Localization Experiments Using Different 2D Ambisonics Decoders

(Lokalisationsversuche mit verschiedenen 2D Ambisonics Dekodern)

Matthias Frank, Franz Zotter**, Alois Sontacchi***

* Graz University of Technology, student, matfrank@sbox.tugraz.at
**Institute of Electronic Music and Acoustics, {zotter,sontacchi}@iem.at

Abstract

This study presents the results from localization experiments of virtual sound sources using a 12 channel, nearly circular 2D Ambisonics system. The perceived direction of the sound and a subjective rating of the localization accuracy has been assigned to each virtual source. As playback methods, Ambisonics decoders with different order and spatial smoothing (basic, max r_E, in-phase) are evaluated. In each case, the evaluation has been carried out over two listening positions: within and outside the Ambisonics listening area. The analysis shows the reproduction accuracy of different Ambisonics variants within the studied playback situation, and allows for comparison. Furthermore, the test includes an investigation concerning the presence of a compensation of loudspeaker signal delays to the center.

1. Introduction

There are various spatial sound reproduction systems, each one of which exhibiting its characteristic limits and errors. So there is the need for evaluation concering the audible resolution/artifacts. Several studies exist on VBAP, for instance [1], and WFS, e.g. [2]. There are, however, only few currently emerging studies on the performance of Ambisonics restitution systems, see [3, 4, 5]. Most studies have been carried out in anechoic rooms, in order to minimize artifacts due to the room acoustics. Some works [6, 7] have investigated the localization in ordinary (reverberant) rooms, but primarily using monophonic sound sources. Therefore, the motivation for this study was a combination: evaluation of Ambisonics within an "ordinary" listening room and only near-circular setup. This paper presents results of a listening test that studies the effect of different decoder variants and listening positions, as well as the implications of an appropriate delay compensation for the loudspeaker positions.

2. Ambisonics

One of the most comprehensive works about Ambisonics is [8]. The following paragraphs recapitulate the small cross-section of the theory applied within this study.

2.1. Encoding

For the 2D Ambisonics system to describe an angle of incidence, a delta distribution at the angle ψ is decomposed into its circular Fourier coefficients. The coefficients $\mathbf{c}(\psi)$ truncated to

the order M, constitute the *encoder* of a signal $x[n]$ into a vector $\boldsymbol{\chi}[n]$ of Ambisonics signals

$$\begin{aligned}\boldsymbol{\chi}[n] &= \mathbf{c}(\psi)\; x[n] \\ &= \left[\frac{1}{\sqrt{2}},\; \cos(\psi),\; \sin(\psi),\; \cos(2\psi),\; \sin(2\psi),\; \ldots\; \cos(\mathrm{M}\psi),\; \sin(\mathrm{M}\psi)\right]^{\mathrm{T}} x[n],\end{aligned} \tag{1}$$

which has limited angular resolution and therefore allows spatial discretization without losses. In principle, this decomposition of the delta distribution is the angular Green's function. The factor of the radial Green's function is considered given due to the circular playback situation.

2.2. Decoding

Assume a circular loudspeaker setup with the angles $\{\psi_1,\; \ldots,\; \psi_{\mathrm{L}}\}$ and its signals $\mathbf{y}[n]$. The resulting vector of loudspeaker Ambisonics signals $\boldsymbol{\Upsilon}[n]$ is described by decomposition of the l^{th} loudspeaker at the angle ψ_l in its circular harmonics representation. Just as described above, this is the M-truncated transform of a delta distribution located at ψ_l. Putting the coefficients $\mathbf{c}(\psi_l)$ into a matrix $\mathbf{C}$ for all loudspeakers $l = 1, \ldots, \mathrm{L}$, we obtain

$$\boldsymbol{\Upsilon}[n] = \mathbf{C}\,\mathbf{y}[n] = [\mathbf{c}(\psi_1),\; \mathbf{c}(\psi_2),\; \ldots \mathbf{c}(\psi_{\mathrm{L}})\;]\;\mathbf{y}[n]. \tag{2}$$

The task of the *decoder* $\mathbf{D}$ is now to derive the loudspeaker signals $\mathbf{y}[n]$ from the Ambisonics encoded input signal $\boldsymbol{\chi}[n]$

$$\mathbf{y}[n] = \mathbf{D}\;\boldsymbol{\chi}[n], \tag{3}$$

so that the vector of the loudspeaker Ambisonics signals $\boldsymbol{\Upsilon}[n]$ matches exactly the input $\boldsymbol{\chi}[n]$

$$\boldsymbol{\Upsilon}[n] \stackrel{!}{=} \underbrace{\mathbf{C}\,\mathbf{D}}_{\stackrel{!}{=}\mathbf{I}}\;\boldsymbol{\chi}[n], \tag{4}$$

i.e. $\mathbf{D}$ shall be inverse to $\mathbf{C}$. Since for arbitrary layouts $\{\psi_1,\; \ldots,\; \psi_{\mathrm{L}}\}$ the matrix $\mathbf{C}$ is in general neither orthogonal nor squared, a suitable right-inverse ($\mathbf{C}\,\mathbf{C}^{\dagger} = \mathbf{I}$) is computed

$$\Longrightarrow \mathbf{D} = \mathbf{C}^{\dagger} = \mathbf{C}^{\mathrm{T}}(\mathbf{C}\mathbf{C}^{\mathrm{T}})^{-1}. \tag{5}$$

In order to control the main and side lobes emerging from circular harmonics truncation, a weighting vector $\mathbf{w}$ is applied to the harmonics domain, and regarded as a part of the decoder.

Finally, we arrive at the complete synthesis equation with encoding $\mathbf{c}(\psi)$ and decoding $\mathbf{D}$ (without distance coding)

$$\mathbf{y}[n] = \mathbf{D}\;\mathrm{diag}\,\{\mathbf{w}\}\;\mathbf{c}(\psi)\,x[n]. \tag{6}$$

Table 1 shows the decoder weights $\mathbf{w}$ used in this study.

decoder	basic	max r_E	in-phase
weight $w[m]$	1	$\cos\left(\frac{m\pi}{2\mathrm{M}+2}\right)$	$\frac{\mathrm{M}!^2}{(\mathrm{M}+m)!(\mathrm{M}-m)!}$

Table 1: Decoder weights from [8]: $\mathbf{w}^{\mathrm{T}} = (w[0],\; w[1],\; \cdots,\; w[\mathrm{M}])$.

3. The Experiment

3.1. Task

The remainder of this paper gives a characterization of the perceived direction applying the reproduction principle in the environment described below. The evaluation task of the subjects consists of two different things: a perceived direction of the sound and a subjective rating of the localization accuracy [2]. From the subjective rating, a mean opinion score (MOS) is computed.

3.2. Test Environment

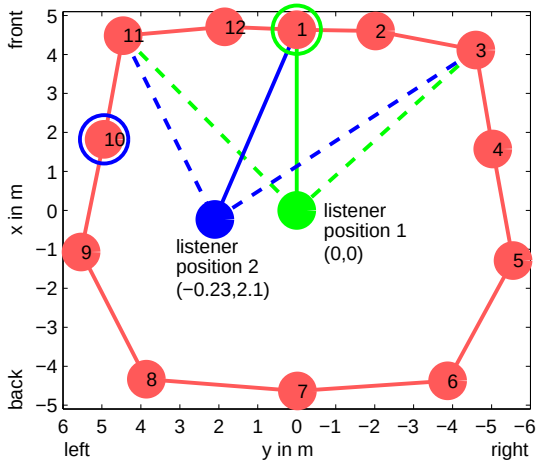


Figure 1: Loudspeaker and listener positions.

As test environment, the "CUBE" at the University of Music and Dramatic Arts Graz was chosen, with a configuration depicted in figure 1. Red dots indicate the positions of the non-equispaced loudspeakers, the green and blue spots show the two listening positions, and the circles mark the nearest loudspeakers. The compensation of the loudspeaker signal delays at listenting position 1 are given in table 2. The room is $10\text{m}\times12\text{m}\times4\text{m}$ with parquet floor and RT-60 < 1s (broadband). There was no possibility to curtain the loudspeakers, so they were visible during the experiment. In order to reduce acoustic floor-reflections and simulate other listeners, stage molleton has been spread. During the experiments, the orientation of the subjects has been adjusted aiming towards the first loudspeaker, at both listening seats (see solid green and blue line in figure 1). This orientation is supposed to be the ordinary use case in performance situations. Orientation and position (also height) of the subjects have been monitored using a head tracking system, to stay within the limits of ±4cm and $\pm10°$ while listening.

speaker	1	2	3	4	5	6	7	8	9	10	11	12
delay [ms]	4.76	3.56	0.16	2.68	1.36	0.86	4.51	1.32	1.90	3.04	0.00	3.61

Table 2: Delay compensation (rounded to integer samples).

3.3. Method

The perceived direction of the sound is measured with a pointing device. It was decided to use a toy-gun which is tracked by a 15 camera infrared motion capture system (also used for monitoring the head position). In order to compare them with the target, the angles pointed at by the subjects are converted into the polar (spherical) coordinate system of the playback setup. The subjective rating of the localization accuracy has been given on a 5-point-scale. All parameters are stored by pressing buttons on the pointing device. The laser pointer mounted on the toy-gun proved useless for the aiming task, because of the tiny point size, so the subjects used the ironsights. The overall error, when aiming at visual objects has been found to be less than $0.5°$ and therefore being sufficiently accurate [9].

3.4. Stimulus

Broadband pink noise has been chosen as the stimulus. Because of its large frequency range, many localization cues are available [9]. The stimulus is divided into 4 periods. This division is

based on other localization experiments [2, 1] and our own preliminary tests. Each period has a fade-in and fade-out time of $100\mathrm{ms}$, as well as $200\mathrm{ms}$ of unattenuated noise in between. The periods are separated by $100\mathrm{ms}$ of silence (see figure 2), and the entire stimulus lasts $2\mathrm{s}$.

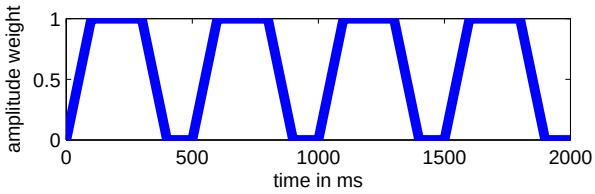


Figure 2: Envelope of the stimulus.

The decoder variants under test (weighting, order) and spatialization angles are listed in table 3. Regardless of the decoder order, all loudspeakers are used for reproduction. The angles lie within the interval between $\pm 40°$ quantized to $5°$ steps. For each decoder variant, 5 angles have been selected: $0°$, and randomly: one left and one right near $0°$, and two farther left and right. The subjects were presented the stimuli in a mixed chronological order including the spatial angle, decoder variant, and 3 real sources (loudspeakers 1, 2, 12).

decoder	**order**	**delay compensation**	**no. of angles**
basic	1	no	5
basic	1	yes	5
basic	3	no	5
basic	3	yes	5
basic	5	no	5
basic	5	yes	5
max r_E	1	no	5
max r_E	1	yes	5
max r_E	3	no	5
max r_E	3	yes	5
max r_E	5	no	5
max r_E	5	yes	5
in-phase	1	yes	5
in-phase	3	yes	5
in-phase	5	yes	5
real source		no	3
		TOTAL	**78**

Table 3: Set of the 78 stimuli per subject and position.

3.5. Listeners

Fifteen subjects participated in this experiment. The population included 2 females and 13 males, ranging in age from 23 to 35 years (median age was 28).

3.6. Experiment procedure

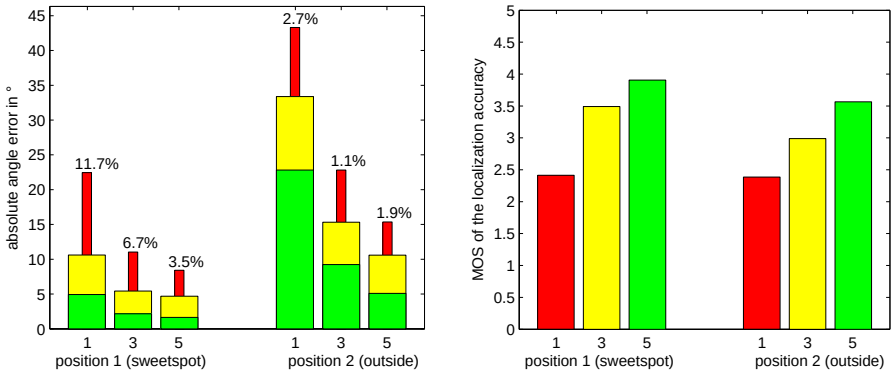
Although the toy-gun is handy in terms of sufficiently accurate pointing, each subject did an aiming exercise to become familiar with the pointing device. In order to get an idea of the range in which to rate the localization accuracy, the subjects were presented 11 stimuli at both listening positions before the experiment. Furthermore, there was a 10 minutes break prior to shifting to the second listening position. After the break, the 11 example stimuli were repeated at both

positions to maintain consistency. Every single measurement took about 10s and comprised listening, aiming and rating. If desired, the stimulus could be repeated.

4. Analysis

4.1. Effect of decoder order

As the order of the decoder increases, the angle error (difference between perceived and reproduced angle) decreases and the subjective rating of the localization accuracy improves, see figure 3. This fact is independent of the listening position. In terms of the listening positions, position 1 (*sweetspot*) exhibits less errors than position 2 and an improved rating of the accuracy.



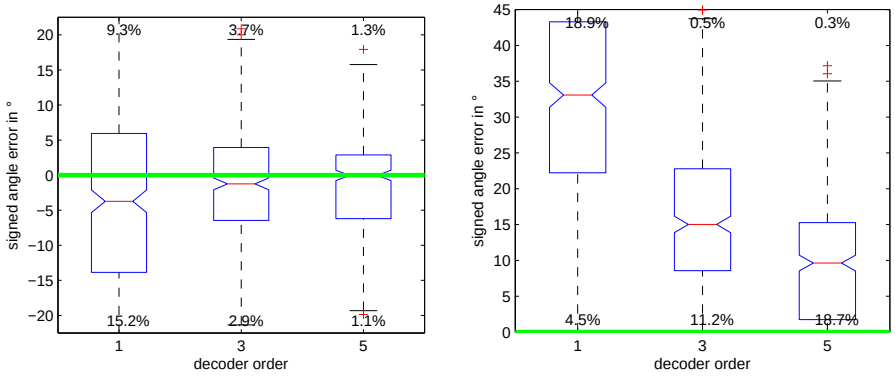
(a) Absolute angle error: colored boxplot (median separates yellow from red), percentages show outliers (error > highest quartil + 1.5·IQR).

(b) MOS (mean opinion score: average of the subjective rating) of the localization accuracy.

Figure 3: Effect of the decoder order at both listening positions (using all decoders).

4.1.1. Position 1 (sweetspot)

That the results at position 1 are better is also evident regarding the signed angle errors, see figure 4. At position 1, the median only shows a small offset, even for the lowest order. At higher orders, the interquartile range (IQR) as well as the number of outliers decreases.



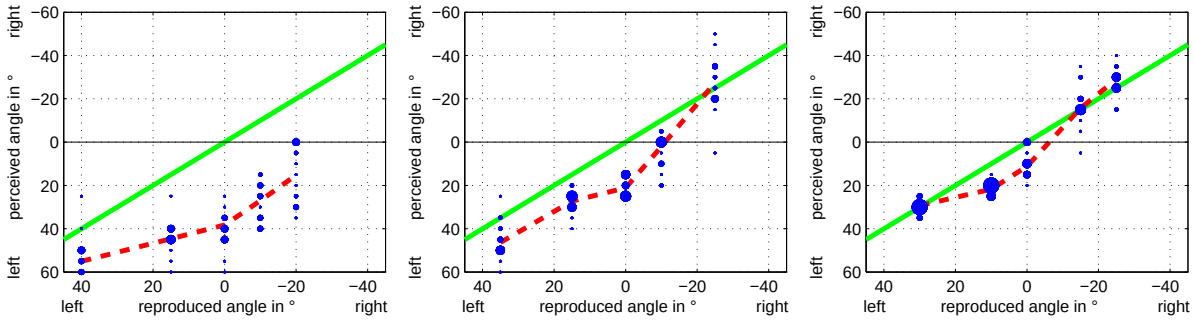
(a) Position 1.

(b) Position 2.

Figure 4: Signed angle errors at both listening positions (using all decoders): boxplot, percentages show values outside the plot range. The 2 plot ranges have different limits but are in same scale.

4.1.2. Position 2 (outside sweetspot)

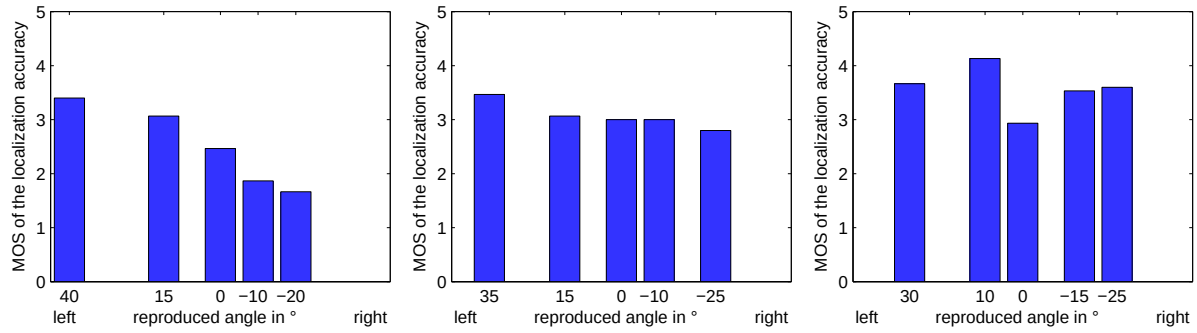
At position 2, there is a large bias of the median localized angle towards the left. This bias gets smaller at higher orders. Figure 5 provides an overview over this behavior, plotting the perceived angle as a function of the reproduced target angle.



(a) 1^{st} order. (b) 3^{rd} order. (c) 5^{th} order.

Figure 5: Effect of increasing orders at position 2: angle mapping, histogram for each reproduced angle, bubble size indicates the number of values per $5°$ division, the dashed line plots the medians (example for max r_E decoders without delay compensation).

The large bias towards the left most probably results from the proximity of the listener to the loudspeakers on the left, see figure 1. As the wide main lobe of the 1^{st} order nearly covers the semi-circle, this proximity even affects target angles on the right. For higher orders, the main lobe gets narrower, and the effect described above diminishes. For the 5^{th} order, the bias only affects target angles near $0°$.



(a) 1^{st} order. (b) 3^{rd} order. (c) 5^{th} order.

Figure 6: Effect of increasing orders at position 2: MOS (mean opinion score: average of the subjective rating) of the localization accuracy (example for max r_E decoders without delay compensation).

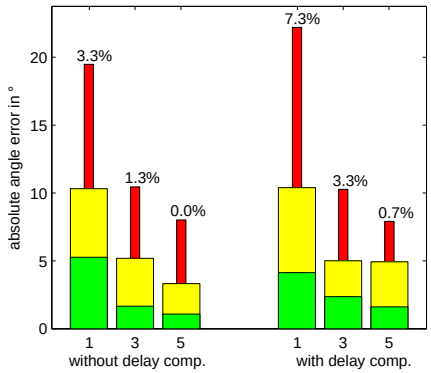
For the lowest decoder order, the MOS (subjective rating of localization accuracy) decreases towards the right, see figure 6. This effect is attributed to the distances between the loudspeakers and the listener, too. At higher orders, the average rating improves and the dependancy on the reproduction angle decreases.

4.2. Effect of the delay compensation

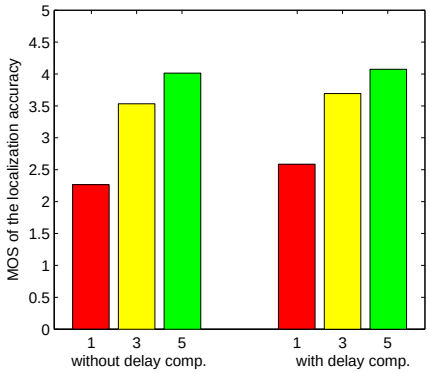
4.2.1. Position 1 (sweetspot)

Using delay compensation, the amount of front/back confusion grows at listening position 1, see figure 7. Regarding the medians of the absolute angle error, the value without delay com-

pensation is only smaller at the highest order. For 1^{st} and 3^{rd} order, however, the subjective rating (MOS) is slightly higher with the compensation, even significantly for the 1^{st} order. This over estimation is probably due to phase distortions in the sound, which can be perceived as sound coloration. The awareness of the subjects not to rate sound quality might lead to this bias.



(a) Absolute angle error: colored boxplot (median separates yellow from red), percentages show errors > 120° (indicates front/back confusion).

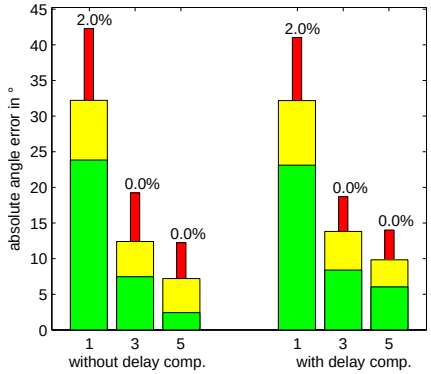


(b) MOS (mean opinion score: average of the subjective rating) of the localization accuracy.

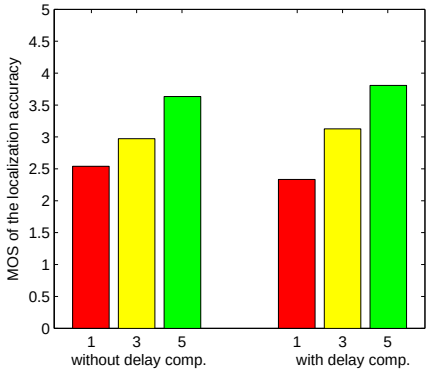
Figure 7: Effect of the delay compensation at position 1 (using max r_E and basic decoders).

4.2.2. Position 2 (outside sweetspot)

The absolute angle error shows reduced front/back confusion compared to listening position 1, see figure 8. The effect of the delay compensation is not evident in the amount of confusion, here. Apart from that, detection of the delay compensation based on the median angle error is possible at higher orders. The ratings (MOS) are not significantly different.



(a) Absolute angle error: colored boxplot (median separates yellow from red), percentages show errors > 120° (indicates front/back confusion).



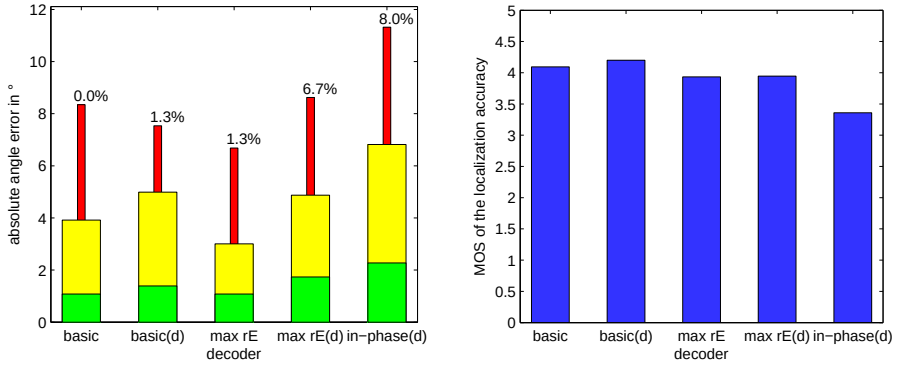
(b) MOS (mean opinion score: average of the subjective rating) of the localization accuracy.

Figure 8: Effect of the delay compensation at position 2 (using max r_E and basic decoders).

4.3. Best decoder

As shown in section 4.1, the best results, i.e. the smallest angle errors and the best MOS ratings, are achieved by using the highest order decoders, at both positions. Consequently, to find the best decoder for each listening postion, it is sufficient to concentrate on 5^{th} order.

4.3.1. *Position 1 (sweetspot)*



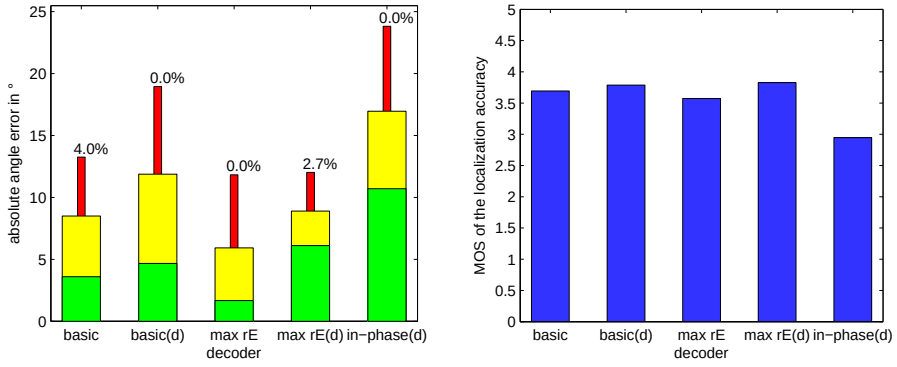
(a) Absolute angle error: colored boxplot (median separates yellow from red), percentages show outliers (error > highest quartil + 1.5·IQR).

(b) MOS (mean opinion score: average of the subjective rating) of the localization accuracy.

Figure 9: Absolute angle error and MOS for 5^{th} order decoders (d = with delay compensation) at position 1.

Concerning the absolute angle error, the best decoder at position 1 is the max r_E ahead of the basic decoder, both without delay compensation, see figure 9. The decoders with delay compensation yield bigger errors than their counterparts. Inversely, they are given a slightly higher subjective rating of the localization accuracy (MOS). Worst of all, the in-phase decoder causes the poorest results for both, angle error and MOS. Regarding MOS, the basic decoder is a little bit better than the max r_E. But the differences between both are insignificant (64.2% for classification by signed angle error and 71.2% for classification by subjective rating).

4.3.2. *Position 2 (outside sweetspot)*



(a) Absolute angle error: colored boxplot (median separates yellow from red), percentages show outliers (error > highest quartil + 1.5·IQR).

(b) MOS (mean opinion score: average of the subjective rating) of the localization accuracy.

Figure 10: Absolute angle error and MOS for 5^{th} order decoders (d = with delay compensation) at position 2.

For this listening position, the same tendencies as for position 1 hold true, see figure 10. Whereas the subjective rating of the localization accuracy is not distinguishable (50.3% significance) for the max r_E and the basic decoders, the signed angle error of the max r_E decoder is definitely smaller (99.5% significance).

Therefore, the max r_E decoder is the best decision for the present listening setup.

5. Conclusion

For the test environment (see figure 1) with a constant number of 12 active loudspeakers, the present listening test meets the following expectations:

- The localization improves at higher orders.
- Localization at the central listening postition is more accurate than at the off-center position.

Against our expectations, the experiment indicates:

- Surprisingly, the compensation of the loudspeaker signal delays to the center causes confusion and worsens the results.
- The in-phase decoder is the worst candidate for every Ambisonics order of the test set at both listening positions.

The decoder with the best overall performance within this experiment is the max r_E decoder without delay compensation. Generally, the degradation at position 1 using delay compensation could be due to pronounced phase distortions outside the listening area, i.e. for radii $r > \mathrm{R}$. The radius $\mathrm{R} = \lambda\,\mathrm{M}/2\pi$ at the order $\mathrm{M} = 5$ is smaller than the head for frequencies above 2.2kHz allowing ± 4cm off-center shifts. $\mathrm{M} \geq 17$ would provide a sufficiently large area with a ± 4cm center. However, why the *uncompensated* delays perform well is subject to future studies.
Regarding other studies [3] that have been carried out under acoustically well-conditioned circumstances, the above discussed angle errors are comparable, desipite the non-ideal conditions. Consequently, a certain degree of robustness to real acoustic environments could be attributed to the reproduction principle.

6. References

[1] V. Pulkki, "Localization of Amplitude-Panned Virtual Sources II: Two- and Three-Dimensional Panning," in *Journal of the Audio Engineering Society, Vol. 49, No. 9*, 2001.

[2] T. Huber, "Zur Lokalisation akustischer Objekte bei Wellenfeldsynthese," Master's thesis, Technische Universität München, 2002.

[3] S. Bertet, J. Daniel, E. Parizet, and O. Warusfel, "Investigation on the restitution system influence over perceived Higher Order Ambisonics sound field: a subjective evaluation involving from first to fourth order systems," in *Acoustics-08, Paris*, 2008.

[4] S. Bertet, J. Daniel, L. Gros, E. Parizet, and O. Warusfel, "Investigation of the Perceived Spatial Resolution of Higher Order Ambisonics Sound Fields: A Subjective Evaluation Involving Virtual and Real 3D Microphones," in *30th AES Int. Conference*, 2007.

[5] G. Marentakis, N. Peters, and S. McAdams, "Auditory resolution in virtual environments: Effects of spatialization algorithm, off-center listener positioning and speaker configuration," in *Acoustics-08, Paris*, 2008.

[6] E. Bates, G. Kearney, F. Boland, and D. Furlong, "Monophonic Source Localization for a Distributed Audience in a Small Concert Hall," in *10th Int. Conference on Digital Audio Effects, Bordeau, France*, 2007.

[7] W. Hartmann, "Localization of sound in rooms," in *Journal of the Acoustical Society of America, vol. 78*, 1985.

[8] J. Daniel, "Représentation de Champs Acoustiques, Application à la Transmission et à la Reproduction de Scènes Sonores Complexes dans un Contexte Multimédia," Ph.D. dissertation, Université de Paris, 2000.

[9] J. Blauert, *Räumliches Hören*. Hirzel, 1974.