

10ème Congrès Français d'Acoustique

Lyon, 12-16 Avril 2010

Perception sonore de la variabilité de positionnement d'un casque audio

Mathieu Paquier, Vincent Koehl

UEB – UBO – LISyC EA 3883, Centre Européen de Réalité Virtuelle, 25 rue Claude Chappe, 29280 Plouzané, France. Mathieu.Paquier@univ-brest.fr

La restitution au casque est utilisée dans de nombreux tests de perception sonore ainsi que dans des applications de spatialisation sonore utilisant les techniques binaurales. De même, les ingénieurs du son utilisent régulièrement des casques pour leurs travaux de prise de son ou de mixage. La problématique de la fidélité de la restitution par les transducteurs a été largement traitée par la littérature. Cependant, la variabilité de positionnement du casque sur la tête de l'auditeur n'est généralement pas prise en compte, bien que des mesures montrent que de légères modifications de la position du casque peuvent engendrer des différences spectrales très importantes. La perception du timbre et de l'espace sonore est susceptible d'être altérée par ces modifications. Cette étude a pour but d'évaluer l'audibilité de ces variations spectrales. Des enregistrements ont été effectués à l'entrée des conduits auditifs d'une tête artificielle sur laquelle différents types de casque (diffusant de la musique ou du bruit rose) ont été repositionnés de nombreuses fois. Un panel d'auditeurs devait ensuite évaluer la similarité entre ces enregistrements présentés sur un casque fixe.

1 Introduction

La restitution sonore sur casque est utilisée dans de très nombreuses applications (tests de perception sonore, restitution binaurale, mais aussi travaux des ingénieurs du son et utilisation domestique). Pour chacune de ces applications une attention particulière est portée au choix du modèle de casque, et notamment à la qualité des transducteurs. En revanche le couplage entre le casque et l'oreille d'un auditeur donné n'est généralement pas considéré. La fonction de transfert regroupant la réponse du casque et son couplage avec l'oreille d'un auditeur est appelée HPTF (HeadPhone Transfer Function). Dans certains cas de restitution binaurale (suite à une prise ou une synthèse binaurale), les HPTF sont mesurées à plusieurs reprises (avec plusieurs repositionnements de casque), moyennées, puis compensées –pour chaque auditeur- par filtrage inverse.

En résumé, que les HPTF soient compensées par filtrage inverse d'une moyenne de plusieurs mesures (pour plusieurs positions de casque) ou pas compensées du tout, la variabilité due aux différences de positionnement du casque sur la tête de l'auditeur n'est pas prise en compte. Pourtant, de légères modifications de la position du casque peuvent engendrer des différences spectrales très importantes. Cet effet semble assez limité en basses fréquences, mais il devient significatif au delà de 6kHz [1,2]. La déviation standard des HPTF est de l'ordre de 10dB pour des fréquences comprises entre 9 et 14kHz [3]. La perception d'une scène sonore (timbre et localisation des sources) pourrait être altérée par de tels changements spectraux. Certaines études ont indiqué que le lissage effectué par les filtres cochléaires était susceptible de gommer une partie de ces changements [1]. Néanmoins, en audiologie, Green [4] a observé des variations des seuils audiométriques de l'ordre de 15dB suite à des mauvais placements du casque sur les oreilles de patients.

Le but de cette étude est donc d'évaluer l'audibilité de variations spectrales dues à des changements réalistes de positions de casques.

Pour cela nous avons diffusé des extraits sonores (musique et bruit rose) sur quatre casques, repositionnés chacun à 8 reprises sur une tête artificielle équipée de microphones en entrée de conduits bloqués. Les séquences enregistrées sur la tête artificielle ont ensuite été présentées (avec un casque unique, fixe et compensé) à des sujets experts, qui devaient différencier les enregistrements issus de positions différentes d'un casque donné sur la tête artificielle.

2 Matériel et méthode

2.1 Extraits sonores

Trois extraits sonores ont été utilisés :

- 1) Bruit rose – durée 3,5s
- 2) Ben Harper (album The will to live) : I want to be ready (piste 6) - durée 5s (01'40'' à 01'45'').
- 3) Leonard Bernstein (avec le Los Angeles Philharmonic Orchestra) : West side story, Rumble, Molto allegro (piste 10) – durée 4s (00'54'' à 00'58'').

Les deux extraits musicaux ont été découpés à partir d'enregistrements du commerce (CD 44100Hz, 16 bits). Les trois extraits étaient monophoniques (seule la voie gauche des enregistrements musicaux a été conservée).

L'utilisation d'extraits monophoniques a été choisie pour que les éventuelles différences perçues par les auditeurs soient dues à des modifications de timbre, et non à des possibles altérations de l'espace stéréophonique.

2.2 Casques

Riederer [5] a montré que la reproductibilité des mesures de HPTF était meilleure avec des casques circum-auraux (dont les coussins entourent les pavillons de l'auditeur) qu'avec des casques supra-auraux (dont les coussins sont appuyés sur les pavillons).

Dans cette étude 4 modèles différents ont été utilisés:

- A) Sennheiser HD497, supra-aural
- B) Sony MDR CD580, circum/supra-aural
- C) Sennheiser HD600, circum-aural
- D) Sony MDR CD2000, circum/supra-aural

Ces quatre casques sont de bonne qualité, notamment les modèles C et D, régulièrement utilisés par les professionnels du son et dans des expériences psychoacoustiques.

Les deux casques Sony sont constitués de coussins laissant une large ouverture (cas d'un casque circum-aural), mais cette ouverture est coiffée d'un très léger tissu (cas d'un casque supra-aural). Ce tissu épousant le contour du pavillon de l'auditeur, nous avons considéré ces casques à la fois circum- et supra-auraux.

Aucun casque fermé n'a été testé car ceux-ci n'auraient pas permis une prise de son fiable en entrée de conduit fermé [6].

2.3 Enregistrements

Les enregistrements ont été effectués grâce à une tête artificielle Neumann KU 100 équipée de ses propres microphones, placés en entrée de conduit auditif externe bouché (mesure dite "conduit bloqué").

Idéalement, il aurait fallu enregistrer les séquences jouées dans les casques dans les oreilles des auditeurs participant aux tests de discrimination ultérieurs. Néanmoins enregistrer les séquences dans chaque paire d'oreilles compliquait énormément l'expérience, pour un gain potentiellement faible. En effet, Møller et al. [6] ont montré que les positionnements des casques étaient plus répétables lorsque des sujets posaient eux même les casques sur leur propre oreilles ; néanmoins McAnally et Martin [1] ont indiqué que la variabilité de HPTF mesurées plusieurs fois en entrée de conduit fermé était très semblable quand les enregistrements étaient effectués sur des sujets ou quand ils étaient effectués sur une tête artificielle.

Plusieurs études ont montré les avantages de l'enregistrement en entrée de conduit fermé : d'une part à partir de ce point toutes les informations de direction sont comprises, et la variabilité interindividuelle est réduite (car principalement liée aux différences de formes de conduits entre les sujets [7]), d'autre part cette technique permet l'utilisation de microphones aux diaphragmes assez larges, et donc avec des meilleurs rapports signal sur bruit [8].

L'utilisation d'une sonde (tube de très petit diamètre inséré dans le conduit, avec un petit microphone à son autre extrémité) permet des mesures en conduit ouvert, éventuellement assez proche du tympan, néanmoins Møller et al. [6] ont indiqué que la reproductibilité des mesures réalisées avec cette technique était assez faible, même pour un seul sujet (notamment à cause de la difficulté à placer le tube à un endroit précis dans le conduit).

En revanche, Møller et al [6] insistent sur le fait qu'une mesure avec un conduit fermé n'est valable que pour des casques vérifiant les conditions FEC (« Free air Equivalent Coupling to the ear ») : cela signifie que la charge

acoustique du conduit auditif par le casque est négligeable. Si le casque ne vérifie pas ces conditions (ce qui est le cas de tous les casques fermés), la mesure doit être effectuée en conduit ouvert (prenant donc en compte la modification de l'impédance de l'ensemble tympan+conduit par la "fermeture" de l'oreille par le casque. En acceptant une charge acoustique du conduit par le casque entraînant une variation de 4dB des mesures en conduit ouvert avec et sans casque (définie par Møller et al. [6] comme *Pressure Division Ratio*), beaucoup (mais pas tous) de casques dit "ouverts" valident les conditions FEC. Cette approximation n'est pas abusive puisque les différences inter- et intra-individuelles de HPTF sont largement supérieures à 4dB.

Des casques ouverts, considérés FEC, ont donc été utilisés, et les enregistrements ont été effectués en entrée des conduits fermés de la tête artificielle Neumann KU100.

Chacun des casques a été enlevé et remplacé à 8 reprises sur la tête artificielle (2 manipulateurs ont effectué chacun 4 repositionnements).

Pour chaque position, les trois extraits ont été joués dans le casque et ont été enregistrés, via les microphones de la tête, sur un ordinateur MacBookPro équipé d'une carte son MBox2 ($f_c=44100$, 16 bits). Les enregistrements ont été effectués en cabine audiométrique (figure 1).

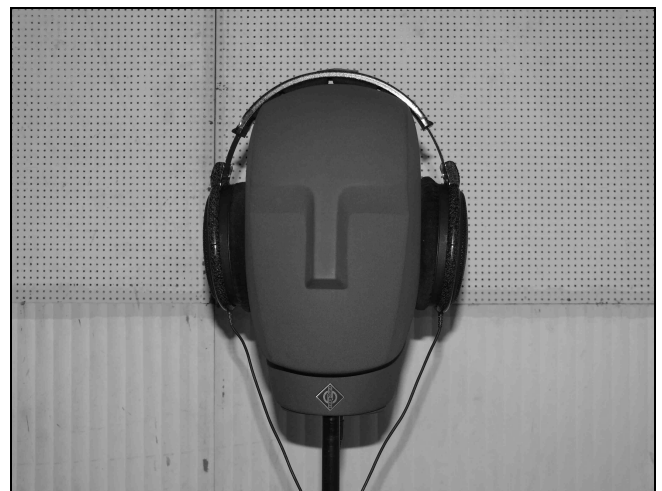


Figure 1 : Casque Sennheiser HD600 posé sur la tête Neumann KU 100

2.4 Normalisation des niveaux des enregistrements

Bien que dans le test perceptif les différents casques ne soient pas comparés, les niveaux moyens des casques ont été normalisés. De cette façon si l'audibilité de variabilité de positionnement est différente selon le casque, cette différence ne peut être due à une modification du volume.

Le volume de chaque enregistrement a été normalisé pour que la moyenne des volumes sur les 8 positions soit identique pour chaque casque. Cela revient en fait à compenser les sensibilités différentes des quatre casques. Par contre les volumes des 8 enregistrements issus d'un même casque n'ont pas été compensés entre eux, puisque leurs différences sont dues à la position du casque, et constituent donc un indice d'audibilité de la variabilité de positionnement.

2.5 Compensation du casque de restitution (utilisé pendant le test perceptif)

Le casque utilisé pour la restitution sonore pendant le test perceptif était le Sony MDR CD2000, déjà utilisé pour les enregistrements, et montrant des HPTF particulièrement stables et large bande.

Ce casque a été compensé par filtrage inverse, en considérant la moyenne de 8 mesures (figure 2). Pour ces mesures, le casque a été déposé sur la tête artificielle, et un sinus glissant a été diffusé et enregistré via les microphones de la tête.

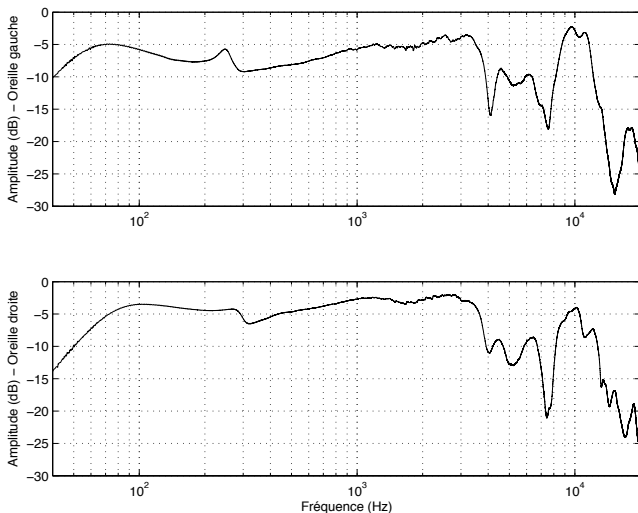


Figure 2 : Réponse moyenne (sur 8 mesures) du casque Sony MDR CD2000 pour l'oreille gauche (en haut) et l'oreille droite (en bas)

La réponse des casques est généralement caractérisée en basses fréquences par des légères fluctuations, et en hautes fréquences par des pics/creux plus importants, à forts coefficients de qualité, et plus individualisés. Ici la compensation du casque de restitution n'a pas été individuelle (en considérant les HPTF mesurées sur chaque sujet) ; en effet, Møller et al. [6] ont indiqué que même si une compensation individuelle est idéale (mais compliquée à mettre en œuvre), une compensation générique est suffisante, et est généralement utilisée dans la plupart des applications.

Par ailleurs, puisque le test perceptif sera une comparaison entre plusieurs enregistrements dont les artefacts de restitution seront toujours identiquement compensés, une compensation parfaite n'est pas indispensable.

2.6 Protocole de test

Dix auditeurs ont participé à l'étude. Ceux-ci étaient étudiants en formation aux métiers du son (Master Image & Son de l'Université de Bretagne Occidentale). Bien que n'ayant pas d'expérience des tests perceptifs, ces étudiants, formés à l'écoute analytique, peuvent être considérés comme "experts" [9].

Les auditeurs étaient placés dans une cabine audiométrique, face à un écran, et disposaient d'une souris. Ils étaient équipés du casque Sony MDR CD2000, compensé, qu'ils devaient placer eux-mêmes sur leur tête.

Après avoir positionné le casque confortablement, ils avaient pour consigne de ne plus modifier son placement une fois le test entamé.

La procédure utilisée ici était un protocole 3I-3AFC (3-Interval 3-Alternative Forced Choice) : trois séquences étaient successivement présentées au sujet. Deux de ces séquences étaient identiques et la troisième était différente (issue d'une position de casque différente). Les trois séquences étaient présentées dans un ordre aléatoire ; la séquence présentée une seule fois (nommée ensuite "séquence unique") pouvait donc être en première, deuxième, ou troisième position. Après l'écoute, le sujet devait indiquer laquelle des trois séquences lui semblait être différente des deux autres, en cliquant sur une des trois cases indiquées à l'écran. Une fois sûr de son choix, l'auditeur devait valider sa réponse en cliquant sur la case "suivant", qui provoquait la lecture du trio de séquences suivant.

4 casques ont été utilisés pour les enregistrements.

Pour chaque casque, 8 positionnements différents ont été effectués, et pour chaque position 3 extraits sonores ont été joués dans le casque et enregistrés par la tête artificielle.

Dans le test perceptif, pour chaque trio de séquences présenté, un couple de séquences était tiré au sort parmi les 28 couples possibles parmi 8 séquences ($8 \times (8-1)/2 = 28$), pour un casque et un extrait donné. Dans un trio, une des séquences du couple était donc présentée une fois, et l'autre séquence du couple était présentée deux fois.

Les 28 trios concernant un extrait (déterminé aléatoirement parmi les trois possibles) étaient tout d'abord présentés, puis les 28 suivants (pour le deuxième extrait), et enfin les 28 derniers trios (pour le troisième extrait). Pour un casque donné, $28 \times 3 = 84$ trios étaient donc présentés.

Le test était composé de deux sessions de 45 minutes. Au cours d'une session, deux casques (tirés au sort) étaient testés. Pour chaque session $84 \times 2 = 168$ trios étaient donc présentés, et en tout chaque auditeur a entendu $168 \times 2 = 336$ trios. Le sujet prenait une pause de 5 minutes à mi-session, après avoir évalué les 84 trios relatifs au premier des deux casques. La première des deux sessions démarrait par un pré-test de 5 minutes environ, où l'auditeur pouvait se familiariser avec l'interface de réponse et les stimuli à discriminer.

Le volume du casque pendant le test, identique pour tous les auditeurs, a été fixé par les expérimentateurs de façon à ce que le niveau d'écoute soit réaliste pour les extraits musicaux acoustiques.

3 Résultats

3.1 Audibilité de la variabilité de positionnement

Avec le protocole 3I-3AFC, un sujet répondant au hasard a une chance sur trois (soit 33,33%) de donner la bonne réponse. Le test statistique utilisé ici compare donc les distributions de réponses des sujets (pour des conditions données) à la valeur nominale 33,33%, et vérifie ainsi que la probabilité de réponse au hasard est (ou n'est pas) inférieure à 5%.

Le test statistique comparant la valeur nominale 33,33% à la distribution des réponses des sujets a indiqué que quel que soit le casque considéré et quel que soit l'extrait, les sujets ont significativement détecté les séquences uniques ("unique" signifiant présentée une seule fois dans le trio), et donc perçu la variabilité de positionnement des casques ($p=0,001$ pour les conditions les plus critiques, figure 3).

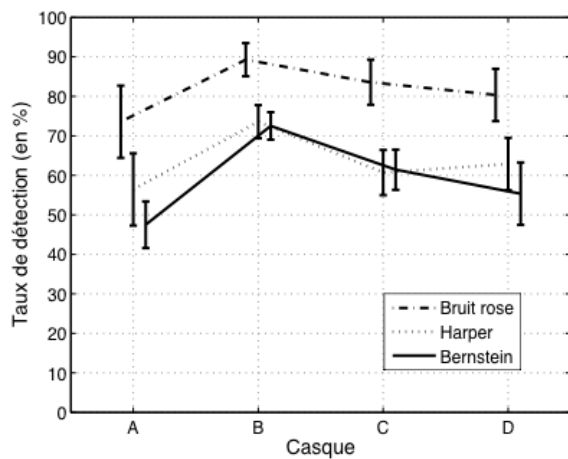


Figure 3 : Taux de détection (moyennes et intervalles de confiance à 95%) de la séquence unique par les sujets, en fonction du casque et de l'extrait.

3.2 Effet du casque

Le test statistique comparant la valeur nominale 33,33% à la distribution des réponses des sujets a indiqué que pour chacun des casques considérés et tous extraits confondus, les sujets ont significativement détecté les séquences uniques, et donc perçu la variabilité de positionnement du casque considéré ($p < 0,0001$, figure 4).

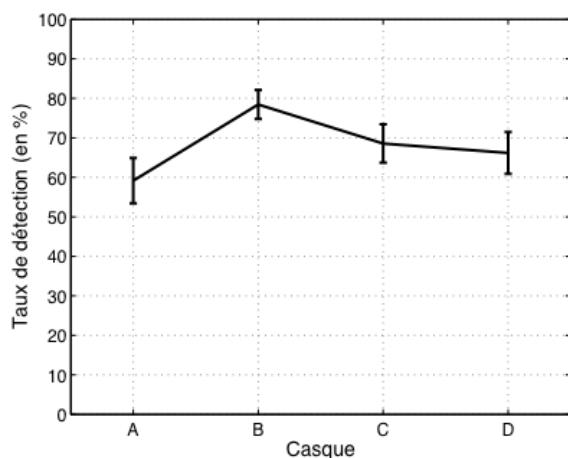


Figure 4 : Taux de détection (moyennes et intervalles de confiance à 95%) de la séquence unique par les sujets, en fonction du casque (tous extraits confondus).

Par ailleurs, l'ANOVA a permis d'observer un effet significatif du casque utilisé pour l'enregistrement ($F(3,108)=20,199$, $p < 0,0001$), et le post-hoc LSD de Fisher (comparaison des casques deux à deux) a indiqué que les casques étaient tous significativement différents ($p < 0,01$) sauf pour les casques C et D, pour lesquels $p=0,345$. La tâche de détection de la séquence unique a été plus difficile avec le casque A (Sennheiser HD497) qu'avec les autres casques. Au contraire, l'audibilité de la variabilité de positionnement a été significativement facilitée avec le casque B (Sony MDR CD580).

3.3 Effet de l'extrait

Le test statistique comparant la valeur nominale 33,33% à la distribution des réponses des sujets a indiqué que pour chacun des extraits considérés et tous casques confondus, les sujets ont significativement détecté les séquences uniques, et donc perçu la variabilité de positionnement des casques ($p < 0,0001$, figure 5).

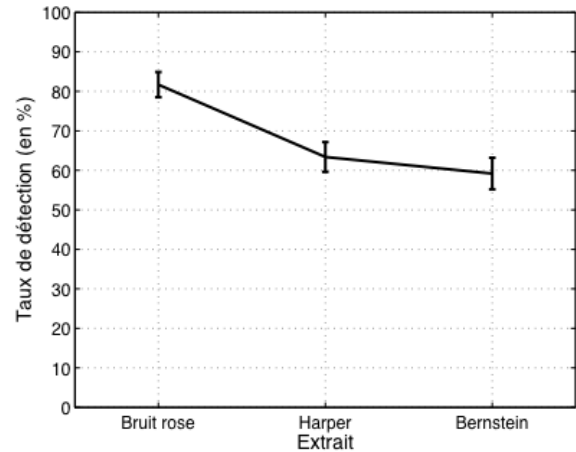


Figure 5 : Taux de détection (moyennes et intervalles de confiance à 95%) de la séquence unique par les sujets, en fonction de l'extrait (tous casques confondus).

Par ailleurs, l'ANOVA a permis d'observer un effet significatif de l'extrait ($F(2,108)=60,603$, $p < 0,0001$), et le post-hoc LSD de Fisher (comparaison des extraits deux à deux) a indiqué que les extraits musicaux étaient significativement différents du bruit rose ($p < 0,0001$), et quasiment significativement différents l'un de l'autre ($p=0,056$). Ainsi, l'audibilité de la variabilité de positionnement a été largement plus facile avec le bruit rose qu'avec les extraits musicaux.

4 Discussion

Le principal résultat de cette étude est la confirmation de l'audibilité de la variabilité de positionnement d'un casque. Même si les taux de détection de la séquence unique (présentée une seule fois dans le trio) varient d'un casque à l'autre et d'un extrait à l'autre, ceux-ci sont toujours significativement et largement supérieurs à la chance. Comme le laissaient prévoir les mesures réalisées dans les études passées [1,3], déposer un casque et le replacer modifie –d'un point de vue objectif mais aussi perceptif– le signal. Le lissage fréquentiel effectué par les filtres cochléaires [1] ne semble donc pas suffire à gommer les différences induites par les repositionnements successifs.

Dans cette étude, les signaux issus des casques ont été enregistrés sur une tête artificielle, et les casques ont donc été manipulés par les expérimentateurs. Il est possible que les repositionnements successifs des casques aient été plus répétables si des sujets avaient eux-mêmes effectué cette opération sur leurs propres oreilles [6]. Néanmoins la significativité des résultats et l'audibilité évidente des repositionnements de casques sur la tête artificielle laissent penser que même si l'expérience avait été menée avec des auditeurs-manipulateurs, l'audibilité des repositionnements successifs aurait été conservée.

Le fait que les différences induites par les positionnements successifs des casques soient perceptibles ne signifie pas qu'une compensation soit indispensable. Concernant la synthèse/restitution binaurale, des études ont montré que l'importance des pics des HRTF (nécessaires à la synthèse binaurale) était telle que la variabilité des HPTF était secondaire [1]. Concernant les travaux des ingénieurs du son, il est probable que les choix de microphones, d'égalisation, de traitement, etc... entraînent des modifications du signal au moins aussi grandes que celles engendrées par les différences de positions de casques.

Le fait que différentes positions de casque engendrent des sons restitués différents (physiquement et perceptivement) est d'ailleurs un problème difficile à résoudre : une compensation en temps réel d'un casque et de sa position signifierait aussi une mesure en temps réel de sa réponse in situ (dans les oreilles de l'auditeur). Cette mesure ne pourrait être possible qu'avec une sonde puisqu'il ne serait pas possible de bloquer le conduit en conditions réelles d'utilisation. Or les mesures avec sonde sont difficiles à mettre en œuvre : difficulté de positionnement précis de la sonde, calibration de la sonde au prix d'une égalisation radicale et potentiellement dégradante pour le signal, microphones en bout de sonde non adaptés aux mesures large bande etc... D'ailleurs les principales utilisations d'une sonde ont été menées avec des prototypes non commercialisés [10].

En résumé, cette étude constitue surtout une mise en garde et la consigne en résultant pour les utilisateurs de casques est une simple prise en compte de la grande variabilité perceptive des positionnements de casques.

L'audibilité de la variabilité de positionnement des casques était plus flagrante avec le bruit rose qu'avec les extraits musicaux. Cette observation peut avoir plusieurs explications : d'une part le bruit rose est quasi-stationnaire, ce qui facilite l'effort de mémoire auditive, non négligeable avec les extraits musicaux, même s'ils sont très courts. D'autre part l'exhaustivité du contenu fréquentiel du bruit rose assure que les modifications spectrales liées à la position du casque (aux coefficients de qualité parfois très élevés) concernent des zones fréquentielles énergétiques du signal joué. Ce n'est pas forcément le cas avec les extraits musicaux, au contenu spectral plus irrégulier, et de plus fluctuant.

5 Conclusion

Cette étude a confirmé l'audibilité des différences spectrales induites par la variabilité de positionnement des casques, et déjà observées par des mesures.

Même si les différences perçues ont varié d'un casque à l'autre et d'un contenu sonore à l'autre, celles-ci se sont toujours avérées significatives, quelles que soient les conditions.

Remerciements

Les auteurs tiennent à remercier Rozenn Nicol (Orange Labs Lannion) pour le prêt des casques, Brice Jantzen (Centre auditif Entendre Brest) pour ses conseils et son aide logistique, Simeon Delikaris-Manias pour son assistance en traitement du signal, la formation Image & Son pour le prêt de la tête artificielle et de la cabine audiométrique, ainsi que tous les auditeurs ayant participé à l'expérience.

Références

- [1] McAnally K.I., Martin R.L., "Variability in the Headphone-to-Ear-Canal Transfer Function", *Journal of the Audio Engineering Society*, 50(4), 263-266 (2002).
- [2] Toole F.E., "The Acoustics and Psychoacoustics of Headphones", 2nd AES International Conference, Preprint C1006 (May 1984).
- [3] Kulkarni A., Colburn H.S., "Variability in the characterization of the headphone transfer-function", *Journal of the Acoustical Society of America*, 107(2), 1071-1074 (2000).
- [4] Green D.S., "Pure tone air conduction threshold ", In: Katz J., "Handbook of clinical audiology ", chapter 5.
- [5] Riederer K.A.J., "Repeatability Analysis of Head-Related Transfer Function Measurements", 108th AES Convention, Preprint 4846 (September 1998).
- [6] Møller H., Hammershøi D., Jensen C.B. and Sørensen M.F., "Transfer Characteristics of Headphones Measured on Human", *Journal of the Audio Engineering Society*, 43(4), 203-217 (1995).
- [7] Hammershøi D., Møller H., "Sound transmission to and within the human ear canal", *Journal of the Acoustical Society of America*, 100(1), 408-427 (1996).
- [8] Møller H., Sørensen M.F., Jensen C.B. and Hammershøi D., "Binaural technique: do we need individual recordings", *Journal of the Audio Engineering Society*, 44(6), 451-469 (1996).
- [9] ISO 8586-2, "Sensory analysis – General guidance for the selection, training and monitoring of assessors – Part 2: Experts", *International Organization for Standardization*, (1984).
- [10] Toole F.E., " Binaural Record/Reproduction Systems and Their Use in Psychoacoustic Investigations", 91st AES Convention, Preprint 3179 (October 1991).