectifs de l'expérience. Ces travaux étaient par ailleurs guidés par la constatation que, si intéressants qu'ils puissent être en soi, il importait de perdre le minimum de temps possible dans leur création. Une matière sonore accessible, manipulable et de bonne qualité fut donc recherchée.

Stimulus musique actuelle (nom de code « rock »)

Ce premier stimulus a été réalisé grâce à la collaboration de Baptiste Palacin, (étudiant à Louis Lumière, section Son, promotion 2013), qui mit à disposition de l'expérience l'enregistrement multipistes du morceau de rock *Sans Visage*, composé et joué par son groupe de musique Les Quenelles de Requins. L'enregistrement était suffisamment bon, et les instruments suffisamment nombreux et distincts pour les besoins du test. Le gros du travail a donc été de mixer ce multipistes complexe (plus de dix pistes de guitares différentes) en 5.0, puis pour le casque.

Le mixage en binaural, en particulier en vue de contenu expérimental, soulève plusieurs problématiques, esthétiques et techniques, auxquelles il fut décidé de répondre par le choix d'un mixage simple (pour ne pas perturber les sujets), partant des techniques acquises pour la stéréo puis étendues au 5.1 ; il fallait cependant explorer, en parallèle, toutes les ressources de l'espace binaural en vue du test. Ainsi, dans le cas du stimulus musique actuelle, un mixage globalement L-R (gauche-droite) fut tout d'abord réalisé, sur enceintes. Ce mixage fut ensuite étendu au 5.0 discret, en ajoutant donc un centre et des arrières. Cette étape fut réalisée en étudiant les mixages multicanaux déjà sortis dans le commerce, et d'après les conseils d'Hervé Déjardin et du compositeur du morceau, Baptiste Palacin. Les arrières étaient ainsi nourris avec des instruments complémentaires de ceux de la façade avant (guitares rythmiques à l'arrière répondant à celles de l'avant), et de la réverbération. Les instruments étaient souvent mixés avec une réverbération d'espace avant, et une autre, un peu retardée et mise hors-phase, panoramiques gauche-droite inversés, envoyée vers les canaux arrière Ls et Rs, et dosée à l'oreille, suivant une technique suggérée notamment par Bernard Lagnel [12]. Enfin, une fois l'équilibre du mixage à peu près mis en place, trois instruments furent jugés pertinents pour le test (ceux que les sujets devraient localiser au sein du stimulus) : la **voix**, la **guitare mélodique**, la **caisse claire**, et leur localisation fut rendue suffisamment précise (la réverbération de la voix fut recentrée, des chorus de guitare trop confus furent enlevés, la caisse claire fut montée en niveau en traquant ses reprises dans les autres pistes de la batterie, qui auraient pu induire des sources fantômes).

Comme pour les autres stimuli, le placement des sources à localiser dépendait soit des conventions de mixage les plus classiques (c'est ainsi qu'aucun instrument n'était spatialisé sur les arrières pour le stimulus musique classique), soit d'une volonté de les répartir au mieux dans l'espace (ainsi pour les ambiances, pour le bruit rose et pour les voix).

Pour ce stimulus comme pour les autres, la réalisation consista donc en un mixage minutieux progressivement étendu au 5.0 en évitant tout effet susceptible de perturber l'écoute. Ainsi fut obtenu le mixage de référence, qui allait

être binauralisé. Par la suite, ce mixage fut repris en opérant deux autres spatialisations des sources (souvent une spatialisation « conventionnelle », et une autre « non-conventionnelle »), afin d'obtenir deux autres versions « leurres » pour chaque stimulus 5.0, en vue de la première phase du test.

Stimulus musique classique (nom de code « classique »)

La difficulté avec ce stimulus était de trouver un enregistrement qui comprenne suffisamment de pistes pour permettre de manipuler précisément la spatialisation des sources en évitant des sources fantômes. Après quelques tentatives malheureuses, notamment des enregistrements d'orchestres avec une microphonie trop restreinte, fut récupéré, par le biais de Brian Katz, un multipistes de laboratoire : un extrait d'un opéra de Mozart, avec soprano, clarinette, flûte, basson, cors, et cinq pistes de cordes, enregistré en chambre anéchoïque, instrument par instrument (enregistrements réalisés par l'université Paris Sud). Cette configuration permettait théoriquement une libre manipulation de la spatialisation des sons ; toutefois, elle ne permettait pas d'aborder la problématique de la captation globale, qui met en jeu des relations de phase dont les conséquences en binaural pouvaient être intéressantes. Hervé Déjardin, de Radio France, estima pour sa part pertinent d'utiliser ce multipistes, à partir du moment où d'autres stimuli emploieraient des prises de son faites au couple à capsules noncoïncidentes : c'est par leur biais qu'on aborderait la problématique des relations de phase.

Le travail s'opéra donc à partir de ce multipistes anéchoïque, dont le mixage posa plusieurs problèmes. Le souffle de l'enregistrement, cumulé sur toutes les pistes, était très gênant, ce qui nécessita un denoising. Puis a commencé un travail de réverbération : après quelques essais, une acoustique fut recrée par la combinaison d'une réverbération « de largeur » des sources, avec un decay court (employant la ReaVerbate de Reaper), différente selon la source, et d'une réverbération d'espace avant, plus longue, constituée par un préset légèrement modifié de Chamber, de la réverbération hardware Bricasti ; l'acoustique était complétée par une réverbération arrière constituée par une réverbération identique, mais délayée de 50 ms, mise en opposition de phase, et inversée gauche-droite, puis envoyée dans les arrières, comme pour le stimulus musique actuelle. Le dosage des sources en niveau et de nombreuses égalisations permirent de donner aux sources un côté brillant et précis pour éviter de les noyer dans la réverbération. Afin de donner la sensation définitive d'un lieu et nourrir davantage les arrières, un bruissement de public et un son de toux réverbéré, enregistrés lors de l'entracte d'un concert dans une église, y furent envoyés.

Pour ce stimulus, la localisation de la flûte, du basson, et de la voix, seraient demandés aux sujets. La clarinette fut en revanche retirée à cause de son timbre trop proche de la flûte.

Stimulus ambiance (nom de code « ambiance »)

Ce stimulus était constitué de chants d'oiseaux enregistrés dans les sous-bois, au point du jour. L'équipement employé, deux Oktava MK012 avec capsule

cardioïde disposés en couple ORTF, permettaient d'introduire une corrélation de phase. Ces sons furent répartis dans l'espace 5.0. Un corbeau, un craquement de branche et un mouvement de buisson enregistrés au même moment, et avec le même système microphonique, furent ajoutés comme signaux à localiser.

Stimulus voix parlées (nom de code « voix »)

L'idée était ici d'enregistrer un texte dit par quatre voix réparties dans l'espace sonore : une voix d'homme mûr, d'homme jeune, de femme mûre, de femme jeune (respectivement joués par Pascal Chédeville, Pierre Hugonnet, Nathalie Palazot et Margot Castel, employés, apprenti ou stagiaire à DMS). Les enregistrements ont été réalisés dans le studio de DMS, avec un bruit de fond variable (la climatisation ne pouvant être contrôlée complètement). Le microphone employé était un modèle OceanWay, connecté à un préampli Alto Esotar et de là, à une carte son Avid HD I/O. Le texte choisi était neutre, sans dramaturgie, pour ne pas distraire l'auditeur dans son analyse de la localisation ; en l'occurrence, il s'agissait d'un extrait du poème *Le hareng saur* de Charles Cros [18].

Les comédiens enregistrèrent les deux premières strophes du poème, qui furent montées de façon que les quatre voix s'y relaient, sans logique avec la construction du texte ; l'objectif étant que l'auditeur ne puisse prévoir les changements de voix. Chacune d'elle était spatialisée en un endroit fixe de l'espace sonore.

On pourra ici nous reprocher de n'avoir pas choisi un stimulus représentatif d'une prise de voix concrète, c'est-à-dire animée par une dramaturgie ou une narration, et accompagnée d'une réelle construction sonore. Pour ce travail, l'hypothèse fut cependant émise que si des altérations de localisation n'étaient pas sensibles à l'écoute d'un texte sans dramaturgie et sans ambiances, elles ne le seraient sans doute pas davantage dans un texte mélangé à d'autres sons et porteur d'une intrigue, celle-ci retenant l'attention de l'auditeur.

FIGURE 2.2 – Une cabine speak légère pour s'isoler des bruits environnants. La qualité de l'enregistrement des voix a probablement souffert d'un environnement mal adapté.

FIGURE 2.3 – Un « hamac acoustique » qui permettait d'étouffer le bruit de la ventilation.

Stimulus bruit rose (nom de code « bruit rose »)

Celui-ci fut réalisé à l'aide du plug-in JS :Liteon/pinknoisegen sous Reaper : plusieurs pistes de bruit rose (non-dupliqué, donc différent), à niveau égal, alimentaient les 5 enceintes, en fond sonore. Trois salves monophoniques de bruit rose, toutes de même niveau et spatialisées en trois endroits différents fixes, devaient être localisées par le sujet.

Procédure de mixage et spatialisation

Chaque stimulus fut mixé en 5.0 discret (ITU), la spatialisation étant faite au moyen du logiciel SpherAudio en mode VBAP 5.0 (voir les pages consacrées au fonctionnement du logiciel). Ces mixages 5.0 furent ensuite binauralisés en basculant SpherAudio en mode binaural, avec les paramètres suivants :

- HRTF employée : Best Matching 2 ;
- égalisation gauche-droite (L/R Equalization) activée ;
- Room activée en ordre 3 (sauf pour le stimulus bruit rose où elle était désactivée, de manière à obtenir un stimulus de laboratoire anéchoïque) ;
- niveau de la room à 0.2 (valeur trouvée empiriquement comme étant la plus efficace pour l'acoustique et la plus transparente du point de vue du timbre, pour l'ensemble des stimuli).

Lors des exports le niveau était adapté, sur le master, afin que le son binauralisé et le son en 5.0 discret aient sur le Vu-mètre un niveau moyen équivalent au dB près.

2.2.2 La mise en place du studio

En arrivant à DMS l'un des objectifs de l'auteur de ce mémoire était de contribuer à la mise en place d'un studio, équipé *in fine* en 28.2 (L, C, R, Ls, Cs, Rs, inter-L, inter-R, top-L, top-C, top-R, top-Ls, top-Cs, top-Rs, bas-L, bas-R, wide-L, wide-R, wide-Ls, wide-Rs, top-wide-L, top-wide-R, top-wide-Ls, top-wide-Rs, 4 enceintes zénithales au plafond, 2 subwoofers), permettant l'enregistrement, le mixage et la diffusion, avec à terme le projet d'intégrer également un écran pour une projection vidéo, le tout dans un bureau de 5.20 x 7.40 x 2.30 m.

La station de travail comprenait un Mac Pro, des cartes d'entrées-sorties Avid HDIO et HD MADI, et les logiciels Pro Tools, Cubase et Reaper (SpherAudio ne fonctionnant que sur ce dernier logiciel) (voir le synoptique définitif du studio en annexe A.1, aménagé pour le mixage en 5.0 et les expériences).

Le mixage s'opéra tout d'abord en 5.0 cinéma, quand les premières enceintes furent installées, puis dès que possible, elles furent replacées en **5.0 ITU** (fig. 2.6) (en accord avec les prérequis du mode VBAP), et elles firent l'objet d'une égalisation par un processeur cinéma Trinnov 24+. Le mixage fut retouché afin de conserver les équilibres dans l'installation définitive de l'expérience. Les enceintes étaient situées à environ 1.20 m du sol (distance recommandée par l'ITU BS1116 [21]), et leur pavillon situé à 1 m des murs (sauf pour les enceintes arrière pour

FIGURE 2.4 – Le studio 28.2 de DMS, à un stade avancé de l'installation.

FIGURE 2.5 – L'espace de travail.

lesquelles ce ne fut pas possible au vu des dimensions de la pièce). Notre cercle ITU avait un rayon de 2.23 m.

En situation de calme (personne ne discutant dans le couloir, pas de voiture passant sur le parking), le **bruit de fond** mesuré dans le studio était de **34 dBspl**. La rareté de la circulation automobile, et le silence recommandé aux employés de DMS quand des expériences avaient lieu, font que cette mesure peut probablement être considérée comme valable pendant les tests.

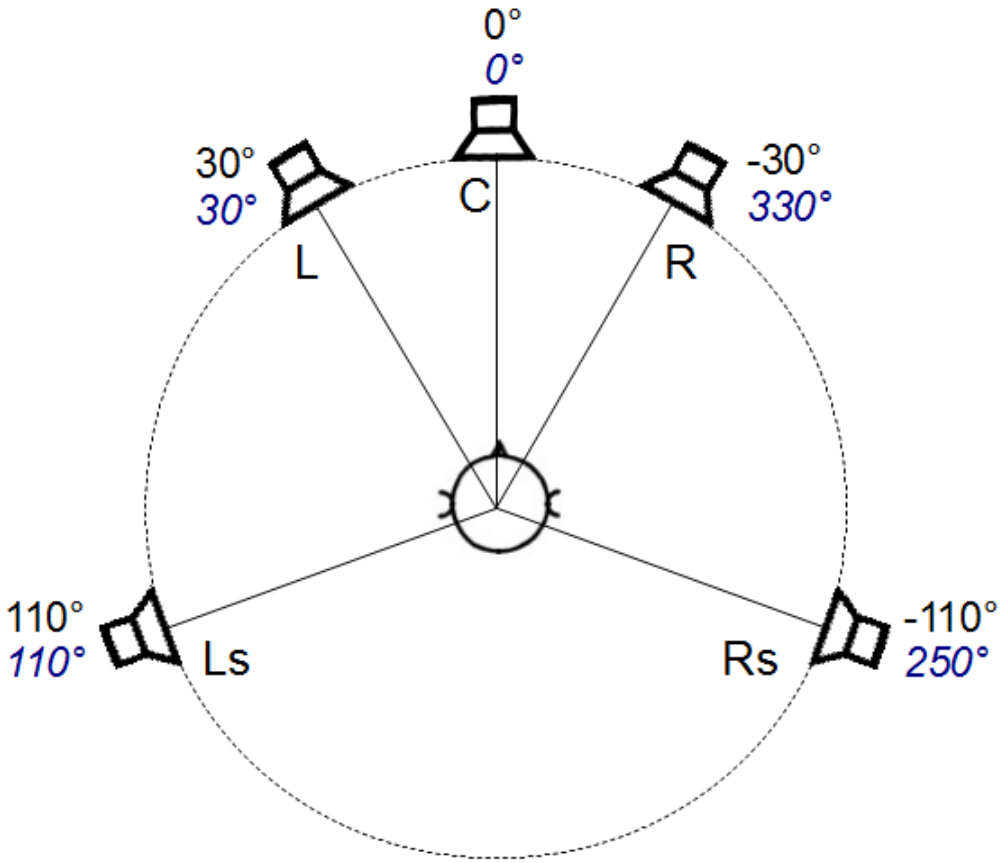


FIGURE 2.6 – Rappel de la configuration 5.0 ITU.

Couper l'influence du visuel

Le mur gauche était orné d'une rangée de fenêtre dont les stores ne coupaient pas toute la lumière, or pour éviter d'influencer les sujets on désirait leur masquer la vue des enceintes, et donc plonger la pièce dans le noir : on fixa donc des rideaux noirs sur les fenêtres, à l'aide de gaffeur (voir fig. 2.7). Pour donner aux sujets un repère constant leur indiquant l'« avant », Lucas McCauley et Mathieu Devillers, de DMS, soudèrent une LED rouge que l'on pouvait brancher sur secteur au moyen d'un transformateur 12V ou 9V. Cette LED fut disposée juste au-dessus de l'enceinte centrale, un peu en retrait par rapport au pavillon,

et exactement à 0 degré d'azimut. La puissance de la LED n'était pas suffisante pour éclairer la moindre enceinte.

Le dispositif fut arrangé de telle manière que, sauf à dévier la lumière de la lampe qui éclairait leur feuille, les sujets ne pouvaient pas distinguer les enceintes tandis qu'ils effectuaient l'expérience. La **pénombre** obtenue semble donc suffisante pour estimer que, sauf accident, les sujets n'ont pas été influencés au cours du test par la vision des enceintes.

Test d'étalonnage lumineux

Dans l'optique d'étalonner les réponses des sujets, un **test introductif lumineux** fut mis en place, qui avait lieu juste avant le test sur enceintes. Il s'agissait de demander au sujet de localiser, sur un schéma-réponse similaire à tous ceux qu'il allait avoir par la suite, cinq points lumineux dans l'espace autour de lui. Ces points lumineux étaient projetés au moyen d'un laser rouge manipulé par l'auteur de ce mémoire, sur cinq repères de moins de trois centimètres de largeur chacun, disposés respectivement à -15 degrés/265 cm, 30 degrés/219 cm, 55 degrés/322 cm, 155 degrés/211 cm et -88 degrés/260 cm par rapport au « sweet spot » (voir fig. 2.7). Cette opération permettait d'évaluer la capacité du sujet à retranscrire, sur son schéma, l'azimut qu'il visualisait. Ainsi, si par exemple un sujet plaçait le point situé à 30 degrés, à 75 degrés sur son schéma, nous pouvions légitimement penser qu'il aurait tendance à exagérer de même les positions gauche et droite des sons qu'il entendrait lors de leur représentation.

La diffusion des stimuli

Au terme de la création des stimuli, nous avions donc quinze stimuli pour le système discret, soit cinq types de stimulus ayant subi chacun **trois mixages** (trois spatialisations) différentes :

- stimulus « ambiance » (numéroté 1), mixages M1, M2, M3
- stimulus « musique actuelle » ou « rock » (numéroté 2), mixages M1, M2, M3
- stimulus « voix » (numéroté 3), mixages M1, M2, M3
- stimulus « musique classique » (numéroté 4), mixages M1, M2, M3
- stimulus « bruit rose » (numéroté 5), mixages M1, M2, M3

Vingt-cinq stimuli étaient également obtenus pour le système binaural, soit cinq types de stimuli sous quatre formes chacun : **original (A)**, **compression antenne (B)**, **réduit en MP3 192 (C)**, **réduit en AAC 192 (D)**, plus une rediffusion de l'original comme **référence cachée (E)**. Soit en tout :

- stimulus « ambiance » (numéroté 1), A, B, C, D, E
- stimulus « musique actuelle » ou « rock » (numéroté 2), A, B, C, D, E
- stimulus « voix » (numéroté 3), A, B, C, D, E
- stimulus « musique classique » (numéroté 4), A, B, C, D, E
- stimulus « bruit rose » (numéroté 5), A, B, C, D, E

Ces stimuli étaient organisés en **6 séries** pseudo-aléatoires : séries **1, 2, 3** pour les **stimuli discrets**, séries **4, 5, 6** pour les **stimuli binauraux**. En réalité

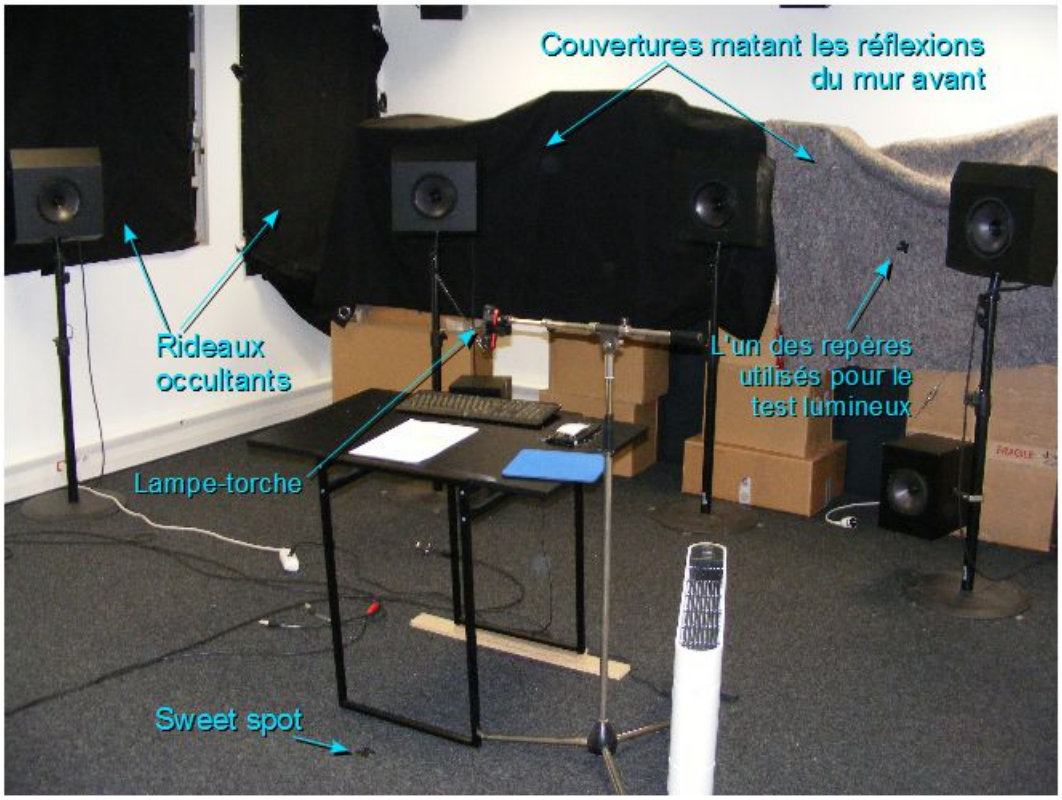


FIGURE 2.7 – Le studio en configuration de test. Noter les couvertures face au mur avant, les rideaux occultant la lumière des fenêtres, la table surélevée pour éviter des réflexions de la lampe sur les enceintes. Le sujet était assis devant la feuille, sur un fauteuil en plastique au dossier troué. Il manque ici l'amplificateur casque, et le casque, connectés aux câbles longeant la table. Noter enfin que certaines enceintes à l'image ne font pas partie du 5.1 ITU.

ces séries avaient été définies en agençant les stimuli dans un ordre quelconque, selon le principe du carré latin, respectant cependant quelques règles :

- les interventions des stimuli étaient à peu près équilibrées dans le temps
- pour les séries binaurales, la première exposition de chaque stimulus n'était pas en traitement B (compression antenne), afin de ne pas trop déstabiliser l'auditeur par un stimulus très dégradé
- les références cachées survenaient en fin de série.

Pour chaque série, une session Reaper a ensuite été élaborée, dans laquelle les stimuli étaient insérés dans leur ordre « aléatoire ». Chaque session débutait par un commentaire d'introduction enregistré rappelant le fonctionnement du test (texte reproduit en annexe A.2). Pour les séries en système discret, ce commentaire était suivi d'une injonction à garder la tête la plus immobile possible. Pour les séries en binaural, il était suivi d'une courte initiation au binaural : une voix faisait le tour de la tête de l'auditeur dans le casque en partant de la droite, en signalant sa position en terme d'heure : « Trois heures, quatre heures... » jusqu'à « douze heures » (le texte était dit par Cloé Chope, de l'école Louis Lumière).

Ensuite, chaque stimulus était précédé d'une voix rappelant son appellation : par exemple « 12M1 » pour « série 1, stimulus 2 (« rock »), mixage 1 » ou encore « 65C » pour « série 6, stimulus 5 (« bruit rose »), traitement C (mp3 192) ». Ces appellations, reportées sur les feuilles-réponses, permettaient à l'auditeur de vérifier que celles-ci s'enchaînaient dans le bon ordre et qu'il n'y avait pas d'erreur. Dans le cas des stimuli « rock », classique, et ambiance, cette annonce était toujours suivie par des **extraits des sources à localiser**, mises en **solo** et placées au **centre**, précédées de leur nom (par exemple : « Vous devrez localiser la flûte » - extrait de flûte - « le basson » - extrait de basson ; la voix était jugée trop reconnaissable pour nécessiter un solo). Ce procédé devait permettre de s'assurer que les sujets identifiaient bien les sources à repérer (tous les auditeurs n'étant pas forcément familiers du son d'un basson, ou d'une flûte). L'annonce du stimulus et sa diffusion étaient séparées par une seconde, et la fin du stimulus et l'annonce suivante par deux secondes.

Pour être importés dans la session Reaper, il est à noter que les fichiers en AAC 192 devaient être reconvertis en WAV 44.1 kHz, 24 bits (opération effectuée avec Quick Time Pro 7), ce qui normalement ne changeait en rien leur contenu fréquentiel. En revanche, Reaper acceptant de lire le mp3 (et utilisant lui-même Lame encoder), il n'y avait pas de problème pour importer les fichiers MP3 tels quels.

Les sujets écoutaient le test binaural au moyen d'un casque Sennheiser HD650 branché à un ampli casque Roll. Ce casque fut choisi pour sa réponse en fréquence relativement plate, notamment dans l'aigu : cela permettait au sujet de disposer du maximum d'informations fréquentielles dans son jugement de la localisation, et de donner ainsi « toutes ses chances » au binaural.

La disposition employée était la suivante : l'auditeur, placé au sweet spot sur un siège (dossier au niveau de la nuque, en matière plastique souple et percée), se trouvait face à une table lui présentant un clavier, et une souris sans fil. L'écran avait été transporté dans le nodal pour permettre, sans déranger le sujet, de surveiller le bon déroulement du test. Avec la barre d'espace, le sujet pouvait actionner le mode play/pause et ainsi réguler la lecture comme il le souhaitait, tandis que les touches Ctrl + flèche de gauche et Ctrl + flèche de droite lui permettaient de ré-écouter un stimulus ou de passer les instruments mis en solo. Le curseur de la souris avait été placé sur le master de sortie de Reaper, et la souris avait été surélevée et scotchée sur un support, de façon que l'auditeur ne puisse la déplacer. En actionnant la molette tactile de la souris, vers le haut ou vers le bas, il pouvait cependant respectivement monter, ou baisser le volume général (master), et ainsi contrôler son niveau d'écoute à tout moment. La molette tactile était complètement lisse et sans aucun repère visuel, de sorte que le réglage du niveau se faisait ainsi uniquement à l'oreille.

Débat : le niveau d'écoute réglable et la tête mobile

On pourra arguer qu'avoir laissé le sujet libre de régler lui-même à tout moment son niveau d'écoute, pénalise la rigueur du test. Cette option fut cependant choisie pour privilégier le confort du sujet, dans une expérience déjà longue et pénible. Fixer le niveau d'écoute présentait un risque : trop fort, il risquait de stresser et fatiguer inutilement le sujet. Trop faible, il pouvait le priver d'éléments importants dans son jugement (la perception des hautes fréquences par exemple, pour un sujet âgé). Tous les fichiers audio d'une même session de test étaient simplement mixés à un niveau équivalent sur le peak-mètre de Reaper.

Néanmoins, le niveau d'écoute a été observé et noté sur le master de Reaper depuis l'écran du nodal, pendant que les sujets passaient l'expérience. Or dans la quasi-totalité des cas, à DMS du moins, les sujets se fixaient en début de session un niveau d'écoute qu'ils ne retouchaient pratiquement plus par la suite (et rarement de plus de quelques dB). Et cependant les niveaux choisis variaient sensiblement d'une personne à l'autre (d'un ambitus que l'on peut estimer à une dizaine de dB). Cela peut laisser penser qu'il était effectivement préférable de laisser les sujets régler eux-mêmes leur niveau d'écoute pour leur confort, un niveau imposé ayant statistiquement risqué de gêner une majorité d'entre eux.

C'est pour la même raison que la tête de l'auditeur n'a pas été fixée : Blauert ([3], p.95) considère qu'une simple injonction à garder la tête droite suffit pour conserver la rigueur d'un test perceptif, mais uniquement pour des stimuli courts (moins d'une seconde). Au-delà, il préconise l'emploi d'un support pour la tête des sujets. Toutefois sur un test d'une quarantaine de minutes, un tel dispositif aurait pu accentuer la fatigue de l'auditeur, aussi le risque a-t-il été pris de lui faire confiance quant à la position de sa tête (qui ne pouvait être vérifiée en cours de test). Une recommandation à conserver la tête fixe a donc semblé être le meilleur compromis entre rigueur du test et confort du sujet.

2.2.3 Le mode d'évaluation

Le mode de réponse a été choisi de manière à brimer le moins possible le sujet, en accord avec les cours d'Étienne Hendrickx. A la suite de nombreuses discussions et de plusieurs essais, fut choisi un **mode de réponse par dessins sur papier**. Très simple à mettre en œuvre, le test consisterait dès lors, pour le sujet, à tracer, sur un espace vu de dessus figuré sur papier par des cercles concentriques, les limites de la zone où il croyait localiser le son. Il indiquerait lui-même à quelle source correspondaient ses différents tracés. Ainsi, il était complètement libre de délimiter la zone qui lui paraissait la plus pertinente.

L'analyse des résultats en revanche, promettait d'être beaucoup plus fastidieuse : on mesurerait les angles délimitant la figure obtenue, son éloignement par rapport au point central représentant la tête (point le plus proche et le plus éloigné), et sa largeur. En marge de chaque cercle, seraient indiquées les sources à localiser, tandis que plusieurs **échelles** devaient permettre d'évaluer l'appréciation du sujet concernant :

- le timbre (coloration : grave/aiguë) ;
- la précision de la localisation ;

- l'appréciation générale de l'enregistrement ;
- le sentiment d'immersion ;
- la lisibilité.

Ces critères seraient notés en plaçant une croix sur une échelle allant de 0 à 7 (voir un modèle de feuille-réponse en annexe A.5). Cette partie du test était notamment inspirée des travaux de Nick Zacharov [8] sur le vocabulaire et l'évaluation en test subjectif.

Un rappel des définitions, et des touches de clavier utiles, était imprimé sur une feuille posée à côté du clavier devant le sujet. Une feuille-réponse type se trouvait au même endroit, donnant un exemple de réponse pour aider le sujet à appréhender les schémas (ces deux documents sont reproduits en annexes A.3 et A.4).

On remarquera qu'un tel mode d'évaluation ne permettait pas d'étudier le sentiment d'**élévation**. Pourtant un sentiment d'élévation peut se faire sentir selon le contenu fréquentiel des sources, en 5.0 discret comme en 5.0 binauralisé (voir les bandes directionnelles [3]). Une telle étude a cependant été jugée trop compliquée : par commodité autant que par souci de rigueur, l'étude de l'azimut a donc été privilégiée.

Dans le cas de l'**évaluation du (des) HRTF** le(s) plus adaptée(s) à chaque sujet, ce sont une mise en place et un mode d'évaluation légèrement différents qui ont été préférés. 8 HRTF ont été considérées parmi celles proposées par SpherAudio (Best Matching 2, utilisée pour les stimuli du test principal, et Min Subset 1 à 7). Un bruit rose continu a d'abord été binauralisé en un mouvement de tour de tête : il partait de la droite (-90 degrés), passait derrière la tête (-180 degrés) et revenait devant (0 degré), en un mouvement homogène, régulier, d'une durée de 30 secondes environ. La binauralisation était effectuée pour chacune des HRTF retenues (égalisation L/R activée, Room désactivée), et 8 séquences sonores étaient donc obtenues en tout. L'affinité des sujets avec chacune des HRTF concernées, serait mesurée en leur demandant de dessiner, sur le même schéma circulaire utilisé tout au long de l'expérience, le mouvement que chaque bruit rose effectuait selon eux. Le résultat attendu était donc un cercle qui partait de leur droite et passait derrière leur tête pour revenir devant.

Pour des raisons de durée, il fut décidé de placer ce dernier volet de l'expérience au terme de la séance en 5.0 (la plus courte) : une voix demandait alors au sujet de placer un casque sur sa tête, et les huit séquences sonores s'enchaînaient, énumérées au fur et à mesure par la même voix pour que le sujet s'y repère. Il avait la possibilité de ré-écouter les séquences autant de fois qu'il le souhaitait. Au terme de l'expérience étaient donc obtenus huit feuillets supplémentaires indiquant les mouvements ressentis par le sujet et d'éventuels commentaires.

2.2.4 Les Pré-tests

Conformément aux conseils donnés notamment par Étienne Hendrickx et Gérard Pelé, une fois le protocole de test suffisamment avancé, quelques **pré-tests**, partiels (une partie sur les deux) ou complets (deux parties), avaient été réali-

sés. L'auteur de ce mémoire, mais aussi Alexandre Dazzoni et Pierre Hugonnet, travaillant à DMS, avaient ainsi pu éprouver le bon fonctionnement de l'expérience, et entraîner des améliorations significatives du procédé : rappel régulier des solos, simplification du déroulement de l'expérience, etc. Ces pré-tests ont eu une importance primordiale dans la bonne mise en marche de l'expérience, et donc, dans la rigueur de ses résultats.

2.3 L'expérience

Du 2 avril au 12 avril 2013, les tests ont été menés au sein de l'entreprise Digital Media Solutions, dans le studio aménagé à cet effet. 20 sujets ont suivi l'expérience complète, un sujet l'expérience complète moins l'étalonnage lumineux (pour des raisons de faisabilité), un autre a suivi l'expérience complète moins l'évaluation des huit HRTFs (pour cause de fatigue), un sujet a suivi la partie sur enceintes du test, avec la partie au casque prévue à l'école Louis Lumière (pour des raisons logistiques) et un sujet a suivi la partie binaurale uniquement du test (pour des raisons d'emploi du temps).

2.3.1 Le déroulement des expériences

Assez vite, un rythme de croisière des expériences a été mis en place, qui se résumait en trois phases : trouver des sujets, mener les expériences proprement dites, exploiter les résultats.

Comme il a été mentionné, la majorité des sujets étaient des employés ou stagiaires de DMS. Chaque sujet participant recevait à la fin du test une boîte de chocolats, dont la promesse avait pour but de les motiver - décision basée sur le constat que les sujets sont plus concentrés lorsqu'ils gagnent quelque chose à l'issue du test. Quatre sujets extérieurs à DMS se se sont également présentés.

Chaque sujet se voyait expliqué oralement un minimum de choses au moment de s'installer dans la pénombre, notamment l'information qu'une voix préenregistrée lui apporterait tous les détails nécessaires. Le sujet était placé au sweet spot, en lui indiquant qu'il se trouvait alors à sa position de référence (dans le cas du test sur enceintes). La recommandation lui était faite de bien lire les feuilles qu'il avait à sa disposition sur la table, et d'appeler en cas de problème. S'il n'avait pas de questions, il pouvait commencer. L'opérateur se rendait alors dans le nodal voisin pour suivre sur l'écran le déroulé de l'expérience.

L'écran déplacé dans le nodal, et une souris complémentaire branchée sur l'unité centrale, permettaient de surveiller où en étaient les sujets, éventuellement repérer une erreur dans la session, et intervenir rapidement si un problème se présentait. Lorsqu'un sujet avait une question ou rencontrait un problème, l'intervention pouvait ainsi être immédiate. Lorsqu'un sujet avait fini, l'opérateur lui posait une série de questions :

- Tout s'était-il bien passé ?
- Avait-il ressenti une sensation de fatigue, et si oui, à partir de quel moment ?

– Selon sa manière d'indiquer la provenance des sons (une zone, un nom, une croix), quel point de référence constituait le centre de gravité du son qu'il avait entendu ? (le centre de la zone ? Le centre du mot ? L'initiale ? Le chiffre, par exemple dans "Voix 3" ?)

Ensuite, l'intégralité de sa feuille-réponse était rapidement observée, pour vérifier qu'il n'avait rien oublié, l'interroger sur des réponses qui n'étaient pas claires, etc. Puis l'heure de début et de fin du test était notée sur la feuille, ainsi que le niveau d'écoute, sur le master fader de Reaper (ainsi que les éventuelles évolutions de ce niveau remarquées en cours de test, mais qui restaient généralement très faibles, comme déjà mentionné). Lorsqu'il s'agissait de la deuxième phase du test, le sujet se voyait remettre une boîte de chocolats avant de s'en aller.

Le temps mis par les sujets pour suivre l'expérience variait beaucoup : si les sujets les plus rapides ont pu mettre 25 minutes pour une partie, d'autres mettaient 55 minutes, voire une heure et demie (l'auteur a personnellement mis 35 minutes pour faire l'expérience au casque). Au-delà d'une heure, l'opérateur intervenait pour s'assurer qu'il n'y avait pas de problème, et leur demandait s'ils n'étaient pas fatigués : à chaque fois ils ont répondu par la négative, expliquant qu'ils avaient l'habitude d'écoutes longues, et il pouvaient alors continuer jusqu'au bout. Globalement, l'expérience sur enceintes était pour eux plus courte, d'environ 10 minutes, que celle au casque.

2.3.2 Le relevé des résultats

Dès le début, il était clair que l'exploitation des résultats sur papier serait longue et fastidieuse. Il fut décidé de relever des éléments bien précis : à savoir, sur les dessins, les angles significatifs (aux extrêmes, dans le sens trigonométrique, et aux centres des zones dessinées par les sujets, ce que l'on nomma angle min, angle max, angle centre de gravité), et les distances significatives (la plus proche, la plus lointaine, la distance du centre de gravité) (fig. 2.8). Les valeurs choisies sur les échelles par les sujets furent également relevées, et les erreurs de localisation réalisées furent observées globalement, en vue de dégager des axes d'analyse des résultats.

Ce travail se faisait au rapporteur, à la règle et au crayon de papier. Il fallait compter en moyenne 45 minutes pour le relevé complet d'une liasse de feuilles-réponses. Pour l'analyse statistique des résultats, via Matlab, il fallait les entrer en traitement de texte : la recopie des données prenait encore environ vingt minutes par liasse. Ces travaux de relevé furent menés autant que possible en parallèle des tests, dans le but de prendre de l'avance et de ne pas perdre de temps. Ils furent néanmoins trop longs pour me permettre d'exploiter les données récoltées par l'expérience à l'école Louis-Lumière.

2.3.3 L'expérience à Louis-Lumière

Les samedi 13, lundi 15, mardi 16, mercredi 17 et samedi 20 avril 2013, une série de tests subjectifs furent menés dans le studio radio de l'école Louis-Lumière. L'objectif de cette deuxième volée de tests était double : tout d'abord,

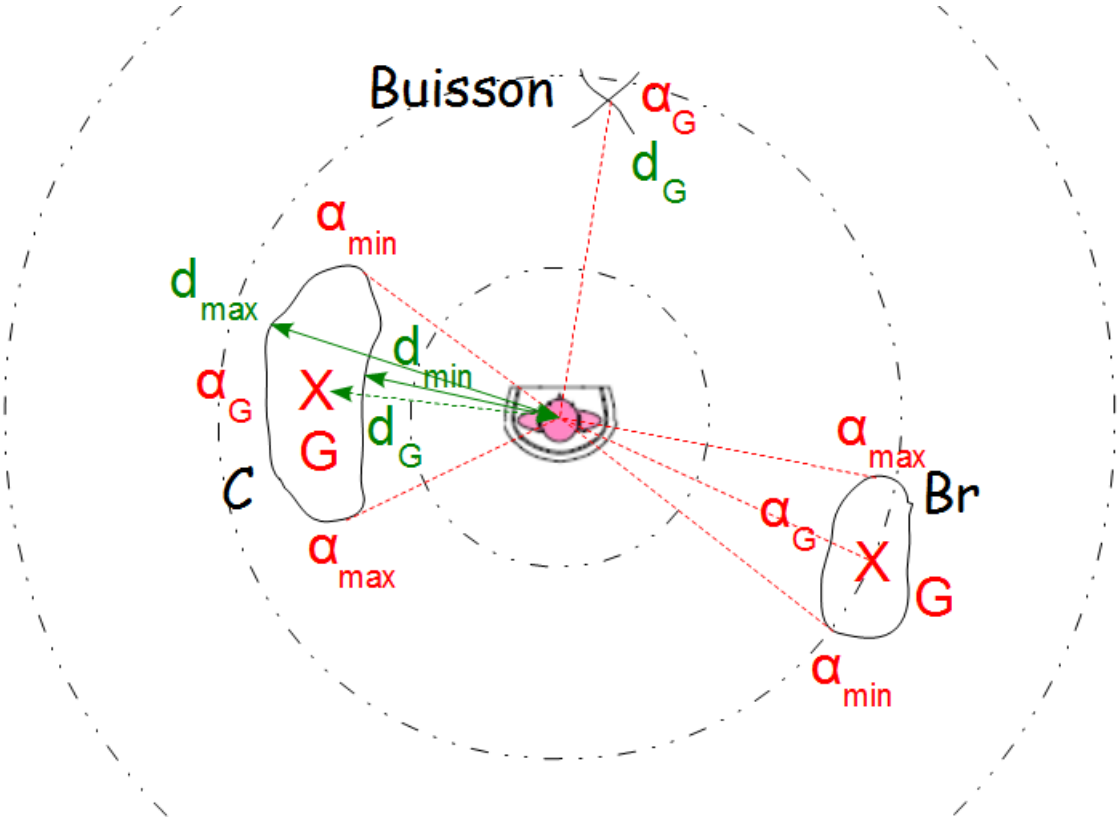


FIGURE 2.8 – Un exemple de relevé des résultats sur une feuille-réponse. « G » : estimation du centre de gravité de la figure. Son relevés les angles extremum (alpha min et max), les distances extrêmes (d min et max) et l'angle et la distance du centre de gravité (alpha G et dG). Pour le buisson, les trois angles et les trois distances se confondent.

réunir davantage de sujets, notamment en permettant aux étudiants de l'école de suivre les tests sans avoir à se rendre au siège de DMS à Noisiel. Ensuite, augmenter plus précisément la proportion de sujets « experts », en faisant suivre le test à de futurs ingénieurs du son, donc des personnes ayant une certaine expérience de l'écoute et disposant d'une audition *a priori* correcte.

Bien entendu, cette deuxième série était vouée à être traitée, dans un premier temps, séparément de la première, les conditions expérimentales n'étant pas les mêmes, et notamment en ce qui concerne le système d'écoute. Mais dans le cas où la différence de lieu ne semblerait pas avoir une influence significative sur les réponses obtenues, l'idée était de pouvoir mélanger les résultats des deux séries pour en tirer des conclusions plus précises. Comme mentionné plus haut, les résultats de l'expérience menée dans le studio radio de Louis Lumière ne purent pas être exploités dans le cadre de ce mémoire, principalement en raison du temps que prenaient les relevés. Leur mise en place ne sera donc pas décrite ici, si ce n'est pour l'évaluation des HRTF, pour lesquelles les données de l'école furent conservées : les sujets, plongés dans la semi-pénombre (mais qui n'empêchait pas tout à fait de distinguer les enceintes), disposaient en effet du même casque, connecté sur le même amplificateur casque qu'à DMS, relié à deux sorties ligne

de la DM2000 via le patch analogique. Les conditions expérimentales ont donc paru suffisamment proches sur cette partie du test pour permettre d'inclure les résultats sur HRTF dans le relevé général.

2.4 L'exploitation des résultats

L'exploitation des résultats s'est faite intégralement au moyen du logiciel Matlab, grâce à un script développé conjointement avec Matthieu Aussal, avec l'aide de Mathieu Coïc et d'Alexandre Dazzoni, de DMS. Au vu de la masse de données récupérées lors du test, la première étape a été de définir les objectifs de l'analyse par rapport au temps restant, et de chercher le moyen le plus simple et le plus efficace de représenter les résultats pour en tirer des conclusions. L'étude des réponses pour les mixages M2 et M3 de la session de test sur enceintes fut laissée de côté, principalement par manque de temps. De même, l'analyse des stimuli passés par la compression antenne fut écartée : tous ces éléments, quoique intéressants, auraient demandé trop de temps de travail par rapport aux délais du mémoire, et il fut décidé de se concentrer sur les traitements qui rentraient directement dans le cadre de la problématique de ce travail.

L'analyse des résultats porterait donc sur les éléments suivants du test :

- localisation des points lumineux lors du test d'étalonnage (« lum ») ;
- localisation en 5.0 discret, mixage 1 (mixage principal) (« HP ») ;
- localisation en binaural référence (« binoRef ») ;
- localisation en binaural référence cachée (« binoRefCach ») ;
- localisation en binaural AAC (« binoAAC ») ;
- localisation en binaural MP3 (« binoMP3 ») ;
- évaluation des échelles : précision, immersion, lisibilité, timbre/coloration, appréciation ;
- évaluation de la préférence des HRTF.

Cette analyse allait s'accompagner de modes de représentation spécifiques.

Des schémas-réponses pour visualiser

Le support utilisé pour la représentation de la localisation, fut constitué de graphiques ayant pour fond le même schéma circulaire que celui présent sur les feuilles-réponses des sujets, copié à la même échelle. La seule modification apportée par rapport au schéma d'origine fut l'ajout de la position des cinq **enceintes** du système d'écoute, représentées à l'échelle par rapport au personnage-repère, afin de visualiser plus facilement l'éventuelle influence des hauts-parleurs. Grâce à Matlab, on superposerait sur ce schéma une modélisation des zones dessinées par chaque sujet, à l'échelle, ce qui devait permettre de visualiser de façon globale et immédiate sa perception de l'espace.

Le premier type de représentation choisi consistait à représenter les centres de gravité des figures dessinées par les sujets, ainsi que le centre de gravité moyen (voir fig. 2.8). Un schéma pouvait ainsi être obtenu pour chaque stimulus

de chaque traitement (« voix binoAAC » par exemple), avec une couleur différente par stimulus, afin de comparer très rapidement la répartition entre deux traitements sur un même stimulus.

Des ellipses pour représenter

La seconde étape consistait à représenter les zones dessinées par les sujets. L'observation des feuilles-réponses avait montré que dans leur grande majorité les sujets représentaient les zones où ils situaient les sons au moyen d'**ellipses** (fig. 2.9), ou de cercles (cas particulier d'ellipse avec deux foyers confondus), ou de points (ellipse nulle).

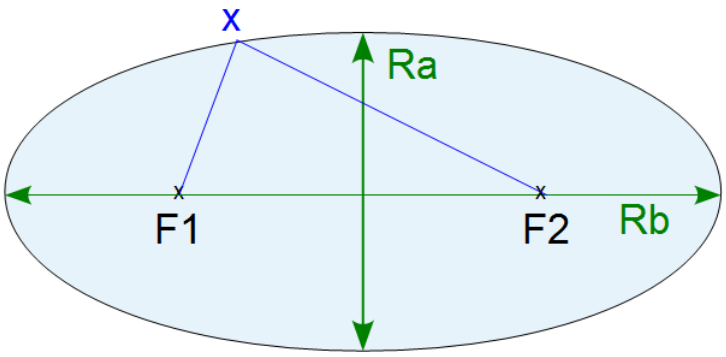


FIGURE 2.9 – Le modèle de l'ellipse : une courbe dessinée par un point x autour de deux foyers F1 et F2, vérifiant la condition : à chaque point, la somme de la distance de x aux deux foyers reste constante, soit (xF1 + xF2) constant.

Le script Matlab fut donc paramétré pour représenter des ellipses en définissant leur largeur (Rb) à partir des deux angles extrêmes relevés, et leur profondeur (Ra) à partir des distances extrêmes relevées. Un nuage d'ellipses pouvait ainsi être représenté, permettant d'observer de manière plus complète la répartition de la localisation.

D'autres ellipses permirent de représenter :

- sur le nuage d'ellipses, l'**ellipse moyenne**, tracée à partir d'une moyenne des données utiles à la formation de toutes les autres ellipses, et qui donnait un aperçu de la localisation globale des sons par les sujets ;
- sur les centres de gravité, une **ellipse de dispersion**, dont le demi-rayon Ra était proportionnel à la variance de la distance des centres de gravité, et dont le demi-rayon Rb correspondait à la variance en azimut des centres de gravité ; cette représentation permettait d'appréhender rapidement la dispersion des résultats, tant en azimut qu'en distance.

Des « box plots » pour comparer

A ces représentations de points et d'ellipses, a été ajoutée une série de représentation par « box plots », ou diagrammes de Tukey [3]. Rappelons que ce type de représentation permet d'afficher, pour une série de réponses données : la médiane des réponses, les premiers et troisièmes quartiles, les réponses extrêmes (fig. 2.10). Ces diagrammes étaient particulièrement indiqués pour une comparaison des résultats selon les systèmes.

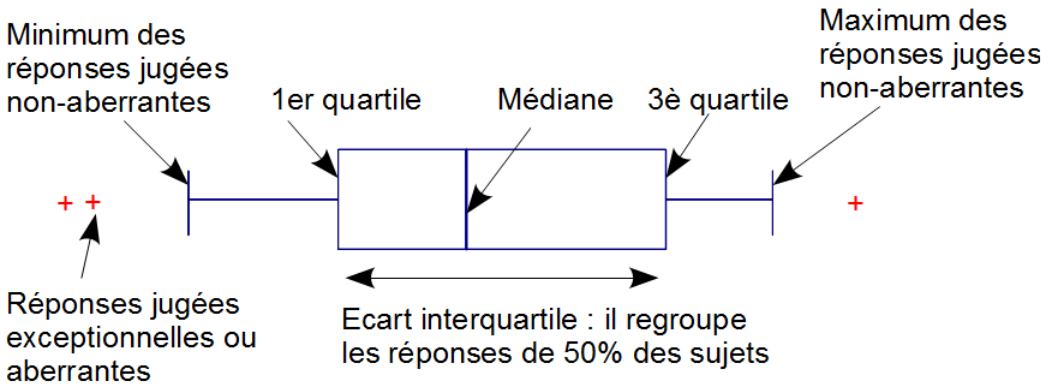


FIGURE 2.10 – Le modèle du box plot ou « diagramme de Tukey ».

Les représentations en box plots ont permis d'afficher les azimuts et distances obtenus par stimulus et par système, ainsi que l'erreur absolue en azimut et en distance, et les réponses aux échelles.

A partir de ces schémas, consultables en annexe (B.1 et suivantes ; ils ne seront pas tous reproduits ici pour des questions de place et de pertinence), des conclusions intéressantes ont pu être dégagées, qui seront exposées ici par système, et dans l'ordre suivant :

- résultats du test lumineux (« lum ») ;
- résultats du test sur enceinte (mixage principal) (« HP ») ;
- résultats du test binaural de référence (« binoRef ») ;
- résultats du test binaural référence cachée (« binoRefCach ») ;
- résultats du test binaural AAC (« binoAAC ») ;
- résultats du test binaural MP3 (« binoMP3 ») ;
- résultats des échelles (tous systèmes confondus) ;
- résultats du test HRTF.

L'estimation de l'azimut occupera une place prépondérante dans cette analyse, l'appréciation de la distance dépendant beaucoup du mixage. Cette dernière sera évoquée en **centimètres**, les distances mesurées étant celles des **figures** sur les **feuilles-réponses** des sujets.

3. (encore appelés diagrammes à boîtes ou à moustaches)

2.4.1 Les résultats du test lumineux (« lum »)

Le schéma 2.11 nous montre le nuage de points obtenus (tous les sujets ayant représenté les points lumineux par un point) ainsi que les ellipses de variance correspondantes, et les positions voulues, pour le test lumineux. Rappelons que ce test consistait pour les sujets à localiser dans l'espace des points positionnés au laser rouge autour d'eux dans la pénombre. Le but était alors d'évaluer leur représentation de leur espace péri-personnel. Rappelons également que lors du test l'obscurité les empêchait de distinguer les enceintes représentées sur le schéma (et qui étaient bien entendu absentes de leurs propres feuilles-réponses). La distance n'a pas été prise en compte dans cette analyse, tant la localisation en distance d'un laser est difficile.

En observant les résultats, on remarque que la précision de la localisation azimutale semble **dépendre** de l'**azimut de départ** du point lumineux : les angles paraissent davantage exagérés, et avec davantage de dispersion pour un stimulus se situant à -15 degrés (lumière 1, bleue, 15 degrés d'erreur en moyenne), qu'à 30 degrés (lumière 2, rouge, 5 degrés d'erreur) ou à 55 degrés (lumière 3, verte, 5 degrés). La représentation d'une lumière arrière offre un cas intéressant (lumière 4, noire, 15 degrés d'erreur en moyenne), avec ce qui semble être un repliement vers l'avant, donc une sous-estimation de l'azimut réel. Enfin, la localisation d'une source sur le côté, à -88 degrés (lumière 5, magenta, 2 degrés d'erreur en moyenne) semble plus facile, du fait sans doute du repère constitué par le dessin central : il doit être aisé pour les sujets de comprendre que le point lumineux est exactement à leur droite, plus que de représenter l'azimut d'un point situé entre 0 et 55 degrés. De même, on peut supposer que la précision de localisation reste bonne un peu au-delà de 90 degrés, mais que l'erreur augmente ensuite (le test actuel ne permet pas de dire si elle se serait de nouveau réduite en se situant exactement derrière les sujets, où on pourrait supposer que ceux-ci auraient pu la localiser avec davantage de précision).

On observe donc que les positions indiquées par les sujets sur ce test lumineux révèlent une erreur qui semble dépendre de la position de départ des sources lumineuses, erreur pouvant aller jusqu'à 15 degrés, que nous nommerons **deltaR** et qu'il nous faudra donc prendre en compte lors des analyses suivantes. Les conclusions de cette analyse sont résumées dans la figure 2.12.

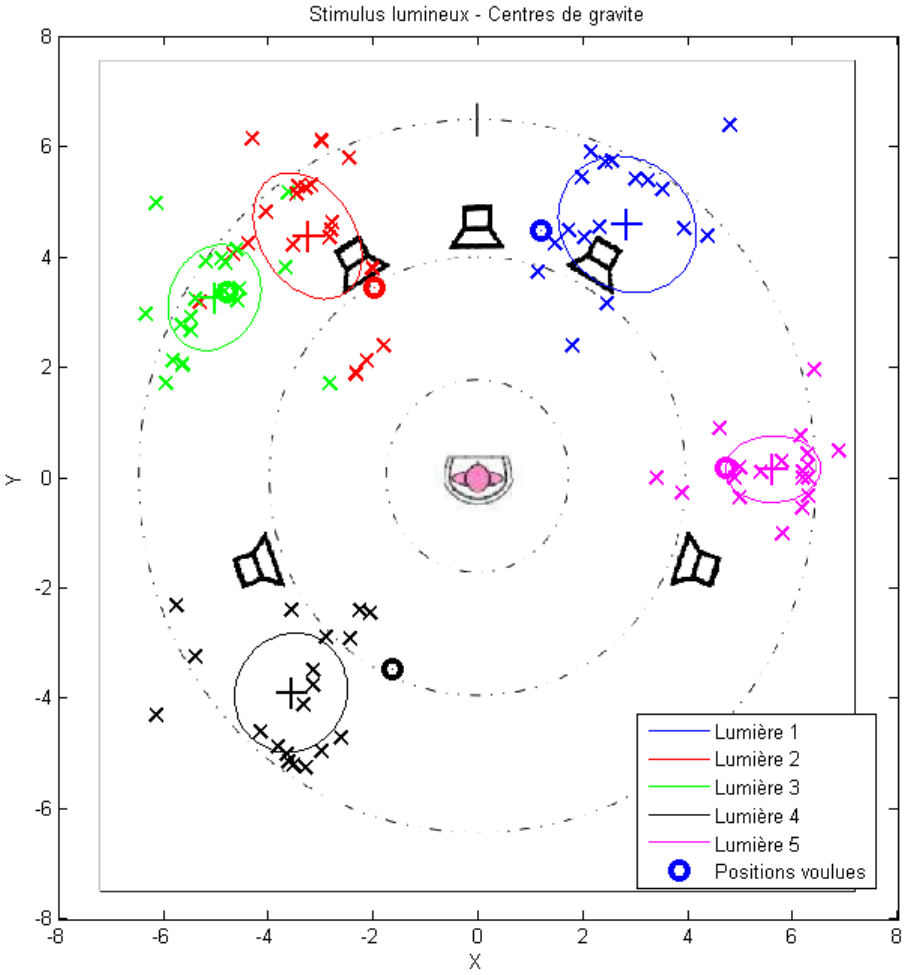


FIGURE 2.11 – Représentation de la localisation par les sujets de points lumineux. Les « + » de taille supérieure indiquent le point moyen obtenu, centre de l'ellipse de dispersion, dont la largeur est proportionnelle à la variance en azimut, et dont la profondeur est proportionnelle la variance en distance.

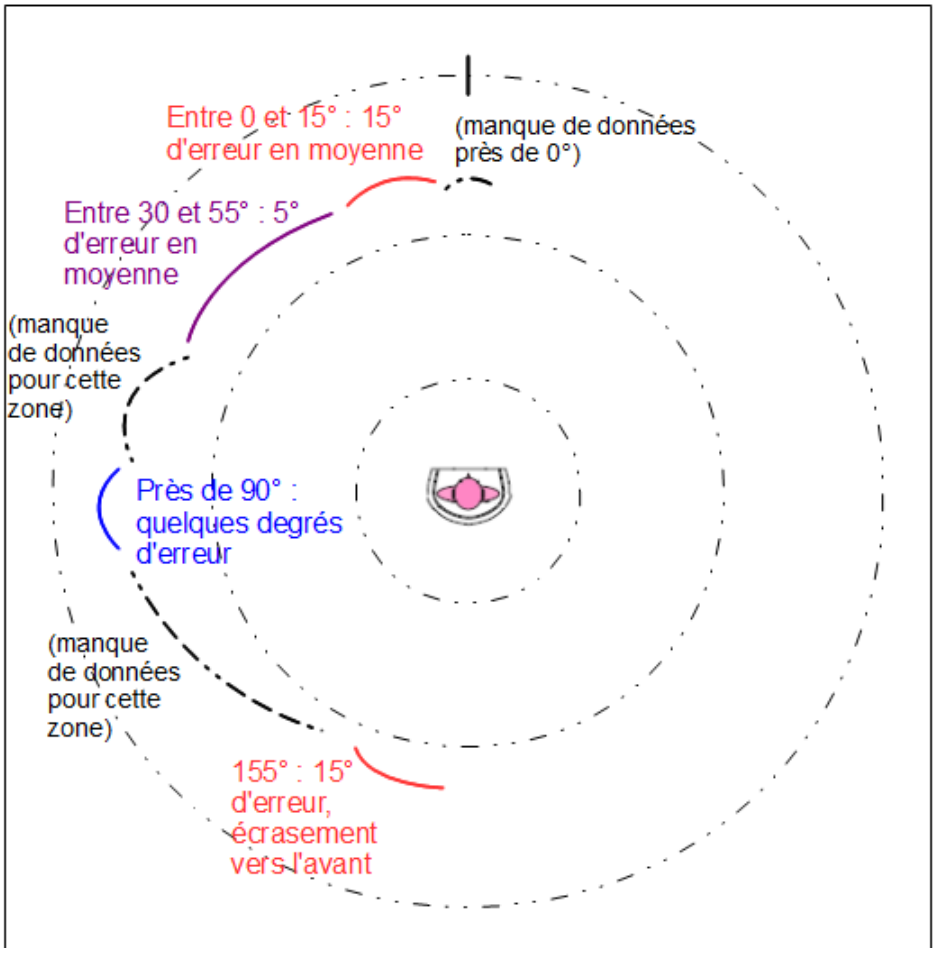


FIGURE 2.12 – Diagramme résumant l'erreur en azimut relevée pour le **test de localisation de points lumineux**, en fonction de l'azimut attendu. Les conclusions ne sont représentées que pour une moitié de l'espace péri-personnel azimutal du sujet, mais elles sont supposées valables pour l'autre moitié par simple symétrie axiale.

Introduction à l'analyse des résultats du test de localisation sonore

Les résultats des tests de localisation sonore seront résumés dans une série de **tableaux**, indiquant à chaque fois :

- le stimulus considéré (ambiance, rock, voix, classique, bruit rose) ;
- l'élément sonore considéré (corbeau, branche, buisson, caisse claire etc.) ;
- l'azimut attendu au mixage pour cet élément sonore (présenté sous la forme : azimut L/azimut R dans le cas des sons stéréophoniques) ;
- la médiane de l'azimut obtenu pour cet élément sonore (d'après les box plots d'azimut, en annexe B.3) ;
- l'écart interquartile entourant la médiane (valeur de l'écart avec précision entre parenthèses des premier et troisième quartiles) ;
- des remarques sur la localisation obtenue. La colonne des remarques comprendra également, le cas échéant, un rappel de l'erreur de représentation **deltaR** révélée par le test lumineux et correspondant à la zone de l'espace où l'élément sonore a été localisé.

Ces tableaux permettront donc de tirer des conclusions quand à l'azimut des sources. L'examen des résultats sera également à chaque fois précédé d'une reproduction d'un schéma circulaire, indiquant, à titre d'exemple, pour le stimulus ambiance, les centres de gravité des dessins des sujets, le centre de gravité moyen, et l'ellipse de variance. Les schémas concernant l'intégralité des stimuli sont reproduits en annexe B.1, et la modélisation des ellipses dessinées par les sujets est visible sur les schémas de l'annexe B.2.

Les observations concernant la distance seront basées notamment sur les box plots des distances reproduits en annexe B.6. Les box plots des distances pour le stimulus ambiance sont reproduits ci-après à titre d'exemple (fig. B.59). Rappelons que toutes les distances évoquées sont celles mesurées sur les feuilles-réponses des sujets, entre le centre de la feuille et le centre de gravité de leurs dessins, et qu'elles sont donc notées en centimètres.

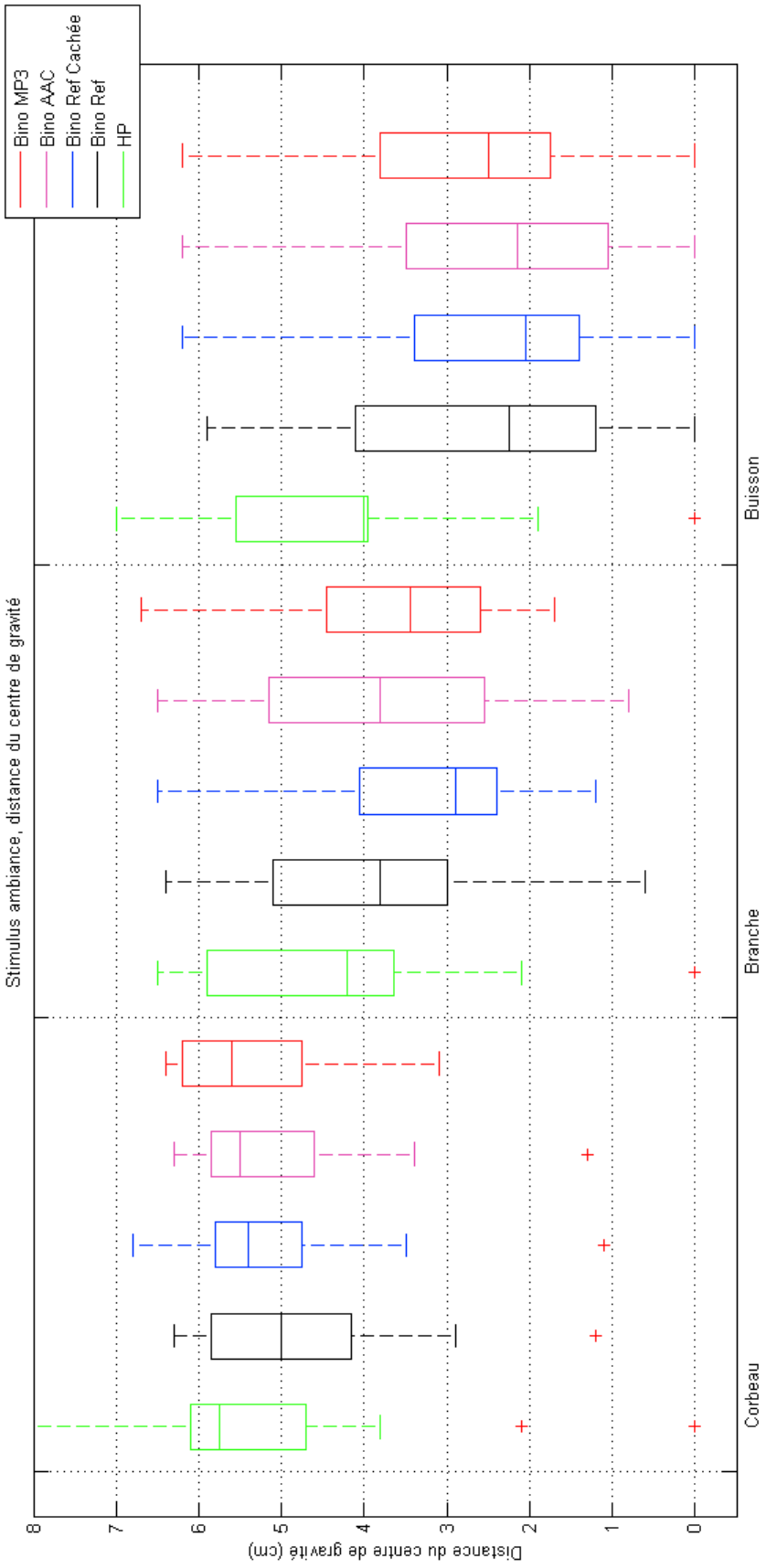


FIGURE 2.13 – Box plots des distances : « ambiance »

2.4.2 Les résultats du test sur enceintes (« HP »)

La démarche d'analyse employée pour ce travail est illustrée ici par le schéma montrant les centres de gravité des sources, le centre de gravité moyen, et l'ellipse de variance (de largeur proportionnelle à la variance en azimut, de profondeur proportionnelle à la variance en distance), pour le stimulus ambiance. Des schémas complémentaires se trouvent en annexe (B.1 à B.3).

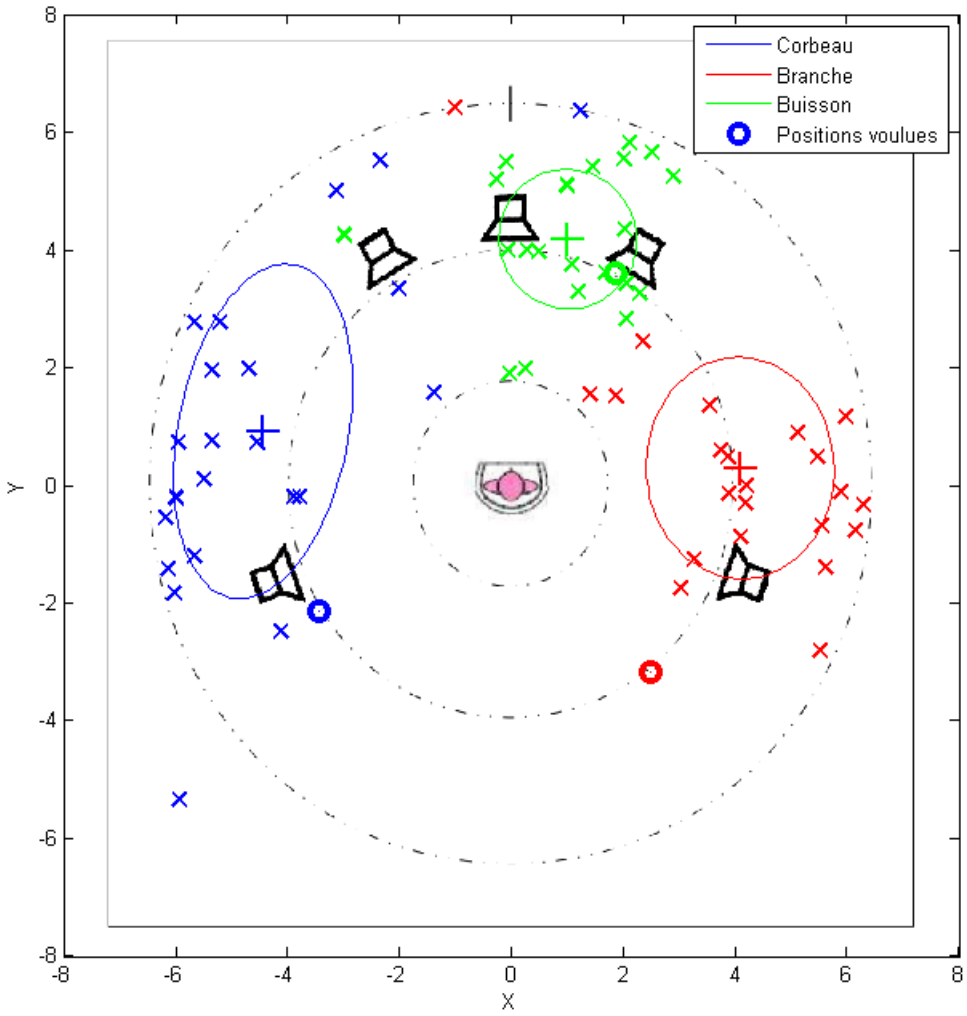


FIGURE 2.14 – Centres de gravité et ellipses de variance obtenus pour le stimulus « ambiance » , sur hauts-parleurs.

TABLE 2.1 – Azimuts et écarts interquartiles obtenus sur hauts-parleurs

Stimulus	Élément	Azimut (deg)		Écart (deg)	Remarques
		attendu	obtenu	et quartiles	
Ambiance	Corbeau	103/141	80	42 (50 à 92)	Ecrasement par sous-estimation de l'azimut (! deltaR = 15 degrés dans cette zone)
	Branche	-177/-106	-88	40 (-75 à -95)	Sous-estimation de l'azimut (! deltaR = 15 degrés)
	Buisson	-12/-42	-15	23 (-2 à -25)	Remis sur HP droit (! deltaR = 15 degrés)
Rock	Caisse	7.5	2	17 (-1 à 16)	Semble remis sur HP central
	Guitare	-12.9	-28	30 (-45 à -15)	Remis sur HP droit (! deltaR = 15 degrés)
	Voix	-4.5	0	4 (-4 à 0)	Localisation très bonne
Voix	V1 (h jeune)	112.5	97	18 (89 à 107)	Sous-estimation de l'azimut (! deltaR = 15 degrés)
	V2 (f jeune)	-35.22	-33	22 (-45 à -27)	Bonne localisation (mais remis sur le HP)
	V3 (h mûr)	23.55	32	11 (27 à 38)	Remis sur HP gauche (! deltaR = 5 à 15 degrés)
	V4 (f mûre)	-117.05	-105	24 (-112 à -88)	Sous-estimation de l'azimut (! deltaR = 15 degrés)
Classique	Flûte	14.7	29	22 (13 à 35)	Semble remis sur HP gauche (! deltaR = 15 degrés)
	Basson	-15.3	-33	16 (-22 à -38)	Semble remis sur HP droit (! deltaR = 15 degrés)
	Voix	0	0	4 (-2 à 2)	Localisation très bonne

TABLE 2.2 – Azimuts et écarts interquartiles obtenus sur hauts-parleurs - suite

Stimulus	**Élément**	**Azimut (deg)**		**Écart (deg)**	**Remarques**
		attendu	**obtenu**	**et quartiles**	
Bruit rose	Salve 1	-61.39	-37	16 (-32 à -48)	Rapproché du HP droit (! deltaR = 5 à 10 degrés)
	Salve 2	122.74	98	24 (88 à 112)	Ecrasement par sous-estimation de l'azimut (! deltaR = 15 degrés, donc peut-être une simple remise sur le HP Ls)
	Salve 3	-1.22	0	4 (-2 à 2)	Localisation très bonne

Analyse des résultats sur enceintes

Les résultats résumés dans les tableaux 2.1 et 2.2, et dans les box plots de distance (voir fig. B.59) permettent les remarques suivantes :

- que la diffusion sur système 5.0 ne semble pas retranscrire un espace complètement homogène sur le plan horizontal : les sources sonores ne sont pas toujours localisées là où elles ont été voulues lors du mixage ;

- que la localisation semble très précise (erreur de localisation inférieure à 5 degrés en écart interquartile, donc pour 50% des sujets) pour des sources situées exactement à l'avant (0 à 5 degrés d'azimut environ) (voir le « rock » : voix, le « classique » : voix, ou encore le « bruit rose » : salve 3) ;

- que la localisation des sources proches des hauts-parleurs semble subir une **aspiration** par les hauts-parleurs (voir sur l'« ambiance » : le buisson, qui semble remis sur les hauts-parleurs les plus proches de ses canaux gauche et droite ; sur le « rock » : la voix, la guitare ; sur le stimulus « voix » : voix 2, voix 3 ; sur le « classique » : la flûte, le basson) : le son semble alors localisé à son point d'émission physique (HP) le plus proche ; ce phénomène occasionne jusqu'à 15 degrés d'erreur sur le système LCR, davantage si le son à localiser est situé entre les enceintes avant et arrière (24 degrés d'erreur pour le « bruit rose », salve 2, pas d'exemple de son qui ait été spatialisé davantage vers l'arrière sans se rapprocher à moins de 20 degrés de Ls) ;

- que les sources spatialisées à l'arrière (entre Ls et Rs) semblent subir un **écrasement vers l'axe interaural**, donc en l'occurrence une sous-estimation de l'azimut qui ramène le son vers l'avant (42 degrés d'erreur pour le corbeau de l'« ambiance », pouvant être ramenés à 30 ou 20 degrés, si l'on prend en compte deltaR = 15 degrés dans cette zone d'après le test lumineux) ;

- que l'erreur de spatialisation et la dispersion des résultats semblent **dépendre du stimulus considéré** : une source stéréophonique semble donner lieu à davantage d'incertitude qu'une source monophonique, et deux sources monophoniques au contenu différent (voix d'homme jeune et salve de bruit rose) ne semblent pas tout à fait localisées avec une dispersion identique (écart de 18 degrés pour la voix 1 du stimulus « voix », attendue à 112.5 degrés, localisée à 97 degrés ; écart de 24 degrés pour la salve 2 de bruit rose, attendue à 122.7 degrés, localisée à 98 degrés) ;

- que la distance moyenne des sources est à peu près celle des hauts-parleurs, mais avec une dispersion très variable selon le stimulus considéré (voir : « ambiance », la dispersion de représentation en distance sur le corbeau : 1.3 cm, sur la branche : 2.3 cm, sur le buisson : 1.6 cm ; pour les « voix » en revanche, la dispersion tourne davantage autour de 2.2 cm).

– qu'on ne relève aucun problème d'internalisation des sources, ou IHL.

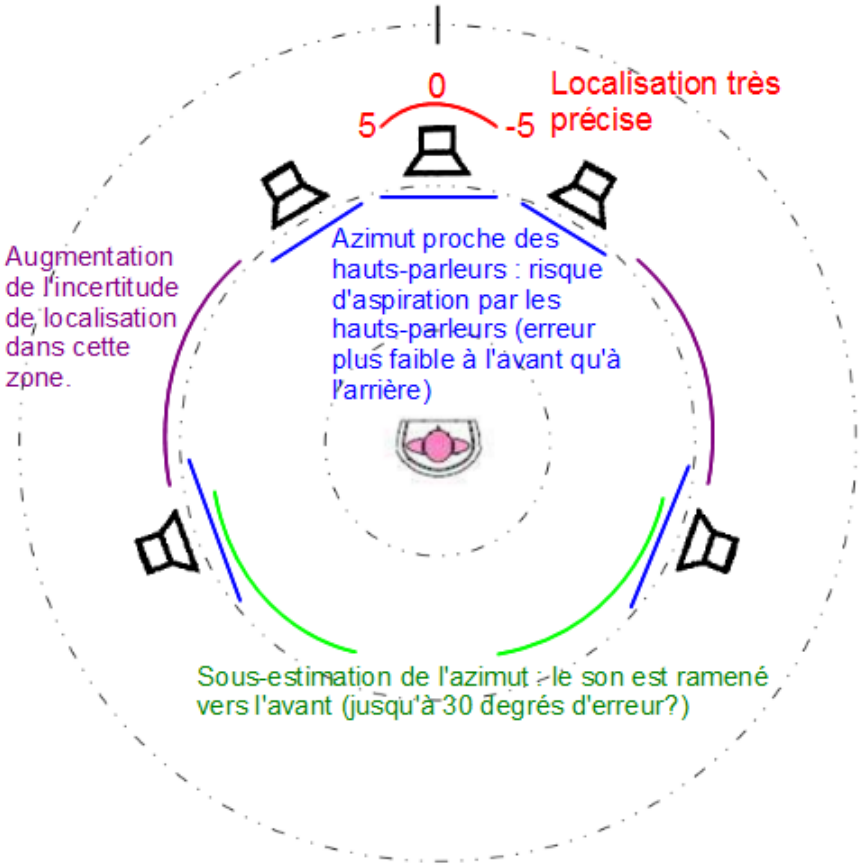


FIGURE 2.15 – Diagramme récapitulant les conclusions concernant la localisation en 5.0 sur enceintes.

2.4.3 Les résultats du test en binaural de référence (« binoRef »)

Rappelons que ce test consistait pour les sujets à juger la spatialisation de certains éléments de notre mixage 5.0, binauralisé, sans autre traitement.

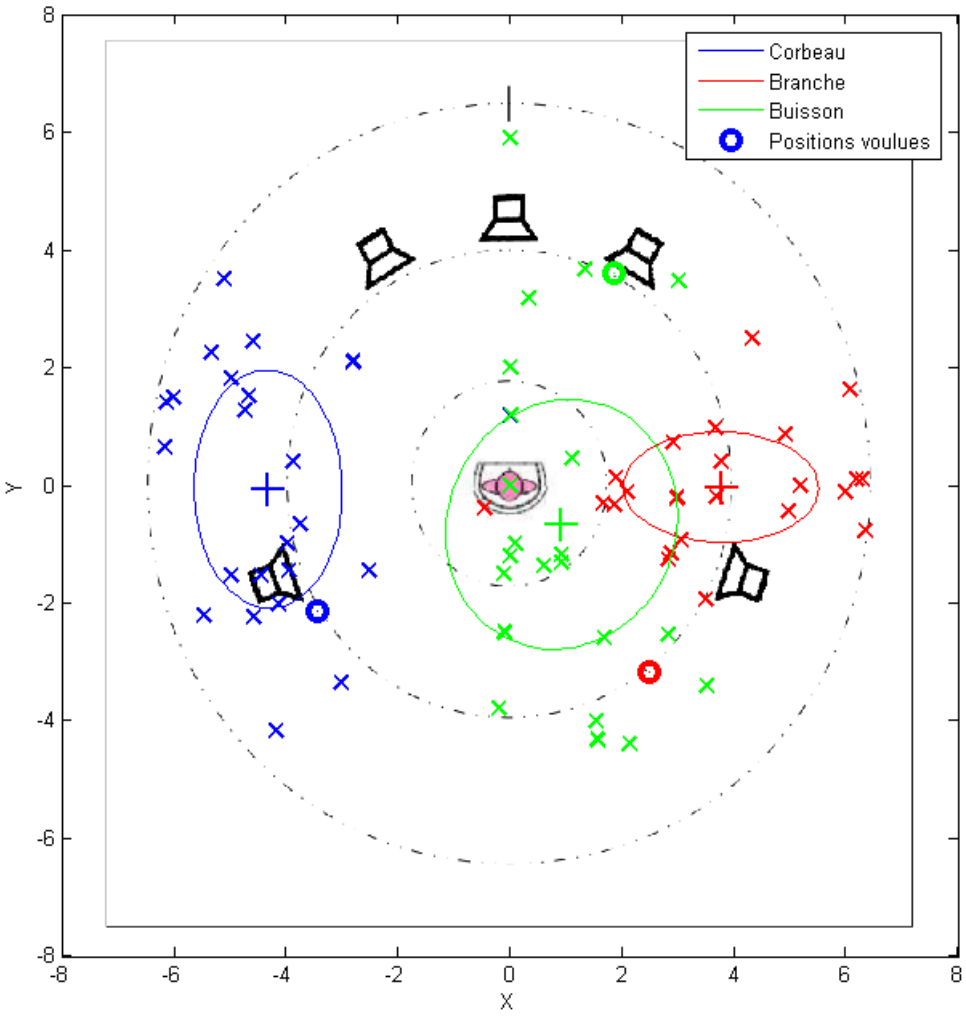


FIGURE 2.16 – Centres de gravité et ellipses de variance, « ambiance » , binoRef.

TABLE 2.3 – Azimuts et écarts interquartiles obtenus en binoRef

Stimulus	Élément	Azimut (deg)		Écart (deg)	Remarques
		attendu	obtenu	et quartiles	
Ambiance	Corbeau	103/141	92	43 (70 à 113)	Sous-estimation de l'azimut, mais moindre que sur HP (! deltaR = 15 degrés dans cette zone)
	Branche	-177/-106	-92	20 (-100 à -80)	Sous-estimation de l'azimut, moindre qu'en HP (! deltaR = 15 degrés)
	Buisson	-12/-42	-38	147 (-147 à 0)	Ecrasement par exagération de l'azimut, variance importante, confusion avant/arrière (! deltaR = 5 à 10 degrés)
Rock	Caisse	7.5	48	90 (17 à 107)	Ecrasement, variance importante (! deltaR = 5 à 15 degrés)
	Guitare	-12.9	-62	76 (-32 à -108)	Ecrasement, variance importante (! deltaR = 5 à 15 degrés)
	Voix	-4.5	0	20 (-18 à 2)	Localisation assez bonne mais confusions avant/arrière et internalisations.

TABLE 2.4 – Azimuts et écarts interquartiles obtenus en binoRef - suite

Stimulus	Élément	Azimut (deg)		Écart (deg) et quartiles	Remarques
		attendu	obtenu		
Voix	V1 (h jeune)	112.5	92	18 (84 à 102)	Sous-estimation de l'azimut, à peu près comme sur HP (! deltaR = 15 degrés)
	V2 (f jeune)	-35.22	-75	33 (-90 à -57)	Ecrasement (! deltaR = 5 à 10 degrés)
	V3 (h mûr)	23.55	45	27 (39 à 66)	Ecrasement (! deltaR = 5 à 10 degrés)
	V4 (f mûre)	-117.05	-100	22 (-115 à -93)	Sous-estimation de l'azimut, à peu près comme sur HP (! deltaR = 15 degrés)
Classique	Flûte	14.7	42	32 (31 à 63)	Ecrasement (! deltaR = 5 à 10 degrés)
	Basson	-15.3	-77	72 (-112 à -40)	Ecrasement (! deltaR = 5 à 10 degrés)
	Voix	0	0	7 (-5 à 2)	Localisation assez bonne mais confusions avant/arrière et internalisations.
Bruit rose	Salve 1	-61.39	-90	20 (-103 à -83)	Ecrasement, confusion avant/arrière (! deltaR = 5 à 10 degrés)
	Salve 2	122.74	100	23 (90 à 113)	Sous-estimation de l'azimut, à peu près comme sur HP (! deltaR = 15 degrés)
	Salve 3	-1.22	1	16 (-3 à 13)	Localisation assez bonne mais confusions avant/arrière et internalisations.

Analyse des résultats pour le binaural de référence : l'azimut

Les tableaux 2.3 et 2.4 nous permettent les observations suivantes :

- On observe une **dégradation globale** de la retranscription de l'espace sonore par rapport à une diffusion sur **enceintes**, avec notamment des **erreurs d'azimut** plus **nombreuses** et plus **importantes**.

- La **précision de localisation** semble bien **moindre** que sur enceintes, en azimut comme en distance, comme on peut en juger par les fortes variances, et sur les box plots par l'augmentation des écarts interquartiles (voir les box plots généraux d'azimut en annexe), sans que l'on puisse cependant donner une moyenne chiffrée pertinente de cette augmentation (en « rock », sur la caisse claire, l'écart interquartile est passé de 17 degrés sur enceintes à 90 degrés, sur la guitare de 30 à 76 degrés, pour la voix de 4 à 20 degrés ; sur l'« ambiance », pour le buisson, il est passé de 23 à 147 degrés).

- La précision de localisation **dépend** de l'**azimut** des sources : des sources situées entre **0 et 5 degrés** d'azimut restent localisées globalement à l'avant avec assez **peu de variance**, en revanche elles occasionnent de nombreux cas d'**internalisation** (que l'on peut voir sur les schémas d'ellipses en annexe B.2) ou de **confusion avant-arrière** (en « rock » et en « classique », voir la voix ; sur le « bruit rose », la salve 3) ; les sources ne venant pas du centre n'occasionnent que très rarement d'internalisation.

- Les sources situées entre **0 et 90 degrés** subissent un **écrasement**, c'est-à-dire un rapprochement de la **ligne interaurale** (coupant la tête selon une ligne 90 -90 degrés) (voir en « rock » : la caisse claire, la voix ; sur les « voix » : voix 2, voix 3 ; en « classique » : la flûte, le basson), qui peut parfois la dépasser (on a alors un cas de repliement ou confusion avant-arrière, comme pour le buisson, dans l'« ambiance »).

- Les sources situées **au-delà de 90 degrés** subissent un **écrasement** symétrique, qui les **ramène** vers la ligne interaurale (voir dans l'« ambiance » : le corbeau, la branche ; dans les « voix », les « voix » 1 et 4 ; pour le « bruit rose », la salve 2), avec parfois un **repliement arrière-avant** (voir le corbeau, la branche) ; certaines observations pourraient nous laisser croire que l'écrasement est d'autant plus important que le son de départ est spatialisé dans les arrières (donc qu'il est proche de l'azimut 180 degrés), mais le manque de données et l'incertitude de placement des sources arrière (montrée par le test lumineux) ne nous permettent pas de confirmer ces observations (voir cependant le corbeau de l'« ambiance », les « voix » 1 et 4 du stimulus « voix », la salve 2 de bruit rose). En revanche, l'écrasement observé dans les arrières n'est pas nécessairement plus important en binaural que sur enceintes (« ambiance », corbeau : attendu en moyenne à 120 degrés, perçu à 80 degrés sur enceintes et à 92 en binaural référence).

- La **précision** de localisation **dépend du stimulus considéré** : ainsi dans la musique classique, la flûte (14.7 degrés) et le basson (-15.3) sont en symétrie, mais la flûte est perçue à 32 degrés (idem sur enceintes) et le basson à -77 degrés (-30 sur enceintes) (voir aussi les « voix » : voix 1 et 4).

Il faut cependant relativiser ces résultats, en gardant à l'esprit que la représentation des points lumineux occasionnait une erreur deltaR de 15 degrés pour un son entre 0 et 30 degrés ou situé à l'arrière, qui se réduisait à moins de 5 degrés entre 30 et 90 degrés.

Analyse des résultats pour le binaural de référence : la distance

- Les sources sonores en binaural sont ici toujours perçues plus **proches** que sur hauts-parleurs, avec un rapprochement moyen que l'on peut évaluer à **1.3 - 1.7 cm**, soit dans l'espace réel un rapprochement d'un mètre environ pour notre système. Le rapprochement maximal observé est celui du « bruit rose » : salve 3 (de 4.8 à 1.7 cm).

- Tout cumulé, il semble que le seul paramètre demeurant à peu près fixe dans le passage de nos sons des enceintes au binaural, soit leur **projection sur l'axe interaural**, autrement dit leur **valeur d'abscisse en x**, dans un repère (0, x, y) où (0x) serait l'axe interaural (projection du plan frontal sur le plan horizontal) et (0y) la ligne avant/arrière (projection du plan médian sur le plan horizontal) (voir les figures 2.17, 2.18, 2.19 et l'annexe B.4).

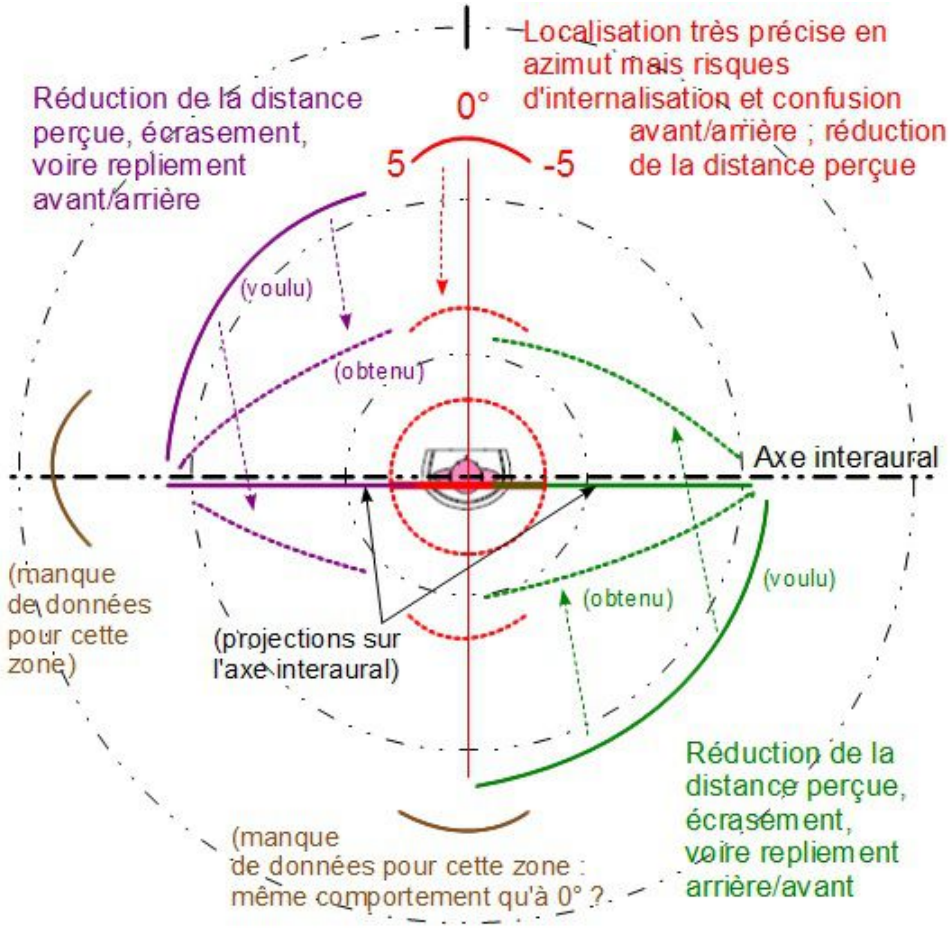


FIGURE 2.17 – Diagramme récapitulatif de la localisation en 5.0 binauralisé. On constate que l'un des seuls paramètres qui n'est pratiquement pas touché par la binauralisation est la projection sur l'axe interaural de la position de la source.

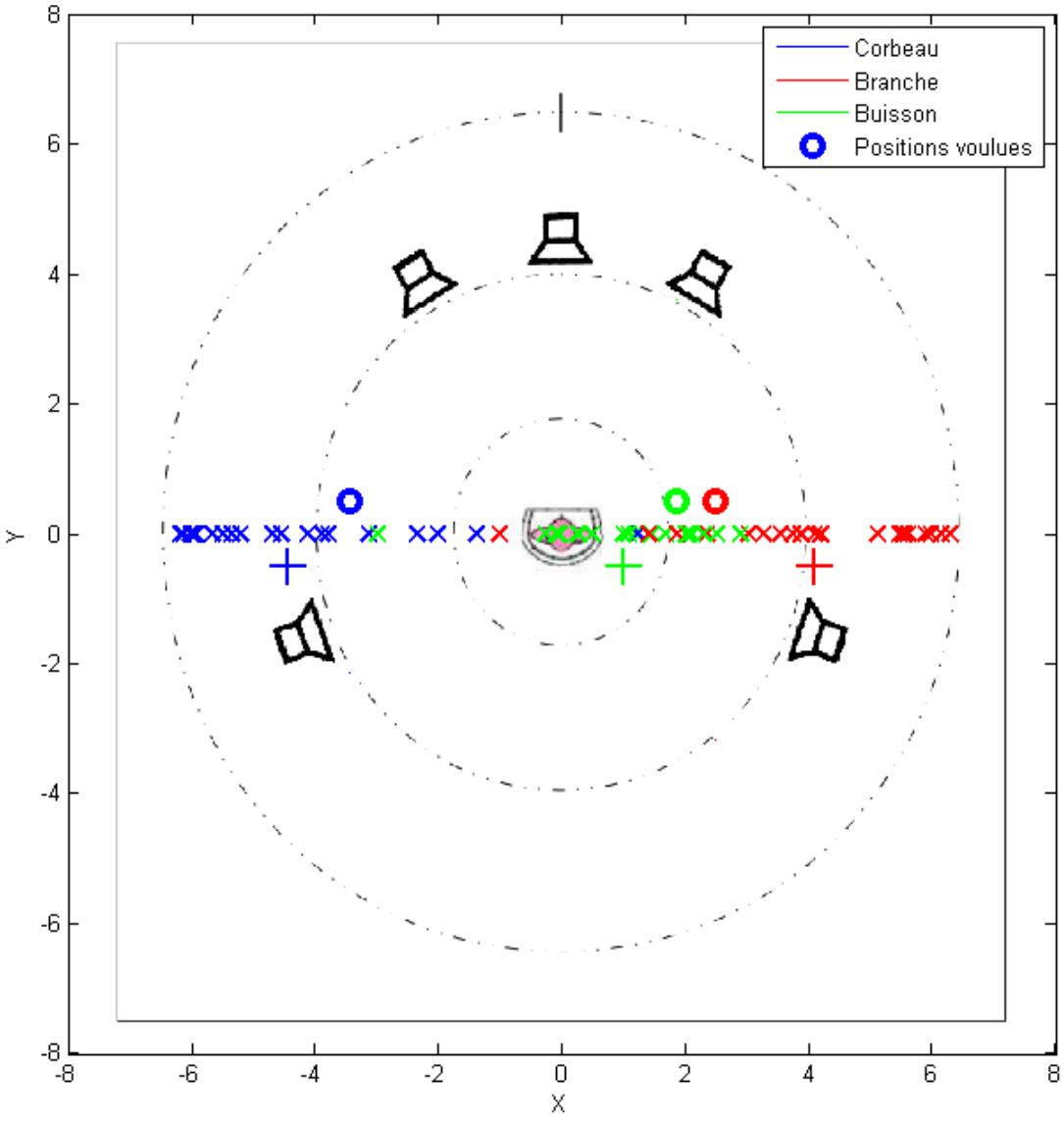


FIGURE 2.18 – Projection sur l'axe interaural des centres de gravité des réponses des sujets, « ambiance », hauts-parleurs.

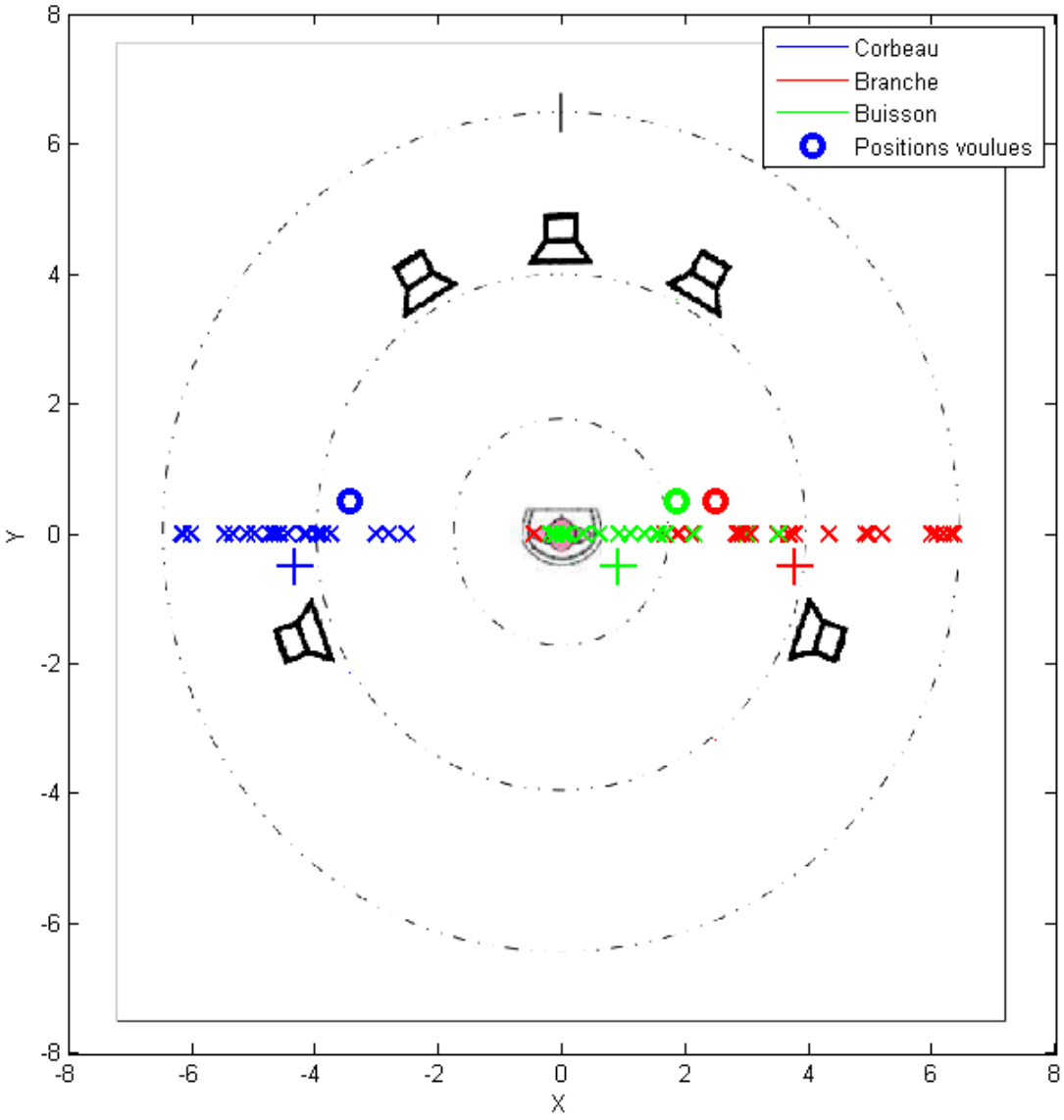


FIGURE 2.19 – Projection sur l'axe interaural des centres de gravité des réponses des sujets, « ambiance », binoRef.

2.4.4 Les résultats du test en binaural référence cachée (« binoRefCach »)

La référence cachée était une copie conforme du stimulus binaural de référence (binoRef), qui était donc diffusé une seconde fois au sujet, plutôt vers la fin du test, tandis que le binaural de référence pouvait survenir n'importe quand dans le courant de l'expérience. Théoriquement, on devrait donc retrouver pour la référence cachée des résultats identiques à ceux du binaural de référence.

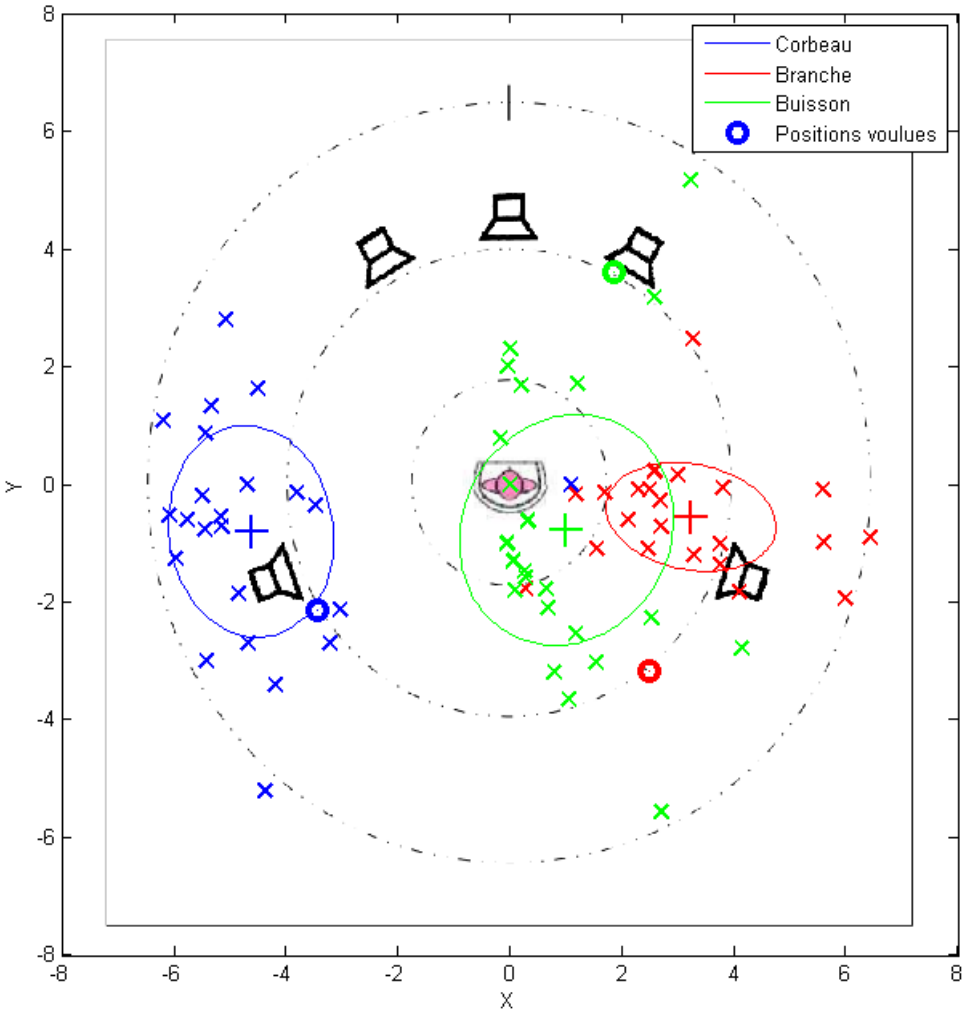


FIGURE 2.20 – Centres de gravité et ellipses de variance, « ambiance », référence cachée.

TABLE 2.5 – Azimuts et écarts interquartiles obtenus en binoRefCach

Stimulus	Élément	Azimut (deg)		Écart (deg) et quartiles	Remarques
		attendu	obtenu		
Ambiance	Corbeau	103/141	97	33 (80 à 113)	Ecrasement (sous-estimation de l'azimut) un peu moindre (de 3 deg), avec écart un peu plus réduit, qu'en BinoRef (! deltaR = 15 degrés dans cette zone)
	Branche	-177/-106	-97	20 (-109 à -91)	Sous-estimation de l'azimut, moindre qu'en BinoRef (de 5 deg) (! deltaR = 15 degrés)
	Buisson	-12/-42	-147	137 (-170 à -33)	Ecrasement (exagération de l'azimut), écart important, plus de confusion avant/arrière qu'en BinoRef (! deltaR = 5 à 15 degrés)
Rock	Caisse	7.5	58	48 (30 à 78)	Face à BinoRef, écrasement supérieur (Ref : 48), et écart moindre (Ref : 90) (! deltaR = 5 à 15 degrés)
	Guitare	-12.9	-86	43 (-102 à -59)	Ecrasement plus important (-86 contre -62) mais écart moindre (43 contre 76) qu'en BinoRef (! deltaR = 5 à 15 degrés)

TABLE 2.6 – Azimuts et écarts interquartiles obtenus en binoRefCach - suite

Stimulus	Élément	Azimut (deg)		Écart (deg) et quartiles	Remarques
		attendu	obtenu		
(Rock)	Voix	-4.5	-5	36 (-34 à 2)	Localisation assez bonne (meilleure qu'en BinoRef, à -5 contre 0) mais confusions avant/arrière et internalisations. Ecart plus large que BinoRef (36 contre 20).
Voix	V1 (h jeune)	112.5	94	21 (89 à 110)	Sous-estimation de l'azimut comparable à BinoRef (94 à 92) (! deltaR = 15 degrés)
	V2 (f jeune)	-35.22	-70	51 (-96 à -45)	Ecrasement un peu moindre, mais écart plus large qu'en BinoRef (51 contre 33) (! deltaR = 5 à 10 degrés)
	V3 (h mûr)	23.55	45	22 (41 à 63)	Ecrasement ; écart un peu moindre qu'en BinoRef (22 contre 27) (! deltaR = 5 à 10 degrés)
	V4 (f mûre)	-117.05	-112	18 (-115 à -97)	Sous-estimation de l'azimut moindre qu'en BinoRef (-112 contre -100), écart moindre de 6 degrés (! deltaR = 15 degrés)

TABLE 2.7 – Azimuts et écarts interquartiles obtenus en binoRefCach - suite

Stimulus	Élément	Azimut (deg)		Écart (deg)	Remarques
		attendu	obtenu	et quartiles	
Classique	Flûte	14.7	45	25 (32 à 57)	Ecrasement ; écart moindre qu'en BinoRef (25 contre 32) (! deltaR = 5 à 10 degrés)
	Basson	-15.3	-77	66 (-106 à -40)	Ecrasement ; écart moindre de 6 deg qu'en BinoRef (! deltaR = 5 à 10 degrés)
	Voix	0	0	5 (-2 à 3)	Localisation assez bonne mais confusions avant/arrière et internalisations.
Bruit rose	Salve 1	-61.39	-92	10 (-97 à -87)	Ecrasement, confusion avant/arrière, écart un peu moindre qu'en BinoRef (10 contre 20) (! deltaR = 5 à 10 degrés)
	Salve 2	122.74	108	19 (94 à 113)	Sous-estimation de l'azimut un peu moindre (108 contre 100), avec écart un peu plus réduit (19 contre 23), qu'en BinoRef (! deltaR = 15 degrés)
	Salve 3	-1.22	0	28 (-13 à 15)	Ecart plus large que BinoRef (28 contre 16), assez bonne mais confusions avant/arrière et internalisations,.

Analyse des résultats pour le binaural référence cachée : l'azimut

Les tableaux 2.5, 2.6 et 2.7 nous permettent les conclusions suivantes :

- Les résultats en azimut pour la référence cachée, ne sont similaires qu'avec une **tolérance** qui nous oblige à relativiser les différences constatées et à venir. Des tendances générales restent : sur les **confusions avant-arrière** (« ambiance » : buisson ; « rock » : voix ; « classique » : voix ; « bruit rose » ; salve 3), sur l'**internalisation** (à peu près en quantité semblable, et toujours majoritairement pour les sources venant de devant) (« rock » : voix ; « classique » : voix ; « bruit rose » : salve 3), sur l'**écrasement avant et arrière** en **fonction de l'azimut**. En revanche, les angles exacts et les variances subissent des modifications : sous-estimation de l'azimut moindre pour l'« ambiance » : le corbeau, la branche ; pour les « voix » : la voix 4 ; cependant, on a une sous-estimation équivalente pour la voix 1. De même, l'écart est plus large sur la voix 2, ou la salve 3 du « bruit rose », mais moindre sur les salves 1 et 2, sur la voix 3 du stimulus « voix », sur le corbeau de l'« ambiance »... L'écrasement à l'avant est plus important, en « rock » : caisse claire et guitare, mais il est moindre sur les « voix » : voix 2 et 3.

- Les azimuts des centres de gravité sont dans leur grande majorité conservés à **10 degrés près** en moyenne, par rapport à la référence, et ce, quel que soit l'azimut attendu. Des différences supérieures à 10 degrés sont exceptionnelles (voir le « rock » : guitare, et la « voix » : voix 4). Il faudra donc être prudent dans la suite de notre analyse, si l'on relève des différences d'azimut inférieures ou égales à 10 degrés, avant de les estimer ou non significatives.

- Ces différences semblent **dépendre du stimulus considéré**, puisque par exemple la guitare du stimulus « rock » témoigne d'une forte différence (plus de 20 degrés) tandis que le basson du stimulus « classique » conserve un angle strictement identique au degré près. Les deux sons étaient pourtant monophoniques et spatialisés dans la même zone (-12.9 et -15.3 degrés). Les variations d'azimut entre la référence et la référence cachée dépendraient donc en partie des **caractéristiques** du stimulus qui rendront celui-ci plus ou moins **précis** pour le sujet (par son spectre, sa durée, sa réverbération etc.), de sorte que deux écoutes lui laisseront la sensation de deux placements plus ou moins semblables dans l'espace.

- Les variances semblant globalement réduites sur la référence cachée, laissent penser que le sujet a reconnu une spatialisation proche et a moins hésité lors de l'écoute de la référence cachée (qui intervenait toujours en deuxième écoute, après la référence), et/ou qu'il a été victime à la longue d'un phénomène d'**automatisme** dans ses réponses (comme certains sujets l'ont eux-mêmes confié après le test). Quelques contre-exemples surviennent cependant ponctuellement (voix du « rock », voix 2 des « voix »), sans doute

liés là encore à la nature du stimulus lui-même.

Analyse des résultats pour le binaural référence cachée : la distance

Les box plots de distance (voir fig. B.59) nous permettent plusieurs observations :

- Les distances ne semblent **pas** suivre une **loi fixe**, ayant tendance tantôt, à se rapprocher sur la référence cachée (de plus d'un demi-centimètre sur la branche de l'« ambiance », la guitare pour le « rock », la voix 2 pour les « voix »...), tantôt à rester identiques à 0.2 cm près (sur l'« ambiance » : buisson, sur le « bruit rose » : salve 1, salve 2, en « classique » : basson, sur le « rock » : caisse claire, voix, ou encore sur les « voix » : voix 4...), tantôt à s'éloigner (de plus d'un demi-centimètre, sur l'« ambiance » : corbeau, sur les « voix » : voix 1...). La tendance générale semble plutôt être un léger rapprochement (de 0.3 cm ?), avec une variance un peu réduite. Peut-être l'effet de la fatigue a-t-il donné au sujet l'impression que les sons étaient plus proches lors de cette deuxième écoute. L'expérience pourrait alors être jugée trop longue, ou générer un effet d'apprentissage. Mais peut-être, comme nous l'avons nous-même ressenti lorsque nous avons effectué le test, cette deuxième écoute a-t-elle été l'occasion pour le sujet d'entendre des **éléments du son** qu'il n'avait pas **perçus** la première fois : un accent, une note, qui lui ont donné des indices supplémentaires plaidant pour une distance différente par rapport à sa première écoute. Quoi qu'il en soit, ces observations montrent qu'il nous faudra être extrêmement **prudents** dans nos conclusions sur les **distances** par la suite. L'appréciation de la distance est de toute manière tellement asujettie au **mixage** (voir les effets de plans sonores déjà possibles dans un mixage en mono ou en stéréo) que les différences observées ne sauraient relever uniquement des capacités du moteur binaural ou du codec utilisé.

2.4.5 Les résultats du test en binaural AAC (« binoAAC »)

Ces stimuli étaient, comme nous l'avons dit plus haut, le résultat d'une simple réduction de débit du binaural de référence, en AAC 192kbps.

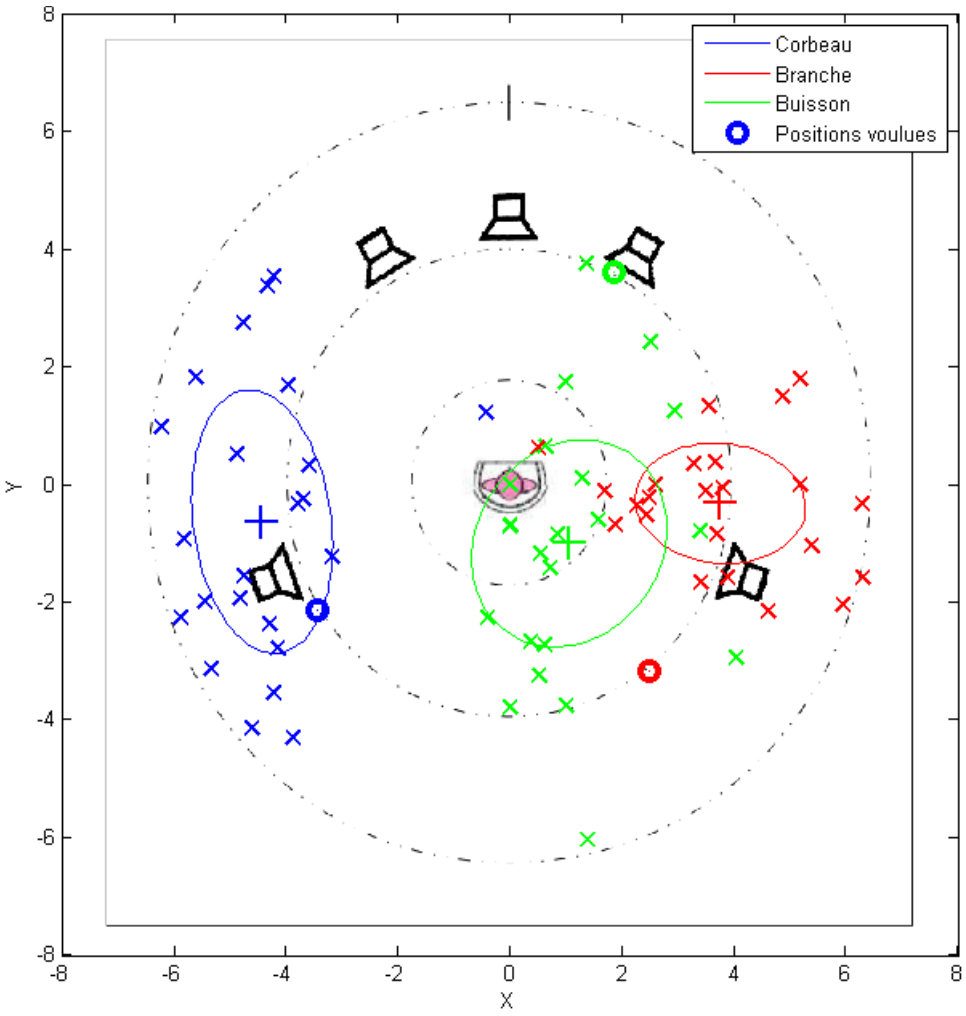


FIGURE 2.21 – Centres de gravité et ellipses de variance, « ambiance », binaural AAC.

TABLE 2.8 – Azimuts et écarts interquartiles obtenus en binoAAC

Stimulus	Élément	Azimut (deg)		Écart (deg) et quartiles	Remarques
		attendu	obtenu		
Ambiance	Corbeau	103/141	97	47 (67 à 114)	Sous-estimation de l'azimut comparable à BinoRefCach, écart plus grand qu'en BinoRef (47 contre 43) (! deltaR = 15 degrés dans cette zone)
	Branche	-177/-106	-92	20 (-108 à -88)	Azimuts et écart compris entre ceux de BinoRef et BinoRefCach (! deltaR = 15 degrés)
	Buisson	-12/-42	-75	153 (-153 à 0)	Ecrasement et écart compris entre ceux de BinoRef et BinoRefCach (écart comparable à BinoRef) (! deltaR = 5 à 15 degrés)
Rock	Caisse	7.5	45	42 (23 à 65)	Ecrasement moindre qu'en BinoRef (48) et RefCach (58) et écart moindre que BinoRefCach (48) (! deltaR = 5 à 15 degrés)
	Guitare	-12.9	-72	80 (-108 à -28)	Azimut compris entre BinoRef et BinoRefCach, écart supérieur (76 pour BinoRef) (! deltaR = 5 à 15 degrés)
	Voix	-4.5	-1	21 (-18 à 3)	Localisation assez bonne, mais confusions avant/arrière et internalisations. Azimut et écart compris entre BinoRef et BinoRefCach.

TABLE 2.9 – Azimuts et écarts interquartiles obtenus en binoAAC - suite

Stimulus	Élément	Azimut (deg)		Écart (deg) et quartiles	Remarques
		attendu	obtenu		
Voix	V1 (h jeune)	112.5	90	16 (87 à 103)	Azimut et écart compris entre BinoRef et BinoRefCach (! deltaR = 15 degrés)
	V2 (f jeune)	-35.22	-57	43 (-88 à -45)	Ecrasement moindre, mais écart compris entre BinoRef et BinoRefCach (! deltaR = 5 à 10 degrés)
	V3 (h mûr)	23.55	41	12 (38 à 50)	Ecrasement et écarts moindres qu'en BinoRef (45 et 27) et RefCach (45 et 22) (! deltaR = 5 à 10 degrés)
	V4 (f mûre)	-117.05	-105	26 (-115 à -89)	Ecrasement compris entre BinoRef et RefCach ; écart supérieur (26 contre 18 et 22) (! deltaR = 15 degrés)
Classique	Flûte	14.7	44	21 (39 à 60)	Ecrasement et écart compris entre BinoRef et BinoRefCach (! deltaR = 5 à 10 degrés)
	Basson	-15.3	-57	53 (-93 à -40)	Ecrasement moindre qu'en BinoRef (-77) et RefCach (-77) ; écart compris entre les deux (! deltaR = 5 à 10 degrés)
	Voix	0	0	6 (-5 à 1)	Azimut compris entre BinoRef et RefCach, mais confusions avant/arrière et internalisations.

TABLE 2.10 – Azimuts et écarts interquartiles obtenus en binoAAC - suite

Stimulus	**Élément**	**Azimut (deg)**		**Écart (deg)** et quartiles	**Remarques**
		attendu	**obtenu**		
Bruit rose	Salve 1	-61.39	-90	25 (-97 à -72)	Azimut compris entre BinoRef et BinoRefCach, écart plus important (25 contre 10 et 20) (! deltaR = 5 à 10 degrés)
	Salve 2	122.74	112	32 (93 à 125)	Moins d'écrasement qu'en BinoRef et RefCach (100 et 108), écart plus grand (32 contre 23 et 19) (! deltaR = 15 degrés)
	Salve 3	-1.22	0	187 (-15 à 172)	Localisation assez bonne mais confusions avant/arrière et internalisations, écart bien plus large que BinoRef et RefCach (187 contre 16 et 28).

Analyse des résultats pour le binaural AAC : l'azimut

On peut effectuer les observations suivantes d'après les tableaux 2.8, 2.9 et 2.10 :

- On relève plusieurs différences par rapport au binaural de référence, et à la référence cachée : que ce soit en azimut, comme sur le buisson du stimulus « ambiance », perçu plus à l'arrière en AAC (-75 degrés contre -38 degrés), ou sur la guitare, en « rock » (perçue à -72 degrés en AAC, contre -62 degrés pour la référence) ; ou pour les écarts interquartiles (voir la flûte, en « classique » : 25 degrés d'écart contre 32 pour la référence, ou la voix 2 du stimulus « voix » : 43 degrés au lieu de 33). Cependant, dans la majorité des cas les azimuts et les écarts relevés sont situés **entre la référence et la référence cachée** (ainsi du buisson, de la guitare, de la flûte...), et les cas où l'AAC se démarque clairement sont rares (azimut de la voix 2, azimut et écarts de la voix 3, azimut du basson...), et ces différences sont rarement supérieures à 10 degrés (exceptées la voix 2 : différence de 13 degrés d'avec la référence cachée, ou encore le basson : différence de 20 degrés d'avec la référence et la référence cachée). Or 10 degrés est l'écart constaté entre la référence et la référence cachée, écart en-deçà duquel les différences constatées ne sont donc **pas forcément significatives**.

- Les quelques différences relevées ne permettent pas de **dégager des axes clairs** : l'**écrasement** paraît dans l'ensemble **légèrement moindre** en AAC (voir « rock » : la caisse claire, stimulus « voix » : les voix 2 et 3, en « classique » : le basson, « bruit rose » : la salve 2) mais ce n'est souvent que de moins de 10 degrés.

- Les **écarts** montreraient quant à eux une légère tendance à être **plus importants** en AAC (voir la guitare du « rock », la voix 4 des « voix », les salves 1, 2 et 3 du « bruit rose »), mais ce n'est pas toujours vrai (voir la caisse claire du « rock », la voix 3 des « voix » ; pour le reste l'écart est compris entre ceux de la référence et de la référence cachée), et les différences constatées ne semblent là encore **pas forcément significatives** (pour le corbeau, la guitare, les salves 1 et 2 de bruit rose : l'écart est inférieur à 10 degrés).

- La différence la plus visible (voir notamment les box plots d'azimuts en annexe) concerne l'écart interquartile de la salve 3 de bruit rose, beaucoup plus grand en AAC (187 degrés contre 16 et 28). Mais cette différence impressionnante s'explique par le taux de confusions avant/arrière légèrement plus grand : la salve 3, située devant, provoquait déjà des cas de confusion avant/arrière (7 pour la référence, 9 pour la référence cachée) mais ils étaient notés sur le box plot comme données exceptionnelles. Une légère augmentation de ces confusions en AAC les a inclus dans les maximums et les quartiles ; et d'ailleurs la médiane reste la même en AAC qu'en PCM (0 degré contre 1 degré). Le taux de confusion avant/arrière n'est pour le

reste pas forcément plus grand en AAC qu'en PCM.

- En résumé, les **tendances** restent les **mêmes** en binaural AAC que pour la référence et la référence cachée, par rapport à la diffusion sur hauts-parleurs, et les différences qui pourraient être relevées entre l'AAC et la référence, pour l'azimut et pour l'écart, sont contredites d'un stimulus à l'autre et sont souvent comprises entre les valeurs de la référence et de la référence cachée. **La différence ne semblerait donc pas clairement perceptible**, au point de vue de l'azimut perçu des sources et de leur écart interquartile, entre le **binaural PCM** et le **binaural AAC 192**.

Analyse des résultats pour le binaural AAC : la distance

On peut tirer plusieurs observations des box plots de distance :

- Les réponses pour le binaural AAC semblent montrer une tendance à être **légèrement plus externalisées** qu'en binaural de référence : les sources seraient perçues plus loin. Et les box plots de distance montrent effectivement plusieurs cas où les sources en AAC sont perçues, en moyenne, à une distance soit identique, à 0.2 cm près (« voix » : voix 2, voix 3 ; « rock » : caisse claire, voix ; « classique » : flûte ; « ambiance » : branche, buisson), soit plus grande (« voix » : voix 1, distance augmentée de 0.8 cm en moyenne en AAC, voix 4 : de 0.6 cm, « classique » : voix, de 0.3 cm, « bruit rose » : salve 3, de 1.3 cm, « ambiance » : corbeau, de 0.6 cm). Seule la guitare, dans le stimulus « rock », est en moyenne rapprochée de 0.4 cm. Mais il faut relativiser cette tendance en comparant l'AAC avec la référence cachée : éloignée de 0.8 cm par rapport à la référence sur la voix 1 (stimulus « voix »), la source perçue en binaural AAC n'est éloignée que de 0.1 cm par rapport à la référence cachée. De même sur le corbeau (« ambiance »), où AAC et référence cachée se côtoient au 0.1 cm près, alors que la référence était plus proche de 0.6 cm. Il ne reste plus alors qu'une **très légère tendance générale** de l'AAC à être plus externalisé que le binaural PCM ; les **incertitudes** sur la distance, révélées par la comparaison entre la référence et la référence cachée, rendent difficile une conclusion plus tranchée.

- De même, le **taux d'internalisation**, ou de **confusion avant/arrière** pour les sources venant de devant, ne **permet pas** de dégager des **lois générales** (trois cas d'internalisation en plus en AAC que sur la référence, pour l'« ambiance », mais un cas de moins en « classique », et deux cas de moins en « rock »).

2.4.6 Les résultats du test en binaural MP3 (« binoMP3 »)

Rappelons que les stimuli « binoMP3 » sont le résultat d'une réduction de débit du binaural de référence, en MP3 192 kbps.

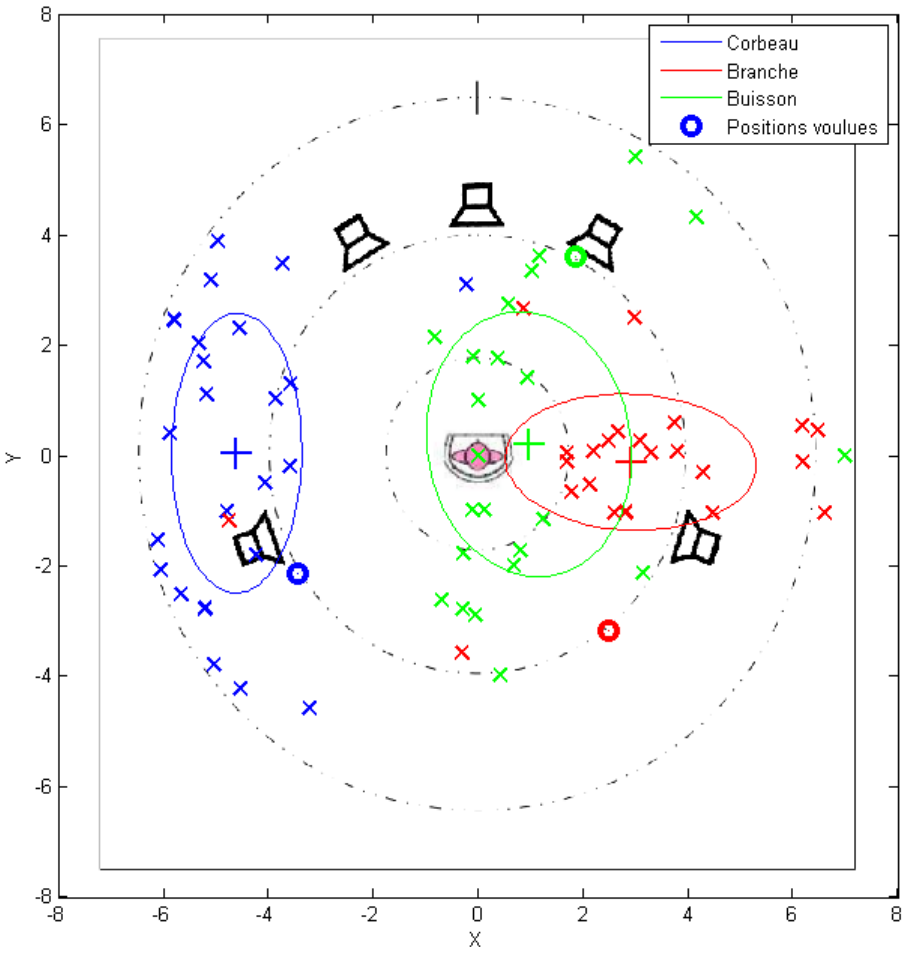


FIGURE 2.22 – Centres de gravité et ellipses de variance, « ambiance », binaural MP3.

TABLE 2.11 – Azimuts et écarts interquartiles obtenus en binoMP3

Stimulus	Élément	Azimut (deg)		Écart (deg) et quartiles	Remarques
		attendu	obtenu		
Ambiance	Corbeau	103/141	90	47 (65 à 112)	Sous-estimation de l'azimut moindre que les autres systèmes binauraux (92 en binoRef), écart supérieur à BinoRef (47 contre 43) (! deltaR = 15 degrés dans cette zone)
	Branche	-177/-106	-88	20 (-106 à -85)	Ecrasement et même repliement, plus que sur les autres systèmes binauraux (-92 pour BinoRef). Ecart comparable (! deltaR = 15 degrés)
	Buisson	-12/-42	-17	141 (-128 à 13)	Sous-estimation de l'azimut, écrasement inférieur à BinoRef (-17 contre -38) (! deltaR = 5 à 15 degrés)
Rock	Caisse	7.5	43	55 (15 à 70)	Ecrasement moindre qu'en BinoRef (48) et RefCach (58) et écart compris entre les deux (! deltaR = 5 à 15 degrés)
	Guitare	-12.9	-71	79 (-112 à -33)	Azimut compris entre BinoRef et BinoRefCach, écart comparable à BinoRef (76) (! deltaR = 5 à 15 degrés)
	Voix	-4.5	0	32 (-32 à 0)	Localisation assez bonne, mais confusions avant/arrière et internalisations. Azimut et écart compris entre BinoRef et RefCach.

TABLE 2.12 – Azimuts et écarts interquartiles obtenus en binoMP3 - suite

Stimulus	Élément	Azimut (deg)		Écart (deg) et quartiles	Remarques
		attendu	obtenu		
Voix	V1 (h jeune)	112.5	91	24 (88 à 112)	Azimut comparable à BinoRef (92), et écart un peu supérieur à BinoRefCach (21) (! deltaR = 15 degrés)
	V2 (f jeune)	-35.22	-60	42 (-84 à -42)	Azimut moindre (-60 pour -70 en BinoRefCach) et écart compris entre BinoRef et BinoRefCach (! deltaR = 5 à 10 degrés)
	V3 (h mûr)	23.55	45	22 (34 à 56)	Ecrasement et écart compris entre BinoRef et BinoRefCach (! deltaR = 5 à 10 degrés)
	V4 (f mûre)	-117.05	-108	33 (-125 à -92)	Ecrasement compris entre BinoRef et BinoRefCach, écart supérieur (33 contre 18 et 22) (! deltaR = 15 degrés)

TABLE 2.13 – Azimuts et écarts interquartiles obtenus en binoMP3 - suite

Stimulus	Élément	Azimut (deg)		Écart (deg) et quartiles	Remarques
		attendu	obtenu		
Classique	Flûte	14.7	34	29 (27 à 56)	Ecrasement moindre qu'en BinoRef (42) et RefCach (45), mais écart compris entre les deux (! deltaR = 5 à 10 degrés)
	Basson	-15.3	-63	64 (-100 à -36)	Ecrasement moindre qu'en BinoRef (-77) et RefCach (-77), mais écart compris entre les deux (! deltaR = 5 à 10 degrés)
	Voix	0	0	3 (-3 à 0)	Localisation comprise entre BinoRef et RefCach, mais confusions avant/arrière et internalisations.
Bruit rose	Salve 1	-61.39	-88	15 (-95 à -80)	Ecrasement et écart compris entre BinoRef et BinoRefCach (! deltaR = 5 à 10 degrés)
	Salve 2	122.74	100	26 (88 à 114)	Ecrasement compris entre BinoRef et RefCach, écart plus grand (26 contre 23 et 19) (! deltaR = 15 degrés)
	Salve 3	-1.22	0	22 (-20 à 2)	Localisation assez bonne mais confusions avant/arrière et internalisations, azimut et écart compris entre BinoRef et RefCach.

Analyse des résultats pour le binaural MP3 : l'azimut

D'après les résultats retranscrits dans les tableaux 2.11, 2.12 et 2.13 :

- Les différences constatées en azimut, entre la version MP3 et les versions de référence et de référence cachée, semblent **faibles** : dans une majorité de cas, la médiane des azimuts et l'écart interquartile sur les box plots d'azimut, restent en MP3 compris entre ceux pour la référence et la référence cachée (voir en « rock » : la guitare, la voix ; stimulus « voix » : voix 3 ; classique : voix ; « bruit rose » : salves 1 et 3), et dans la quasi-totalité des cas ces différences sont inférieures ou égales à 10 degrés (exceptions notables : le buisson, le basson), ce qui ne les rend **pas forcément significatives**. Les deux exceptions relevées nous permettent donc de conclure que, pour les sources situées à l'avant, l'**écrasement** vers l'axe interaural en MP3 semblerait soit égal, soit **légèrement moindre** qu'en binaural de référence et pour la référence cachée.

- Pour le reste, de même qu'en AAC, **la différence ne semblerait pas clairement perceptible**, au point de vue de l'azimut perçu des sources et de leur écart interquartile, entre le **binaural PCM** et le **binaural MP3 192**.

Analyse des résultats pour le binaural MP3 : la distance

- On observe, sur plusieurs stimuli, une **légère tendance** à l'**externalisation**, en MP3 plus qu'en binaural de référence : sur le corbeau (« ambiance »), à 5.6 cm contre 5 cm pour la référence ; sur le « bruit rose » : salve 3, 3 cm contre 1.8 cm ; sur la voix 1 (stimulus « voix »), 4.5 cm au lieu de 3.9 cm. Mais cette tendance est contredite ailleurs : les médianes des box plots sont à 4.2 cm sur la guitare, en « rock », sur la référence, et à 3.9 cm en MP3. Si la salve 3 du « bruit rose », située devant, s'éloigne bel et bien, passant de 1.8 cm sur la référence à 3 cm en MP3, la voix du « rock », elle aussi devant, se rapproche, passant de 3.1 cm sur la référence à 2.5 en MP3. Dans une majorité de cas, la distance moyenne perçue en MP3 reste comprise entre celle perçue sur la référence et sur la référence cachée (pour les médianes comme pour les écarts interquartiles), ce qui laisse penser une fois de plus que les **différences** constatées en **distance** en MP3 par rapport aux autres versions ne seraient **pas significatives**.

- Le **taux d'internalisation** et de **confusion avant/arrière** reste lui aussi **semblable** à celui constaté pour la référence et la référence cachée (un cas de plus en MP3 sur les ambiances et le « rock », mais un cas de moins sur le « bruit rose », autant de cas sur le « classique », toujours aucun cas sur les « voix »).

2.4.7 Conclusion pour les tests de localisation

A la lumière des différents résultats observés jusqu'ici, on peut estimer que, en azimut comme en distance, des différences claires apparaissent entre la diffusion sur **enceintes** et la diffusion en **binaural** : **écrasement** des sources vers l'axe interaural en binaural, donc **réduction des distances** observées et **exagération** ou **sous-estimation** des **azimuts**, **selon l'azimut** attendu pour la source au mixage. On relève également un risque élevé de **confusion avant/arrière** et d'**internalisation**, particulièrement pour les sources situées devant (azimut proche de zéro degré, à environ 7 degrés près). Par rapport aux modifications relevées entre les enceintes et le binaural de références, les différences observées entre la **référence**, les stimuli en **AAC** et les stimuli en **MP3** sont **faibles**, et sont dans l'ensemble **difficilement significatives**, d'autant plus lorsqu'on les compare avec la **référence cachée**. Il ne faut pas non plus perdre de vue l'**incertitude** de représentation de l'espace révélée par le test lumineux. Enfin, les résultats obtenus semblent dépendre de la **nature** du stimulus considéré.

2.4.8 Les résultats pour les échelles

Rappelons que les échelles, graduées de 0 à 7, permettaient aux sujets de noter : la précision, l'immersion, la lisibilité, le timbre/coloration, l'appréciation, pour les différents stimuli [4]. Ces échelles ne permettaient de noter à chaque fois, pour chaque système, que l'ensemble du stimulus : « ambiance », « rock », « voix » , « classique », « bruit rose », et non pas les éléments sonores de chaque stimulus séparément (corbeau, branche, buisson, caisse claire, guitare, voix etc.). Les échelles obtenues sont reproduites sur les pages suivantes.

De nombreux sujets ont fait part de leur perplexité face à ces échelles, pour plusieurs stimuli. Plusieurs d'entre eux ont avoué avoir noté un peu à l'instinct, sans être sûrs qu'ils auraient donné les mêmes notes s'ils avaient dû refaire le test à un autre moment. On sait par ailleurs que les sujets ont tendance à éviter, par réflexe, de mal noter un stimulus [5]. Il convient donc d'analyser ces échelles avec prudence.

4. Pour un rappel des définitions que nous avons données aux échelles, voir l'annexe A.3, et pour un exemple de ces échelles sur la feuille-réponse voir les annexes A.4 et A.5.

5. Le fait qu'on n'ait aucune note à 0 ne veut donc pas forcément dire qu'aucun stimulus ne l'aurait mérité.

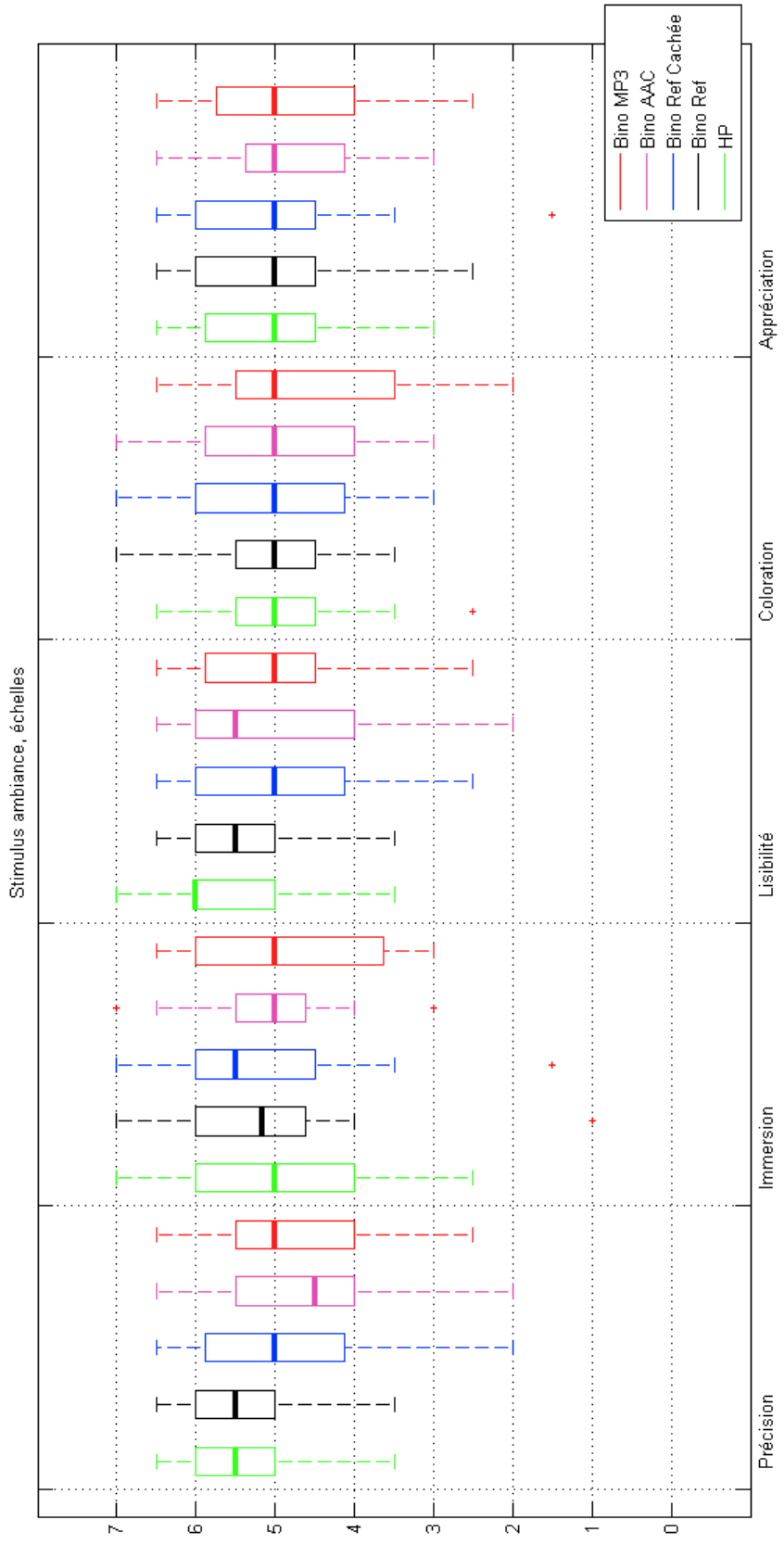


FIGURE 2.23 – Box plots des échelles : « ambiance »

Résultats pour l'échelle : la précision

Le sentiment de précision était noté de 0 : « mal défini », à 7 : « bien défini ». Les hauts-parleurs offrent généralement le plus de précision : avec une médiane à 5.5 points sur l'« ambiance », le « bruit rose », le « classique », 6 sur les « voix », mais 4.3 points seulement sur le « rock ». Toutefois, les réponses en binaural ne sont jamais notées très différemment : 5.5 points sur l'« ambiance » en binaural de référence (à égalité avec les enceintes), le minimum étant atteint avec 4.5 en AAC ; sur le « bruit rose », on a 5 points pour le binaural référence, référence cachée, AAC, mais 5.5 points pour le MP3 (à égalité avec les enceintes). Sur le stimulus « classique », le binaural MP3 est à 0.5 point derrière les enceintes (avec 5 points pour 5.5), référence et référence cachée sont à 4.5 points et la valeur minimum est de 4 pour l'AAC. Pour le stimulus « rock », les enceintes à 4.3 points sont légèrement dépassées par les systèmes binauraux à 4.5 points, sauf l'AAC qui a la valeur minimum avec 4 points. Sur les « voix » en revanche, c'est la référence qui est la dernière avec 4.5 points, les autres systèmes sont à 5 points (et les enceintes à 6).

On observe donc que **les enceintes** restent le système offrant pour les sujets le **plus de précision**. Mais cette notation **dépend du stimulus** : la différence est plus prononcée sur les « voix » ou le « classique », que sur le « bruit rose » et l'« ambiance », et devient presque négative pour le « rock ». Parmi les systèmes binauraux, la comparaison entre la référence et la référence cachée est intéressante : théoriquement les résultats devraient être identiques, or si les médianes sont de même valeur sur le « bruit rose », le « classique » et le « rock », la référence obtient 0.5 point de plus que la référence cachée sur les ambiances, et 0.5 point de moins sur les « voix ». Les écarts interquartiles quant à eux restent toujours compris entre 1 et 2.2 points. Ces observations laissent penser que des différences d'échelles entre deux systèmes, qui seraient inférieures ou égales à **0.5 point**, ne seraient **pas forcément significatives**.

Le binaural **AAC** a tendance à être perçu comme **légèrement moins précis** (0.5 points de moins que la référence et la référence cachée, en « classique » et en « rock » ; pour les ambiances : 1 point de moins que la référence mais 0.5 de moins que la référence cachée), sauf sur les « voix » où la référence est en arrière de 0.5 point sur l'AAC et le MP3 (mais la référence cachée est à égalité avec eux). Le **MP3** montrerait une tendance à être **légèrement plus précis** que les autres systèmes binauraux : il obtient les meilleurs résultats d'entre eux sur le « bruit rose » et le « classique » (avec 0.5 point d'avance sur la référence), et n'est dépassé que sur l'« ambiance », de 0.5 point, par la référence (mais reste égal à la référence cachée). Cependant, puisque ces différences ne dépassent jamais 0.5 point, eu égard à nos réflexions sur la comparaison référence - référence cachée, elles doivent être considérées avec réserve.

Tous systèmes confondus, la précision semble cependant meilleure sur l'« ambiance », le « bruit rose », et les « voix », que les musiques classique et « rock » (0.5 à 1.5 point d'écart en moyenne selon le système considéré). Cela peut paraître logique du point de vue du mixage, les musiques présentant des éléments sonores fortement réverbérés et plus flous, donc potentiellement plus difficiles à localiser avec précision.

Résultats pour l'échelle : immersion

Le sentiment d'immersion était noté de 0 : « mauvais », à 7 : « excellent ». Le « bruit rose » n'était pas concerné par ce critère, qu'il aurait été difficile d'évaluer pour ce stimulus. Cette fois-ci les **enceintes** n'obtiennent **pas toujours** les **meilleurs résultats** : si elles sont encore en tête sur les « voix », à égalité avec la référence cachée et l'AAC (à 4.5 points, référence et MP3 étant derrière à 4 points), elles sont partout ailleurs dépassées : sur les ambiances, par la référence et la référence cachée (à respectivement 5.2 et 5.5 points contre 5 pour les enceintes), AAC et MP3 étant à égalité à 5 points ; en « classique », les enceintes sont dépassées par le MP3 (à 5 points contre 4.5), les autres systèmes étant à égalité avec les hauts-parleurs (à 4.5 points) ; en « rock », elles sont dépassées par la référence (à 4.5 contre 4 points pour les enceintes), la référence cachée et le MP3 étant à égalité avec les hauts-parleurs (4 points) et l'AAC étant derrière avec 3.5 points.

Entre les systèmes binauraux, on retrouve un écart référence - référence cachée pouvant atteindre 0.5 point, dans un sens ou dans l'autre (sur l'« ambiance », le « rock », les « voix »). La référence, ou la référence cachée sont plusieurs fois en tête (sur l'« ambiance », le « rock »), parfois à égalité (avec l'AAC, sur les « voix »), mais jamais de plus de 0.5 point par rapport au système ayant obtenu la plus mauvaise note. Le système le plus mal noté, d'ailleurs, n'est pas toujours le même : l'AAC est bon dernier sur le « rock » avec 3.5 points (4.5 pour la référence, 4 pour la référence cachée), et dernier sur l'« ambiance », *ex aecquo* avec le MP3 (à 5 points, 5.2 pour la référence, 5.5 pour la référence cachée) mais il est en tête sur les « voix » avec 4.5 points, à égalité avec la référence cachée. A l'opposé, le MP3 est en tête sur le « classique » avec 5 points (4.5 pour les autres), mais partout ailleurs il est dans la moyenne des autres.

Tous systèmes confondus, l'immersion semble globalement meilleure sur le **stimulus « ambiance »** (comme l'ont confirmé d'ailleurs les commentaires écrits par les sujets) avec une moyenne comprise entre 5 et 5.5 points et plusieurs maximums à 7 points. Elle est moins bonne sur le « classique » (comprise entre 4.5 et 5 points), encore moins sur les « voix » (entre 4 et 4.5 points), encore moins sur le « rock » (entre 3.5 et 4.5 points). Les commentaires des sujets laissent penser que la musique classique, avec son ambiance de salle et son bruissement de public, leur semble encore immersive ; ils parlent souvent, en revanche, de leur difficulté à ressentir une bonne immersion sur un stimulus constitué de 4 voix séparées ; pour le « rock » enfin, il semble que beaucoup aient eu du mal à se sentir immergés dans la scène sonore, d'après leurs commentaires. Mais dans ce cas, ils évoquent davantage des caractéristiques de mixage, que du système considéré. Par ailleurs puisque les différences entre enceintes et binaural, et plus encore entre les systèmes binauraux, excédent rarement 0.5 point, elles peuvent difficilement être considérées comme clairement significatives.

En résumé, l'immersion est globalement jugée un peu **meilleure** avec les **systèmes binauraux** que pour les enceintes. Les différences entre systèmes binauraux ne permettent pas d'instituer une hiérarchie claire entre eux. En revanche, il apparaît que les résultats sont **liés au stimulus**, à sa nature et à son

mixage.

Résultats pour l'échelle : lisibilité

Celle-ci était notée de 0 : « confuse », à 7 : « distincte ». Le « bruit rose » diffusé sur hauts-parleurs ne comprenait pas d'échelle pour ce critère, que les pré-tests avaient montré difficilement quantifiable pour ce mode de diffusion.

Là encore, les enceintes ne semblent pas forcément offrir les meilleurs résultats : en tête sur le « ambiance » et les « voix » (à 6 points, les meilleurs systèmes binauraux n'arrivant qu'à 5.5), elles sont à égalité avec tous les autres systèmes sur le « rock » (à 5 points, avec qui plus est des écarts interquartiles assez comparables, de 2 points en moyenne), et sont moins bien notées sur le « classique » (à 5 points, contre 5.5 points pour la référence cachée par exemple).

La référence et la référence cachée diffèrent encore de 0.5 point, dans un sens ou dans l'autre (sur le « ambiance » , le « bruit rose », le « classique »). Dans l'ensemble les systèmes binauraux restent très proches, à moins de 0.5 point d'écart, et un système qui est mieux noté sur un stimulus (AAC et référence sur les ambiances, à 5.5 points), sera moins bien noté ailleurs (AAC sur le « classique », à 5.5 points).

La différence n'est jamais très marquée non plus entre les stimuli, même si elle existe : les valeurs pour les « voix » sont en tête, comprises entre 5.5 et 6 points (tous systèmes confondus) ; pour l'« ambiance » elles sont comprises entre 5 et 6 points ; sur le « bruit rose », entre 5 et 5.5 points, sur le « rock », tous les systèmes sont à 5 points, et sur le « classique » les résultats s'échelonnent entre 4.5 et 5.5 points. On pourrait donc en conclure que les différences de lisibilité sont faibles ; cependant il peut sembler étrange que quatre voix monophoniques paraissent aussi lisibles qu'une ambiance de forêt, que des salves de bruit rose, qu'une musique rock, et qu'une musique classique avec de longues réverbérations. Peut-être alors peut-on suspecter que ce critère n'a pas été toujours très bien compris par les sujets, ce qui pourrait mettre en cause notre définition, peut-être trop imprécise. Il nous faut donc rester prudent dans nos conclusions par rapport à ce critère.

Résultat pour l'échelle : timbre/coloration

Pour ce critère, noté de 0 : « sombre », à 7 : « brillant » (attention, il ne s'agit donc pas ici d'un jugement de valeur « bon/mauvais »), on observe clairement une baisse des résultats des systèmes binauraux par rapport aux hauts-parleurs : ceux-ci arrivent presque toujours en tête (à 5.5 points sur le « bruit rose », avec 1 point d'écart sur le suivant qui est la référence ; à 5 points sur le « rock » et les « voix », avec une avance de 0.5 point sur le suivant, qui est respectivement la référence, et l'AAC), ou en tête à égalité (sur le « classique » : à 5 points, égalité avec le MP3 ; sur l'« ambiance », à 5 points, à égalité avec les autres systèmes). On observe donc une tendance à considérer que le timbre des sons en binaural est plus **sombre**, donc avec une prépondérance de la partie basse du spectre sur la partie aiguë, que sur enceintes.

Entre les systèmes binauraux, la différence varie : référence et référence cachée présentent les mêmes valeurs, sauf en « bruit rose » (référence à 4.4 points, référence cachée à 4 points). En « classique » et en « rock », c'est l'AAC qui paraît le plus sombre, avec tout de même 1 point de moins que la référence et la référence cachée (à 3.5 contre 4.5), mais sur les « voix », il est en tête de 0.5 point (à 4.5 points), et le reste du temps il est compris entre les valeurs de la référence et de la référence cachée. Le MP3 est moins bien noté que la référence et la référence cachée, avec un écart 0.5 point sur le « rock » (4 points contre 4.5) et sur les « voix » (3.5 contre 4 points), mais il est en tête sur le stimulus « classique » (à 5 points, à égalité avec les enceintes, et avec 0.5 point d'avance sur la référence et la référence cachée).

En résumé, les sons paraissent globalement plus **sombres** en **binaural** que sur enceintes, encore que cela **varie** selon le **stimulus considéré** : aucune différence entre enceintes et binaural n'est ainsi relevée pour l'« ambiance » , tandis que la différence varie de 0 à 1.5 point en « classique » ; elle varie de 0.5 à 1.5 point en « rock » et sur les « voix », et de 1 à 1.5 point en « bruit rose » (stimulus pour lequel la différence est donc la plus sensible). Aucune différence très significative n'émerge entre les différents systèmes binauraux, même si des différences sont régulièrement perçues entre le binaural PCM, le MP3 et l'AAC, dans un sens (PCM plus brillant) ou dans l'autre (PCM plus sombre).

Résultat pour l'échelle : appréciation générale

Ce critère était noté de 0 : « mauvais » à 7 : « excellent ». Il s'agit sans doute du critère le plus **important**, dans la mesure où un système qui permettrait une retranscription parfaite de l'espace 5.0 mais serait jugé désagréable à l'écoute par tous les sujets n'aurait sans doute qu'un avenir commercial limité. Il apparaît que les systèmes binauraux ne sont pas systématiquement jugés moins bons que le 5.0 sur enceintes : pour l'« ambiance » , tous les systèmes sont notés à égalité (à 5 points), avec un écart interquartile un peu plus défavorable pour l'AAC. Pour le « bruit rose », les enceintes sont même au minimum (4 points), à égalité avec la référence, quand la référence cachée et l'AAC sont à 4.5 et le MP3 à 5 points. Pour le stimulus « classique », tous les systèmes sont à 4.5 points sauf la référence (à 4 points), et pour le « rock », les enceintes, la référence et la référence cachée sont notées à 4.5 points, avec 3.5 points pour l'AAC et 4 points pour le MP3. Il n'y a en définitive **que pour les « voix »** que les **enceintes** sont **clairement favorisées** : leur médiane est à 5, quand elle est à 4.5 pour tous les systèmes binauraux.

L'écart de 0.5 point entre référence et référence cachée est toujours présent (pour le « bruit rose » et la musique classique), même si c'est toujours ici en faveur de la référence cachée. Aucune tendance claire ne se dégage pour départager les différents systèmes binauraux : le MP3 est certes jugé meilleur pour le « bruit rose », et son écart interquartile le favorise pour le stimulus « classique », mais il est jugé moins bon que la référence pour le stimulus « rock » (4 points contre 4.5 pour la référence et la référence cachée). Pour ce même stimulus, l'AAC est jugé le moins bon (avec 3.5 points), mais partout ailleurs il est à égalité avec la référence ou avec la référence cachée.

Enfin, une observation de l'appréciation tous systèmes confondus, montre que l'« ambiance » est globalement la mieux notée (5 points en moyenne, avec beaucoup de réponses comprises entre 5 et 6 points) ; les « voix » sont notées à 4.5 points pour le binaural, 5 points pour les enceintes ; le « bruit rose » est noté entre 4 et 5 points (mais l'appréciation d'un tel stimulus est difficile à évaluer par les sujets, comme ceux-ci l'ont confié ou écrit dans leurs commentaires), le stimulus « classique » est noté entre 4 et 4.5 points, et le « rock » entre 3.5 et 4.5 points.

Les sujets semblent donc plus sévères dans leur notation sur les stimuli musicaux : musique classique et « rock », que pour les stimuli de voix ou d'ambiances, pour lesquels ils semblent assez tolérants. Cela vient peut-être de leur expérience d'écoute : il est probable que la majorité des sujets a plus souvent écouté de la musique que des voix seules ou une ambiance ; ils auraient donc davantage de points de comparaison dans ce domaine. Néanmoins il est difficile de généraliser davantage, car nous ne disposons que d'un stimulus par catégorie. Les différences entre les systèmes ne semblent pas laisser entrevoir de tendance générale, ni entre binaural et diffusion sur enceintes (où seules les « voix » témoigneraient d'une certaine préférence pour les hauts-parleurs), ni entre les différents systèmes binauraux.

Conclusion pour les échelles :

L'observation des échelles, avec les précautions que les conditions d'expérience imposent, nous permettrait de conclure que les systèmes binauraux sont parfois **aussi bien** voire **mieux notés** par les sujets que les hauts-parleurs, sur plusieurs critères. Ainsi, l'**immersion** offerte par les **systèmes binauraux** a été jugée au moins aussi **bonne** que celle proposée par les hauts-parleurs, sur l'« ambiance », le « classique » et le « rock » (sauf l'AAC). La **lisibilité** demeure apparemment meilleure avec les enceintes sur l'« ambiance », mais sur d'autres stimuli les notes du binaural sont égales, voire même meilleures en « classique » (sauf l'AAC) et en « rock ». Enfin, l'**appréciation** n'est clairement notée en faveur des enceintes que sur la « voix », et peut-être sur le « rock ». La **précision** a tendance à être mieux notée pour les enceintes, mais au moins un système binaural lui fait **concurrence**, sur l'« ambiance », le « bruit rose », et le « rock ». Si on laisse de côté la coloration, la grande majorité des notes pour les enceintes ne dépassent jamais celles des systèmes binauraux de plus de **0.5 à 1 point**, et les rares exceptions n'excèdent jamais 1.5 point (comme pour la précision en musique classique).

L'échelle de **coloration** montre que le passage en **binaural** présente une **altération manifeste du spectre**, qui est jugé plus **sourd**, plus **sombre**, pour l'ensemble des stimuli, exceptée l'« ambiance ».

Ces différents résultats montrent donc que plusieurs caractéristiques importantes du son sont globalement préservées, voire préférées, en binaural par rapport aux enceintes. De plus la **comparaison** entre les différents **systèmes binauraux** ne permet pas de dégager de **différence manifeste**, même si l'AAC et le MP3 semblent récolter plus d'une fois des notes distinctes de la référence, comme de la référence cachée. Ces résultats semblent cependant **dépendre**

du stimulus considéré (voir notamment les échelles de coloration, pour les « voix », le « rock », et le « classique »).

2.4.9 Les résultats sur les HRTF

Rappelons que le test sur les HRTF consistait en la diffusion dans le casque des sujets d'un bruit rose effectuant un tour de tête qui partait de leur droite, passait derrière leur nuque et revenait devant eux à vitesse constante. Ce trajet intervenait 8 fois : une première fois, convolué par la HRTF Best Matching 2 utilisée pour le reste du test en binaural, et sept fois, convolué par les HRTF Min subset 1 à 7 proposées par SpherAudio. La préférence des sujets pour l'une des Min subset est montrée dans la figure 2.24.

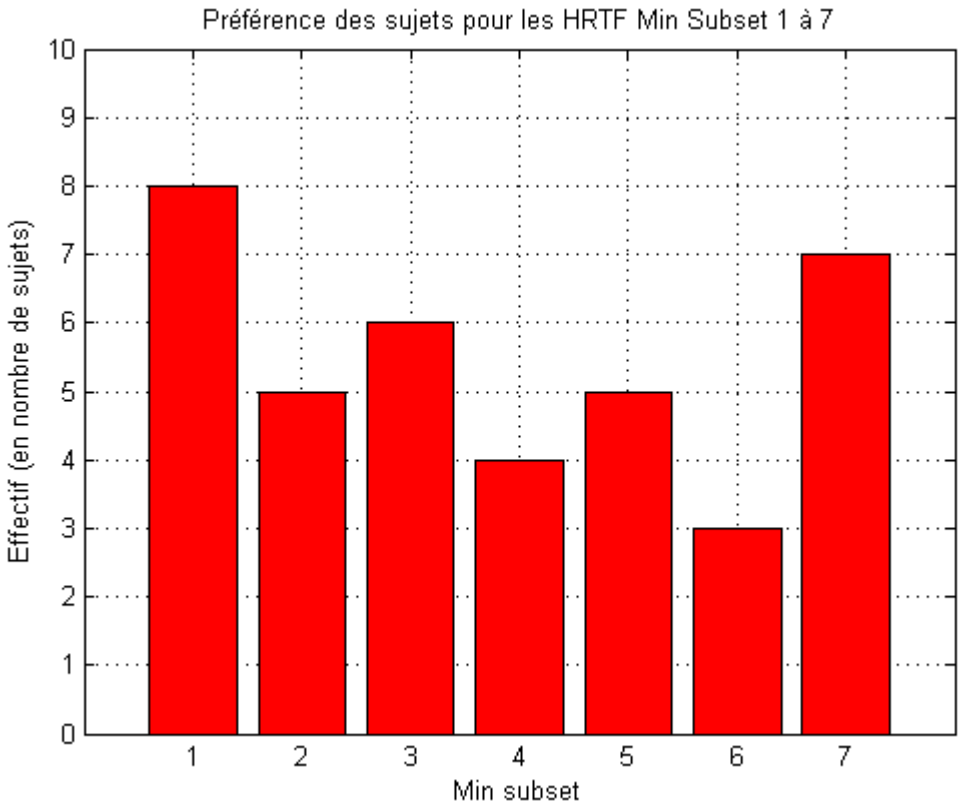


FIGURE 2.24 – Représentation de la préférence, par les sujets de l'expérience, pour les différentes HRTF Min subset 1 à 7 proposées par SpherAudio.

Cet histogramme représente le nombre de sujets pour lesquels chaque Min subset semble être la HRTF qui fonctionne le mieux (la HRTF Best Matching 2 mise à part). Elle regroupe les réponses des sujets de DMS et de l'école Louis Lumière, pour lesquels on a considéré que la différence de conditions expérimentales n'empêchait pas d'inclure leurs préférences dans cette analyse. La préférence pour telle ou telle HRTF était observée à l'oeil d'après les trajets dessinés par les sujets : pour chaque sujet, on a ainsi évalué lequel des huit trajets dessinés était le plus proche de celui paramétré dans SpherAudio. Les résultats nous

montrent que les effectifs sont assez **répartis** : si la Min subset 1 semble en tête avec 8 sujets, la numéro 7 regroupe 7 sujets, et la numéro 3, 6 sujets. La numéro 6 arrive dernière mais avec tout de même 3 sujets. Ces chiffres nous montrent, d'une part, que la **question** de l'**individualité des HRTF** est **réelle**, puisqu'il serait difficile de décider qu'une de ces 8 HRTF n'est pas utile et ne convient à personne ; d'autre part, qu'il n'y a pas eu d'apprentissage, de **phénomène de compréhension**. C'est-à-dire, pour reprendre l'expression de Blauert ([3], p. 417), qu'il ne semble pas y avoir eu de « compréhension de la mélodie spatiale » (« *picking up the spatial melody* »), qui aurait permis au sujet de comprendre en cours de test le mouvement voulu pour le bruit rose. Si un tel apprentissage avait eu lieu, le sujet aurait probablement réalisé, au fur et à mesure des écoutes, des dessins de plus en plus proches de la trajectoire paramétrée dans le logiciel. Les Min subset étant toujours présentées de la numéro 1 à la numéro 7, on aurait alors vu une croissance de l'effectif associé aux Min subset de 1 à 7, or ce ne semble pas être le cas (on aurait même plutôt l'image globale d'une décroissance de l'effectif de la Min subset 1 à 6, suivie d'une remontée sur la 7).

On remarquera que la somme des effectifs évoqués ne totalise que 38 sujets, trois sujets n'ayant pas pu faire l'expérience sur HRTF pour des raisons de fatigue ou d'emploi du temps.

Le second schéma (fig. 2.25) nous montre le pourcentage des sujets (DMS et Louis Lumière confondus) préférant la HRTF **Best Matching 2** devant toutes les Min subset. Une distinction a été faite, selon qu'ils avaient effectué le test HRTF avant (1ère session) ou après (2è session) le test binaural : on voulait ainsi observer l'influence de l'**apprentissage** avec des HRTF étrangères.

Les résultats nous indiquent que seulement **17%** des sujets ont manifesté une préférence pour la Best Matching 2 par rapport à toutes les Min subset, toutes sessions confondues. **10%** des sujets ont manifesté cette préférence alors qu'ils ont réalisé le test HRTF en **première session** (**avant la session binaurale** ; on peut donc considérer que Best Matching 2 est pour ces sujets la HRTF qui leur correspond le **mieux**, parmi celles proposées par SpherAudio), tandis que seulement **7%** ont manifesté cette préférence alors qu'ils avaient réalisé la **session binaurale avant** le test HRTF. Cela signifie que les sujets ayant d'abord suivi une session de test de 40 minutes durant laquelle ils ont côtoyé la HRTF Best Matching 2, ne se sont pas forcément sentis plus à l'aise avec cette HRTF lors du test HRTF qui a conclu le test sur enceintes. Il n'y aurait donc, visiblement, **pas eu d'effet d'apprentissage au cours de notre session binaurale**. Le fait que la proportion de sujets ayant préféré la Best Matching 2 toutes sessions confondues ne dépasse pas les 17% montre par ailleurs que même cette HRTF, censée convenir à un certain nombre de personnes, ne peut **remplacer** pour tout le monde l'usage d'une des Min subset. La problématique de l'individualisation des HRTF reste donc entière.

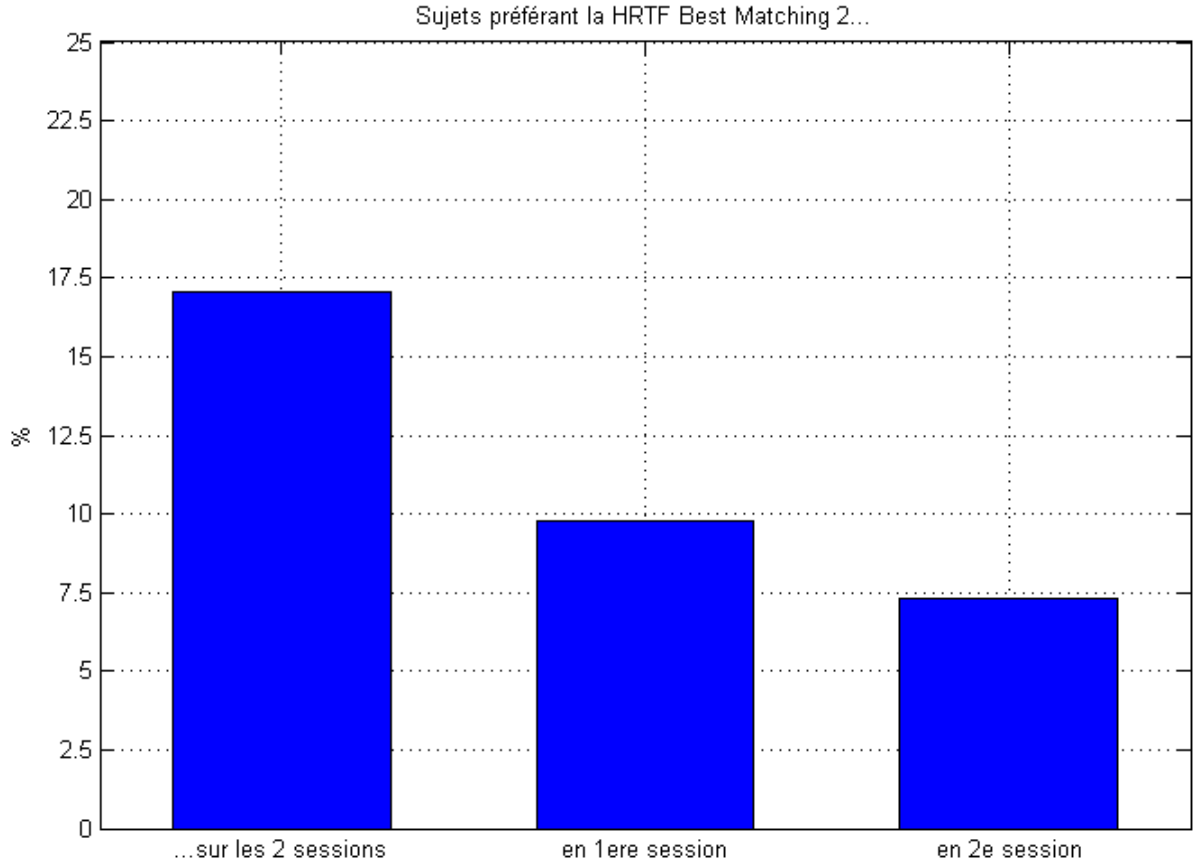


FIGURE 2.25 – Représentation du pourcentage des sujets ayant préféré la HRTF Best Matching 2 à l'ensemble des Min subset qui leur ont été présentées. « En 1ère session » : ils ont expérimenté les HRTF avant d'avoir fait la session de test binaurale ; « en 2è session » : ils ont expérimenté les HRTF après avoir fait la session de test binaurale.

2.5 Limites du test réalisé dans le cadre de ce mémoire

Ce test a été réalisé avec le logiciel SpherAudio, et les résultats sont donc dépendants des performances de ce logiciel, et des paramètres choisis lors de son utilisation, en terme d'ajout de room, de choix de la HRTF (Best Matching 2), d'égalisation gauche-droite... Ils ne sont également valables que pour les stimuli employés, et notamment les choix de mixage réalisés (or on a vu que le mixage semblait avoir une influence sur les réponses des sujets, notamment en ce qui concernait la perception de la distance).

Parmi les principales limites de notre test, on peut citer que celui-ci a été réalisé avec des stimuli statiques. Il est probable que l'utilisation de stimuli sonores en mouvements aurait donné des résultats différents, notamment en terme de précision de localisation. En effet, la combinaison des HRTF nécessaires à la

convolution joue dans ce cas le rôle d'une combinaison de HRTF par mouvements de tête dans le monde réel, et permet potentiellement de ce fait, d'affiner la localisation des sources.

Nos tests ont été réalisés sans image, ils ont donc écarté les nombreuses problématiques que fait probablement surgir l'ajout d'une image, eu égard aux modifications que le visuel implique sur la localisation des sources.

La diffusion a été réalisée sur un système 5.0 ITU aménagé à DMS, et sur un seul modèle de casque (Sennheiser HD650). La liberté laissée aux sujets de contrôler leur niveau d'écoute peut constituer une limitation supplémentaire à la généralisation de nos résultats.

On pourra émettre l'hypothèse que des échelles graduées, impaires, influent davantage sur les réponses des sujets qu'une échelle paire (qui ne permet pas de juste milieu) ou qu'une échelle non-graduée. Les graduations ont été ajoutées pour faciliter la quantifications des résultats, mais nous avons pris en compte les réponses placées par les sujets à des valeurs non-entières entre deux graduations.

Par ailleurs, la comparaison des résultats des échelles pour les enceintes et pour les système binauraux, pourrait être considérée comme partiellement biaisée, par le fait que la notation des échelles pour une diffusion sur enceintes et au casque avait lieu lors de sessions séparées.

Nous n'avons pas réalisé de tests statistiques à partir de nos résultats d'études : Anova ou « t » test de Student, test de Kolmogorov-Smirnov pour vérifier la distribution de nos résultats... Ces tests auraient pourtant pu apporter des précisions complémentaires pour notre analyse. Des contraintes de temps justifient en partie cette lacune, mais nous avons aussi décidé de laisser de côté ces tests lorsque nous avons constaté que nos résultats ne suivaient pas toujours une distribution normale (ce qui rendait difficile l'application d'une Anova). Par ailleurs, ce genre de test se base sur la théorie des grands nombres, alors que nous ne disposions que de 24 sujets pour nos résultats. Nous n'étions donc pas convaincus de la pertinence de ces tests pour notre étude.

Conclusion générale

Les différents tests réalisés dans le cadre de ce mémoire ont permis de montrer certaines limites et performances du binaural pour la diffusion de stimuli en 5.0 au casque par rapport à une diffusion sur enceintes en 5.0 ITU. Ils ont également permis de comparer le rendu en binaural PCM, AAC et MP3. Les résultats de nos tests ont montré que l'espace sonore 5.0 ITU semble loin d'être retranscrit tel quel en binaural : la binauralisation semble provoquer un **écrasement** des sources vers la **ligne interaurale**, à l'**avant** comme à l'**arrière**, voire un **repliement** par **confusion avant-arrière**, un risque d'**internalisation** pour les sources venant de **devant**, un **rapprochement** des sources... Les distorsions en azimut et en distance semblent **dépendantes de l'azimut** du stimulus considéré. Elles semblent aussi dépendre de la **nature** de ce stimulus : non seulement son enveloppe (stimulus riche en transitoire, ou au contraire constitué de sons tenus), mais aussi son spectre, et ce que l'on pourrait appeler son « contexte » , c'est-à-dire la cohérence de sa position par rapport au reste de la scène sonore

(par le même phénomène qui fait percevoir une élévation sur des sons d'oiseaux diffusés en stéréo). Plus particulièrement, les distorsions en distance en binaural semblent très liées au **mixage** lui-même : par l'usage de filtres, de variations du niveau sonore, de réverbération.

Cependant ces constats sont à pondérer, d'une part du fait des défauts de représentation par les sujets de leur environnement, mis en exergue par le test lumineux ; d'autre part, par les défauts de la diffusion en 5.0 ITU sur enceintes, qui montrerait une certaine tendance à l'**aspiration** des sources par les hauts-parleurs, et des incertitudes de localisation dès lors que la source est située entre les hauts-parleurs avant et arrière ou entre les hauts-parleurs arrière, incertitudes se manifestant par des écrasements d'azimut voire des repliements de l'arrière vers l'avant.

Entre le binaural PCM, AAC 192 et MP3 192, **aucune différence** n'a pu être constatée avec **certitude**, surtout si l'on prend en compte les écarts constatés d'avec la référence cachée. Notre expérience tendrait donc à montrer que la conversion du signal binaural en AAC 192 ou en MP3 192 n'aurait **pas d'influence significative** sur la perception de l'**espace sonore**.

Les échelles nous ont permis de constater que le binaural, s'il se révèle globalement inférieur à la diffusion sur hauts-parleurs pour des critères de précision ou de lisibilité (avec déjà quelques exceptions), peut donner un sentiment d'immersion légèrement supérieur et obtenir à plusieurs reprises une meilleure note quant à l'appréciation globale. En revanche, une modification de la coloration, caractérisée par un renforcement de la partie basse du signal en binaural par rapport à la partie haute, semble clairement audible comparée aux hauts-parleurs. Ces constatations ne sont cependant que des tendances générales. Là encore, entre les différents formats de sons binauraux, on ne constate pas de différence qui semble particulièrement significative.

On pourrait donc conclure, en lien avec les problématique du binaural d'aujourd'hui :

- que la binauralisation d'un master 5.0 a des conséquences variables, en termes d'image sonore et de qualité de son. Ces conséquences dépendent de plusieurs paramètres, notamment l'azimut attendu, la nature du stimulus traité, et les caractéristiques du mixage ; le mixeur peut donc conserver un rôle de premier plan dans la recréation de l'espace sonore en binaural à partir de son mixage 5.0 (à noter que les ingénieurs du son de Radio France avaient déjà remarqué une tendance à un écrasement à l'arrière en binaural, qu'ils avaient entrepris de corriger au moment de la binauralisation ; on pourrait estimer que leur démarche serait donc pertinente) ;

- que la conversion du signal binaural en AAC 192 (tel que le fait Radio France dans le cadre du projet NouvOson), ou en MP3 192, **ne semble pas altérer significativement** le signal ni sur le plan de la localisation obtenue, ni sur les plans de la précision, de la lisibilité, de l'immersion, de la coloration, et de l'appréciation générale par les auditeurs ;

- qu'en dépit de ses défauts dans la retranscription de l'espace 5.0, le binaural semble apporter des solutions **satisfaisantes** en terme d'**immersion** et de **plaisir d'écoute**, alors même que les sujets n'étaient pas forcément **habitués** à ce **mode d'écoute**, qui requiert sans doute, comme tout nouveau système, un temps d'accoutumance ;

- plus largement, que les faiblesses apparentes du binaural ne doivent pas constituer un obstacle à son utilisation ; si la retranscription de l'espace est encore imparfaite, ce système possède un potentiel intéressant et des capacités propres à inspirer la création sonore et la réalisation d'effets pour le divertissement, au-delà de ses difficultés pour restituer en toute rigueur un espace sonore précis ;

- que les grandes limites du binaural aujourd'hui demeurent les difficultés dans la combinaison des HRTF (limite peut-être repoussée dans un avenir proche par le développement du head tracking), et l'individualisation des HRTF (problématique dont toute la complexité a été partiellement montrée par notre test HRTF).

Nous pensons donc que le binaural possède un avenir porteur : non seulement il demeure une technologie 3D facile à insérer dans les habitudes d'écoute et d'équipement du consommateur (en regard de systèmes complexes tels que la Wave Field Synthesis, l'ambisonie ou les formats multicanaux comme l'Auro 3D), mais il semble robuste aux traitements de réduction de débit, et constitue un espace sonore à explorer qui nous semble particulièrement riche : il ne faut pas oublier que le 5.0 lui-même est loin d'être exempt de défauts sur le plan de la restitution de l'espace sonore, ce qui ne l'empêche pas d'avoir été adopté pour de nombreuses utilisations et de permettre la création d'une expérience sonore spécifique ; de la même manière, les limitations du binaural actuel ne nous semblent pas devoir empêcher l'exploration des possibilités sonores qu'il offre.

Bibliographie

[1] Forum AVS. "Dolby Atmos Theatre System", p.10, Juin 2012. www.avsforum.com/t/1407030/dolby-atmos-theatre-system/270, consulté en avril 2013.

[2] Durand R. BEGAULT. *3-D Sound for Virtual Reality and Multimedia.* NASA, Ames Research Center, Moffett Field, California, 2000.

[3] Jens BLAUERT. *Spatial Hearing, Revised Edition.* The MIT Press, 1995.

[4] Pierre BOMPY. *Contrôle au casque d'une prise de son multicanale*, mémoire de fin d'étude, sous la direction de Benjamin BERNARD et Mohammed ELLIQ. Master's thesis, École Nationale Supérieure Louis Lumière (section Son), 2008.

[5] Florent CASTELLANI. *Adaptation du mixage cinéma à l'écoute au casque*, mémoire de fin d'étude, sous la direction de Benjamin BERNARD et Claude GAZEAU. Master's thesis, École Nationale Supérieure Louis Lumière (section Son), 2011.

[6] Michel CHION. *Un art sonore, le cinéma.* éd. Cahiers du Cinéma Essais, 2003.

[7] Brian F.G. KATZ et Fabien PREZAT. "The Effect of Audio Compression Techniques on Binaural Audio Rendering". In *Audio Engineering Society Convention 120*, 2006.

[8] Nick ZACHAROV et Kalle KOIVUNIEMI. "Unravelling the perception of spatial sound reproduction : Techniques and experimental design". *AES*, 2001.

[9] Christian HUGONNET et Pierre WALDER. *Théorie et pratique de la prise de son stéréophonique.* Eyrolles, 2005.

[10] Ville PULKKI et Tapio LOKKI. "Creating Auditory Display with Multiple Loudspeakers Using VBAP : A Case Study with DIVA Project". *ICAD*, 1998.

[11] Alain GOYÉ. "La perception auditive". Cours Telecom Paris, 2002.

[12] Bernard LAGNEL. "Prise de son multicanale et binaurale". In *Audio Engineering Society*, janvier 2013. Conférence à Radio France.

[13] Fritz MENZER. "Efficient Binaural Audio Rendering Using Independant Early and Diffuse Paths". In *Audio Engineering Society Convention 132, Budapest*, avril 2012.

[14] Nerds and Art. "Espace sonore et localisation", décembre 2010. nerdsandart.blogspot.fr/2010/12/espace-sonore-et-localisation.html.

[15] Rozenn NICOL. *Représentation et perception des espaces auditifs virtuels*, Mémoire d'Habilitation à Diriger des Recherches. Master's thesis, Juin 2010.

[16] Baptiste PALACIN. *Le rôle du son dans la représentation de la peur dans le jeu vidéo*, mémoire de fin d'étude, sous la direction de Sébastien GENVO et Claude GAZEAU. Master's thesis, École Nationale Supérieure Louis Lumière (section Son), 2013.

[17] Gaëtan PARSEIHIAN. *Sonification binaurale pour l'aide à la navigation, thèse de doctorat.* PhD thesis, Univeristé Pierre et Marie Curie, LIMSI, 2012.

[18] *Poème à dire, choisis par Daniel Gélin.* éd. Seghers, 2003.

[19] Ville PULKKI. "Virtual Source Positionning Using Vector Base Amplitude Panning". *AES*, 1997.

[20] Raffi KRIKORIAN, Samidh CHAKRABARTI, Sharmila SINGH, and Boris ZBARSKY. "Localization of Sound in Micro-Gravity". URL : web.media.mit.edu/ raffik/zero-g/aup/final.html, consulté en avril 2013.

[21] *Recommandation UIT-R BS.1116-1, "Méthodes d'évaluation subjective des dégradations faibles dans les systèmes audio y compris dans les systèmes sonores multivoies"*, 1994-1997.

[22] *Recommandation UIT-R BS.1534-1, "Méthode d'évaluation subjective du niveau de qualité intermédiaire des systèmes de codage"*, 2001-2003.

[23] Guillaume ROUX-GIRARD. *L'écoute de la peur : une étude du son dans les jeux vidéo d'horreur.* PhD thesis, Université de Montréal, Département d'histoire de l'art et d'études cinématographiques, Faculté des arts et des sciences, Décembre 2009.

[24] Sylvia SIMA. *HRTF Measurements and Filter Design for a Headphone-Based 3D-Audio System, Bachelorarbeit.* PhD thesis, Department Informatik der Fakultät Technik und Informatik der Hochschule für Angewandte Wissenschaften Hamburg, 2008.

[25] Site officiel d'Auro 3D. www.auro-technologies.com/.

[26] Site officiel de Flux, Ircam HEar. www.fluxhome.com.

[27] Site officiel de Digital Media Solutions. www.dms-cinema.com/fr/technologies.html.

[28] Site officiel de DMS, BP84. www.dms-cinema.com/fr/products/pro/audio/processeurs.html.

[29] Site officiel de Dolby, Dolby Headphone. www.dolby.com/us/en/consumer/technology/home-theater/dolby-headphone.html.

[30] Site officiel de Longcat, Longcat 3D. www.longcat.fr/web/fr/prods/h3d-plugin.

[31] Site du projet Listen, Ircam.
http ://recherche.ircam.fr/equipes/salles/listen/, consulté en avril 2013.

[32] Günther THEILE. "Principles and Applications of Stereophony, Binaural Techniques and Wave Field Synthesis". In *Tonmeistertagung Leipzig*, 2006.

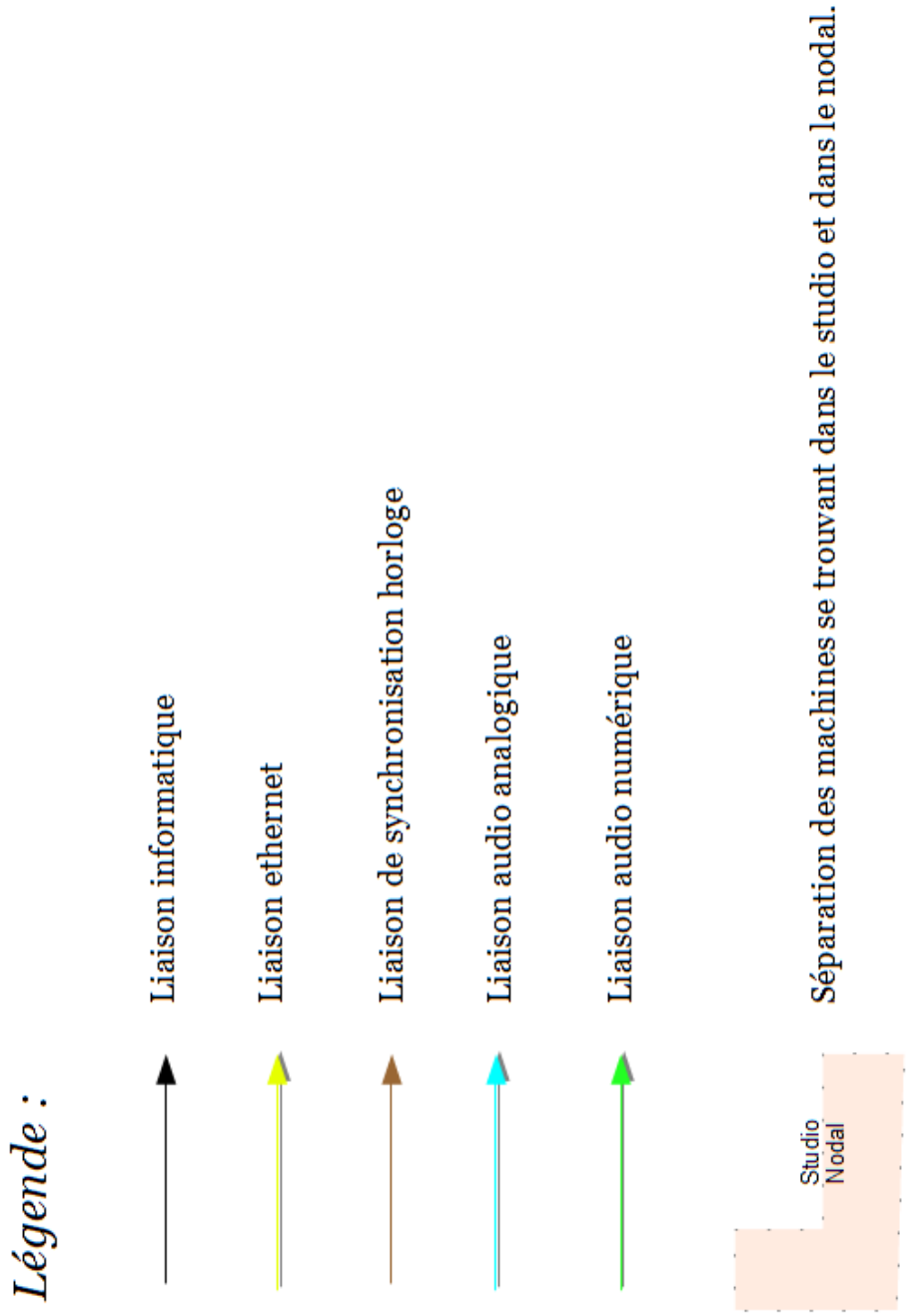
Table des figures

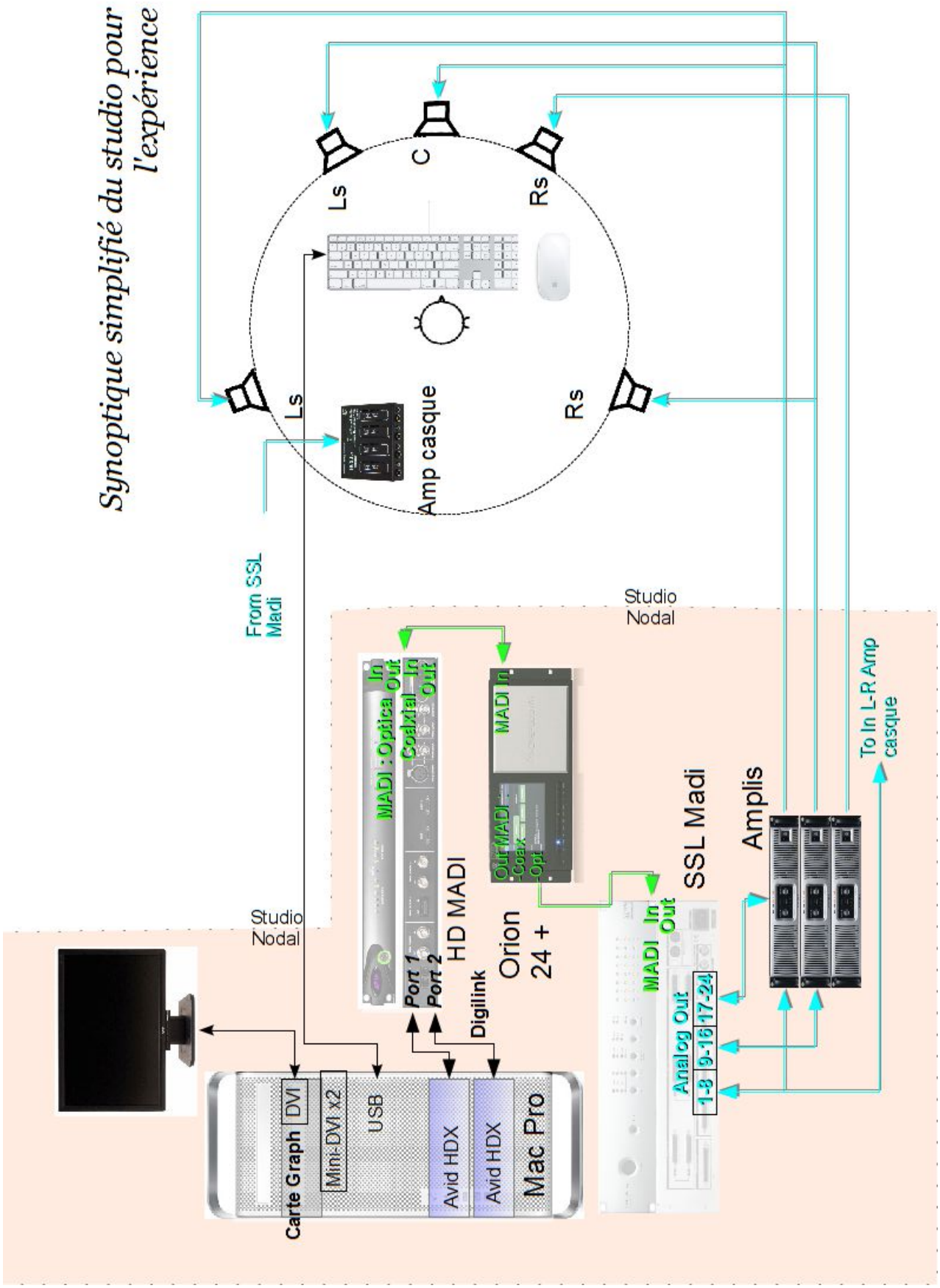
Liste des tableaux

Annexe A

Mise en place de l'expérience

A.1 Synoptique du studio




Synoptique simplifié du studio pour l'expérience
Ls
C
Rs
Ls
Rs
Amp casque
From SSL Madi
Studio Nodal
MADI : Optica
In Out
Coaxial In Out
Port 1
Port 2
HD MADI
Digilink
Orion 24 +
MADI In
Out MADI
Coax
Opt
SSL Madi
MADI In Out
Analog Out
1-8 9-16 17-24
Amplis
To In L-R Amp casque
Studio Nodal
Carte Graph
DVI
Mini-DVI x2
USB
Avid HDX
Avid HDX
Mac Pro

A.2 Texte enregistré accompagnant l'expérience :

En introduction :

« Au cours de ce test, vous devrez localiser différents éléments sonores au sein de différents stimuli. Vous indiquerez, sur le schéma circulaire, les zones de l'espace dans lesquels ces éléments se situent, selon vous, par rapport à votre tête. Les éléments à localiser seront à chaque fois précisés à côté du schéma circulaire.

« Ainsi, vous entendrez une ambiance de forêt, au sein de laquelle on vous demandera de localiser : le corbeau (+ solo corbeau), la branche qui craque (+ solo), le buisson (+ solo).

« Vous entendrez un extrait d'un morceau rock, dans lequel on vous demandera de localiser la caisse claire (+ solo), la guitare mélodique (+ solo), et la voix.

« Dans la récitation d'un poème, il s'agira de localiser les 4 voix successives que vous entendrez.

« Dans un extrait de musique classique, on vous demandera de localiser la flûte (+ solo), le basson (+ solo), et la voix.

« Ces solos vous seront répétés avant chaque stimulus. Pour les passer, vous pourrez appuyer une fois sur la touche Entrée.

« Enfin, dans un fond de bruit rose, on vous demandera de localiser les 3 salves successives de bruit rose (+ salve).

« Quand vous aurez localisé ces sons, il vous sera demandé d'indiquer votre appréciation de certains critères sonores, en plaçant une croix sur une échelle de 0 à 7. Si vous ne savez pas comment noter un stimulus, vous pouvez mettre un point d'interrogation, à côté de l'échelle correspondante. Les définitions des critères à noter sont idiqués sur une feuille à votre disposition. Une feuille-réponse type est également posée sur votre table pour vous aider.

« Vous pourrez lancer ou arrêter la lecture en mode play/pause, avec la barre d'espace du clavier. Vous pouvez ré-écouter un stimulus en actionnant les touches « ctrl flèche de gauche », ou « ctrl flèche de gauche flèche de gauche » si vous êtes déjà passé au stimulus suivant.

« Vous pouvez adapter votre niveau d'écoute en glissant votre doigt vers le haut ou vers le bas sur la souris. »

Complément d'intro pour les sessions sur enceintes (séries 1, 2, 3) : « Pendant les écoutes, nous vous demanderons autant que possible d'éviter de tourner la tête, même si vous êtes en train d'écrire. Une lumière rouge est posée devant vous pour vous servir de référence. Vous n'avez pas à tenir compte de la position des enceintes dans vos réponses.

« Le test se conclura par une expérience au casque. »

Complément d'intro pour les sessions de test binaurales (séries 4, 5, 6) : « En introduction à cette expérience, nous allons vous initier rapidement

au binaural. Une vois va lentement faire le tour de votre tête, en vous signalant sa position au fur et à mesure en terme d'heure (+ voix « 3h... 12h »). »

Fin de l'introduction : « Si vous avez maintenant des questions, appuyez sur la barre d'espace et venez m'appeler. Pour commencer directement le test, appuyez une fois sur la touche Entrée. »

Avant chaque stimulus :

Indication de la référence du stimulus « 11M1, 52B », etc.

Avant les stimulus ambiance, rock, classique :

« En solo : » (+ le nom des solos correspondants et le solo proprement dit)

2è partie de la session de test sur enceintes (séries 1, 2, 3) :

« Dans la dernière phase du test, nous vous demanderons de mettre le casque audio posé sur la table. Attention au sens : la bague jaune et verte sur le câble indique l'oreille gauche. Quand ce sera fait, appuyez une fois sur la touche Entrée.

« Vous allez entendre 8 déplacements successifs d'un bruit rose qui partiront de votre droite. Nous vous demanderons d'indiquer sur le schéma circulaire, en traçant une flèche, le trajet que ces bruits roses ont réalisé selon vous. Vouspouvez ré-écouter chaque bruit rose en appuyant sur Ctrl flèche de gauche. Vous pouvez régler votre niveau d'écoute grâce à la barrette numéro 1 du boîtier casque, repérée par une bague jaune et verte. Pour commencer le test, appuyer une fois sur la touche Entrée. »

Avant chaque bruit rose : Indication de la référence du bruit rose, de « 1 » à « 8 ».

En conclusion :

« Le test est terminé. Merci de votre participation. »

A.3 Feuille de définitions à la disposition des sujets

Définition des critères demandés :

Sentiment de précision : La position des sources à localiser est-elle fixe, invariante au cours du temps, facile à délimiter (donc : **sentiment bien défini**) ou au contraire floue, changeante, incertaine (donc : **sentiment mal défini**) ?

Lisibilité : La scène sonore dans son ensemble est-elle explicite, comprend-on tout ce qui s'y passe (lisibilité **distincte**), ou les sons sont-ils noyés en un ensemble confus que l'on n'appréhende pas très bien (lisibilité **confuse**) ?

Sentiment d'immersion : A-t-on l'impression d'être ancré dans la scène sonore, « comme si on y était » (sentiment d'immersion **excellent**), ou au contraire, d'en être détaché, de « ne pas y être » (sentiment d'immersion **mauvais**) ?

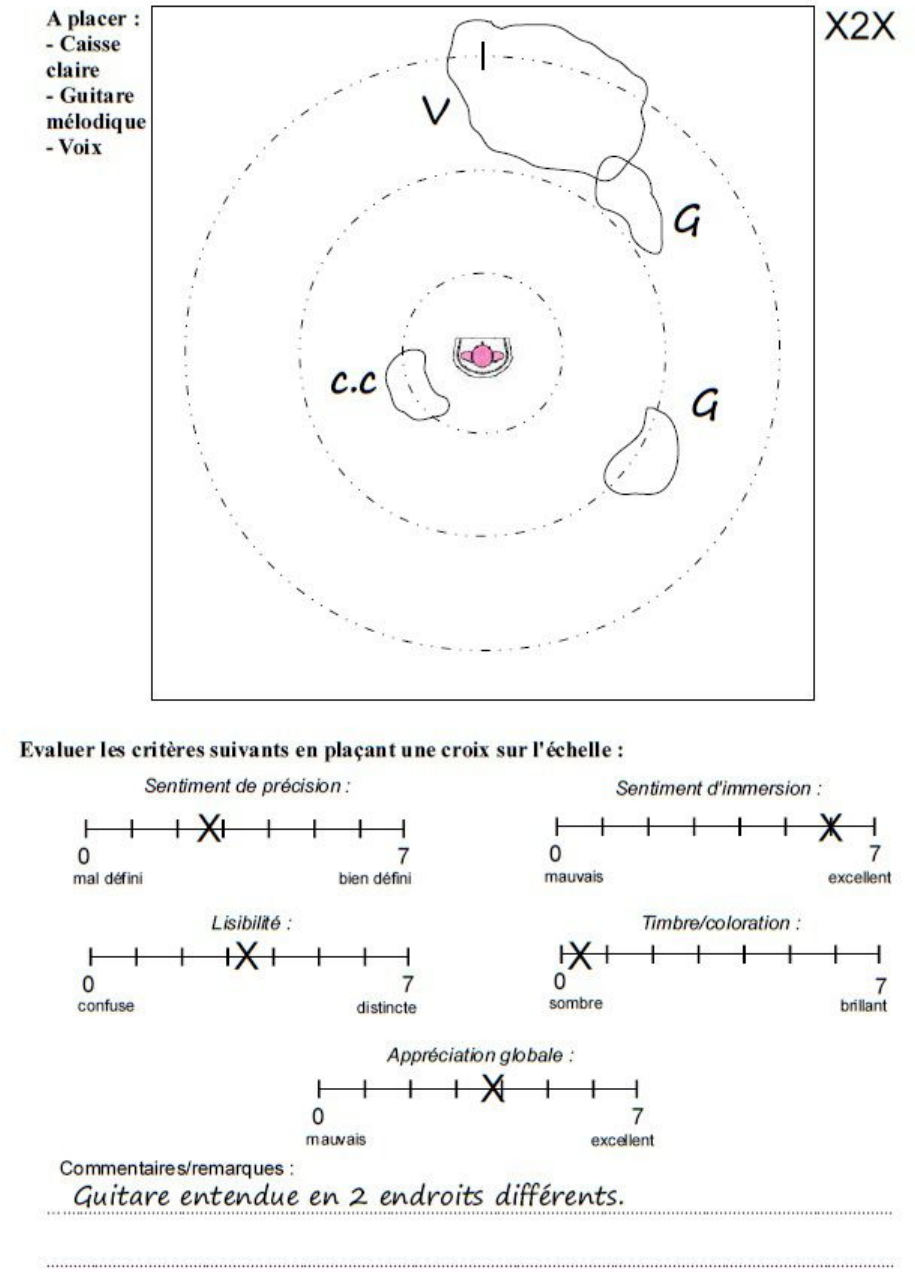
Timbre/coloration : Le timbre global de la scène sonore donne-t-il l'impression d'être plutôt sombre, un peu sourd, voire étouffé (timbre/coloration **sombre**) ou au contraire très clair, présent, brillant (timbre/coloration **brillant**) ?

Appréciation globale : L'enregistrement que vous venez d'entendre vous paraît-il plutôt mauvais ou plutôt excellent ?

Rappel :
Barre d'espace : Play/pause
Ctrl + flèche de gauche pour ré-écouter ce stimulus du début
Ctrl + flèche de gauche + flèche de gauche pour ré-écouter le stimulus précédent.
Entrée : passer les solos.

*Casque : la bague jaune et verte indique l'oreille **gauche**.*
***Sweet spot** du système d'enceintes : le corps calé contre la table, les jambes entre les arceaux de la table.*

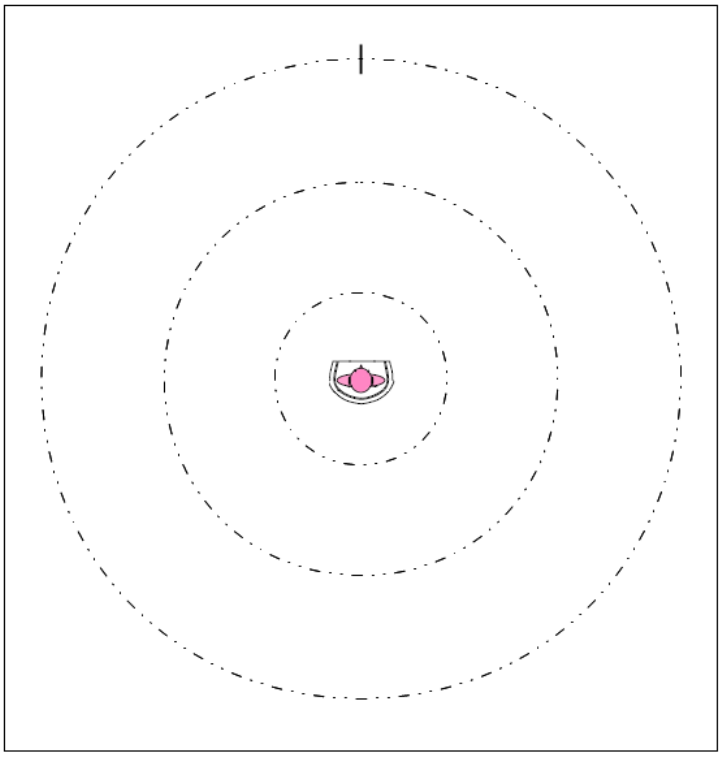
A.4 Feuille-réponse exemple à la disposition des sujets



A.5 Exemple de feuille-réponse vierge :

A placer :
- Corbeau
- Branche
- Buisson

41A



Evaluer les critères suivants en plaçant une croix sur l'échelle :

Sentiment de précision :

0 — 7
mal défini — bien défini

Sentiment d'immersion :

0 — 7
mauvais — excellent

Lisibilité :

0 — 7
confuse — distincte

Timbre/coloration :

0 — 7
sombre — brillant

Appréciation globale :

0 — 7
mauvais — excellent

Commentaires/remarques :

Annexe B

Schémas réponses des sujets

B.1 Centres de gravité et ellipses de variance

B.1.1 Sur hauts-parleurs

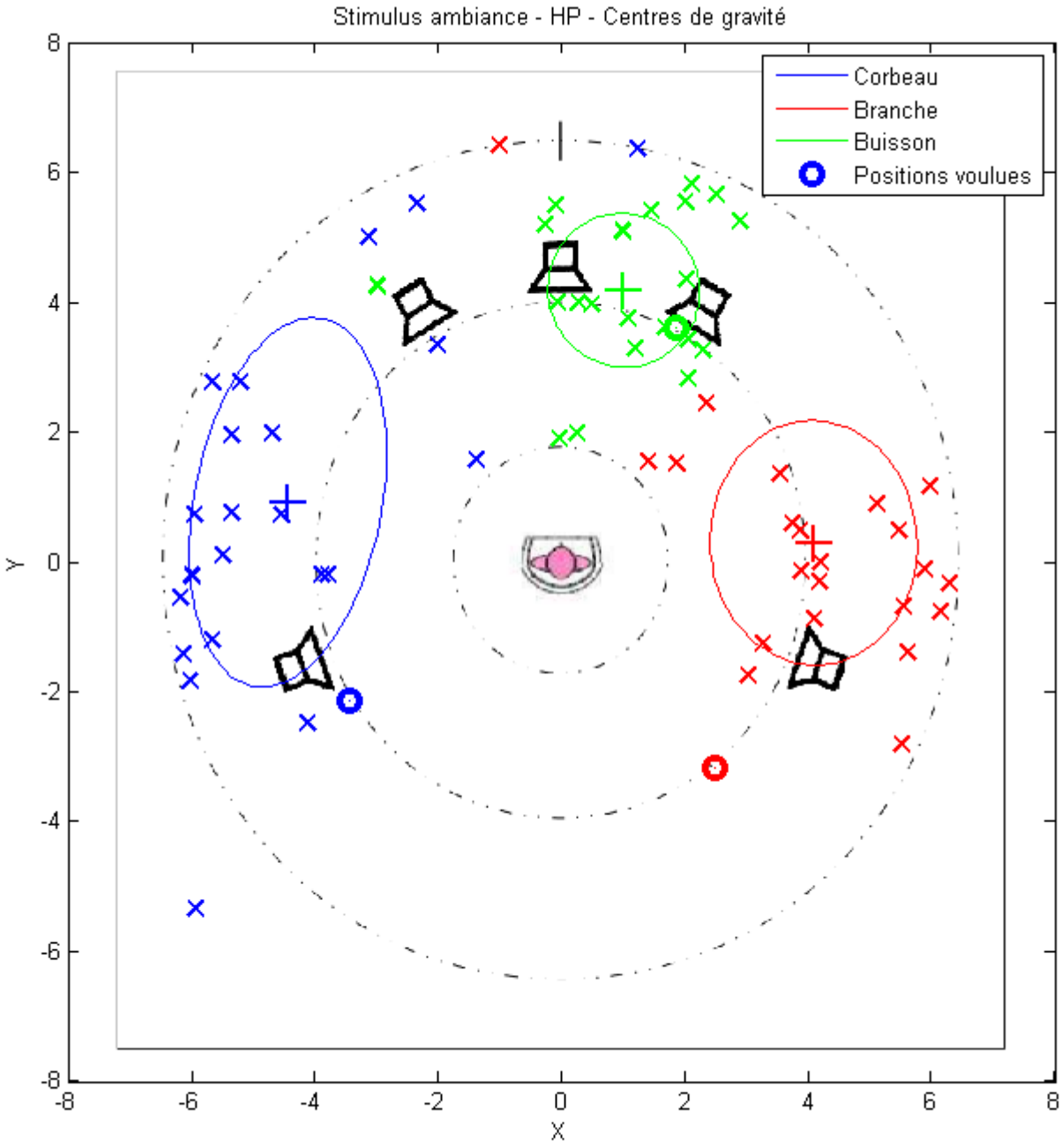


FIGURE B.1 – Centres de gravité et ellipses de variance obtenus pour l'ambiance, sur HP.

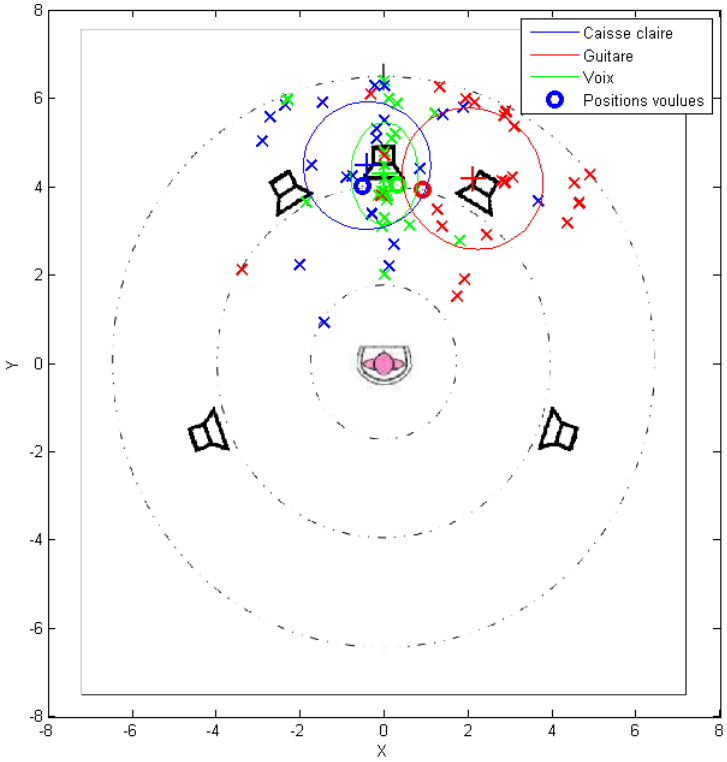


FIGURE B.2 – Centres de gravité et ellipses de variance, rock, HP.

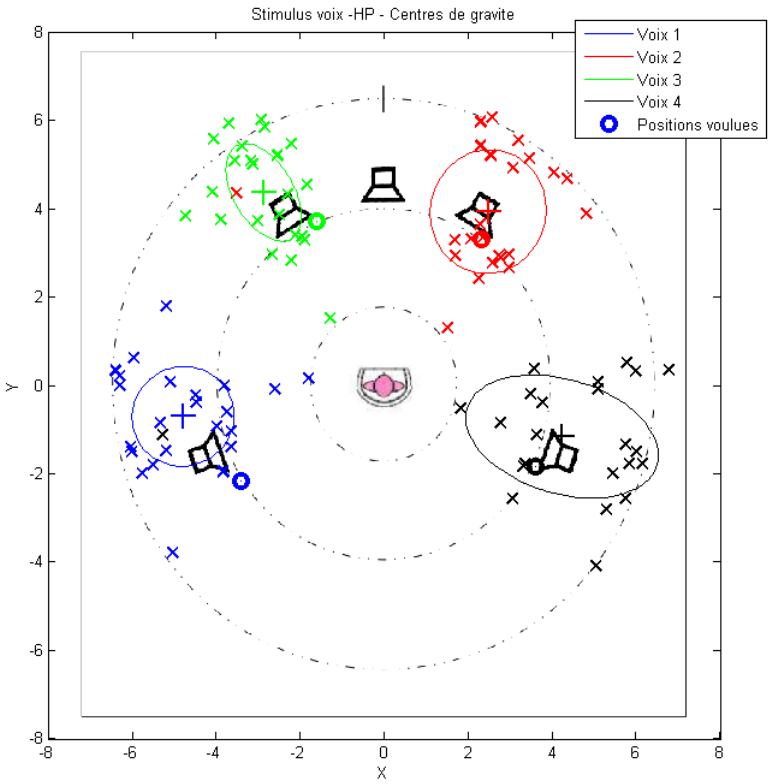


FIGURE B.3 – Centres de gravité et ellipses de variance, voix, HP.

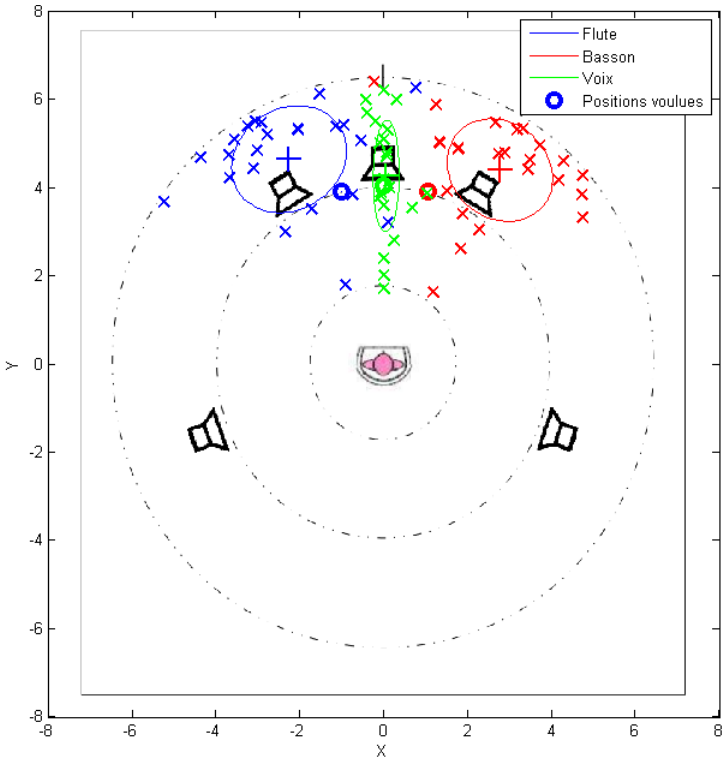


FIGURE B.4 – Centres de gravité et ellipses de variance, classique, HP.

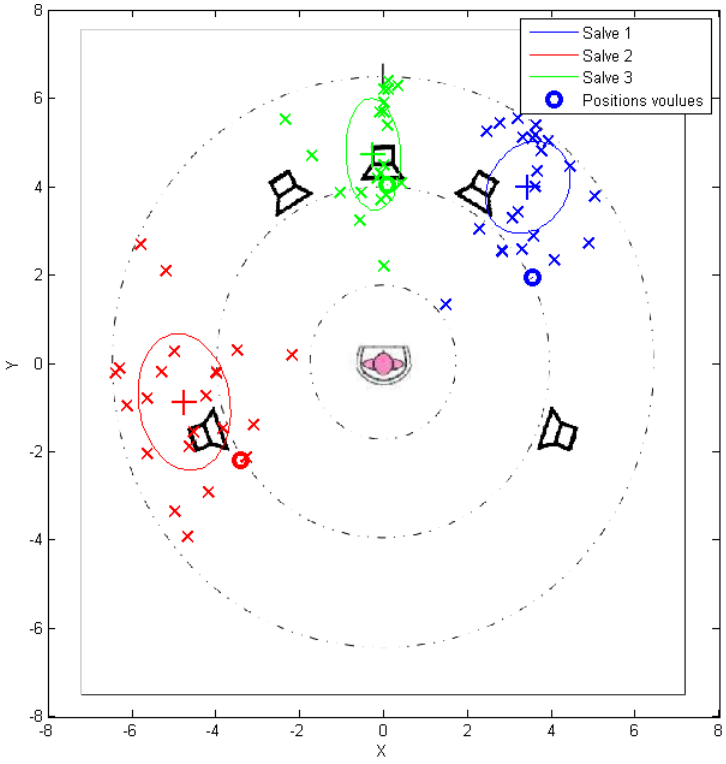


FIGURE B.5 – Centres de gravité et ellipses de variance, bruit rose, HP.

B.1.2 En binaural référence

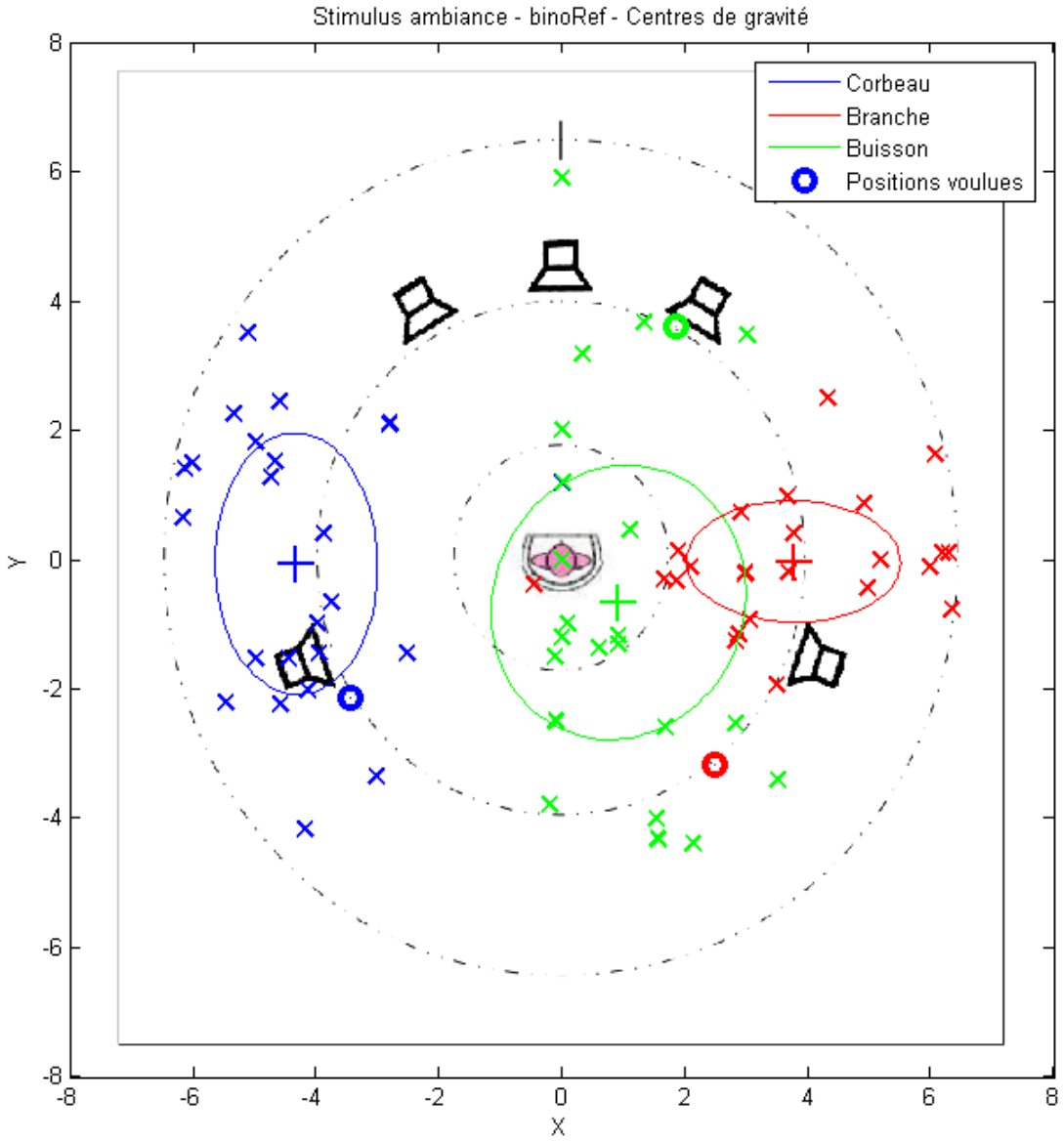


FIGURE B.6 – Centres de gravité et ellipses de variance obtenus pour l'ambiance, binaural référence.

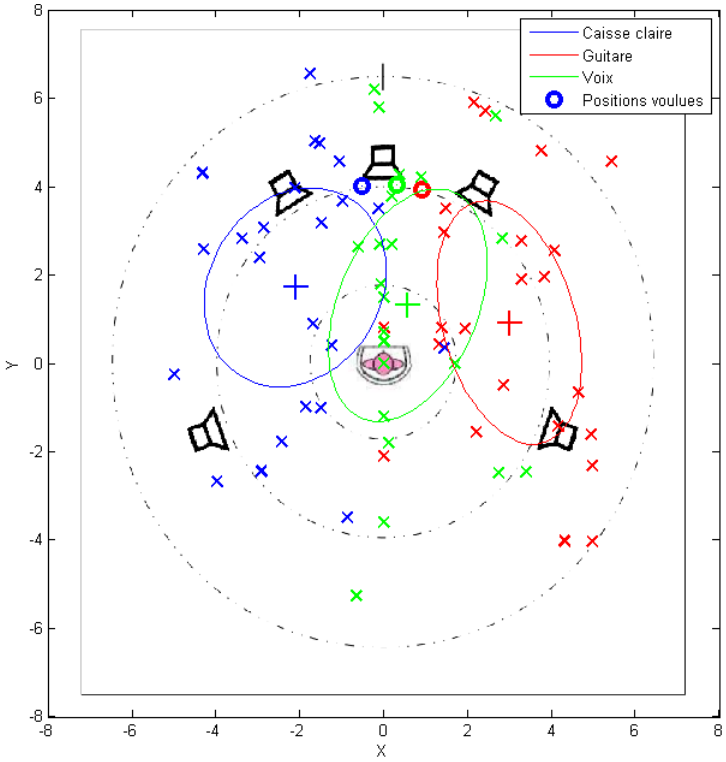


FIGURE B.7 – Centres de gravité et ellipses de variance, rock, binaural référence.

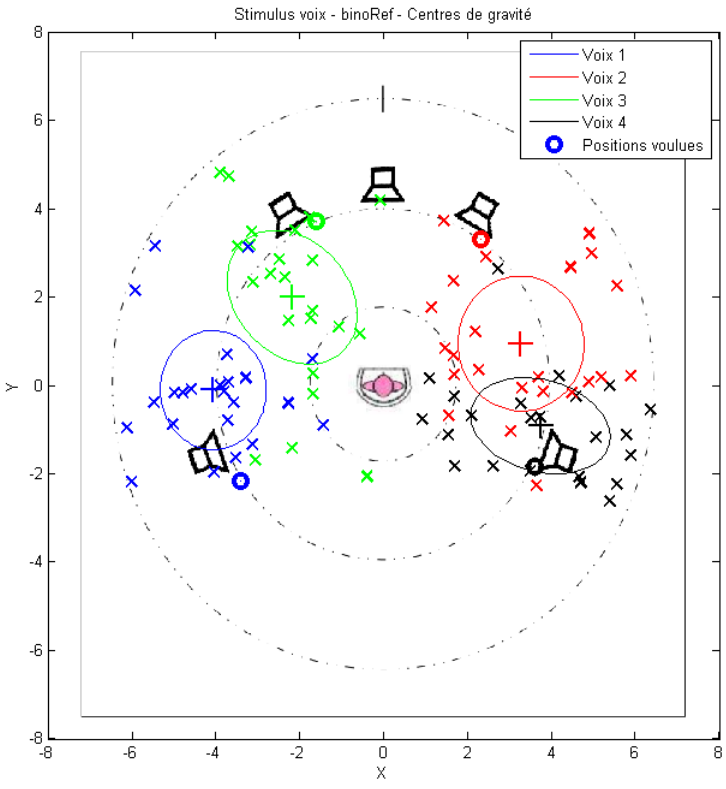


FIGURE B.8 – Centres de gravité et ellipses de variance, voix, binaural référence.

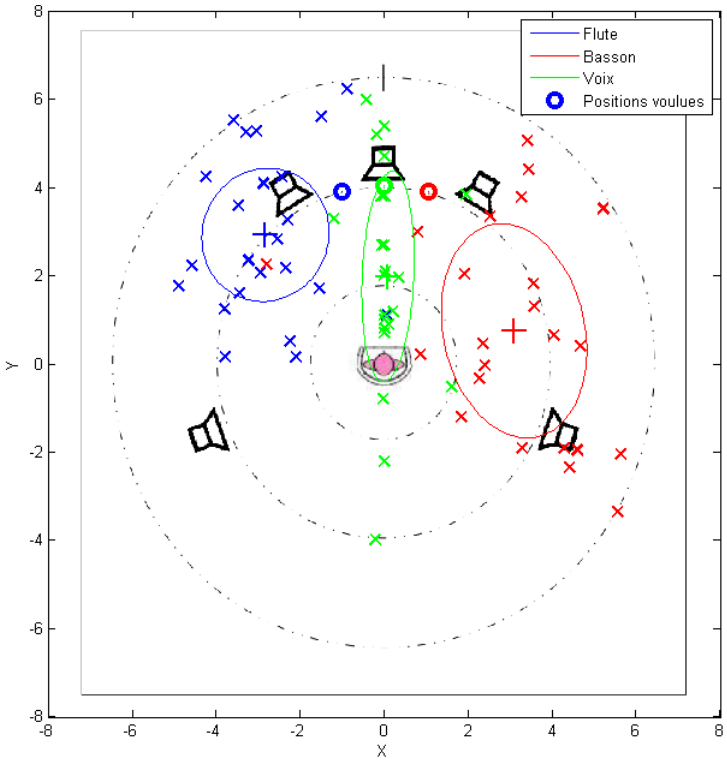


FIGURE B.9 – Centres de gravité et ellipses de variance, classique, binaural référence.

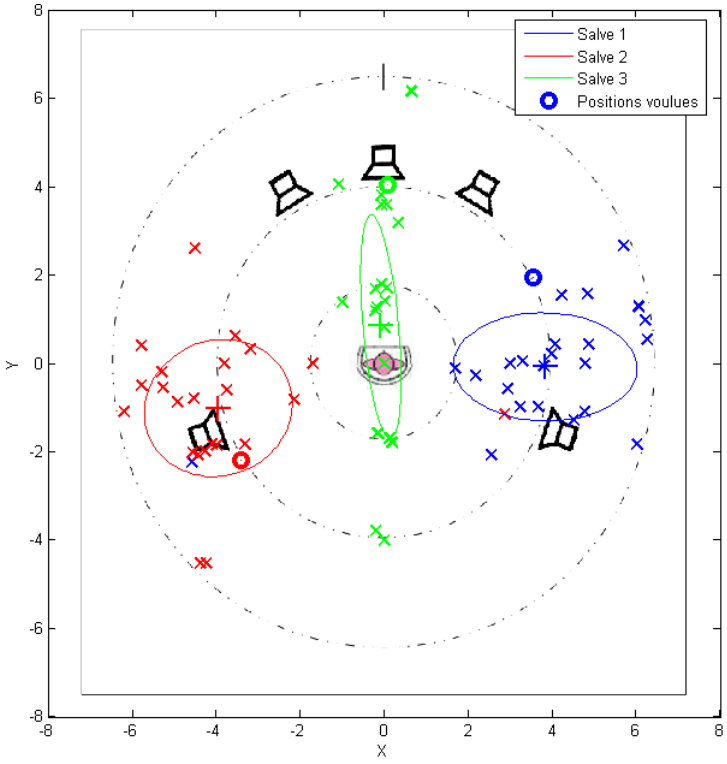


FIGURE B.10 – Centres de gravité et ellipses de variance, bruit rose, binaural référence.

B.1.3 En binaural référence cachée

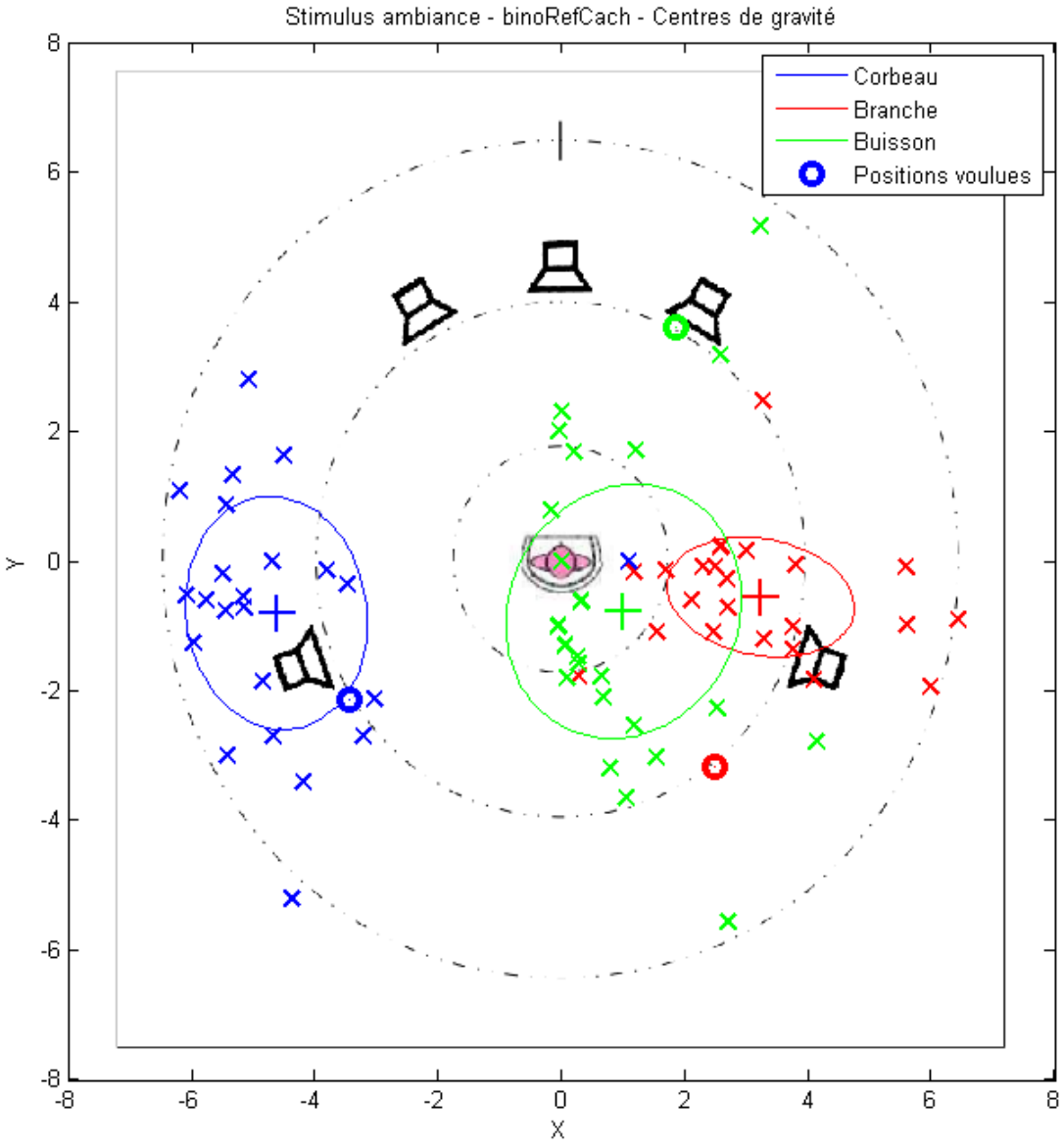


FIGURE B.11 – Centres de gravité et ellipses de variance obtenus pour l'ambiance, binaural référence cachée.

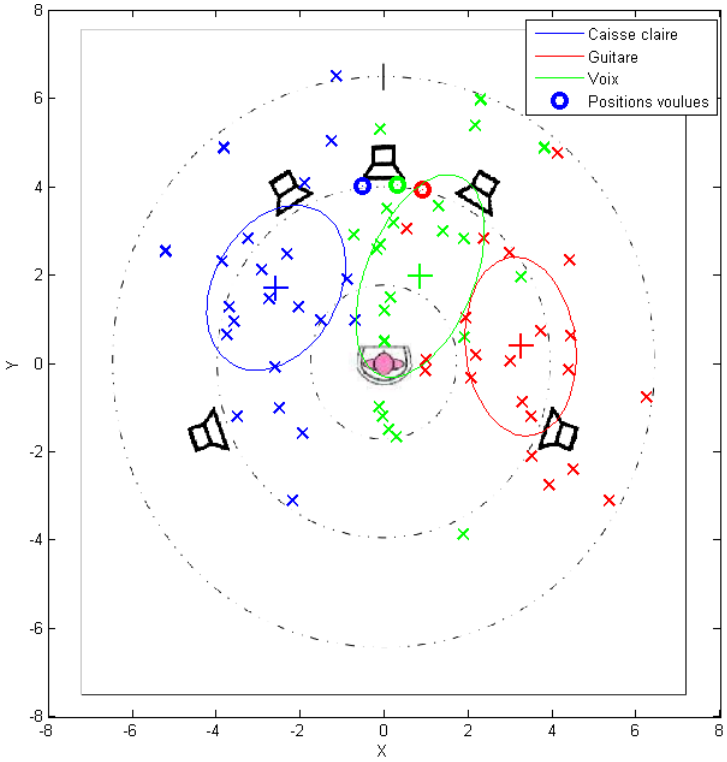


FIGURE B.12 – Centres de gravité et ellipses de variance, rock, binaural référence cachée.

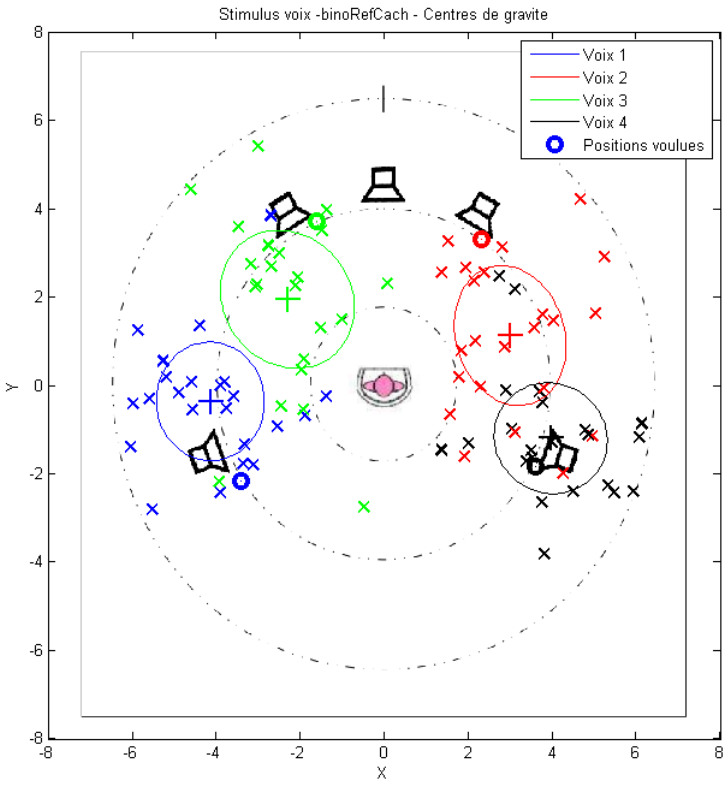


FIGURE B.13 – Centres de gravité et ellipses de variance, voix, binaural référence cachée.

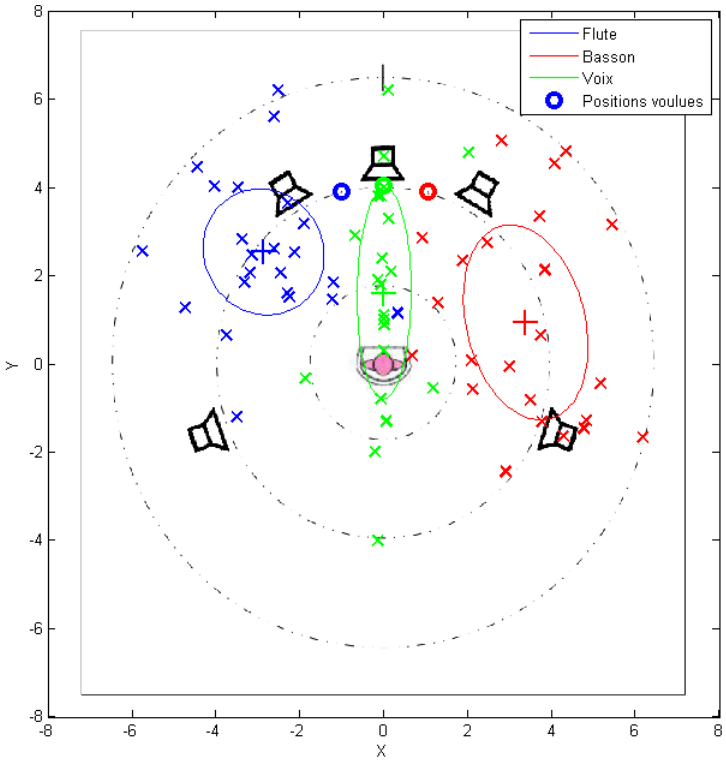


FIGURE B.14 – Centres de gravité et ellipses de variance, classique, binaural référence cachée.

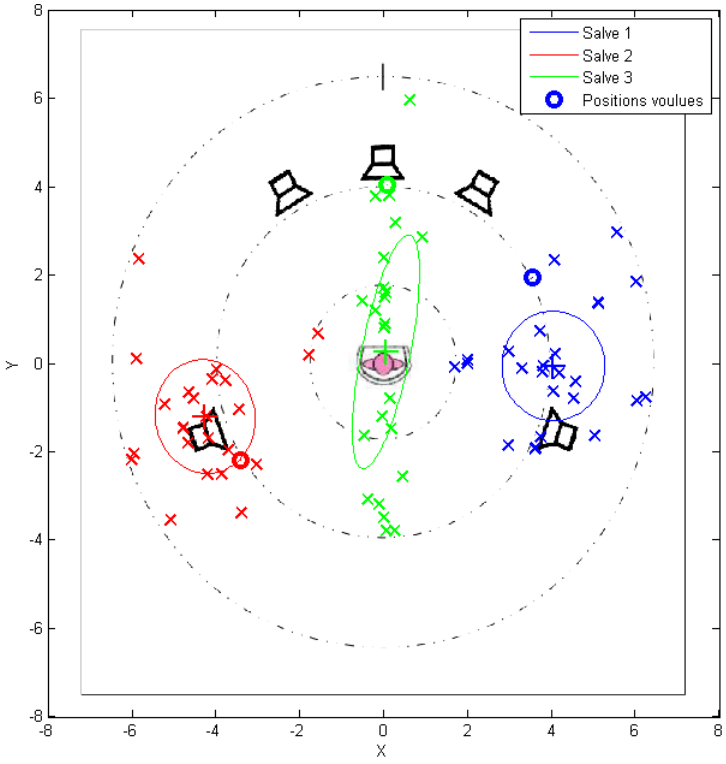


FIGURE B.15 – Centres de gravité et ellipses de variance, bruit rose, binaural référence cachée.

B.1.4 En binaural AAC

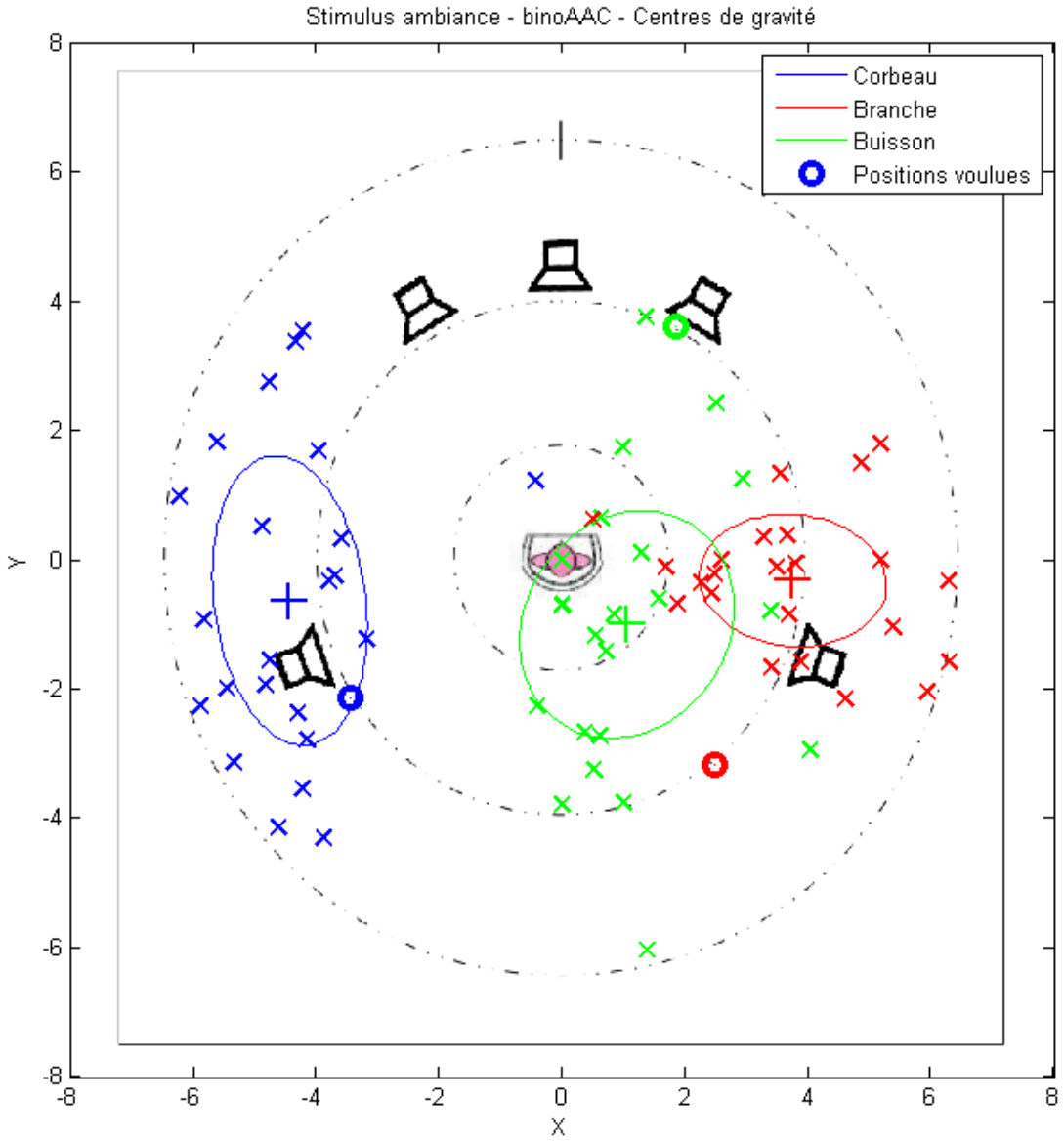


FIGURE B.16 – Centres de gravité et ellipses de variance obtenus pour l'ambiance, binaural AAC.

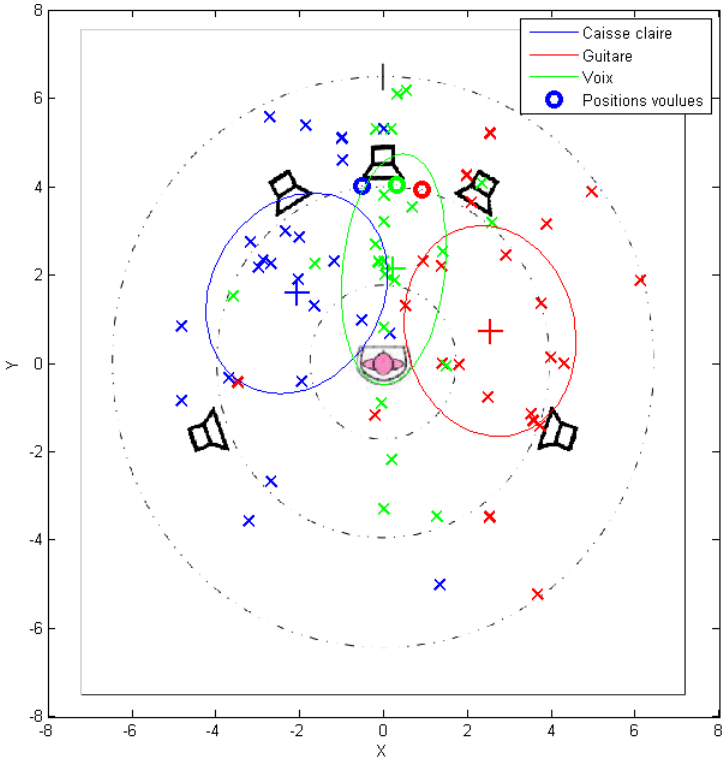


FIGURE B.17 – Centres de gravité et ellipses de variance, rock, binaural AAC.

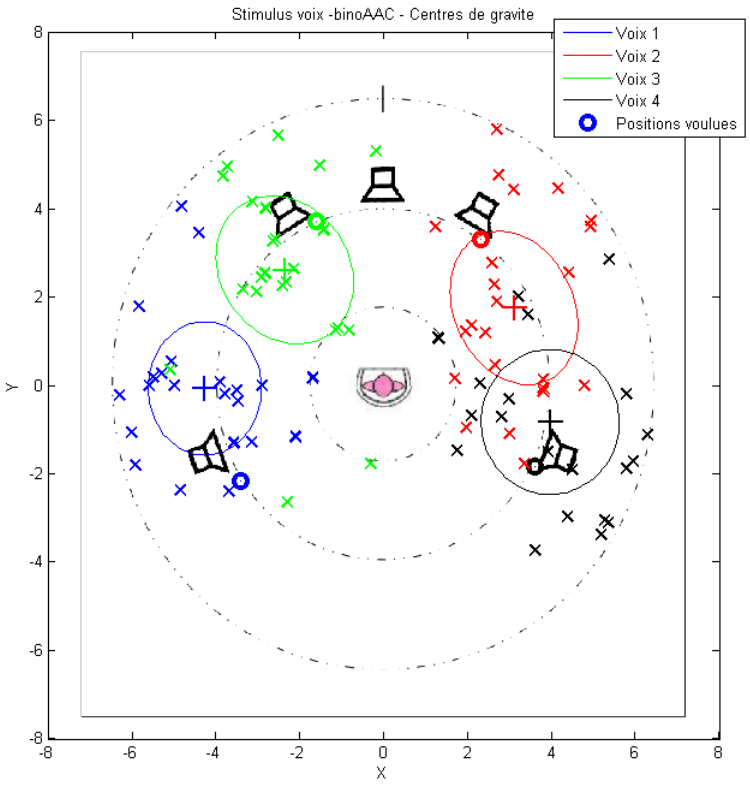


FIGURE B.18 – Centres de gravité et ellipses de variance, voix, binaural AAC.

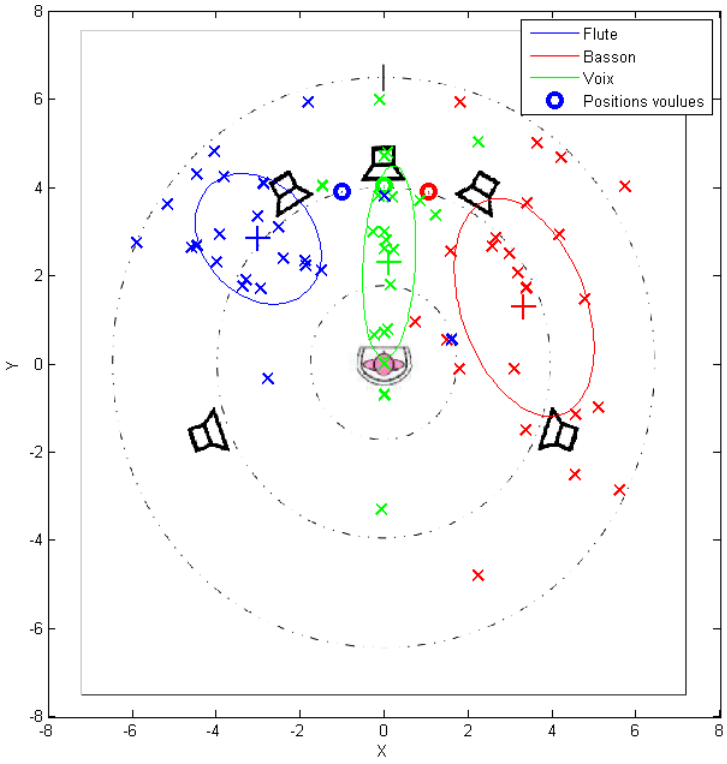


FIGURE B.19 – Centres de gravité et ellipses de variance, classique, binaural AAC.

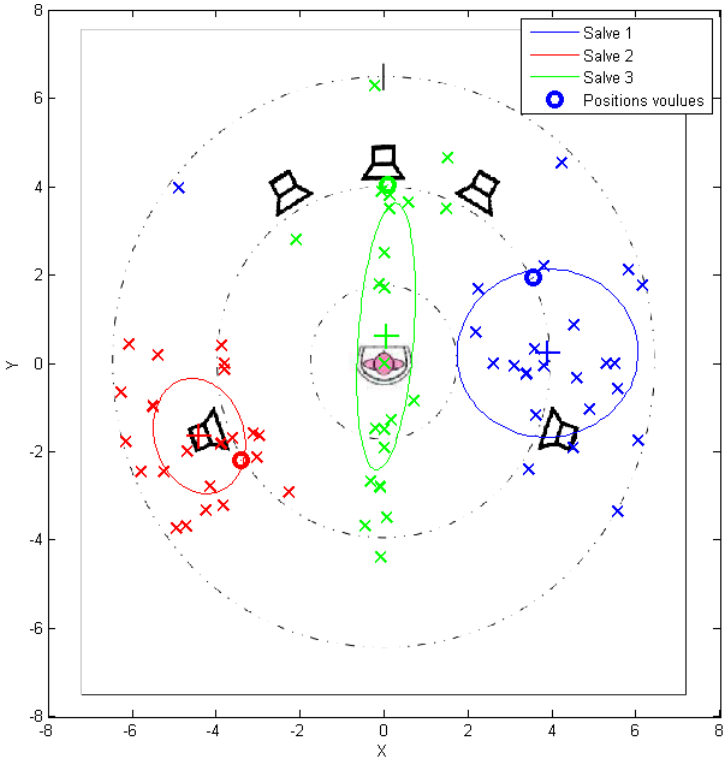


FIGURE B.20 – Centres de gravité et ellipses de variance, bruit rose, binaural AAC.

B.1.5 En binaural MP3

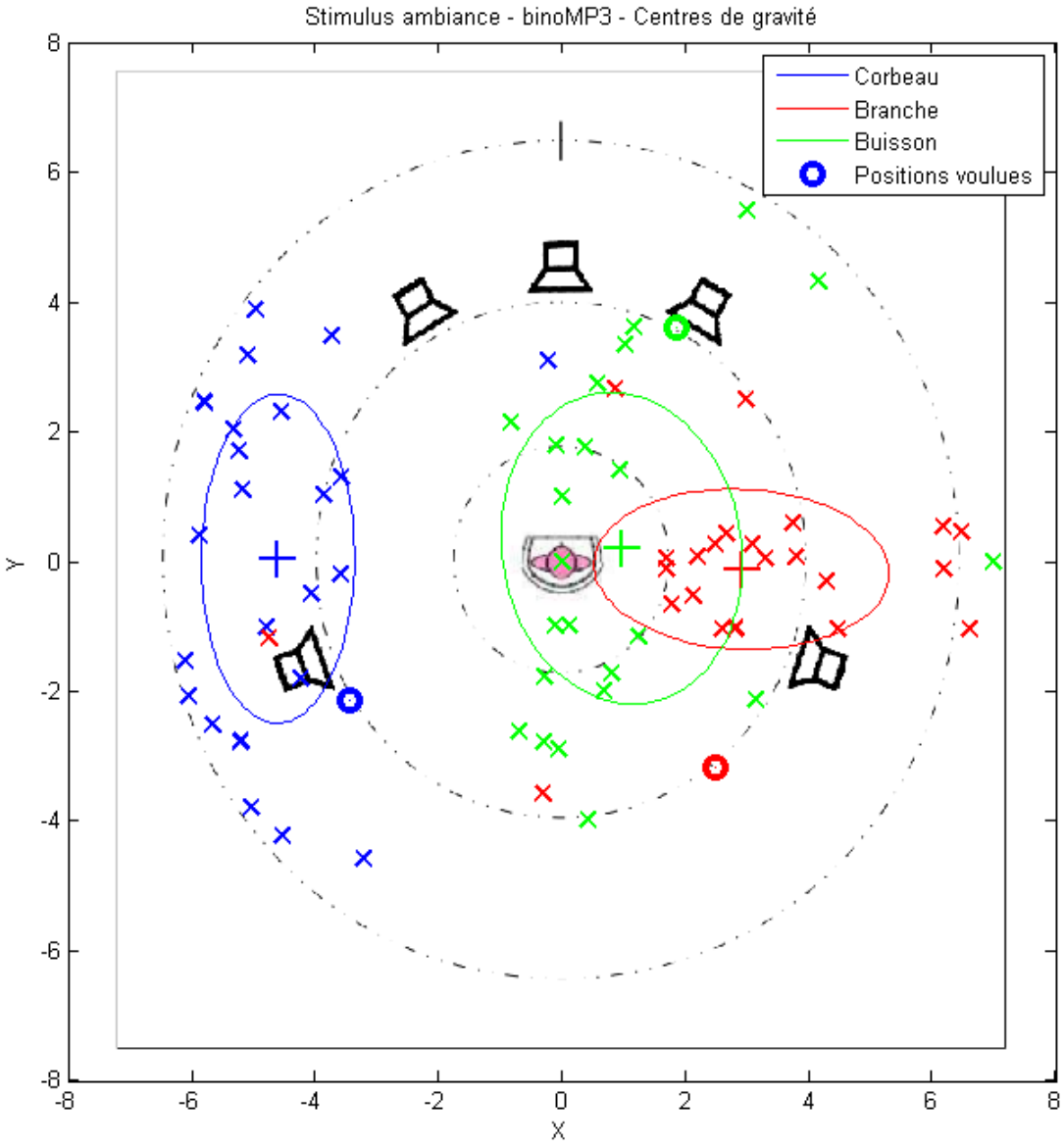


FIGURE B.21 – Centres de gravité et ellipses de variance obtenus pour l'ambiance, binaural MP3.

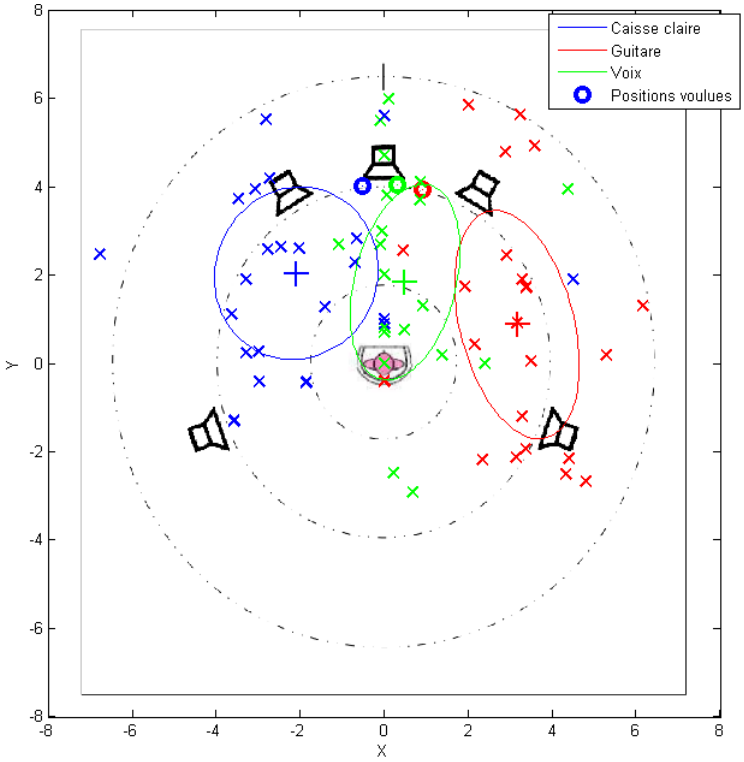


FIGURE B.22 – Centres de gravité et ellipses de variance, rock, binaural MP3.

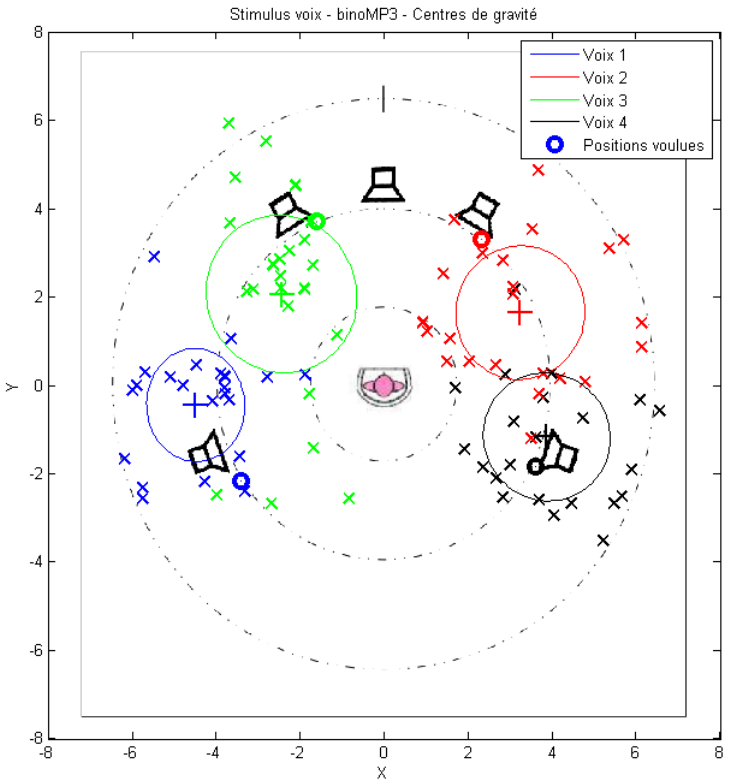


FIGURE B.23 – Centres de gravité et ellipses de variance, voix, binaural MP3.

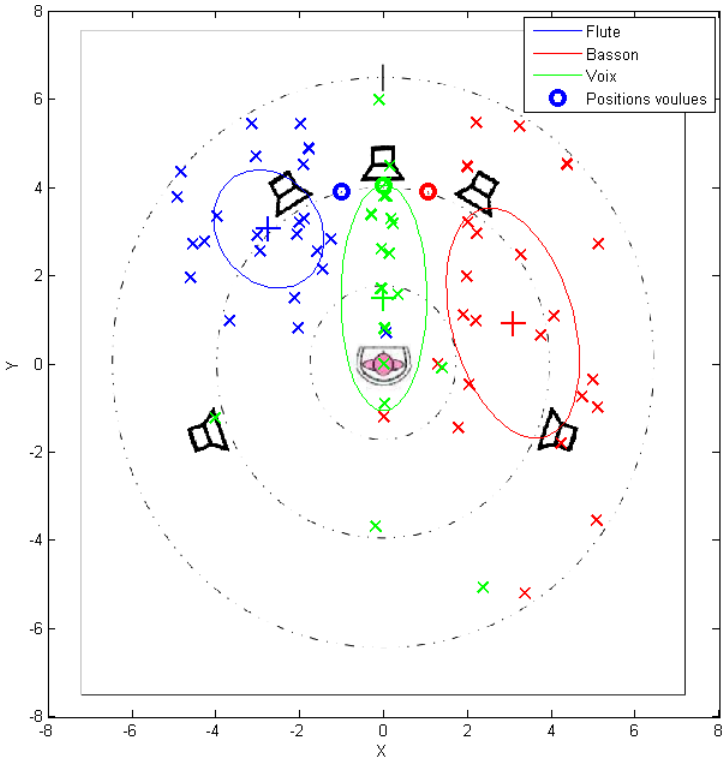


FIGURE B.24 – Centres de gravité et ellipses de variance, classique, binaural MP3.

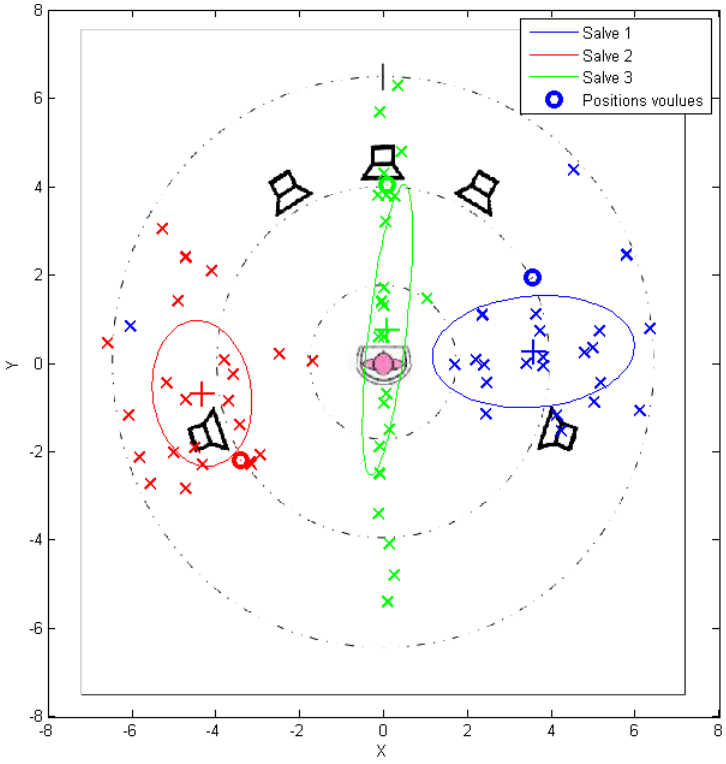


FIGURE B.25 – Centres de gravité et ellipses de variance, bruit rose, binaural MP3.

B.2 Ellipses et ellipses moyennes

B.2.1 Sur hauts-parleurs

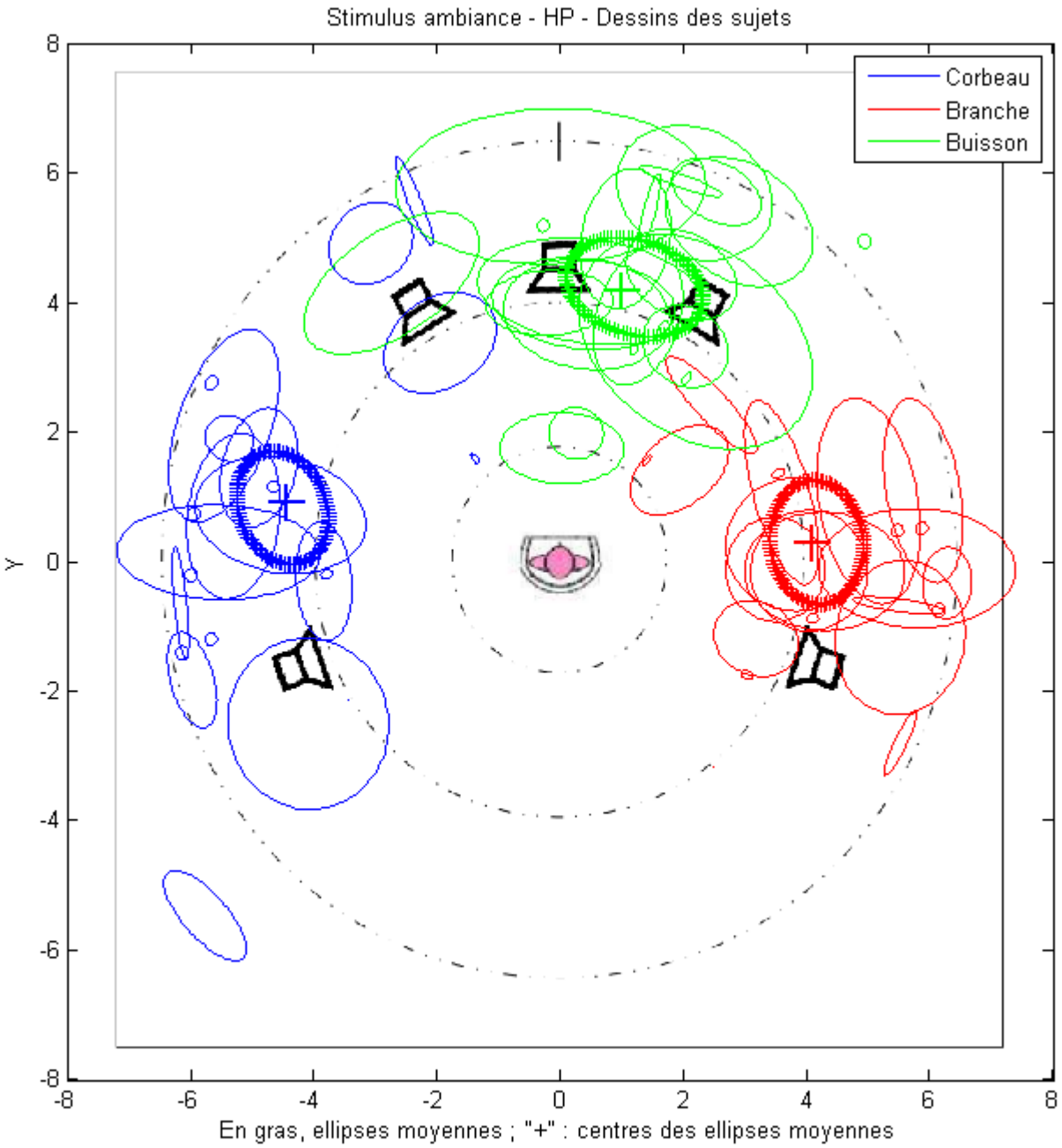


FIGURE B.26 – Stimulus ambiance, HP ; ellipses dessinées par les sujets.

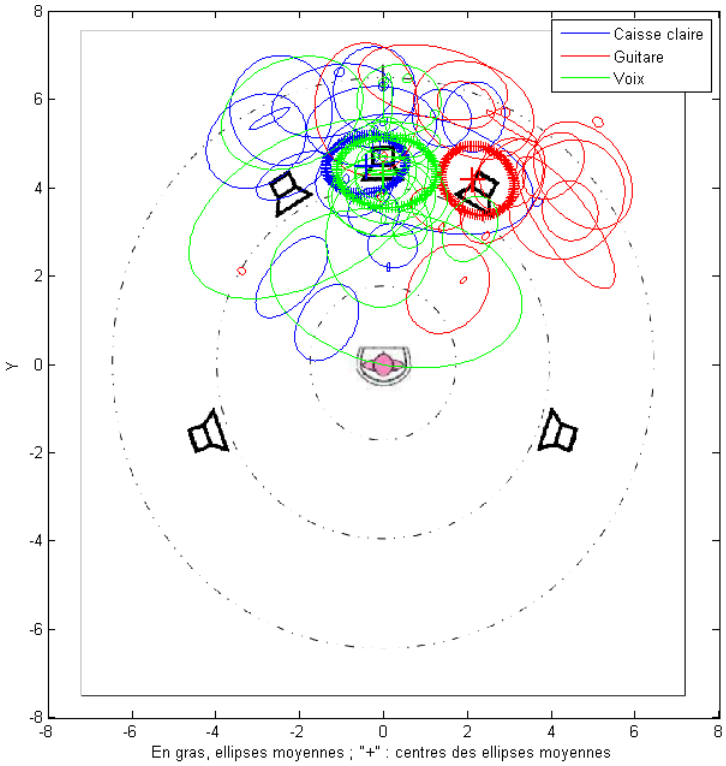


FIGURE B.27 – Stimulus rock, HP ; ellipses dessinées par les sujets.

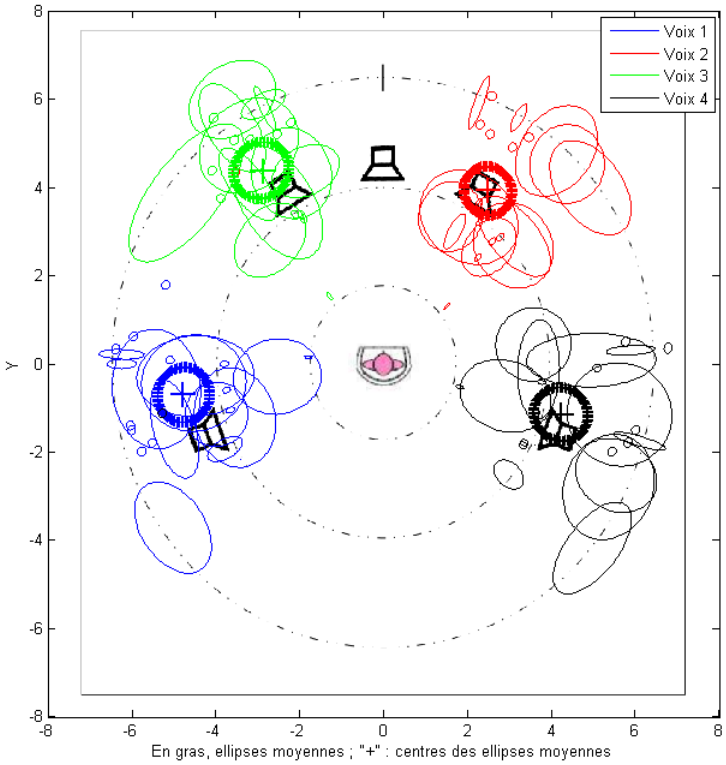


FIGURE B.28 – Stimulus voix, HP ; ellipses dessinées par les sujets.

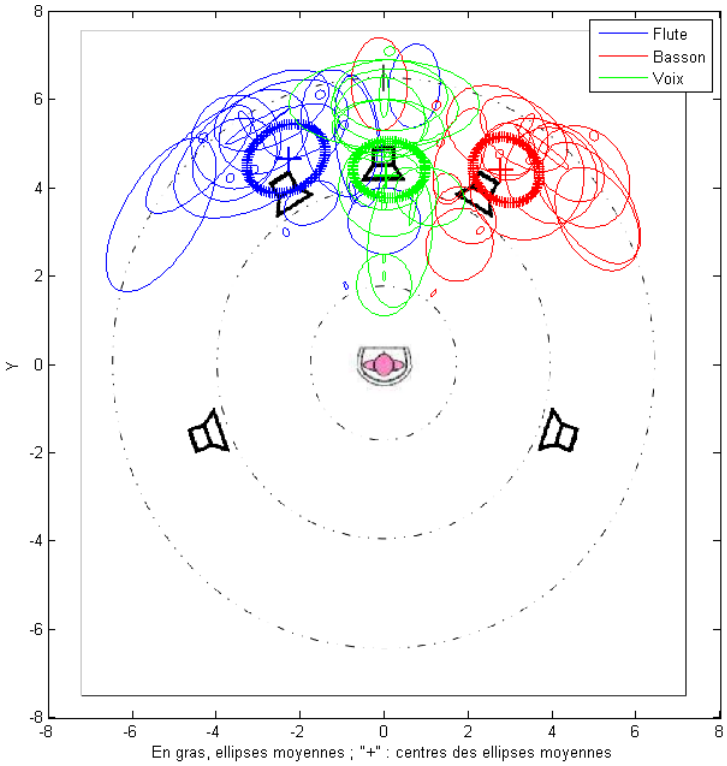


FIGURE B.29 – Stimulus classique, HP ; ellipses dessinées par les sujets.

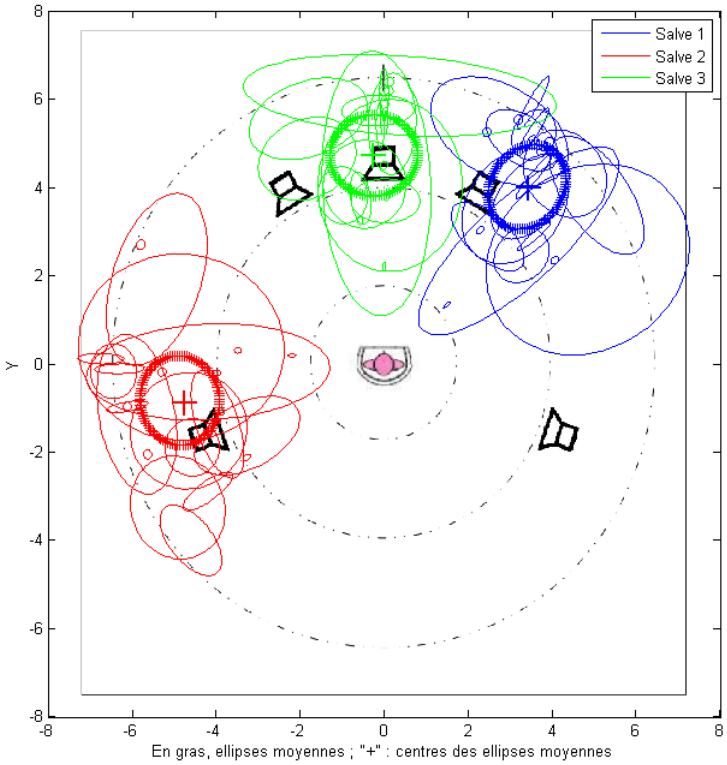


FIGURE B.30 – Stimulus bruit rose, HP ; ellipses dessinées par les sujets.

B.2.2 En binaural référence

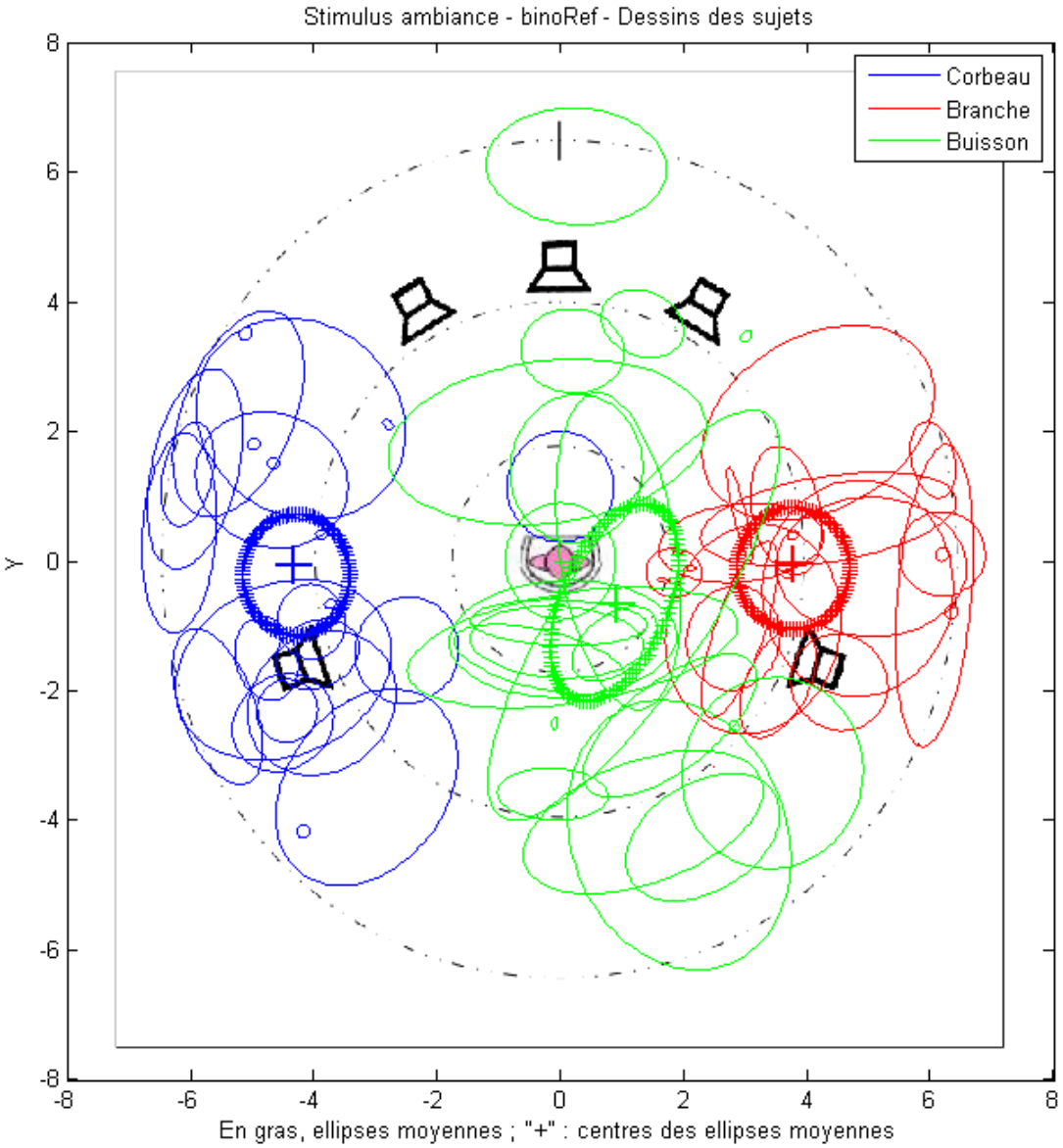


FIGURE B.31 – Stimulus ambiance, binaural référence ; ellipses dessinées par les sujets.

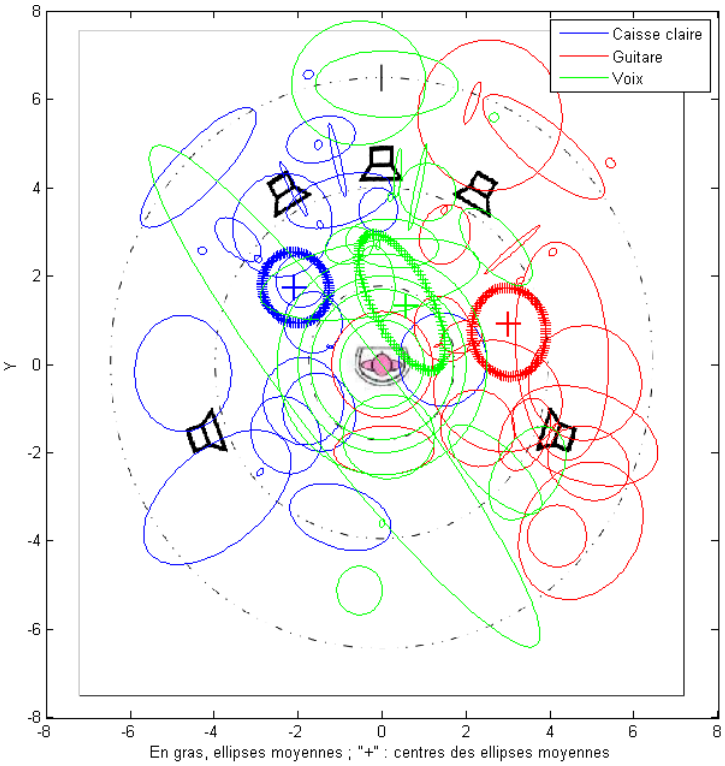


FIGURE B.32 – Stimulus rock, binaural référence ; ellipses dessinées par les sujets.

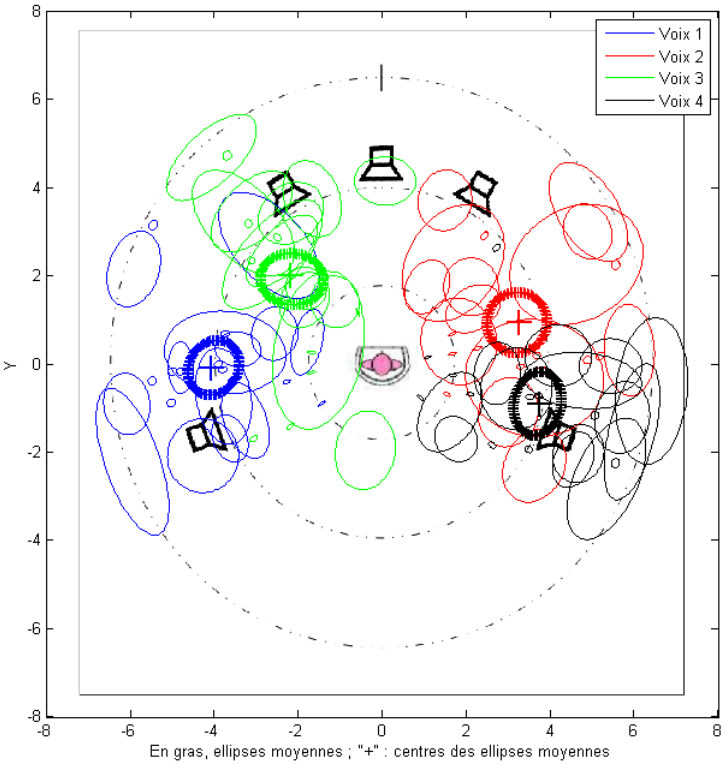


FIGURE B.33 – Stimulus voix, binaural référence ; ellipses dessinées par les sujets.

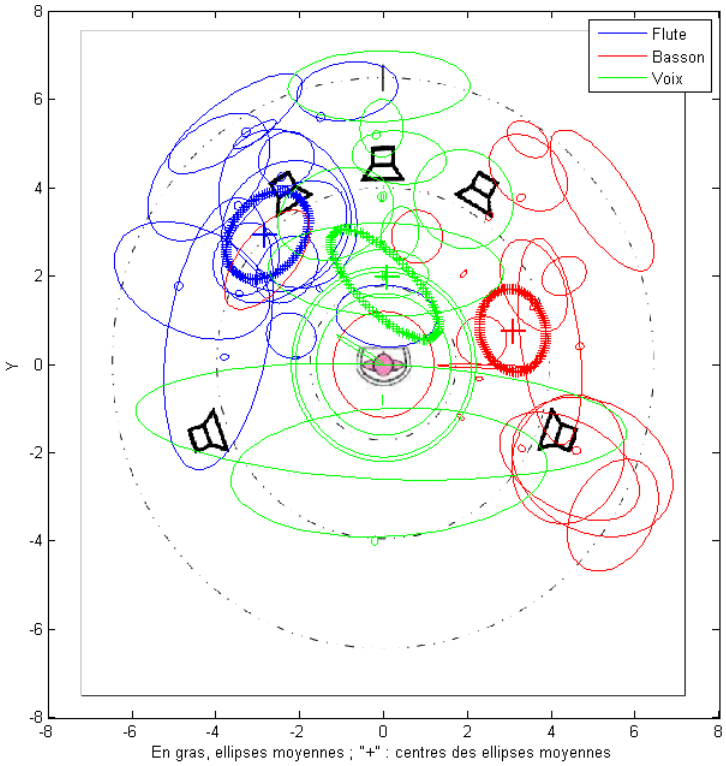


FIGURE B.34 – Stimulus classique, binaural référence ; ellipses dessinées par les sujets.

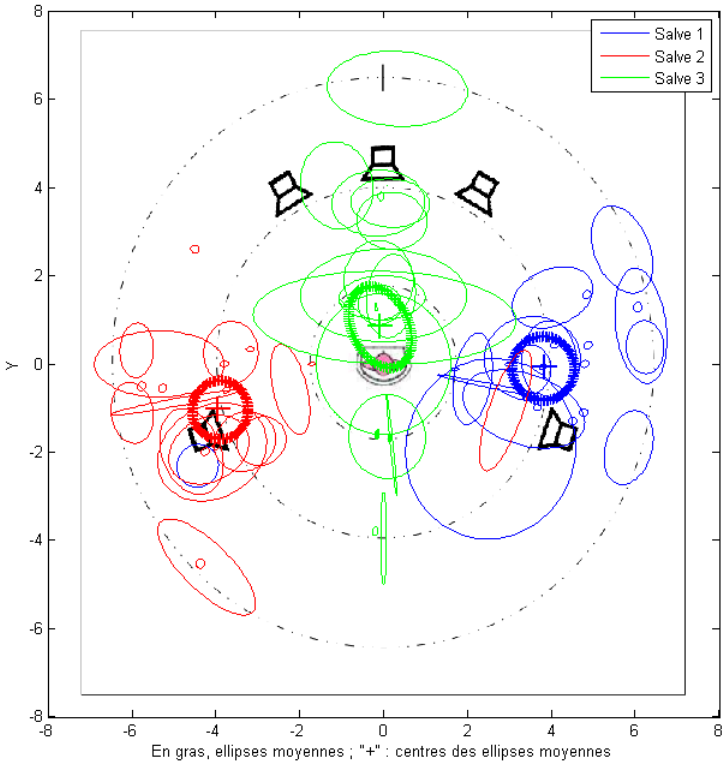


FIGURE B.35 – Stimulus bruit rose, binaural référence ; ellipses dessinées par les sujets.

B.2.3 En binaural référence cachée

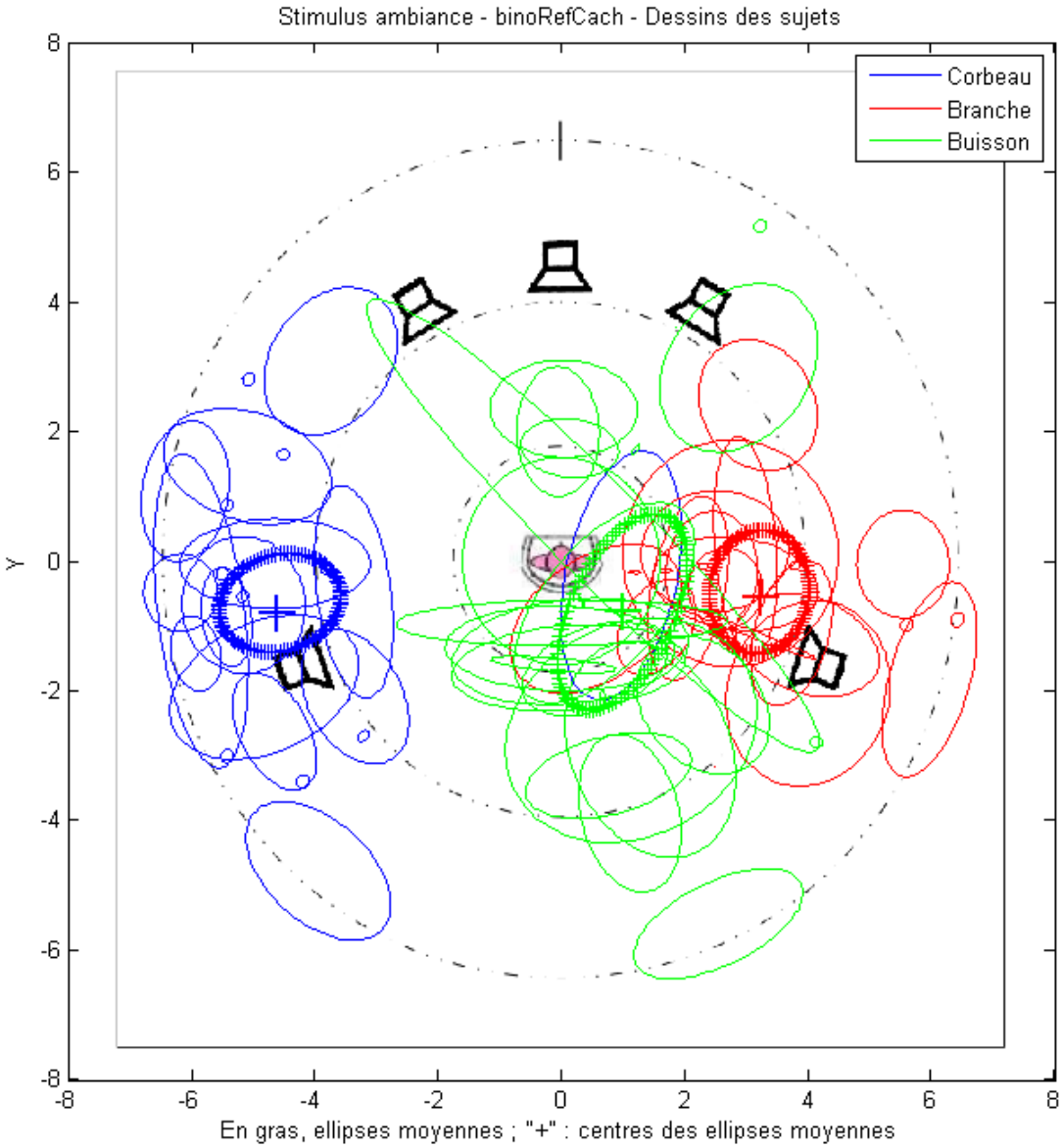


FIGURE B.36 – Stimulus ambiance, binaural référence cachée ; ellipses dessinées par les sujets.

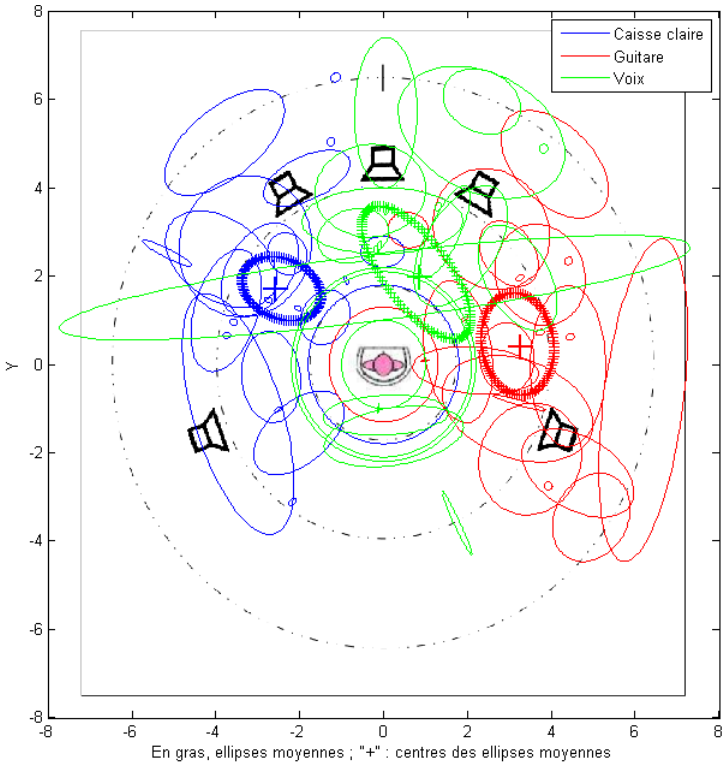


FIGURE B.37 – Stimulus rock, binaural référence cachée ; ellipses dessinées par les sujets.

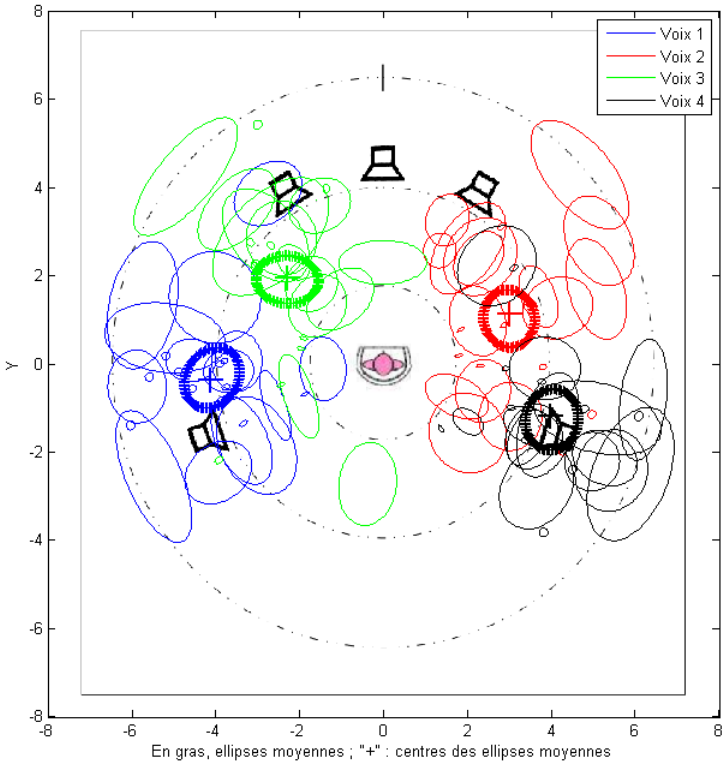


FIGURE B.38 – Stimulus voix, binaural référence cachée ; ellipses dessinées par les sujets.

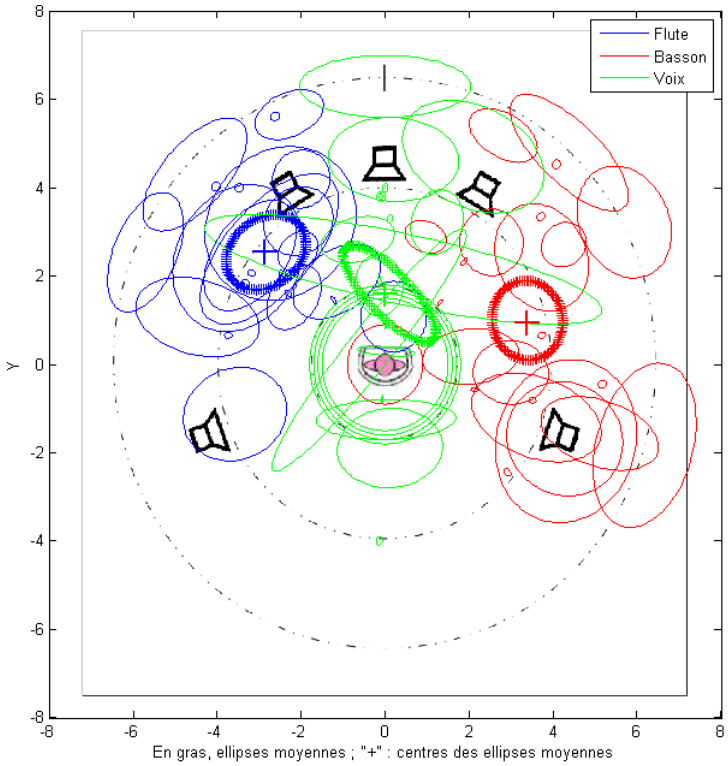


FIGURE B.39 – Stimulus classique, binaural référence cachée ; ellipses dessinées par les sujets.

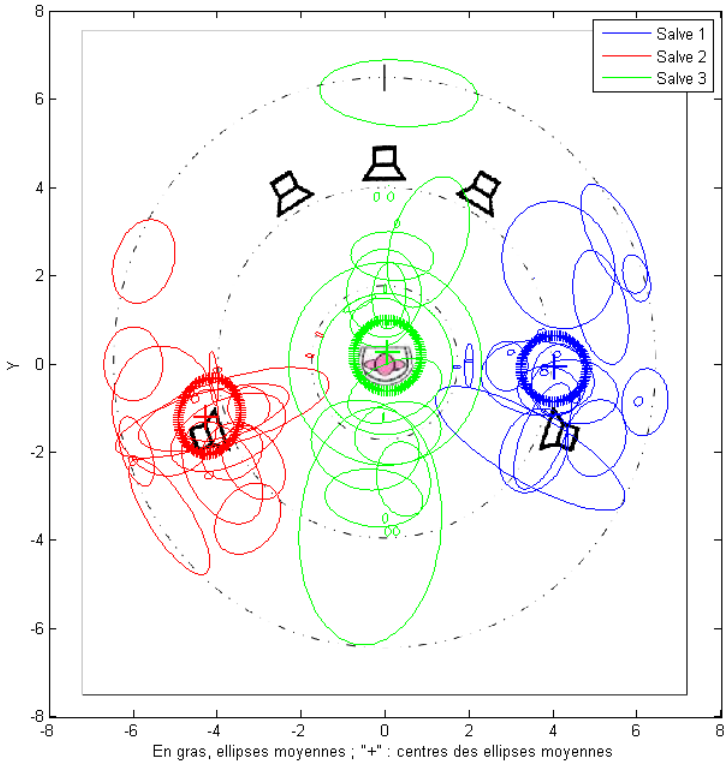


FIGURE B.40 – Stimulus bruit rose, binaural référence cachée ; ellipses dessinées par les sujets.

B.2.4 En binaural AAC

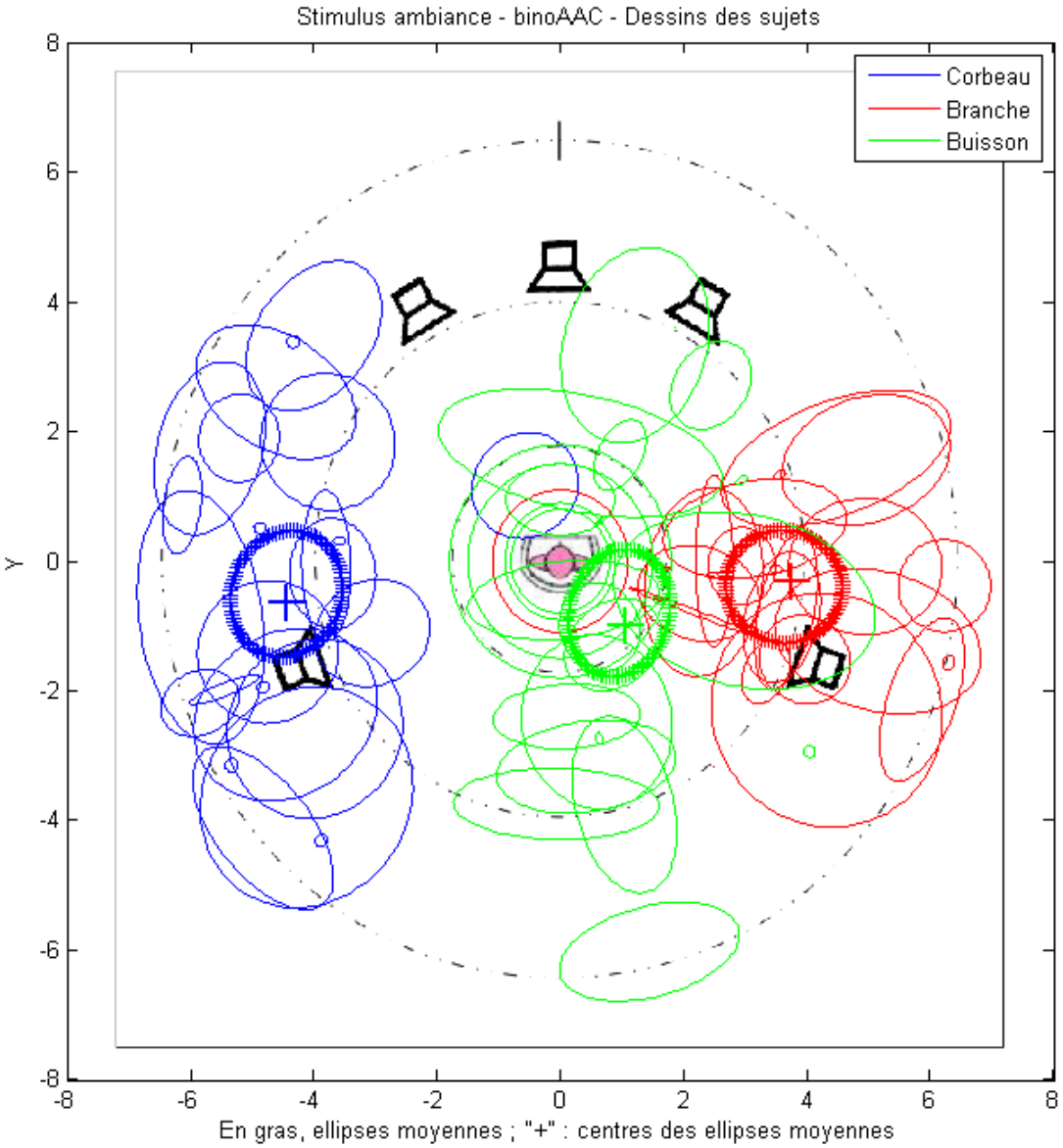


FIGURE B.41 – Stimulus ambiance, binaural AAC ; ellipses dessinées par les sujets.

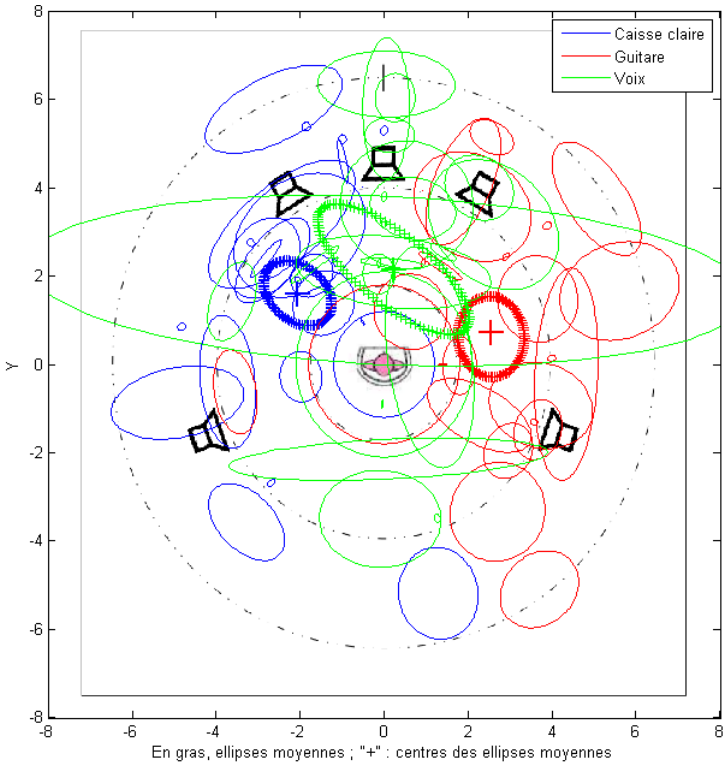


FIGURE B.42 – Stimulus rock, binaural AAC ; ellipses dessinées par les sujets.

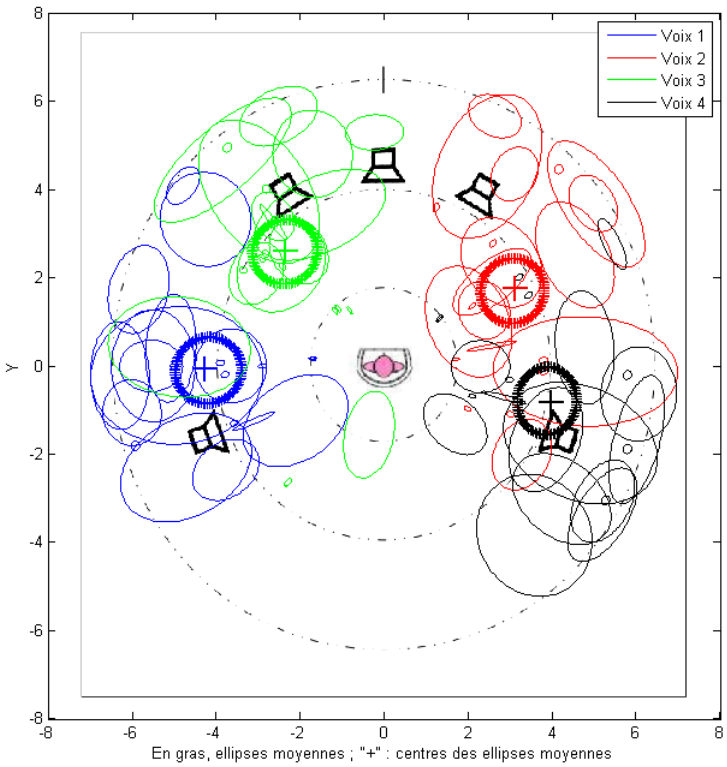


FIGURE B.43 – Stimulus voix, binaural AAC ; ellipses dessinées par les sujets.

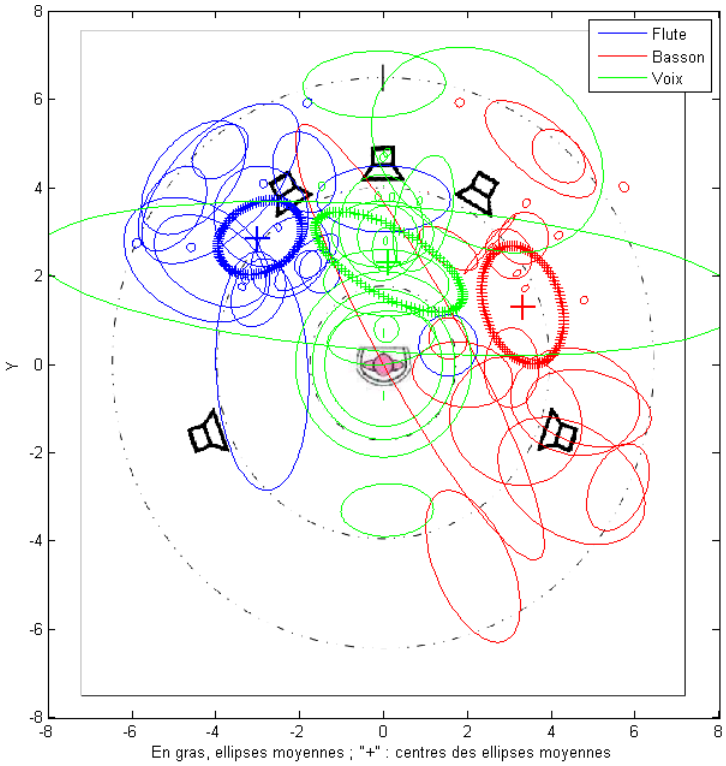


FIGURE B.44 – Stimulus classique, binaural AAC ; ellipses dessinées par les sujets.

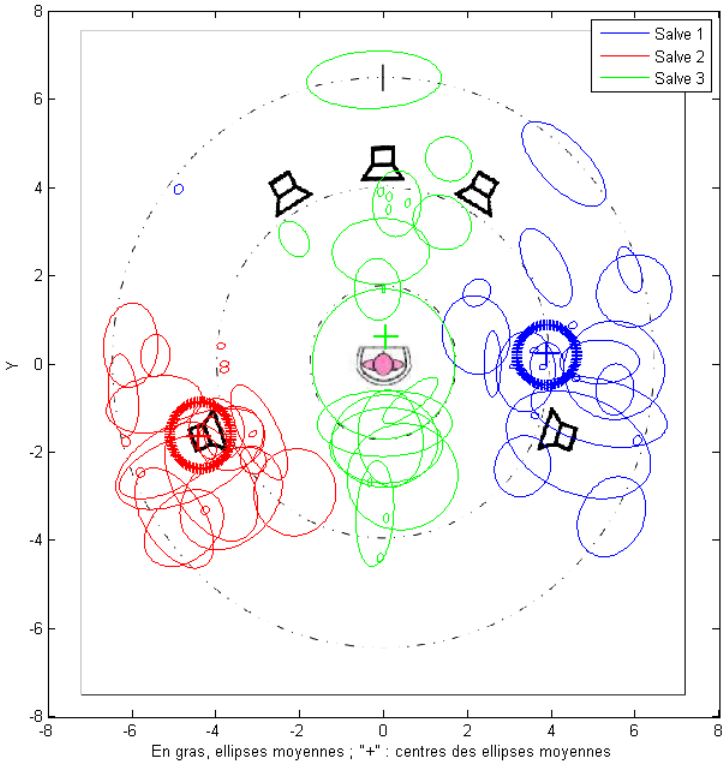


FIGURE B.45 – Stimulus bruit rose, binaural AAC ; ellipses dessinées par les sujets.

B.2.5 En binaural MP3

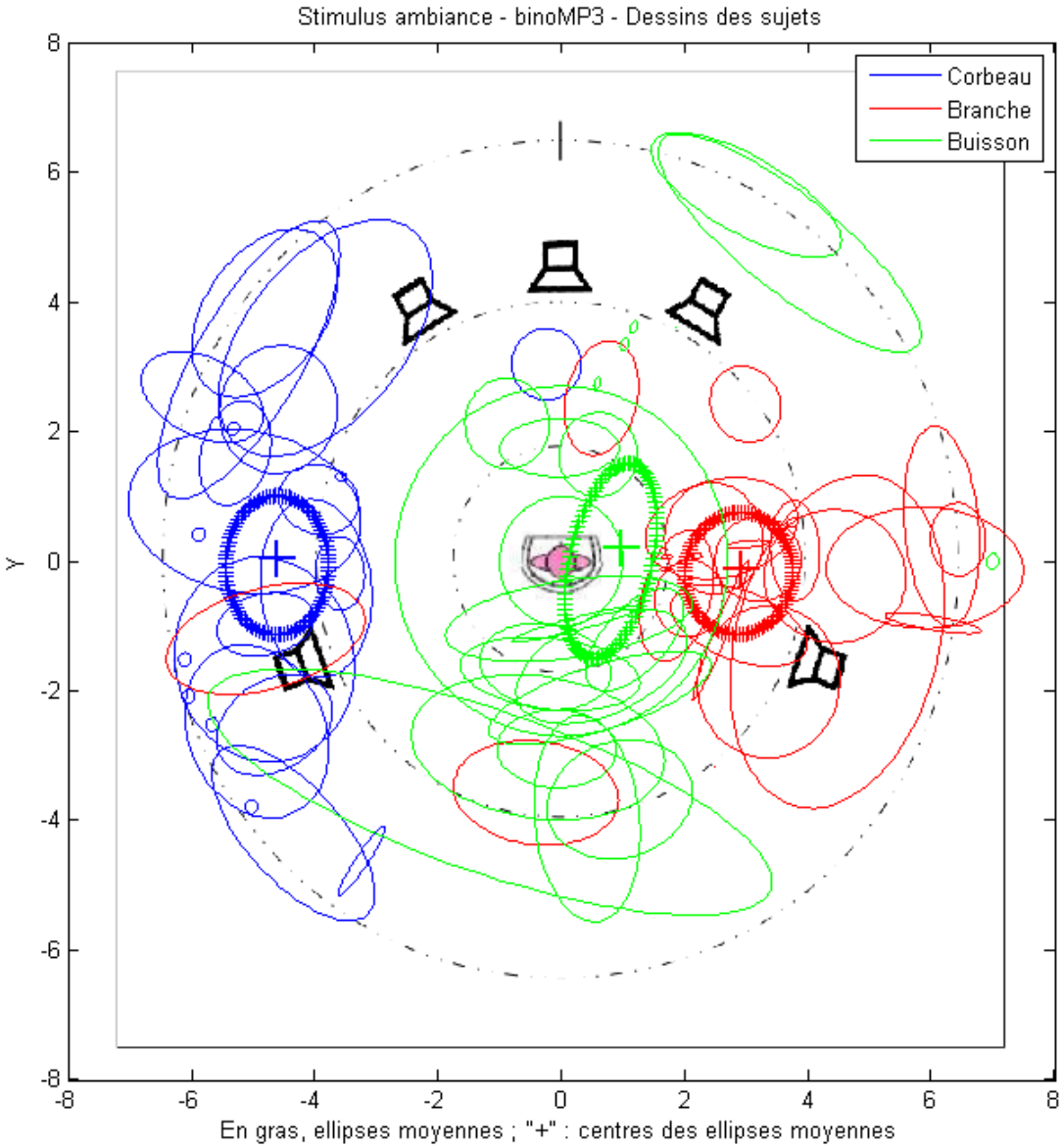


FIGURE B.46 – Stimulus ambiance, binaural MP3 ; ellipses dessinées par les sujets.

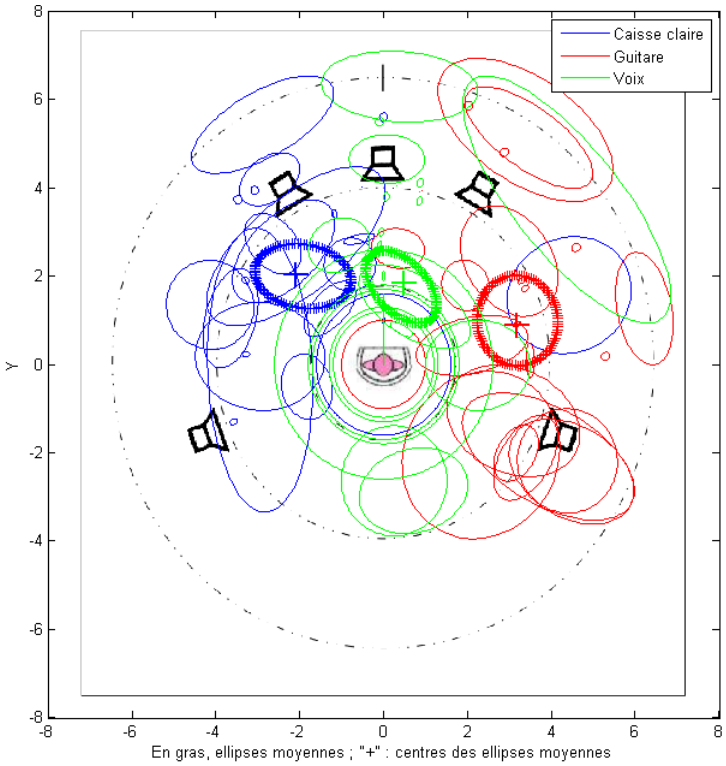


FIGURE B.47 – Stimulus rock, binaural MP3 ; ellipses dessinées par les sujets.

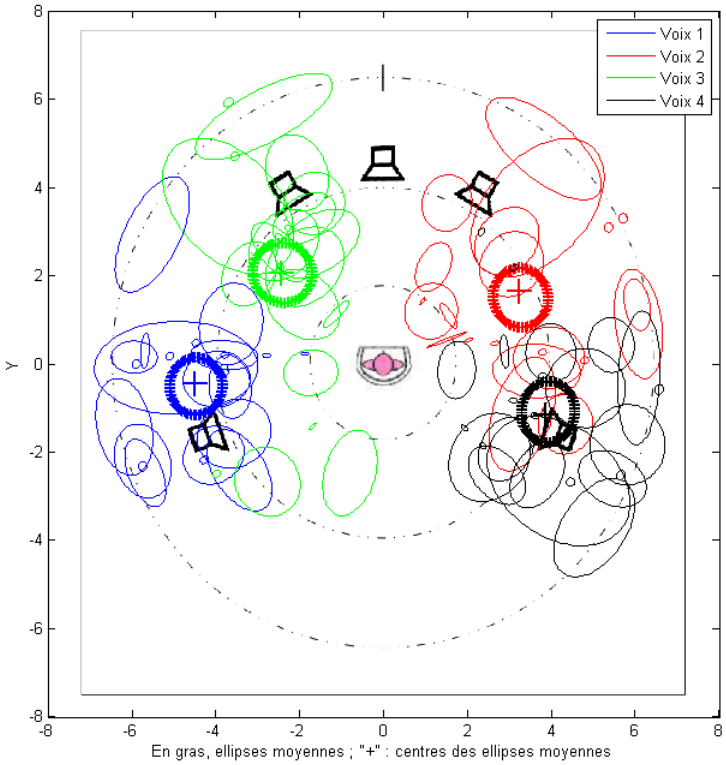


FIGURE B.48 – Stimulus voix, binaural MP3 ; ellipses dessinées par les sujets.

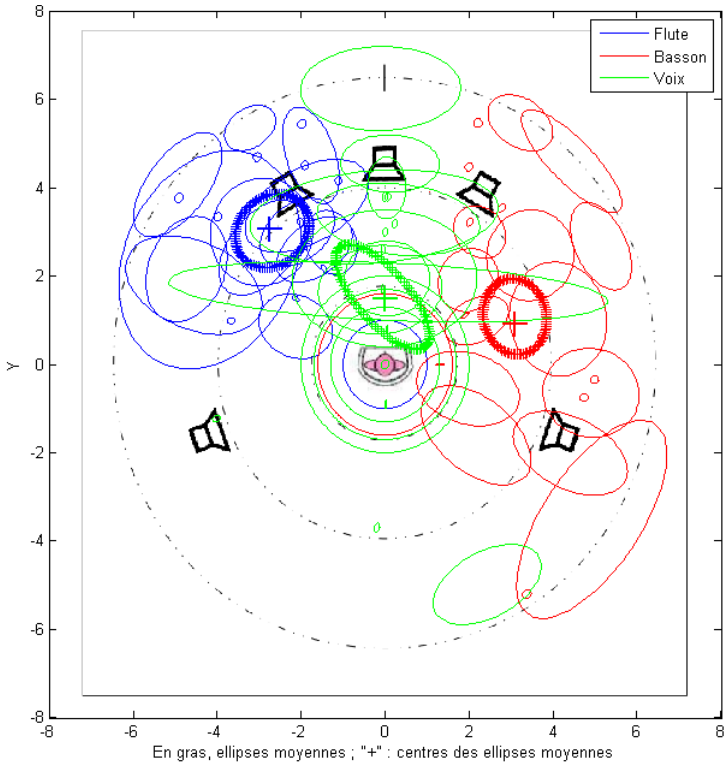


FIGURE B.49 – Stimulus classique, binaural MP3 ; ellipses dessinées par les sujets.

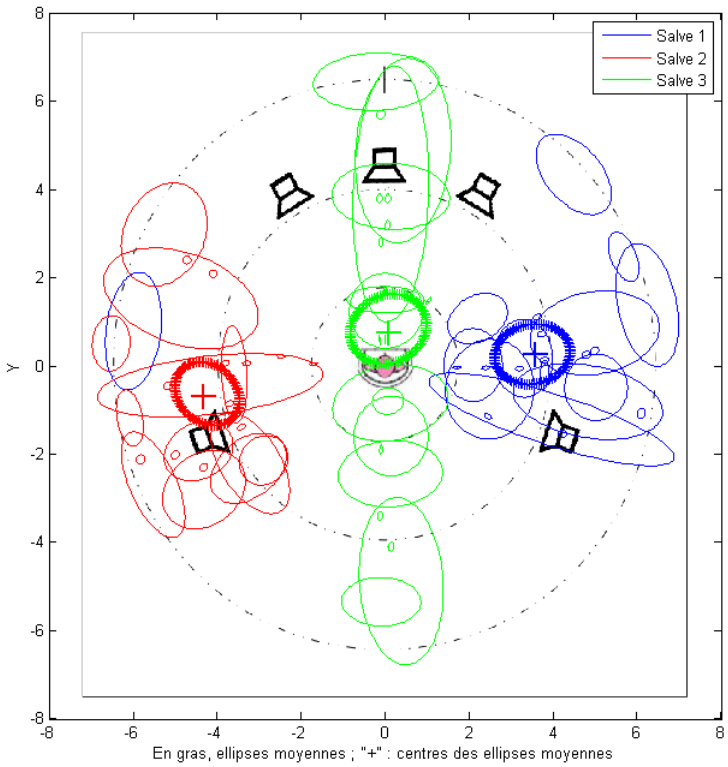
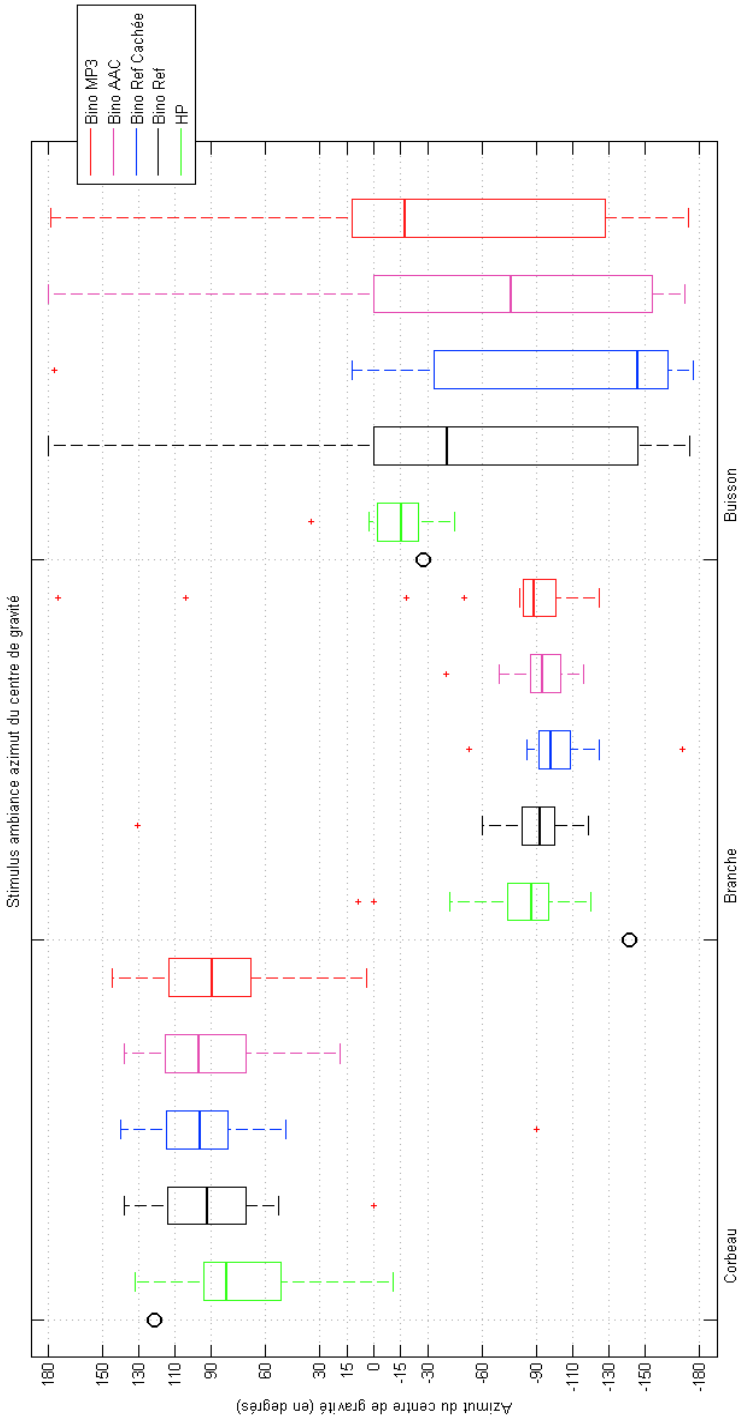
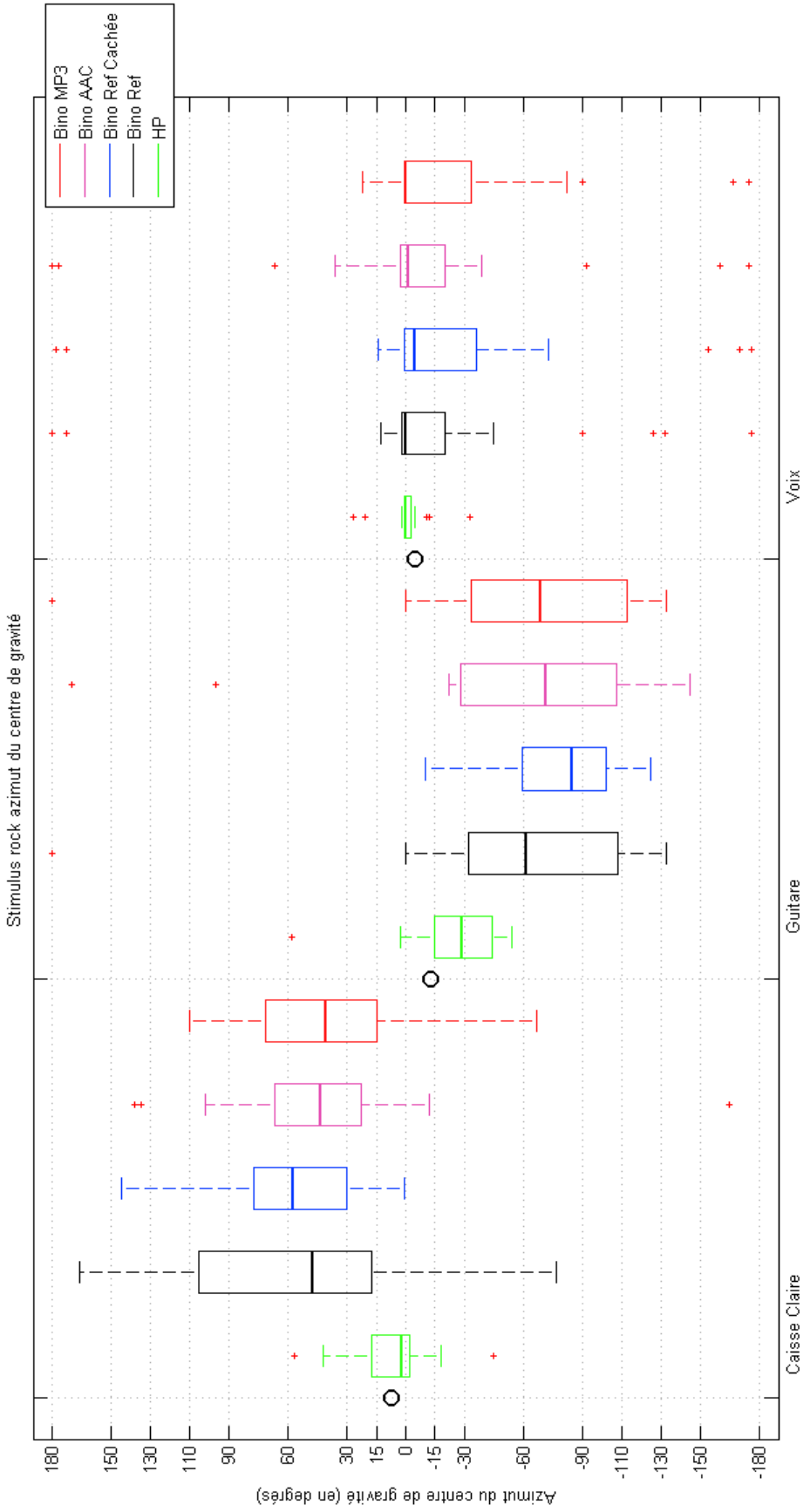
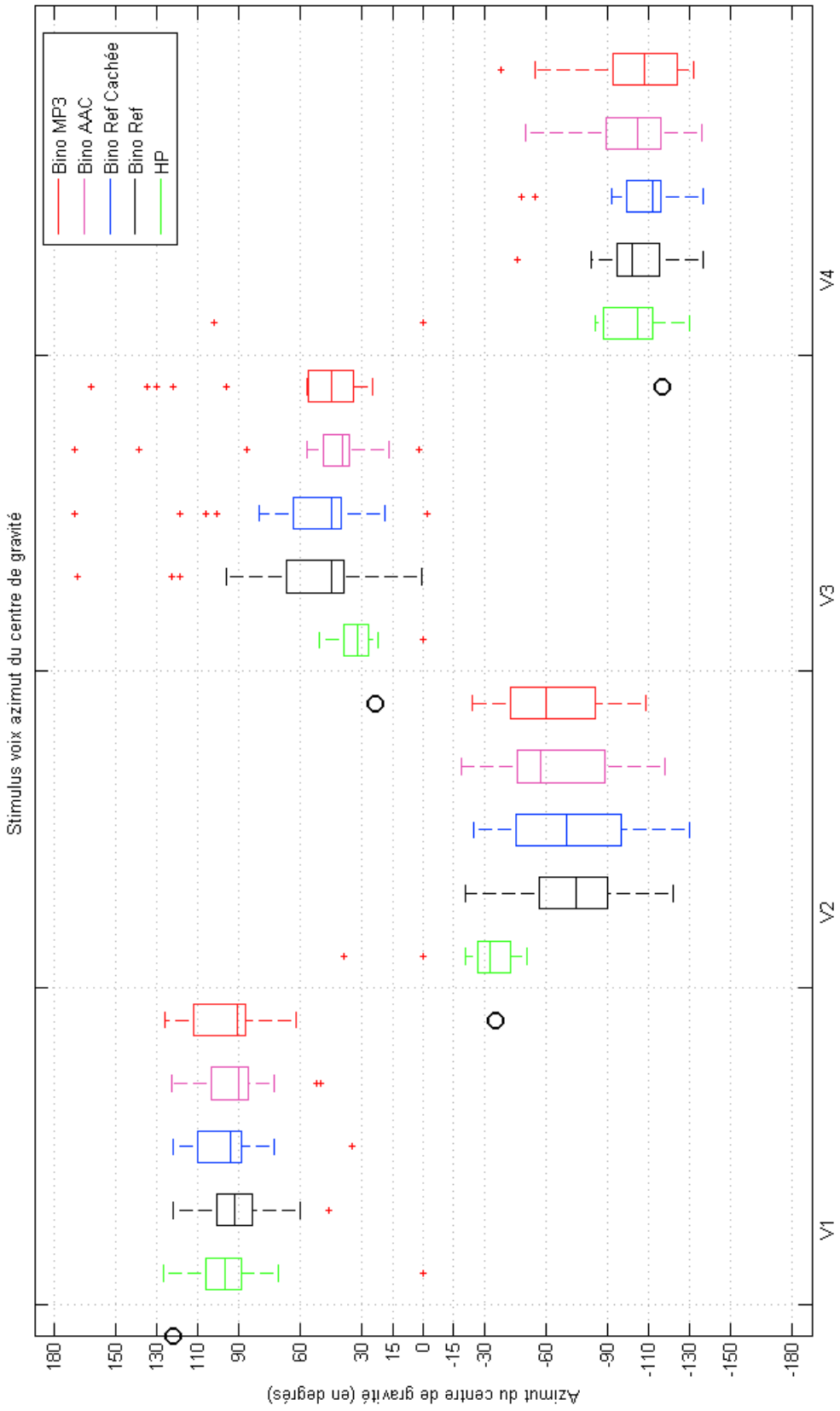


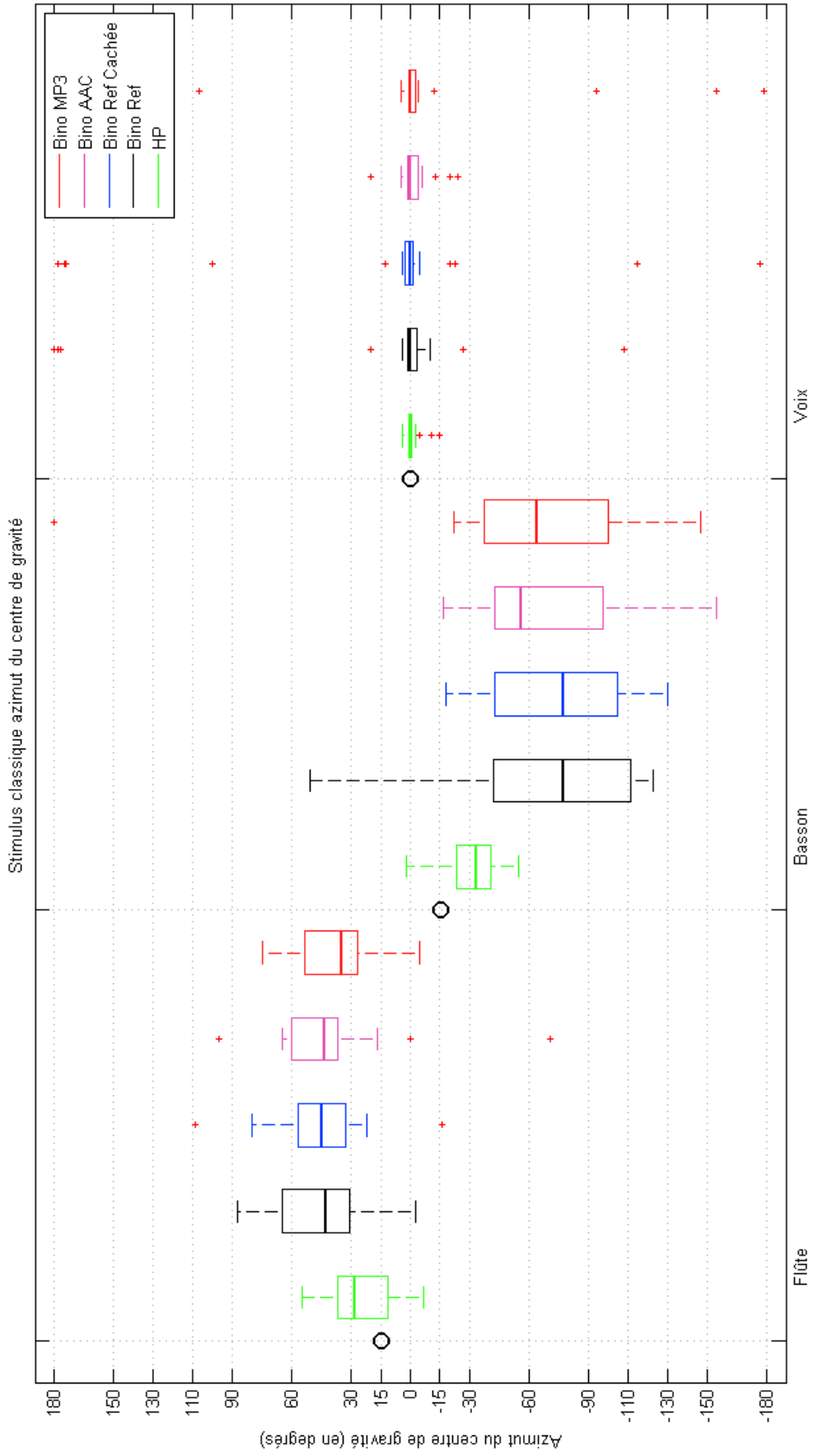
FIGURE B.50 – Stimulus bruit rose, binaural MP3 ; ellipses dessinées par les sujets.

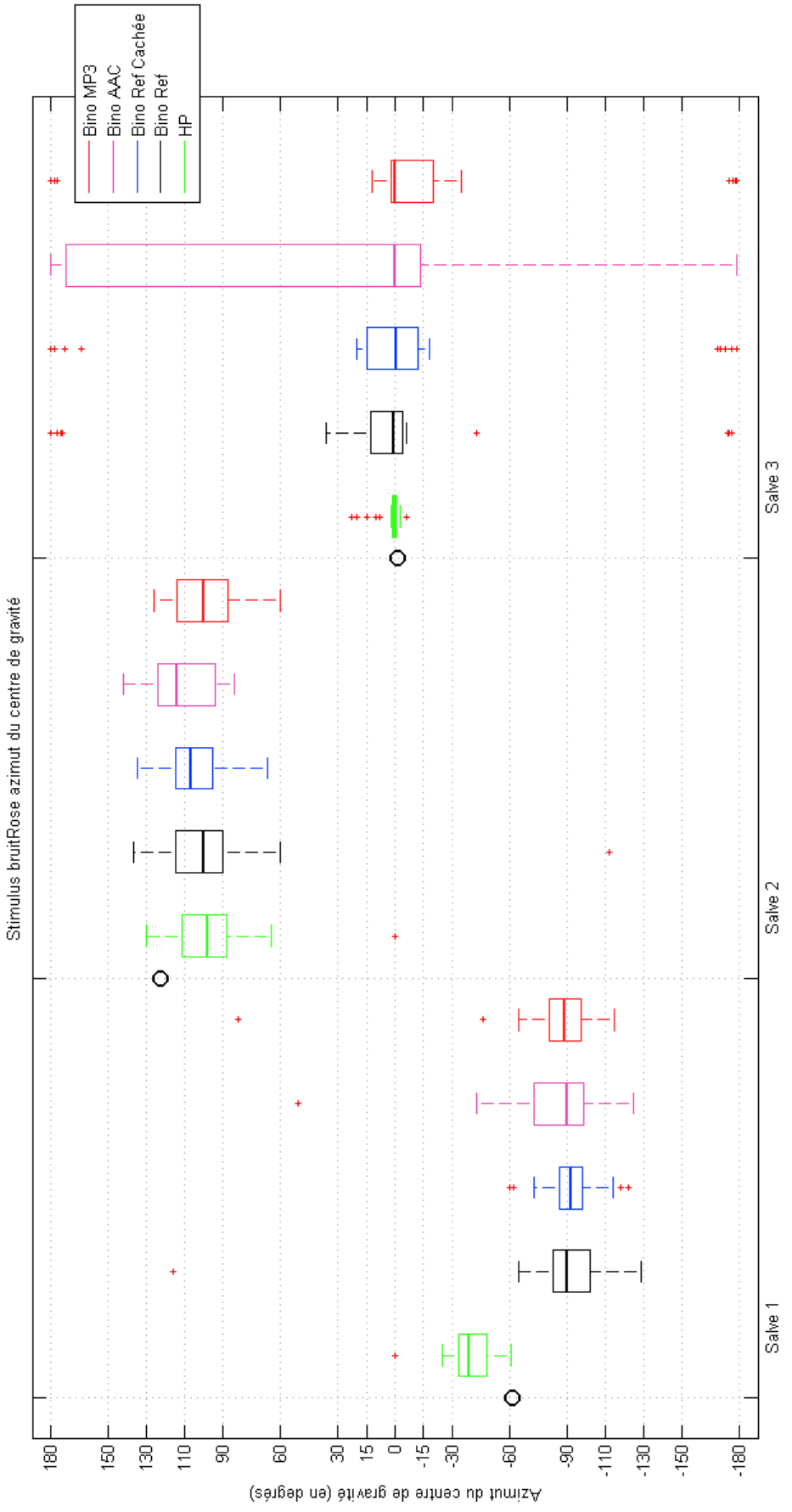
B.3 Box plots des azimuts




Bino MP3
Bino AAC
Bino Ref Cachée
Bino Ref
HP
Stimulus rock azimut du centre de gravité
Voix
Guitare
Caisse Claire
Azimut du centre de gravité (en degrés)
180
150
130
110
90
60
30
15
0
-15
-30
-60
-90
-110
-130
-150
-180


Bino MP3
Bino AAC
Bino Ref Cachée
Bino Ref
HP
Stimulus voix azimut du centre de gravité
Azimut du centre de gravité (en degrés)
V1
V2
V3
V4
180
150
130
110
90
60
30
15
0
-15
-30
-60
-90
-110
-130
-150
-180


Bino MP3
Bino AAC
Bino Ref Cachée
Bino Ref
HP
Stimulus classique azimut du centre de gravité
Voix
Basson
Flûte
Azimut du centre de gravité (en degrés)
180
150
130
110
90
60
30
15
0
-15
-30
-60
-90
-110
-130
-150
-180


Bino MP3
Bino AAC
Bino Ref Cachée
Bino Ref
HP
Stimulus bruitRose azimut du centre de gravité
Salve 1
Salve 2
Salve 3
Azimut du centre de gravité (en degrés)
180
150
130
110
90
60
30
15
0
-15
-30
-60
-90
-110
-130
-150
-180

B.4 Projection sur l'axe interaural : comparaison des résultats en HP et en binoRef

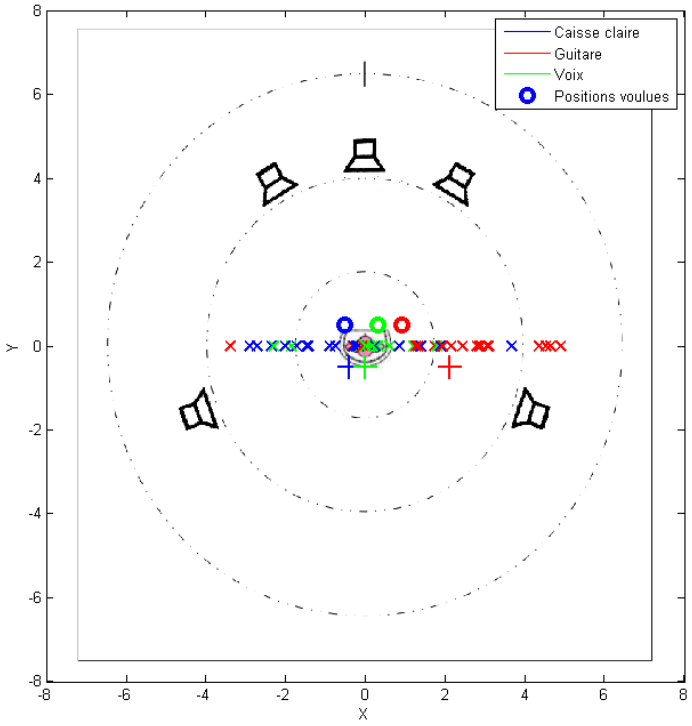


FIGURE B.51 – Stimulus rock, HP ; projection des résultats sur l'axe interaural.

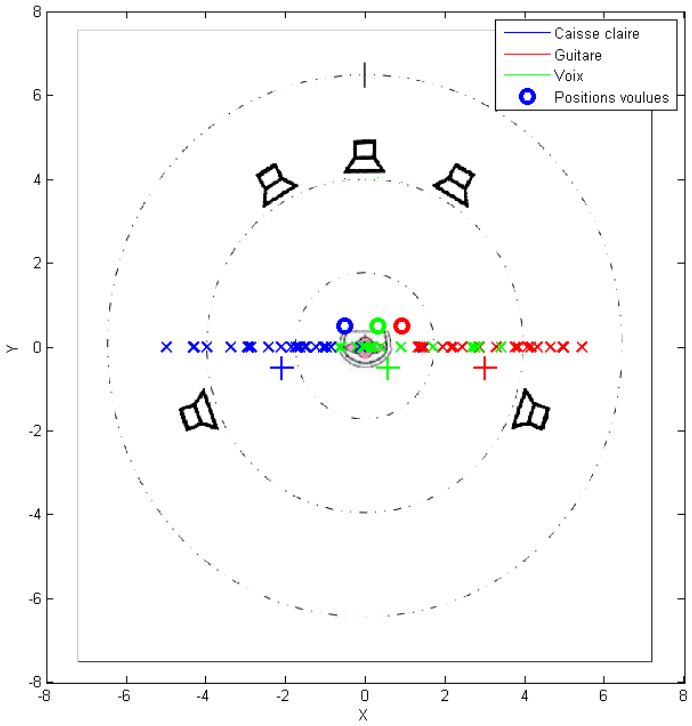


FIGURE B.52 – Stimulus rock, binoRef ; projection des résultats sur l'axe interaural.

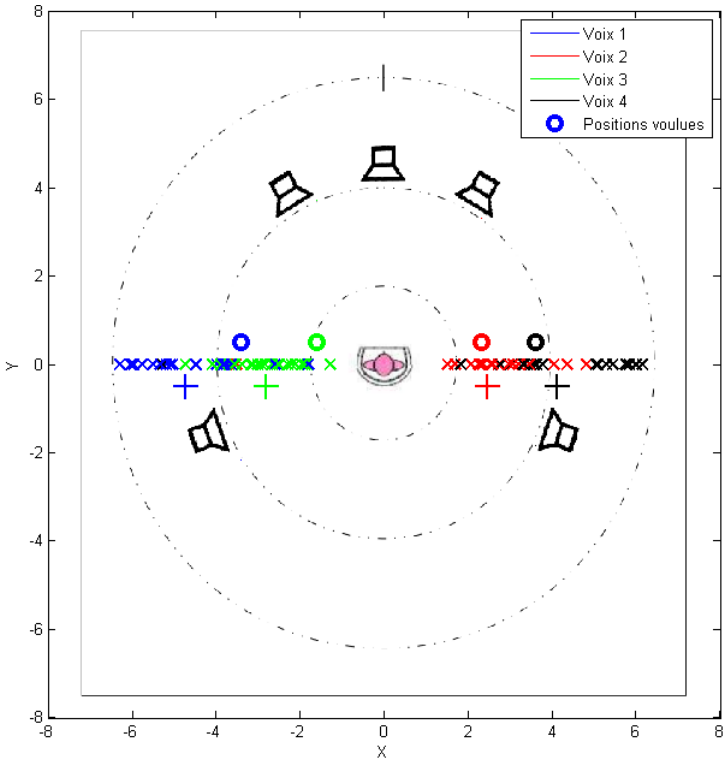


FIGURE B.53 – Stimulus voix, HP ; projection des résultats sur l'axe interaural.

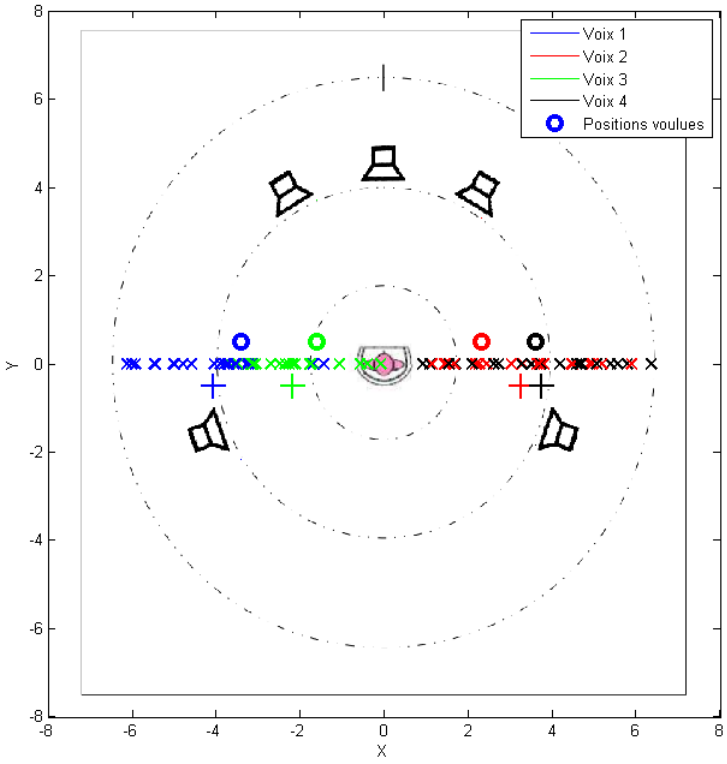


FIGURE B.54 – Stimulus voix, bino-Ref ; projection des résultats sur l'axe interaural.

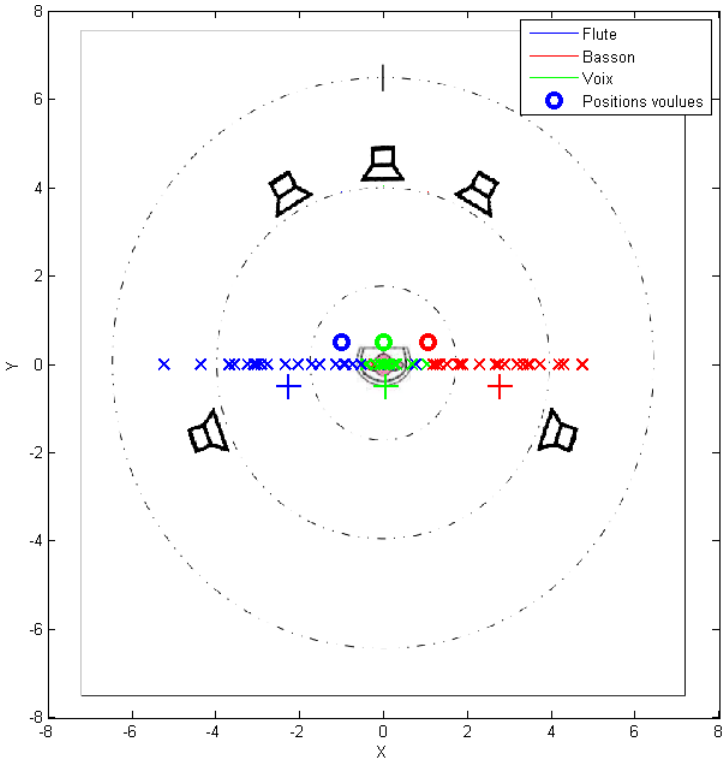


FIGURE B.55 – Stimulus classique, HP ; projection des résultats sur l'axe interaural.

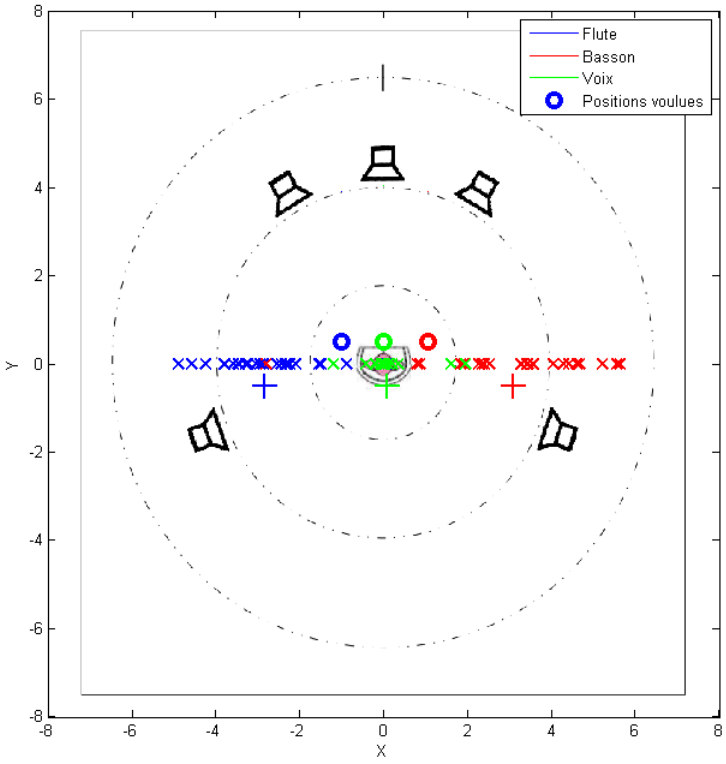


FIGURE B.56 – Stimulus classique, binoRef ; projection des résultats sur l'axe interaural.

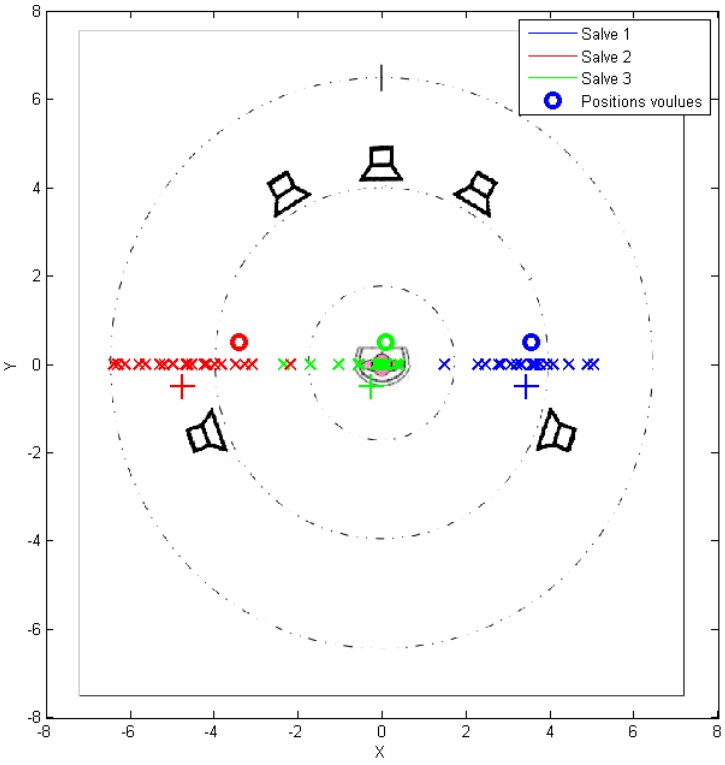


FIGURE B.57 – Stimulus bruit rose, HP ; projection des résultats sur l'axe interaural.

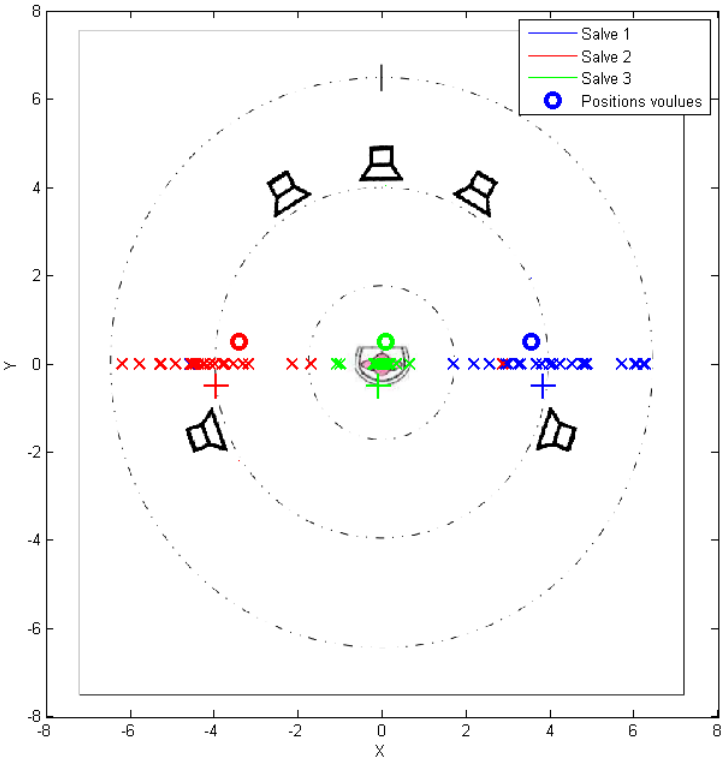


FIGURE B.58 – Stimulus bruit rose, binoRef ; projection des résultats sur l'axe interaural.

B.5 Box plots des distances

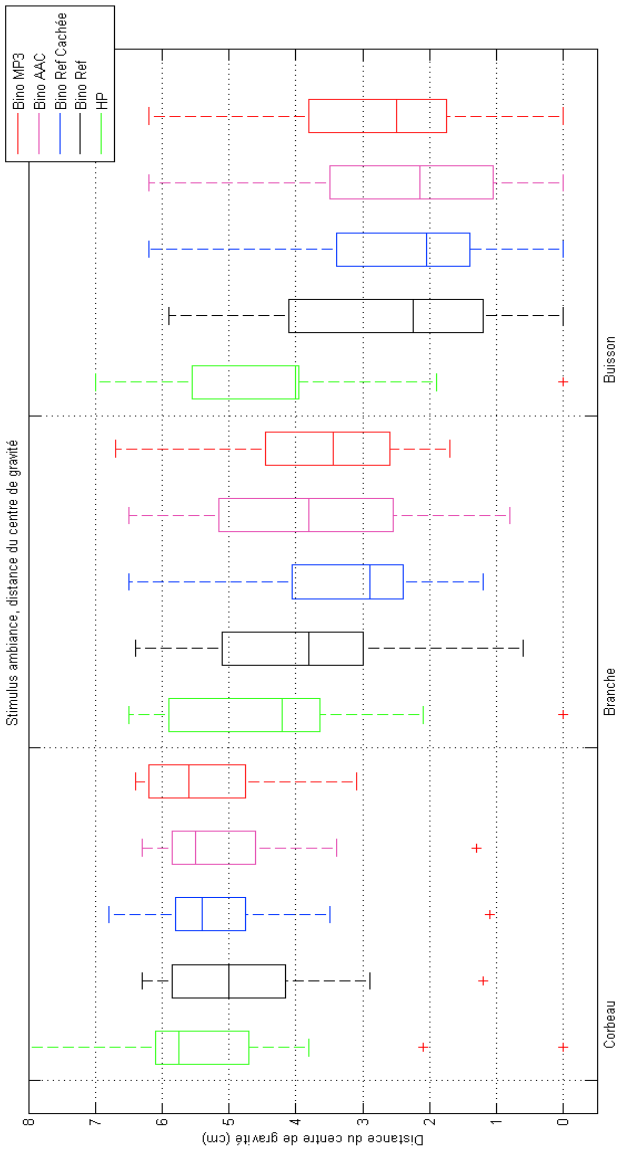


FIGURE B.59 – Box plots des distances : ambiance

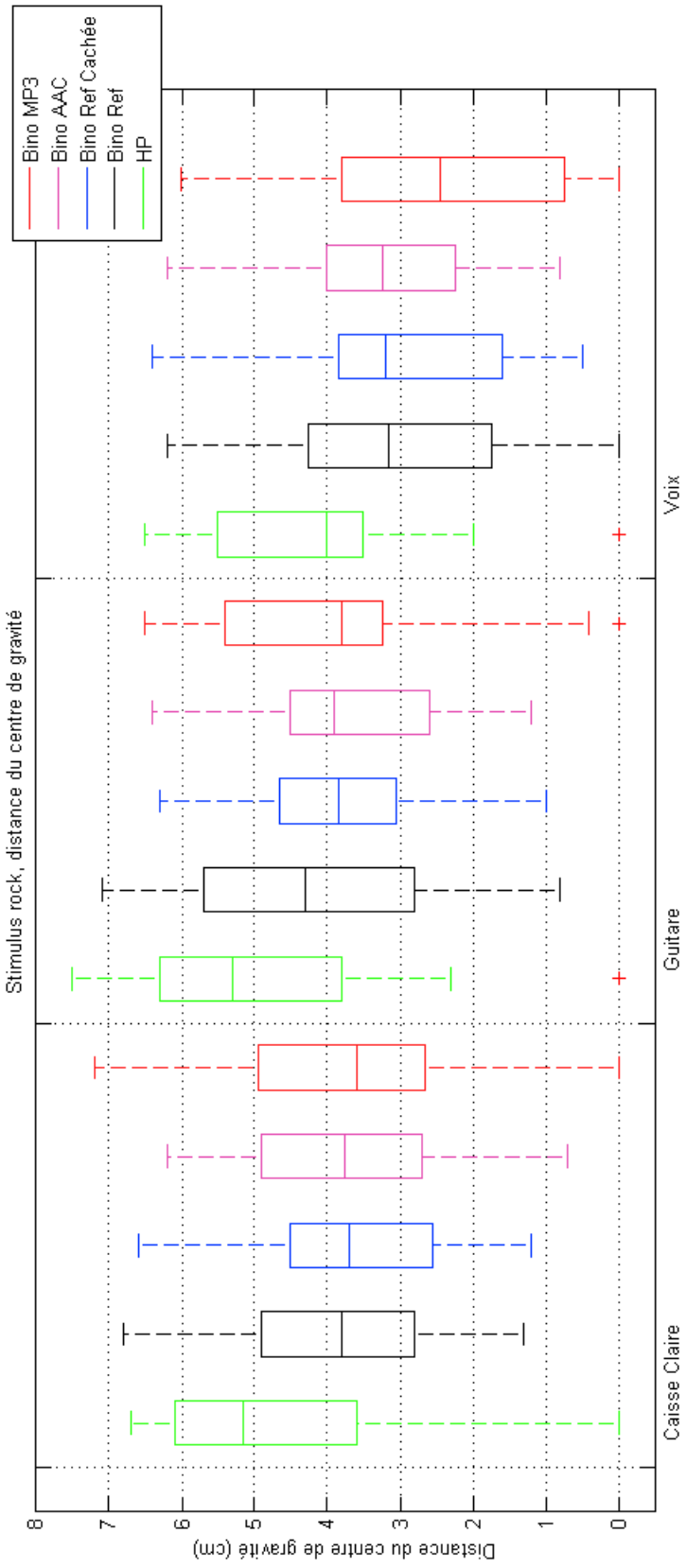


FIGURE B.60 – Box plots des distances : rock

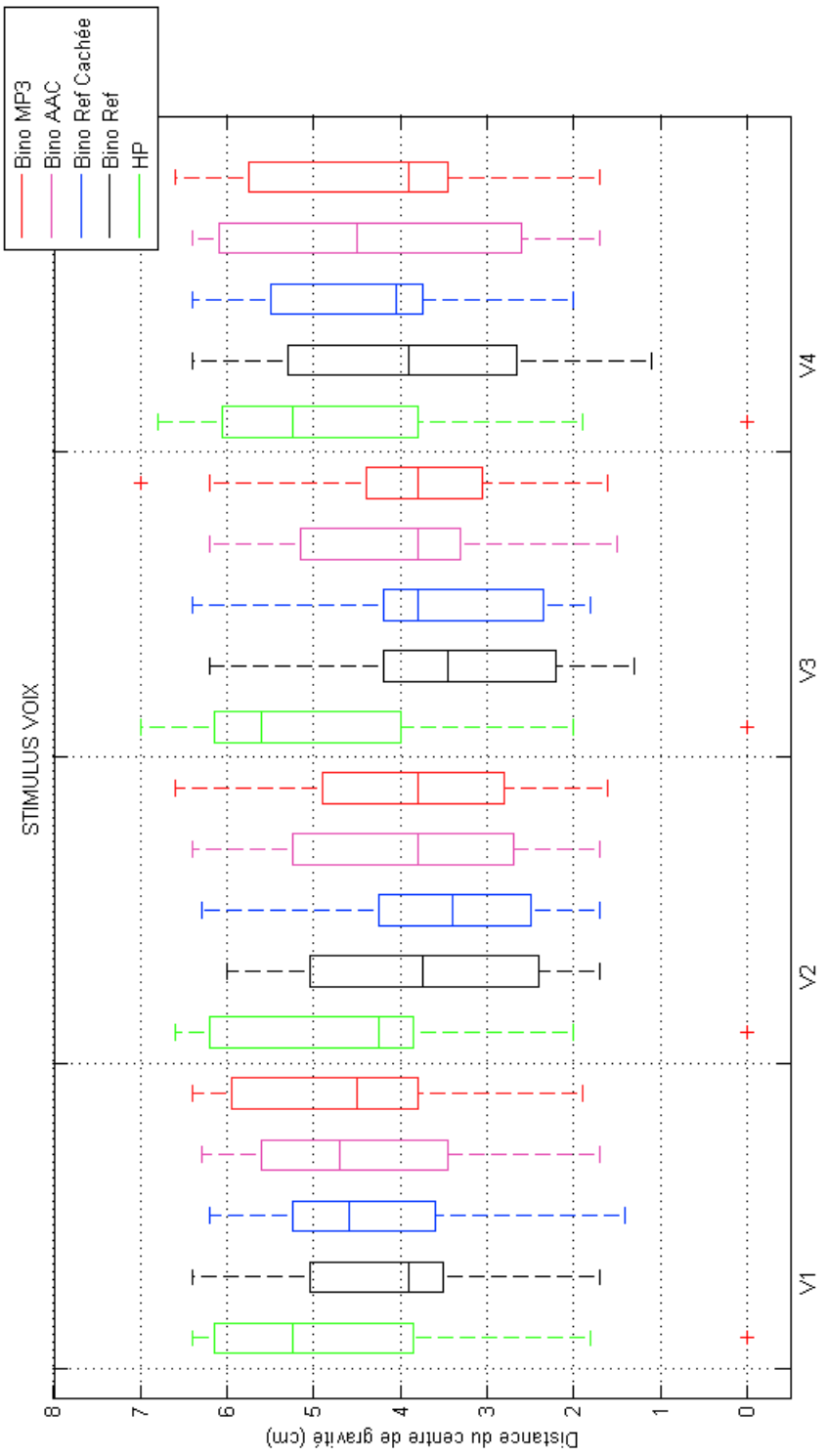


FIGURE B.61 – Box plots des distances : voix

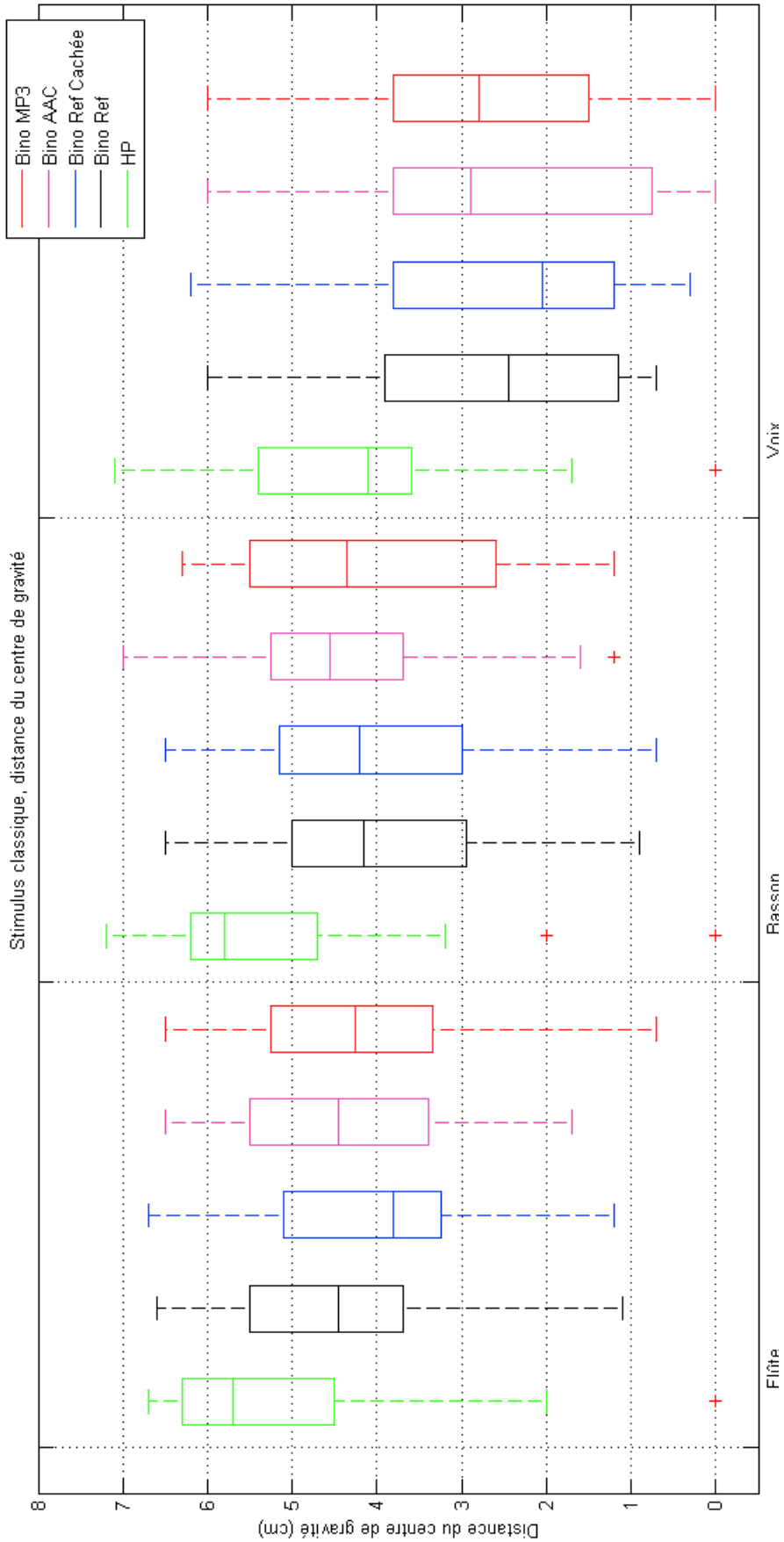


FIGURE B.62 – Box plots des distances : classique

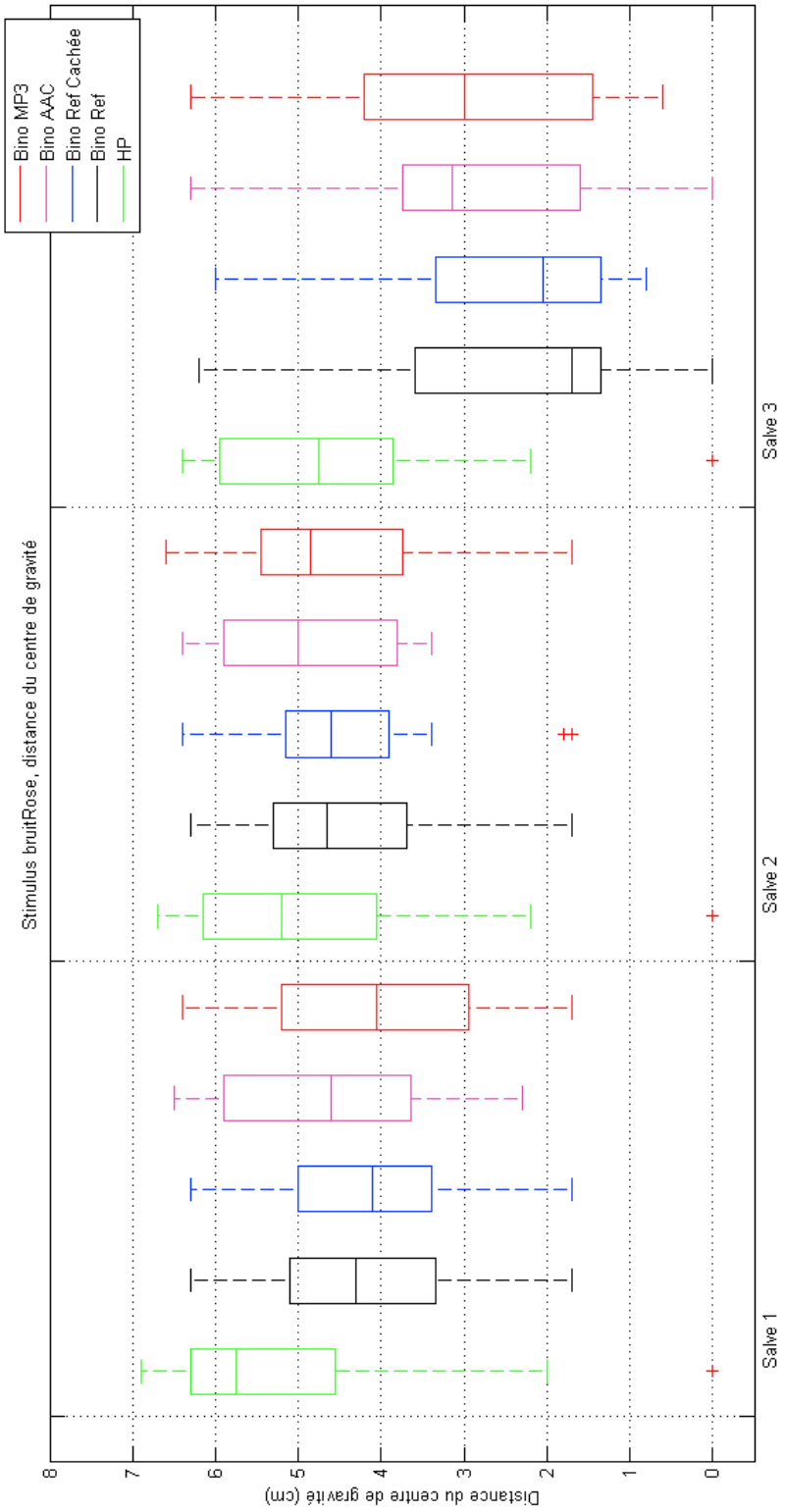
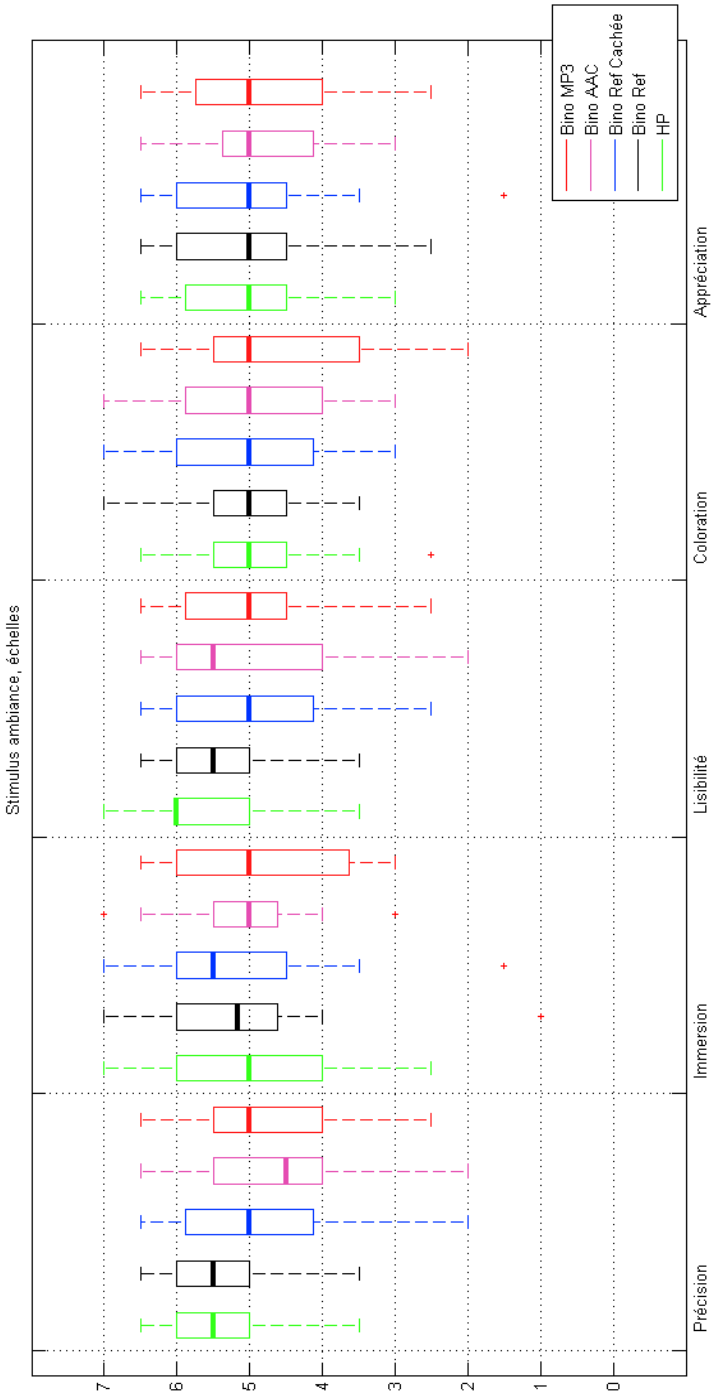
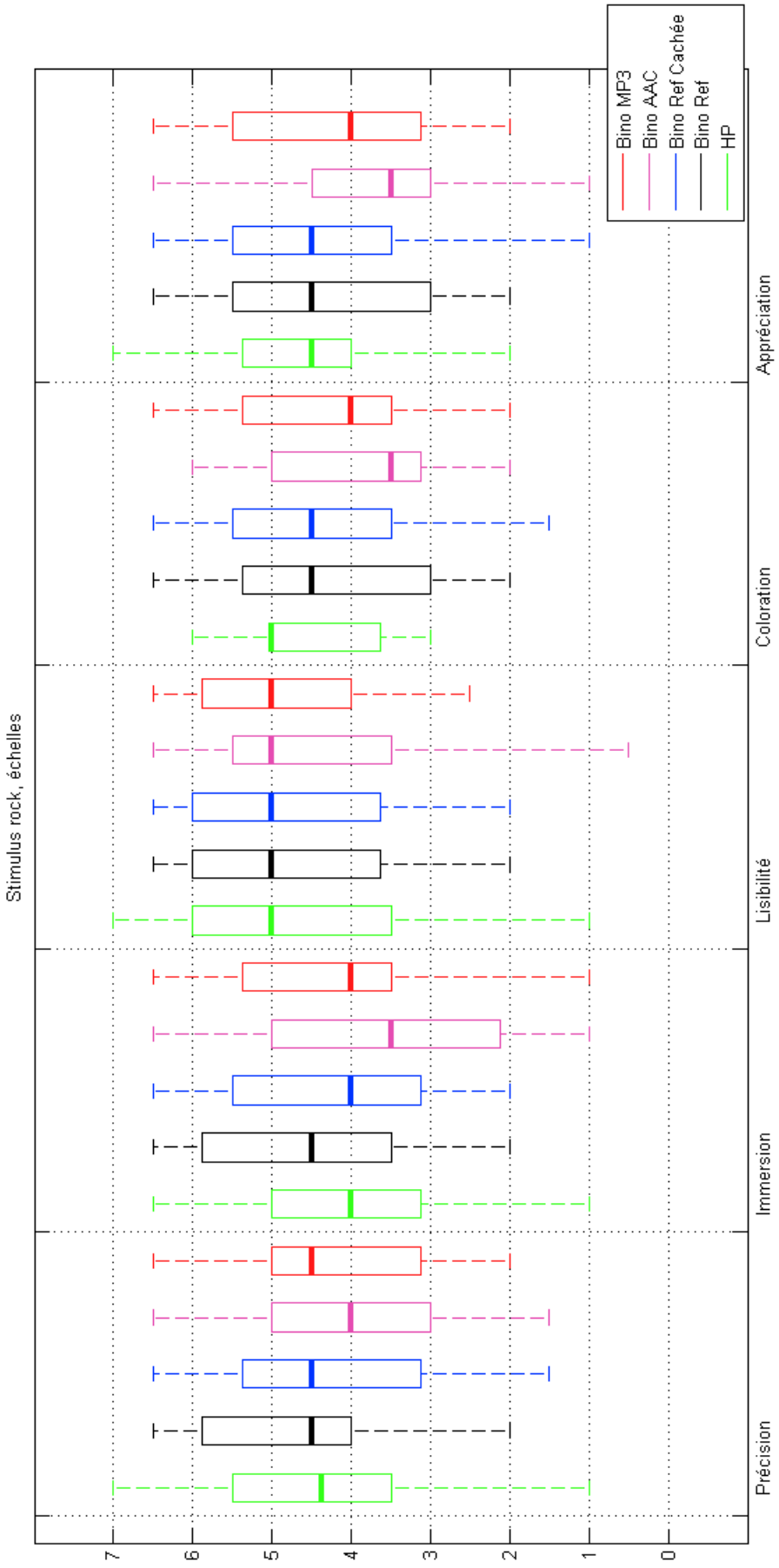
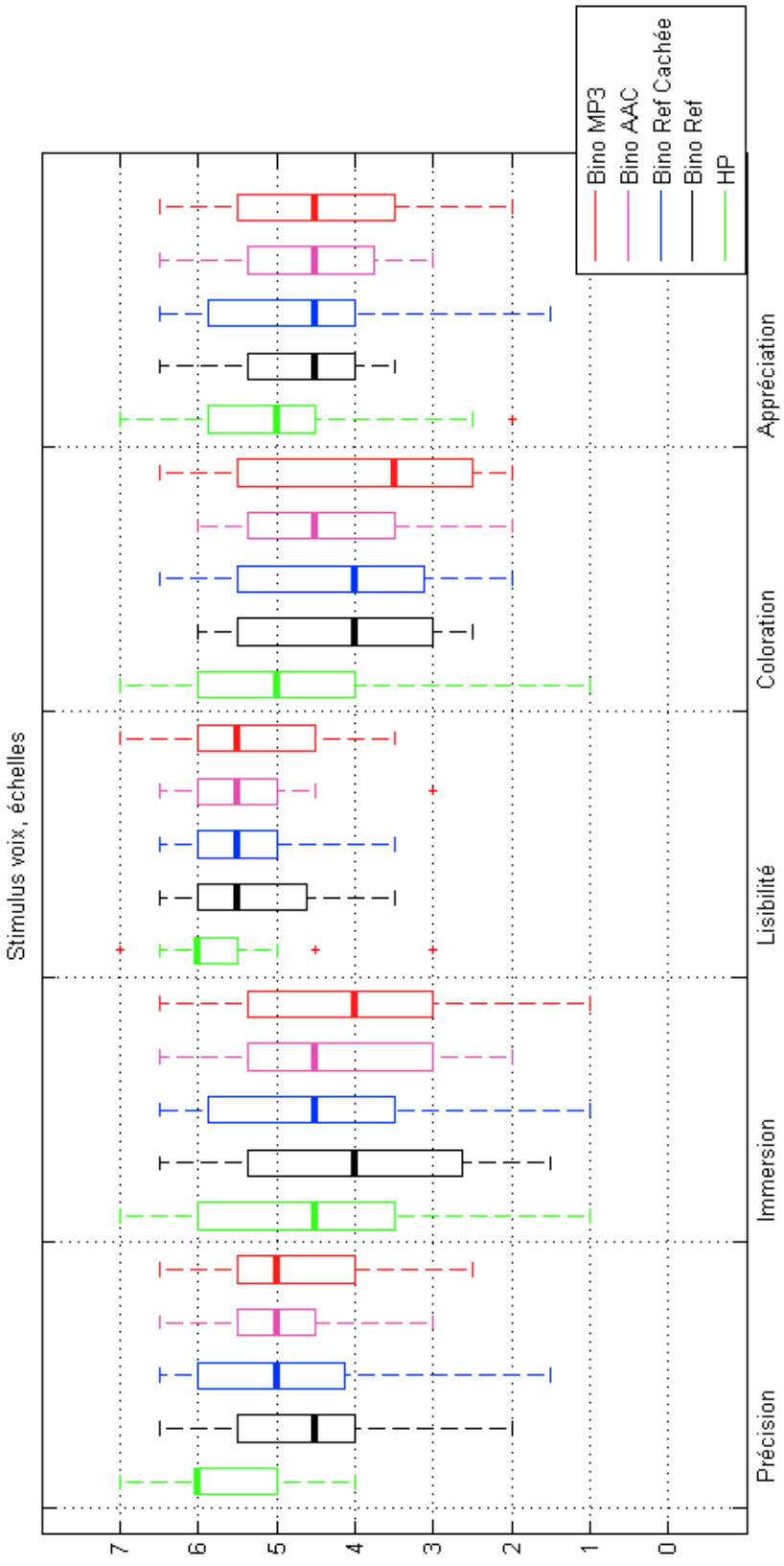


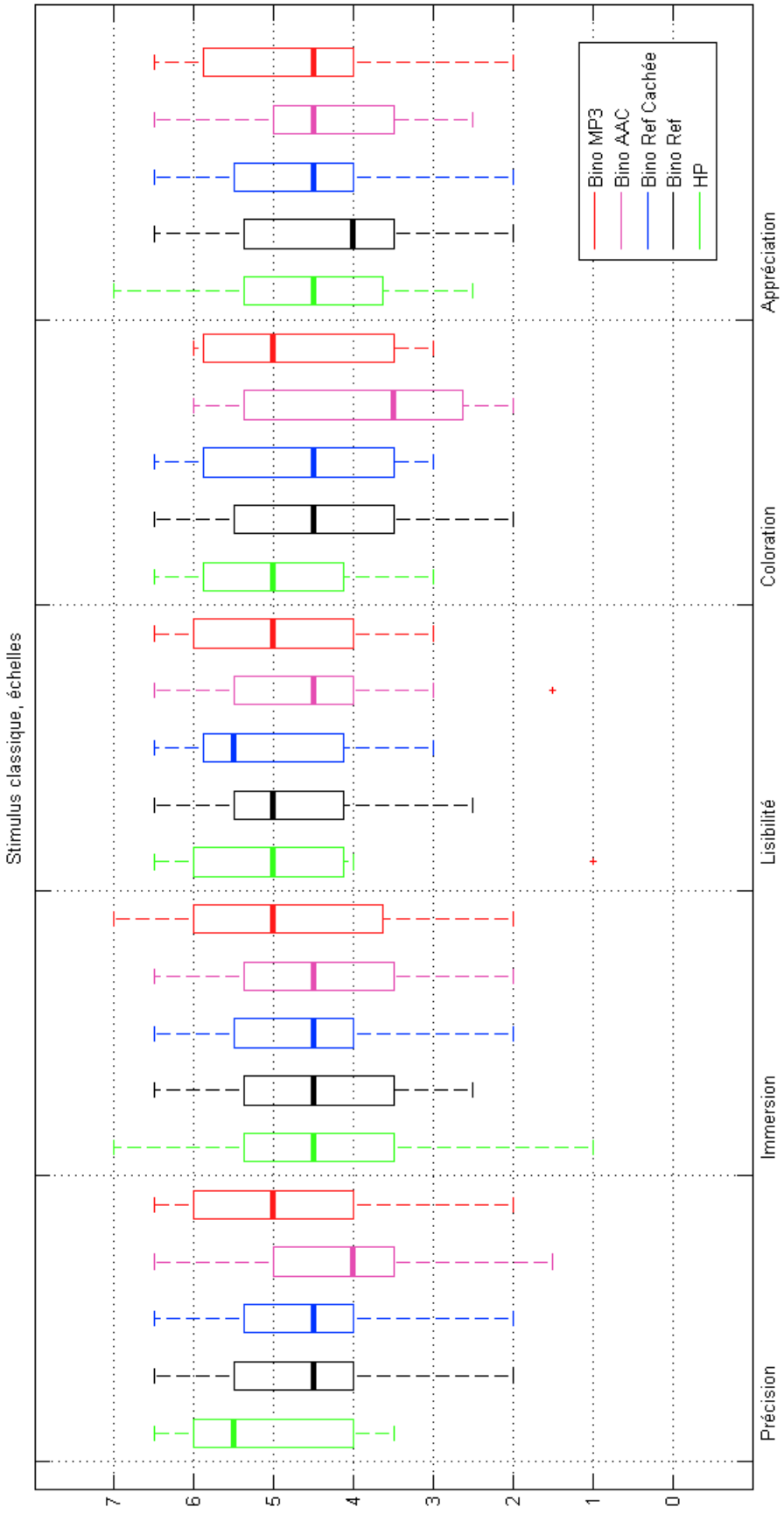
FIGURE B.63 – Box plots des distances : bruit rose

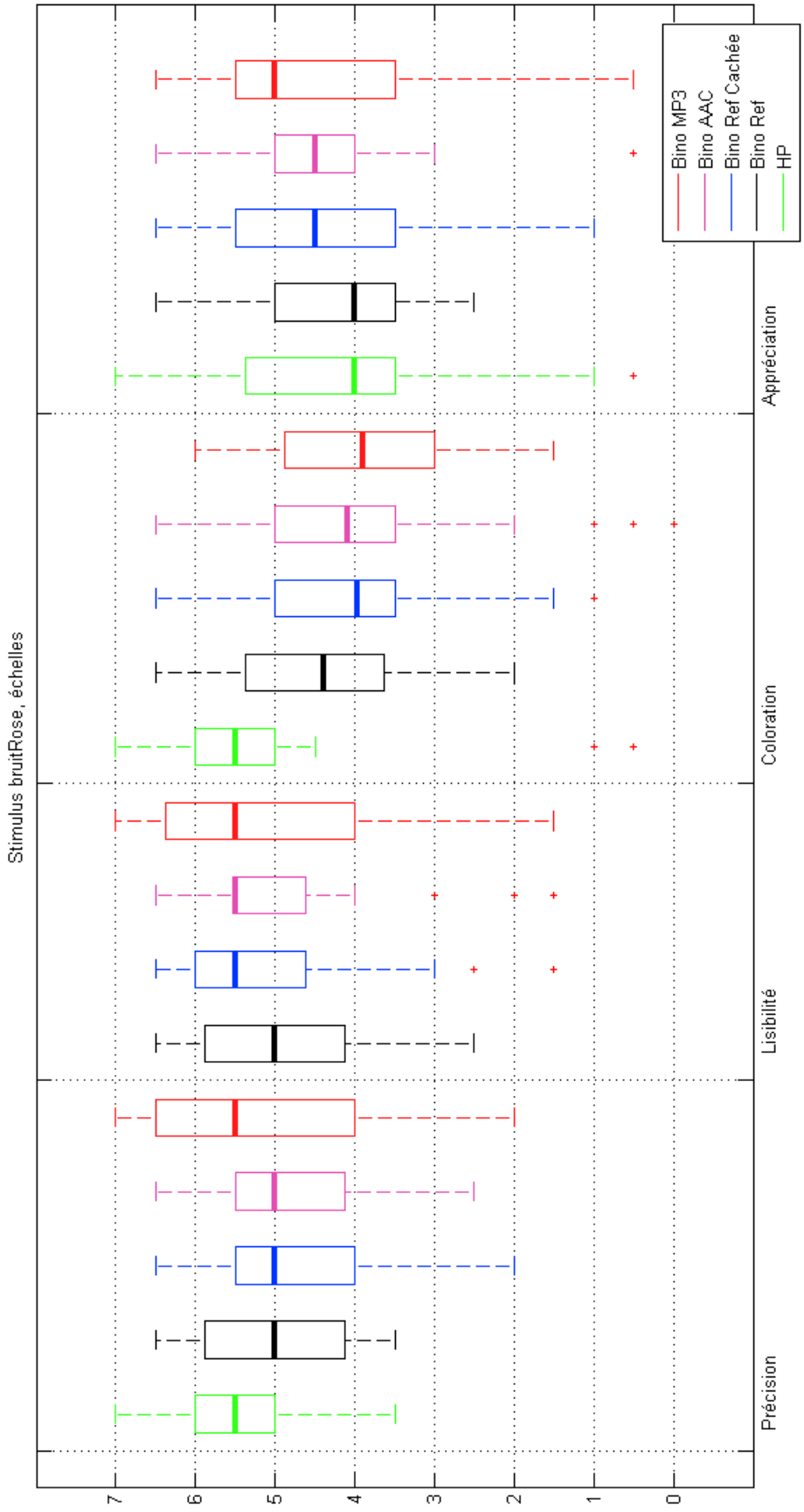
B.6 Echelles


Bino MP3
Bino AAC
Bino Ref Cachée
Bino Ref
HP
Stimulus ambiance, échelles
Précision
Immersion
Lisibilité
Coloration
Appréciation
7
6
5
4
3
2
1
0


Bino MP3
Bino AAC
Bino Ref Cachée
Bino Ref
HP
Stimulus rock, échelles
Appréciation
Coloration
Lisibilité
Immersion
Précision
7
6
5
4
3
2
1
0


Bino MP3
Bino AAC
Bino Ref Cachée
Bino Ref
HP
Stimulus voix, échelles
Précision
Immersion
Lisibilité
Coloration
Appréciation
0
1
2
3
4
5
6
7


Bino MP3
Bino AAC
Bino Ref Cachée
Bino Ref
HP
Stimulus classique, échelles
Précision
Immersion
Lisibilité
Coloration
Appréciation
0
1
2
3
4
5
6
7


Stimulus bruitRose, échelles
Précision
Lisibilité
Coloration
Appréciation
Bino MP3
Bino AAC
Bino Ref Cachée
Bino Ref
HP
7
6
5
4
3
2
1
0