Forum Acusticum 2011 - Aalborg - Denmark

Evaluation of a mixed-order planar and periphonic Ambisonics playback implementation

Johannes Käsbach and Sylvain Favrot
Center of Applied Hearing Research (CAHR), Technical University of Denmark (DTU), Ørsteds Plads Building 352, 2800 Kgs. Lyngby.

Jörg Buchholz
National Acoustic Laboratories, 126 Greville Street, Chatswood NSW 2067, Australia.

Summary
Planar (2D) and periphonic (3D) higher-order Ambisonics (HOA) playback systems are widely used in multi-channel audio applications. For a given Ambisonics order, 2D systems require far less loudspeakers and provide a larger spatial resolution but cannot naturally reproduce elevated sound sources. In order to combine the benefits of 2D and 3D systems, a higher order 2D playback system can be mixed with a lower order 3D system. In the present study, a mixed-order Ambisonics playback system was realised by extending the spherical harmonics decomposition of a 3D sound field with additional horizontal components. The performance of the system was analysed by considering a small and a large loudspeaker setup, allowing for different combinations of 2D and 3D Ambisonics orders. An objective evaluation showed that the systems provided a high spatial resolution for horizontal sources while producing a smooth decrease in spatial resolution with increasing source elevation until 3D performance is reached. This observation was confirmed by a listening test (simulated concert scenario), which showed that in comparison to a conventional 3D system the perceived spatial resolution for sources in the horizontal plane can be significantly increased by adding 2D components and thereby approaching 2D system's performance. Simultaneously, frequency spectrum properties of horizontal sound sources were restored and did not show a low pass filtering effect as it is present in 3D HOA systems.

PACS no. 43.38.Md, 43.60.Dh

1. Introduction

Practical periphonic Ambisonics playback setups often have a non-regular distribution of loudspeakers. A typical layout contains a regular loudspeaker ring in the horizontal plane and has additional elevated loudspeakers to represent elevated sources. Such a setup in combination with an appropriate mixed-order Ambisonics (MOA) coding strategy [1] can be used to enhance the spatial resolution in the horizontal plane. Such MOA approach is further encouraged by the fact that most sound sources that are of interest in hearing or hearing aid research are roughly located in the horizontal plane. Examples include a speaker (during a conversation), traffic noise, or music (as musician or audience). Since human auditory localization is much more accurate in the horizontal plane than in the vertical plane [2], a mixed-order approach is also highly motivated by an auditory perception perspective. Within this study, a MOA system was developed in respect to the following goals, i.e. (1) improving the localisation in the horizontal plane while maintaining periphonic properties for elevated sources, (2) providing a smooth perceptional transition between both representations, and (3) minimising coloration effects. Thereby, the portability of Ambisonic systems, which is the independence of the encoding and decoding stage, was maintained.

2. Implementation

The MOA scheme is based on the spherical harmonic decomposition of the three-dimensional sound field. A conventional 3D HOA system is built up to the periphonic order M_{3D}. The directional selectivity in the horizontal plane is then increased by adding horizontal components ($m = n$) up to the planar order M_{2D}. The spherical Fourier-Bessel series describing the pressure sound field according to [3] is truncated in respect to the defined orders and becomes

$$p(r,\theta,\delta) = \sum_{m=0}^{M_{3D}} j^m j_m(kr) \sum_{0\leqslant n \leqslant m, \sigma=\pm1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta,\delta)$$
$$+ \sum_{m=M_{3D}+1}^{M_{2D}} j^m j_m(kr) \sum_{\sigma=\pm1} B_{mm}^{\sigma} Y_{mm}^{\sigma}(\theta,\delta), \quad (1)$$

with $j_m(kr)$ being the Bessel functions in respect to the spatial frequency kr, B_{mn}^{σ} the Ambisonic channels and $Y_{mn}^{\sigma(N3D)}(\theta,\delta)$ the spherical harmonic functions[1] defined as

$$Y_{mn}^{\sigma(N3D)}(\theta,\delta) = \underbrace{\sqrt{2m+1} N_{mn} P_{mn}(\sin\delta)}_{a_{mn}(\delta)} .$$
$$\underbrace{\begin{cases} \cos n\theta \text{ if } \sigma = +1 \\ \sin n\theta \text{ if } \sigma = -1 \text{ (ignored if n=0)}, \end{cases}}_{b_n(\theta)} \quad (2)$$

where θ describes the azimuthal and δ the elevation angle. They are defined with the common Schmidt semi-normalisation

$$N_{mn} = \sqrt{(2-\delta_{0,n})\frac{(m-n)!}{(m+n)!}}, \quad (3)$$

where $\delta_{0,n}$ defines the Kronecker symbol. For the extended horizontal components (with $m = n$), $a_{mn}(\delta)$ are the weighted Legendre functions $P_{mm}(\sin\delta)$ that determine the elevation-dependent contribution of the 2D components. Two schemes are hereby suggested. The *mix* scheme includes higher order Legendre functions up to order $m = M_{2D}$, whereas the *mixT* scheme truncates the Legendre functions at order $m_T = M_{3D}$. The latter provides a smoother decay of the 2D components with increasing source elevation and improves the orthonormality properties of the spherical harmonic functions as outlined in the following.

3. Loudspeaker array and orthonormality

In this section, the maximal orders M_{2D} and M_{3D} are discussed for an exemple loudspeaker array shown in Figure 1. This setup contains 30 loudspeakers including a horizontal ring with 16 equidistantly spaced loudspeakers.
The maximum order for regular loudspeaker arrays is usually determined by the number of loudspeakers $L = (M+1)^2$ for 3D and $L = (2M+1)$ for 2D Ambisonics. Due to the non-regular design of the present array, considerations of the orthonormality properties

[1] The given definition refers to the full normalisation 3D (N3D) convention.

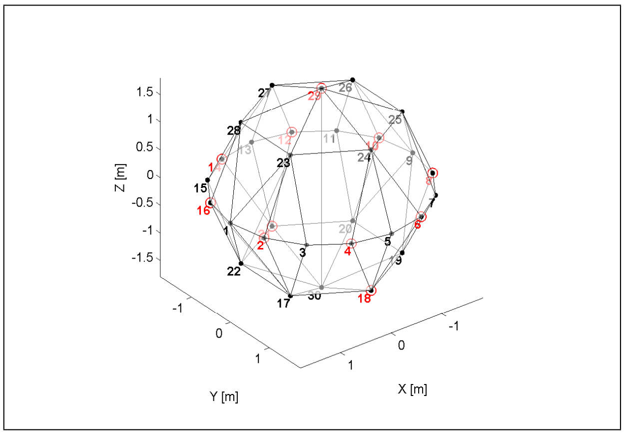

Figure 1. Illustration of the 30 loudspeaker array (idealised design of the Spacelab at DTU facilities). Elevated loudspeaker rings are placed at $\delta = \pm 45°$.

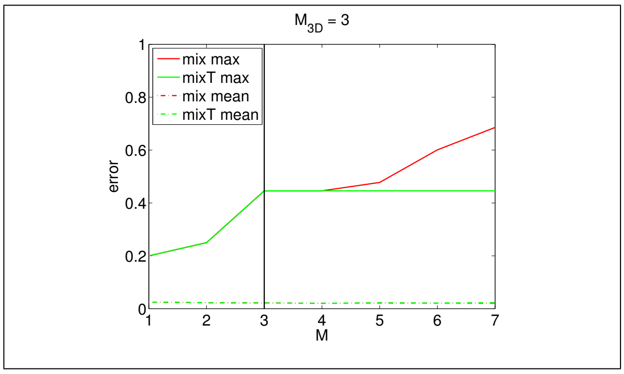

Figure 2. Orthonormality errors (maximum and mean) for the 30LS array in a *mix* MOA and *mixT* MOA application as a function of order M. An example system of $M_{3D} = 3$ is considered.

of the spherical harmonic functions are necessary in order to determine the maximum orders. Following [4] an orthonormality error is calculated by

$$U = I_k - \frac{1}{L} C C^T, \quad (4)$$

with I_k being a KxK identity matrix, K the number of spherical harmonic components and C the re-encoding matrix. The mean and maximum orthonormality error of the mix and mixT schemes are shown in Figure 2 as a function of order M for an example system of $M_{3D} = 3$.
When using the entire array in a conventional 3D HOA reproduction, results show that the maximal periphonic order is limited to $M_{3D} = 3$ (with a maximal error of 45%) since otherwise for a higher order ($M_{3D} = 4$) an orthonormality error of 100% occurs and a correct sound field reconstruction cannot be guaranteed. The horizontal loudspeaker ring itself (in a conventional 2D HOA reproduction) allows for a planar order $M_{2D} = 7$ without any occurrence of errors. Extending the present 3D system to

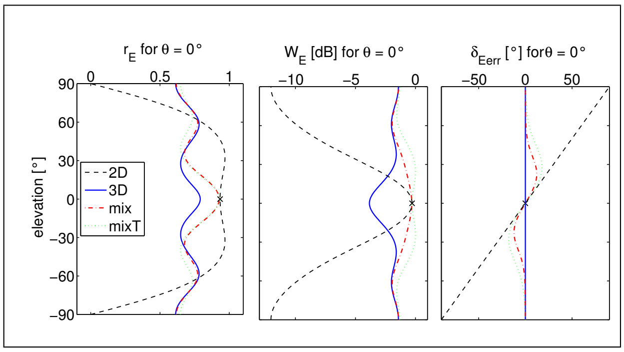

Figure 3. Investigations of Gerzon's energy vector (magnitude r_E, total energy W_E and elevation error $\delta_{E_{err}}$) for different source elevation angles reproduced in the 30LS array in case of the schemes under study for $M_{2D} = 7$ and $M_{3D} = 3$.

a MOA system by adding additional horizontal components (with $m > M_{3D}$) leads to an increased maximal orthonormality error for the *mix* scheme (69% at $M = 7$) and a maximal error for the *mixT* scheme that does not exceed 45%. In the latter case, the maximal orthonormality error is thereby limited according to the chosen periphonic order M_{3D}.

4. Objective Evaluation

The performance of the two MOA schemes are objectively analysed and compared to conventional 3D HOA and 2D HOA in the following. Plots of Gerzon's energy vector [5] are presented in Figure 3. Note that a very weak dependency on the source azimuth can be observed, which is omitted in the following. The energy vector's magnitude r_E shows the desired improvement of both MOA implementations compared to 3D HOA for sources in the horizontal plane reaching an almost identical performance to 2D HOA. 3D HOA performance is then obtained for fully elevated sources with a transition area in between. The total energy W_E for 3D HOA reveals a dip in the horizontal plane due to the non-regular array design. The two MOA schemes lead to a more balanced energy distribution with a smooth transition towards the conventional 3D HOA for elevated sources. The *mixT* implementation shows thereby a slightly smoother transition than the *mix* one. A clear drawback of MOA is an error in the reproduction of elevated sound sources $\delta_{E_{err}}$ in between $\delta = \pm 60°$ due to an asymptotic behaviour to 2D HOA in the transition area.
Investigations of the directivity pattern (Figure 4), obtained by plotting loudspeaker gains respective to their angular position, further highlight the energy vector results. A more focussed energy beam towards the source can be produced by the MOA implementations in the horizontal plane as compared to 3D HOA with a mainlobe that is identical to 2D HOA. The *mixT* implementation additionally reduces the side-

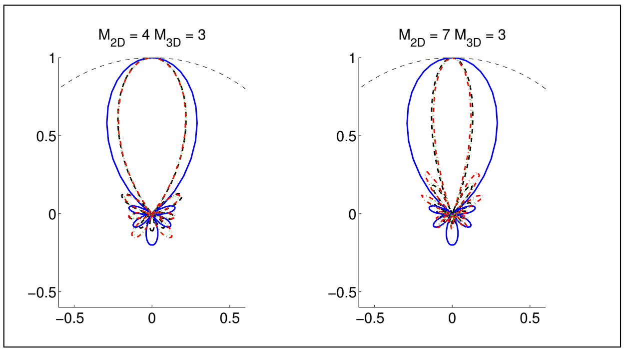

Figure 4. Directivity pattern for the coding techniques under study illustrated in the xy plane. Legend as in Figure 3.

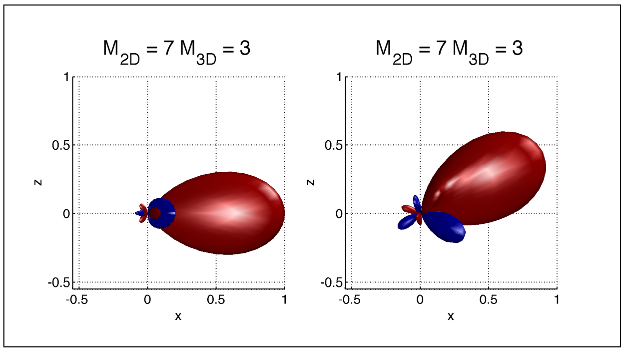

Figure 5. Vertical expansion of the directivity pattern for *mixT* MOA for a source at $\delta = 0°$ (left) and $\delta = 30°$ (right). Red marks positive and blue negative loudspeaker gains.

lobes in contrast to *mix* especially when a larger difference between orders is considered (right plot). The planar order M_{2D} thereby determines the degree of focus in the horizontal plane whereas the periphonic order M_{3D} terminates the directivity pattern's vertical expansions (Figure 5). A transformation of the pattern takes place, when elevating the source until 3D performance is reached at the zenith.
Simulated magnitude responses for a centered head and torso simulator (HATS) and a frontal sound source with two different elevation angles are shown in Figure 6. In the case of a horizontal source ($\delta = 0°$), 3D HOA has an attenuated mid- and high-frequency spectrum, whereas all other schemes exhibit a similar spectrum compared to a single loudspeaker representing a single sound source. This impairment of 3D HOA vanishes for an elevated source ($\delta = 37°$), so that a similar behaviour is obtained for all implementations except 2D HOA. Elevated sources can be reproduced with 2D HOA by projecting them into the horizontal plane (using the horizontal components of the spherical harmonic functions). Due to the decreasing contribution of the 2D components with increasing the elevation angle, 2D HOA shows an attenuated response in this case.

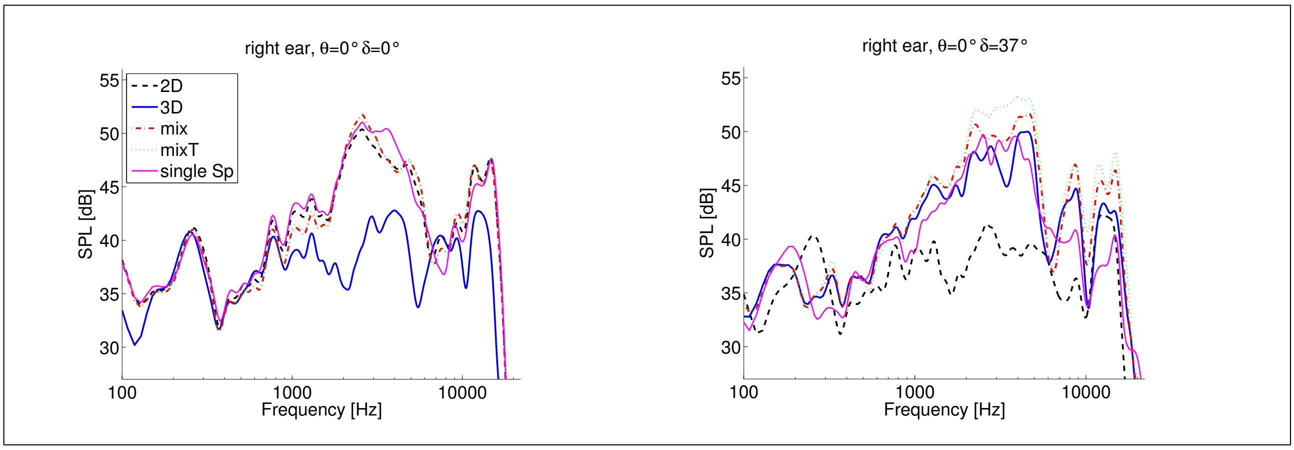

Figure 6. Simulated magnitude responses for a centered HATS and two source positions applying different Ambisonic coding strategies. Right HRTFs are shown.

5. Subjective Evaluation

5.1. Procedure and stimuli

A subjective evaluation of the proposed MOA methods was performed by simulating a concert listening situation. The listening experiment's procedure applied a MUSHRA-like test [6] but without anchor and reference. The task was to rank the 5 coding techniques 3D HOA, 2D HOA, *mix* MOA, *mixT* MOA and Nearest Loudspeaker (NLS)[2] on a given scale from 0 to 100 according to 3 attributes *spatial resolution*, *clarity* and *distance* (defined in the appendix).
To find adequate attributes, [7] and [8] were taken into account and preliminary studies were conducted in [1]. The Spacelab at the facilities of the Technical University of Denmark (DTU) was used as playback system containing 29 loudspeakers. It corresponds to the layout in Figure 1 without the bottom loudspeaker and changed placements of the two elevated loudspeaker rings to $\delta = 36.5°$ and $\delta = -34°$. This array allows for a high-order playback with $M_{2D} = 7$ and $M_{3D} = 3$. Moreover, a low-order setup ($M_{2D} = 3$, $M_{3D} = 1$ after orthonormatlity considerations) was tested, using a reduced version of the array with 11 loudspeakers (marked with red circles in the same Figure). The simulation was realised by calculating an acoustic room model of the large concert hall at the Royal Danish Academy of Music (DKDM) in ODEON by considering 7 sources on the stage (in the horizontal plane) and one listening position for a seated listener 8 m away from the stage. The model was auralised with the LoRA Toolbox [9]. The direct sound and early reflections were encoded with the five different schemes under test. The late reflections, represented by energy envelopes, were auralised with first order Ambisonics (2D for 2D HOA and 3D for the others) and multiplied with Gaussian noise uncorrelated for each loudspeaker. The resulting 7 mRIRs (multi-channel room impulse responses) were convolved with the corresponding anechoic instrument recordings (1 min snippet) of a pop-song [10] including vocals (male), piano, organ, guitar, bass, drums and sound effects (cello or bells). In the experiment, the stimuli and attributes appeared in randomised order. The two systems, high- and low-order, were tested separately from each other and an equal number of subjects started with either system. One repetition was included in the procedure.
In total 12 normal hearing listeners (8 male, 4 female) participated in the experiment. All subjects were experienced in psychoacoustic experiments and were trained for the procedure.

[2] This scheme selects the loudspeaker closest to the origin of sound and is therefore not a physical sound field reproduction approach. It is included for comparison.

5.2. Results

The average ratings across subjects are presented in Figure 7 and are obtained by performing a multiple comparison test [11]. They reflect the mean values with confident-intervals based on an overall level of significance. Significant differences between coding strategies are present for non-overlapping intervals. Considering the ranking for the attribute "spatial resolution", the performance of both MOA schemes is significantly improved compared to 3D HOA and is as good as 2D HOA. This is true for both the high- and low-order system with the exception that a weaker performance is given for the *mix* than the *mixT* scheme in the low-order system. The best performance is achieved by the NLS coding.
For the attribute clarity a similar hierarchy can be observed. The 3D HOA system is characterised by having a more muffled sound compared to all other systems. In the low-order system there is also a significant difference between 2D HOA and NLS which is not seen in the high-order system.

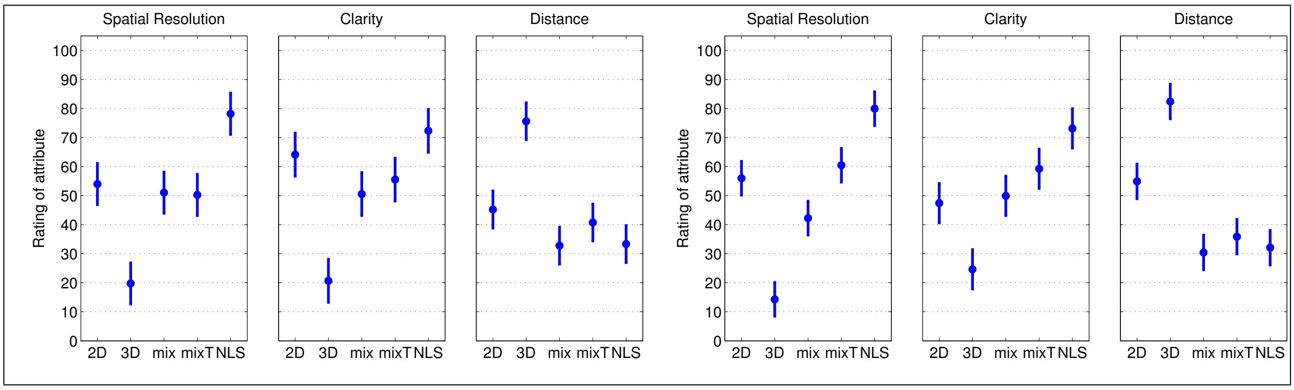

Figure 7. Averaged results from the subjective evaluation. Left: high-order ($M_{2D} = 7$, $M_{3D} = 3$), 29LS. Right: low-order ($M_{2D} = 3$, $M_{3D} = 1$), 11LS.

For the distance perception a significant difference to the 3D HOA system is shown as well. The musicians appear to be further away for this system compared to the others. Furthermore, the two MOA schemes produced a similar distance perception as the NLS system for both playback systems.
In addition, a principal component analysis (PCA) was performed. A one-dimensional representation of the tested attributes was thereby revealed, showing that there was a consensus among subjects for each attribute.

5.3. Discussion

The subjective evaluation confirms the results of the objective evaluation. Simulated results of r_E and directivity (Figure 3 and 4) are in good agreement with the apparent spatial resolution, i.e. an increase in rE value (or a narrower directivity pattern) is linked to an increase in apparent spatial resolution. It is interesting to see that even the seventh-order 2D HOA system can not achieve the spatial resolution provided by the NLS system. Hence, rather high Ambisonics orders are required for perfectly representing spatial scenes.
The perception of the attribute clarity can be directly linked to the simulated frequency spectrum in Figure 6: The more muffled sound of 3D HOA is due to its low pass characteristic.
The distance perception is mainly influenced by the Direct to Reverberant ratio (D/R). Considering again Figure 3, the dip in the horizontal plane of the total energy W_E for the 3D HOA system leads to a reduced direct sound level. The reduced D/R could have created a distance perception of the musicians where they appear further away. The similarity in the apparent distance between MOA and NLS in both systems indicates a correct D/R representation in the MOA implementations.
Although the room reverberation was presented in 3D (where applicable), the current listening test mainly evaluated the advantage of the MOA approach on horizontal sound reproduction. Hence, the periphonic qualities of the MOA systems were not really tested.
In [12] the benefit of with-height representations (first-order 3D) in contrast to horizontal-only playback (first-order 2D) and a hybrid approach (first-order 3D with omitted height information) was investigated in a subjective evaluation procedure in terms of the more general attribute "enjoyment of listening experience". However, no significant differences could be observed. Hence, further comparisons between 2D, 3D, and mixed-order systems should be performed considering different acoustic scenes that systematically address 2D and 3D qualities. Moreover, meaningful and well suited attributes need to be applied and listeners should be well trained to reliably indicate even slight differences between the different representations.

6. Conclusions

In the present study, a mixed-order Ambisonics (MOA) scheme was proposed based on the extension of spherical harmonic functions by additional horizontal components. Thereby two implementations, *mix* and *mixT*, were suggested. Both schemes were tested and compared to conventional 3D HOA and 2D HOA codings in an objective and subjective evaluation. From this case study, it can be concluded that significant improvement is achieved by the two suggested MOA schemes in respect to the originally defined aims, i.e.
(1) Objective (r_E and directivity) and subjective evaluations confirmed the improved spatial resolution of a MOA system in contrast to conventional 3D HOA.
(2) The energy distribution of MOA matches the 2D HOA performance in the horizontal plane and follows a smooth transition to 3D HOA for zenithal sources.
(3) MOA does not exhibit the low pass filter characteristic of 3D HOA.
These observations were made for high-order and low-order MOA where 2 exemplary loudspeaker setups were tested. From the 2 proposed MOA schemes, the *mixT* implementation is favourable to the *mix* one because (a) the maximum orthonormality error is lower

(according to M_{3D}), (b) a smoother energy transition between horizontal and elevated sources is provided, and (c) an increased apparent spatial resolution is obtained at least in case of the low-order system.
The present study contributes to the research of an entire MOA chain, that covers the process from recording to playback. In future studies, the suggested concepts should be applied to appropriate microphone arrays. Investigations of different microphone/loudspeaker arrays with their respective Ambisonics orders will be helpful for the optimisation of the MOA chain. Drawbacks such as a modified elevation angle reproduction should be further improved.

References

[1] Käsbach J.: Development and evaluation of a mixed-order Ambisonics playback system. Master's thesis, CAHR, Technical University of Denmark (DTU), 2010.

[2] Blauert J.: Spatial Hearing: The psychophysics of human sound localization. Revised edition, MIT Press, 1997.

[3] Daniel J., Nicol R. and Moreau S.: Further investigations of High Order Ambisonics and Wavefield Synthesis for holophonic sound imaging. Presented at the 114th convention of the Audio Engineering Society, Amsterdam, Netherlands, 2003.

[4] Moreau S., Daniel J. and Bertet S.: 3D sound field recording with Higher Order Ambisonics - Objective measurements and validation of spherical microphone. Presented at the 120th convention of the Audio Engineering Society, Paris, France, 2006.

[5] Gerzon M. A.: General metatheory of auditory localisation. Presented at the 92nd convention of the Audio Engineering Society, Vienna, Austria, March 1992.

[6] Recommendation ITU-R BS.1534-1: Method for the subjective assessment of intermediate quality level of coding systems. The ITU Radiocommunication Assembly, 2001-2003.

[7] Choisel S. and Wickelmaier F.: Extraction of auditory features and elicitation of attributes for the assessment of multichannel reproduced sound. Journal of the Audio Engineering Society, 54(9), 2006.

[8] Choisel S. and Wickelmaier F.: Evaluation of multichannel reproduced sound: scaling auditory attributes underlying listener preference. Journal of the Acoustical Society of America 121(1), 2007.

[9] Favrot, S. and Buchholz, J. M. LoRA - A loudspeaker-based room auralization system. Acta Acustica united with Acustica, 96(2), 2010.

[10] Homesick Hank, "Brother". www.myspace.com/homesickhankmusic. Last update: 28.01.2011.

[11] Johnson R.: Miller and Freund's probability and statistics for engineers. Seventh edition. Pearson Education, 2005.

[12] Baume C. and Churnside A.: Upping the auntie: A broadcaster's take on Ambisonics. Presented at the 128th convention of the Audio Engineering Society, London, United Kingdom, 2010.

Appendix

Definition of the attributes in the subjective evaluation:
Spatial resolution: A sound has a high spatial resolution when it is *distinct* (100), i.e. you can easily 'draw a map' of the musicians on the given stage, and a low spatial resolution when it is *blurry* (0).
Clarity: A sound has a high clarity when it is *brilliant* (100) and a low clarity when it is *muffled* (0) [8].
Distance: The distance to the musicians might change with the stimuli. Some sounds might appear to be *closer* (0) to you, whereas others seem more *far away* (100). Evaluate the distance of the musicians to you.