EBU
OPERATING EUROVISION AND EURORADIO

TECHNICAL REVIEW

How to make immersive audio available for mass-market listening

14 July 2016

Rozenn Nicol, Marc Emerit, Edwige Roncière, Hervé Déjardin
(nouvOson)

FOREWORD

The main purpose of an EBU Technical Review is to critically examine new technologies or developments in media production or distribution. All Technical Reviews are reviewed by 1 (or more) technical experts at the EBU or externally and by the EBU Technical Editions Manager. Responsibility for the views expressed in this article rests solely with the author(s).

To access the full collection of our Technical Reviews, please see: tech.ebu.ch/publications

If you are interested in submitting a topic for an EBU Technical Review, please contact: tech@ebu.ch

Abstract

In March 2013, Radio France launched a new part of its website, called nouvOson [1], to convey 5.1[1] surround sound and binaural[2] audio. In November 2015, an updated version of the website was released that included new functionalities such as a binaural filters selection. The binaural technique was initially chosen to reach people who do not have a home theatre set for 5.1 reproduction as well as for mobile applications.

At the same time, Radio France became a founding member of the collaborative research project in binaural listening BILI [2], whose aim is to find an accessible way to personalize Head-Related Transfer Functions (HRTF) for mass market applications.

This article discusses the progress of the nouvOson player project since 2013 and its future outlook.

[1] 5.1 surround sound consists of five main loudspeaker channels (Left, Centre, Right, Left surround and Right surround) and a low frequency effects channel (LFE – the "0.1" in 5.1 that helps reinforce the impact of loud noises such as explosions in a movie, and is often reproduced in a subwoofer). Public Service Broadcasters' programming isn't confined to movies and they usually omit the LFE and broadcast 5.0.

[2] Binaural audio is a two channel signal that contains enough information (subtle variations in delay, amplitude and phasing of the two channels) to convey a three-dimensional soundscape when listened to on a pair of headphones. It was first observed in 1881 in Paris, when Inventor Clement Ader placed microphones on the Paris Opera House stage, for remote reproduction on a pair of telephone transducers (http://histv2.free.fr/theatrophone/theatrophone.htm). A practical problem with binaural reproduction is that every person has an individual head shape that requires a specific "flavour" of the binaural signal to optimize this 3D effect. The technical description for this "flavour" is the HRTF, as explained later in the article.

How to make immersive audio available for mass-market listening

1. Introduction

In March 2013, Radio France launched a new part of its website, called nouvOson [1], to convey 5.1 and binaural sound. In November 2015, a second version of the website was released including new functionalities such as a binaural filters selection. A binaural technique was initially chosen to reach people who do not have a home theatre set for 5.1 reproduction as well as for mobile applications.

At the same time, Radio France became a founding member of the collaborative research project in binaural listening BILI [2], whose aim is to find an accessible way to personalize Head-Related Transfer Functions (HRTF) for mass market applications.

This Technical Review article discusses the progress of the nouvOson player project since 2013 and its future outlook.

2. Initial architecture

NouvOson was launched to allow experimentation in the production and distribution of spatial sound. New modes of listening and especially the resurgence of headphone use convinced us of the value of binaural audio. For this reason, in the context of our research and innovation, we focused on the best way to distribute binaural audio via the web.

This is what motivated us to develop a player for audio production for broadcast in 5.1 and binaural.

A total of 98% of the binaural sound provided by nouvOson is created by encoding a 5.1 sound file with a generic HRTF. When someone listens to a programme on the site, they can choose between 5.1 or binaural rendering. In 2013, the 5.1 audio codec that was available on nearly all web browsers and on all operating systems (Windows, MacOS, iOS, Android...) was the MPEG HE-AAC codec. Today OPUS, Dolby Digital or Dolby Digital+ are alternative codecs that are now also supported by web browsers. Whilst OPUS and HE-AAC are supported by many web browsers, the Dolby solutions are supported in Microsoft Edge on Windows 10 and Safari on MacOS, only.

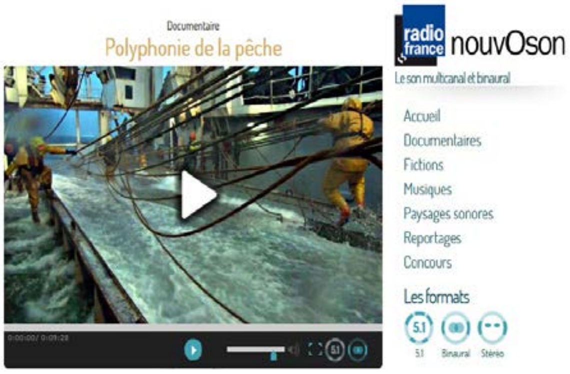


Figure 1 – Initial version of the nouvOson player

The binaural encoding parameters were chosen based on the expertise of Francois Ragenard, a Radio France technical supervisor. The Fraunhofer MPEG HE-AAC codecs were used at 192 kbit/s for 5.1 audio files and AAC at the same rate for binaural files, both with constant bit rate (CBR) coding.

For the moment, nouvOson's 5.1 and binaural audio files are predominantly available only as podcasts, but occasionally, live events are available in binaural.

The loudness and level of the audio files is EBU R128 compliant[3] [3]:

1) The 5.1 files, which are assumed to be listened to in a quiet home theatre environment, are set to –23 LUFS.
2) The binaural files are set to -15 LUFS to fit in with mobile practices.

Two points are important to note in this configuration:

1) For each programme, Radio France had to produce two audio files, one in 5.1 and another encoded in binaural at different loudness levels, as noted above.
2) Whilst the use of a generic HRTF was a good way to launch the player, it is not sufficient to provide an excellent spatial experience for everyone.

3. Introduction of the web audio API in nouvOson

With the initial architecture, Radio France was trying to solve the problems of the use of generic HRTF by using a library of HRTF for a first approach to personalization.

In order to improve on this first approach, it was necessary to introduce the HTML5 Web Audio API[4] to nouvOson that can allow the amplification, filtering, routeing and convolving of audio samples decoded by the browser. Using these tools in the API, Marc Emerit and Michael Pontiggia from Orange Labs have constructed a binaural engine that Radio France is now introducing into its latest nouvOson player.

[3] See EBU Technology & Innovation: tech.ebu.ch/loudness

[4] See Web Audio API: http://webaudio.github.io/web-audio-api/

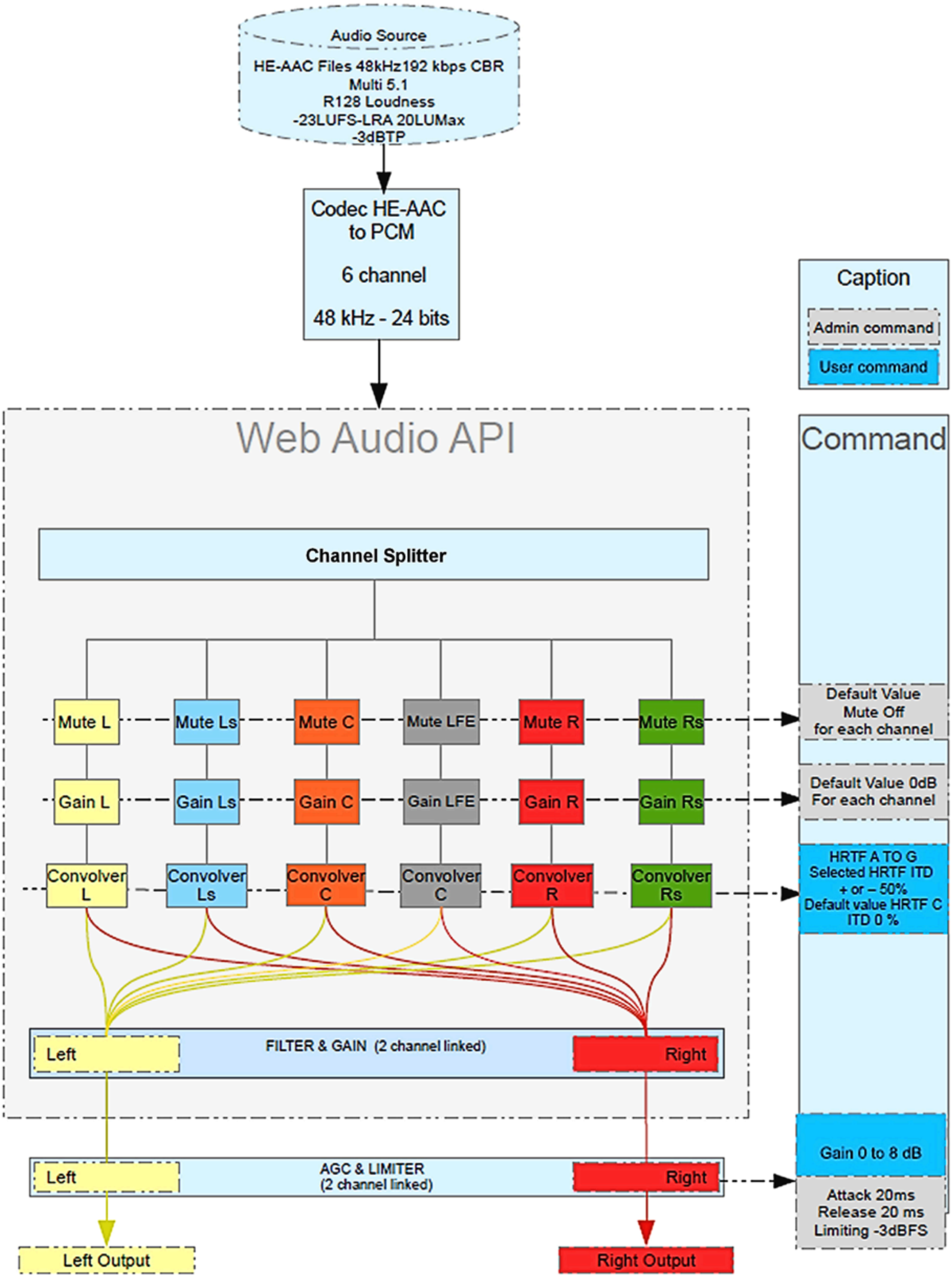
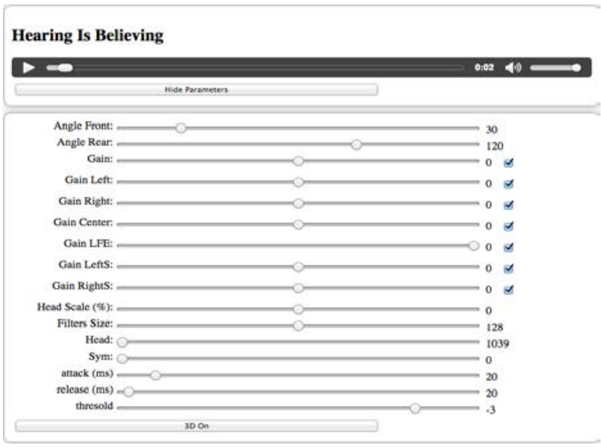


Figure 2 – Functional replay diagram of the nouvOson V2binaural encoder

With this process, it is only necessary for Radio France to produce a 5.1 audio file. The binaural file is created on-the-fly in the listener’s HTML5 capable browser. The first prototype was very basic and was used only for internal trials.



Hearing Is Believing

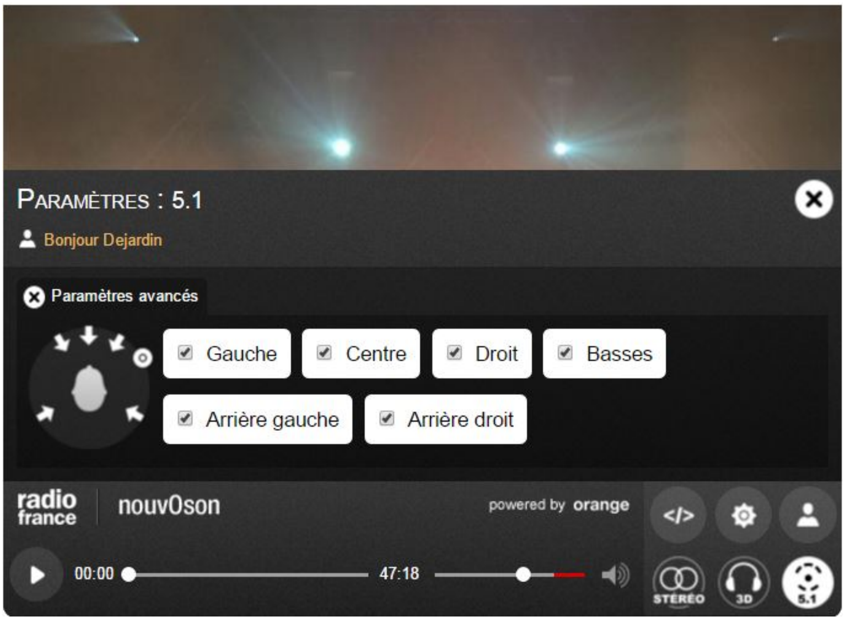
Angle Front: 30
Angle Rear: 120
Gain: 0 ☑
Gain Left: 0 ☑
Gain Right: 0 ☑
Gain Center: 0 ☑
Gain LFE: 0 ☑
Gain LeftS: 0 ☑
Gain RightS: 0 ☑
Head Scale (%): 0
Filters Size: 128
Head: 1039
Sym: 0
attack (ms) 20
release (ms) 20
thresold -3

Figure 3 – First prototype of the player offering customization

Once the quality of the player was assessed and approved, a better integration of the API was implemented, thanks to Julien Decaudin, Jean-François Mougnot and Guillaume Baret from Radio France.

With this player, if the listener does not select the listening format, the WebAudio API automatically detects if a 5.1 loudspeaker system or a pair of headphones is being used. In the latter case, a binaural version of the 5.1 programme is rendered in the browser and a generic HRTF is automatically applied.

In the case of 5.1 rendering, the listener can use an online tool to check that his loudspeakers are receiving the correct channels for reproduction, as shown in the following figure.



PARAMÈTRES : 5.1

Bonjour Dejardin

Paramètres avancés

☑ Gauche ☑ Centre ☑ Droit ☑ Basses
☑ Arrière gauche ☑ Arrière droit

Figure 4 – Online tool to set up 5.1 home theatre systems, stereo and binaural parameters

The LFE (low frequency extension; the 0.1 in the 5.1 system, often reproduced in a subwoofer) is separate and the listener can manually apply the +10 dB gain on this channel that is required for proper reproduction in home theatre systems.

If needed, there is a classical stereo down-mix available:

Left downmix= L + C-3dB + Ls-3dB

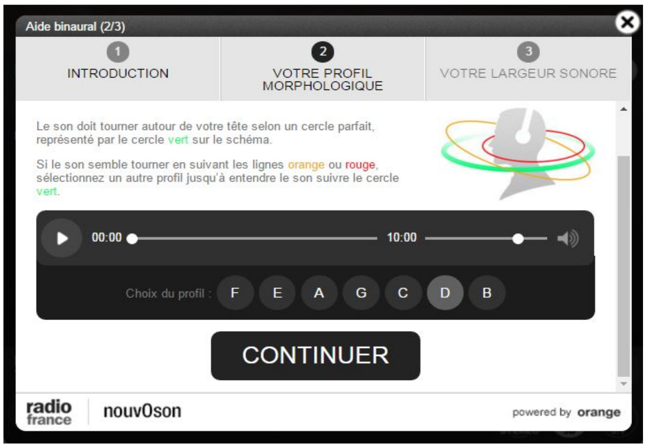
Right downmix=R + C-3dB + Rs-3dB

4. Level management in the Web Audio API

The level reproduced by the new player defaults to 23 LUFS for both 5.1 and binaural reproduction. On the volume slider there is a mark between the light grey and the red zone that indicates -23 LUFS. The listener can adjust the volume cursor downwards and upwards to vary the gain to a maximum of –15 LUFS. In order to stay compliant with EBU R128 in mobility, there is a smart AGC (Automatic Gain Control) inside the Web Audio API that follows the signal peaks on every sample and ensures that it remains under -3 dB peak (true peak is not yet available).

5. Evolution of the Web Audio API for binaural listening

If the listener wishes to personalize (and therefore improve) his binaural rendering, he can perform a short localization test by clicking on the cog-wheel button seen in Figure 4, which opens the following panel.



Le son doit tourner autour de votre tête selon un cercle parfait, représenté par le cercle vert sur le schéma.

Si le son semble tourner en suivant les lignes orange ou rouge, sélectionnez un autre profil jusqu'à entendre le son suivre le cercle vert.

Figure 5 – Test window to find the best HRTF for the listener

The user is presented with the choice of seven different HRTFs (termed morphological profiles, labelled A to G) that he chooses between to maximise the immersive experience obtained with a short test sound that orbits the head of the listener. On the right side of the panel, the green circle in the graphic shows the desired binaural rendering of the test. The red and orange circles indicate false HRTF situations to help the listener understand what they have to achieve to optimize their experience of nouvOson.

Hervé Dejardin designed this test environment. It was done using spectral and temporal properties that allow easy location of a sound orbiting the head for about five seconds. It took many different trials with different sounds to discover the most appropriate test sequence.

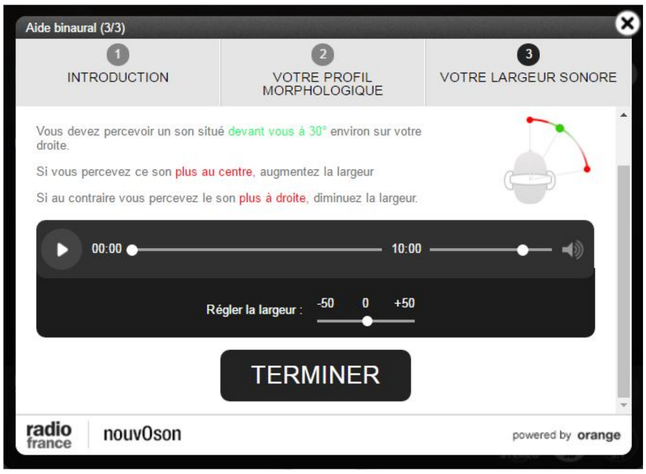
Hervé found that the required sound must fill a wide spectrum in order to stimulate both the Interaural Time Difference (ITD) under 2.5 kHz and the Interaural Level Difference (ILD) above 2.5 kHz.

On the advice of Rozenn Nicol from Orange Labs, he added some artificial reverberation, which gives depth and more medium spectrum components to improve the test. Nicol also chose the seven HRTF from the IRCAM HRTF database.[5]. Her selection process is described below.

The choice of a continuous moving source instead of several discrete fixed audio points in the same space compensates for the absence of head tracking.

The test is constructed with the virtualization of eight loudspeakers around the head, only, and it is very important that the speed of movement of the sound is consistent with this.

By clicking the "continuer" button (see Figure 5) the second panel of the personalization test appears (see Figure 6). Here the listener can adapt the ITD of his chosen HRTF by adjusting the "Régler la largeur" (set width) slider. The listener must perceive the sound at 30° to the right as indicated by the green point on the graphic.



Vous devez percevoir un son situé devant vous à 30° environ sur votre droite.

Si vous percevez ce son plus au centre, augmentez la largeur

Si au contraire vous percevez le son plus à droite, diminuez la largeur.

Figure 6 – Test window to adjust the ITD suited to the listener

This bit of the test is not the easiest to adjust because even though sound engineers can understand and hear the difference it makes, it is not clear that a non-professional listener will be able to do so with the same ease.

[5] http://recherche.ircam.fr/equipes/salles/listen/download.html

After the test, the listener saves his preferences to a personal profile that is stored in a cookie for use every time nouvOson is visited with that browser.

6. Future outlook

Several issues are currently being addressed, for instance,

- The LFE rendering in the binaural case is achieved by manually adding +10 dB in the Web Audio API.
- The 80 Hz low pass filtering is inside the MPEG HE-AAC encoder.
- The LFE channel is not "binauralised".

We have begun to observe that listeners are starting to expect a more personalized experience. 5.1 surround sound is not the end of the story and there will be the need for many more channels for UHDTV and the commensurate need to manage many more virtual sources in binaural as well. All this has led us to object-oriented mixing and new ways of producing audio (possibly the subject of another article).

The Web Audio API will be a good tool to manage all these future applications.

7. HRTF personalization for mass-market binaural applications: a "ready-to-wear" solution based on a selection of seven HRTF sets

As a first attempt to propose a tool for HRTF personalization in the context of mass-market applications, we chose to start with a simple solution, both easy to implement and easy to use. A selection of seven sets of HRTFs was presented to the listener who identified his(her) preferred set by comparative listening. But, to provide proper customization in this manner, the critical issue is to get the "ultimate" selection of HRTFs from which to choose. It is still not clearly known how non-individual HRTFs are perceived, e.g. what are the perceptual dimensions underlying the discrimination between two sets of HRTFs?

Most often, HRTFs are assessed by localization tests[6], which show that the use of non-individual HRTFs leads to poor externalization (and even In-Head-Locatedness[7]), poor localization in elevation (where the pinna is the main morphological contribution), an increase in the rate of front-back and top-bottom reversals[8], and a spatial shift of frontal sources. These results account only for the perception of spatial information. The potential other dimensions responsible for perceived differences between HRTFs are therefore missed. For instance, timbral artefacts are not taken into account. In addition, most of the previous studies did not consider a large number of non-individual HRTF sets.

To give a deeper insight into this question, an experiment was carried out to investigate the perceptual attributes governing the perception and the discrimination

[6] Wenzel 1993, Moller 1996, Begault 2001

[7] Blauert 1996

[8] Wenzel 1993

of HRTFs. It relied on dissimilarity judgments between stimuli in order to compute a dissimilarity matrix that allowed us to build up the perceptual space in which the stimuli are placed in accordance with their perceived dissimilarities. To achieve this, Multi-dimensional Scaling (MDS) [12] analysis was used. The sound stimuli used in the experiment were derived from an original multichannel 5.1 audio excerpt which was converted to binaural format. In order to achieve the most representative space, it is required that the set of stimuli covers the whole range of potential variations that can be expected from the variable under assessment, which means, in our case, the use of a HRTF database composed of a large number of individuals. Forty-six different versions of this excerpt were obtained by convolving it by the forty-six sets of HRTFs of the IRCAM[9] "Listen" database.

Before, HRTFs were pre-processed the response of the measurement system was compensated for, and the individual Interaural Time Delay (ITD) was replaced by an average ITD which was computed as the mean of the 46 individual ITDs for each direction. Finally, the loudness of each binaural stimulus was corrected to achieve equal loudness for all the versions, in accordance with the "N10" indicator proposed by Zwicker and Fastl.[10]

The experiment aimed at measuring the dissimilarity between all possible pairs taken from the set of M=46 stimuli, which leads to a total of N=1035 pairs. Instead of comparing all these pairs, which would require a prohibitive amount of time, an alternative method, the "Similarity Picking with Permutation of References" (SPPR)[11], was used. The idea of SPPR is to present a reference stimulus and a set of P stimuli, which are all taken from the M stimuli under assessment, and to ask the subject to identify among the set of P stimuli, the K stimuli which he (she) perceives as the most similar to the reference.

One advantage of the method is that each stimulus plays the role of the reference. In the case of visual stimuli, the (P+1) stimuli (the set of P stimuli in addition to the reference stimuli) are presented simultaneously, which allows the comparison of a large number, P, of stimuli in one trial. Thus, the total number of trials required to compare all the M stimuli is considerably decreased, expediting the overall experiment. However, in the case of audio stimuli, the comparison cannot be simultaneous and is inevitably sequential. What's more, because of the limitation of the auditory memory, the number P of stimuli under comparison in one trial should be not too high. In *Michaud 2013*[11], it is shown that P=3 sound stimuli is a good compromise to achieve a significant reduction of the total duration of the experiment, while preserving the reliability of the judgment. The participant's task is to select the K=1 stimulus that he (she) judges the most similar to the reference.

[9] http://recherche.ircam.fr/equipes/salles/listen/

[10] Zwicker 1999

[11] Michaud 2013 (SPPR)

Ten subjects took part in the experiment. They can be considered expert listeners; indeed the group was composed of sound engineers from Radio France, France Télévisions, the Conservatoire National Supérieur de Musique et de Danse de Paris (CNSMDP) and researchers in spatial audio from Orange Labs and the Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI, University of Orsay). They are therefore used both to critical listening and are familiar with 3D audio content, and particularly binaural sound. The experiment ran in five different locations in Paris (Radio France, CNSMDP and LIMSI), in Rennes and in Lannion. The audio equipment used for rendering the binaural sound stimuli (Focusrite Scarlett 6i6 sound card and Sennheiser HD 650 headphones) was identical.

An individual dissimilarity matrix was obtained for each participant. All of the individual dissimilarity matrices were then averaged, leading to one single mean dissimilarity matrix, which could then be processed by Multi-dimensional Scaling (MDS) [12] analysis.

The objective of MDS is to propose a spatial distribution of the stimuli under comparison. In this distribution it was intended that the distances between the stimuli are representative of their dissimilarities. By iterative process, the MDS algorithm builds up the spatial distribution which best fits the dissimilarity judgments. Thus this spatial arrangement (organization) of the stimuli reveals the structure of the perceptual space governing their perception.

The perceived dissimilarities between all the pairs of these 46 sets of HRTFs can be used by a hierarchical analysis to build up a dendogram (see Figure 7) which helps to visualize the dissimilarities between the HRTFs. In a dendogram, the length of the path between two sets is proportional to their perceived dissimilarity; the shorter the path, the more similar the sets are. We proposed to use this dendrogram to eliminate the most similar sets of HRTFs, in order to extract the most dissimilar sets.

Our objective was to select a reduced number of sets which are still representative of the variability of the 46 sets, but with minimal redundancy. Arbitrarily, it was decided to select only 7 sets. A threshold of minimal dissimilarity was defined and fixed to 0.6 (see Figure 7), which allowed us to isolate 7 clusters of HRTF sets. Within each cluster, the dissimilarities between the HRTF set is below this threshold. For each cluster, a representative set is extracted by choosing the set which is the closest (in terms of perceived dissimilarities) to the barycentre of the members of the cluster.

This resulted in a selection of 7 sets of HRTF which represent the extent of perceived dissimilarities of non-individual HRTFs, and which is assumed to be able to provide one suitable set of HRTFs to any listener taken from a wide range.

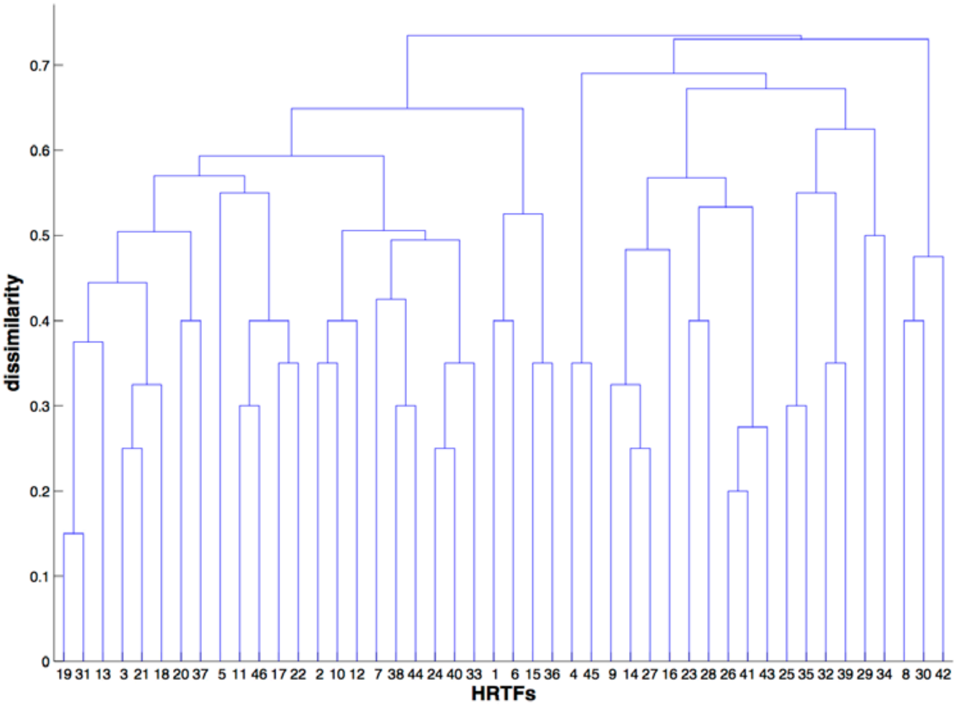


Figure 7– Dendogram built from the dissimilarity matrix of the 46 sets of HRTF

8. Assessment of the Quality of Experience (QoE) through a questionnaire

Beginning in November 2015, we asked visitors to the nouvOson website to fill in a questionnaire, to collect feedback about their perception of the processing.

First, they were asked to give three words to describe their experience. From a total of 272 words collected up until now (for 121 participants), 90% are positive. 13% of the words refer to "immersion", 11% to “interest”, 10% to "space", and 6% to "surprise". Second, several criteria are rated (see Figure 8). The overall rating is highly positive, except for the externalization, which scores a bit lower.

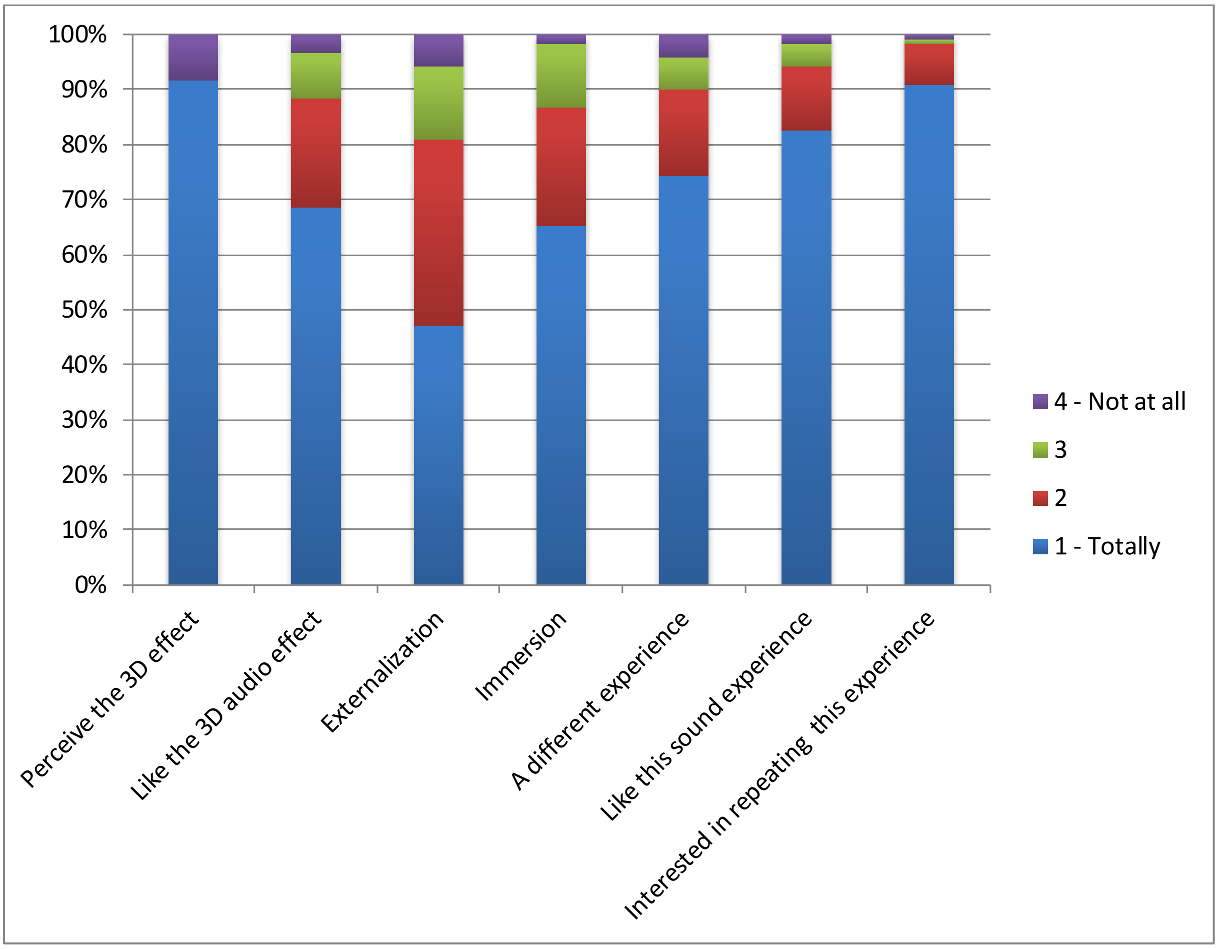


Figure 8 – Percentage achieved for each criteria of QoE

The HRTF personalization tool was also assessed (see Figure 9). Only 82% of the participants used the personalization and the results are restricted to these latter. The personalization is judged “easy” (48% for the HRTF selection, 45% for the ITD adjustment) or “pretty easy” (26% for the HRTF selection, 24% for the ITD adjustment) to adjust by more than 70% of the participants. Once the personalization is done, an improvement of the sound reproduction is perceived by two-thirds of the participants (37%: strong improvement, 28%: moderate improvement). In addition, information about the listener profile and his equipment is collected. It should be noticed that the participation is well balanced between the generations.

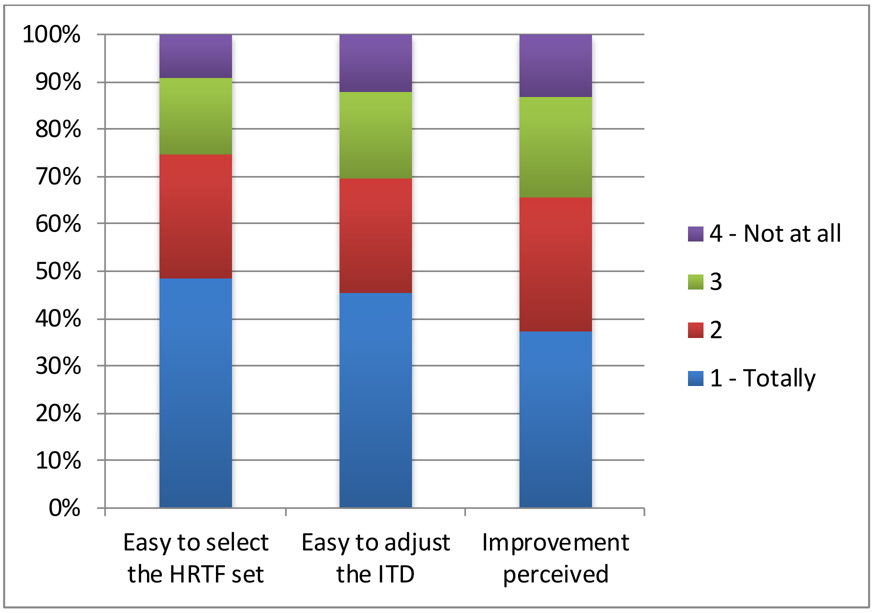


Figure 9 – Percentage achieved by parameters of HRTF customization

9. Conclusion

The new version of the nouvOson website from Radio France has been online since November 2015 and new content is regularly added. It demonstrates that it is possible ‘nowadays to broadcast 5.1 content over the web and to apply binaural down-mix for headphone listening within the browser without any prior software installation. This makes the end user experience much simpler to achieve.

In order to improve the perceived quality of immersion, we proposed to the user to customize their experience by selecting a set of binaural HRTF filters between 7 sets; the "Ready-To-Wear" concept. The Quality of Experience of the users was subsequently evaluated through an online questionnaire. The results are very positive and encouraging since the binaural rendering of 5.1 content over headphones is appreciated by 87% of the users and the HRTF personalisation has improved the experience of 67% of users. Studies into binaural listening personalization will continue within the BiLi project to improve these scores.

10. References

[1] Nouvoson: http://nouvoson.radiofrance.fr/ (Accessed on 27 June 2016)

[2] BiLi Project: http://www.bili-project.org/ (Accessed on 27 June 2016)

[3] EBU Technology & Innovation: https://tech.ebu.ch/loudness (Accessed on 27 June 2016)

[4] GitHub: http://webaudio.github.io/web-audio-api/ (Accessed on 27 June 2016)

[5] http://perso.limsi.fr/katz/Katz_publist_web.html : B. Katz and G. Parseihian, “Perceptually based head-related transfer function database optimization,” J. Acoust. Soc. Am., vol. 131, no. 2, pp. EL99–EL105, 2012, (doi:10.1121/1.3672641)

[6] IRCAM: http://recherche.ircam.fr/equipes/salles/listen/download.html (Accessed on 04 July 2016)

[7] Wenzel 1993, E. M. Wenzel, D. J. Kistler, and F. L. Wightman, “Localization using non-individualized head-related transfer functions," J. Acoust. Soc. Am., vol. 94, pp. 111-123, 1993.

[8] Moller 1996, H. Moller, “Binaural technique: Do we need individual recordings?" J. Audio Eng. Soc., vol. 44, pp. 451-469, 1996.

[9] Begault 2001, D. R. Begault, E. M.Wenzel, and M. R. Anderson, “Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," J. Audio Eng. Soc., vol. 49(10), pp. 451-469, 2001.

[10] Blauert 1996, J. Blauert, “Spatial Hearing - Revised Edition: The Psychophysics of Human Sound Localization”. The MIT Press, 1996.

[11] Zwicker 1999, E. Zwicker and H. Fastl, “Psychoacoustics: Facts and models.” Springer-Verlag, 1999.

[12] Michaud 2013, P.-Y. Michaud, S. Meunier, P. Herzog, M. Lavandier, G. Drouet d’Aubigny, “Perceptual Evaluation of Dissimilarity Between Auditory Stimuli: An Alternative to the Paired Comparison”, Acta Acustica united with Acustica, vol. 99, pp. 806-815, 2013.

10. Author(s) biographies

Dr Rozenn Nicol

After studying sound engineering at the Ecole Louis Lumière, Rozenn Nicol received a Diploma of Engineering (CNAM) in physics and acoustics and a Master of Science degree from the University of Maine (Le Mans) in 1996. In 1999, she received a PhD on a thesis entitled "Sound Spatialization over an Extensive Area: Application to Telepresence and Videoconferencing". In 2000, she joined Orange Labs, the R&D department of France Telecom, as a research engineer in spatial audio. Her work mainly concerns binaural synthesis, wave-field synthesis, and Higher Order Ambisonics. She takes part in the development and integration of sound spatialization technologies for spatial audio conferencing, audio enhancement, music or audiovisual content delivery, human computer interfaces, and virtual reality.

Marc Emerit

Marc Emerit is an experienced scientist in digital audio signal processing, with a proven track record in technology innovation and technology transfer, IP and telecommunication service development, leadership, coaching highly skilled engineers and scientists. Expert in digital audio effects, spatial audio, voice/audio coding and audio processing architecture for telecommunication service, consumer electronics, computer music and gaming.

Edwige Roncière

In 1982 Edwige Roncière began working in Radio France as a Sound Engineer. She graduated from Ecole Nationale Supérieure Louis Lumière, Conservatoire National des Arts et Métiers and Conservatoire National Supérieur de Musique et de Danse de Paris. She is currently Head of the Quality and Innovation Department in the Radio France Production Office. She leads the Radio France scientific and technical responsibilities in the BILI (Binaural Listening) and EdiSon3D R&D consortiums.

Hervé Déjardin

Herve Déjardin is a sound engineer and he currently works in the Quality and Innovation Department of Radio France for the development of multichannel and binaural sound. He also contributes to the work of the Bili (Binaural Listening) research consortium, which includes Radio France as a partner.

Published by the European Broadcasting Union, Geneva, Switzerland

ISSN: 1609-1469

Editor-in-Chief: Simon Fell

Managing Editor: Shannon Frame

E-mail: frame@ebu.ch

Responsibility for views expressed in this article rests solely with the author(s).