



HAL
open science

Controlling Programme Loudness in Individualized Binaural Rendering of Multi-Channel Audio Contents

Emmanuel Ponsot, Hervé Déjardin, Edwige Roncière

► **To cite this version:**

Emmanuel Ponsot, Hervé Déjardin, Edwige Roncière. Controlling Programme Loudness in Individualized Binaural Rendering of Multi-Channel Audio Contents. AES Convention, Jun 2016, Paris, France. hal-01429679

HAL Id: hal-01429679

<https://hal.science/hal-01429679>

Submitted on 9 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Audio Engineering Society Convention e-Brief

Presented at the 140th Convention
2016 June 4–7 Paris, France

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.

Controlling Programme Loudness in Individualized Binaural Rendering of Multi- Channel Audio Contents

Emmanuel Ponsot^{1,2}, Hervé Déjardin², and Edwige Roncière²

¹STMS Lab (Ircam-CNRS-UPMC), 1 place Igor Stravinsky, 75004 Paris, France
Emmanuel.Ponsot@ircam.fr

²Radio France, 75016 Paris, France
herve.DEJARDIN@radiofrance.com
Edwige.RONCIERE@radiofrance.com

ABSTRACT

For practical reasons, we often experience multi-channel audio productions in binaural context (e.g. headphones on mobile devices). To make listeners benefit from optimal binaural rendering (“BiLi project”), Radio France developed *nouvOson* (<http://nouvoson.radiofrance.fr/>), an online audio platform where listeners can select HRTFs and ITDs that best fit them. The goal of the present study was to control the programme loudness (measured according to the ITU-R BS.1770 / R128 recommendations) after binauralization. To this end, we examined the influence of various parameters such as the audio content (synthetic vs. real broadcast audio), HRTFs and ITDs on loudness. We propose a dynamic process, which adapts the gain in the binauralization chain so as to control the output loudness of virtual surround audio contents.

1. INTRODUCTION

Recent technical advances in audio production and restitution systems have been accompanied by an increase in the production of multi-channel contents. Listeners can now easily experience multi-channel contents at home using surround audio restitution systems such as 5.1 systems and thus benefit from immersive audio rendering. However, listening through headphones with portable media devices such as

smartphones or laptops is also very widespread [1, 2]. In these contexts, the multi-channel productions are typically stereo-downmixed, that is to say that the n -channel signals are converted into stereo 2-channels signals. Such stereo-downmixing significantly degrades the rendering and thus the quality of the original immersive content as designed by the composer and/or the sound engineer. Instead, binaural processing techniques can be used to “simulate” the multi-channel rendering by recreating the signals of the spatialized n -speakers as they would reach listeners’ ears in a real surround context (see Figure 1).

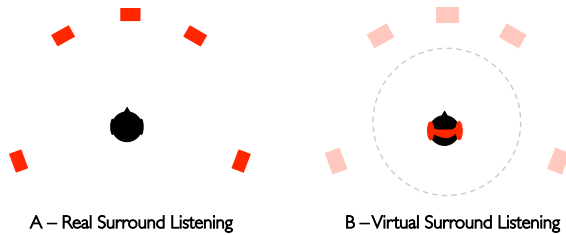


Figure 1 - Schematic representations of (A) a *real* surround listening situation with multiple speakers (here, 5) and (B) a *virtual* surround listening situation where binaural synthesis is used to recreate the immersive experience through headphones

The rendering quality at the output of such binauralization processing depends on the type of HRTFs employed. Although virtual surround reproduction is already quite good with generic HRTFs (i.e. filters supposed to be appropriate to the largest panel of listeners), near-to-perfect surround rendering can be attained if individual HRTFs are used. However, the use of individual HRTFs in online audio streaming platforms is not realistic yet, because it would require remote HRTFs measurements for anyone in the world who would like to listen to these contents.

In the context of the BiLi (Binaural Listening) project, Radio France developed *nouvOson*, an online audio platform where binaural rendering is optimized for everybody: listeners can select HRTFs from a set of pre-defined filters and adjust the ITDs so as to best fit their individual perception (<http://nouvoson.radiofrance.fr>). This procedure aims at providing listeners with an optimal rendering quality of multi-channel audio contents when they use headphones instead of real surround audio systems and when their individual morpho-characteristics (size of the head, of the torso, shape of the ears, etc.) are unknown. Because listeners are free to choose their own processing parameters, an important issue to deal with is the control of the “programme loudness” of the output signal. Indeed, the loudness being modified by the filtering process, it is necessary to compensate for this modification so that the output signal still meets current loudness standards requirements for audio broadcasting, such as ITU-R BS.1770 / R128 [3, 4].

The aim of the present study was to qualitatively assess the effects of this binauralization process on the loudness of 5.1 audio contents. Computational tests were conducted to evaluate the *gain* required to provide

binaural output signals that satisfy loudness standards for broadcasting (see Fig. 2) in different conditions resulting from the listeners’ selections (HRTF, ITD). Putting it differently, if the programme loudness of the 5.1 input signals is at -23 LUFS (the standard value in Europe), what is the gain providing a binaural signal with the same -23 LUFS loudness?

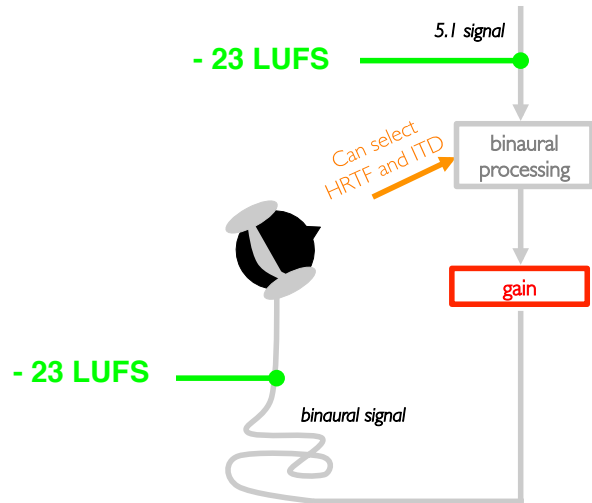


Figure 2 - The goal of this study was to evaluate the gain to apply at the end of the binaural processing chain so as to obtain an output binaural signal at the same programme loudness (in LUFS) as the input 5.1 signal.

2. COMPUTATIONAL EVALUATION UNDER VARIOUS CONDITIONS

2.1. Materials

This evaluation process was conducted with various audio contents. Real broadcast audio contents were tested for pragmatic “engineering” purpose and well-controlled synthetic stimuli were used to gain further insight into the influence of the spectral content of the production. The real broadcast stimuli employed for these tests were 5.1 musical and speech audio contents lasting around 3-5 minutes ($M = 249$ s, $SD = 50$). The synthetic stimuli were white noise and 400-Hz narrow-band noises with different center frequencies, ranging from 300 Hz to 18 kHz (the noises had constant intensity profiles and were identical on the five channels). Because we were interested in the input/output loudness difference caused by the *linear* binauralization processing chain, the loudness of the input stimuli were not a concern.

2.2. Parameters considered in the present study

Different binauralization processing conditions were examined:

- 7 different HRTFs were tested, which were *a priori* selected for being adapted to a large panel of listeners' morphologies.
- Each HRTF was tested at 5 resolutions (i.e., number of points), ranging from 512 to 32 points. It was motivated by the fact that online streaming audio platforms might use HRTFs of various resolutions. The low-resolution versions of the standard HRTFs were constructed from the initial 512-points filters by smoothing procedures not detailed in this paper. All these HRTFs were equalized in diffuse field.
- 5 different head sizes were simulated, by applying modifications $\{-50\%; -25\%; 0\%; +25\%; +50\%\}$ to the standard ITDs of each HRTF (similarly to both ears), as shown in Fig. 3.

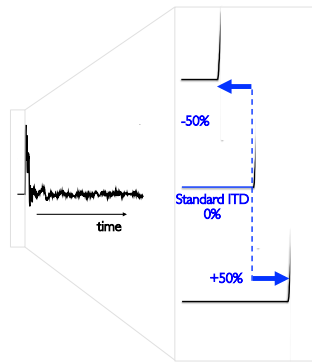


Figure 3 - The ITD values were varied by increasing or decreasing the number of null samples before the impulse response, in a similar fashion for both ears.

2.3. Description of the binauralization processing chain

The 5.1 input signals were converted into 2-channel binaural signals using standard binauralization procedures. The binaural signal in the right ear b_{right} was calculated as follows:

$$b_{right}(t) = \sum_{i=1}^5 Hr_i(t) * s_i(t) \quad (1.)$$

where s_i is the signal in the i^{th} channel, convolved with the corresponding right-ear filter Hr_i having the same azimuth as the i^{th} channel. This convolution was made for the 5 surround channels (the LFE-channel was removed), which were then summed together. A similar processing led to the signal in the left ear. The standard values of the ITDs associated to each HRTF were modified by increasing or decreasing (at a given percentage, from -50% to $+50\%$) the number of zero samples before the impulse response, as illustrated in Fig. 3. Finally, the resulting binaural signals were high-shelf filtered using a byquad filter with a 440-Hz cut-off frequency and a 2-dB gain.

2.4. Loudness standards for 5.1 and 2-channel (stereo) audio signals

To assess the loudness at the input (5.1 signal) and the output (binaural signal) of the processing chain, loudness normalization algorithms were implemented as proposed by the ITU-R BS.1770 / R128 [3, 4]. The input content was thus evaluated using the multichannel version of the algorithm, where 1.5-dB gains are applied to left and right surround channels before summation of all channels [3].

3. MAIN RESULTS AND DISCUSSION

We observed significant variations with the binaural processing parameters and input materials tested, as shown by the histograms in Figure 4 and the more detailed results plotted in Figure 5.

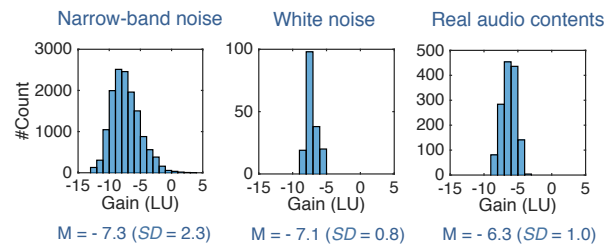


Figure 4 - Distributions of the gains (in LU) required in the processing to control input/output loudness equality as a function of the stimulus category.

On average, the required gain was found to be smaller with real broadcast stimuli ($M = -6.3$ LU) than with white noise stimuli ($M = -7.1$ LU). The difference between a musical live concert and a typical radio

drama could be as large as 2 LU (see Fig. 5, upper panel). A more detailed analysis of these results showed significant effects and interactions between every tested parameter (the HRTF, its resolution, the ITD); some of them are illustrated in Figure 5. No systematic rules or trends emerged from these patterns. Therefore, the search of a “high-level function” controlling the gain as a function of the values taken by these parameters was not successful.

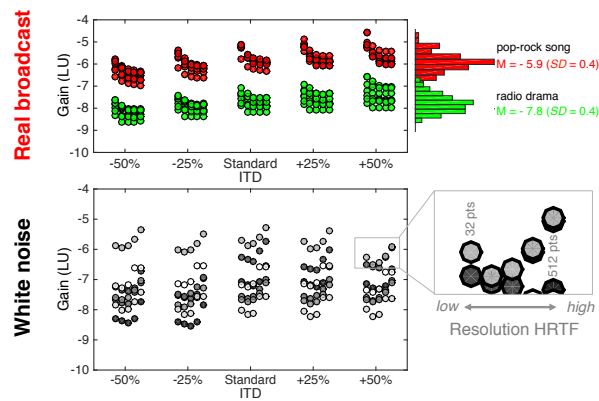


Figure 5 - Further insights into the effect of binauralization processing parameters on loudness: the ITD (horizontal axis), the HRTF (lower panel, different gray scales), the broadcast category (upper vs. lower panel) and genre (upper panel, green vs. red colors) and the HRTF-resolution (x -coordinates at a given change of ITD)

4. CONCLUSION

To conclude, our results show that a gain of about -7 dB is required when binauralizing typical 5.1 audio broadcast contents to provide binaural output signals that satisfy loudness standards for broadcasting (according to standardized evaluation methods in audio broadcasting). However, to accurately meet required standards that tolerate a ± 1 LU error only, we show that this gain must be adapted on a case-by-case basis, as revealed by the large variance obtained with both critical narrow-band noises and real audio broadcast contents. This demonstrates that a simple static gain is not sufficient. Consequently, a dynamic control (e.g. as implemented in a AGC-like circuit) appears to be the most relevant solution in this context. Such a dynamic approach is adapted to accurately target a given loudness value not only after binaural processing (as presented here) but also after stereo downmixing, which presents the similar issue in certain cases [6]. The

approach proposed in the present paper could easily be applied to object-oriented audio contents.

5. ACKNOWLEDGEMENTS

This work was done in the framework of the BiLi project. BiLi is a French FUI research project with France Televisions, Radio France, IRCAM, Orange Labs, CNSMDP, CNRS LIMSI, Trinnov audio, Arkamys and A-volute. <http://www.bili-project.org/>. We would like to thank Marc Emerit for his help and for sharing the HRTFs used in this study.

6. REFERENCES

- [1] Wolters, M., Mundt, H., & Riedmiller, J. (2010, May). Loudness Normalization in the Age of Portable Media Players. In *Audio Engineering Society Convention 128*. Audio Engineering Society.
- [2] Pike, C., & Melchior, F. (2013, May). An assessment of virtual surround sound systems for headphone listening of 5.1 multichannel audio. In *Audio Engineering Society Convention 134*. Audio Engineering Society.
- [3] ITU, “ITU-R-BS.1770: Algorithms to Measure Audio Programme Loudness and True-peak Audio Level BS Series,” tech. rep., International Telecommunications Union, 2006.
- [4] EBU, “Tech Doc 3341 Loudness Metering: EBU Mode Metering to Supplement Loudness Normalisation in Accordance with EBU R 128,” Tech. Rep. August, European Broadcast Union, Geneva, 2010.
- [5] Norcross, S. G., & Lavoie, M. C. (2011, October). The Effect of Downmixing on Measured Loudness. In *Audio Engineering Society Convention 131*. Audio Engineering Society.