

# Simulation Study on the Effect of (Non-)Individual HRTFs and Ambisonics on Median Plane Localization

Matthias Frank<sup>1</sup>, Stefan Riedel<sup>1</sup>

<sup>1</sup> *Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Austria*

*Email: frank@iem.at, riedel@iem.at*

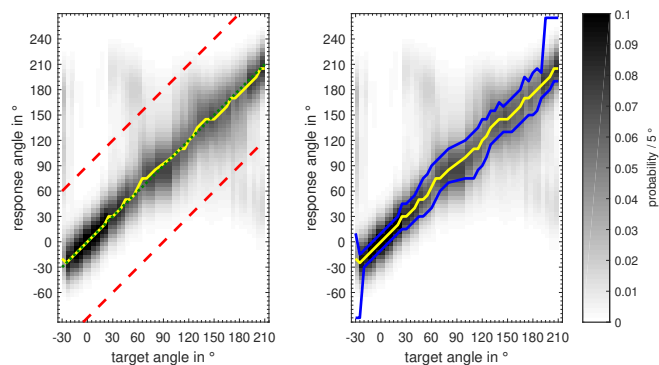
## Introduction

Ambisonics [1] is a scene-based audio format that provides a production workflow that is largely independent of the final playback system and thus enables playback on headphones [2, 3] and arbitrary loudspeaker systems [4]. However, in practice the headphone rendering sometimes suffers from an impaired elevation perception, i.e. from an under- or overestimation of source elevation. It is not obvious, whether such vertical localization errors stem from Ambisonics processing itself or from the typically employed non-individual head-related transfer functions (HRTFs), e.g. from other humans or a dummy head. On the one hand, Ambisonics decoding to loudspeakers causes vertical localization errors, especially for low orders [5], and on the other hand non-individual HRTFs degrade vertical localization performance in comparison to individual HRTFs [6, 7, 8].

This study aims to identify the effect of individual, non-individual, and dummy head HRTFs, as well as binaural Ambisonics decoder type and order on localization in the median plane, where no interaural cues are available. Localization performance is evaluated in terms of local error, localization uncertainty, and quadrant error rate using an open-source auditory model [9]. HRTF datasets include individual measurements from 23 human listeners and a KU100 dummy head. As studies showed general quality differences between different Ambisonics decoder types and orders [10, 11], decoding varies the order between 1 and 15 and uses the magnitude-least-squares approach (magLS) [2, 3], as well as a simple sampling approach with sparse and dense directional grids.

## Model and Measures

This study employs the *baumgartner2014* model [9] from the auditory modeling toolbox (AMT) [12] to predict vertical localization. It is a probabilistic, functional model of sagittal-plane localization using directional transfer functions (DTFs), i.e. HRTFs without the direction-independent characteristics, to emphasize high-frequency components. The model compares the spectral gradient of an incoming sound with the spectral gradient of template DTFs. The primary output of the model is a probability mass vector of polar response angles for each target angle, cf. Figure 1. In principle, the model accounts for listener-specific sensitivities and includes binaural weighting to extend prediction from the median plane to all other sagittal planes. However, this study employs the default settings without any specific sensitivities and is employed for median-plane localization only.



**Figure 1:** Exemplary model result indicating ideal localization (dotted green line, left), outer limit for responses that cause quadrant errors (red dashed line, left), localized direction (solid yellow line), and uncertainty region around that direction (blue solid line, right).

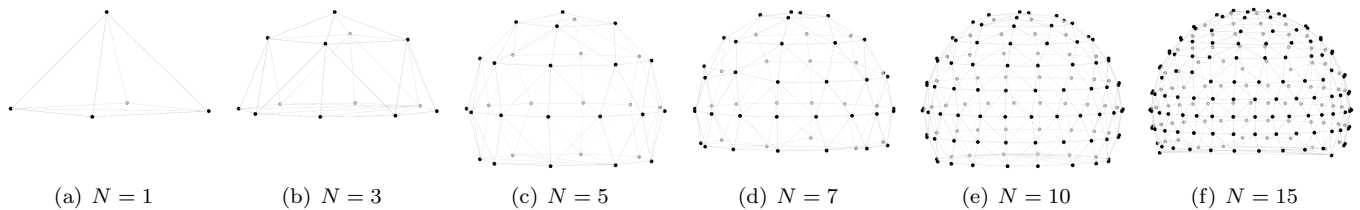
From these probability distributions, we calculate several performance measures. First, the *quadrant error rate* (in %) is calculated by summing up the probabilities of all responses that deviate by more than  $90^\circ$  from the target direction, cf. responses outside the dashed red lines in Figure 1. For the *local error* (in  $^\circ$ ), the target angle (green dotted line) is subtracted from the response angle with the maximum probability (yellow solid line), i.e. the 'localized direction'. The absolute error value is reported as an unsigned error metric and the signed error is reported to assess systematic directional bias. As an estimate for *localization uncertainty* (in  $^\circ$ ), probabilities for polar angles above the localized direction are summarized until their sum reaches 25%. The same is done below resulting in an angular region that includes 50% of the probability, cf. blue line in Figure 1. This measure is similar to an interquartile range. To keep the study compact, the measures are averaged within two polar regions: *front* with target angles in the range from  $0^\circ$  to  $45^\circ$  and *top* with target angles in the range from  $45^\circ$  to  $135^\circ$ .

## Conditions

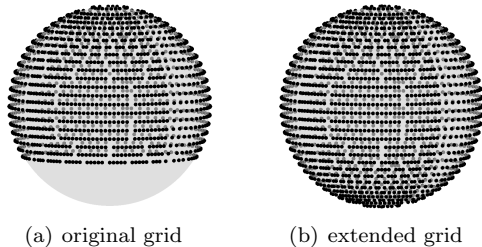
The simulation study considers: 1) a comparison of three different Ambisonics binaural decoders using individual HRTFs and 2) a comparison of individual, non-individual human, and non-individual dummy head HRTFs employing only the magLS decoding approach. The following paragraphs provide details about the HRTF datasets and the different Ambisonics decoders.

### (Non-)Individual HRTFs

Measured HRTFs from 23 human listeners that are available by default in the AMT were employed. The HRTFs



**Figure 2:** Grid of HRTF directions for sparse sampling in dependence of Ambisonics order  $N$ .



**Figure 3:** Grid of HRTF directions.

were measured from 1550 directions with a constant resolution of  $5^\circ$  in elevation between  $-30^\circ$  and  $80^\circ$  and a resolution of  $5^\circ$  or finer in azimuth, cf. Figure 3 (a). The KU100 dummy head was measured at the same lab (nh172) and with the same resolution. Note that the model input are actually DTFs, but for the sake of simplicity, we use the term HRTF when referring to different conditions of the study.

In all evaluations, the reference template is the individual HRTF without any Ambisonics processing. The conditions named *individual* use the individual HRTFs as stimulus, however processed by magLS Ambisonics of different orders. Thus, the magLS-processed HRTFs are compared to the original ones. The case *dummy head* evaluates how well each of the 23 human listeners can localize with the KU100 dummy head HRTFs, again with different orders of magLS as well as without Ambisonics processing. The dataset *non-individual human* simulates listening through other human ears, thus each of the 23 listeners listened to all other 22 human HRTFs. From these 22 HRTF sets, the *best non-individual human* condition selects the HRTF dataset with the smallest local error at a single frontal direction ( $0^\circ$  azimuth,  $0^\circ$  elevation) for each of the 23 listeners. This case is thought to represent a simple HRTF selection criterion that could be performed efficiently in practice.

### Binaural Ambisonics Decoders

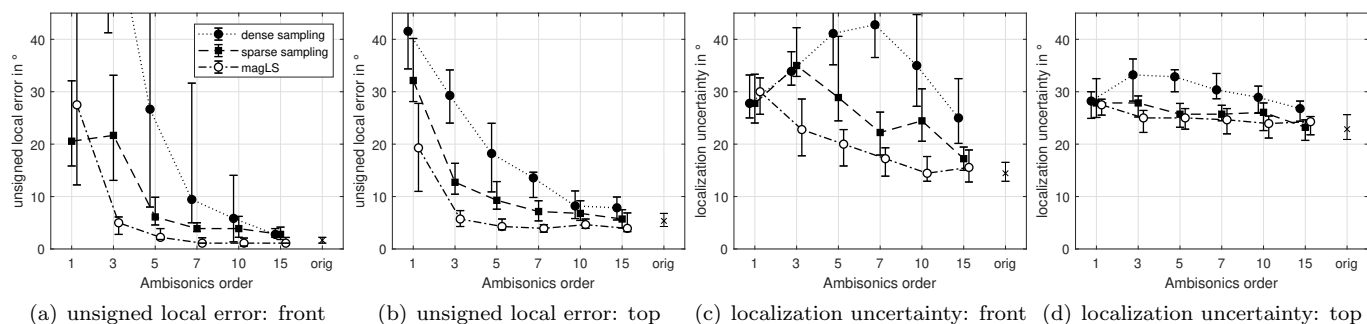
The simplest Ambisonics decoder uses the sampling approach, where no matrix inversion is required [1]. The decoder matrix comprises the spherical harmonics evaluated at all grid directions. To enable such a decoder for binaural playback, the decoded virtual loudspeaker signals are convolved with their respective head-related impulse responses (HRIRs, the time-domain equivalent to HRTFs). The *dense sampling* decoder used here employs all 1550 available HRTF directions independent of the Ambisonics order  $N$ .

As this might introduce coloration, especially at lower orders, a sparser, order-dependent subset might be useful [10]. A simple procedure to create an optimal set of directions was presented in [13] and is based on  $2N + 2$  evenly distributed directions at  $0^\circ$  elevation and a vertical separation of  $180^\circ/(N + 1)$  between the height layers, cf. Figure 2. For an order  $N = 1$ , the directional set consists of four directions at  $0^\circ$  elevation and one at  $90^\circ$ . For  $N = 3$ , these loudspeakers are shifted upwards, so that the layers are located at  $45^\circ$  and  $90^\circ$ . Finally, an additional layer with 8 directions at  $0^\circ$  elevation is added. For  $N = 5$ , the three layers are again shifted upwards to  $30^\circ$ ,  $60^\circ$ , and  $90^\circ$  and a  $0^\circ$ -layer is added with 12 directions. For this order, also a layer with negative elevation can be used based on the available negative elevation angles in the HRTF dataset. In general, the desired directions are rounded to the nearest available position in the HRTF dataset. Note that this procedure is defined for odd orders only, which are typically employed in practice. Nonetheless, for even orders the direction set for the next larger odd order can be used. This procedure of creating an Ambisonics decoder with a sparse set of directions is labeled as *sparse sampling*, here.

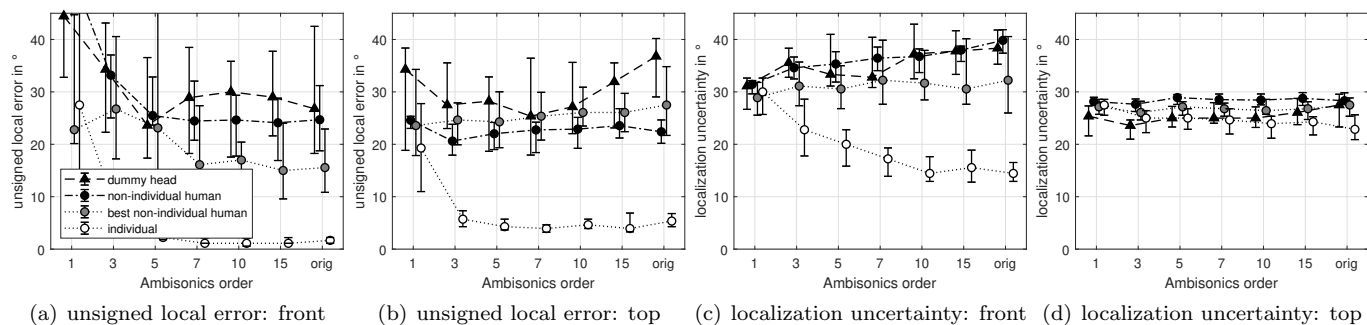
A more advanced binaural decoder is the magnitude-least-squares (*magLS*) [2, 3] approach that time-aligns the HRTF above a frequency  $N \cdot 650$  Hz that increases with the Ambisonics order  $N$ . It employs the pseudo inverse and optimizes the magnitude response of the frequency-dependent directivity pattern of the HRTF. To stabilize the inversion, the directional grid was extended with one additional direction at the zenith and nadir. The corresponding HRTFs were calculated by simply averaging the nearest available HRTFs. Moreover, nine circular layers between  $-35^\circ$  and  $-75^\circ$  elevation were added and their HRTFs were generated by linear interpolation between those at  $-30^\circ$  and the nadir, cf. Figure 3 (b). Note that the study only tests target angles equal or larger than  $0^\circ$  elevation to limit the effect of the artificially extended HRTF grids.

### Results

The results of the simulation study are presented as median values and interquartile ranges across the 23 'listeners' for the different HRTF/Ambisonics conditions, cf. Figures 4 to 6. In a previous step, the defined error measures were averaged across all additional parameters (e.g. target angles per polar region, non-individual human target HRTFs) resulting in 23 data points, i.e. one for each human 'listener'.



**Figure 4:** Median values and interquartile ranges of error metrics using individual HRTFs in dependence of Ambisonics decoder and order in comparison to original HRTF without any Ambisonics processing.



**Figure 5:** Median values and interquartile ranges of error metrics using different HRTFs and Ambisonics orders for magLS in comparison to original HRTF without any Ambisonics processing.

### Effect of Binaural Ambisonics Decoders

Figure 4 shows the results for individual HRTFs with different Ambisonic decoders and orders 1, 3, 5, 7, 10, and 15, as well as the original HRTF without Ambisonics processing. For the unsigned local error in front and top directions, Figures 4 (a) and (b) indicate a clear decrease towards higher orders. In order to estimate the minimum required order to achieve results that are not distinguishable from those of the original HRTFs, two measures are used: (i) p-value of a Wilcoxon signed-rank test  $> 0.05$  and (ii) a non-parametric Cohen's D effect size  $< 0.5$ .

For both the dense and the sparse sampling, even an order of 15 was found to be not enough, except for the top directions, where that order was sufficiently high when using the sparse sampling. In contrast, magLS required just orders of 7 (front) and 3 (top), respectively. It is thus not surprising that magLS achieves the smallest local errors for all orders  $N > 1$  ( $p < 0.001$ ). For order 1, where the error is large anyway, there was no significant difference between magLS and sparse sampling ( $p = 0.21$ ).

For localization uncertainty at frontal directions, cf. Figure 4 (c), there are similar trends: Both dense and sparse sampling never reach the values of the original HRTF, while magLS requires at least an order of 10. Again, for most orders, magLS is the best decoder. Similarity between decoders is large for the localization uncertainty at top directions, where magLS and sparse sampling are not significantly different for orders 1 and 5 ( $p \geq 0.16$ ). However, for orders 3, and 7 to 15, magLS is significantly better. While dense sampling never achieves the uncertainty (precision) of the original HRTFs, sparse sampling requires an order of 15 and magLS an order of 10.

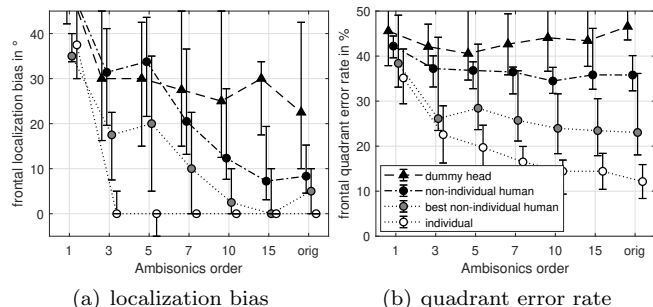
### Effect of (Non-)Individual HRTFs

Figure 5 shows the results for magLS decoding at orders 1 to 15 and without Ambisonics processing for dummy head, non-individual human, best non-individual human, and individual HRTFs. For the frontal localization, cf. Figure 5 (a), the individual HRTFs perform best ( $p < 0.001$ ) for orders  $N > 1$ . Until order 7, all other HRTFs perform comparable among each other. Starting at order 10, the best non-individual human HRTFs achieve significantly smaller errors than the dummy head ( $p \leq 0.03$ ). The non-individual human HRTFs lie between dummy head and best non-individual human. Within each HRTF condition, the minimum required order to indicate no effect of Ambisonics processing is 5 for dummy head and non-individual human, and 7 for best non-individual human.

Smallest localization errors for top directions were achieved by the individual HRTFs for orders  $N > 1$  ( $p < 0.001$ ), cf. Figure 5 (b). The other three HRTF conditions perform similarly and show no increased error due to Ambisonics processing, i.e. already an order of 1 is similar to the original HRTF. However, comparing the HRTFs without Ambisonics processing, there is significant ranking ( $p \leq 0.02$ ) from non-individual human over best non-individual human to dummy head.

The individual conditions achieve the smallest localization uncertainty at front directions for orders  $N > 1$ , cf. Figure 5 (c). The other three HRTF sets are again quite similar and there is no reduction of the uncertainty for higher orders. There is no clear trend when comparing the localization uncertainty at top directions, cf. Figure 5 (d). However, the individual

HRTFs perform best ( $p \leq 0.02$ ) for orders  $N \geq 10$ , and best non-individual human gets second place without Ambisonics processing. Again, an order of 1 is similar to the original HRTF for all conditions except the individual HRTF.



**Figure 6:** Median values and interquartile ranges of error metrics at single frontal direction ( $0^\circ/0^\circ$ ) using different HRTFs and Ambisonics orders for magLS and original HRTF.

To compare our model results to recent listening experiments [14], Figure 6 shows the localization bias and quadrant error rate for a single frontal direction ( $0^\circ$  azimuth,  $0^\circ$  elevation). There is no bias when using individual HRTFs for orders  $N \geq 3$ . All other HRTF conditions exhibit an upward shift. This shift strongly increases from (best) non-individual to dummy head HRTF. The same ranking is visible for the quadrant error rate. While for dummy head and non-individual HRTFs, there is no benefit of higher orders, there is one for the other two HRTF sets, especially for the individual.

## Discussion and Conclusion

Our study investigated the effect of (non-)individual HRTFs and Ambisonics on median plane localization using an auditory model. Localization error and uncertainty revealed a clear advantage of the magLS approach over the sampling decoder. Sparse sampling clearly outperformed dense sampling, as in [10]. Towards higher orders, the differences between the decoder strategies are reduced. For orders between 3 and 7, localization errors with magLS were similar to the original HRTF without Ambisonics, agreeing with [15]. Comparing individual to non-individual and dummy head HRTFs indicated a strong advantage of individual HRTFs. This could also be shown with regard to localization bias and quadrant error rates with similar results to experiments in [8, 14]. While acquisition of individual HRTFs might become more accessible based on numerical calculation from geometry scans [16], an interesting alternative could be an appropriate selection from a database of human HRTFs. Our proposed simple approach based on matching a single frontal direction showed some potential. Interestingly, only individual (or properly selected) human HRTFs benefit from Ambisonic orders  $\geq 3$ . While dummy head HRTFs and 3<sup>rd</sup> order might be enough to achieve plausibility in simple scenarios [17], individual HRTFs seem beneficial whenever accurate vertical localization is important, e.g. when virtualizing 3D loudspeaker studios [14].

## References

- [1] F. Zotter and M. Frank, *Ambisonics*, ser. Springer Topics in Signal Processing. Springer, 2019.
- [2] M. Zaunschirm et al., “Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint,” *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3616–3627, 2018.
- [3] C. Schörkhuber et al., “Binaural rendering of ambisonic signals via magnitude least squares,” in *Proc. of the DAGA*, vol. 44, 2018, pp. 339–342.
- [4] F. Zotter and M. Frank, “All-round ambisonic panning and decoding,” *J. Audio Eng. Soc.*, vol. 60, no. 10, pp. 801–820, 2012.
- [5] M. Frank, “How to make ambisonics sound good,” in *Forum Acusticum*, Krakow, 2014.
- [6] E. M. Wenzel et al., “Localization using nonindividual head-related transfer functions,” *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 111–123, 1993.
- [7] Z. Ben-Hur et al., “Localization of virtual sounds in dynamic listening using sparse hrtfs,” in *Proc. of the AES Int. Conf. on Audio for Virtual and Augmented Reality*, Online, 2020.
- [8] S. Riedel et al., “Localization of real and virtual sound sources in a real room: effect of auditory and visual cues,” *Submitted to J. Acoust. Soc. Am.*, 2024.
- [9] R. Baumgartner et al., “Modeling sound-source localization in sagittal planes for human listeners,” *J. Acoust. Soc. Am.*, vol. 136, no. 2, pp. 791–802, 2014.
- [10] B. Bernschütz et al., “Binaural reproduction of plane waves with reduced modal order,” *Acta Acustica*, vol. 100, no. 5, pp. 972–983, 2014.
- [11] G. Kearney and T. Doyle, “Height perception in ambisonic based binaural decoding,” in *Audio Engineering Society Convention 139*, Oct 2015.
- [12] P. Majdak et al. “AMT 1.x: A toolbox for reproducible research in auditory modeling,” *Acta Acustica*, vol. 6, p. 19, 2022.
- [13] M. Frank et al. “Equalizing the coloration of different ambisonic order weightings,” in *Fortschritte der Akustik DAGA*, Hannover, Germany, 2023.
- [14] S. Riedel and M. Frank, “Spatial perception of multi-source scenarios in real and virtual loudspeaker arrangements,” in *Proc. of the DAGA*, Hanover, Germany, 2024.
- [15] H. Lee et al. “Spatial and timbral fidelities of binaural ambisonics decoders for main microphone array recordings,” in *AES International Conference on Immersive and Interactive Audio*, Mar 2019.
- [16] H. Ziegelwanger et al. “Mesh2hrtf: Open-source software package for the numerical calculation of head-related transfer functions,” in *Proc. of 22nd int. congress on sound and vibration*, 2015.
- [17] K. Enge et al., “Listening experiment on the plausibility of acoustic modeling in virtual reality,” in *Proc. of the DAGA*, Online, 2020.