

<https://www.lesonbinaural.fr>

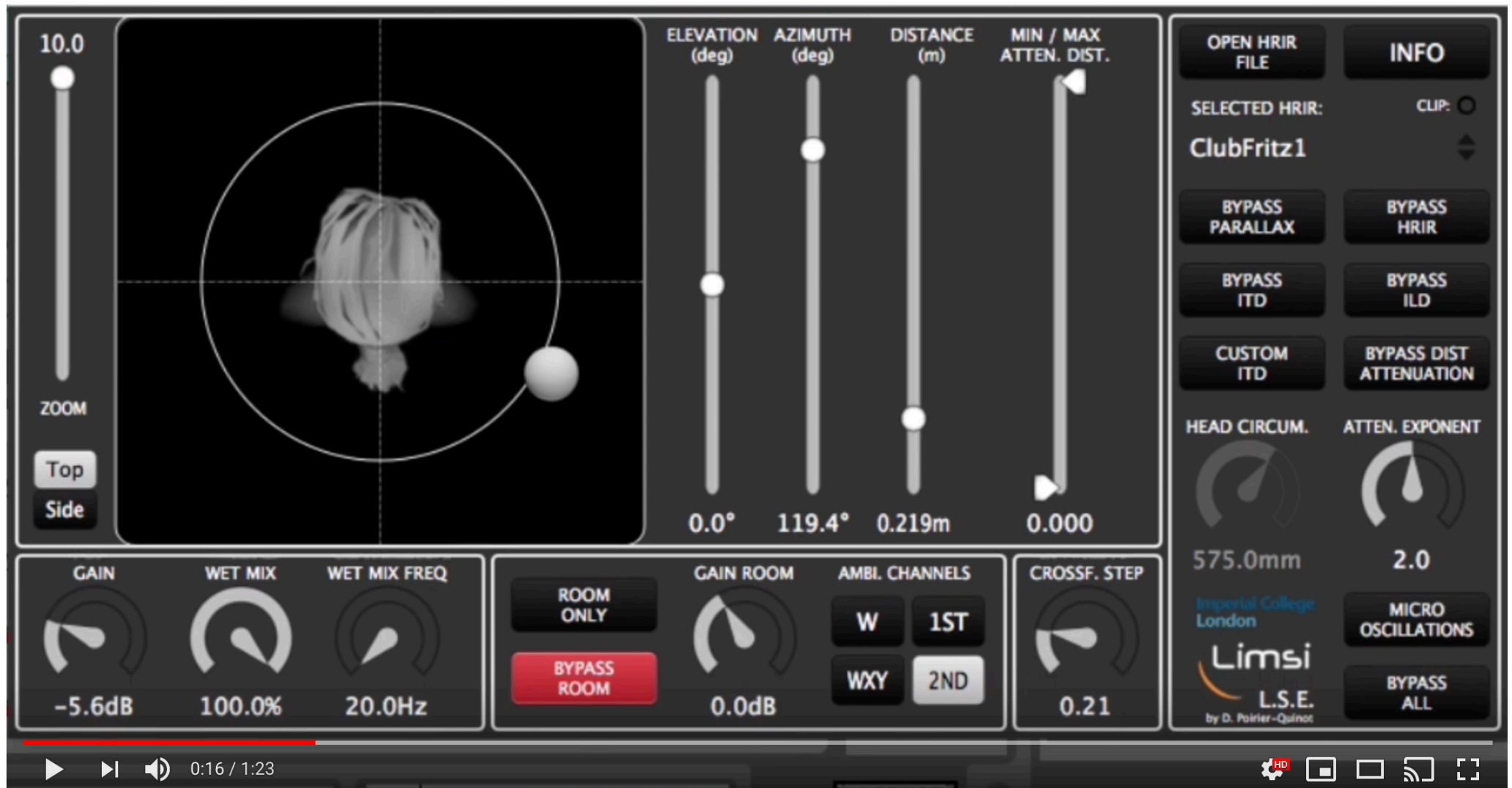
Anaglyph

Plugin VST d'effet de spatialisation Binaural

GRATUIT !

Bernard Lagnel
Novembre 2018

ANAGLYPH 0.9.2 beta



2016

David Poirier-Quinot a lancé **Anaglyph**, un effet de spatialisation binaurale gratuit de haute qualité (à usage non commercial) au format [plug-in VST](#) pour [stations](#) de [travail audionumériques](#) sur PC et Mac. Le développeur a récemment partagé le concept d'Anaglyph avec le public (la vidéo de démonstration présentée ci-dessous a été publiée en 2016), mais le plugin lui-même n'était pas disponible en téléchargement jusqu'à récemment. Ce n'est toujours pas une version stable officielle, car la version actuelle du plugin est la *0.9.2 beta*. Néanmoins, la version bêta actuellement disponible a très bien fonctionné pendant notre test, sans aucun problème de stabilité perceptible.

L'anaglyphe est un effet binaural avancé basé sur les résultats des recherches en cours sur l'audition spatiale. Je ne suis en aucun cas un expert dans ce domaine; votre meilleure solution pour comprendre les concepts et la science derrière le plugin serait donc de lire la documentation fournie sur la page du produit. En bref, cet outil gratuit vous permettra de simuler le positionnement en trois dimensions des sources audio dans votre mix en utilisant certaines des techniques les plus avancées actuellement disponibles.

L'interface utilisateur peut sembler un peu chargée au début, mais les paramètres de contrôle les plus importants sont faciles à identifier et à utiliser. Le pavé tactile XY situé à gauche contrôle le positionnement de base de la source audio, tandis que des ajustements supplémentaires, tels que la hauteur et le zoom, sont effectués à l'aide des curseurs verticaux à gauche et à droite. Malheureusement, aucun manuel de référence détaillé n'est disponible pour le moment. Vous pouvez consulter la vidéo de démonstration ci-dessous pour en savoir plus sur le schéma de contrôle du plugin.

La version actuelle est compatible avec tous les hôtes de plug-ins VST 64 bits sur PC et Mac (les stations de travail audio numériques 32 bits ne sont pas prises en charge pour le moment sur PC). Une version Linux du plugin est actuellement en préparation et devrait être publiée prochainement. Le plugin est publié en tant que freeware pour une utilisation non commerciale. Le développeur suggère aux utilisateurs de les contacter pour discuter des conditions de licence pour une utilisation commerciale potentielle. Un grand merci à notre lecteur, James, qui a écrit pour nous faire connaître Anaglyph.

Anaglyph est disponible en téléchargement gratuit sur [le site Web de David Poirier-Quinot](#) (*taille de téléchargement de 25,5 Mo, archive ZIP, format de plug-in VST 32 bits et 64 bits pour Windows, macOS, Linux*) .

Mise à jour du moteur de spatialisation binaurale Anaglyph VST vers la version 0.9.2c 29 octobre 2018

Le plug-in d'effet de spatialisation binaural de haute qualité [Anaglyph VST](#) récemment publié a été mis à jour vers la version 0.9.2c pour inclure la prise en charge de FL Studio et certains correctifs.

Après la sortie publique de l'Anaglyph VST à l'AES Milan en mai et notre récente exposition (grâce à [rekkerd.org](#), [bedroomproducersblog.com](#) et [www.synthtopia.com](#)), nous avons rapidement constaté un intérêt croissant pour Anaglyph. C'est très excitant pour nous. Au moment de la rédaction de cet article, nous avons déjà plus de 2300 téléchargements. Avec autant de personnes testant rapidement le VST sur autant de plates-formes, nous avons bien entendu été informés de quelques bugs inconnus jusqu'à présent.

Donc, pour commencer, nous sommes déjà prêts à annoncer une mise à jour * critique *, version v0.9.2c (<http://anaglyph.dalembert.upmc.fr/index.html#download>). Cette mise à jour inclut le support de FL Studio et des corrections importantes concernant la convolution binaurale. Veuillez mettre à jour votre version actuelle immédiatement.

La mise à jour est maintenant disponible au téléchargement pour Windows et macOS.

Les développements à venir comprennent la prise en charge du format de plug-in Pro Tools et Audio Unit, l'optimisation de l'UC, la réduction des temps de chargement des plug-ins et la résolution d'un problème lié à Reaper DAW sur les ordinateurs Windows.

Plus d'informations : [Anaglyphe](#)

Traduction Google

ANAGLYPH version v0.9.2c

6.1

ZOOM

Top

Rear

ELEVATION (deg)

AZIMUTH (deg)

DISTANCE (m)

MIN / MAX ATTEN. DIST.

0.0°

40.4°

0.280m

0.1, 10

PERSONALIZATION

CURRENTLY ACTIVE HRIR:

listen_irc_1008

CUSTOM ITD

HEAD CIRCUM.

579.4mm

NEAR FIELD ENHANCEMENTS

BYPASS PARALLAX

BYPASS ILD

DISTANCE ATTENUATION

BYPASS DIST ATTENUATION

ATTEN. EXPONENT

1.0

MICRO OSCILLATIONS

INFO

INPUT CHANNEL MAPPING:

Stereo Downmix

BYPASS ALL

CLIP

GAIN

0.0dB

WET MIX

100.0%

STUDIO ONLY

BYPASS STUDIO

GAIN STUDIO

0.0dB

H 0.0dB

M 0.0dB

L 0.0dB

SPATIAL DETAIL

OMNI

2D

3D 1ST

3D 2ND

CURRENTLY ACTIVE STUDIO:

room_02_studio_dry

Anaglyph

<https://github.com/polarch/JSAmbisonics>

<http://anaglyph.dalembert.upmc.fr/>

10 / 2018

===== ANAGLYPH =====

Anaglyph fait partie d'un effort de recherche en cours sur les capacités perceptuelles et techniques du rendu binaural. Anaglyph Audio Binaural Engine est un plug-in audio autonome / VST pour la spatialisation binaurale intégrant les résultats de près de deux décennies de recherche sur l'audition spatiale. Anaglyph a été conçu comme un plug-in audio, à la fois pour soutenir les efforts de recherche en cours et pour rendre accessible le fruit de cette recherche aux ingénieurs du son par le biais des environnements DAW traditionnels existants. Parmi ses fonctionnalités, Anaglyph inclut un modèle ITD morphologique personnalisable, des corrections ILD en champ proche et une sélection d'HRTF, un amplificateur de localisation, un amplificateur d'externalisation et un support de fichier SOFA HRIR.

DESCRIPTION DES MODULES

Personnalisation HRIR / HRTF: Personnalisation des paramètres de rendu binauraux .
Sélectionnez ici le HRTF spécifique (oreilles) en choisissant un fichier HRTF / SOFA approprié.
Adapter la taille de la tête pour optimiser les repères latéraux.
Améliorations du champ proche : Améliorations du rendu qui améliorent la sensation de distance pour les positions de source proches.
La correction ILD modifie les niveaux droit / gauche, en fonction de la fréquence, pour tenir compte de l'effet d'ombre des objets sonores à proximité.
Externalisation Booster : Lors du rendu de sons au casque, un artefact commun est que les sons semblent être très proches, voire même à l'intérieur de la tête des auditeurs. Cela ne se produit généralement pas lors de l'écoute sur des enceintes de contrôle dans un studio. Ce module ajoute les indices acoustiques subtils d'un système de reproduction dans un studio à contrôle acoustique, offrant une externalisation améliorée. Cela ne doit pas être confondu avec un plug-in de réverbération.
Les niveaux GAIN, EQ et SPATIAL DETAIL peuvent être ajustés pour optimiser l'effet du paysage sonore donné afin d'améliorer la transparence tout en fournissant ces repères supplémentaires.
Stabilisation Enhancer : Dans le cas d'un rendu binaural non suivi, les positions des sources statiques sont parfois difficiles à localiser correctement, des confusions avant / arrière se produisant souvent pour les sources dans certaines régions.

Pour réduire de telles confusions et aider à stabiliser la position perçue de l'objet sonore, la position rendue réelle varie légèrement, offrant une «redondance ajoutée» des signaux de localisation au cerveau, réduisant ainsi la confusion de localisation.

DESCRIPTIONS DES PARAMETRES DE CONTROLE (ordre alphabétique)

Remarque : utilisez les touches alt + clic ou double-clic pour réinitialiser les valeurs de paramètre des éléments de l'interface graphique.

AMBI. CANAUX : Définit le nombre de canaux Ambisonic utilisés pour la réverbération d'une pièce. W est omni, WXY est omni + gauche / droite avant / arrière, le 1er est Ambisonic de premier ordre (WXYZ avec Z: haut / bas), le 2ème est Ambisonic de second ordre.

ATTEN. EXPOSANT : Définir l'exposant d'atténuation de distance (d) utilisé:

(gain = $1 / d^{\text{ATTEN.EXPONENT}}$). (0 = 0 dB, d x 2 :::: 1 = - 6 dB, d x 2 :::: 2 = - 12 dB, d x 2)

AZIMUT : Azimut de la source sonore, en degrés. Contourner tous : Ignore tous les effets et génère un tampon d'entrée intact.

ATTENUATION DE DISTRIBUTION DE BYPASS : Désactiver le modèle d'atténuation de distance (gain = $1 / d^{\text{ATTEN.EXPONENT}}$). (0 = 0 dB, d x 2 :::: 1 = - 6 dB, d x 2 :::: 2 = - 12 dB, d x 2)

BYPASS HRIR : Filtrage de la réponse impulsionnelle liée à la tête de dérivation.

BYPASS ILD : Désactiver le modèle de différence de niveau interaural en champ proche.

BYPASS ITD : Contourner l'application du décalage horaire interaural sur les canaux gauche / droit.

BYPASS STUDIO : Désactiver l'effet Booster d'externalisation.

ITD PERSONNALISÉ : Personnalisez le DTI HRIR d'origine en le remplaçant par un DTI personnalisé basé sur HEAD CIRCUM.

modèle DISTANCE : Distance de la source sonore, en mètres.

ÉLÉVATION : Élévation de la source sonore, en degrés. 0 degré est directement devant, les angles positifs augmentent l'élévation.

ACTIVER PARALLAX : Activez la parallaxe HRIR, c'est-à-dire la sélection des HRIR en fonction de la direction réelle de l'arrivée de la source «vue» par chaque oreille plutôt que celle calculée par rapport à l'origine de la tête.

EQ L / M / H : Les boutons contrôlent les gains d'un égaliseur appliqué au booster d'externalisation STUDIO.

GAIN : Gain général du plug-in.

GAIN STUDIO : Gain appliqué à l'amplificateur d'externalisation STUDIO (distinct du codage binaural direct).

HEAD CIRCUM : Tour de tête, en mm, utilisé par le modèle ITD personnalisé lorsque CUSTOM ITD est activé.

CARTOGRAPHIE DES CANAUX D'ENTRÉE : Définit comment une entrée stéréo est utilisée par le plug-in. (Stéréo downmix, Left, right)

MICRO OSCILLATIONS : Activez l'amplificateur de stabilisation en imprimant de petites oscillations à la position de l'objet sonore.

MIN / MAX ATTEN. DIST : Définit les limites inférieure et supérieure du modèle d'atténuation de distance ($\text{gain} = 1 / (\text{distance})^{\text{ATTEN.EXPONENT}}$).

HRIR SÉLECTIONNÉ : HRIR actuel utilisé par le plug-in, sélectionné dans la base de données disponible.

STUDIO UNIQUEMENT : Couper le traitement binaural du son direct en ne laissant que l'effet STUDIO Externalization Booster (destiné au test).

WET MIX : Pourcentage d'effet de spatialisation appliqué

REMERCIEMENTS-----

Anaglyph a été développé dans le cadre de nombreux projets de recherche académique avec le CNRS et l'Université de la Sorbonne, avec la participation de la Dyson School of Engineering de l'Imperial College London (anciennement LSE). Le moteur audio binaural Anaglyph est publié en tant que logiciel gratuit à usage non commercial, sous licence CC BY-NC-ND 3.0. Plus d'informations peuvent être trouvées à <http://anaglyph.dalembert.upmc.fr>. Anaglyphe VST v0.9.2c

Traduction Google

AES ENGINEERING BRIEFS FORUM

Le moteur audio binaural anaglyphe

Anaglyph fait partie d'un effort de recherche en cours sur les capacités perceptuelles et techniques du rendu binaural. Le moteur audio binaural Anaglyph est un plug-in audio VST pour la spatialisation binaurale intégrant les résultats de plus d'une décennie de recherche sur l'audition spatiale. Anaglyph a été conçu comme un plug-in audio, à la fois pour soutenir les efforts de recherche en cours et pour rendre accessible le fruit de cette recherche aux ingénieurs du son par le biais des environnements DAW traditionnels existants. Parmi ses fonctionnalités, Anaglyph inclut un modèle ITD morphologique personnalisable, des corrections de parallaxe ILD et HRTF en champ proche, un amplificateur de localisation, un amplificateur d'externalisation et un support de fichier SOFA HRIR. L'architecture de base et la mise en œuvre de chaque composant lié à l'audio sont présentées ici.

Auteurs: Poirier-Quinot, David; Katz, Brian FG

Affiliation: Sorbonne Université, CNRS, Paris, France

Convention AES: 144 (mai 2018)

Compte rendu: 431

Date de publication: le 14 mai 2018

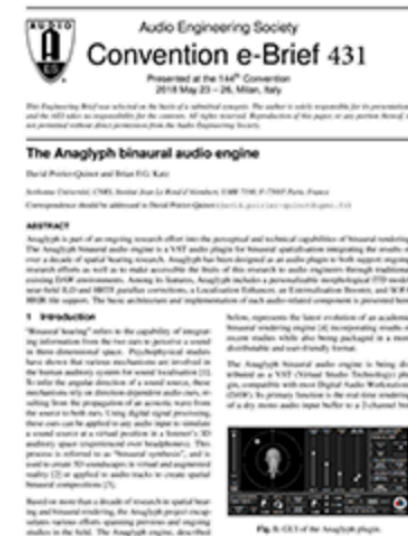
Sujet: Spatial Audio



Télécharger maintenant (266 KB)

Ce document est **Open Access**, ce qui signifie que vous pouvez le **télécharger gratuitement**.

Aucun membre de l'AES n'a encore commenté ce document.





Audio Engineering Society Convention e-Brief 431

Presented at the 144th Convention
2018 May 23 – 26, Milan, Italy

This Engineering Brief was selected on the basis of a submitted synopsis. The author is solely responsible for its presentation, and the AES takes no responsibility for the contents. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Audio Engineering Society.

The Anaglyph binaural audio engine

David Poirier-Quinot and Brian F.G. Katz

Sorbonne Université, CNRS, Institut Jean Le Rond d'Alembert, UMR 7190, F-75005 Paris, France

Correspondence should be addressed to David Poirier-Quinot (david.poirier-quinot@upmc.fr)

ABSTRACT

Anaglyph is part of an ongoing research effort into the perceptual and technical capabilities of binaural rendering. The Anaglyph binaural audio engine is a VST audio plugin for binaural spatialisation integrating the results of over a decade of spatial hearing research. Anaglyph has been designed as an audio plugin to both support ongoing research efforts as well as to make accessible the fruits of this research to audio engineers through traditional existing DAW environments. Among its features, Anaglyph includes a personalizable morphological ITD model, near-field ILD and HRTF parallax corrections, a Localisation Enhancer, an Externalisation Booster, and SOFA HRIR file support. The basic architecture and implementation of each audio-related component is presented here.

1 Introduction

“Binaural hearing” refers to the capability of integrating information from the two ears to perceive a sound in three-dimensional space. Psychophysical studies have shown that various mechanisms are involved in the human auditory system for sound localisation [1]. To infer the angular direction of a sound source, these mechanisms rely on direction-dependent audio cues, resulting from the propagation of an acoustic wave from the source to both ears. Using digital signal processing, these cues can be applied to any audio input to simulate a sound source at a virtual position in a listener’s 3D auditory space (experienced over headphones). This process is referred to as “binaural synthesis”, and is used to create 3D soundscapes in virtual and augmented reality [2] or applied to audio tracks to create spatial binaural compositions [3].

Based on more than a decade of research in spatial hearing and binaural rendering, the Anaglyph project encapsulates various efforts spanning previous and ongoing studies in the field. The Anaglyph engine, described

below, represents the latest evolution of an academic binaural rendering engine [4] incorporating results of recent studies while also being packaged in a more distributable and user-friendly format.

The Anaglyph binaural audio engine is being distributed as a VST (Virtual Studio Technology) plugin, compatible with most Digital Audio Workstations (DAW). Its primary function is the real-time rendering of a dry mono audio input buffer to a 2-channel bin-

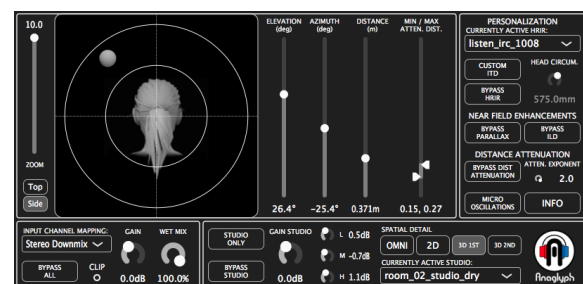


Fig. 1: GUI of the Anaglyph plugin.

aural stereo audio output. A cross between a research platform and a professional solution, Anaglyph is conceived as a proof of concept test for state-of-the-art spatialisation techniques to be compared against production environment expectations. The Anaglyph VST plugin is made available as a freeware for non-commercial use, and can be downloaded from the project webpage: anaglyph.dalembert.upmc.fr.

Anaglyph is implemented in C++, using the JUCE framework for cross-platform compilation, Graphical User Interface (GUI) design, audio classes (buffers, filters, etc.), and standard VST features (automations, presets, I/O management, etc.). The plugin GUI is illustrated in Figure 1. The architecture of the plugin is described in the next section while a detailed description of its main components is given in Section 3.

2 General Architecture

The overall architecture of the Anaglyph plugin is shown in Figure 2. The input audio buffer is encoded by the binaural module into a stereo buffer. The binaural module is composed of a Head Related Impulse Response (HRIR) convolution engine, an Interaural Time Difference (ITD) modifier, and an Interaural Level Difference (ILD) modifier. ITD, ILD, and HRIR respectively represent the time, amplitude, and spectral cues resulting from the propagation of a sound wave from a point source, travelling around the listener's head, arriving at the ear canal entrance.

Parallel to the binaural encoding, the input audio is fed to an Externalization Booster, based on a modified

3D room reverberation simulator [5]. The Finite Impulse Response (FIR) filter is loaded in the form of a 2nd order Ambisonic Room Impulse Response (RIR). The resulting Ambisonic buffer is decoded using a virtual speaker approach [6] and rendered binaurally to a stereo buffer, which is then combined with the output of the direct sound binaural encoder.

3 Detailed Components

This section details the main components of the Anaglyph plugin.

3.1 Binaural Encoding

Whenever a new position is defined, through DAW automation lines or via the plugin's GUI, the convolution engine converts the pair of HRIR filters returned by the MySofa library into FIR filters, applying both right and left Impulse Responses (IR) in the frequency domain to reduce CPU consumption. The resulting stereo buffer is passed to the ITD modifier. This module adjusts the delay of the right or left input buffer to correspond to the estimated personalized ITD of the user relative to the ITD of the selected HRIR dataset [7, 8] for the current position. The user's head circumference is used to provide a reliable ITD model to personalize the ITD values [9]. If disabled, the original HRIR's ITD is used.

3.2 Distance Attenuation

The default distance attenuation model is based on the inverse square law. The law exponent can be modified, as can be both start and stop attenuation distances.

3.2.1 HRIR Near-Field Parallax

Due to technical constraints, most HRIR sets are measured for a unique source to head distance, providing an HRTF for the surface of given sphere. Anaglyph implements a "parallax HRIR correction" option to compensate for variable distances without the need for multiple HRIR sets at different distances by simulating the impact of source distance on the HRIR [10]. The parallax correction is illustrated in Figure 3. The correction affects source positions close to the subject compared to the HRIR measurement ring. The ring radius value is set based on the measurement distance indicated in the loaded SOFA HRIR.

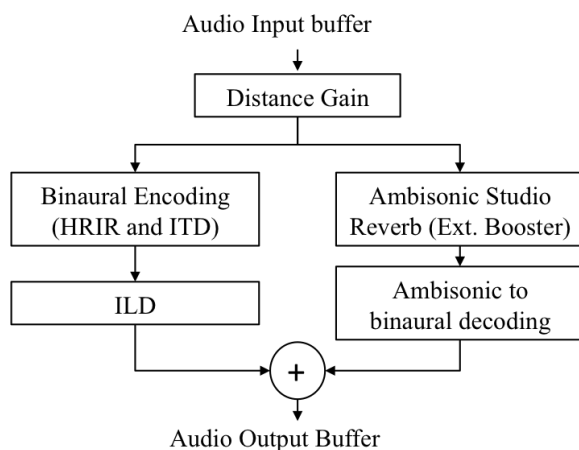


Fig. 2: Simplified architecture of the Anaglyph plugin.

3.3 ILD Near-Field Model

After binaural encoding, the stereo buffer is passed to the ILD modifier. The ILD modifier applies a frequency-dependant adjustment to both left and right buffers, using on a pair of cascaded biquad filters. The ILD modifier is applied for the current audio source position (azimuth, elevation, distance), reproducing ILD variations (via a spherical head model [11]).

3.4 Externalization Booster

In essence, the externalization booster is an Ambisonic room reverberation module with the direct sound removed, as this is processed by the binaural encoding module. As the name suggests, this module is used to improve source externalization, separate from any artificial reverberation added for artistic goals. Based on FIR convolution, this ‘reverb’ is applied on the mono input buffer, parallel to the binaural encoding. The FIR is interpolated from a set of 2nd order Ambisonic RIRs. A spherical array of source positions surrounding the listener allows the convolution reverberator to provide

coherent early reflection patterns (in time and space) with respect to the actual source direction used in the Binaural Encoding module. Such reflection patterns help support the sense that the source is located in a space away from the listener while also reinforcing the spatial cues for source localization.

The resulting Ambisonic buffer is decoded based on a virtual speaker approach [6] prior to summation with the binaural buffer. The virtual speakers used for Ambisonic to binaural decoding represent a 12 channel array, uniformly distributed on a sphere.

3.5 SOFA Loader

Upon startup, the plugin loads HRIR, ITD, ILD, and room IR SOFA files [12] using the MySofa C++ library [13]. Prior to loading the SOFA HRIR, a Matlab utility script is used for time-alignment of the HRIR while the ITD value is stored in the “Delay” section of the SOFA file. The MySofa library internal routines are used for linear interpolation between neighbouring IRs, (HRIR, room IR), time (ITD), and biquad (ILD) values.

3.6 Localisation Enhancer

The localisation enhancer module applies a subtle figure-of-8 movement to the audio source in the azimuth-elevation plane (fixed distance). The radius of these micro-oscillation can be modified, set to a default 1 degree amplitude in both dimensions. This oscillation is designed to compensate for the absence of listener’s head micro-movements in non-tracked conditions, proved to improve localisation accuracy [14, 15]. The optimization of the control parameters and the quantification of the effect through formal studies of this module is currently the subject of study.

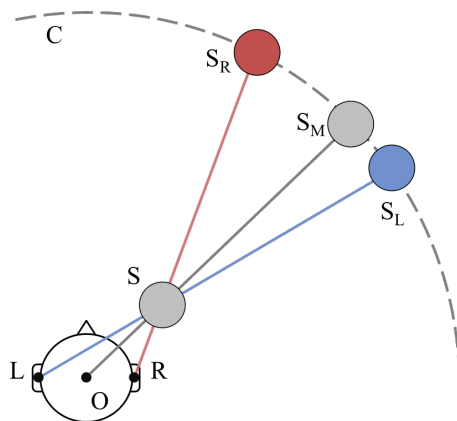


Fig. 3: Illustration of the parallax HRIR correction principle. **O** is the center of the subject’s head, **L** and **R** are the positions of the subject’s left and right ears respectively. **S** is the position of the audio source as defined in the plugin GUI, **C** represents the HRIR measurement ring, and **S_M** the projection of **S** on **C**. **S_M** is the HRIR position that would be selected without the parallax correction option enabled. **S_L** and **S_R** are the positions of the HRIRs actually selected for the left ear and right ear respectively when the parallax correction is enabled.

4 Conclusion

Anaglyph, and its previous incarnations, have been used in numerous research projects [16, 17, 18, 19, 20] as well as several production tests involving spatialised music¹, radio drama fiction², and immersive video game audio³. Anaglyph is currently being use by Bloober Team in several upcoming suspense/horror game projects. Anaglyph will be updated based on results of the authors and peers in binaural spatialisation.

¹www.youtube.com/watch?v=R9Vx4PaElZA

²www.youtube.com/watch?v=V6N-6DxAVQw

³www.youtube.com/watch?v=q6muds1qW-w

5 Acknowledgements

Some upstream development support was provided through funding from the European Union's Horizon 2020 program under grant agreement No 644051.

References

- [1] Blauert, J., *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT press, 1997.
- [2] Begault, D. R. and Trejo, L. J., *3-D Sound for Virtual Reality and Multimedia*, Academic Press, 2000.
- [3] Rumsey, F., *Spatial Audio*, CRC Press, 2012.
- [4] Katz, B., Rio, E., and Picinali, L., "LIMSI Spatialisation Engine," Intl Deposit Digital Number: IDDN.FR.001.340014.000.S.P.2010.000.31235., 2010.
- [5] Guastavino, C. and Katz, B., "Perceptual evaluation of multi-dimensional spatial audio reproduction," *J Acous Soc of Am*, 116(2), pp. 1105–1115, 2004.
- [6] Noisternig, M., Musil, T., Sontacchi, A., and Holdrich, R., "3D binaural sound reproduction using a virtual Ambisonic approach," in *Virtual Environments, Human-Computer Interfaces and Measurement Systems*, pp. 174–178, IEEE, 2003.
- [7] Andreopoulou, A. and Katz, B., "Identifying a perceptually relevant estimation method of the inter-aural time delay," *J Acous Soc of Am*, 141(5), pp. 3635–3635, 2017.
- [8] Katz, B. and Noisternig, M., "A comparative study of interaural time delay estimation methods," *J Acous Soc of Am*, 135(6), pp. 3530–3540, 2014.
- [9] Aussal, M., Alouges, F., and Katz, B., "HRTF interpolation and ITD personalization for binaural synthesis using spherical harmonics," in *Aud Eng Soc Conf* 25, 2012.
- [10] Otani, M., Hirahara, T., and Ise, S., "Numerical study on source-distance dependency of head-related transfer functions," *J Acous Soc of Am*, 125(5), pp. 3253–3261, 2009.
- [11] Duda, R. O. and Martens, W. L., "Range dependence of the response of a spherical head model," *J Acous Soc of Am*, 104(5), pp. 3048–3058, 1998.
- [12] Majdak, P., Iwaya, Y., Carpentier, T., Nicol, R., Parmentier, M., Roginska, A., Suzuki, Y., Watanabe, K., Wierstorf, H., Ziegelwanger, H., et al., "Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions," in *Aud Eng Soc Conv* 134, 2013.
- [13] Hoene, C., Patino Mejia, I. C., and Cacerovschi, A., "MySofa – Design Your Personal HRTF," in *Aud Eng Soc Conv* 142, 2017.
- [14] Nykänen, A., Zedigh, A., and Mohlin, P., "Effects on localization performance from moving the sources in binaural reproductions," in *Intl Cong and Exposition on Noise Control Eng*, volume 4, pp. 3193–3201, 2013.
- [15] Hendrickx, E., Stitt, P., Messonnier, J.-C., Lyzwa, J.-M., Katz, B., and De Boishéraud, C., "Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis," *J Acous Soc of Am*, 141(3), pp. 2011–2023, 2017.
- [16] Katz, B., Rio, E., Picinali, L., and Warusfel, O., "The effect of spatialization in a data sonification exploration task," in *Intl Conf on Auditory Display*, pp. 1–7, Paris, 2008.
- [17] Katz, B. F., Kammoun, S., Parseihian, G., Gutierrez, O., Brilhault, A., Auvray, M., Truillet, P., Denis, M., Thorpe, S., and Jouffrais, C., "NAVIG: augmented reality guidance system for the visually impaired," *Virtual Reality*, 16(4), pp. 253–269, 2012, doi:10.1007/s10055-012-0213-6.
- [18] Parseihian, G. and Katz, B., "Rapid head-related transfer function adaptation using a virtual auditory environment," *J Acous Soc of Am*, 131(4), pp. 2948–2957, 2012.
- [19] Parseihian, G., Jouffrais, C., and Katz, B. F., "Reaching Nearby Sources: Comparison Between Real and Virtual Sound and Visual Targets," *Frontiers in Neuroscience*, 8, pp. 269:1–13, 2014, doi:10.3389/fnins.2014.00269.
- [20] Menelas, B.-A. J., Bourdot, P., Picinali, L., and Katz, B. F. G., "Non-visual identification, localization, and selection of entities of interest in a 3D environment," *J. Multimodal User Interfaces*, pp. 1–14, 2014, doi:10.1007/s12193-014-0148-1.

Mémoire de fin d'études
École Nationale Supérieure Louis Lumière
**Adaptation rapide à une spatialisation
binaurale non-individualisée : cas de la
réalité virtuelle appliquée au jeu vidéo**

Guillaume ANDRE
Directeur interne : Alan BLUM
Directeur externe : Renaud SÉGUIER
Rapporteurs : Laurent MILLOT et Éric URBAIN

promotion SON 2017

3.4.2 Phase de sélection d'HRIR








Les HRIR sélectionnées pour cette première partie du test sont issues de la base de données LISTEN de l'IRCAM, il s'agit des 7 golden ears, des HRIR qui sont supposées couvrir un panel de morphologie suffisamment large pour qu'on en trouve une qui corresponde à n'importe quel sujet. Il s'agit des réponses *irc 1008, irc 1013, irc 022, irc 1031, irc 1032, irc 1048 et irc 1053*, qui sont toutes disponibles au format SOFA sur le site internet sofaconventions.org. Le choix de cette banque parmi d'autres se justifie par la présence de l'étude sur les golden ears² d'une part, mais également par la citation de cette banque de données comme une de celles utilisables dans le manuel d'Oculus.³

2. Schönstein D., Katz B., "Sélection de HRTF dans une Base de Données en Utilisant des Paramètres Morphologiques pour la Synthèse Binaurale.", Société Française d'Acoustique, 10ème Congrès Français d'Acoustique, Lyon, France. 2010


3. *Oculus Audio SDK Documentation*, manuel à destination des développeurs pour la plateforme Oculus, v1.1.4, p12, 2017

Les 7 HRTF disponibles de la base LISTEN IRCAM

:

	listen_irc_1008.sofa	27/10/2018 19:35	1,1 Mo	Document
	listen_irc_1013.sofa	27/10/2018 19:35	984 Ko	Document
	listen_irc_1022.sofa	27/10/2018 19:35	1,2 Mo	Document
	listen_irc_1031.sofa	27/10/2018 19:35	1 Mo	Document
	listen_irc_1032.sofa	27/10/2018 19:35	886 Ko	Document
	listen_irc_1048.sofa	27/10/2018 19:35	918 Ko	Document
	listen_irc_1053.sofa	27/10/2018 19:35	867 Ko	Document

MIN / MAX
ATTEN. DIST.



1, 8

PERSONALIZATION

CURRENTLY ACTIVE HRIR:

listen_irc_1053

- listen_irc_1008
- listen_irc_1013
- listen_irc_1022
- listen_irc_1031
- listen_irc_1032
- listen_irc_1048
- ✓ listen_irc_1053

BYPASS DIST ATTENUATION

MICRO OSCILLATIONS


INFO

ATIAL DETAIL

DMNI 2D 3D 1ST 3D 2ND

CURRENTLY ACTIVE STUDIO:

oom_02_studio_dry



Configuration des mesures D'HRTF

Système de Reproduction ANAGLYPH version v0.9.2c :

The screenshot shows the ANAGLYPH v0.9.2c software interface. The main window is divided into several sections:

- Top Left:** A vertical slider labeled 'ZOOM' with a value of 0.3. Below it are buttons for 'Top' and 'Rear' views.
- Top Center:** A circular HRTF plot with a red arrow indicating a rotation. The plot is labeled 'FRONT' at the top and 'BACK' at the bottom. A yellow 'L' is visible on the left side.
- Top Right:** Four vertical sliders for 'ELEVATION (deg)', 'AZIMUTH (deg)', 'DISTANCE (m)', and 'MIN / MAX ATTEN. DIST.'. The 'DISTANCE' slider is set to 2.000m, which is circled in red. The 'MIN / MAX ATTEN. DIST.' slider is set to 1, 8.
- Right Panel:** A 'PERSONALIZATION' section with a dropdown menu showing 'listen_irc_1008' (circled in red). Below it are buttons for 'CUSTOM ITD', 'BYPASS HRIR', 'HEAD CIRCUM.' (set to 526.3mm), 'NEAR FIELD ENHANCEMENTS' (with 'BYPASS PARALLAX' and 'BYPASS ILD' buttons), 'DISTANCE ATTENUATION' (with 'BYPASS DIST ATTENUATION' and 'ATTEN. EXPONENT' set to 0.0), 'MICRO OSCILLATIONS', and an 'INFO' button.
- Bottom Left:** 'INPUT CHANNEL MAPPING' section with a dropdown set to 'Left', 'BYPASS ALL' button, 'CLIP' button, 'GAIN' slider (0.0dB), and 'WET MIX' slider (100.0%).
- Bottom Center:** 'STUDIO ONLY' button, 'GAIN STUDIO' slider (0.0dB), and 'BYPASS STUDIO' button.
- Bottom Right:** 'SPATIAL DETAIL' section with buttons for 'OMNI', '2D', '3D 1ST', and '3D 2ND'. Below it is a dropdown for 'CURRENTLY ACTIVE STUDIO' set to 'room_02_studio_dry'.

Système de Captation :

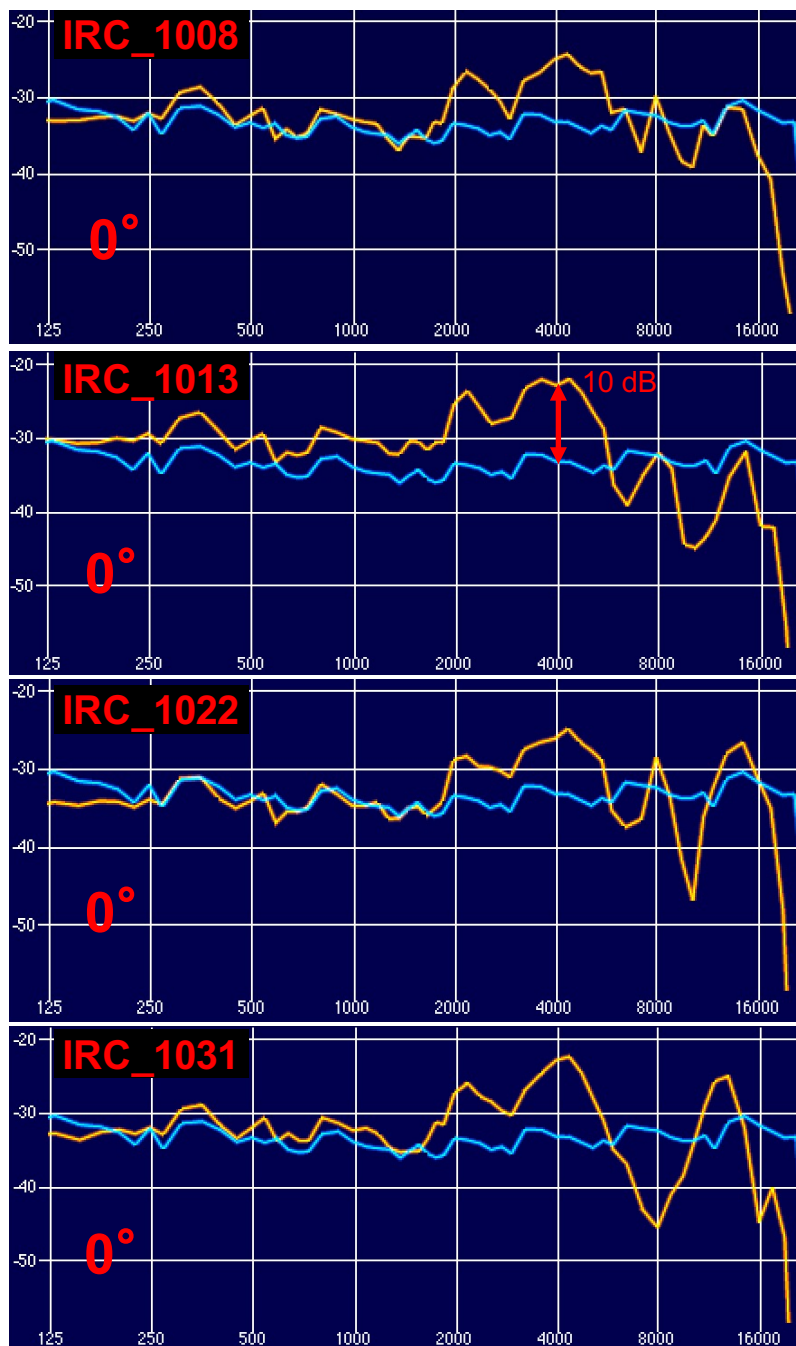
- LOGICIEL :
- NOM : Max / MSP Measurement Toolkit
- VERSION : 1.2
- MICROPHONE :
- TYPE : Knowles FG332 set 3
- NOTES :
- ADCONVERTER : RMC multiface
- DACONVERTER : Multiface RME

- AMPLIFICATEUR : Yamaha PC 1002
- HAUT - PARLEUR :
- LOUD_TYPE : Tannoy System 600
- LOUD_NUMBER : 292930P
- DISTANCE : 1,95
- MEASURING_TYPE : oreille bloquée
- MICROPHONE_DEPTH : 0.0
- SIGNAL_TYPE : Balayage logarithmique
- HEADTRACKER :
- INTERET : Pas de correction, juste la position mesurée

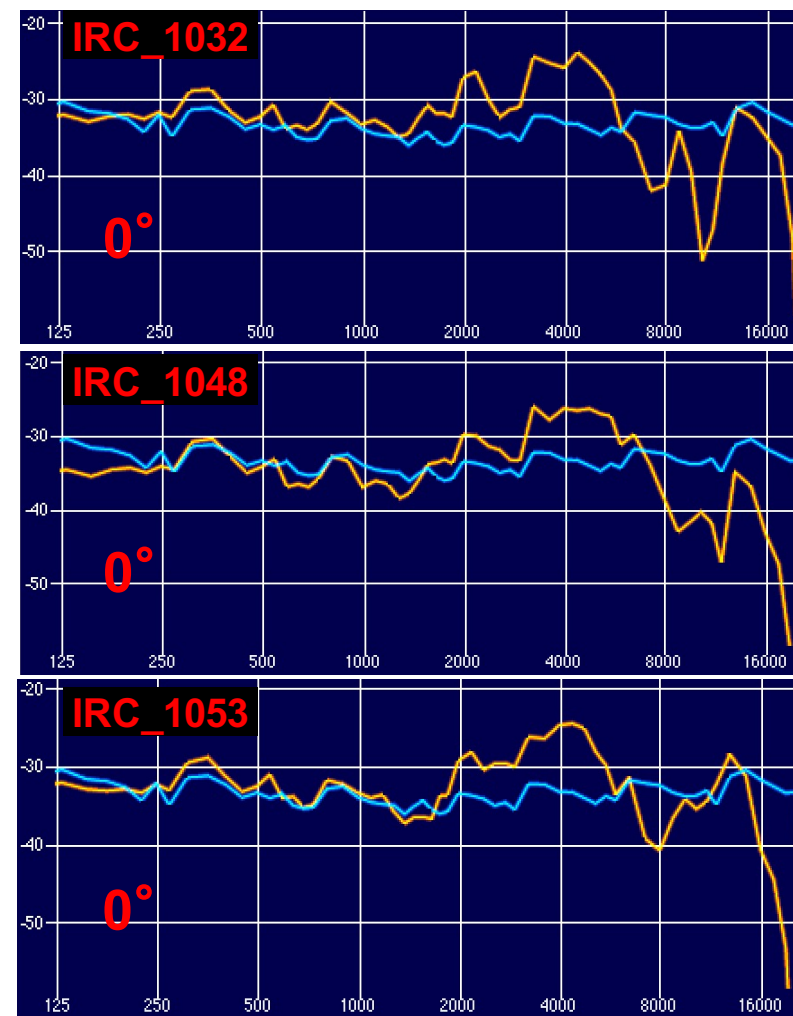
PIÈCE

- ID : salle anéchoïque
- DIMENSION :
- L : 5,9
- W : 4.4
- H : 4,25

7 HRTF : ANAGLYPH à 0° Front

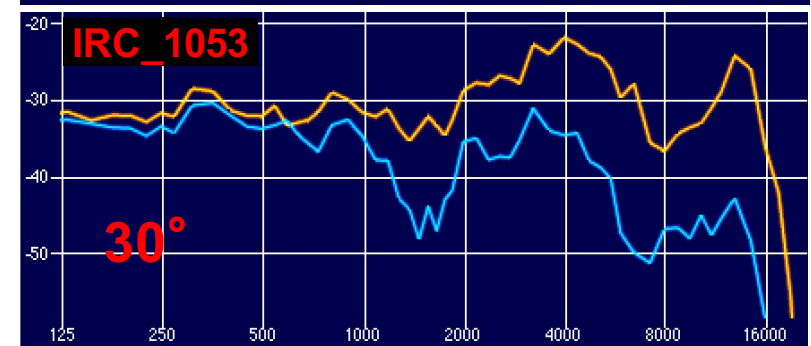
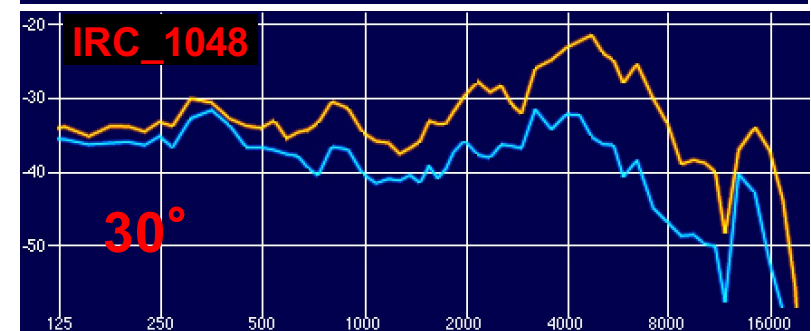
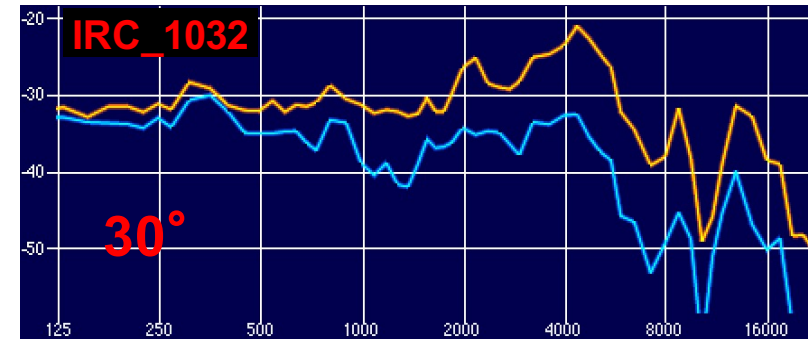
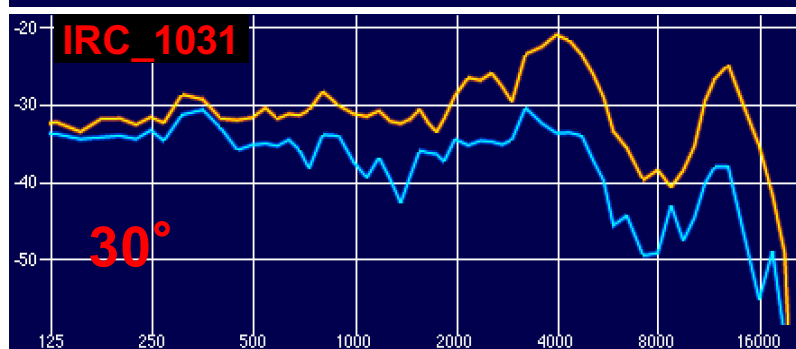
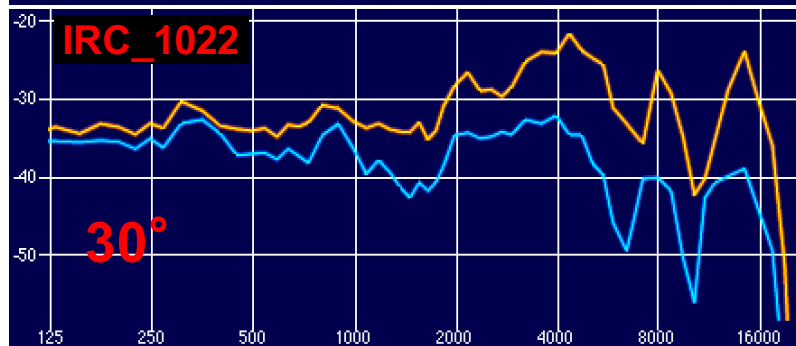
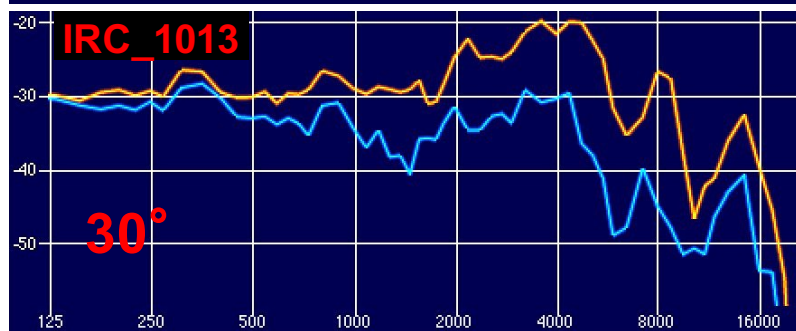
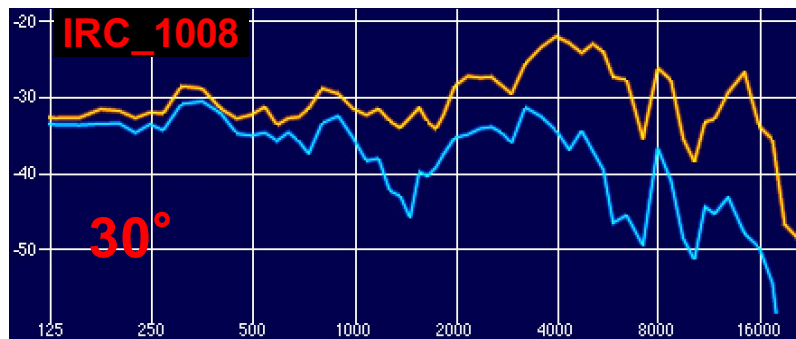


Pas de correction « Champ Diffus »...



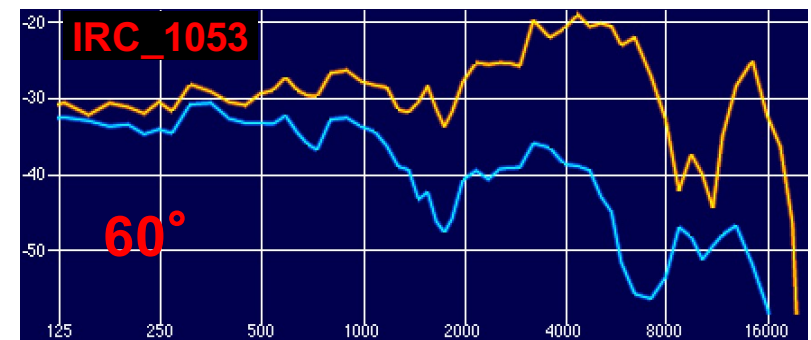
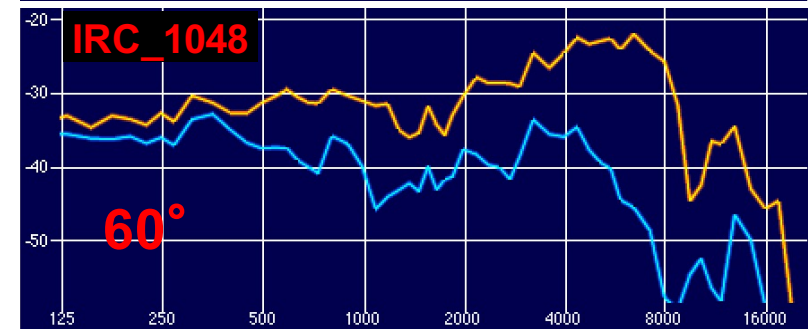
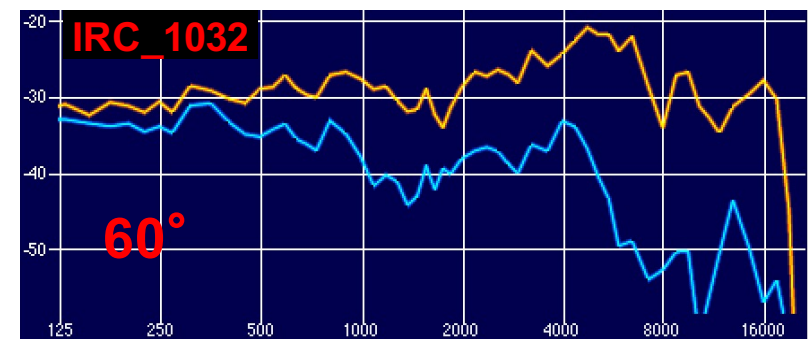
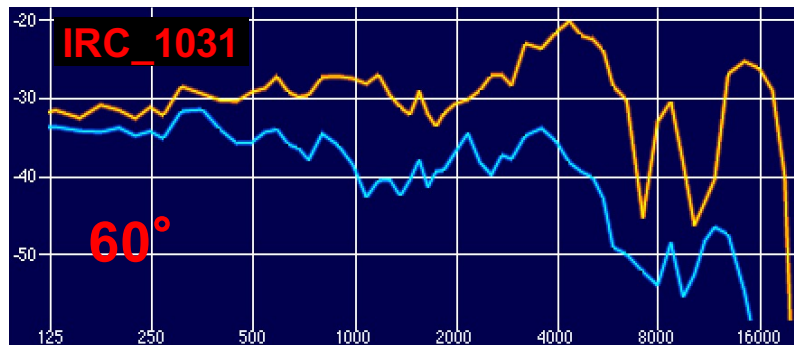
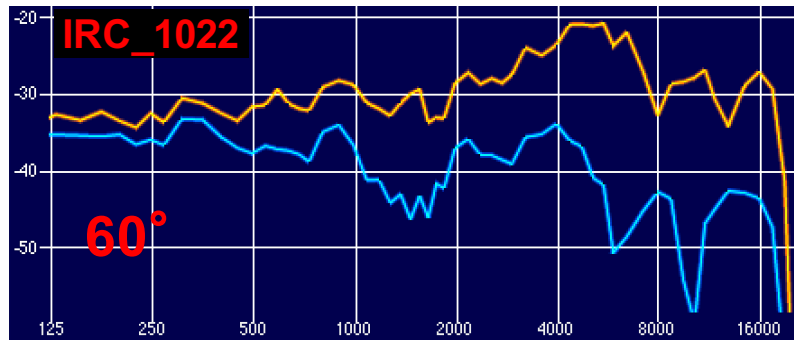
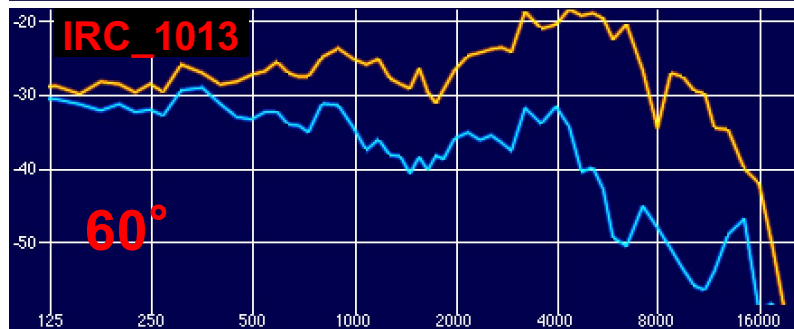
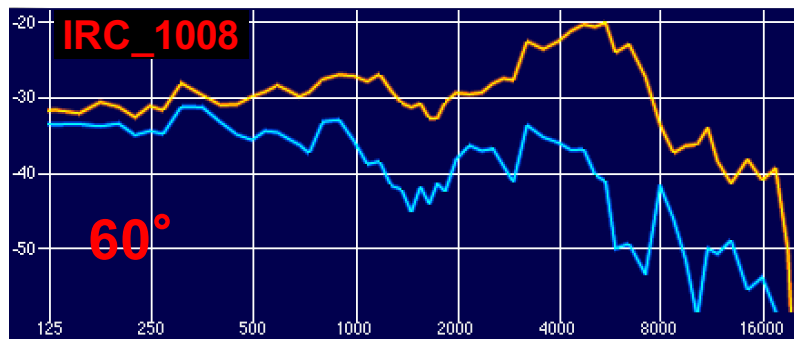
— Oreille Ipsilatéral L
— BRUIT ROSE IN

7 HRTF : ANAGLYPH à 30°



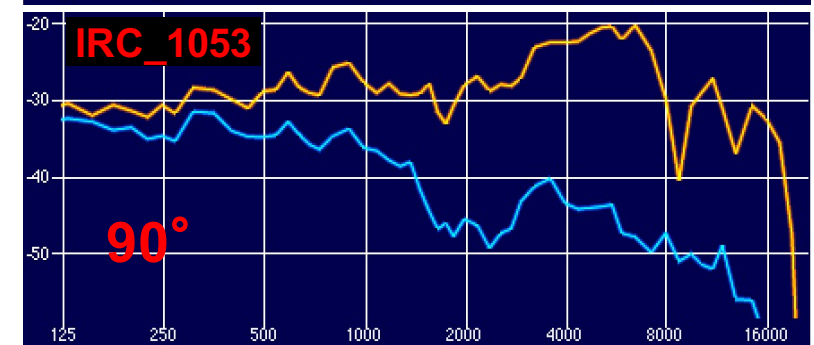
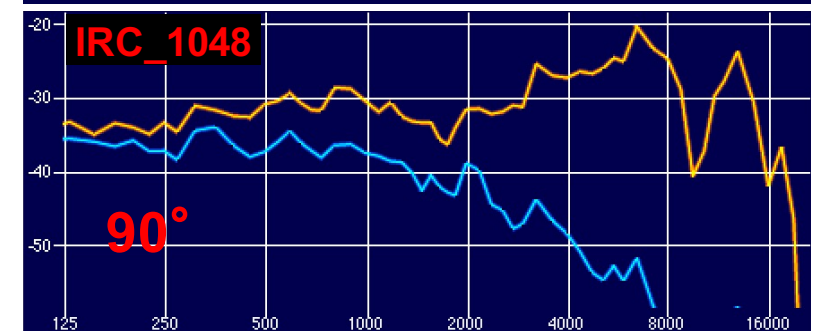
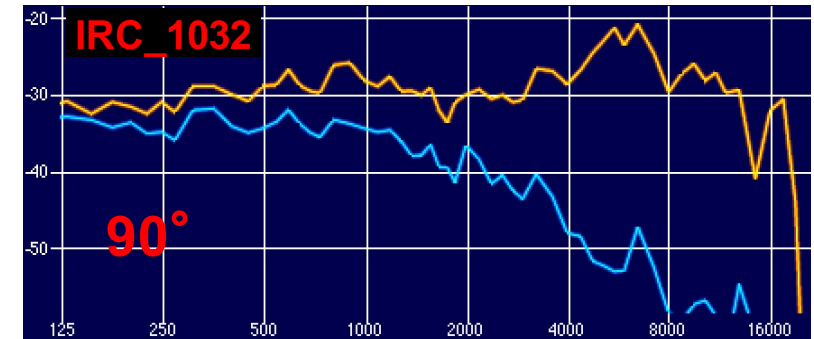
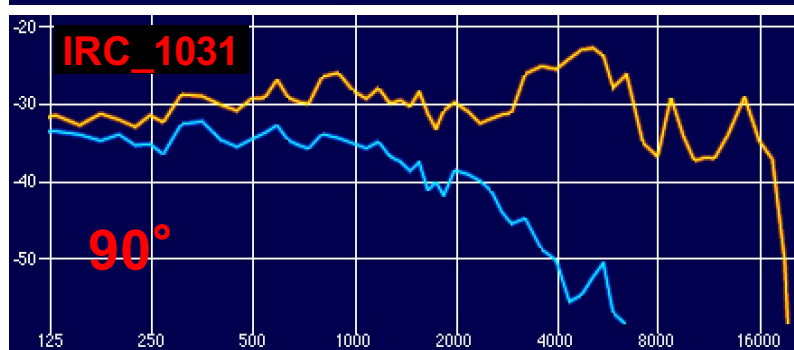
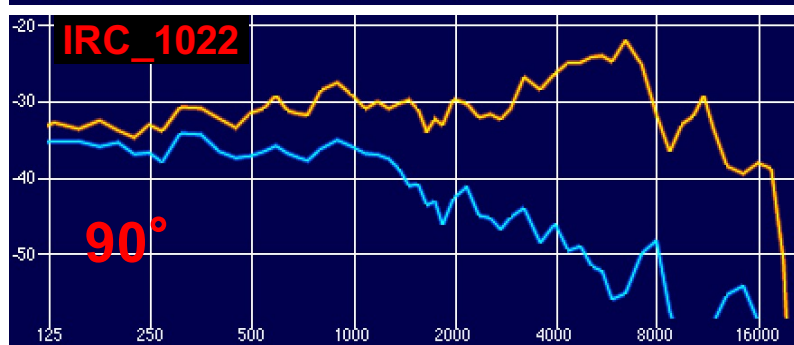
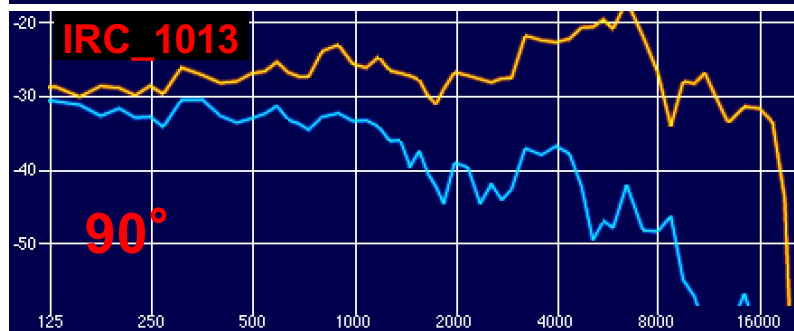
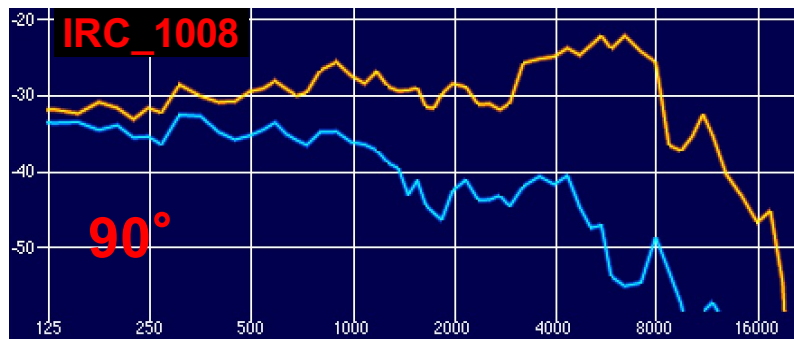
— Oreille Ipsilatéral **L**
— Oreille Contralatéral

7 HRTF : ANAGLYPH à 60°



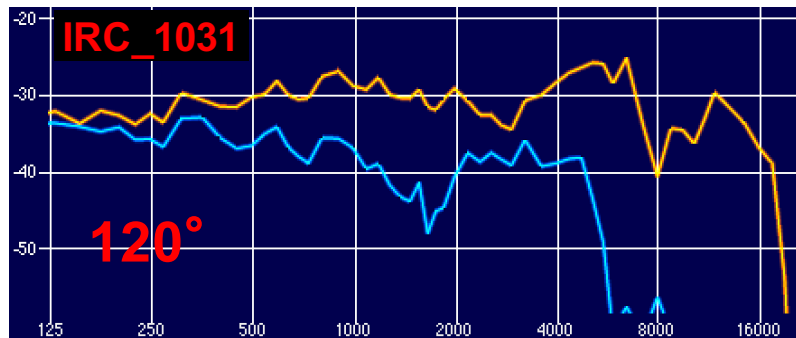
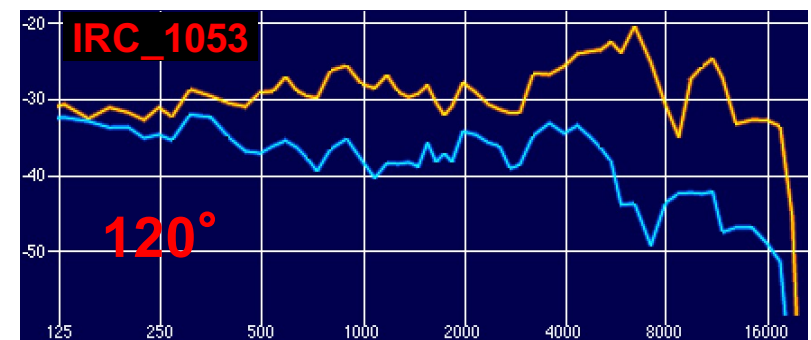
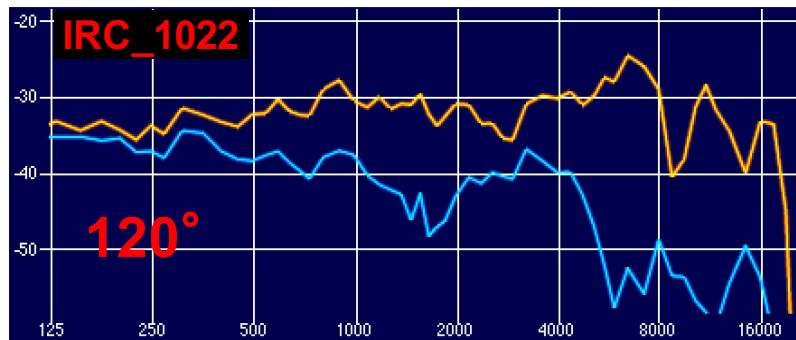
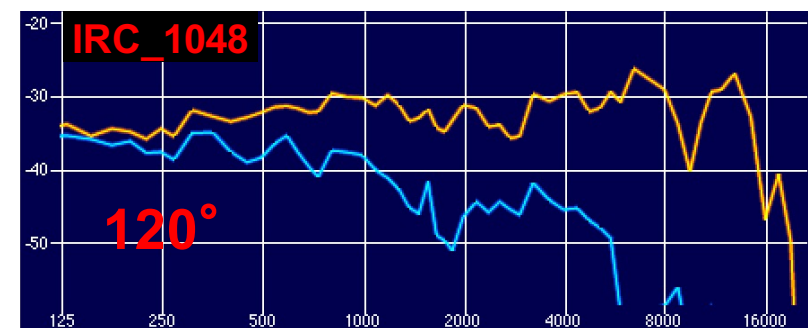
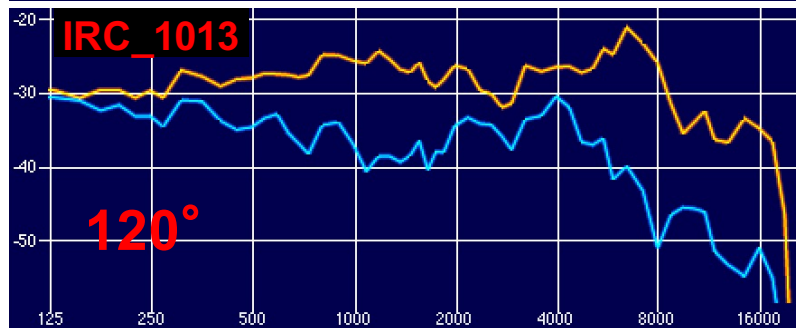
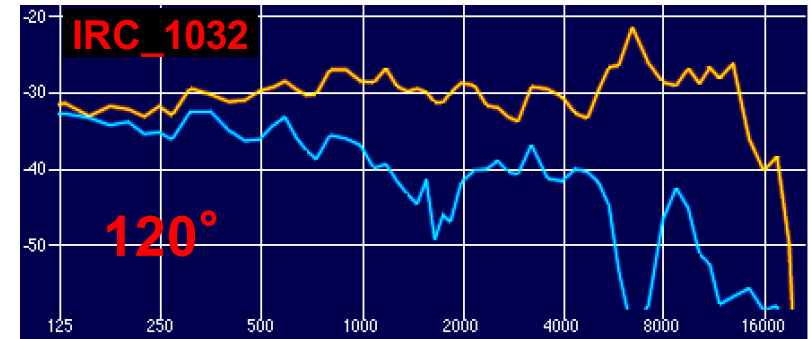
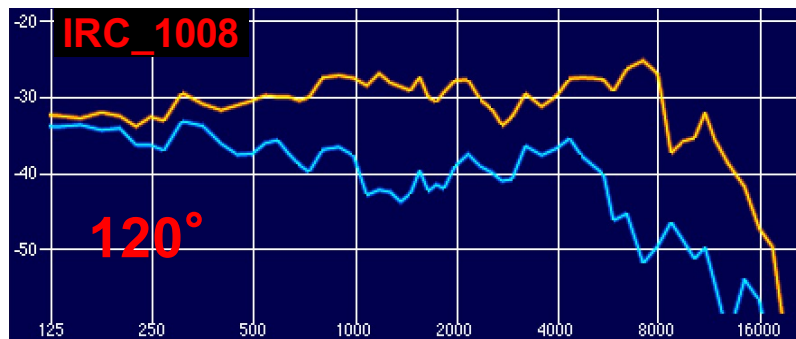
— Oreille Ipsilatéral L
— Oreille Contralatéral

7 HRTF : ANAGLYPH à 90°



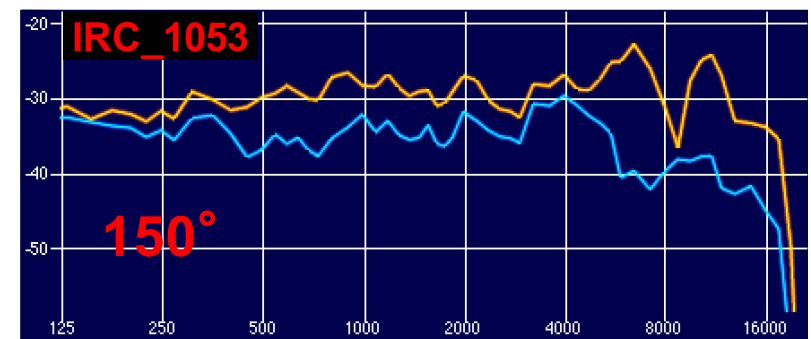
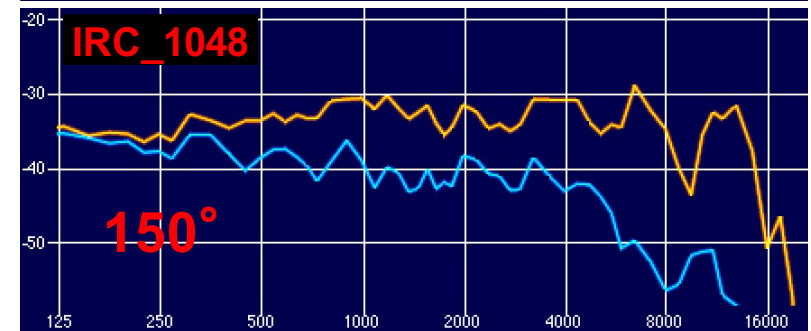
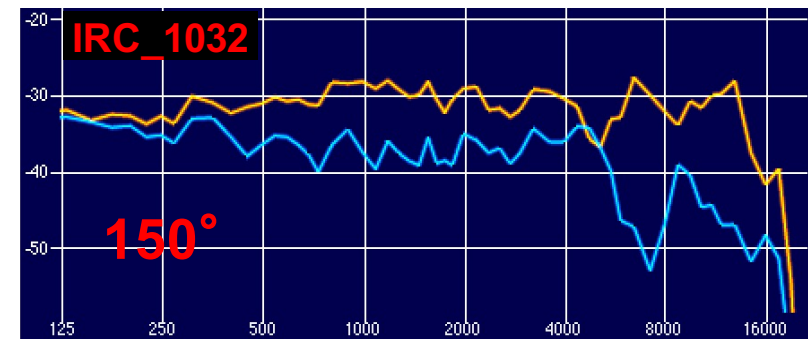
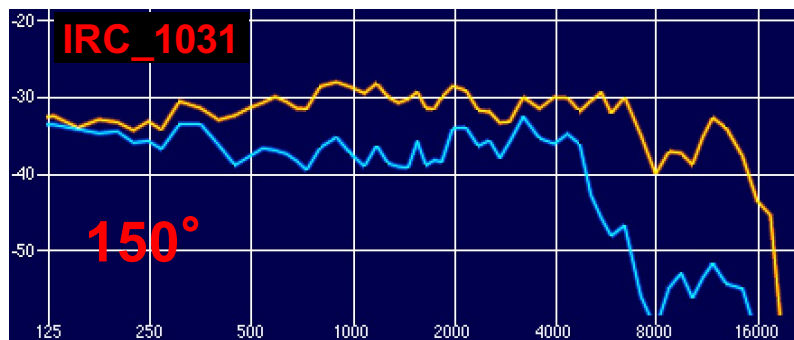
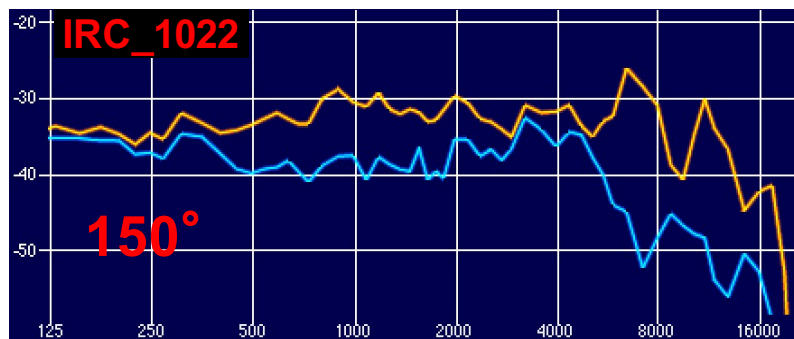
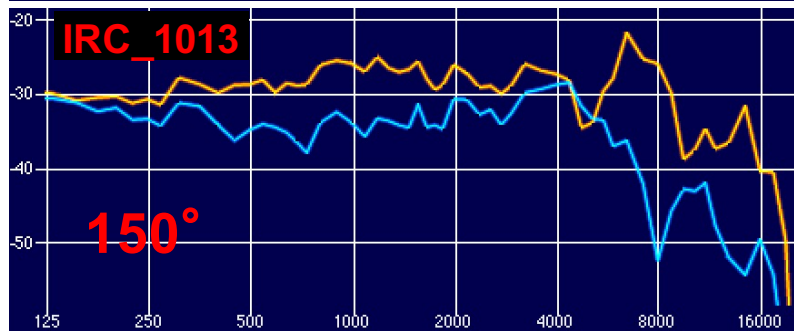
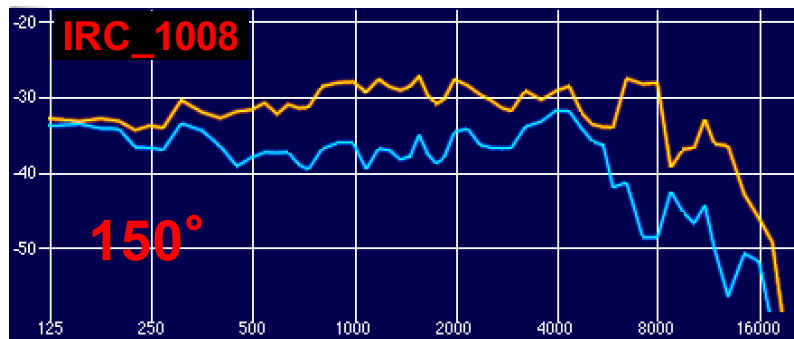
— Oreille Ipsilatéral L
— Oreille Contralatéral

7 HRTF : ANAGLYPH à 120°



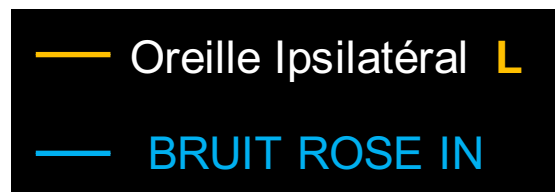
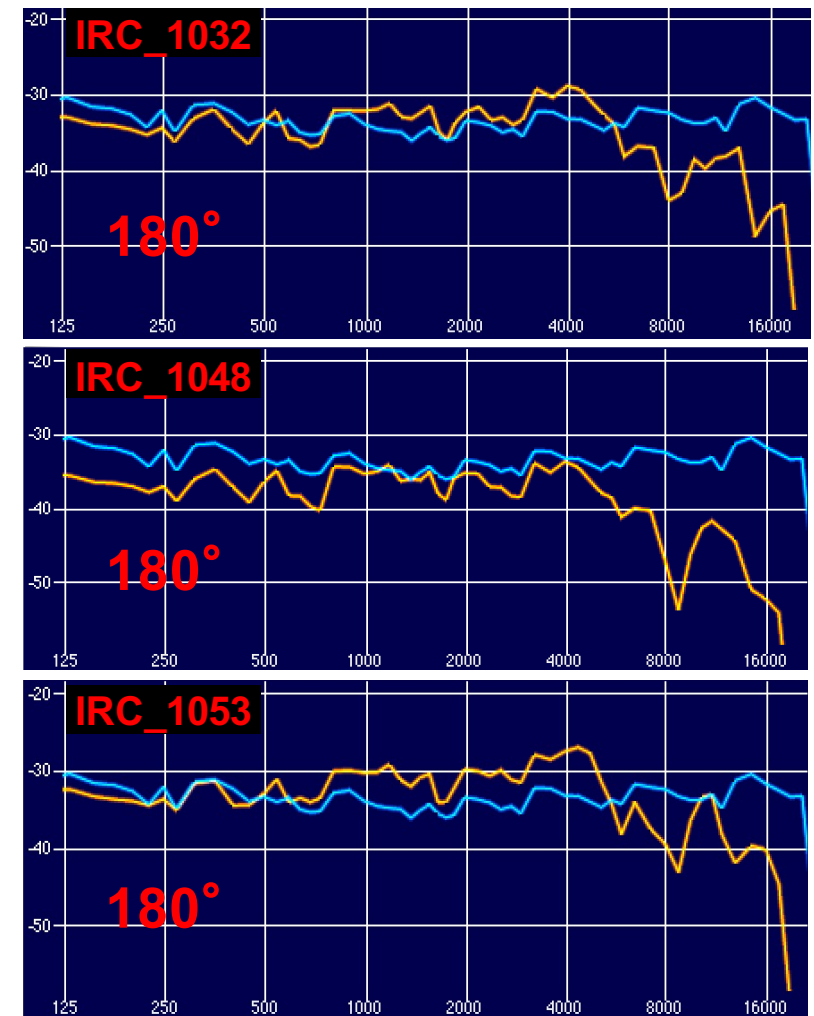
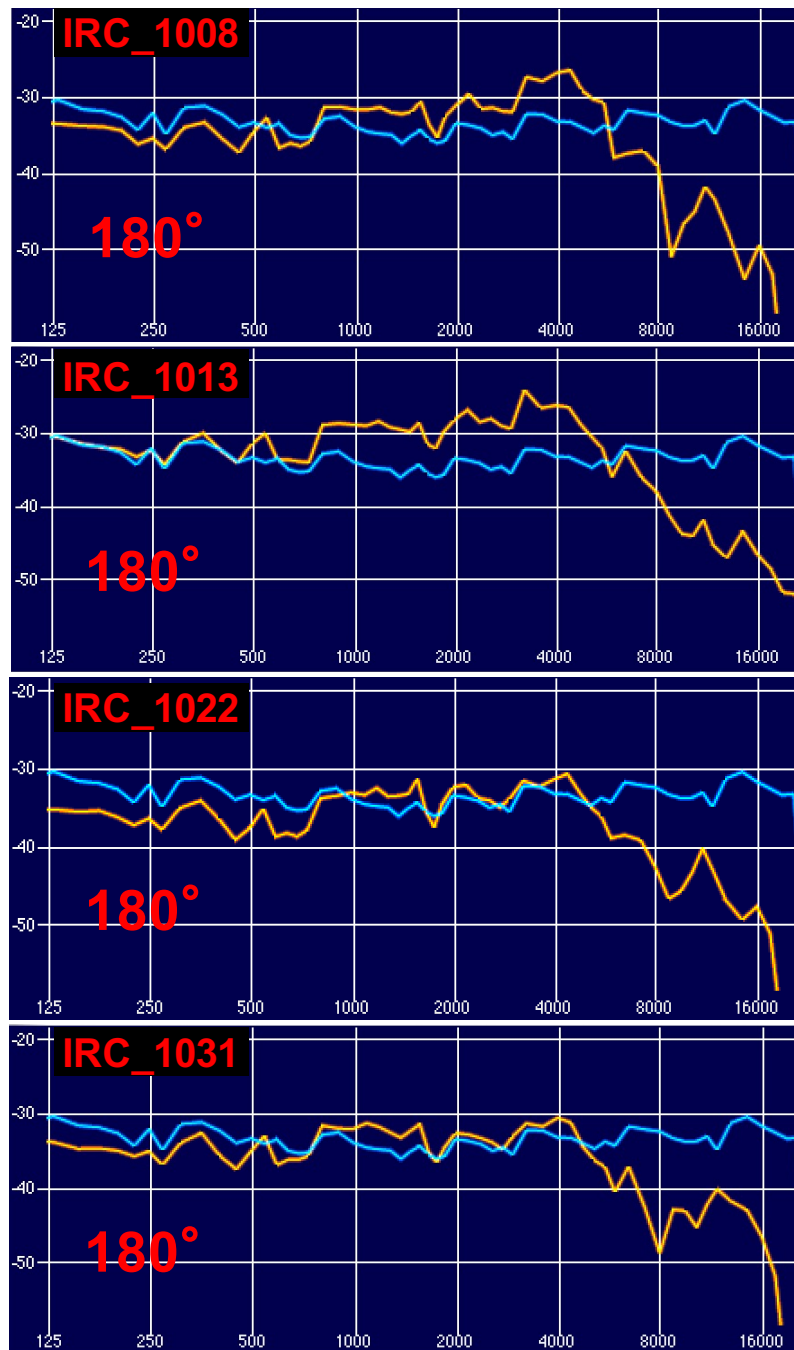
— Oreille Ipsilatéral L
— Oreille Contralatéral

7 HRTF : ANAGLYPH à 150°



— Oreille Ipsilatéral L
— Oreille Contralatéral

7 HRTF : ANAGLYPH à 180° Back



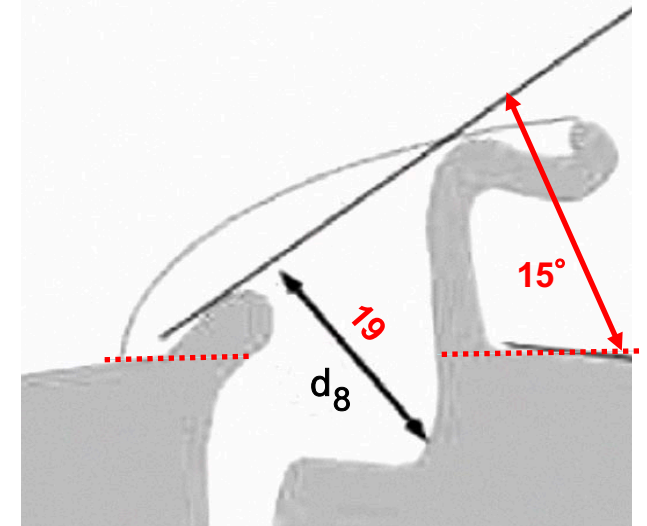
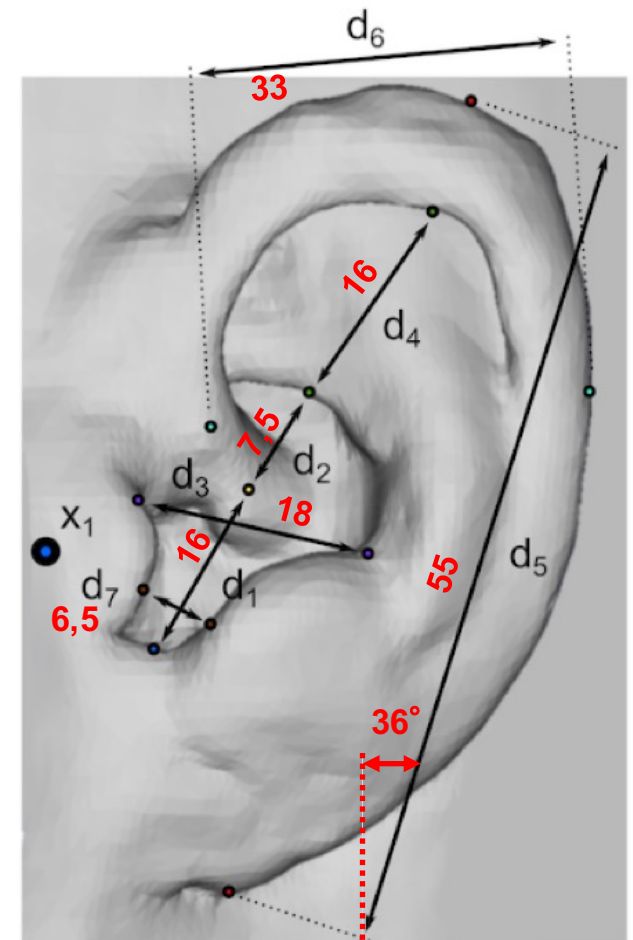
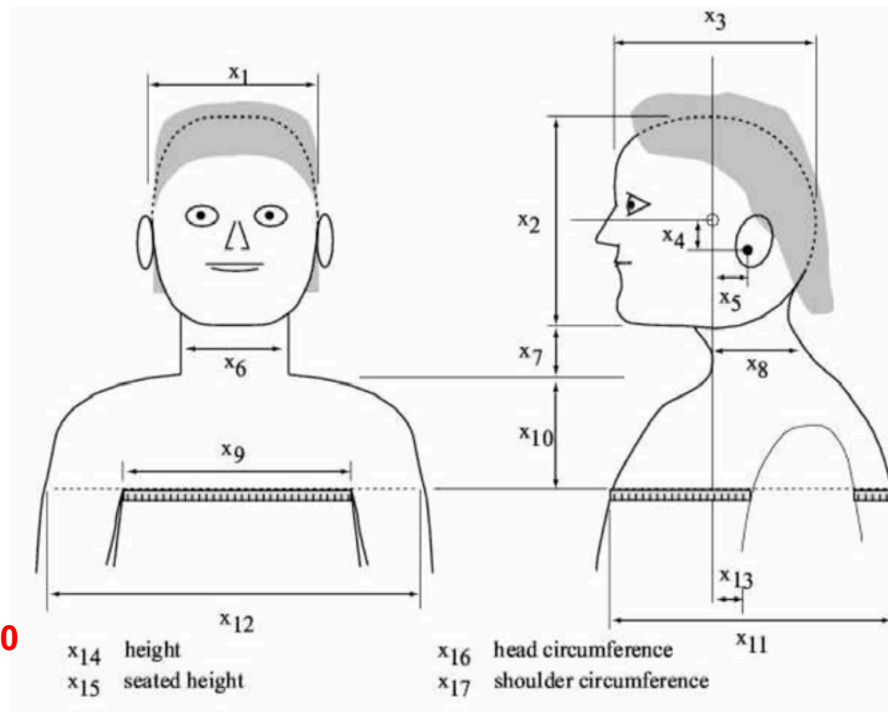
IRC_1008

ID : IRC08

Été 2002

- Homme
- Cheveux court

x_1 head width **155**
 x_2 head height **217**
 x_3 head depth **200**
 x_4 pinna offset down **28**
 x_5 pinna offset back **24**
 x_6 neck width **102**
 x_7 neck height **63**
 x_8 neck depth **102**
 x_9 torso top width **410**
 x_{10} torso top height **160**
 x_{11} torso top depth **162**
 x_{12} shoulder width **470**
 x_{13} head offset forward
 x_{14} height
 x_{15} seated height
 x_{16} head circumference **550**
 x_{17} shoulder circumference **1070**



Moyenne pavillon CIPIC

d_1	d_2	d_3	d_4	d_5	d_6	d_7
18	8	18	19	64	33	5
16	7,5	18	16	55	33	6,5

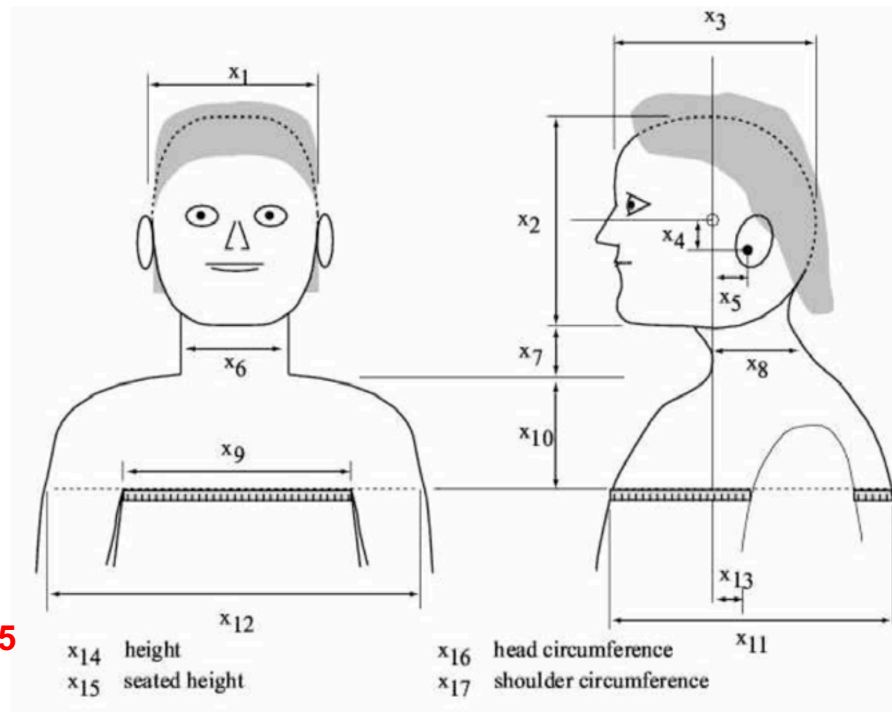
IRC_1013

ID : IRC13

Été 2002

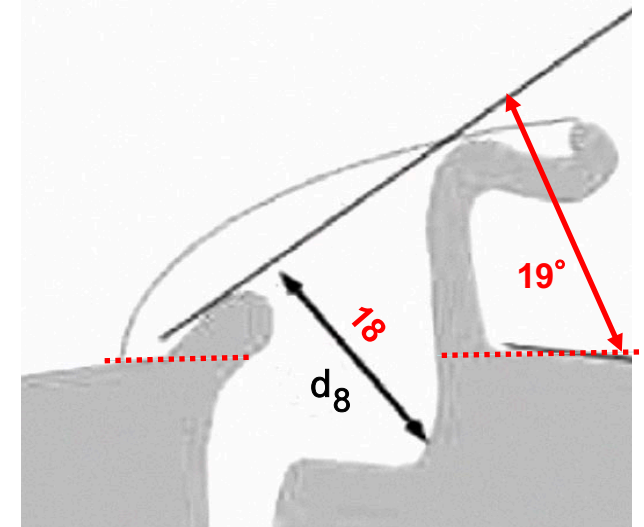
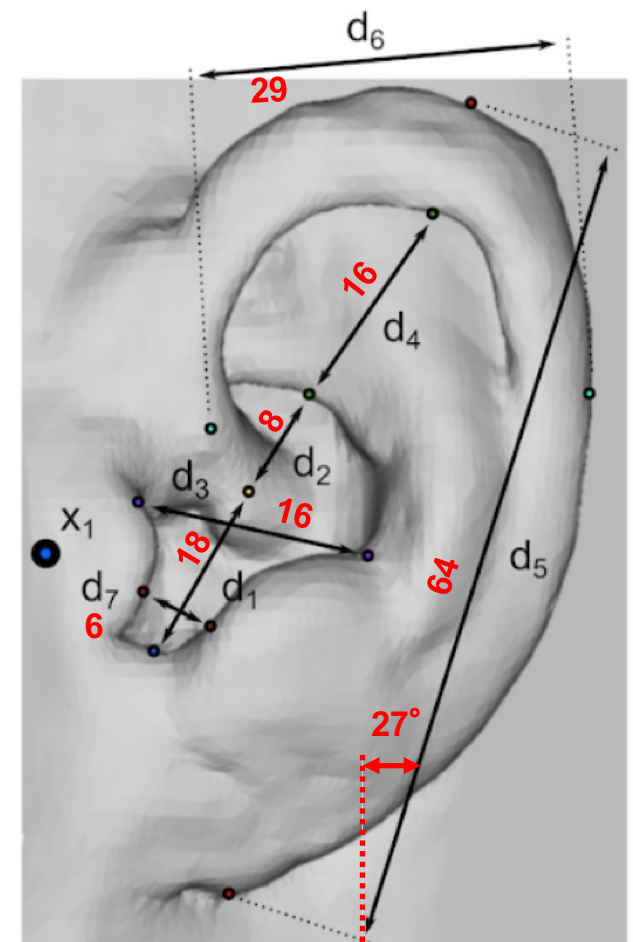
- Homme
- Coiffure courte

x_1 head width **143**
 x_2 head height **227**
 x_3 head depth **201**
 x_4 pinna offset down **8**
 x_5 pinna offset back **43**
 x_6 neck width **103**
 x_7 neck height **110**
 x_8 neck depth **115**
 x_9 torso top width **320**
 x_{10} torso top height **90**
 x_{11} torso top depth **230**
 x_{12} shoulder width **470**
 x_{13} head offset forward
 x_{14} height
 x_{15} seated height
 x_{16} head circumference **565**
 x_{17} shoulder circumference **1175**



Moyenne pavillon CIPIC

d_1	d_2	d_3	d_4	d_5	d_6	d_7
18	8	18	19	64	33	5
18	8	16	16	64	29	6



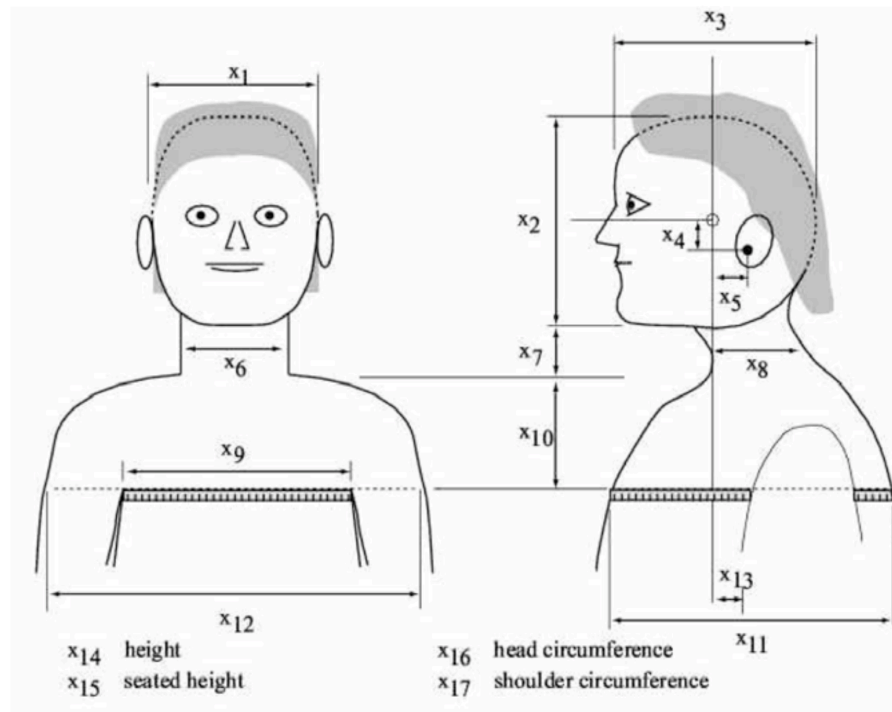
IRC_1022

ID : IRC22

Été 2002

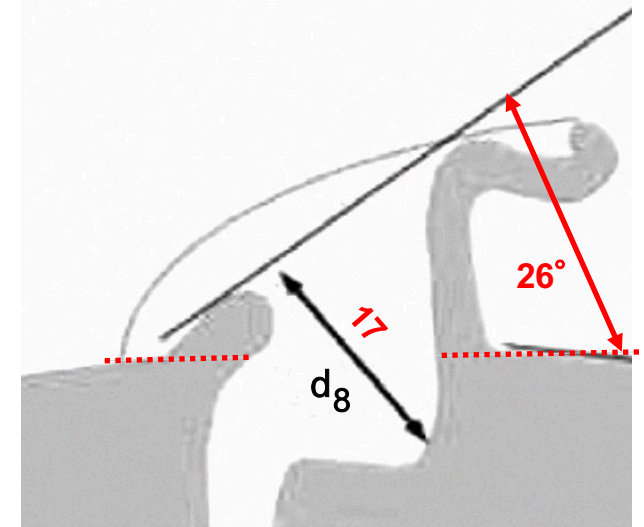
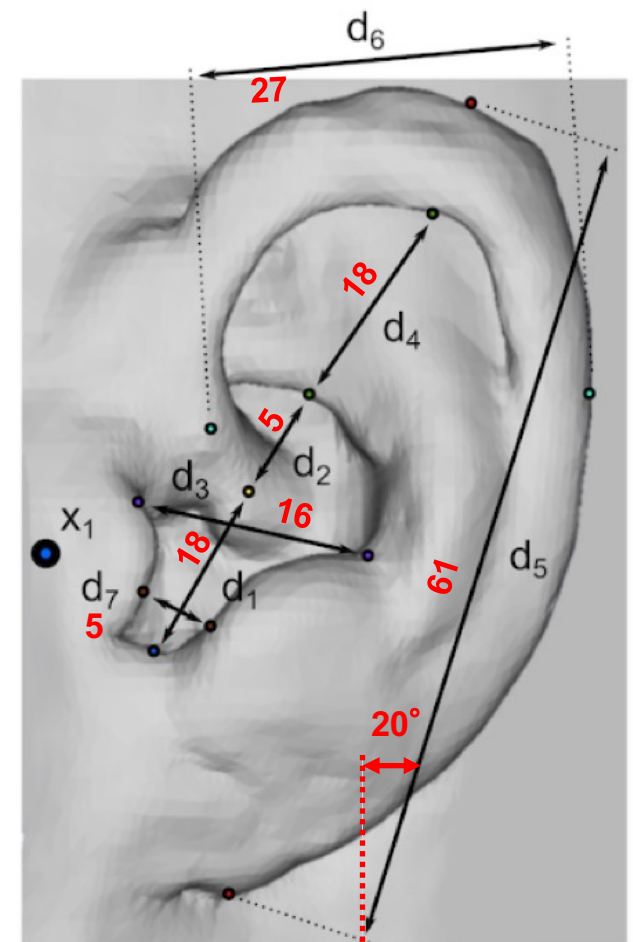
- Femme
- Coiffure : longueur moyenne

x_1 head width **150**
 x_2 head height **212**
 x_3 head depth **183**
 x_4 pinna offset down **34**
 x_5 pinna offset back **36**
 x_6 neck width **87**
 x_7 neck height **103**
 x_8 neck depth **90**
 x_9 torso top width **275**
 x_{10} torso top height **65**
 x_{11} torso top depth **205**
 x_{12} shoulder width **390**
 x_{13} head offset forward
 x_{14} height
 x_{15} seated height
 x_{16} head circumference **550**
 x_{17} shoulder circumference **950**



Moyenne pavillon CIPIC

d_1	d_2	d_3	d_4	d_5	d_6	d_7
18	8	18	19	64	33	5
18	5	16	18	61	27	5



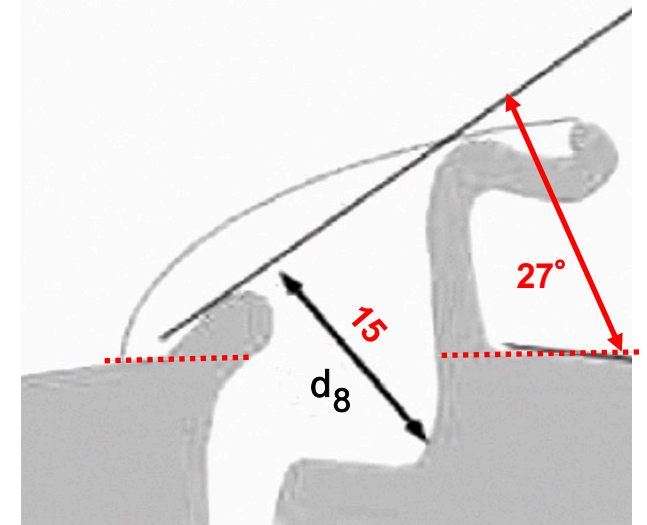
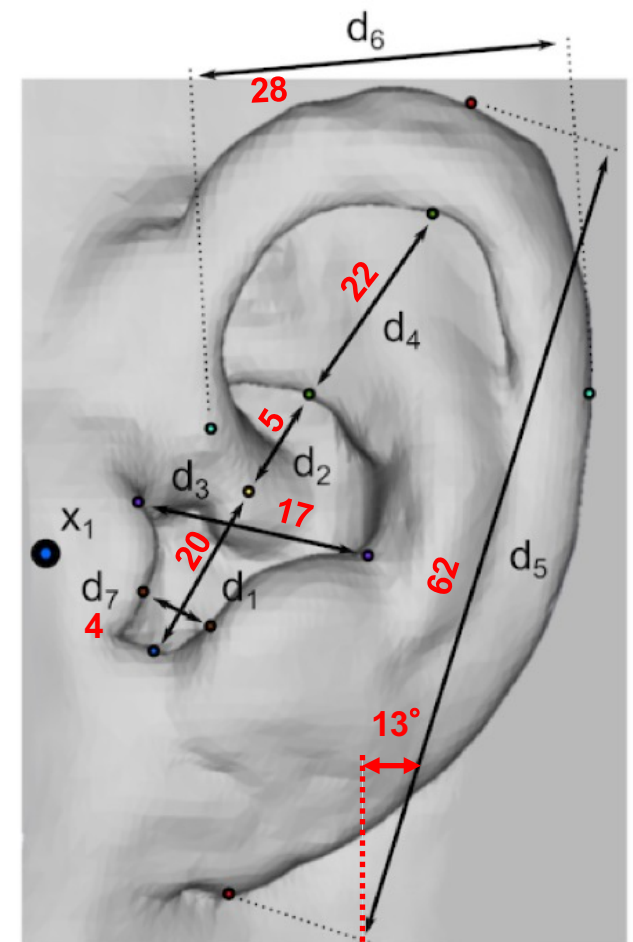
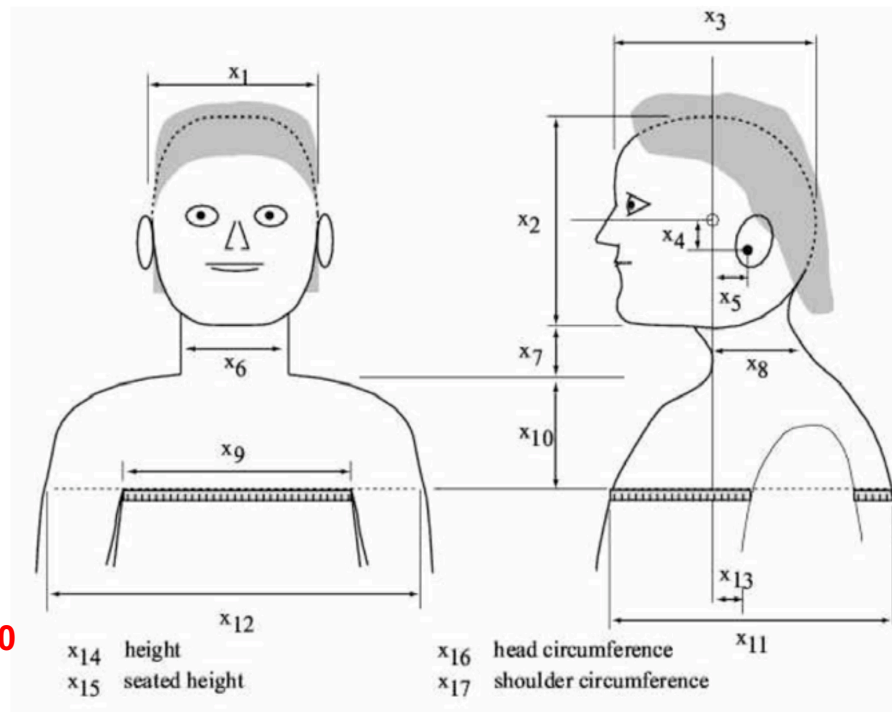
IRC_1031

Automne 2002

ID : IRC31

- Homme
- Coiffure : longue et volumineuse

x_1 head width **153**
 x_2 head height **232**
 x_3 head depth **216**
 x_4 pinna offset down **25**
 x_5 pinna offset back **10**
 x_6 neck width **106**
 x_7 neck height **57**
 x_8 neck depth **106**
 x_9 torso top width **330**
 x_{10} torso top height **66**
 x_{11} torso top depth **192**
 x_{12} shoulder width **530**
 x_{13} head offset forward
 x_{14} height
 x_{15} seated height
 x_{16} head circumference **602**
 x_{17} shoulder circumference **1130**



Moyenne pavillon CIPIC

d_1	d_2	d_3	d_4	d_5	d_6	d_7
18	8	18	19	64	33	5
20	5	17	22	62	28	4

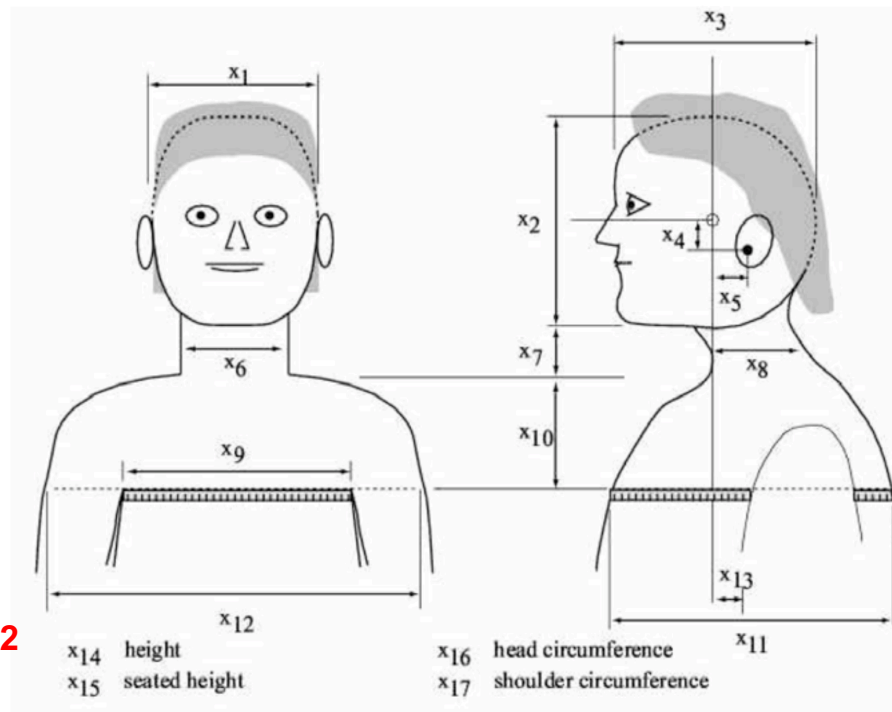
IRC_1032

Automne 2002

ID : IRC32

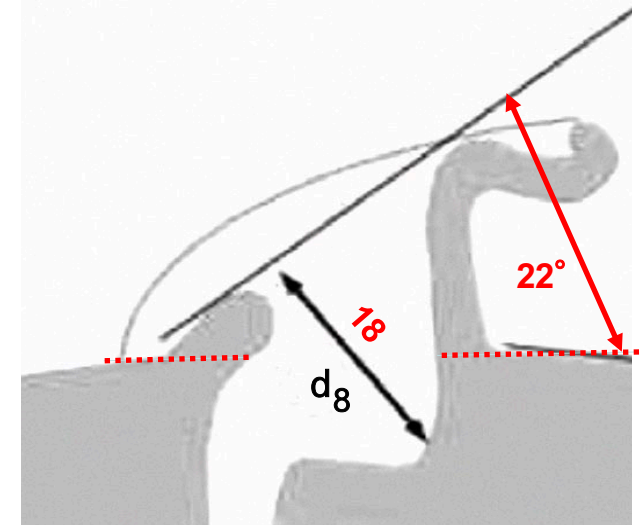
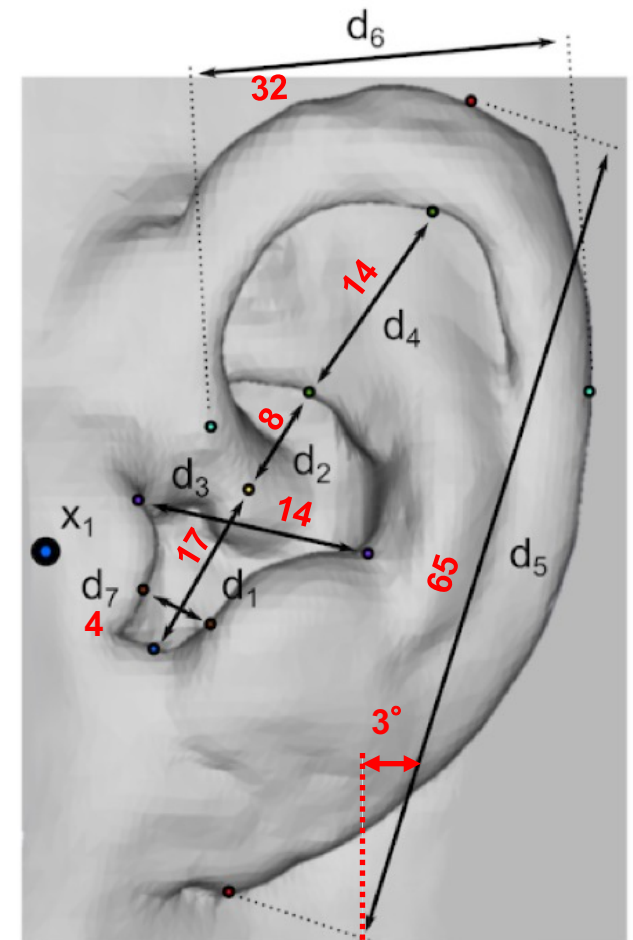
- Homme
- Coiffure : courte

x_1 head width **157**
 x_2 head height **212**
 x_3 head depth **201**
 x_4 pinna offset down **29**
 x_5 pinna offset back **-12**
 x_6 neck width **97**
 x_7 neck height **96**
 x_8 neck depth **110**
 x_9 torso top width **390**
 x_{10} torso top height **64**
 x_{11} torso top depth **192**
 x_{12} shoulder width **480**
 x_{13} head offset forward
 x_{14} height
 x_{15} seated height
 x_{16} head circumference **575**
 x_{17} shoulder circumference **1022**



Moyenne pavillon CIPIC

d_1	d_2	d_3	d_4	d_5	d_6	d_7
18	8	18	19	64	33	5
17	8	14	14	65	32	4



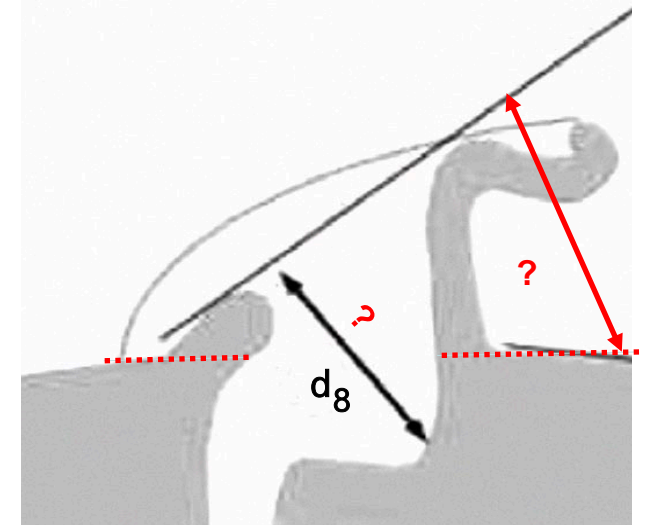
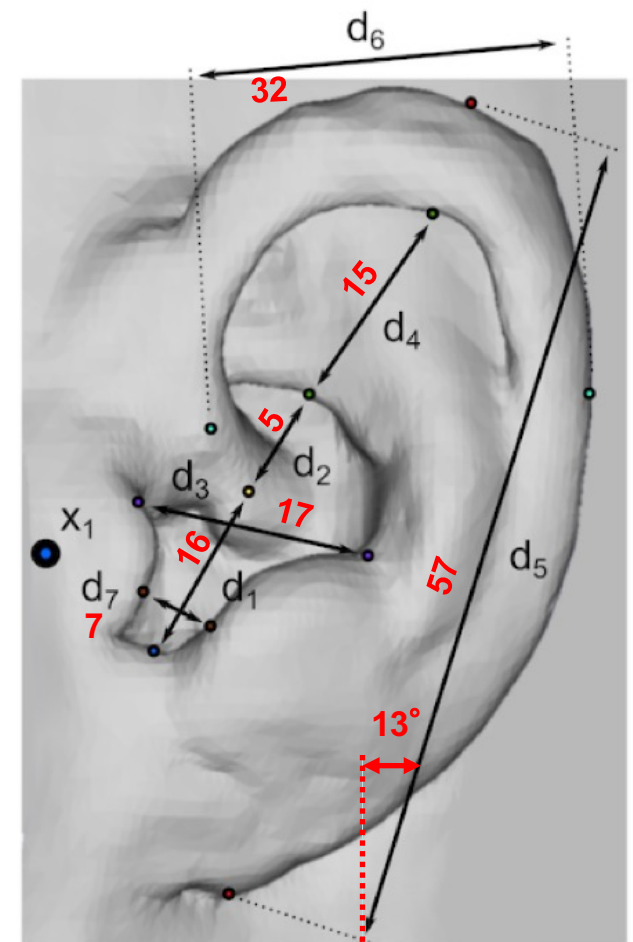
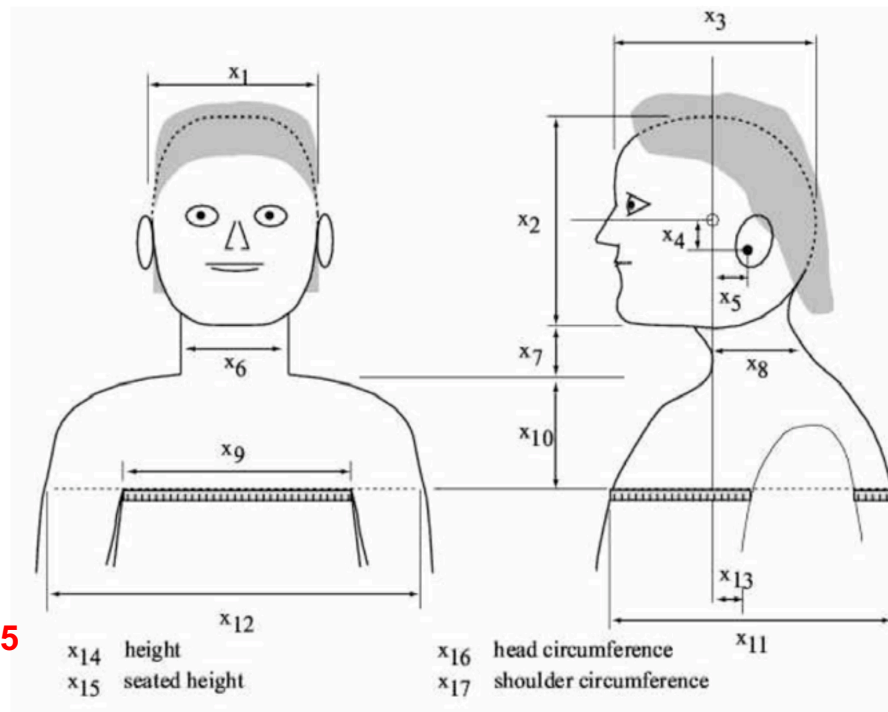
IRC_1048

Automne 2002

ID : IRC48

- Femme
- Coiffure : longue

x_1 head width **150**
 x_2 head height **212**
 x_3 head depth **190**
 x_4 pinna offset down **22**
 x_5 pinna offset back **-2**
 x_6 neck width **90**
 x_7 neck height **115**
 x_8 neck depth **110**
 x_9 torso top width **330**
 x_{10} torso top height **75**
 x_{11} torso top depth **220**
 x_{12} shoulder width **430**
 x_{13} head offset forward
 x_{14} height
 x_{15} seated height
 x_{16} head circumference **555**
 x_{17} shoulder circumference **1025**



Moyenne pavillon CIPIC

d_1	d_2	d_3	d_4	d_5	d_6	d_7
18	8	18	19	64	33	5
16	5	17	15	57	32	7

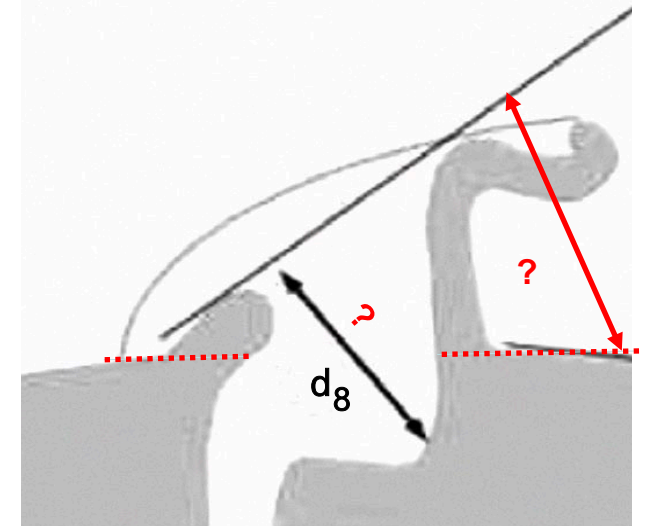
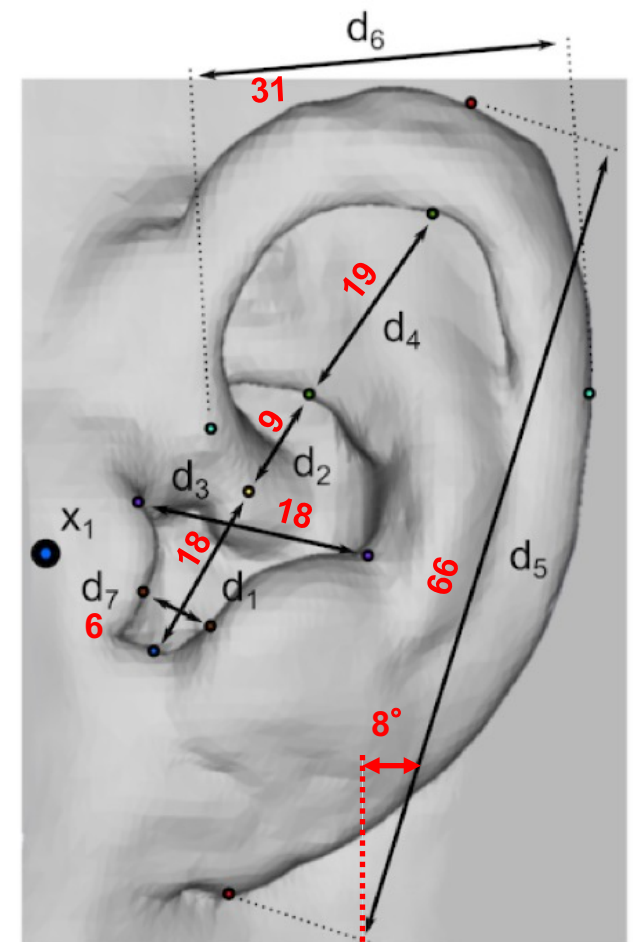
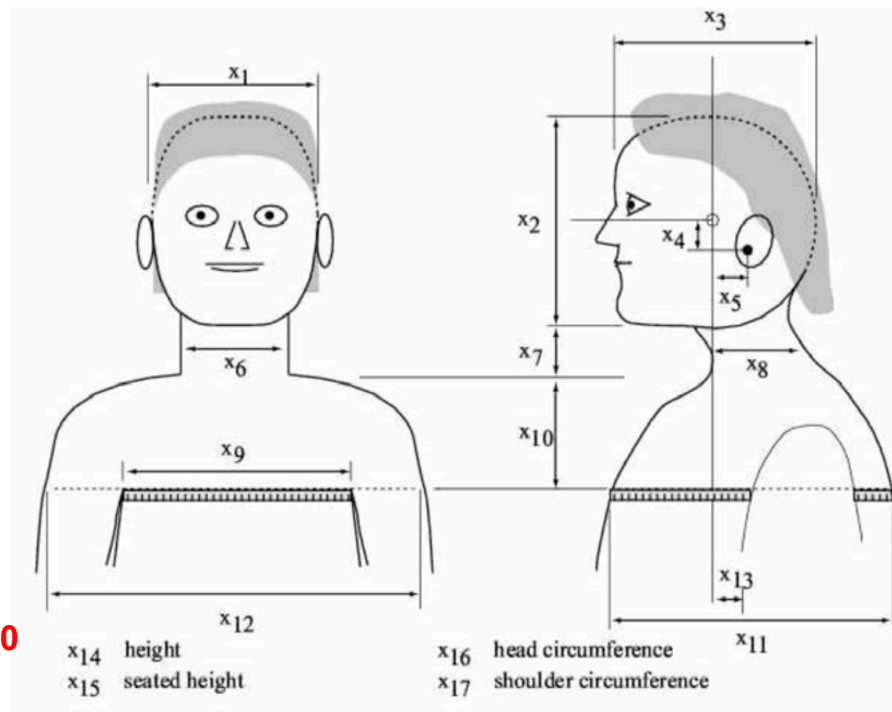
IRC_1053

ID : IRC53

Automne 2002

- Femme
- Coiffure : courte


































x_1 head width **151**
 x_2 head height **225**
 x_3 head depth **185**
 x_4 pinna offset down **15**
 x_5 pinna offset back **15**
 x_6 neck width **94**
 x_7 neck height **102**
 x_8 neck depth **102**
 x_9 torso top width **260**
 x_{10} torso top height **65**
 x_{11} torso top depth **190**
 x_{12} shoulder width **415**
 x_{13} head offset forward
 x_{14} height
 x_{15} seated height
 x_{16} head circumference **570**
 x_{17} shoulder circumference **1030**



Moyenne pavillon CIPIC

d_1	d_2	d_3	d_4	d_5	d_6	d_7
18	8	18	19	64	33	5
18	9	18	19	66	31	6

Index of /data/database/listen

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
<hr/>			
 Parent Directory	-		
 irc 1002.sofa	2014-03-20 16:26	701K	
 irc 1003.sofa	2014-03-20 16:26	699K	
 irc 1004.sofa	2014-03-20 16:26	696K	
 irc 1005.sofa	2014-03-20 16:26	698K	
 irc 1006.sofa	2014-03-20 16:26	698K	
 irc 1007.sofa	2014-03-20 16:26	702K	
 irc 1008.sofa	2014-03-20 16:26	698K	
 irc 1009.sofa	2014-03-20 16:26	699K	
 irc 1012.sofa	2014-03-20 16:26	699K	
 irc 1013.sofa	2014-03-20 16:26	697K	
 irc 1014.sofa	2014-03-20 16:26	697K	
 irc 1015.sofa	2014-03-20 16:26	699K	
 irc 1016.sofa	2014-03-20 16:26	705K	
 irc 1017.sofa	2014-03-20 16:26	704K	
 irc 1018.sofa	2014-03-20 16:26	703K	
 irc 1020.sofa	2014-03-20 16:26	701K	
 irc 1021.sofa	2014-03-20 16:26	704K	
 irc 1022.sofa	2014-03-20 16:26	702K	
 irc 1023.sofa	2014-03-20 16:26	703K	
 irc 1025.sofa	2014-03-20 16:26	704K	
 irc 1026.sofa	2014-03-20 16:26	699K	
 irc 1028.sofa	2014-03-20 16:26	705K	
 irc 1029.sofa	2014-03-20 16:26	702K	
 irc 1030.sofa	2014-03-20 16:26	698K	
 irc 1031.sofa	2014-03-20 16:26	703K	
 irc 1032.sofa	2014-03-20 16:26	698K	
 irc 1033.sofa	2014-03-20 16:26	700K	
 irc 1034.sofa	2014-03-20 16:26	704K	
 irc 1037.sofa	2014-03-20 16:26	698K	
 irc 1038.sofa	2014-03-20 16:26	700K	
 irc 1039.sofa	2014-03-20 16:26	697K	
 irc 1040.sofa	2014-03-20 16:26	698K	
 irc 1041.sofa	2014-03-20 16:26	699K	
 irc 1042.sofa	2014-03-20 16:26	700K	
 irc 1043.sofa	2014-03-20 16:26	698K	
 irc 1044.sofa	2014-03-20 16:26	697K	
 irc 1045.sofa	2014-03-20 16:26	699K	
 irc 1046.sofa	2014-03-20 16:26	698K	
 irc 1047.sofa	2014-03-20 16:26	699K	
 irc 1048.sofa	2014-03-20 16:26	700K	
 irc 1049.sofa	2014-03-20 16:26	700K	
 irc 1050.sofa	2014-03-20 16:26	702K	
 irc 1051.sofa	2014-03-20 16:26	701K	
 irc 1052.sofa	2014-03-20 16:26	699K	
 irc 1053.sofa	2014-03-20 16:26	702K	
 irc 1054.sofa	2014-03-20 16:26	699K	
 irc 1055.sofa	2014-03-20 16:26	694K	
 irc 1056.sofa	2014-03-20 16:26	703K	
 irc 1057.sofa	2014-03-20 16:26	701K	
 irc 1058.sofa	2014-03-20 16:26	704K	
 irc 1059.sofa	2014-03-20 16:26	701K	
<hr/>			
Apache/2.4.29 (Ubuntu) Server at sofacoustics.org Port 80			

<http://sofacoustics.org/data/database/listen/>

Sélection de HRTF dans une Base de Données en Utilisant des Paramètres Morphologiques pour la Synthèse Binaurale

David Schönstein¹, Brian FG Katz²

¹ARKAMYS, 31 rue Pouchet, 75017, Paris, dschonstein@arkamys.com

²LIMSI-CNRS, BP 133, Université Paris XI, F-91403, Orsay Cedex, brian.katz@limsi.fr

Les environnements virtuels auditifs deviennent de plus en plus pertinents pour des applications comme la téléconférence, les prothèses auditives, les jeux vidéo, et l'écoute immersive en général. Afin de permettre une interprétation de haute fidélité d'une scène sonore dans de tels environnements, le contenu audio doit être traité avec les filtres acoustiques de l'auditeur, nommés *Head-Related Transfer Function* (HRTF). Le défi actuel, étant donné la difficulté de mesurer les HRTF d'un auditeur, est d'être capable d'effectuer une sélection de HRTF, à partir d'une base de données préexistante de HRTF, et de donner à l'auditeur les HRTF qui ressemblent le plus possible à ses propres HRTF en utilisant des données simples, comme uniquement une photo de l'oreille. Un processus est décrit dans lequel une base de données de HRTF mesurant 46 auditeurs différents est réduite au moyen de différentes techniques telles que l'analyse en composantes principales ou un décalage des HRTF dans le domaine fréquentiel. De multiples régressions linéaires sont effectuées sur les paramètres morphologiques les plus importants afin de prédire les HRTF préférées d'un auditeur. Les HRTF prédites sont ensuite comparées à un classement subjectif des sujets dans la base de données selon leurs préférences. Les résultats montrent que ce processus est en mesure de prédire les HRTF pour un auditeur de façon significativement meilleure que si les HRTF avaient été choisies au hasard.

1 Introduction

Les humains ont la capacité d'encoder l'information directionnelle d'un son incident avec une fonction de transfert acoustique. La tête, les oreilles externes, le corps d'un auditeur transforment l'information spectrale d'un son dans l'espace par ce qu'on appelle *Head-Related Transfer Function* (HRTF), et ceci nous permet de percevoir notre environnement acoustique en fonction de la position, distance, etc. des sources sonores.

Pour les applications de la synthèse binaurale comme la téléconférence, les prothèses auditives, les jeux vidéo, et l'écoute immersive en général, un rendu haute fidélité de la scène sonore est préférable. L'utilisation des HRTF qui ressemblent le plus possible aux filtres acoustiques de l'auditeur permet ce rendu haute fidélité. Plusieurs études dans la littérature démontrent l'intérêt de ce qu'on appelle des HRTF individualisées [1-4], surtout en termes de précision dans les tests de localisation. En revanche, il existe des études qui montrent que les propres HRTF d'un auditeur ne sont pas toujours préférées par rapport à d'autres HRTF [5].

Les HRTF individualisées peuvent être obtenues via des mesures avec microphones dans les oreilles de l'auditeur [6-8], ou même via une simulation numérique [9, 10]. Malgré la qualité de leur rendu, ces méthodes restent encore aujourd'hui très laborieuses. Il existe d'autres possibilités orientées vers une utilisation grand public comme la mesure rapide des HRTF via le principe de réciprocité [11] ou interpolation [12], l'adaptation de HRTF non individualisées [12, 13], ou la sélection des meilleures HRTF dans une base de données [14-18].

Cette étude vise à poursuivre la troisième méthode, notamment inspirée par des techniques qui utilisent les paramètres morphologiques [17-20]. La plupart de ces

techniques se focalisent sur une validation avec test de localisation. L'étude actuelle cherche à valider deux différentes méthodes pour la sélection de HRTF dans une base de données, non pas avec des tests de localisation, ni avec des HRTF interpolées, mais avec une évaluation qui est plus adaptée aux applications citées : en utilisant un test d'écoute et les résultats des préférences des sujets.

2 Méthode

2.1 Test d'écoute

L'objet de cette étude était de trouver un système de sélection d'une HRTF parmi les éléments d'une base de données. Cette base de données est composée de HRTF et des paramètres morphologiques des sujets associés. Pour mieux évaluer les méthodes de sélection, un classement subjectif des HRTF dans la base de données a été effectué.

Les HRTF utilisées venaient de la base de données publique du projet LISTEN [21]. 45 sujets ont participé à cette évaluation. Les mesures « brutes » (sans égalisation du champ diffus) de 46 sujets (un sujet n'a pas participé dans le test) ont été choisies pour ce test. Les HRTF ont été décomposées pour extraire les composants « phase-minimum » et « phase-excès ». La partie « phase-excès » a été remplacée par un retard pur, qui représente l'*Interaural Time Difference* (ITD) de chaque sujet. L'ITD a été déterminée selon le calcul du maximum du coefficient de corrélation croisée *interaural* (maxIACC) des enveloppes énergétiques.

De cette façon, le délai est apparié pour le sujet et l'information spectrale de filtrage sur la HRTF est variée. Cela revient, en simulation, à changer les oreilles de chaque sujet tout en gardant une même géométrie de la tête.

Le signal utilisé pour le test était une synthèse binaurale d'un bruit blanc à large bande de 0,23 secondes modelé en utilisant une fenêtre de Hanning. Le signal de test a été présenté à des positions fixes le long de deux trajectoires présentées en séquence :

1. Un cercle dans le plan horizontal (élévation = 0°) en incréments de 30°. La trajectoire commence à 0° azimuth et 0° élévation et on effectue deux rotations (durée 6 secondes).
2. Un arc dans le plan médian (azimut = 0°) à partir de l'élévation -45° à l'avant jusqu'à -45° à l'arrière en incréments de 15°. La trajectoire commence à l'avant à élévation -45°, on fait varier l'élévation vers l'arrière et on revient par le même chemin à la position initiale (durée 9 secondes).

Pour chaque sujet, 46 signaux, correspondant aux 46 différentes HRTF de la base de données (incluant la propre HRTF du sujet), ont été présentés en utilisant une interface graphique. Pour chaque HRTF, les sujets ont dû faire un choix : *excellent*, *moyen* ou *mauvais*. Les sujets ont été invités à sélectionner de 5 à 10 HRTF comme *excellent*, afin d'obtenir une statistique utile pour l'analyse des résultats. Les sujets ne savaient pas quelle HRTF parmi les 46 correspondait à leur propre mesure. Les sujets ont été autorisés à écouter les signaux à plusieurs reprises, et dans n'importe quel ordre. La durée du test était d'environ 35 minutes.

La Figure 1 présente les résultats des tests d'écoute pour tous les sujets. Les sujets sont représentés sur l'axe horizontal et les HRTF classées sur l'axe vertical. La couleur des cercles indique le classement des HRTF. La diagonale montre clairement que tous les sujets, sauf 3, ont classé leurs propres HRTF comme *excellent*.

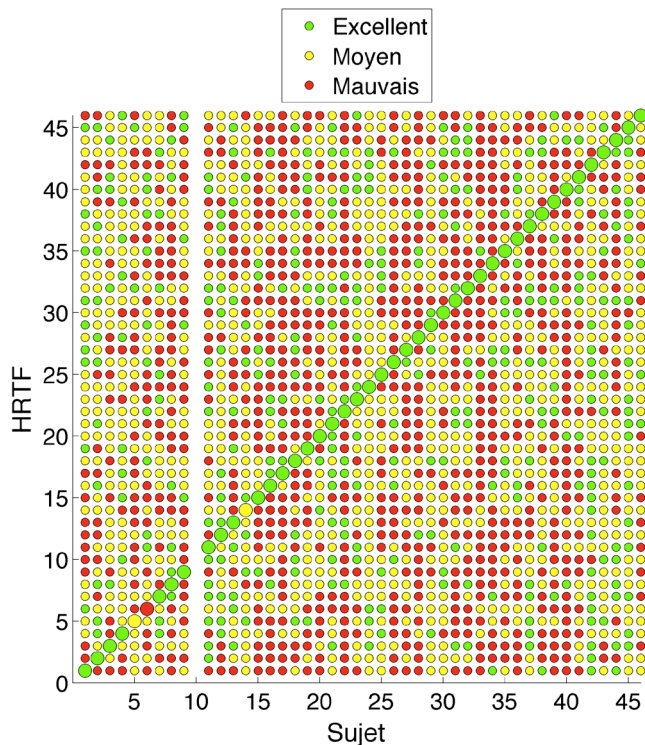


Figure 1 : Classement des HRTF par les sujets. Le sujet numéro 10 n'a pas classé les HRTF.

2.2 Analyse en composante principale

Les mesures de HRTF LISTEN [21] ont été prises à des positions dans l'espace correspondant à des angles d'élévation compris entre -45° et 90° par incréments de 15° et angles d'azimut commençant à 0° par incréments de 15°. Les incréments en azimut ont été progressivement augmentés pour les angles d'élévation au-dessus de 45°, afin d'échantillonner l'espace de façon égale, pour un total de 187 positions. Aucune interpolation n'a été utilisée dans cette étude puisque il était considéré comme important de ne pas avoir la qualité d'interpolation comme variable dans les données, ce qui veut dire que les HRTF mesurées n'échantillonnaient pas l'espace d'une manière parfaitement égale. Les mesures brutes de HRTF ont été normalisées par la racine carrée moyenne entre les oreilles gauche et droite pour toutes les positions, et ensuite normalisées par la racine carrée moyenne pour toutes les positions pour les oreilles gauche et droite sur tous les sujets.

Les HRTF ont été transformées ensuite en ce que l'on appelle la *Directional Transfer Function* (DTF) [22], qui contient uniquement la part des HRTF qui présentent une dépendance directionnelle : les DTF sont les spectres d'amplitude des HRTF divisés par la moyenne des spectres pour toutes les positions dans l'espace pour chaque oreille. De plus, la résolution fréquentielle limitée du système auditif est prise en compte par un lissage des DTF en bande critique. Dans cette étude, la distribution spatiale des mesures de HRTF dans l'espace n'a pas été prise en compte pour le calcul de la DTF.

Les DTF traitées de 512 points, ont été ensuite concaténées pour toutes les positions pour chaque oreille. Les vecteurs des spectres concaténés de chaque oreille ont été ensuite eux même concaténés, ce qui a donné un vecteur par sujet, formant ainsi une rangée d'une matrice de dimension 46x191488 contenant tous les sujets. Les colonnes de cette matrice représentent les magnitudes en fréquence pour toutes les DTF. Une analyse en composante principale (ACP) a été réalisée sur cette matrice. L'objectif de l'ACP était de réduire la dimensionnalité de la base de données de HRTF qui présente une certaine redondance, tout en conservant la part la plus significative de la variabilité observée [23, 24]. L'analyse a donné une nouvelle matrice de données (les scores) qui représentaient les données originales projetées sur de nouveaux axes (les composantes principales) dans les dimensions de la matrice originale qui redisent la variance totale pour chaque colonne de magnitude en fréquence. Chaque rangée de cette nouvelle matrice de scores représentait toujours un sujet, et chaque colonne représentait les données pour chaque axe dans les dimensions de l'espace classé en fonction de la quantité de variance des données originales décrites. Chaque colonne dans la matrice des scores a été utilisée ensuite comme des coordonnées dans une dimension de l'espace. En prenant par exemple trois colonnes, les sujets ont pu être représentés dans un espace tridimensionnel. La distance euclidienne entre les sujets représente une mesure de leurs similarités en fonction des HRTF mesurées.

2.3 Choix des paramètres morphologiques les plus importants

La base de données LISTEN inclut, pour chaque sujet, 22 paramètres morphologiques définis en [25] (Figure 2). Cela représente tous les paramètres morphologiques sauf : x_{13} , x_{14} , x_{15} , d_8 et θ_2 . En utilisant une méthode de *machine*

learning nommée *feature selection*, on souhaitait trouver un sous-ensemble de caractéristiques pertinentes, les paramètres morphologiques les plus importants. Après des évaluations préliminaires, une meilleure performance de *feature selection* a été trouvée en groupant les résultats en deux catégories : les HRTF classées comme *excellent* et *moyen* versus les HRTF classées comme *mauvais*.

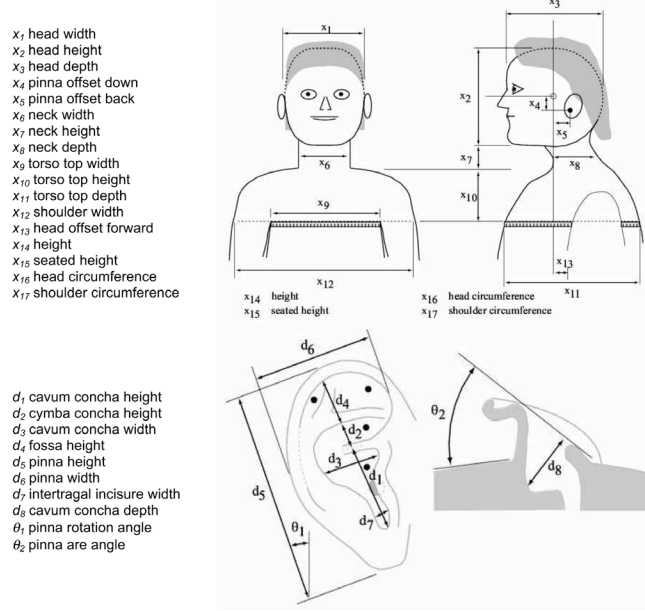


Figure 2 : Paramètres morphologiques de la base de données CIPIC dont 22 ont été utilisés dans cette étude.
Figure prise de [25].

Une normalisation des données morphologiques a ensuite été réalisée pour convenir à la méthode de *feature selection*. Les valeurs morphologiques de l'ensemble des HRTF ont été divisées par les données de chaque sujet. Avec cette normalisation, les valeurs représentaient le pourcentage d'un paramètre morphologique d'un sujet par rapport à un autre. Ces valeurs étaient liées avec les classements de HRTF pour chaque pair de sujet pour l'analyse.

La méthode de *feature selection* choisie pour trouver les paramètres morphologiques les plus importants a été celle des machines à vecteur de supports (*support vector machines* - SVM) qui construit un ensemble d'hyperplans dans un espace de dimension élevée pour classer les données normalisées. Cette technique a été utilisée par exemple pour trouver des gènes de cancer [26]. Avec cette méthode, les paramètres ont donc été classés du plus au moins important. Une variable dans la classification avec SVM est la valeur de complexité C , qui contrôle la tolérance d'erreurs de classifications dans l'analyse, et introduit une fonction de pénalité [27]. Une valeur de C nulle indique que la fonction de pénalité n'est pas prise en compte, et une valeur de C élevée ($C \rightarrow \infty$) indique que la fonction de pénalité est dominante. Selon les différentes valeurs de C , le classement des paramètres morphologiques changeait. Une valeur de $C = 1$ a été utilisée pour la classification avec SVM, choisie pour avoir une tolérance élevée des erreurs de classifications et pour donner les meilleurs résultats (section 3.1). En utilisant cette méthode, les dix premiers paramètres morphologiques les plus importants, du plus important au moins important, correspondaient à : x_{11} , x_2 , x_8 , d_5 , x_3 , d_4 , x_{12} , d_2 , d_1 et x_6 .

3.1 Analyse du *Scaling* Fréquentiel

[13] a proposé une technique d'adaptation de HRTF qui applique un décalage global en fréquence sur la HRTF entière pour que les creux et pics correspondent entre les HRTF du sujet et les HRTF non-individualisées. Cette méthode repose sur l'idée que les variations entre les HRTF d'un sujet à l'autre sont liées à la taille des cavités du pavillon qui jouent sur les résonances dans l'oreille. Une augmentation ou diminution de la taille d'une oreille par un facteur modifie ces résonances et décale les creux et pics vers les basses ou hautes fréquences respectivement.

La méthode de [15] a été utilisée pour calculer le degré de décalage (*scaling factor* global) entre deux sujets de la base de données. Afin de ne pas sur-échantillonner dans les hautes fréquences, les DTF ont été traitées avec une banque de filtres passe-bande (70 par octave) pour échantillonner les composantes en fréquence sur des intervalles égaux en échelle logarithmique base 2, soit en échelle d'octave. Pour chaque direction, la DTF d'un sujet est soustraite, fréquence par fréquence, de la DTF de l'autre sujet, à la même direction pour calculer ce qu'on appelle l'*Inter-Subject Spectral Difference* (ISSD). La variance de cette différence est calculée sur la bande de fréquence de 3.7 kHz à 12.9 kHz et l'ISSD globale est finalement définie comme la moyenne, pour toutes les directions, des ISSD directionnelles. Le *scaling factor* est donc la valeur qui donne le minimum ISSD global.

Une fois le *scaling factor* calculé pour chaque pair de sujet, une matrice de dissemblance a été créée et utilisée pour une analyse par *Multi-Dimensional Scaling* (MDS). Comme il n'y avait pas de notion de direction de transformation avec les *scaling factor*, seulement la magnitude a été prise en compte. Le *scaling factor*, nommé alpha (α), peut être en forme de α ou $1/\alpha$ correspondant à un décalage de HRTF A vers B ou HRTF B vers A. Vu que les deux formes de *scaling factor* représentent la disparité fréquentielle entre HRTF A et B, et que le choix sera arbitraire, le facteur $\alpha > 1$ a été utilisé pour l'analyse.

Le MDS est une technique statistique qui a donné en effet le même type de matrice de coordonnées euclidiennes pour décrire les distance entre les HRTF dans un espace multidimensionnel, comme pour l'ACP. Dans l'analyse, cette méthode a seulement été utilisée pour l'oreille droite, puisque dans notre étude et celle de [13] il existait une forte corrélation (coefficient de corrélation de 0,83 et 0,95 respectivement) entre les *scaling factor* des deux oreilles.

3 Résultats

3.1 Validation

La validation des espaces multidimensionnels a été faite pour chaque sujet en classant les HRTF des sujets dans la base de données en fonction des distances euclidiennes pour toutes les dimensions possibles. Le classement des DTF pour chaque sujet a été comparé avec les résultats du test d'écoute. Pour chaque sujet, le rang a été calculé (une valeur entre 1 et 45, avec la valeur 1 représentant la meilleure HRTF) pour les HRTF classées comme *excellent*, *moyen* ou *mauvais* en utilisant l'espace multidimensionnel. Ensuite, les valeurs de rang de toutes les HRTF classées soit comme *excellent*, *moyen* ou *mauvais* pour tous les sujets ont été groupées ensemble et la distribution a été analysée pour chaque dimension. Comme les distributions des valeurs de

rang ne suivaient pas une distribution normale, une analyse de variance Kruskal-Wallis a été utilisée pour tester l'égalité des médianes.

La Figure 3 montre les distributions des rangs les plus statistiquement différents pour les deux méthodes de réduction. Les 45 premières dimensions ont été utilisées dans l'ACP, ce qui représentait 79 % de la variance des données des DTF. Pour l'analyse par MDS, les 26 premières dimensions ont été prises, ce qui représentait une reconstruction des données de la matrice de dissemblance. L'erreur maximale de la reconstruction était de 0,13 avec une valeur maximale de 0,23 dans la matrice de dissemblance. La distribution dans la figure pour *excellent* par exemple représente les valeurs de rang pour les HRTF classées comme *excellent* pour tous les sujets. Les distributions devenaient légèrement plus statistiquement significatives en augmentant le nombre de dimensions puisque les composants principaux (pour la méthode de l'ACP) décrivaient plus de variance des données. Globalement, ces deux méthodes ont des résultats comparables mais traitent les données des DTF très différemment. Il est intéressant que l'information contenue dans les *scaling factor*, qui représente les différences entre les sujets sur seulement un aspect de la DTF, puisse être utilisée dans les analyses des données pour créer un espace multidimensionnel performant. Cet espace est aussi cohérent que l'espace utilisant toutes les informations du spectre (ACP des DTF).

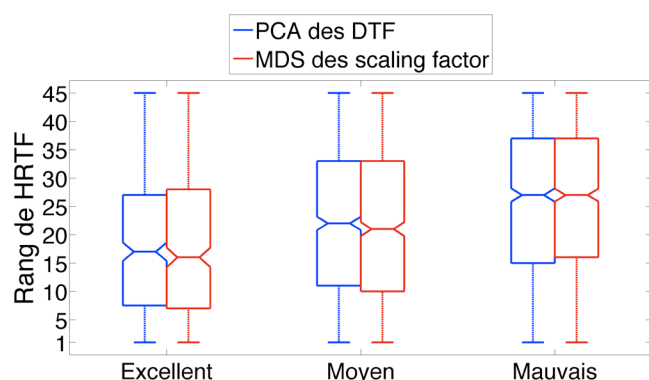


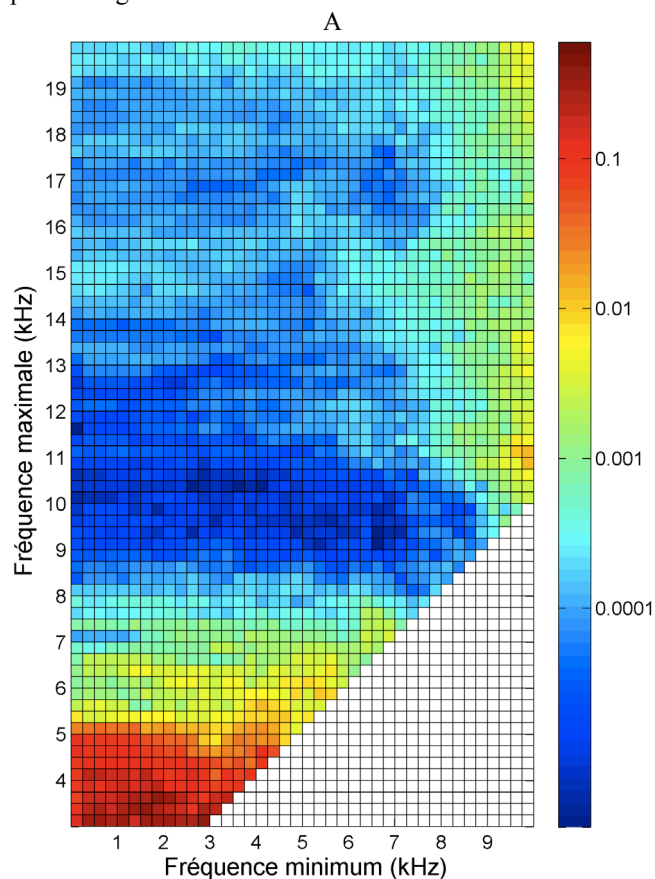
Figure 3 : distributions des rangs des HRTF pour les deux méthodes : l'ACP des DTF et l'analyse par MDS des *scaling factor*.

Étant plus intéressés par l'application de cette méthode pour la prédiction des meilleure HRTF pour un sujet donné, nous avons utilisé une métrique de validation plus subtile et mieux adaptée pour des applications de la synthèse binaurale. Cette métrique est calculée en vérifiant le pourcentage de HRTF classées dans le test d'écoute comme *excellent*, parmi les dix meilleures HRTF du classement de l'espace multidimensionnel de l'ACP des DTF. Ce pourcentage peut être comparé avec le pourcentage global de HRTF classées dans le test d'écoute comme *excellent* pour chaque sujet; correspondant au pourcentage de HRTF classées comme *excellent* qui existera dans une sélection de dix HRTF choisies au hasard. Cette comparaison a été faite pour toutes les dimensions et pour chaque sujet. Un test Student a été utilisé pour vérifier si les moyennes des distributions de valeurs de pourcentage étaient statistiquement différentes pour chaque nombre de dimension différent. La *p-value* des tests Student, nommé *p-value excellent*, représente la probabilité d'obtenir une

statistique de test au moins aussi extrême que celle qui a été effectivement observée, en supposant que l'hypothèse nulle (que les valeurs de pourcentage viennent de la même population) est vraie.

La *p-value excellent* a été utilisée pour trouver les bornes de fréquence des DTF pour créer la matrice de l'ACP donnant la valeur la plus petite, soit la plus statistiquement significative. La Figure 4(A) montre les minimums des valeurs *p-value excellent* pour toutes les dimensions en fonction des différentes bornes de fréquence utilisées. La fréquence minimum est représentée sur l'axe horizontal et la fréquence maximale sur l'axe vertical. Les bornes de fréquence optimales sont : 0 Hz et 11500 Hz. Ces bornes correspondent à une valeur de *p-value excellent* de $3,5 \times 10^{-6}$. Ces bornes de fréquence peuvent être interprétées comme la partie du DTF perceptiblement le plus important en termes de préférence des sujets dans le test d'écoute.

La Figure 4(B) montre le nombre de dimensions utilisées pour chaque borne de fréquence avec une valeur maximale de 45. Le nombre de dimensions utilisé pour les bornes de fréquence optimales était 35. Une corrélation existe entre la valeur *p-value excellent* représentée sur la Figure 4(A) et la valeur du nombre de dimensions utilisé dans l'analyse sur la Figure 4(B). Plus la valeur *p-value excellent* est petite, correspondant à un espace multidimensionnel plus performant, plus de dimensions sont utilisées dans l'analyse représentant un plus grand pourcentage de la variance des données.



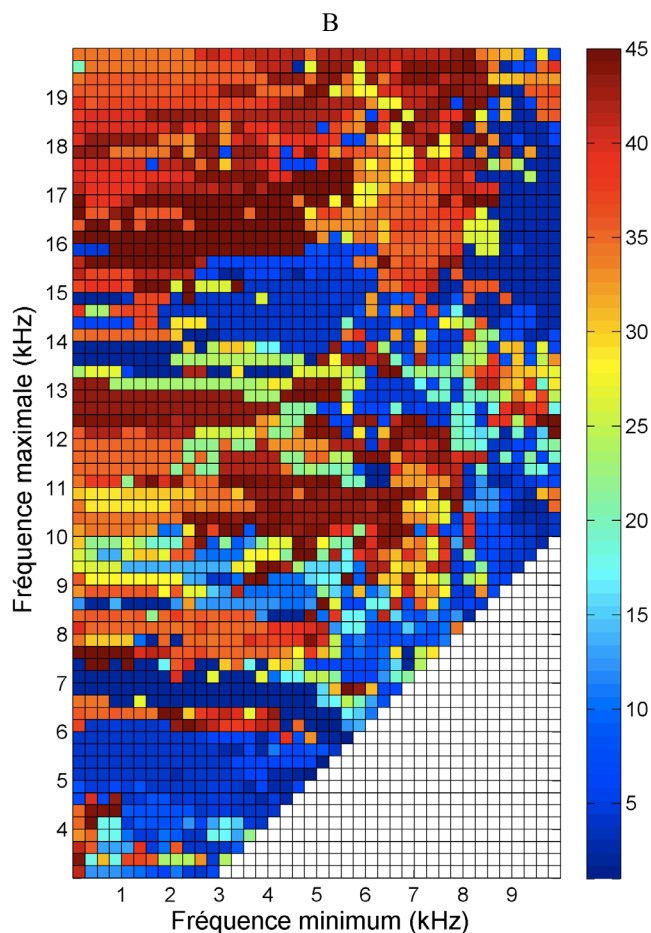


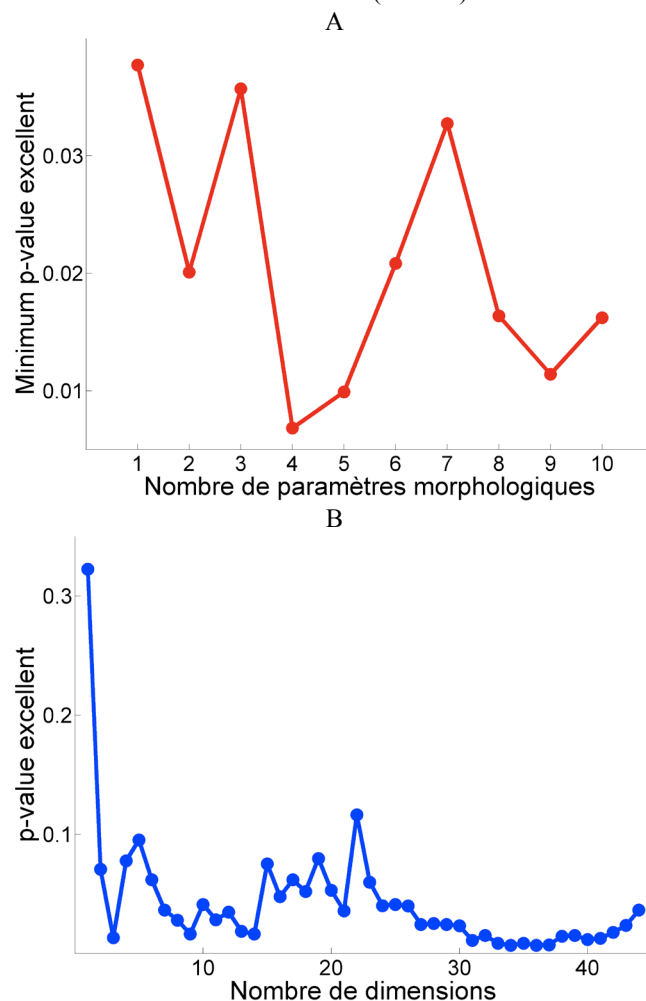
Figure 4 : (A) les minimums des valeurs *p-value excellent* pour les différentes bornes de fréquences. Les valeurs des *p-value excellent* correspondant à une couleur sont affichées sur l'axe vertical sur la droite de la figure sur une échelle logarithmique. (B) Le nombre de dimensions utilisées pour les différentes bornes de fréquence. Les valeurs de dimensions correspondant à une couleur sont affichées sur l'axe vertical sur la droite de la figure.

3.2 Prédiction de HRTF avec paramètres morphologiques

Avec le choix des paramètres morphologiques les plus importants, de multiples régressions linéaires ont été faites afin de voir s'il était possible de prédire le classement des HRTF venant du test d'écoute pour chaque sujet. Pour cela, il a fallu répéter la validation des espaces multidimensionnels pour chaque sujet en retirant leurs données de la matrice des DTF. Les coordonnées de chaque sujet dans l'espace multidimensionnel ont été prédites par une régression entre les paramètres morphologiques les plus importants et chaque dimension (coordonnées), dans l'espace. Une fois que les coordonnées du sujet ont été calculées, la validation s'est faite de la même façon pour chaque nombre de dimension.

La Figure 5 présente les valeurs de la métrique *p-value excellent* pour la validation de la régression en utilisant l'espace multidimensionnel de l'ACP des DTF qui était le plus performant. L'espace multidimensionnel le plus performant correspondait aux bornes de fréquences optimales en utilisant 35 dimensions (Figure 4). Les minimums des *p-value excellent* sur toutes les dimensions pour les dix premiers paramètres morphologiques, classés par la méthode SVM de *feature selection*, sont affichés dans

la Figure 5(A). Le nombre de paramètres qui donnait le meilleur résultat a été calculé égal à 4. Les valeurs *p-value excellent* pour chaque dimension pour ces 4 paramètres morphologiques sont affichées dans la Figure 5(B). Le nombre de dimensions qui donnait le meilleur résultat a été calculé égal à 34. Les 4 paramètres utilisés pour cette analyse correspondaient à : x_{11} , x_2 , x_8 , et d_5 . Ce petit nombre de paramètres morphologiques est donc un premier essai pour réduire l'information nécessaire afin de choisir dans une base de données une HRTF correspondant à un auditeur à partir d'une photo de son oreille. La Figure 5(C) montre la performance de cet espace multidimensionnel avec régression en utilisant les 4 paramètres morphologiques cités. Le pourcentage *global* de HRTF classée comme *excellent* pour 43 sujets (il manquait quelques mesures de paramètres morphologiques pour 2 sujets) est affiché sur la figure en bleu en utilisant les résultats du test d'écoute. Le pourcentage de HRTF classée comme *excellent* dans les dix premières HRTF en utilisant l'espace multidimensionnel est affiché en rouge sur la même figure pour tous les sujets. La comparaison entre les deux courbes dans la Figure 5(C) démontre la performance de la méthode proposée pour un choix de dix HRTF (en rouge), par rapport à un choix de dix HRTF choisies aléatoirement (en bleu).



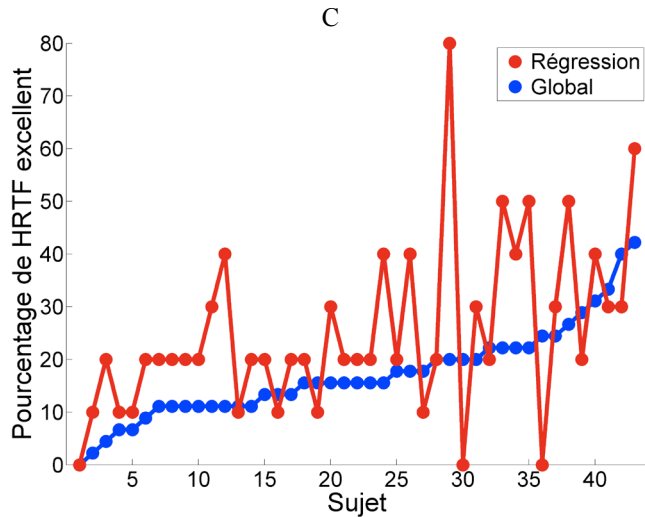


Figure 5 : validation de la régression avec paramètres morphologiques. (A) Minimum p -value excellent en fonction du nombre de paramètres utilisés pour l'ACP. (B) p -value excellent en fonction du nombre de dimensions utilisées pour une régression avec les 4 premiers paramètres morphologiques. (C) Pourcentage de HRTF classée comme excellent sur un choix de dix HRTF en utilisant la méthode proposée (en rouge) par rapport à une sélection aléatoire (en bleu).

4 Conclusion

Cette étude a validé avec succès deux méthodes pour décrire les différences de HRTF dans une base de données. Cette validation a été effectuée avec les résultats d'un test d'écoute qui s'appuie sur les préférences subjectives sans utiliser l'interpolation de HRTF. La technique proposée a été employée pour sélectionner les paramètres morphologiques les plus importants, ainsi que des bornes de fréquence pertinentes. En comparaison au classement aléatoire, la technique décrite dans cette étude a permis de prédire de façon statistiquement plus significative les classements subjectifs des HRTF. Cette méthode, toujours un travail en cours, pourrait donc être utilisée pour faciliter la sélection d'une HRTF dans une base de données. Elle permettra des applications de la synthèse binaurale au service du grand public en utilisant simplement une photo de l'oreille de l'auditeur.

Références

- [1] Moller H. *et al.*, "Binaural technique: Do we need individual recordings?", *Journal of the Audio Engineering Society*, 44, 451-469 (1996).
- [2] Wenzel E. M. *et al.*, "Localization Using Nonindividualized Head-Related Transfer-Functions", *Journal of the Acoustical Society of America*, 94, 111-123 (1993).
- [3] Middlebrooks J. C., "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency", *Journal of the Acoustical Society of America*, 106, 1493-1510 (1999).
- [4] Gardner M. B., Gardner R. S., "Problem of Localization in Median Plane - Effect of Pinnae Cavity Occlusion", *Journal of the Acoustical Society of America*, 53, 400-408 (1973).
- [5] Usher J., Martens W., "Perceived naturalness of speech sounds presented using personalized versus non-personalized HRTFs", *13th International Conference on Auditory Display*, Montreal, Canada, (2007).
- [6] Pralong D., Carlile S., "Measuring the human head-related transfer functions – a novel method for the construction and calibration of a miniature in-ear recording-system", *Journal of the Acoustical Society of America*, 95, 3435-3444 (1994).
- [7] Wightman F., Kistler D., "Measurement and validation of human HRTFs for use in hearing research", *Acta Acustica United with Acustica*, 91, 429-439 (2005).
- [8] Wightman F. L., Kistler D. J., "Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis", *Journal of the Acoustical Society of America*, 85, 858-867 (1989).
- [9] Kahana Y., Nelson P. A., "Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models", *Journal of Sound and Vibration*, 300, 552-579 (2007).
- [10] Katz B. F. G., "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation", *Journal of the Acoustical Society of America*, 110, 2440-2448 (2001).
- [11] Zotkin D. N. *et al.*, "Fast head-related transfer function measurement via reciprocity", *Journal of the Acoustical Society of America*, 120, 2202-2215 (2006).
- [12] Guillon P., "Individualisation des indices spectraux pour la synthèse binaurale : recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF", *thèse de doctorat*, Université du Maine (2009).
- [13] Middlebrooks J. C., "Individual differences in external-ear transfer functions reduced by scaling in frequency", *Journal of the Acoustical Society of America*, 106, 1480-1492 (1999).
- [14] Seeber B., Fastl H., "Subjective selection of non-individualhead-related transfer functions", *9th International Conference on Auditory Display*, Boston, (2003).
- [15] Iwaya Y., "Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears", *Acoustical Science & Technology*, 27, 340-343 (2006).
- [16] Wightman F., Kistler D., "Multidimensional scaling analysis of head-related transfer functions", *IEEE Digital Audio Workshop*, 98-101 (1993).
- [17] Zotkin D. N. *et al.*, "Rendering localized spatial audio in a virtual auditory space", *IEEE Transactions on Multimedia*, 6, 553-564 (2004).
- [18] Jin C. *et al.*, "Enabling individualized virtual auditory space using morphological measurements", *IEEE International Conference on Multimedia Information Processing*, December 13-15, (2000).
- [19] Nishino T. *et al.*, "Estimation of HRTFs on the horizontal plane using physical features", *Applied Acoustics*, 68, 897-908 (2007).
- [20] Xu S. *et al.*, "Improved method to individualize head-related transfer function using anthropometric measurements", *Acoustical Science and Technology*, 29, 388-390 (2008).

- [21] "LISTEN HRTF database", <http://recherche.ircam.fr/equipes/salles/listen/>
- [22] Middlebrooks J. C., Green D. M., "Directional dependence of the interaural envelope delays", *Journal of the Acoustical Society of America*, 87, 2149-2162 (1990).
- [23] Kistler D. J., Wightman F. L., "A model of head-related transfer-functions based on principal components-analysis and minimum-phase reconstruction", *Journal of the Acoustical Society of America*, 91, 1637-1647 (1992).
- [24] Martens W., "Principal components analysis and resynthesis of spectral cues to perceived direction", *Proceedings of the International Computer Music Conference*, 274-281, (1987).
- [25] Algazi V. R. *et al.*, "The CIPIC HRTF Database", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, October 21-24, 99-102, (2001).
- [26] Guyon I. *et al.*, "Gene selection for cancer classification using support vector machines", *Machine Learning*, 46, 389-422 (2002).
- [27] Gunn R. S., "Support Vector Machines for Classification and Regression" *Technical Report*, Image Speech and Intelligent Systems Research Group, University of Southampton, (1997).