An Assessment of Spatial Fidelity of Higher-Order Ambisonics Reproduction Using the SCaLAr System

Thomas Deutsch, Lukas Vollmer and Janina Fels

Institute for Hearing Technology and Acoustics, RWTH Aachen University, 52074 Aachen,
E-Mail: thomas.deutsch@akustik.rwth-aachen.de

Introduction

Surrounding spherical loudspeaker arrays enable the reproduction of virtual acoustic environments by employing a variety of spatial sound reproduction methods. The design and implementation of such a system requires careful consideration of the practical limitations, such as spatial dimensions and placement accuracy. A 68-channel system, named "SCaLAr", with extensive reproduction capabilities was recently developed and installed at the Institute for Hearing Technology and Acoustics, RWTH Aachen University. Details of the implementation and objective evaluation of the system can be found in [1]. In this contribution, a short overview of the most important properties of the SCaLAr system is provided, and the first part of a subjective evaluation by means of a localization experiment comparing different orders of Ambisonics and real sound source reproduction is presented.

The SCaLAr system

SCaLAr is a surrounding spherical cap loudspeaker array, shown in Figure 1, with a sphere radius of 1.35 m. It is a capped array, since there are no loudspeakers installed in or below the grid floor of the anechoic chamber. The array mainly consists of 68 custom-built loudspeakers with individual audio channels, mounted on an equal-area grid. Four additional subwoofers are available for low frequency reproduction. The whole system is installed in the fully anechoic chamber of the Institute for Hearing Technology and Acoustics, RWTH Aachen University.

Figure 1: The SCaLAr loudspeaker array, installed in the anechoic chamber of the Institute for Hearing Technology and Acoustics.

The enclosures of the 68 loudspeakers are built in an octagonal shape and are designed as a bass reflex system. They are equipped with a 4-inch coaxial 2-way driver, with a crossover frequency of 2 kHz. Amplification is realized through "PPA 1000-4-PC" power amplifiers with DSP capabilities by Four Audio. Using these capabilities, every loudspeaker frequency response is equalized with an individual 512-taps FIR-filter.

Communication between the host computer and the DSP power amplifiers is realized via Ethernet. Audio signals are transmitted using a 128-channel Dante interface.

Several different spatial reproduction algorithms have been implemented to be used or evaluated with SCaLAr. These include vector base amplitude panning (VBAP), multiple-direction amplitude panning (MDAP), multi-channel crosstalk cancellation (CTC), and higher-order Ambisonics (HOA). In the case of HOA, several different decoders are available: sampling Ambisonic decoder (SAD), mode-matching Ambisonic decoder (MMAD), energy-preserving Ambisonic decoder (EPAD), and all-round Ambisonic decoding (AllRAD(+)).

To allow tracking of a person or objects inside the spherical array, an optical tracking system was installed, consisting of four "Flex 13" infrared cameras working together with the corresponding software "Motive" by NaturalPoint, Inc.

SCaLAr's main purpose is providing a powerful and flexible tool for the reproduction of complex acoustic environments and the implementation of psychoacoustic experiments, e.g., in the context of auditory selective attention. Additionally, it enables the evaluation and comparison of spatial audio reproduction methods.

Localization experiment

In order to further validate SCaLAr for the intended use cases, a localization experiment was conducted as the first part of the system's subjective evaluation. The devised method comprised the comparison of participants' performance when using real and virtual sound sources modeled with different orders of Ambisonics reproduction. Due to the advantages and the flexibility of the AllRAD algorithm [2] for practical applications, cf. [1], this algorithm was chosen for this assessment. The virtual decoding stage employed a 15-design with 120 points followed by the VBAP stage mapping the virtual to the real loudspeaker array.

In order to assess the resulting gain in localization accuracy, Ambisonics orders 1, 3, 5, 7, and the system's equivalent Ambisonics order 9 [2] were selected for the evaluation.

Our hypothesis for the experiment was that localization accuracy increases with increasing Ambisonics order, while the best performance is achieved when using real sound sources.

Setup

Six additional loudspeakers serving as real reference sources were installed in the frame of the array. Four of the six loudspeakers were installed in the horizontal plane, with azimuth angles of 40° and 90° (to the left of the participant), as well as –18° and –57° (to the right of the participant). The two remaining loudspeakers were installed in the median plane with elevation angles of 20° and 32° (above the horizontal plane).

These additional loudspeakers were the same type of octagonal loudspeaker as the 68 loudspeakers of the SCaLAr system. Furthermore, the same power amplifiers were used, and the speakers were equalized in the same way as the SCaLAr loudspeakers.

A seventh loudspeaker, which is part of the regular array and located at 0° azimuth, 0° elevation (defining the front of the array) was included as a reference sound source as well.

The virtual (point) sound sources were modeled at the same positions as the real sound sources with respect to the center of the array. A constant distance of 1.45 m was used for the virtual sound sources, while the distances of the real sound sources varied slightly due to the given mounting conditions in the frame of the array. However, individual loudspeaker gains and filter delays were adjusted to account for these offsets.

The combination of seven different sound source directions, with real sources and five different orders of Ambisonics reproduction resulted in 42 different test conditions. Each of these conditions was tested four times, resulting in 168 test runs. Stimuli orders were counterbalanced using a uniformly distributed random Latin square design [3].

Stimulus

The stimulus played back in every test run was a sequence of ten pulses of pink noise with a sound pressure level of 60 dB. Every pulse consisted of 300 ms of pink noise with 10 ms fade-in and fade-out using half Hann-windows, followed by 300 ms of silence between two adjacent pulses.

Procedure

Participants received written instructions first and were subsequently familiarized with the setup in the anechoic chamber, guided by the instructor. Before the actual experiment, each participant completed a short training session consisting of three example runs, during or after which remaining questions of the participants were answered.

A calibration procedure for the optical tracking system was performed for each participant before the start of the experiment, using a cross line laser to ensure a correctly aligned head position.

The instructor left the room during the experiment, but the participant could be monitored at all times via a webcam installed in the chamber. An intercom system was available as well, allowing bidirectional verbal communication.

The participant was seated on a chair inside the array (see Figure 2). In order to track head position and orientation using the optical tracking system, participants put on a cap made of soft fabric, fitted with retroreflective markers. To keep participants from being distracted or influenced by visual information, participants wore blinding goggles during the experiment.

Figure 2: A participant sitting inside the array during the experiment, wearing a reflective marker cap and blinding goggles.

Progress during the experiment was controlled by the participant by pushing a single button on a hand-held controller.

Participants were instructed to orient their head towards the front before starting a test run via a button press. After the button was pressed, the stimulus was played back, and the direction of incidence was indicated by the participant using the head pointing method. The head pointing method is known to be very intuitive for participants and can produce more accurate results than other existing options, e.g., verbal reports, graphical interfaces or the so-called GELP method [4][5][6]. Since the experiment's purpose was not to determine human sound source localization ability depending on the direction of incidence, but rather a critical assessment of the spatial fidelity of the reproduction system, participants were allowed to rotate their head and try to locate the sound source right after playback of the stimulus started. When the button was pushed again after the pulsed noise sequence was finished, the tracking data for the current stimulus was logged. It took approximately 25 minutes to complete the main part of the listening test.

With respect to possible sound source directions, participants were only told that no sound source position would be right behind them. They were explicitly allowed to rotate their upper body, should they feel that the mere rotation of the head would not be sufficient to indicate the perceived sound source direction.

Participants

A number of n = 11 participants with normal hearing completed the listening test (ages 23–38 years, mean age

29.2 years). Every participant provided written consent to take part in the study.

Results

To account for slight head center position deviations from the center of the array during a participant's response, an optimum response angle for the current sound source direction was calculated using the tracked head position data for each individual test run. Subsequently, deviation values in degrees were obtained by calculating the difference between the actual response angles and the optimum response angles. This was done for azimuth and elevation angles. Response data of repeated test runs was reduced to the mean response for each participant and condition. Separate means based on absolute deviations and sign dependent deviations were calculated.

The results for the sound sources that were located 90° to the left in the horizontal plane were omitted in the analyses presented in the following, since there seemed to be a physical restraint problem for participants for a source positioned so far sideways.

Azimuth deviations

As can be seen from the boxplot shown in Figure 3, the median values of the azimuth angle deviations are almost 0° for all source types.

A two-way ANOVA for the dependent variable *azimuth deviation* with the two factors *source type* and *source position* supports this initial observation by revealing that there is no statistically significant difference for the means of *source type* ($p = .602$), i.e., the different HOA orders and the real sources. The differences for the factor *source position* are statistically significant ($p < .001$), while there is no statistically significant interaction between the two factors ($p = .991$).

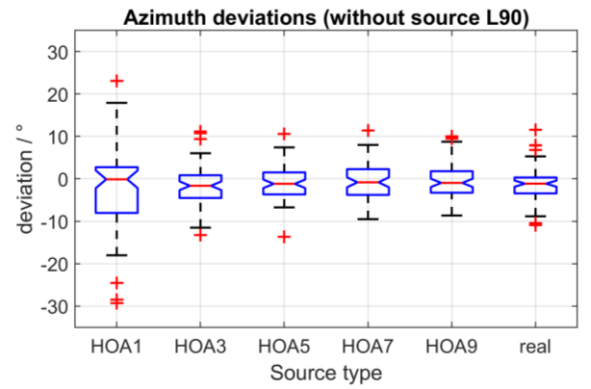

Figure 3: Boxplot showing azimuth deviations in degrees for all source types (HOA orders 1–9 and real sources), including all source positions except 90° left.

The differences in the reliability of the localization for different reproduction types are illustrated more clearly when looking at the absolute azimuth deviations, shown in Figure 4. It can be seen that the deviations are highest for the lowest Ambisonics order 1 (median of 8°), while the median values for all other source types are significantly lower and very similar to each other (median < 4°). The low values in the case of the real sources and HOA orders ≥ 3 are in accordance with the findings in literature concerning the high localization resolution of human binaural hearing for frontal sound incidence [7].

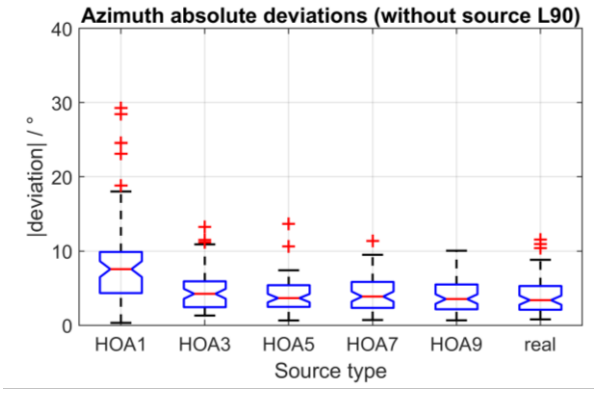

Figure 4: Boxplot showing absolute azimuth deviations in degrees for all source types (including all source positions except 90° left).

Again, an equivalent ANOVA conducted for the dependent variable *absolute azimuth deviation* supports the initial observation by showing that in this case, a statistically significant main effect for the factor *source type* is present ($p < .001$). The Bonferroni test confirms the assumption that this result is caused by the factor level *HOA1*, which significantly differs from all other factor levels ($p < .001$). As before, there is no interaction between factors *source type* and *source position*. It is worth noting that the real source reproduction did not lead to a better performance than HOA orders ≥ 3.

Elevation deviations

While the median values of the elevation deviations are close to 0° for all orders of the HOA reproduction (see Figure 5), this is not the case for the real sources: With a median value of 4°, localization seems to have been too high. Since the real source reproductions represent reference conditions, this slight elevation found in the results is most likely an artifact of the measurement method, and it has to be concluded that the localization in case of the HOA conditions was actually slightly too low. A possible explanation for this is the behavior of the decoder in the case of lower Ambisonics orders, as it distributes comparatively high amounts of energy to the lower loudspeaker ring close to the polar cap.

In the corresponding ANOVA, significant main effects for *source type* ($p = .001$) and *source position* ($p < .001$) are found (with no significant interaction). The Bonferroni test reveals that differences in means are statistically significant between real source reproduction and HOA order 1 ($p < .001$) as well as between real sources and HOA order 3 ($p = .048$), at a significance level of $\alpha = .05$.

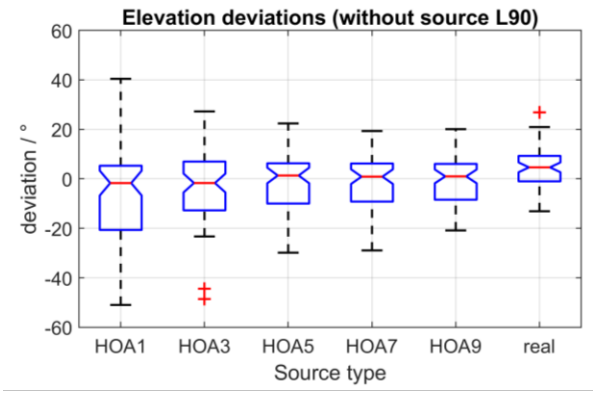

Figure 5: Boxplot showing elevation deviations in degrees for all source types (including all source positions except 90° left).

While the median values for HOA orders ≥ 3 and the real sources were approximately the same in the case of the absolute azimuth deviations, a more gradual decrease in absolute deviation can be observed for the elevation when Ambisonics order is increased, see Figure 6. Here, the real sources seem to exhibit an even smaller deviation than HOA order 9 sources.

Again, significant main effects for *source type* ($p < .001$) and *source position* ($p < .001$) can be found through an ANOVA for the dependent variable *absolute elevation deviation*. For the dataset at hand, the real source reproduction only significantly differs from HOA orders 1 and 3 ($p < .001$). The results for HOA orders 7 and 9 are significantly better than for the lower orders 1 and 3 as well.

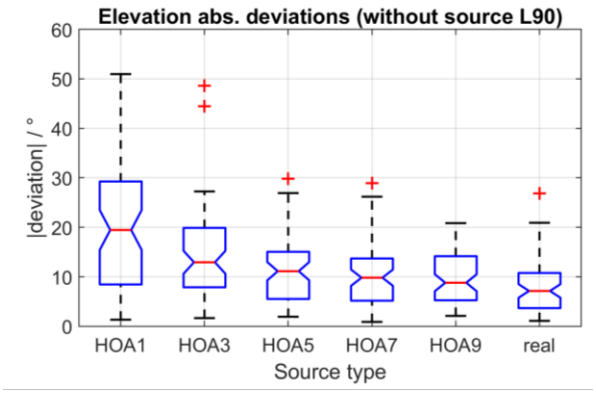

Figure 6: Boxplot showing absolute elevation deviations in degrees for all source types (including all source positions except 90° left).

Discussion and conclusion

A general observation is that deviations are larger in case of the elevation angles when compared to the deviations of the azimuth angles. This might be expected, since localization accuracy in the median plane is usually worse than in the horizontal plane [7].

As expected, none of the HOA orders produced better results than the real source reproduction, though the latter did not prove to be superior in terms of localization accuracy or reliability for azimuth angles when compared to HOA orders ≥ 3. In the case of elevation angles, the performance of HOA reproduction was only significantly worse than real source reproduction for orders < 5, though the data seem to indicate a trend for improved performance when using real source reproduction.

Outlook

As a next step for the validation and assessment process, a spatial release from masking experiment will be conducted. Similar to the localization experiment described in this contribution, the performance of the participants based on speech reception thresholds will be measured and compared when using real sound sources and virtual sound sources with different orders of Ambisonics reproduction.

References

[1] Pausch, F., Behler, G., Fels, J.: SCaLAr – A surrounding spherical cap loudspeaker array for flexible generation and evaluation of virtual acoustic environments. Acta Acustica, vol. 4, no. 5 (2020)

[2] Zotter, F., Frank, M.: All-round Ambisonic panning and decoding. Journal of the audio engineering society, vol. 60, no. 10 (2012)

[3] Jacobson, M., Matthews, P.: Generating uniformly distributed random Latin squares. Journal of Combinatorial Designs, vol. 4, no. 6 (1996), 405-437

[4] Gilkey, R., Good, M., Ericson, M., Brinkman, J., Stewart, J.: A pointing technique for rapidly collecting localization responses in auditory research. Behavior Research Methods, Instruments, & Computers, vol. 27, no. 1 (1995), 1-11

[5] Carlile, S., Leong, P., Hyams, S.: The nature and distribution of errors in sound localization by human listeners. Hearing research, vol. 114, no. 1-2 (1997), 179-196

[6] Djelani, T., Pörschmann, C., Sahrhage, J., Blauert, J.: An interactive virtual-environment generator for psychoacoustic research II: collection of head-related impulse responses and evaluation of auditory localization. Acta Acustica united with Acustica, vol. 86, no. 6 (2000), 1046-1053

[7] Blauert, J.: Räumliches Hören. S. Hirzel Verlag, Stuttgart, 1974