Perceptual effects of ambisonics on room auralization

Sylvain Favrot and Jörg M. Buchholz

Centre for Applied Hearing Research, Department of Electrical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark

Higher-order Ambisonics (HOA) has been widely used for loudspeaker-based reproduction of static and moving sound sources from arbitrary directions. When Ambisonics is used to reproduce anechoic sound sources, it is known that the achievable directionality is limited and coloration artifacts are introduced. However, it is unclear to what extend HOA affects auditory performance measures when it is applied to reproduce reverberant sounds. In order to investigate this issue, a loudspeaker-based room auralization (LoRA) system was utilized, which was recently developed at the Centre for Applied Hearing Research. Within the LoRA system, HOA is used in different ways to auralize the individual parts of a simulated room impulse response (i.e. direct sound, early reflections, and reverberation). In order to study the effect of HOA on the auditory processing of reverberant sounds, first, the effect of the applied Ambisonic order on speech intelligibility in noise was investigated and compared to results obtained with "single loudspeaker" auralization. Second, a distance perception experiment was performed to evaluate the entire LoRA processing in comparison to recorded binaural auralization. For these two experiments, listener performance in HOA-generated virtual environments was similar to that of the reference case which represented the "real" environment.

INTRODUCTION

Virtual auditory environments (VAEs) ideally generate realistic auditory scenes to a listener for applications such as, auditory displays, teleconferencing or hearing research. These environments typically consist of different sound sources placed at different locations in a room. Room auralization systems typically create VAEs by first using an acoustic room model to simulate the interaction of the individual sound sources with the room at the virtual listener position. The result is then conditioned according to the chosen auralization technique, and a set of impulse responses is derived. The impulse responses are finally convolved with an anechoic sound signal and presented to a listener via headphones or loudspeakers.

When applying VAE to auditory research with aided hearing-impaired listeners, a non-individualized VAE is desired that can support head-rotations and facilitates the wearing of hearing aids. For such application, loudspeaker-based auralization methods seem to be most appropriate, e.g., higher-order Ambisonics (HOA) or wave-field synthesis (e.g., Daniel, 2000). In the following, only HOA will be further considered, because it: (i) is a listener-centered method optimized for the sweet spot around the listener, (ii) allows rendering of sound sources from any direction in three-

Proceedings of ISAAR 2009: Binaural Processing and Spatial Hearing. 2nd International Symposium on Auditory and Audiological Research. August 2009, Elsinore, Denmark. Edited by J. M. Buchholz, T. Dau, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN: 87-990013-2-2. The Danavox Jubilee Foundation, 2010.

dimensional (3D) space, and (iii) can be combined with microphone array techniques to playback recorded scenes. For room auralization, HOA can be easily combined with acoustic room modeling software such as ODEON (Rindel *et al.*, 2010) or CATT (Dalenbäck, 1996). In the present study, the loudspeaker-based room auralization (LoRA) system (Favrot and Buchholz, 2010) is applied to realize an example system that combines an acoustic room model (ODEON) with HOA auralization.

When realizing anechoic sound sources with HOA, the limited directivity as well as the spectral artifacts can be (clearly) perceived by a listener (at least for low-order systems), which potentially limits the applicability of HOA in basic hearing research. However, when reverberant sound sources are simulated, human localization accuracy is reduced (e.g., Blauert, 1997) and spectral coloration artifacts can be partially masked by coloration that is introduced by the early reflections. Throughout the present study, a speech intelligibility experiment as well as a distance perception experiment is conducted with normal-hearing listeners to evaluate the potential impact of HOA auralization on auditory perception in simulated reverberant environments.

LOUDSPEAKER-BASED ROOM AURALIZATION

Higher-order ambisonics

In HOA, a (plane-wave) sound field is decomposed into spherical harmonics (or Ambisonics components) up to a given order M (e.g., Gerzon, 1973; Daniel *et al.*, 2003). The sound field is then reconstructed by decoding these components onto a given loudspeaker array and thereby deriving input signals for each individual loudspeaker. This reproduction of the original sound field is only accurate around a reference point O, which typically is the center of the loudspeaker array. The radius r of this sweet spot, where the reproduction error is less than 4%, can be approximated by:

$$r = \frac{Mc}{2\pi f}, \qquad \text{(Eq. 1)}$$

with c being the sound velocity and f the considered frequency. Hence, the size of the sweet spot decreases with increasing frequency f and increases with increasing Ambisonic order M. The applicable Ambisonic order M is restricted by the number of available loudspeakers. A full three-dimensional (3D) HOA representation requires $(M+1)^2$ loudspeakers whereas a horizontal plane (2D) representation only requires $2M+1$ loudspeakers.

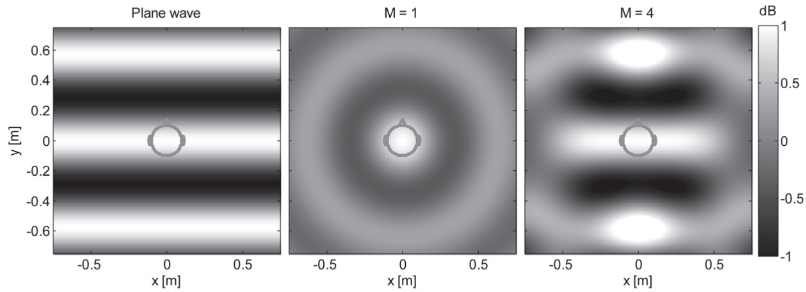

Fig. 1: Sound field reproduction of a plane wave (left panel) with Ambisonics order $M = 1$ (centre panel) and order $M = 4$ (right panel) at frequency f=600 Hz.

Figure 1 (left panel) illustrates the sound field in the horizontal plane for an ideal plane wave traveling towards a listener at a frequency of $f = 600$ Hz. The middle and right panel show the corresponding plane-wave reproduction by 3D HOA with an order of $M = 1$ and $M = 4$, respectively. As the Ambisonic order increases, the reproduced sound field is better described around the listener (i.e., the size of the sweet spot is increased). An alternative way of representing HOA performance is by considering the inherent directivity patterns. These directivity patterns represent the loudspeaker gains as a function of loudspeaker direction. In Fig. 2, horizontal directivity patterns are exemplarily shown for 3D HOA with order $M = 1$ (left panel) and $M = 4$ (right panel) for a (virtual) source directly in front of the listener (i.e., azimuth $\varphi = 0°$). The width of the main lobe decreases (the directivity increases) with increasing Ambisonic order. In consequence, the direction of a virtual HOA source is more accurately perceived with a higher Ambisonic order.

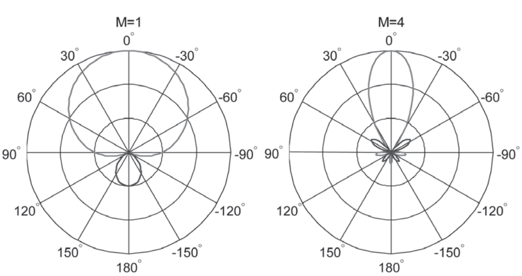

Fig. 2: Two dimensional Ambisonics directivity patterns for order M=1 and M=4.

Room auralization

The loudspeaker-based room auralization (LoRA) system (Favrot and Buchholz, 2010) is applied here as an example system to investigate the perceptual effect of HOA on room auralization. This system relies on the room impulse response (RIRs) description derived by the ODEON room acoustic software (Rindel *et al.*, 2010) as input. For the direct sound and the individual early reflections, detailed information is provided by the acoustic room model: attenuation in 8 octave bands, time and direction of arrival. According to this information, each of these discrete events is reproduced with HOA at the maximum order that can be achieved by the available loudspeaker array. As indicated by the precedence effect (e.g., Blauert, 1997), the direction of the direct sound dominates the overall direction localization. Therefore, it is expected that a high Ambisonic order (i.e., a high directivity) needs to be applied to realize the direct sound in an acceptable manner. For the late reflections, the LoRA system uses the energy and vectorial intensity envelope (in 8 octave bands) of the RIR to calculate 1^{st}-order Ambisonic RIR envelopes. Afterwards, for each individual loudspeaker a separate RIR envelope is derived by applying the Ambisonic decoding principles. These loudspeaker-specific envelopes are then multiplied with white noise which is uncorrelated across loudspeaker channels. In this way a natural diffuseness of the reverberation and very low coloration artifacts are produced. The final RIRs are determined for each loudspeaker by adding the corresponding direct sound, early reflections, and reverberation components.

SPEECH INTELLIGIBILITY

Shirley *et al*. (2007) have shown that speech intelligibility in noise can be deteriorated when the speech is presented by a stereo "phantom" image instead of a single loudspeaker. Here the same principle idea is applied to study the potential effect of HOA-based room auralization on speech intelligibility in diffuse noise. In order to separate the effect on the direct (speech) sound and the early reflections, a similar method as described by Bradley *et al*. (2003) is applied. Speech intelligibility is measured as a function of direct sound energy as well as early reflection energy for different Ambisonics orders.

Method

The experiment was conducted using the LoRA system with a 3D array of 29 loudspeakers in an acoustically-damped room (Favrot and Buchholz, 2010). Nine normal hearing subjects participated in the experiment.

First, a simple auditory scene (one receiver/listener facing one source/talker) was simulated in a classroom with ODEON (See Fig. 3) and the direction, latency and frequency-dependent attenuation of the direct sound (DS) and early reflections (ER) was computed (Favrot and Buchholz, 2009a). The obtained early impulse response was 55 ms long. The direction of the DS and of each ER was then adjusted such that it coincided with the direction of the closest loudspeaker. Each event was auralized either with the single loudspeaker ('0'), 4^{th}-order 3D HOA or standard first-order 3D Ambisonics ('1'). The single loudspeaker auralization method represents infinite-order HOA and provides the reference condition. The sound pressure level of the auralized speech was equalized for these three techniques via an omni-directional microphone placed in the centre of the loudspeaker array.

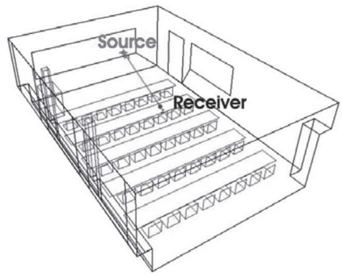

Fig. 3: Source and receiver in the 3D model of the considered classroom.

The speech corpus used in this experiment consisted of the five-word Dantale II Danish Hagerman sentences (Wagener *et al*., 2003). A diffuse speech-shaped noise (SSN) interferer was obtained from cutting the Dantale II monaural SSN into 29 loudspeaker signals. The level of the diffuse SSN was fixed at 60 dB SPL during the whole experiment and the signal-to-noise ratio (SNR) was changed by changing the level of the impulse response. In the first part of the experiment, speech intelligibility scores were measured when only the direct sound was used at four fixed SNRs for each of the three auralization techniques (conditions ‘0’, ‘4’ and ‘1’). In the second part, scores were measured when the direct sound level was fixed and the level of the combined early reflections was changed (conditions ‘00’, ‘11’ and ‘44’). For all conditions, SNRs were normalized according to the individual speech reception threshold (SRT), which was measured for each subject at the beginning of the experiment. Each test-subject participated in a 30-minutes training phase before the actual experiment started. In each run they were asked to select the five words they heard out of ten possible choices (for each word) on a touch screen.

Results

The inter-subject mean intelligibility scores for each part of the experiment are plotted in Fig. 4. The across subject standard deviation is indicated by the error bars. To quantify the measured psychometric function, a sigmoid function was fitted (by optimizing threshold and slope at the inflection point) to the data for each condition.

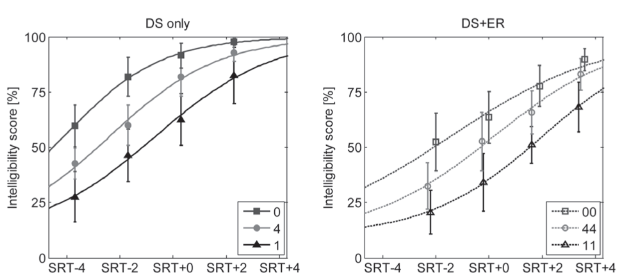

Fig. 4: Mean speech intelligibility scores (markers) with fitted sigmoid function (lines) for the DS-only and DS+ER condition.

For the direct sound only (DS-only) conditions (Fig. 4, left panel), intelligibility scores were significantly dependent on the auralization method. Highest scores were obtained with the single loudspeaker technique and lowest by first-order Ambisonics. The differences were mainly due to a horizontal (threshold) shift of about 2 dB (between conditions '0' and '4') and 4 dB (between conditions '0' and '1'). For the DS+ER conditions (right panel), fitted psychometric curves show a similar significant dependency on auralization technique with a similar 2 dB and 4 dB shift. This decrease of intelligibility when using HOA or when decreasing the Ambisonics order, is likely due to the imperfect sound field reconstruction with Ambisonics which led to degraded spatial cues as well as spectral coloration. When comparing the fitted psychometric functions for the DS-only and the DS+ER conditions, a positive shift as well as a decrease in slope can be observed. Both changes are the same for the three playback conditions. The reduced efficiency when adding ER energy can be interpreted as the consequence of a temporal and spatial spread of energy when the early reflections are added.

Conclusion

This experiment showed that the 4^{th}-order HOA-auralized speech provided less intelligibility than single loudspeaker presented speech and that a decrease in Ambisonic order further decreases speech intelligibility. However, the main difference seems to be a shift of effective speech level which is only dependent on Ambisonic order, but independent of the acoustic condition (i.e., shift is equal for DS-only and DS+ER condition). This indicates that HOA-auralized early reflections provide similar enhancement of speech intelligibility than single loudspeaker reflections. In consequence, the effect of Ambisonics order on speech intelligibility might be simply equalized by an order-specific gain. Pilot listening experiments indicated that this level shift might be linked to a change in loudness, which is not compensated by intensity normalization.

DISTANCE PERCEPTION

Another important characteristic of VAEs is the ability to provide accurate localization of simulated sound sources, both in terms of direction and distance. The following section focuses on distance perception of HOA auralized sources in the frame work of the LoRA system. To evaluate the distance of sound sources, the auditory system makes use of two main auditory cues: the intensity and the direct-to-reverberant energy ratio (Zahorik *et al.*, 2005). The intensity cue is relatively simple to reproduce in a VAE whereas the direct-to-reverberant ratio cue requires that the energy (and maybe the spatial distribution) of the different parts of the impulse response is maintained during the whole processing chain. In the LoRA system, this chain consists in the simulation of the RIR by the acoustic room model, the HOA auralization of the early part of the RIR, and the reverberation auralization method described above.

The aim of this second experiment is to assess the effectiveness of the direct-to-reverberation ratio cue provided by the LoRA system. In order to achieve this goal, distance estimation in the LoRA system needs to be compared to estimations in corresponding real environments. Zahorik (2002) has demonstrated that when recording binaural RIRs (BRIRs) with a dummy-head the same auditory distance perception estimates are measured as when individual BRIRs are used. Following this observation, dummy-head recordings of BRIRs are here used as the real room reference.

Methods

The experiment took place with the LoRA system (Favrot and Buchholz, 2010) and included seven normal-hearing subjects. Two physical rooms at the Technical University of Denmark were considered: a classroom (the same as in Fig. 3) and an auditorium. In each room, BRIRs were recorded at six source-receiver distances with the receiver (i.e., the dummy head) pointing at the source. The considered source-receiver distances were from 0.6 to 6.3 m in the classroom and from 0.5 to 11.2 m in the auditorium. Both rooms were also modeled in ODEON, and RIRs were computed for the same source-receiver positions. These RIRs were then processed by the LoRA system using (i) 4^{th}-order 3D HOA for the direct sound and the early reflections and (ii) the reverberation algorithm described above for the late reflections.

In each simulated room, the sound pressure level (SPL) in the 500 Hz octave band of the derived loudspeaker RIRs were adjusted to match the SPL in the 500 Hz octave band of the recorded BRIRs, for a source-receiver distance of 2.3 m. The loudspeakers and headphones were equalized according to Favrot and Buchholz (2009b). A second condition was derived by normalizing the intensity (IN condition) for all source-receiver distances and thereby suppressing (or at least highly reducing) the effect of intensity on distance perception.

Danish sentences (2 s long) were convolved with the obtained impulse responses. Test-subjects were blind-folded to avoid interference with visual cues. They were asked to orally give an estimate of the sound source distance in meter after each stimulus presentation. For each subject, nine repetitions were measured for all source-receiver distances in eight conditions (i.e., two rooms, two auralization techniques, and with/without intensity normalization).

Results

Individual mean distance estimates are plotted in Fig. 5 for each condition and subject. Error bars indicate the corresponding within-listener standard deviations. In order to model the data, the estimated distances were fitted with a power-law function $r'=kr^{a}$, where r and r' represent the physical and estimated distances, a the power-law exponent, and k a constant. The value of the power-law exponent, a, was always lower than its veridical value of 1, reflecting the well-known compressive behavior of distance perception (e.g., Zahorik *et al*., 2005). The constant k represents the offset in absolute distance estimation and was between 0.5 to 4.7 m for the classroom and from 0.6 to 11.1 m for the auditorium, highlighting the large inter-subject variability.

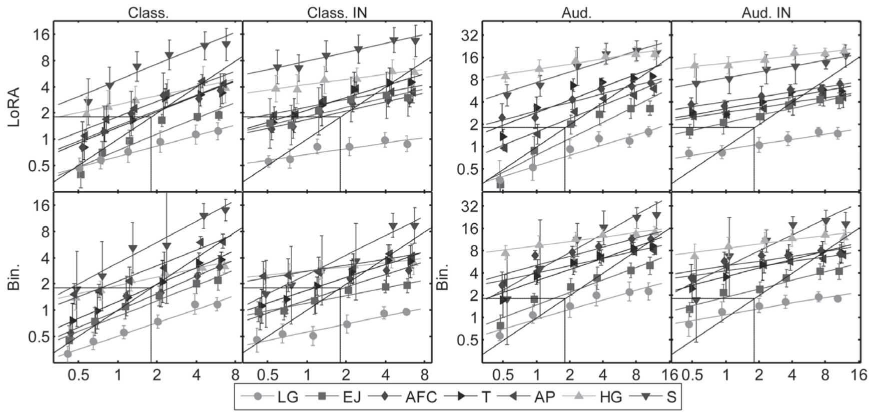

Fig. 5: Mean values and standard deviations of the apparent distance plotted against the physical distance. Axis label units are in meters. Vertical and horizontal dashed lines represent the array radius, and the diagonal dashed lines the physical-estimated identity.

In order to analyse these data, the mean values and confidence intervals of the power-law exponents across subjects are plotted in Fig. 6 for all conditions. The power-law exponents are significantly positive for all conditions, indicating that the subjects were able to perceive a change of distance even when the intensity of the stimuli was normalized with distance. Hence, the LoRA system provides a significant reverberation cue. When comparing both auralization techniques, the difference between exponent *a* and constant *log(k)* was not significantly different from 0 for both rooms, with and without intensity normalization. Hence, the dummy-head recordings and the corresponding LoRA simulations result in similar distance estimates, indicating that the LoRA system provides realistic distance cues.

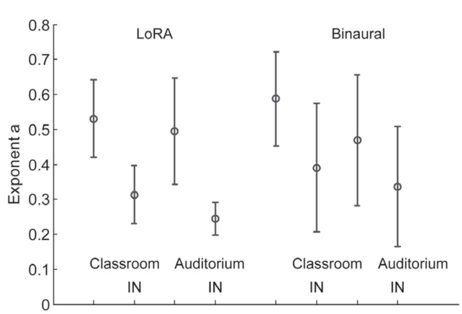

Fig. 6: Inter-subjects mean values and confidence intervals of the fitted power-law exponent *a*.

CONCLUSION AND PERSPECTIVE

Two experiments were conducted to investigate the impact of HOA on auditory perception in simulated reverberant environments: (i) a speech intelligibility experiment and (ii) a distance perception experiment. The results from the speech intelligibility experiment indicated that HOA auralization can have a significant effect on overall speech intelligibility performance. However, this effect might be compensated by the introduction of a simple gain factor that is only dependent on the Ambisonic order. The distance perception experiment revealed that modern room simulation techniques combined with HOA auralization provide significant reverberation cues that result in realistic distance perception. Therefore, it can be concluded that VAEs (as generated by the LoRA system) can be suitable for conducting psychoacoustic experiments, at least for speech intelligibility and distance measures.

In the case that VAEs are applied to test the processing of hearing instruments, VAEs are not only required to provide realistic auditory perception, but also to provide realistic input signals to the hearing device. In this regard, it is unclear to what degree HOA-based auralization might impact the analysis of hearing aid processing strategies in simulated environments, in particular when spatial processing techniques are considered that utilize multiple microphones (e.g., adaptive beamforming). Hence, future research will need to investigate the effect of VAEs on hearing instrument signal processing.

REFERENCES

Blauert, J. (**1997**). *Communication acoustics* (MIT Press, Cambridge, MA).

Bradley, J. S., Sato, H., and Picard, M. (**2003**). "On the importance of early reflections for speech in rooms," J. Acoust. Soc. Am. **113**, 3233-3244.

Dalenbäck, B. I. (**1996**). "Room acoustic prediction based on a unified treatment of diffuse and specular reflection," J. Acoust. Soc. Am. **100**, 899–909.

Daniel, J., Nicol, R., and Moreau, S. (**2003**). "Further investigations of high order Ambisonics and wavefield synthesis for holophonic sound imaging," AES 114th Convention, preprint 5788.

Favrot, S., and Buchholz, J. M. (**2009a**). "Validation of a loudspeaker-based room auralization system using speech intelligibility measures," AES 126th Convention, preprint 7763.

Favrot, S., and Buchholz, J. M. (**2009b**). "Distance perception in loudspeaker-based room auralization," AES 127th Convention, preprint 7854.

Favrot, S., and Buchholz, J. M. (**2010**). "LoRA – A loudspeaker-room auralization system," Acta Acustica/Acustica **96**, 364-375.

Gerzon, M. A. (**1973**). "Periphony - with-height sound reproduction," J. Audio Eng. Soc. **21**, 2-10.

Rindel, J. H., Nielsen, G. B., and Christensen, C. L. (**2010**). "ODEON website," www.odeon.dk, last viewed 10/01/2010.

Shirley, B., Kendrick, P., and Churchill, C. (**2007**). "The effect of stereo crosstalk on intelligibility: comparison of a phantom stereo image and a central loudspeaker source," J. Audio Eng. Soc. **55**, 852-863.
Wagener, K., Josvassen, J. L., and Ardenkjaer, R. (**2003**). "Design, optimization and evaluation of a Danish sentence test in noise," Int. J. Audiology **42**, 10-17.
Zahorik, P. (**2002**). "Auditory display of sound source distance," Proc. Int. Conf. on Auditory Display 326-332.
Zahorik, P., Brungart, D. S., and Bronkhorst, A. W. (**2005**). "Auditory distance perception in humans: A summary of past and present research," Acta Acustica/Acustica **91**, 409-420.