PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

Effects of the order of Ambisonics on localization for different reverberant conditions in a novel 3D acoustic virtual reality system

Hermes SAMPEDRO LLOPIS[(1)*], Finnur PIND[(2)], Cheol-Ho JEONG[(1)]

[(1)]Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark, Denmark
[(2)]Henning Larsen Architects, Denmark

Abstract

This study presents an acoustic virtual reality technique based on geometrical acoustics simulations and high-order Ambisonics. The technique relies on pre-calculated B-format impulse responses in a grid of receiver positions, that are decoded into a virtual array of loudspeakers which follows the listener during runtime. The virtual loudspeaker signals are synthesized for a binaural representation via head-related transfer functions. This tool is assessed in terms of sound localization. The 1st and 2nd order Ambisonics are compared while visual and head movement conditions are varied, to see if they affect the localization performance. Moreover, this study tests the localization in full 3D challenging reverberant conditions (0.6 to 2.0 s). Results show that significant outperformance of the 2nd order technique over the 1st order is found in all head movement and visual conditions. When using 2nd order Ambisonics, allowing head movement and visual are provided, sound localization error is found to be less than the just-noticeable difference.

Keywords: sound localization, order of Ambisonics, acoustic VR, geometrical acoustics simulation, B-format impulse responses

1 INTRODUCTION

For several years, the ability to listen to simulated acoustics of spaces has been possible through auralizations (1), albeit usually only for a non-dynamic, non-interactive, fixed source-receiver set up. Recently, acoustic VR systems have become popular due to their obvious advantages of interactivity and multi-sensory evaluation (2–4). Combining various factors, such as aesthetics, sound, daylight, artificial lighting, space planning, and geometry, into an integrated and immersive VR experience is advantageous in a holistic evaluation of indoor climate. By means of VR, the communication between architects/stakeholders and acousticians becomes more effective, so that it is possible to evaluate various acoustic solutions in the early design phase and also makes it possible to detect acoustic defects, such as echoes. The VR system should work accurately in terms of localization in actual reverberant conditions and should be immersive and realistic during the evaluation.

In previous research (2–4) different solutions are given for fixed positions or established paths where the user is not allowed to move freely. In this study, an acoustic virtual reality (AVR) system is developed where the user can walk freely in the virtual scene and experience VR visuals and auralizations based on accurate acoustic simulations. Moreover, it allows the user to change immediately between different acoustic and visual conditions. This study focuses on the localization performance under different Ambisonics orders, and under different head-movement and visual stimuli conditions. Morever, the localization performance is investigated in virtual rooms with realistic reverberation times, ranging from 0.6 s to 2.0 s. Note that previous localization studies mainly dealt with near-anechoic conditions (5). The novelty of this study is twofold: a) a novel acoustic virtual reality technique based on pre-calculated, accurate B-format room impulse responses is presented and b) localization test results in 3D reverberant rooms using an integrated VR system combining accurate vision and sound stimuli.

*hsllo@elektro.dtu.dk

2 THEORY AND METHODS

2.1 AVR system design

A virtual reality system should be ideally real-time for flexibility and interactivity (6, 7). However, a real-time AVR system requires a lot of assumptions and simplifications during acoustic simulations due to the real-time constraint, so the accuracy will be limited.

For this study, a more accurate acoustic representation is sought. Room impulse responses are pre-calculated on a grid of points across the domain. The audio playback consists of the interpolated nearest room impulse responses, convolved with source signals. This approach comes with the benefit of allowing for the use of highly accurate room acoustic simulations (e.g. wave-based methods). Furthermore, this approach enables for instantaneous switching between different room acoustic configurations during run-time, which can be challenging to realize in real-time implementations due to the time needed to compute the new conditions. However, different scenarios must be pre-computed in advance. Furthermore, handling moving sources is difficult due to the large memory footprint required.

This approach allows the user to walk freely around the domain, integrating dynamic auralizations into VR such that acoustics and visuals can be experienced together. A head-position-tracked auralization system was created using virtual game engine (Oculus Rift (8) and Unity (9)) along with geometrical acoustics simulation methods (provided by Odeon (10)) and real-time audio signal processing (Pure Data (11)). This process allows for the recreation of a virtual, dynamic 3D sound field by using pre-calculated B-Format impulse responses, and decoding them to a specific virtual loudspeaker array that, ultimately, is virtually synthesized for a binaural listening experience through the use of HRTFs. The whole process is summarized in Fig. 1, where a flow chart presents how the different systems are connected.

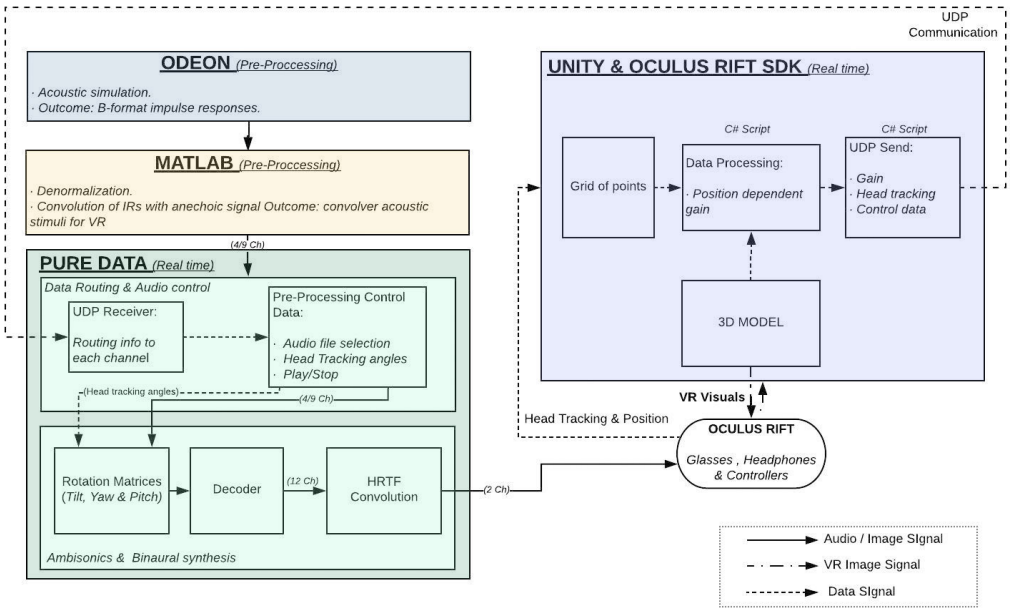


Figure 1. Flow chart illustrating the AVR system.

2.2 Odeon: Acoustic simulation and pre-processing

Odeon is a commercial room acoustic simulation software, which is used in this study for pre-computing the room impulse responses. Odeon relies on a hybrid geometrical acoustics algorithm, which combines the image-source method with a modified ray tracing algorithm. For this study, the receivers are aligned on a grid of equidistant points, see example in Fig. 2a). Odeon provides Ambisonics B-format IRs, which include the directional information of sound incidence that enables accurate acoustic auralization for for any head orientation.

2.3 Unity/Oculus Rift & Data routing

The game engine Unity is used for managing the visuals of the 3D model and for tracking the location and orientation of the listener. The same grid of receiver points used in the acoustic simulation must also be set up in Unity. As the listener walks through the virtual environment, the system will reproduce the audio file (the pre-processed convolved file), which corresponds to the point where the listener is placed. When the listener is located between points, a linear interpolation is used.

A C# script implemented in Unity allows tracking the listener's position and orientation in real-time. This information is passed to the audio processor (Pure Data), which is responsible for playing back the relevant impulse responses. The communication between Unity and Pure Data is done via UDP (User Datagram Protocol).

A Pure Data patch is built, which contains two modules: Data Routing & Audio control and Ambisonics & Binaural Synthesis as shown in Fig. 1. The Data Routing & Audio control part includes a UDP receiver. The information is routed to the subpatches (as many as the number of points) that manage the playback for each receiver position of the grid. The outputs of the subpatch are the Ambisonics signals that Pure Data reads from the stored files. All the Ambisonics channels of all the point subpatches are summed per channel and routed to the Ambisonics & Binaural Synthesis module.

2.4 Ambisonics & binaural synthesis

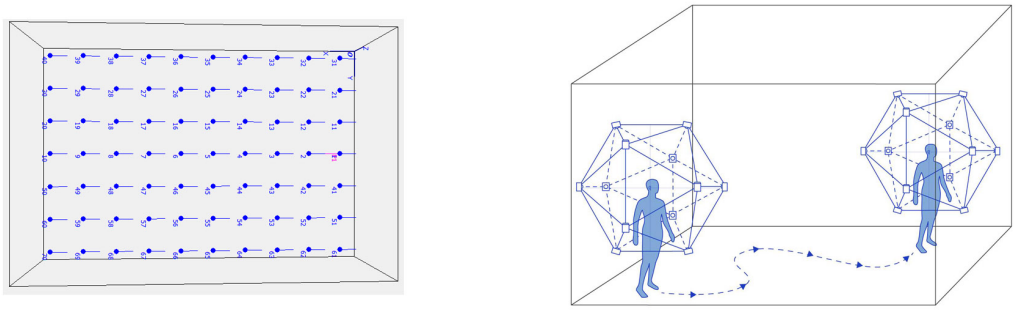
The Ambisonics & Binaural Synthesis module (see Fig. 1) is responsible for taking the B-format channels and processing them such that the sound field can be synthesized in binaural form while taking into account real-time head tracking. The audio files, which contain all the Ambisonics channels, are the inputs of this module. The Ambisonics channels are decoded for playback through a virtual loudspeaker array, which follows the listener. Finally, each loudspeaker signal is convolved with the HRTF that corresponds to the loudspeaker position, to provide the binaural synthesized signal as L/R channels (SADIE Project library is used, in particular, the Subject 001 KEMAR HRTF's (12) which provides a total of 1550 points in space with 5° resolution in azimuth and 10° resolution in elevation). Rotational matrices are applied before the decoding process to rotate the sound field according to the listener's head orientation in real-time.

The chosen geometry for the loudspeaker array was an icosahedron, which has a total number of 12 vertices containing 12 loudspeakers to play back the decoded Ambisonics sound field. Choosing a regular geometry allows one to simplify the problem. The virtual loudspeaker array is always fixed and centered at the listener's head and follows the listener's position and head orientation as illustrated in Fig. 2b).

A *Dual-band* decoder is chosen for this study as it satisfies both, the low and high-frequency localization mechanisms (13). Both criteria can be applied by using two different decoders, basic and max-rE decoders on a low and high frequency band split signal respectively (14, 15).

2.5 Subjective localization test

A subjective test was carried out to assess sound source localization using the AVR system while varying different parameters: the first and second order of Ambisonics, with and without the visuals, with and without of allowance of head movement, in three full 3D challenging reverberation time conditions going from 0.6 s up to 2.0 s.



a) b)

Figure 2. a) An example of a receiver grid in a simple room in Odeon. b) Illustration of the icosahedron loudspeaker array used in the AVR implementation and how it follows the listener in the virtual space.

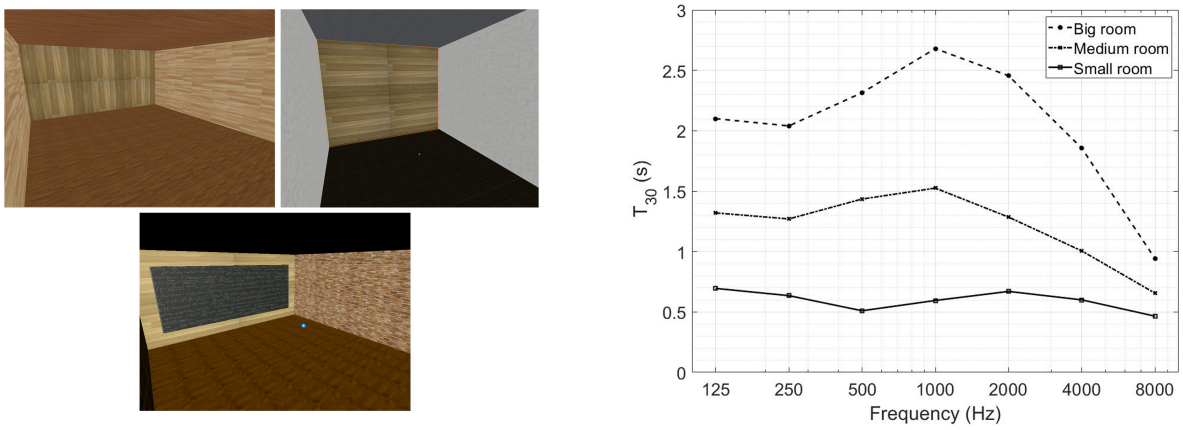
The subjective testing was conducted in an isolated listening booth. The listener sat in a specific position in the booth and was fitted with an Oculus Rift headset (which also contains headphones). The sound source signal used for all subjective experiment is an anechoic speech signal of a female voice. The volume of the headphone reproduction was calibrated such that the sound pressure level at the ears of the test subjects was 60 dB(A). A total of seventeen normal hearing test subjects from both genders, aged 25-32 participated in the experiment. All experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391).

2.5.1 Method

Three rectangular rooms were designed in order to achieve three different acoustic conditions. A big room with $V = 6000$ m^3 and $T_{30} = 2.0$ s (averaged across the 125-4000 Hz octave bands), a medium room with $V = 1376$ m^3 and $T_{30} = 1.2$ s and a small room with $V = 144$ m^3 and $T_{30} = 0.6$ s. Fig. 3a) shows the 3D models made in Unity. The blue point represents the listening position. The simulated reverberation time is presented in Fig. 3b). The primary purpose is to compare the localization performance when using the first order Ambisonics versus the second order Ambisonics in the AVR system. The initial hypothesis regarding this experiment is that the second order Ambisonics would significantly improve the localization performance. In the experiment, the participants were placed in a fixed position in a virtual room. The participants were presented with an auditory and visual stimulus and were asked to point where the sound was coming from, taking advantage of the Oculus Rift headset and controller. The source was randomly positioned inside a range of -120 to 120 degrees for azimuth (with 5 degrees of resolution) and 0 to 45 degrees for elevation (with 15 degrees of resolution). The subjects were presented with the following stimuli conditions: ***a)*** Only audio (A): **A**, ***b)*** Audio and head movement (HM) allowed: **AHM**, ***c)*** Audio and visuals (V) allowed: **AV**, ***d)*** Audio, visuals and head movement: **AVHM**. The stimuli order was randomized during the localization test.

2.5.2 Data analysis

The absolute difference in the angle between the original sound source position and the answer of the test subject was used in order to quantify the localization error. Responses were grouped according to the four scenarios *a), b), c)* and *d)* for the first and the second order Ambisonics. The results were further subjected to a multivariate analysis of variance (ANOVA), using a linear mixed model. The purpose of this statistical analysis is to conclude whether there is a statistically significant difference between different scenarios. The subject was regarded as a random effect on the mean value. The JMP software (16) was used to perform the analysis. The description of each factor and the effect type of the mixed model is summarized in Tab. 1.



a) b)

Figure 3. a) A screen capture from Unity showing the big room (top left), medium room (top right) and small room (bottom). The listening position is shown as a blue sphere. b) Simulated reverberation times for the three room types used in the localization experiment.

Table 1. Factors and effect type of the mixed model for multivariate analysis of variance of the localization test results.

Factor	Levels	Levels names	Effect type
Subject	17	1,2...17	Random effect
Order	2	First, Second	Fixed effect
HM	2	Yes, No	Fixed effect
Visual	2	Yes, No	Fixed effect
T_{30}	3	0.6 s, 1.2 s, 2 s	Fixed effect

3 RESULTS AND DISCUSSION

Fig. 4a) shows the azimuth localization errors, averaged across all test subjects and across the three rooms, together with the 95% confidence interval (CI), for the four different stimuli conditions (a,b,c,d), for both the first order and the second order Ambisonics. In addition, Tab. 2 presents the parameters with statistically significant differences. The largest mean error is around 30 degrees, for the scenario when first order Ambisonics are used and no visuals are provided and head movement is not allowed. On the other hand, 0.78 degree is the lowest mean error for the second order Ambisonics with visuals and head movement are allowed. The just-noticeable difference (JND) in sound localization is 1 degree (17). These results are similar to what other studies have found for the horizontal plane (18).

For all four conditions, the azimuth error is found to be statistically significantly lower for the second order Ambisonics compared to the first order Ambisonics. This result is consistent with other results found in the literature where higher orders provide an improvement in localization accuracy (19).

When visuals are added, or head movement is allowed, the error decreases. Again, these results coincide with what other similar studies have found (20). The visual cues help to resolve localization ambiguities and improve compactness perception (21), which means that the visuals have a direct impact on the sound localization and can improve it as well. Head movement, in particular, seems to be an essential factor for improving the localization accuracy. Notice that in condition *b)* and *c)* the CI is also reduced, which also indicates that the second order works better than the first order.

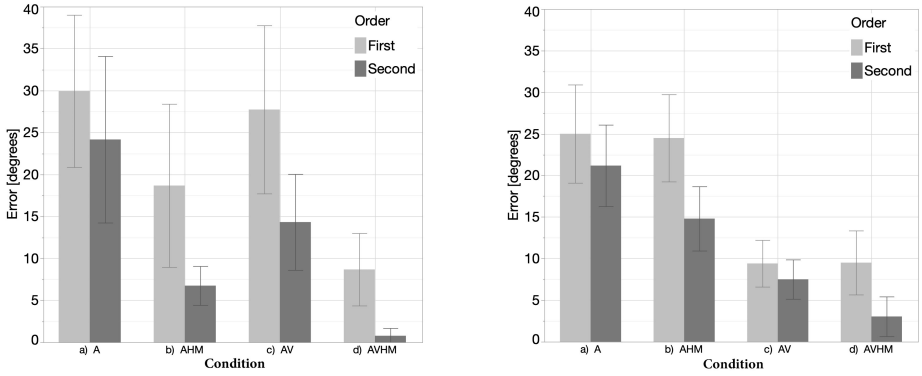


Figure 4. Results of the subjective localization test for the four different stimuli conditions tested. The results are averaged across all test subjects and across the three different acoustic conditions tested. a) Azimuth. b) Elevation.

Table 2. Multivariate ANOVA analysis results using a mixed model with REML method for elevation error data of the localization test. **Left:** Azimuth. **Right:** Elevation.

Source	DF	F Ratio	Prob>F
Order	1	14.2992	0.0002
HM	1	35.4005	< .0001
Visual	1	7.4820	0.0065
Visual×HM×T_{30}	1	5.2179	0.0229

Source	DF	F Ratio	Prob>F
Order	1	14.8452	0.0001
Visual	1	96.3213	< .0001
T_{30}×Order	1	5.4065	0.0206

The same analysis for the elevation is presented in Fig. 4b). Tab. 2 presents the parameters with statistically significant differences. As in the case of the azimuth angle, for all the conditions, the usage of second order Ambisonics results in statistically significantly lower error compared to using the first order Ambisonics.

However, adding head movement does not result in significantly reduce errors for the elevation angle. In this case, we rely mostly on monoaural spectral cues that vary with the elevation angle in a highly individual manner, so the head movement does not help a lot with a non-personalized HRTF (22). The highest error value is 25 degrees, given for condition *a)* (first order). On the other hand, the lowest value is 3 degrees, given for condition *d)* (second order). Again, adding visuals helps to localize the source position.

A one-way ANOVA concluded that there is no statistical difference between the azimuth and elevation results $(F(2,814) = 1.6005, p = 0.2062)$. However, as it is known that the localizing sources in the median plane are much more difficult for the subjects than in the horizontal plane, the resolution for the elevation was chosen coarser than for the azimuth. The differences in the resolution could explain why there is not a statistically significant difference between the two.

The performance between different room types is compared and presented in Fig. 5, where the mean error for different T_{30} values together with the 95% confidence interval is shown. A one-way ANOVA concluded that there is no statistical difference between the different T_{30} values for azimuth $(F(2,930) = 1.2095, p = 0.2994)$. However, the analysis shows a statistical difference for the elevation $(F(2,837) = 3.1216, p = 0.0451)$. This can be due to the difference in the degrees resolution between azimuth and elevation used during the experiment. However, for both azimuth and elevation, the second order Ambisonics presents less mean error compared with the first order.

The averaged across all stimuli conditions (a,b,c,d), all room types and elevation and azimuth angles show a mean error of 19 degrees for the first order and 11 degrees for the second order, which reveals that second order

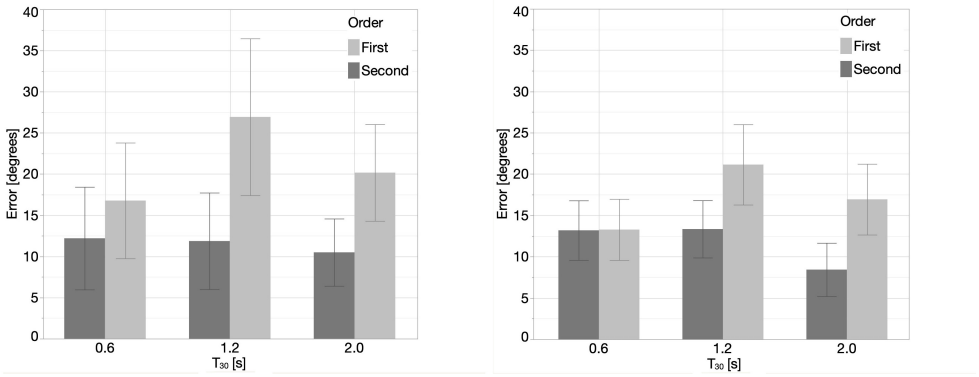


Figure 5. Results of each room type of the subjective localization test, averaged across all stimuli conditions (a,b,c,d). a) Azimuth; b) Elevation.

results in lower error values. This fact is in agreement with the previous analysis. With this in consideration, it is possible to conclude that the second order improves the localization performance significantly in the integrated VR system.

4 CONCLUSIONS

In this study, the localization performance using the first and second order Ambisonics in different visuals and head movement conditions for three different acoustic conditions was tested. The AVR system used is based on pre-calculated B-Format impulse responses from a hybrid geometrical acoustics software, and it allows for the free movement and orientation of the listener in the modeled space.

Results revealed that using the second order Ambisonics results consistently in better localization performances compared to when using the first order Ambisonics. When the second order Ambisonics decoder is combined with high quality VR visuals, and head movement is allowed, the localization error is found to be the lowest, being smaller than the JND threshold. A statistical linear mixed model ANOVA analysis on the localization test results confirms the hypothesis that the second order would significantly improve the localization performance than the first order.

REFERENCES

1. Krokstad A, Strom S, Sørsdal S. Calculating the acoustical room response by the use of a ray tracing technique. Journal of Sound and Vibration, Vol 8 (1), pp 118-125, 1968.

2. Ballestero E, Robinson P, Dance S. Head-tracked auralizations for a dynamic audio expercience in Virtual Reality sceneries. Proceedings of International Congress on Sound and Vibration, London, UK, July 23-27, 2017.

3. Postma B N, Poirier-Quinot D, Meyer J. Virtual Reality performance auralization in a calibrated model of Notre-Dame cathedral. Proceedings of EuroRegio, Porto, Portugal, June 13-15, 2016.

4. Sanchez M, Van Renterghem T, Sun K. Using Virtual Reality for assessing the role of noise in the audio-visual design of an urban public space. Landscape and Urban Planning, Vol 167, pp 98-107, 2017.

5. Bertet S, Daniel J, Gros L, Parizet E, Warusfe O. Investigation of the perceived spatial resolution of higher order Ambisonics sound fields: A subjective evaluation involving virtual and real 3D microphones. Proceed-

ings of Audio Engineering Society Conference: 30th International Conference: Intelligent Audio Environments, Saariselka, Finland, March 15-17, 2007.

6. Pind F, Jeong C-H, Sampedro Llopis H. Acoustic Virtual Reality – Methods and challenges. Proceedings of Baltic-Nordic Acoustic Meeting (BNAM), Reykjavík, Iceland, April 15-18, 2018.

7. Pelzer P, Aspöck L, Schröder D, Vorlaender M. Interactive Real-Time Simulation and Auralization for Modifiable Rooms. Building Acoustics, Vol 21, pp 65-73, 2014.

8. Facebook Technologies LLC. Oculus Rift [Internet]. 2019 [cited May 2019]. Available from: https://www.oculus.com/

9. Technologies Unity. Unity [Internet]. 2019 [cited May 2019]. Available from: https://unity.com/

10. Odeon A/S. Odeon [Internet]. 2019 [cited May 2019]. Available from: https://odeon.dk/

11. Institut für Elektronische Musik und Akustik - IEM. Pure Data [Internet]. 2019 [cited May 2019]. Available from: https://puredata.info/

12. York-University. SADIE, Spatial Audio for Domestic Interactive Entertainment [Internet]. 2018 [cited May 2019]. Available from: https://www.york.ac.uk/sadie-project/

13. Gerzon M A. Surround-sound psychoacoustics. Wireless World, Vol 80, pp 483-486, 1974.

14. Heller A J, Lee R, Benjamin E M. Is My Decoder Ambisonic?. Proceedings of Audio Engineering Society 125th Convention, Moscone center, San Francisco, CA, October 3-5, 2008.

15. Heller A J, Lee L, Benjamin E B. A Toolkit for the Design of Ambisonic Decoders. Proceedings of Linux Audio Conference, Stanford University, California, April 12-15, 2012.

16. SAS Institute Inc. JMP Statistical Discovery [Internet]. 2018 [cited May 2019]. Available from: https://www.jmp.com/

17. Perrott D R, Saberi K. Minimum audible angle thresholds for sources varying in both elevation and azimuth. The Journal of the Acoustical Society of America, Vol 87 (4), pp 1728-31, 1990.

18. Bertet S, Daniel J, Parizet E, Warusfel O. Investigation on Localisation Accuracy for First and Higher Order Ambisonics Reproduced Sound Sources. Acta Acustica united with Acustica, Vol 99 (4), pp 642-657, 2013.

19. Thresh L, Armstrong C, Kearney G. A Direct Comparison of Localization Performance When Using First, Third, and Fifth Ambisonics Order for Real Loudspeaker and Virtual Loudspeaker Rendering. Proceedings of Audio Engineering Society 143rd Convention, Javits convention center, NY, USA, October 18-21, 2017.

20. Wersenyi G. Effect of Emulated Head-Tracking for Reducing Localization Errors in Virtual Audio Simulation. IEEE Transactions on Audio, Speech, and Language Processing, Vol 17 (2), pp 247-252, 2009.

21. Gil Carvajal J C, Santurette S, Cubick J, Dau T. The Influence of Visual Cues on Sound Externalization. Proceedings of 39th Midwinter Meeting of Association of Research in Otolaryngology, San Diego, CA, United States, February 20-24, 2016.

22. Rajendran V G, Gamper H. Spectral manipulation improves elevation perception with non-individualized head-related transfer functions. The Journal of the Acoustical Society of America, Vol 145 (3), pp EL222-EL228, 2019.