



# Audio Engineering Society Convention Paper 9742

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

*This Convention paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This convention paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. 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.*

## Comparative perceptual evaluation between different methods for implementing reverberation in a binaural context

Lorenzo Picinali<sup>1</sup>, Alexander Wallin<sup>2</sup>, Yuli Levto<sup>2</sup>, and David Poirier-Quinot<sup>1</sup>

<sup>1</sup> Dyson School of Design Engineering, Imperial College London, South Kensington, SW7 1NA, London, UK

<sup>2</sup> Reactify Music, London, UK

Correspondence should be addressed to Lorenzo Picinali ([l.picinali@imperial.ac.uk](mailto:l.picinali@imperial.ac.uk))

### ABSTRACT

Reverberation has always been considered of primary importance in order to improve the realism, externalisation and immersiveness of binaurally spatialised sounds. Different techniques exist for implementing reverberation in a binaural context, each with a different level of computational complexity and spatial accuracy. A perceptual study has been performed in order to compare between the realism and localization accuracy achieved using 5 different binaural reverberation techniques. These included multichannel Ambisonic-based, stereo and mono reverberation methods. A custom web-based application has been developed implementing the testing procedures, and allowing participants to take the test remotely. Initial results with 54 participants show that no major difference in terms of perceived level of realism and spatialisation accuracy could be found between four of the five proposed reverberation methods, suggesting that a high level of complexity in the reverberation process does not always correspond to improved perceptual attributes.

### 1 Introduction

Reverberation has always been considered of primary importance in order to improve the realism, externalisation and immersiveness of binaurally spatialised sounds. Different techniques exist for implementing reverberation in a binaural context.

Room reverberation plays a major role in auditory space reproduction, increasing distance perception [1, 2], externalisation [3, 4, 5, 6], and localisation accuracy [7] compared to a dry binaural spatialisation. Room reverb can be either designed with the aim of achieving a reverb physically and perceptually close from the targeted environment [8, 9, 10], or with the aim of having a ‘good sounding room’, i.e. purely based on subjective assessment [11].

The use of Binaural Room Impulse Responses (BRIRs) is one of the most common methods for implementing reverberation in a binaural context. It consists in measuring a Head Related Transfer Function (HRTF) in a reverberant room instead of in an anechoic chamber [12]. Spatialisation is then obtained by convolving an anechoic sound with the BRIR corresponding to the specific location in which the sound needs to be spatialised. Although the quality of the reverberation is very high, this technique can become rather heavy from a computational point of view when a large number of sources need to be spatialised simultaneously.

Other methods for producing reverberation for binaural spatialisation in a more flexible and less computationally expensive manner exist, and are the object of the current study. While the computational complexity of such methods can be easily estimated,

the differences in terms of perceptual attributes can be assessed only through a study involving comparative perceptual evaluation between the various techniques.

Assuming that the binaural spatialisation of the direct signal is performed using anechoic HRTFs, which binaural reverberation method achieves the highest level of realism, externalisation and immersiveness? Which level of complexity, in terms of spatial resolution (e.g. Ambisonic order), is needed in order to achieve a realistic, externalised and immersive binaural spatialisation?

## 2 Reverberation methods

Five different artificial reverberation methods have been implemented for the current study. Each method has been used to produce a ‘wet’ (i.e. 100% reverb, without direct path) audio signal, which has then been summed with a ‘dry’ signal (spatialised using anechoic HRTFs), both with equal weights/levels.

Each reverberation method is implemented by convolving the original non-spatialised mono signal with one or more Room Impulse Responses (RIR), synthesised from a geometrical acoustic model created using CATT Acoustic (<http://catt.se/>). The same model was used for each reverb method. The simulated space is a 10m\*10m\*4m room with an estimated Sabine T30 of approximately 3s. Walls and ceiling were slightly tilted so as to avoid any flutter effect in the simulated RIRs.

The first two methods, namely ‘**ABIR 20**’ and ‘**ABIR 6**’, are based on the virtual-Ambisonic approach [10, 13]. This approach consists of encoding sound sources in the Ambisonic domain, then decoding the Ambisonic channels in a set of speaker channels, each of which are finally converted to binaural by convolving them with the HRIRs corresponding to their position. By using BRIRs (synthesised from the CATT Acoustic model described earlier) instead of HRIRs, the acoustics of different environments can be simulated. This method allows for a certain amount of flexibility (e.g. the complexity of the rendering can be reduced/increased by modifying the Ambisonic

order and the number of speakers used for the decoding). Furthermore, thanks to Ambisonic encoding being a relatively simple operation, its computational weight is not as impacted by the number of spatialised sources as the binaural rendering method.

The ABIR 20 method was implemented by encoding the sound source in the 3<sup>rd</sup> Order Ambisonic domain, and decoding it over a spherical array of 20 virtual loudspeakers around the listener (regular dodecahedron). Each speaker channel has then been convolved with a BRIR synthesised from the CATT Acoustic virtual room model. The ABIR 6 method followed the same implementation, using only 1<sup>st</sup> Order Ambisonic and 6 virtual loudspeakers around the listener (4 in the horizontal plane, 1 above, 1 below).

A further processing optimisation was performed for the Ambisonic-based reverberation methods by calculating the direct transfer function between each Ambisonic channel and the left and right binaural channels, and creating Ambisonic-to-Binaural Impulse Responses (ABIR). This allowed to reduce the number of convolutions needed for performing the auralisation from 40 to 32 for ABIR 20, and from 12 to 8 for ABIR 6. More information about this optimisation can be found in [13].

In order to compare these two methods with simpler ones, two types of mono reverberations have been implemented. In the **MONO D** (Mono Diotic) the reverb was created by synthesizing a mono IR in the CATT Acoustic model, which was then convolved with the signal and summed with the anechoic spatialised sound without any panning (i.e. same level at both ears - diotic). The **MONO P** (Mono Panned) implementation differed simply by the fact that the ‘wet’ reverb signal was summed to the original non-spatialised mono sound, and then spatialised in one single position. Using this method, both direct sound and reverb came from one position only.

Finally, a **STEREO** reverb was implemented by using only the omni (W) and left/right (Y) channels of the 1<sup>st</sup> Order ABIR 6 reverb implementation.

The following list summarises the different reverberation methods considered in this study:

- **ABIR 20.** 3<sup>rd</sup> Order virtual-Ambisonic decoded on 20 virtual loudspeakers.
- **ABIR 6.** 1<sup>st</sup> Order virtual-Ambisonic decoded on 6 virtual loudspeakers.
- **MONO D.** Mono IR presented diotically (same signal at both ears)
- **MONO P.** Mono IR panned together with the anechoic source.
- **STEREO.** Ambisonic-based stereo IR.

Considering that the test platform had to be accessible on the web (see Section 3), all processing was done offline, and only binaural (stereo) audio files were used during the test. The signal processing for creating the files (i.e. convolution with IR and HRIRs, binaural spatialisation, etc.) was performed using Matlab and the IRCAM SPAT [14]. The binaural spatialisation (both HRIR and BRIR) was done using the IRC\_1008 HRTF set from the IRCAM Listen database [15].

### 3 Procedure and test platform

The procedure used for the test had to be particularly simple, in order to allow for a web-based, unsupervised experiment, targeting a large number of subjects.

As exposed in [16] the sound material used for evaluations on spatial hearing perception impacts on subjects preference for ‘multi-dimensional spatial audio reproduction’. Our material has been carefully (yet arbitrarily) chosen to avoid favouring any of the studied reverb methods. A sound scene, composed of a single human speaker slowly moving around the listener, was created using each of the reverberation methods described in the previous section. A female voice was used as the sound source, together with step noise (one each 50cm, approx. every second) positioned 160cm below the voice source.

At the beginning of the test, participants were instructed about the task, and were given a visual reconstruction of the virtual room (see Figure 1), together with written information about the position

and movements of the sound source. After an initial level calibration, participants were presented with the sound scene, and were allowed to change in real-time between two of the five reverb methods (blind AB comparison). The selection of a different reverb method did not cause the audio to stop and restart, therefore participants could perform a very rapid and effective comparison.

Within the same page, participants were then asked the following two questions:

- Which example do you find more convincing?
- Which example gives a better impression of the direction and distance?

For each of the question, they could answer with one of the following statements:

- Definitely A
- Slightly A
- Both equally
- Slightly B
- Definitely B

In order to be able to test each possible pair between the five different reverb methods, each participant had to repeat the task 10 times. AB pairs were presented in a randomised order.



Figure 1. Visual rendering (Blender Cycles) of the virtual room used for creating the artificial reverberation.

A custom web-based application was developed, implementing the test procedure (<http://www.3d-tune-in.eu/reverb-survey>). The reverb survey website was implemented as a single-page application, built using JavaScript, HTML and CSS

languages (see screenshot in Figure 2). The survey would take participants through the entire process in a unidirectional navigation flow, controlled using ‘Next’ buttons (with varying labels), and – as from the concepts of a single-page application – carried out using seamless transitions between views without any page reloads. All audio samples used in the survey were WAVE PCM files to ensure lossless audio encoding. Their total file size amounted to around 125Mb, pre-loaded during the instruction phase of the test. The ten one-to-one AB pairings were randomly created every time the survey was loaded, i.e. on every page reload. Results were stored as JSON objects and submitted to the Firebase cloud database service.

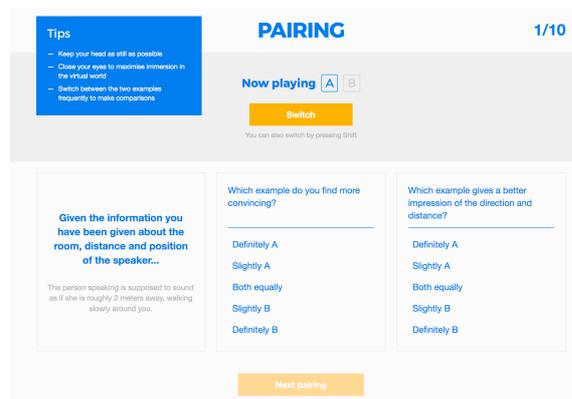


Figure 2. A screenshot of the custom web-based application used for the test (<http://www.3d-tune-in.eu/reverb-survey>).

#### 4 Early results

In this short paper, an initial analysis of the results after the first 54 subjects took the test is presented.

Scores have been assigned to every reverb method following this scheme:

- Definitely A: +2 points for method A and -2 for method B
- Slightly A: +1 point for method A and -1 for method B
- Both equally: 0 points to both methods

Considering the first question (‘Which example do you find more convincing?’), the mean scores for

each reverb method are reported in Table 1, and the boxplot displaying the data distribution, median and means is reported in Figure 3.

These results show that the MONO P reverb method achieved sensibly lower mean score (-1.04, StDev=0.89) than the other methods. After verifying the normal distribution of the data, a One-Way ANOVA was performed, confirming that the mean score for the MONO P reverb is significantly different from the mean scores achieved by the other reverb methods ( $F[4, 275]=33.58$ ;  $p<0.000$ ). No significant difference could be found between each of the other reverb methods.

Reverb method	Mean	StDev
ABIR 20	0.10	0.74
ABIR 6	0.21	0.81
MONO D	0.40	0.63
MONO P	-1.04	0.89
STEREO	0.33	0.73

Table 1. Mean scores and standard deviations for each reverb method related with the question ‘Which example do you find more convincing?’

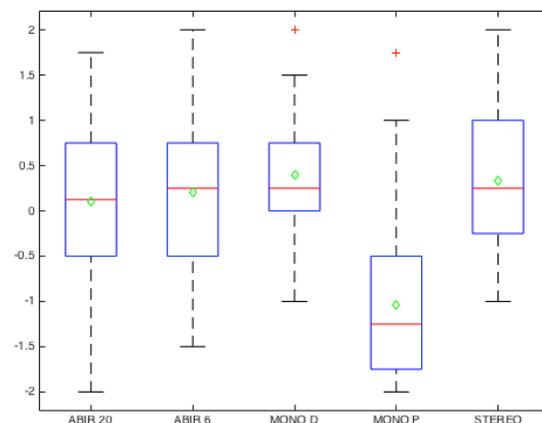


Figure 3. Scores (means reported in the green diamonds) for each reverb method related with the question ‘Which example do you find more convincing?’

Considering the second question ('Which example gives a better impression of the direction and distance?'), the mean scores for each reverb method are reported in Table 2, and the boxplot displaying the data distribution, median and means is reported in Figure 4.

Once again the MONO P reverb method achieved the lowest mean score (-0.39, StDev=1.04) when compared with the other methods, although this time the difference is not as large as for the first question.

After verifying the normal distribution of the data also for this second dataset, the One-Way ANOVA inferential statistics reported that the MONO P reverb method achieved a significantly lower mean score when compared to the other reverb methods ( $F[4, 275]=4.84$ ;  $p=0.009$ ), except for ABIR 20, whose mean does not differ significantly from the means achieved by any of the other reverb methods.

Reverb method	Mean	StDev
<b>ABIR 20</b>	0	0.68
<b>ABIR 6</b>	0.06	0.68
<b>MONO D</b>	0.12	0.75
<b>MONO P</b>	-0.39	1.04
<b>STEREO</b>	0.21	0.78

Table 2. Mean scores and standard deviations for each reverb method related with the question 'Which example gives a better impression of the direction and distance?'

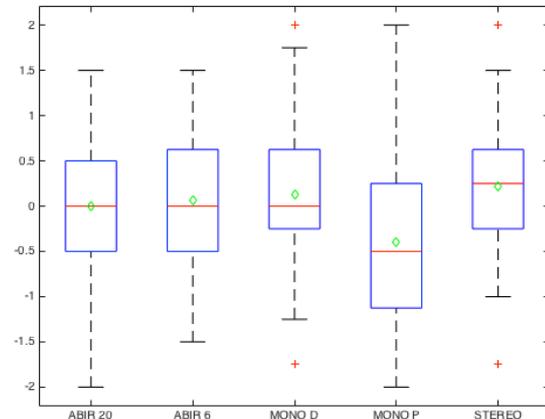


Figure 4. Scores (means reported in the green diamonds) for each reverb method related with the question 'Which example gives a better impression of the direction and distance?'

These initial results seem to suggest that, while the MONO P reverb is clearly being perceived as the worst both in terms of realism and direction/distance impression, the MONO D and STEREO reverb methods are perceived as very similar to the much more complex Ambisonic-based renderings (ABIR 20 and ABIR 6). Furthermore, no significant difference could be found between the scores achieved by the two Ambisonic-based methods, despite these having a rather different complexity (e.g. 3<sup>rd</sup> Order and 20 virtual speakers for one, 1<sup>st</sup> Order and 6 virtual speakers for the other).

In terms of computational complexity, the two mono reverbs required a single convolution to be performed for the auralisation. The STEREO reverb required 4 convolutions (2 Ambisonic channels for each ear), the ABIR 6 a total of 8 convolutions (1<sup>st</sup> order ABIR: 4 per ear), and 32 convolutions (3<sup>rd</sup> order ABIR: 16 per ear) for the ABIR 20 method. Even if we consider that for all methods the direct path signal was spatialised using the same procedure (i.e. convolution with HRIR), it is still surprising that, from a perceptual point of view, no significant differences seem to be present between 4 of the 5 proposed methods.

It is, however, important to consider that the test was carried out on a web-based platform, unsupervised,

and with participants using non-controlled hardware (headphones, DAC, etc.). Factors such as quality (and familiarity) with the playback system, use of different audio content, and good understanding of the questions asked, could have an effect on the result of similar evaluations.

## 5 Conclusions and future works

In this short paper, an initial analysis of the results after the first 54 subjects took the test is presented. The test is still open, and a publication presenting the final test results, with a more in-depth analysis will be soon published. This will include:

- Analysis of additional factors regarding the participants (e.g. type of headphones used, familiarity with spatial audio tests, etc.)
- More in-depth analysis of the pairwise comparisons.
- Larger population (the target is >100 participants)

The impact of such research could be extremely important for the 3D audio community, both in the research and industry domains. It is in fact true that if these initial findings are ultimately confirmed, the need for extremely complex binaural reverberation algorithms could become unjustifiable from a perceptual point of view in certain conditions, e.g. in unsupervised participants with non-controlled hardware.

## 6 Acknowledgments

This research has received funding from the European Community's Horizon2020 Programme (GA No.644051). For more information about the 3D Tune-In project, please visit the project's website - <http://www.3d-tune-in.eu/> - and/or refer to the References section [17, 18]

## References

- [1] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, "Auditory distance perception in humans: A summary of past and present research" *Acta Acustica united with Acustica*, vol. 91(3), pp. 409-420 (2005)
- [2] D. H. Mershon and J. N. Bowers, "Absolute and relative cues for the auditory perception of egocentric distance" *Perception*, vol. 8(3), pp. 311-322 (1979)
- [3] W. M. Hartmann and A. Wittenberg, "On the externalization of sound images" *The Journal of the Acoustical Society of America*, vol. 99(6), pp. 3678-3688 (1996)
- [4] D. R. Begault, "Perceptual effects of synthetic reverberation on three-dimensional audio systems" *Journal of the Audio Engineering Society*, vol. 40(11), pp. 895-904 (1992)
- [5] N. I. Durlach, A. Rigopulos, X. Pang, W. Woods, A. Kulkarni, H. Colburn, and E. Wenzel, "On the externalization of auditory images" *Presence: Teleoperators & Virtual Environments*, vol. 1(2), pp. 251-257 (1992)
- [6] F. Toole, "In-head localization of acoustic images" *The Journal of the Acoustical Society of America*, vol. 48(4B), pp. 943-949 (1970)
- [7] 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" *Journal of the Audio Engineering Society*, vol. 49(10), pp. 904-916 (2001)
- [8] J. H. Rindel and C. L. Christensen, "Room acoustic simulation and auralization-how close can we get to the real room" *Proc. 8<sup>th</sup> Western Pacific Acoustics Conference*, Melbourne (2003)
- [9] B. N. Postma and B. F. Katz, "Creation and calibration method of acoustical models for historic virtual reality auralizations" *Virtual Reality*, vol. 19(3-4), pp. 161-180 (2015)
- [10] L. Picinali, A. Afonso, M. Denis and B. FG Katz, "Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge", *International Journal of Human-Computer Studies*, 72(4), 393-407 (2014)

- [11] W. G. Gardner, "Reverberation algorithms" *Applications of digital signal processing to audio and acoustics*, pp. 85-131, Springer (2002)
- [12] B. G. Shinn-Cunningham, N. Kopco and T. J. Martin "Localizing nearby sound sources in a classroom: Binaural room impulse responses" *The Journal of the Acoustical Society of America* vol. 117(5) pp. 3100-3115 (2005)
- [13] A. McKeag, D. S. McGrath "Sound Field Format to Binaural Decoder with Head Tracking" *Proc. of the 101st AES Convention*, Los Angeles, CA (1996)
- [14] T. Carpentier, M. Noisternig and O. Warusfel "Twenty years of Ircam Spat: looking back, looking forward" *Proc. of the 41st International Computer Music Conference (ICMC)*, pp. 270-277, (2015)
- [15] O. Warusfel "Listen HRTF database" <http://recherche.ircam.fr/equipes/salles/listen/index.html>.
- [16] C. Guastavino and B. F. Katz, "Perceptual evaluation of multi-dimensional spatial audio reproduction" *The Journal of the Acoustical Society of America*, vol. 116(2) pp. 1105-1115 (2004)
- [17] L. Yuli, L. Picinali, L. Simeone and M. D'Cruz, "3D Tune-In: The Use of 3D Sound and Gamification to Aid Better Adoption of Hearing Aid Technologies." *Proc. 140<sup>th</sup> Audio Engineering Society Convention*. Paris (2016)
- [18] R. Eastgate, L. Picinali, H. Patel and M. D'Cruz, "3D Games for Tuning and Learning About Hearing Aids" *The Hearing Journal*, 69(4), 30-32 (2016)