



Audio Engineering Society Convention Paper

Presented at the 142nd Convention
2017 May 20–23, Berlin, Germany

This paper was peer-reviewed as a complete manuscript for presentation at this convention. This paper is available in the AES E-Library (<http://www.aes.org/e-lib>) all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Optimization of interactive binaural processing for object-based audio mixing

François Salmon¹, Matthieu Aussal², Jean-Christophe Messonnier³, and Laurent Millot^{1,3}

¹*École nationale supérieure Louis-lumière, La Cité du Cinéma, 20 Rue Ampère, 93200 La Plaine Saint-Denis*

²*Centre de mathématiques appliquées de l'École Polytechnique, Route de Saclay, 91128 Palaiseau*

³*Conservatoire national supérieur de musique et de danse de Paris, 209 Avenue Jean Jaurès, 75019 Paris*

Correspondence should be addressed to François Salmon (francois.salmon3@gmail.com)

ABSTRACT

Object-based audio mixing can provide a spatial encoding for audio contents to suit various multichannel setups. This versatility requirement may involve several monitoring devices during a post-production to ensure the compatibility of a mix. Head-tracked binaural processing could then be helpful for professionals and be a new device for individuals to listen to spatialized contents given its low cost and its ease of implementation. However, this technology provides significant spectral coloration due to direction-dependent features of the processing and suffers from its current comparison to the reproduction of a stereophonic signal through headphones. Therefore, different methods for designing modified filters are proposed to optimize the binaural processing and achieve a better balance between the externalization of sounds and the timbral coloration. In order to assess the accuracy of such treatments and to identify future paths of development, perceptual tests related to the externalisation and timbre perception are presented.

1 Introduction

Binaural rendering could become a mixing solution for professionals who wish to equip themselves without investing in a complete loudspeaker setup in order to pursue the forthcoming professional changes. This mean of reproduction is also of economic interest for individuals wishing to feel a surrounding sound space while having a rendering close to those they usually encounter: stereophony on speakers or with a headset.

In the current state of technological advances, the binaural can not compete with a natural listening on speakers in terms of perception of space, precision of localization or audio scene stability. Moreover,

with the same listening device (a headphone), for binaural signal noticeable spectral colourations occur in comparison with the stereophonic signal. Yet, with object-based audio mixing, the same audio content is intended for diffusion on several rendering systems. Such reproducing devices must be able to faithfully render the audio content mixed by a sound engineer. Binaural must ensure that fidelity to the mix and provide a sensation of space.

This technology is considered here as a mean to reproduce 3D audio contents and not as *the* technology that reproduces our true sensation of space. It simply presents a powerful restitution tool given its low cost and its ease of implementation.

In the following section of this paper, we question the integration of an interactive - assisted by a head-tracker - binaural monitoring, in the context of an object-based audio production. In section 3, we study how to improve the binaural rendering in order to minimize the spectral colouration introduced by binaural filtering. The last sections consist of discussing listening tests performed over the exposed processing methods in order to identify further axes of development.

2 Binaural monitoring of object-based audio content

Rather than performing a specific mix dedicated to a given setup, object-based audio mixing aims at restoring a mix in a speaker-agnostic format - compatible with a wide variety of reproduction systems - thanks to the use of meta-data containing the coordinates of each sound object. The resulting audio scene is then more or less spatially defined according to the limits of the given playback system.

This spatial encoding of sound sources aims to diffuse a sound production in an accurate and faithful manner despite the variability between the possible listening devices. This requirement of versatility influences the whole production chain, from the sound recording to the monitoring. The latter can require the use of several playback systems in order to check and ensure the compatibility of the mix with different setups. To that extent, a binaural rendering engine could constitute a useful and flexible tool.

This technology allows a precise description of the sound space around the listener thanks to advanced interpolation methods whereas with stereophony, spatialization is less favourably restored due to its low number of channels. One can then fix "bounds" of perception of the mix in which the listener will evolve according to the system he uses : from a very masking setup - stereophony - to a less masking render with binaural.

Binaural rendering could be of valuable help for monitoring an object-based audio production, however when one uses a binaural rendering engine, some phantom sources are not as well perceived as the ones located between speakers, the HRTF are not necessarily personalized and it is not necessarily possible to take into account head movements. Consequently, it is necessary to be vigilant and to take some precautions. With the right operating environment, binaural monitoring can prove to be relevant.

2.1 Head-tracking

A head-tracker allows an increased perception of sounds in terms of localization and externalization. The use of a head-tracker has been shown to greatly reduce front / back confusions as well as mislocalizations for frontal azimuths and elevations [1].

Moreover, it also permits to separate localization information of the HRTF from the spatialized sound object itself. By moving the head, one sweeps a part of space positions involving a certain number of filters. The head movements permit to apprehend the common information in the sound, which is the object itself without localization indices. This phenomenon then allows partial discolouration of the signal.

2.2 Well-suited audio content

Since binaural processing does not simulate perfectly our natural sound perception, phantom sources are not as precise and stable as in a diffusion using speakers. The use of sound involving both ΔI and ΔT time differences, the use of decorrelated objects, or a dense spatial description in the significant zones are all solutions allowing a robustness with respect to these phantom sources [2].

Within the Bili Project (a French collaborative research project in binaural listening), some recording systems have been identified to be relevant for binaural rendering through subjective comparisons [3].

2.3 Anechoic filters and reverberation

When mixing with the headphones, special attention must also be given to reverberation. Indeed, mixing with anechoic HRTF will tend to favour a high level of reverberation which, when rendered on loudspeakers, will be exacerbated and may give rise to problems of compatibility. This effect can be lessened if the binaural engine used includes the simulation of a listening room effect. In addition, the use of a room effect favours externalization [4]. However a trained listener can be able to anticipate the influence of anechoic filters with respect to reverberation.

2.4 Training

Any reproduction system requires minimal training in order to use most of the possibilities offered. Even stereophony requires training in order to gain localization accuracy. In binaural, training is essential in order to assimilate localization distortions and help externalization.

Theile's studies make it possible to suppose that we possess a database of personal localization indices acquired and memorized throughout our life [5]. When we perceive a sound in space it is decoded according to this database in order to determine its incidence. Through training, the plasticity of the auditory system can permit to adapt this decoding for different localization indices. [6].

3 HRTF and timbral effects

During mixing, in order to deal with masking, levels, relief and depth, the frequency content of the sources is an essential component to take into account. The noticeable colouration of the binaural processing seems to be an obstacle for its use in this stage of post-production.

Furthermore, this technology suffers from its comparison with stereophony on headphones. Given our cultural habits with this listening device, it is difficult to accept the loss of timbre quality for the benefit of spatial indices.

Despite being a natural component of the propagation of sound to a listener's ears, these frequency features significantly change the timbre of sound content, which can be problematic for professional applications. Moreover, they do not seem very effective as externalization presents problems in the frontal and rear areas of the space. Without head-tracking the latter is not significantly perceptible outside the areas located on the sides. The colouration thus presents a disadvantage without justifying a real asset. It is then necessary to find a better compromise between the sensation of space and the fidelity to the timbre, by altering the externalization which seems less relevant than timbre for mixing.

3.1 The frontal area

In the context of non-individualized listening, the externalization of sound sources is more difficult to perceive in the frontal and rear areas. The spectral indices contained in these areas do not seem to be sufficiently relevant and the use of head-tracking associated with training can improve externalization.

Besides, sound objects are usually concentrated on the front scene, while the surround or zenithal channels are used for atmospheres, acoustics of the room, effects or artificial reverberations. We therefore assume that the alteration of spectral indices is not necessarily critical for the sensation of space in the frontal area.

From these observations, different methods will be proposed to minimize spectral colouration in the frontal area only.

3.2 Removal of spectral indices

A first treatment of the frontal zone consisted in keeping the interaural time and level differences (ITD and ILD) of the HRTF while eliminating the spectral indices contained in the filters. A unique sample (digital dirac) was retained to produce a flat spectrum filter with the same original filter's ITD and ILD. The amplitude of this single sample is calculated so that the modified spectrum's norm is equal to the norm of the original one, $L_O(\theta, \phi)$:

$$L_O(\theta, \phi) = \sqrt{\frac{1}{\omega_1 - \omega_2} \int_{\omega_1}^{\omega_2} |H_O(\omega, \theta, \phi)|^2 d\omega}. \quad (1)$$

The frequency band $[\omega_1; \omega_2]$ should correspond to frequencies where the human hearing is the most sensitive so that the calculated level has a sense from the point of view of perception.

The frequency content corresponds to a constant whose amplitude is a function of the sound's incidence angle. The resulting HRTF makes it possible to locate the sound sources without providing localization information allowing externalization. It remains to be seen if the brain allows to externalize thanks to head-tracking. These filters were subjected to the appreciation of their externalization and spectral qualities in the listening tests presented Sec. 4.

3.3 Interaural Spectral Differences

Since the use of a unique coefficient to filter sound sources does not present satisfactory externalization, it has been considered to take into account spectral indices in order to improve the feeling of externalization in the frontal area. This approach relies on the use of spectral differences between left and right filters only - without considering their entire spectrum.

According to studies conducted by Morimoto [7] or Hofman [8], beyond the frontal area, the spectral indices can be considered monoral: the auditory system extracts the spectral information from the right and left signals independently. The role of the ipsilateral ear, closest to the sound source, is predominant for azimuths greater than 30° . On the other hand, in the frontal area the auditory system extracts spectral information from both ears. The method described here consists in extracting only the interaural spectral differences (ISD).

Figure 1 represents the interaural spectral difference of left and right HRTF for a position of 20° in azimuth and 0° in elevation. Up to 16 kHz, the amplitude of the left and right spectrum have a dynamic of 19 dB and 18 dB respectively and their level difference reaches 15 dB. Moreover, we observe the rugged appearance of the frequency content from 1 kHz.

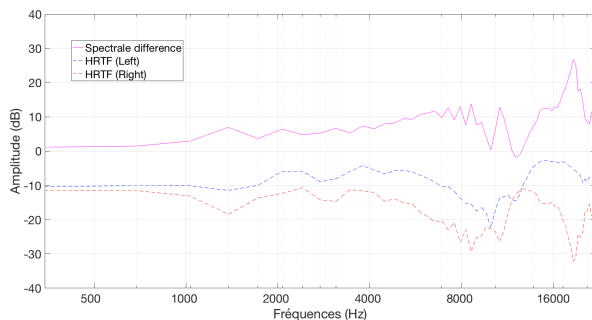


Fig. 1: Illustration of left and right HRTF from incidence (20° , 0°) along with the associated spectral difference.

Figure 2 represents the right and left processed HRTF containing the interaural spectral differences only. These are obtained by allocating half of the spectral difference on the left filter and the other half on the right filter. Let $|H_{OL}|$ and $|H_{OR}|$, be the magnitudes of the original left and right filters. The computing of

the respective modified magnitudes $|H_{FL}|$ and $|H_{FR}|$ is undertaken such as :

$$|H_{FL}| = \frac{1}{L_M(\theta, \phi)} \cdot \sqrt{\frac{|H_{OL}(\omega, \theta, \phi)|}{|H_{OR}(\omega, \theta, \phi)|}} \quad (2)$$

and

$$|H_{FR}| = \frac{1}{L_M(\theta, \phi)} \cdot \sqrt{\frac{|H_{OR}(\omega, \theta, \phi)|}{|H_{OL}(\omega, \theta, \phi)|}} \quad (3)$$

where $L_M(\theta, \phi)$ is the mean of the original left and right filters norms $L_{OL}(\theta, \phi)$ and $L_{OR}(\theta, \phi)$, defined by the equation 1. It has been chosen to use the average norm of the two filters since each of the two modified HRTF can not be standardized independently insofar as the level ratios must remain unchanged in order to respect the original interaural spectral difference.

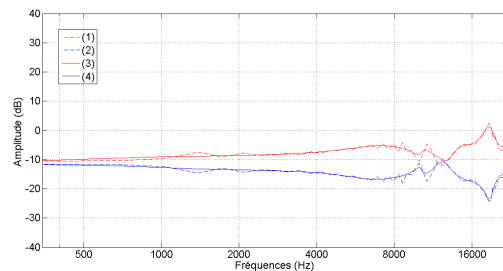


Fig. 2: Spectral difference allocated to left (1) and right (2) filters along with the corresponding 10th order polynomial approximations (3)(4).

In order to smooth further the frequency responses, an approximation by a polynomial of order 10 was used. The approximation uses the least squares method and the Yule-Walker equations to approximate the specified frequency response.

Thus the spectral differences are respected while the filter profile has fewer deviations in amplitude. Figure 3 shows that along different azimuthal incidences, the spectral dynamics are lessened by almost a half.

In order to validate the use of this method, it is necessary to ensure that an auditor is able to externalize sound sources using ISD only. Two common interpretations of the utility of spectral indices consider either

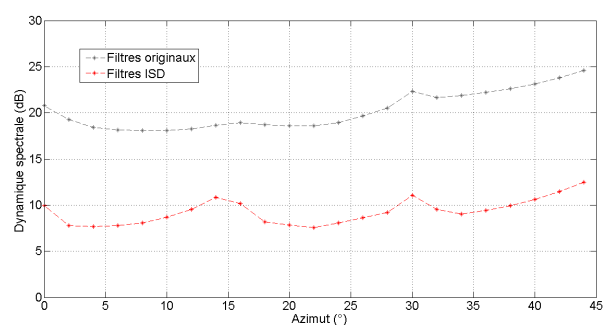


Fig. 3: Spectral dynamics of the contralateral ear's filters (computed up to 16 kHz) as a function of the incidence in the azimuthal plane.

the auditory system analyzes the troughs and bumps of HRTFs [9], or it analyzes the incident spectrum in its entirety [10, 5]. The approach presented here does not respect these two interpretations insofar as the troughs and the peaks are modified and the entire spectrum is not respected. The hypothesis that the use of ISD only in the frontal area, associated with head-tracking, allows a satisfactory externalization has been subjected to the tests presented Sec. 4.

3.4 On the spectral variation of the quadratic mean

Juha Merimaa proposed a method developed in the research laboratory of Sennheiser to reduce the effects of spectral colouration due to binaural processing [11]. This method is based on the reduction of the spectral variation of the quadratic mean of a pair of HRTF while retaining the interaural differences of level and time. Formal listening tests showed that timbral effects can be reduced without adversely affect localization performances.

The hypothesis that the spectral variation of the quadratic mean is largely responsible for the colouration is formulated. In order to parameterize a more or less important modification, a coefficient - namely c - is introduced such that for $c < 1$, the spectral variations of the quadratic mean of the filters decrease with its value. When it is equal to zero, the resulting spectrum corresponds to a constant. The use of a coefficient $c < 1$ not only flattens the frequency response of the quadratic mean but also smoothes the response of the ipsilateral ear to the extent that it contributes more to the quadratic mean than the contralateral ear.

Figure 4 shows several HRTF filters from the ipsilateral ear for an azimuthal incidence of 30° and no elevation. It is indeed observed that the variation of the coefficient greatly influences the flatness of the frequency response. The main spectral accidents are certainly attenuated while remaining present. This method has been undertaken to process HRTF in the frontal area in order to compare its relevance to the other ones.

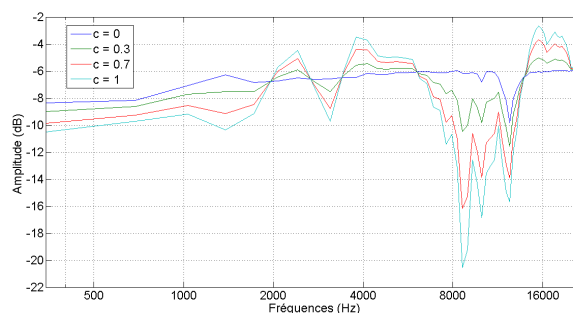


Fig. 4: Network of curves corresponding to a left HRTF processed according to a more or less important reduction of the variation of the quadratic mean of a pair of HRTF filters for a source incidence of 30° in azimuth.

4 Listening tests

Different HRTF processing methods have been undertaken in order to find filters adapted to the use of the interactive binaural as part of an object-based audio post-production. The externalization sensation and the timbre quality of binaural rendering were subjectively evaluated according to different HRTF processing and frontal areas with various widths.

Seven HRTF sets were considered during these two perceptual tests. All of them were obtained from the same filters : those of subject 1040 taken from Ircam's LISTEN database. This decision was guided by the fact that both the *Conservatoire national supérieur de musique et de danse de Paris* (CNSMDP) and the *Centre for applied mathematics of Ecole Polytechnique* (CMAP) were using these particular HRTF.

The seven different sets gathered the three different methods detailed in the previous section. We will name these methods respectively "dirac", "ISD" and "Sennheiser". Two angles of 30° and 75° - both

in azimuth and elevation - were used to define the frontal area for each methods in order to determine to what extent space windowing has an influence. The "Sennheiser" filters with a coefficient $c = 0.5$ were used to check if the results are in agreement with the following hypothesis: the alteration of the corresponding spectral contents is not sufficient to distinguish it from the externalization and colouration of the 1040 filters.

Two stimulus of 21 seconds, one for each test, were used in these evaluations. The relatively short duration of these stimuli allowed us to evaluate HRTF sets on the same musical extract. These were object-based audio mixes realised at the CNSMDP by Jean-Christophe Messonnier. The musical writing as well as the aesthetic project of the mix made it possible to choose these stimuli in coherence with the criteria to be evaluated. The frontal area consisted of 13 sound objects and 11 were used to describe the surrounding acoustics.

The listening tests presented in this section used the interactive binaural engine, *MyBino*, developed under MATLAB at the CMAP. The graphical user interface allowed the subjects to select the stimuli as well as the HRTF sets to listen to during playback. A switch enabled to bypass the binaural processing and thus listen to a stereo downmix.

The stimuli were looped and the maximum time allowed to the subjects was set to 15 minutes per test. They were asked to listen to all HRTF sets before answering the questionnaire in order to apprehend the scale of appreciation. Finally, only head movements of less than 60° from the original position were permitted since the performed processings were concentrated in the frontal area only.

Given that the developments are aimed at professional working conditions, only sound engineers and student sound engineers were among the subjects.

4.1 Perception of externalization

This test was intended to evaluate the sensation of externalization according to the various HRTF sets in comparison with the 1040 HRTF set, standing here as the reference. This comparison aimed at evaluating whether the spectral modifications are critical for the appreciation of space. We expected that the use of a spatial window wider than the frontal area alters

the feeling of externalization, because the sound objects intended to render the acoustics of a room, atmospheres or reverberation would be subject to an alteration of the localization indices. They would therefore be less able to transcribe a sensation of space.

This test was performed in order to identify a limit in the modification of the spectral content beyond which the sensation of externalization is no longer present. Subjects had to determine whether, in comparison with the 1040 HRTF set, externalization was : better, similar, of lesser quality, of much lesser quality or if they did not perceive any externalization at all.

4.2 Perception of timbre

The hypothesis to be verified is that filtering using very little spectral information would improve the similarity between binaural rendering and the stereophonic down-mix. Furthermore, it was interesting to check the influence of the width of the spatial windowing related to the perception of the timbre. *A priori*, the wider the window, the greater the similarity with the stereophony.

The down-mixed stereophonic signal, standing here as the reference, comes from an automatic reduction of the multichannel signal within two channels. This process was carried out by applying the following intensity distribution law:

$$L(\theta_i) = \sqrt{\frac{\sin \theta_i + 1}{2}}, \quad (4)$$

$$R(\theta_i) = \sqrt{1 - \frac{\sin \theta_i + 1}{2}}, \quad (5)$$

$$S(\theta_i) = 10 \log_{10} \left(10^{\frac{L_{dB}(\theta_i)}{10}} + 10^{\frac{R_{dB}(\theta_i)}{10}} \right). \quad (6)$$

Where $L(\theta_i)$ and $R(\theta_i)$ are the weighting coefficients of the sources located at the azimuth θ_i . This repartition allows to obtain a summation of the acoustic powers of the left and right channels, $S(\theta_i)$, equal to 0 dB.

Subjects were asked to measure the timbral difference with the stereophonic signal. Four mentions appeared in the questionnaire in order to qualify the colouration in relation to this reference : similar, little colourful, colourful, colourful, very colourful.

5 Results

Twenty subjects passed the perceptual tests. The mean and medians of the results are represented in figures 5 and 6. The scales of the different mentions of the questionnaire were converted to numerical scales ranging from -3 (for "very colourful" and "of much lesser quality") to 0 (for "similar") and 1 (for "better", in the case of the externalization perception).

A Wilcoxon-Mann-Whitney test was performed to analyse the results. This non-parametric statistical test makes it possible to test the hypothesis that a data distribution is the same for two given observations. This permitted to determine which HRTF sets were significantly different or not, between them and with respect to the references.

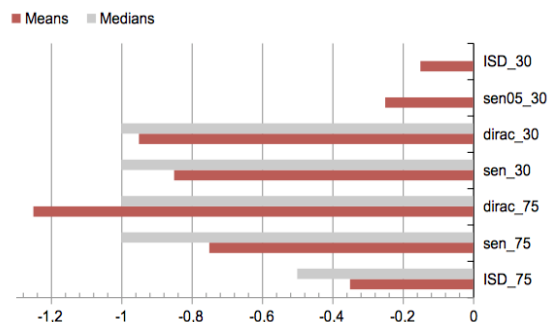


Fig. 5: Means and medians of scores obtained by the various HRTF sets, in the case of the externalization perception

The seven labels on the right of figure 5 refer to the various filters sets. They correspond to the following nomenclature: *method_angle-of-spatial-windowing* - except for the « sen05_30 » label whose "05" corresponds to the coefficient $c = 0.5$ of the "Sennheiser" processing.

The Wilcoxon test informs that the filters labelled "dirac_30", "dirac_75", "sen_30" and "sen_75" are significantly different from the 1040 HRTF set, from the point of view of externalization. The mention attributed to these filters are similar to the term "of lesser quality". The treatments performed with these methods and these spatial windowings thus seem to have altered the localization indices contained in the spectral information. Note, however, that the mean and median of the scores corresponding to externalization are relatively far from the lowest value, -3,

which corresponds to the sensation of not perceiving any externalization. The head-tracking is likely to be responsible for this minimum externalization rate. *A contrario*, the filters labelled "ISD_30", "ISD_75", "sen05_30" are not significantly different from the reference. This is consistent with the assumption that the latter two methods did not sufficiently impact the spectral indices to distinguish them from the externalization of the 1040 filters. The "ISD" method thus seems to retain the necessary localization indices for externalization as compared to other methods.

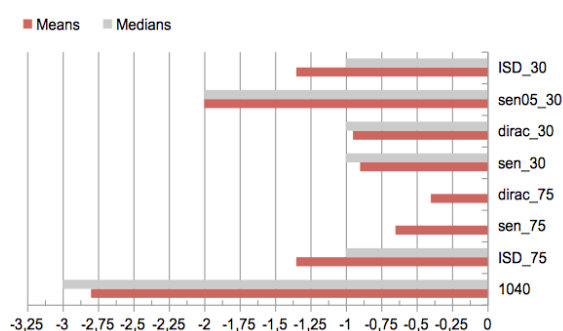


Fig. 6: Means and medians of scores obtained by the various HRTF sets, in the case of the timbre perception

The results related to the perception of timbre show a significant difference between the signals resulting from the binaural filtering and the stereophonic signal for all the HRTF sets used. None of the treatments used therefore enabled to obtain a similarity with the reference. Although HRTF set labelled « dirac_75 » and « sen_75 » have the closest mean and median scores to the mention "similar".

Likewise, the 1040 filters are significantly different to all the other HRTF sets. All the processing carried out allowed a significant modification of the spectral rendering of the 1040 filters set. Figure 6 shows the bad results obtained by the 1040 HRTF during this test.

Concerning spatial windowing, its need appears to depend on the method used. A difference in the frontal area width (30° or 75°) was significant to the «dirac» approach only. Furthermore, the use of small windowing - with an angle of 30° - did not permit to discriminate against processings carried out in the frontal area.

6 Discussion

To summarize the results of the tests, filters with the worst externalization capabilities, namely HRTF sets labelled "dirac_30", "dirac_75" and "Sennheiser", also have better timbral quality. The HRTF sets achieving the best compromise for this test are those resulting from the "ISD" method. We can see that its externalization capacity is not significantly different from the one of the 1040 HRTF set and that it has a quality of timbre close to the stereophonic signal. It remains to determine what degree of externalization we are willing to sacrifice to get closer to the sources' original timbre. It is possible that the minimal externalization observed with the «dirac» HRTF sets - probably due to the use of head-tracking - may constitute sufficient externalization. Training being a key for externalization, regular listening using these HRTF would determine whether they provide a convincing and sufficient sensation of space, avoiding any intracranial perception.

It is interesting to note that the evaluation of externalization was considered the most difficult to achieve as this notion seemed difficult to measure. The perceived differences between the signals appeared to be relatively low and determining a worst or better externalization compared to another does not seem necessarily relevant. Therefore, the use of the 1040 filters as a reference can be called into question, insofar as they do not necessarily constitute an ideal of externalization. Yet, it is necessary to find this possible ideal if it exists. An absolute appreciation of the sensation of space would have been a better choice rather than imposing a reference.

Evaluation of the timbre quality has also led to reflections on the reference that constitutes the stereophonic signal. It turned out that the latter was not necessary the most relevant to stand as reference. Indeed by its capacity of unmasking, binaural rendering - thanks to some HRTF sets - enabled to reveal a timbre that was more in adequation with the natural timbre that the subject had in mind. However, a stereophonic signal can, be regarded as a reference insofar as it is a work carried out by the mixer which implies an aesthetic project, an artistic intention. Thus, the use of an automatic downmix can be called into question given that the interaction between the sources can lead to masking issues.

In order to carry out further tests a mix up and repetitions of each evaluation can be used. Along with an

increased number of stimuli per tests and number of subjects. The number of filters to be studied can be reduced following this first test. It would be interesting to find a limit beyond which spatial windowing is significant for the "ISD" and "Sennheiser" methods. Moreover, for the latter method, the use of coefficients c different but close to 0 could be relevant to improve the externalization capacity while maintaining a low spectral colouration.

7 Summary

Through the use of a flexible compatible format, the interest in object-based audio is to respond to listeners' listening choices in a consistent and faithful manner with respect to the mix realized by the sound engineer; despite the great variability of listening conditions and playback systems. The need for compatibility between devices involves monitoring on different reproduction systems. Using stereophony or binaural rendering is two different approaches in the construction of a sound scene.

The use of these two modes of reproduction permits to review a wide range of devices behaviour related to masking effects and phantom sources. The creation of those is an important point in object-based audio mixing due to its need to adapt to poorly defined devices in terms of space rendering. The use of adequate sound recordings, decorrelated sound objects or a dense spatial description helps to prevent possible negative interference between objects.

Any reproduction system requires minimal training in order to use the most of the possibilities offered by the system. It is important to learn how to use a binaural rendering engine both in terms of audio contents and rendering conditions. This involves using head-tracking, identifying appropriate sound miking techniques and accommodate with the possible absence of listening acoustics. Moreover, given the plasticity of hearing, training is an alternative to the complex problem of individualization. In particular, it makes it possible to familiarize oneself with the tool in order to assimilate the localization distortions and to improve its capacity for externalization.

The frequency content of audio sources is an essential component to be taken into account in the realization of a mix. As we have explained in the previous sections, the binaural has spectral colourations that are

unfavorable for this task in particular, but also with regard to the compatibility with stereophony through headphones, to which it is often compared. We consider the binaural as a mean to reproduce sound which is asked to ensure fidelity to the timbre and to provide a sensation of space. Despite the impact of this colouration on the perception of space, it presents a disadvantage without justifying a real asset due to the poor capacity of externalization in the frontal and rear zones. It was then undertaken to find a better compromise between the feeling of space and the fidelity of the timbre.

Processing methods were then investigated to optimize this compromise. More or less significant spectral changes have been made on the filters located in the frontal area of space. In order to evaluate the relevance of the treatments carried out and to identify paths of development, listening tests related to the perception of externalization and timbre were realized. While the removal of spectral indices significantly alters the capacity of externalization, the use of interaural spectral differences seems to provide a good compromise for a wide frontal area. The parameterization of the method proposed by Sennheiser could be studied further in order to improve the externalization capacity while maintaining a good spectral quality. Some methods have been discarded and others deserve to be included in more formal tests to quantify the contribution of the spectral processing. Many ways of development are therefore possible in order to allow to binaural technology the possibility of constituting a reliable working tool for professionals.

References

- [1] Faure, J. and PALLONE, G., "Evaluation de la synthèse binaurale dynamique," Technical report, Tech. Rep., France Telecom, 2005.
- [2] Messonnier, J.-C., Lyzwa, J.-M., Devallez, D., and de Boishéraud, C., "Object-based audio recording methods," *Journal of the Audio Engineering Society*, 140, 2016.
- [3] Nicol, R., Gros, L., Colomes, C., Roncière, E., and Messonnier, J.-C., "Etude comparative du rendu de différentes techniques de prise de son spatialisée après binauralisation," *CFA / VISHNO*, 2016.
- [4] Kendall, G. S., "The decorrelation of audio signals and its impact on spatial imagery," *Computer Music Journal*, 19(4), pp. 71–87, 1995.
- [5] Theile, G., *On the localisation in the superimposed soundfield*, Ph.D. thesis, Technische Universität Berlin, 1980.
- [6] Blum, A., Katz, B. F., and Warusfel, O., "Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training," in *Proc. CFA/DAGA*, volume 4, 2004.
- [7] Morimoto, M., "The contribution of two ears to the perception of vertical angle in sagittal planes," *The Journal of the Acoustical Society of America*, 109(4), pp. 1596–1603, 2001.
- [8] Hofman, P. and Van Opstal, A., "Binaural weighting of pinna cues in human sound localization," *Experimental brain research*, 148(4), pp. 458–470, 2003.
- [9] Bloom, P. J., "Creating source elevation illusions by spectral manipulation," *Journal of the Audio Engineering Society*, 25(9), pp. 560–565, 1977.
- [10] van Opstal, A. and Esch, T. v., "Estimating spectral cues underlying human sound localization," 2003.
- [11] Merimaa, J., "Modification of HRTF Filters to Reduce Timbral Effects in Binaural Synthesis," *Journal of the Audio Engineering Society*, AES 127th Convention, 2009.
- [12] Salmon, F., *Optimisation du traitement binaural interactif pour le mixage orienté objet*, Master's thesis, École nationale supérieure Louis-Lumière, 2016.