Audio Engineering Society

Convention Paper

Presented at the 144th Convention
2018 May 23 – 26, Milan, Italy

Evaluation of Binaural Renderers: Localization

Gregory Reardon[1], Andrea Genovese[1], Gabriel Zalles[1], Patrick Flanagan[2], and Agnieszka Roginska[1]

[1]*New York University, 35 W. 4th St., New York, NY 10012*
[2]*THX, 1255 Battery St, Suite 100, San Francisco, CA 94111*

Correspondence should be addressed to Patrick Flanagan (patrick@thx.com) or Agnieszka Roginska (roginska@nyu.edu)

ABSTRACT

Binaural renderers can be used to reproduce spatial audio over headphones. A number of different renderers have recently become commercially available for use in creating immersive audio content. High-quality spatial audio can be used to significantly enhance experiences in a number of different media applications, such as virtual, mixed and augmented reality, computer games, and music and movie. A large multi-phase experiment evaluating six commercial binaural renderers was performed. This paper presents the methodology, evaluation criteria, and main findings of the horizontal-plane source localization experiment carried out with these renderers. Significant differences between renderers' regional localization accuracy were found. Consistent with previous research, subjects tended to localize better in the front and back of the head than at the sides. Differences between renderer performance at the side regions heavily contributed to their overall regional localization accuracy.

1 INTRODUCTION

Binaural audio technologies, known in this work as *binaural renderers*, seek to recreate 3D audio scenes over headphones. Recent interest in augmented reality (AR) and virtual reality (VR) applications have highlighted the need for high-quality binaural audio. Not only is coherent spatial audio necessary for plausible virtual and mixed-reality environments, but it also plays a more significant role in orienting the user to their 360° environment, providing necessary audio cues and directing the user's attention.

Binaural renderers take a collection of audio waveforms with associated metadata describing the scene location, reverb characteristics, directivity, etc., of each waveform on the virtual soundstage. By leveraging the psychoacoustic features of human hearing, the representation is ultimately transformed into a binaural audio signal (sometimes passing through ulterior transformation into the spherical harmonic domain). These waveforms and their associated metadata are often referred to as *audio objects*. In traditional fixed media settings, such as surround sound reproduction, the location of sound sources is baked into the transmitted audio recordings. Audio objects are much more flexible and useful in interactive settings because the metadata describing how each audio object should be rendered at runtime can be updated in real-time. Critically, updating the location and orientation of audio objects by tracking the user's head means that sound sources appear as naturally occurring in the environment [1]. Within the context of *virtual auditory displays*, sound localization errors - *localization*, *externalization*, and *reversal errors* - have often been the subject of psychoacoustic investigation.

A methodology for a three-phase evaluation of binaural renderers was proposed and laid out by the authors in a previous publication [2]. The first phase of the test was concerned with evaluating sound localization errors of 3D audio reproduced over headphones. The phase was split into four tasks: *externalization, front/back confusions, up/down confusions*[1] and *localization*. The second phase of the experiment was concerned with evaluating spatial sound quality attributes, while the third phase, consisted of a force-choice ranking of renderers by order of preference. This third phase critically serves as a global assessment of renderer performance. The results of these latter two phases are to follow in future publications. The primary goal of the subjective experiment at large is to analyze the variance in performance between the tested binaural renderers. The secondary goal is to better understand how these different evaluation metrics impact user preference for a renderer. This can inform where specific improvements in the renderers should be focused and elucidate the most salient features for high-level appraisal of auditory scenes. It is beyond the scope of this study to identify the specific renderers that were tested.

This paper focuses on the assessment of renderer localization performance over static headphone reproduction using a source identification paradigm, and subsequent statistical analysis of significant differences. For the goals of this task within the overall test phase, and given the reported difficulty of correct localization on the median plane even in the best of conditions [4, 5, 6], only the horizontal plane was tested. A broadband audio object source was encoded for the most similar locations with each of the tested renderers. However, unequal polar grid resolutions in the encoding softwares meant that it was not guaranteed whether it was possible to render binaural locations for the exact same positions. To address this problem, a set of 12 localization regions were defined across the horizontal plane, also reflecting the non-linear azimuth-dependent limits of human auditory localization accuracy [4]. This decision also allowed a certain degree of tolerance for eventual localization blurs experienced by subjects, acceptable for the entertainment applications for which these renderers are primarily marketed.

1.1 Localization Errors

Whether spatial audio is reproduced over loudspeakers or headphones, humans experience a location-dependent phenomenon called *localization blur* [7]. In support of the notion of an auditory perceptual resolution, localization blur has been observed to be in the order of 5° to 20° [6, 7, 8], while the scatter of responses for azimuth perception was found to be more pronounced in peripheral locations compared to frontal locations [4, 7]. The *Minimum Audible Angle* (MAA) describes the minimum detectable angular difference of two successive sound sources and is often used to quantify perceptual resolution of source location. In fact, several studies found the MAA to be non-linear and location-dependent: about 1° for sounds placed directly in front, and 20° for sounds placed on the interaural axis [9]. Other experiments found the MAA to be 1° at 0° azimuth, non-linearly increasing up to over 40° at an azimuth angle of ±90° [10].

Previous studies have agreed that for binaural audio the average angle of error between a perceived and a target sound source does not substantially differ from that perceived over loudspeaker reproduction. The average difference is on the order of 1° to 2° [8] for an absolute mean ranging from 9.5° of error [11] to 15° of error [6]. However, the cited studies were performed on a very small set of subjects using individualized head-related transfer functions (HRTFs). Errors were found to be as high as 60° when artificial head recordings were used [12]. An interesting, different, approach to the task of source identification consists into limiting the set of possible responses into perceptual "zones." A previous experiment was reported to have observed a range of *neighbor confusion* of 29% to 33% for 12 zones spaced at 30° on the azimuth plane [13]. Other literature looked at perceptual macro-regions by dissecting the azimuth plane into *front, sides* and *back* to group statistical data into areas of interest. An average error angle range of 20° to 25° was found for all three regions [14].

One of the most common types of error in static binaural reproduction over headphones is the reversal error. These errors occur along auditory cones of confusion, where interaural cues are identical for sources placed at opposite sides of the cone [7]. Although an investigation on these types of confusions was specifically addressed by the authors in a related publication [3] which looked at cross-hemisphere trajectories, confusion rates over static binaural presentation are also reported in the present document. Front/back confusions are deemed to be caused by several factors, including lack of plausible source emitter in the visual field [1]

[1]Results of renderer performance on the externalization, front/back, and up/down confusions tests are presented in [3].

and non-ideal spectral content presented at the listener's ears [7], since front/back discrimination is largely dependent on the pinnae spectral distortions. Distinct from localization blur, confusion rates have been reported to increase in binaural reproduction when compared to loudspeaker reproduction. Exact rates depend on the experimental setup, stimuli choice, and HRTF processing. In [8], an average reversal rate of 11% was reported, about double those reported for the free-field conditions. For non-individualized HRTFs, the reversal error was found to range from 28% to 41% using different plastic heads [13]. Further, with artificial heads, maximum reversal rates were found to peak at 0° and 180° azimuth, sometimes reaching 58% [12].

Disregarding reversal errors, localization on the horizontal plane has been shown to be comparable between generalized HRTFs and individualized HRTFs [14]. In fact, interaural time differences and interaural level differences, the main cues for horizontal localization, are fairly consistent among individuals and frequency ranges. On the contrary, spectral cues determined by the pinnae present a considerable variance between subjects, are more relevant for the auditory perception among the vertical plane, and influence front/back confusions [1, 11]. Since commercial renderers do not unanimously provide for HRTF customization, but instead rely on non-optimal general HRTFs from either public or private databases, assessing localization at elevations other than zero would prove to be difficult and possibly an unfair measure of binaural rendering quality.

Another important aspect to keep in mind is that localization performance and confusion rates also depend on the frequency content of the presented stimuli. It has been demonstrated that a broadband stimuli can achieve a confusion rate of 2-10% while a single-octave noise band presents a confusion rate of >20% [4]. This is easily explained by the fact that different cues are effective at different frequency ranges [5], and broadband signals can exploit the full spectrum of binaural cues available.

1.2 Rendering Methodologies

Commercial binaural renderers implement object-based audio in either one of two main modalities: direct virtualization through HRTFs or spherical harmonic decomposition and encoding for First Order Ambisonics/Higher Order Ambisonics (FOA/HOA) virtual speaker configurations. Virtual microphone techniques for HOA use decomposition methods to transform sound into the spherical harmonic domain. Once encoded, the sound can be flexibly decoded for any given loudspeaker, or virtual loudspeaker, configuration. The advantage of the decomposition is a reduction in the number of virtual sources that need to be rendered in real-time, from any arbitrary number, to the number of sources necessary to accurately reproduce the spherical soundfield of the desired order. However, the reduced complexity comes at the cost of a spatial approximation which leads to worse localization accuracy and possible phasing issues between the upmixed B-format streams [15].

Localization accuracy was found to heavily depend on the order of the ambisonics encoding. The order of the spherical encoding process has a direct impact on the resolution of the spatial approximation, meaning that a 4^{th}-order decomposition can yield significantly better localization results than a 2^{nd} or 1^{st}-order one [16]. Other experiments have validated the correlation between ambisonics order and localization accuracy. One experiment found that spatial resolution in ambisonics localization presented a non-linear dispersion response correlated to the system order [17]. An ideal 4^{th}-order system would present a median perceptual error of 5° along the azimuth range of ±90° to ±135°. In contrast, a 1^{st}-order system presented a median error of 10° to 25° in the same range indicating a higher localization blur. Similar results were found in another study which compared ambisonics reproduction on real versus virtual sound sources, without finding significant differences between the two, but validating the correlation of order and resolution [18]. Both studies observed a significant number of front/back confusions and a drift tendency for lateral sources to be localized towards the rear.

2 METHODOLOGY

2.1 Rendering Procedure, Stimuli, and Presentation

Six different commercial renderers were tested comparatively. The renderers are labeled from 00 - 05. Three of the renderers (00, 01 and 05) use HOA for spatialization. Two of the renderers (03 and 04) use FOA. The final renderer (02) uses direct virtualization

Zone	Azimuth Range	Available Positions
0	350° - 10°	0°
1	10° - 30°	20°
2	30° - 60°	30°, 40°
3	60° - 120°	70°, 80°, 90°, 100°, 110°
4	120° - 150°	130°, 140°
5	150° - 170°	160°
6	170° - 190°	180°
7	190° - 210°	200°
8	210° - 240°	220°, 230°
9	240° - 300°	250°, 260°, 270°, 280°, 290°
10	300° - 330°	310°, 320°
11	330° - 350°	340°

Table 1: Details of zone range areas and available locations for trial presentations.

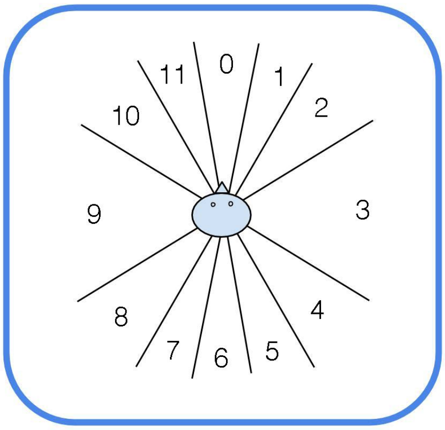

Fig. 1: Graphical depiction of regional divisions of the horizontal plane tested.

through HRTFs. Three dry two-second monophonic drum loops were created in Pro Tools and used for testing. These stimuli are labeled 0 - 2. They were output from Pro Tools at 48 kHz sampling rate and 24 bit-depth. The selection of the stimuli reflects the desire to use the results of this study in commercial applications. And, the stimuli are relatively broadband and therefore able to exploit the full range of auditory cues. The same stimuli were used to assess the regional localization accuracy task as well as the other phase-one tasks portrayed in [3]: externalization, front/back and up/down confusions.

Though each renderer supported headtracking in its native application, for the purposes of the experiment, content was presented in a static condition. Stimuli were placed in the audio scene at a distance of one meter from the audio listener at a chosen set of azimuth angles (see *Table 1*) and at an elevation of 0°. A large set of static binaural content generated with each renderer was used for the evaluation. Subjects were presented with a randomized subset of the full space of rendered stimuli. In order to make more uniform comparisons between renderers and evaluate the base rendering engine of each binaural renderer, all room information was turned off; the early reflection and late reverb modules for each renderer were disabled. All other export settings were set to the highest quality.

Audio was presented over circumaural stereophonic headphones (Sennheiser HD-650) in a soundproof booth (NYU Dolan Isolation Booth). Custom software was used to administer the test and collect data. A 2D graphical user interface (GUI) allowed subjects to play and replay stimuli, indicate and submit their responses, and make comments on each specific trial. Subjects were not given feedback on the outcome of the trial nor were the subjects trained.

2.2 Localization Test

The localization test was concerned with regional localization accuracy on the horizontal plane (0° elevation). Given the non-linearity of the MAA [10] and also as a tolerance measure against localization blur, unequally spaced zones were defined. These subdivisions, which also represent the set of available answer choices, are listed in *Table 1*. No stimuli were presented on the boundary of a zone. Subjects indicated responses directly on the graphical interface. This graphic is pictured in *Fig. 1*.

Each of the seventy-seven participating subjects evaluated 54 trials in the localization test. Each renderer and each stimulus were presented three times (6 renderers, 3 stimuli, and 3 observation times) from randomly picked zone locations. Given that multiple renderers were being tested, an exhaustive test of localization accuracy for each combination of available position, renderer, and stimulus was not feasible. In addition, given the large number of participating subjects, it was deemed unnecessary to present the same subset of azimuth locations.

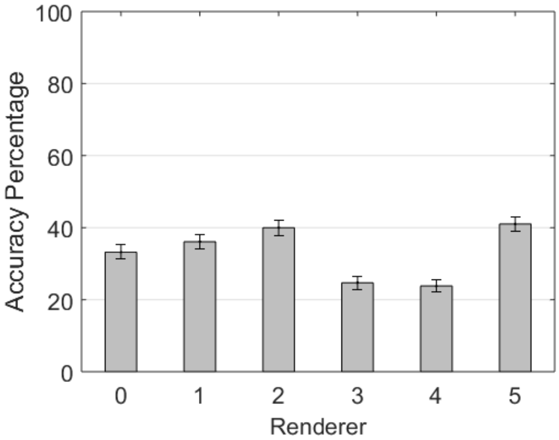

Fig. 2: Localization Accuracy - estimated marginal means and standard errors.

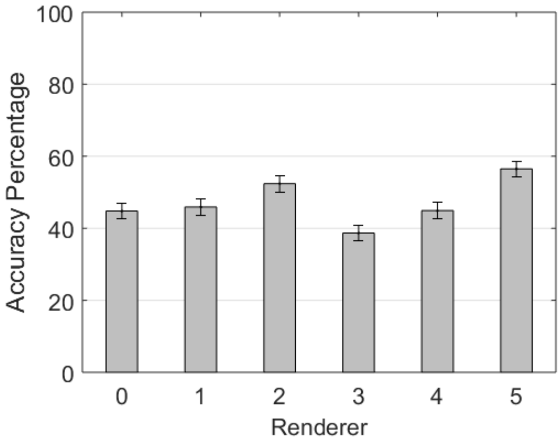

Fig. 3: Confusion-Corrected Localization Accuracy - estimated marginal means and standard errors.

For any given trial, the azimuth location was drawn from the larger pool at random and presented in isolation to the listener. The subject was asked to indicate the zone from which the sound appeared to emanate. Each zone was not sampled uniformly because the authors did not want to bias renderer performance towards a specific azimuth position by virtue of the unequal zone sizes. The renderer, stimulus, trial number, correct answer, and subject answer were recorded for each trial.

3 RESULTS

Seventy-seven subjects participated in the localization test. In a first analysis, responses were first treated as binary outcome (correct or incorrect), with no correction for front/back reversals or respect to distance of error. This scenario was modeled using a generalized linear mixed model (GLMM). The GLMM has an advantage over a repeated-measures analysis of variance (ANOVA) in that it can model binomial outcome variables. A logit link function was used in the specification, making this GLMM specification an extension of a logistic regression, amenable to repeated-measures data. A 6 x 3 x 3 (6 renderers, 3 stimuli, 3 observations) repeated-measures structure was used to analyze the data. The subject-specific effects were modeled as random effects, while renderer, stimulus, and a renderer*stimulus interaction term were treated as fixed effect factors. Because observations were made over a short period of time and the azimuths presented for each trial were randomized, observation time was not modeled as a fixed effect. The main effect for renderer was found to be significant ($F(5,4140) = 17.061$, *p<0.001**). No significant effects for stimulus or renderer*stimulus were found. The estimated marginal mean and standard error for each renderer is presented in *Fig. 2*.

The marginal mean estimates of regional localization accuracy presented in *Fig. 2* have no corrections for front-back and back-front reversals. In order to understand localization performance irrespective of reversal errors, the GLMM procedure was repeated, this time with front-back and back-front reversals marked as correct. That is, if the stimulus was appraised by the subject as appearing to emanate in the zone symmetric across the interaural axis from the true location, the answer is marked as correct. This amounts to folding the zones over the interaural axis and aggregating results. This gives a clearer understanding of localization performance agnostic to reversal errors and their type (front-back versus back-front). Under this scenario, the GLMM also indicated a significant main effect for renderer ($F(5,4140) = 10.800$, *p < 0.001**). No significant effects for stimulus or renderer*stimulus were found. The estimated marginal means and standard errors for confusion-corrected localization accuracy are presented in *Fig. 3*. The mean reversal rate for each renderer is also reported (*Table 2*). The figure was calculated as the ratio of reversal errors over correct and reversed

	Renderer					
	00	01	02	03	04	05
Reversal Rate	25.2 %	20.9 %	22.9 %	34.4 %	45.4 %	26.9 %

Table 2: Mean reversal rate for each renderer.

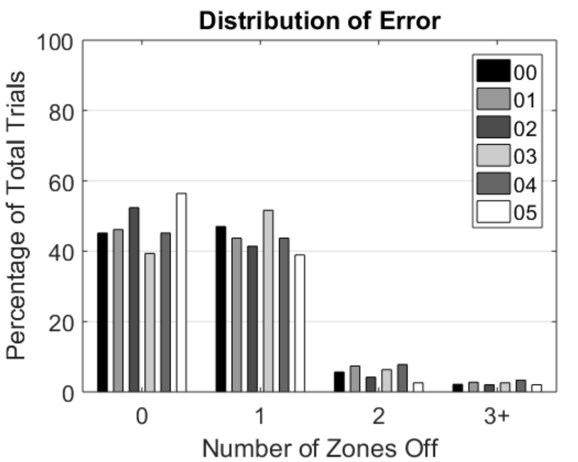

Fig. 4: Absolute error distribution for each renderer with confusions corrected.

judgments. Thus excluding all other errors that are not purely reversal errors from the denominator[2].

Since the study was also concerned with regional localization accuracy, the severity of errors was quantified by the localization zone distance between the subject's response and the correct zone. Confusions were again corrected for to provide a better understanding of performance irrespective of these errors. For example, if the stimulus was located in zone 1 (10° - 30°) but the subject answered zone 4 (120° - 150°), the subject's response would be corrected to zone 2 (30° - 60°) and the error marked as one zone off. The error distribution for each renderer was calculated globally (*Fig. 4*) and for each zone (*Fig. 5*). These global results of error distribution are presented as a percentage of the total number of trials for each renderer. The zone-wide results are calculated as percentages of the total number of trials for each renderer in that specific zone. Given the symmetric nature of the head across the median plane, zones opposite one another along said axis were aggregated. This gives a better sense of trends in the distribution of localization errors as one moves from the front to the sides and to the back of head. As discussed, each region was sampled non-uniformly, so smaller zones had fewer observations.

[2]More details about reversal errors can be found in [3].

4 DISCUSSION

Developing a comprehensive set of evaluation metrics and procedures requires weighing complexity and the length of testing procedures. With this in mind, the authors elected to evaluate regional localization accuracy. This allowed for a shortened procedure which still gathered a rich set of data on localization accuracy for each renderer. This is important for generalizing the methodology to experiments involving any number of binaural renderers.

Results demonstrate that commercially available renderers do perform differently with respect to regional localization accuracy. Without correcting for reversal errors, zone localization accuracy for all renderers is under 50%, with the poorest performance given by the FOA renderers. After correction, overall zone localization accuracy for some renderers improves up to almost 60%. Given the coarser zonal resolution of the task, localization accuracy is measured in terms of percentage rate rather than average angle of error. Although some literature groups azimuth performance within macro-regions of perception [14], those results incorporate data from different elevation angles and are more concerned with measuring source identification error angle. Thus, the cited literature does not provide a direct reference of comparison for the proposed specific general zone accuracy but it can give an indication about the expected angle of error at different zones.

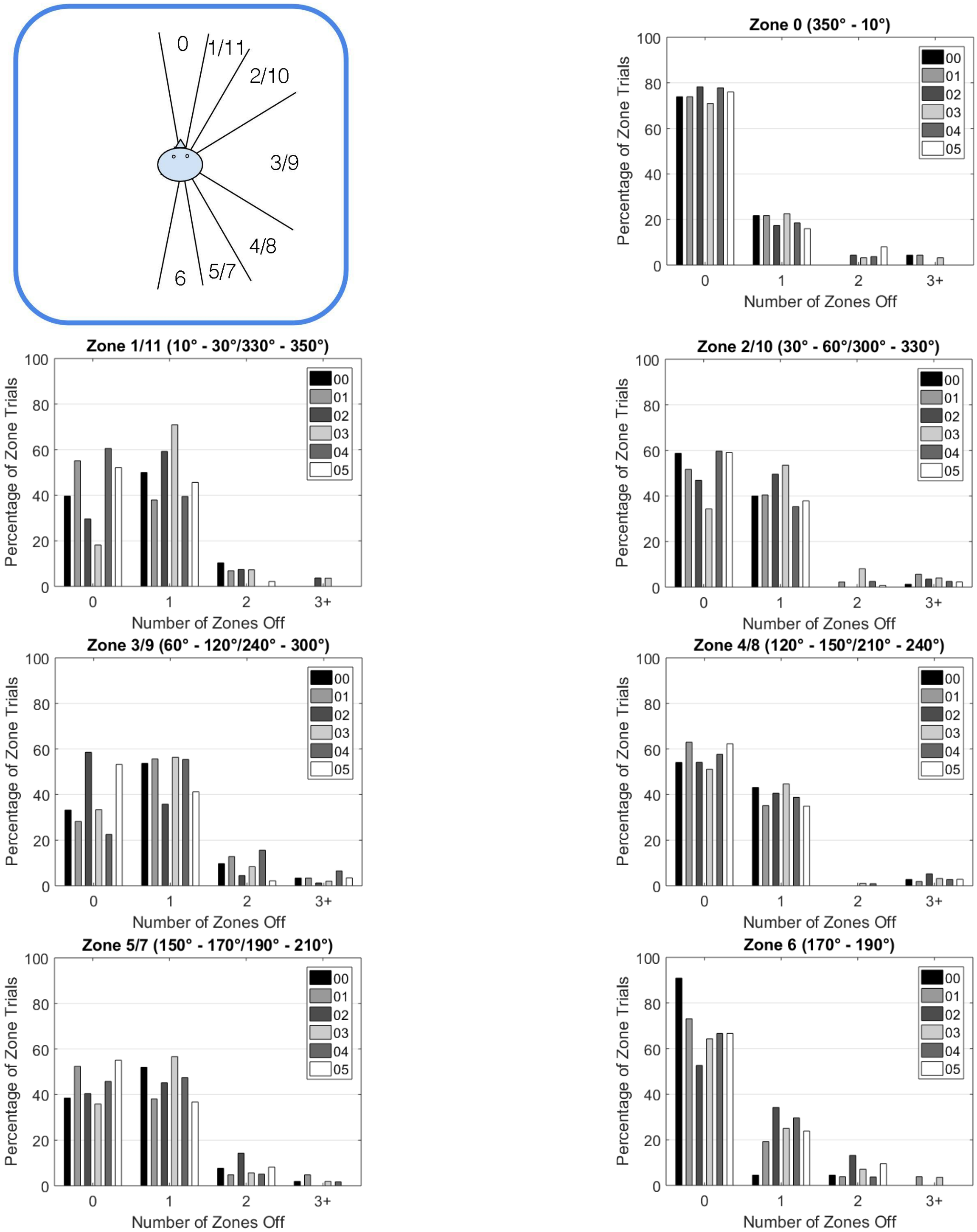

Fig. 5: Graphic of localization regions folded over the median plane (top left) and absolute error distribution by zone for each renderer with confusions corrected.

Reversal error rates are reported for direct comparisons with previous literature. Generally, the error rates agree with the range found in [4, 8, 12] for low-elevation headphone conditions. The results in *Table 2* also suggest that FOA renderers present a larger number of front/back confusions in relation to fixed location binaural presentations, although other factors such as HRTF choice or proprietary upmixing algorithms could interact with the rendering technique. This finding agrees with what was previously found by the authors in a front/back trajectory identification task [3].

Since the reversal rates of binaural renderers were more thoroughly investigated in a previous publication related to the general assessment methodology [3], the main concern of this work was the reversal-agnostic localization performance of renderers. The poorest renderer performance as seen in *Fig. 4* is renderer 03, a FOA renderer. The performance of renderer 02 (direct virtualization) and 05 (HOA) stand out. Not only did these renderers perform best, but almost all errors for these renderers were one zone away.

In agreement with the majority of similar studies, the distribution of error for each zone indicates that subjects tended to localize better at the front and back of the head than at the sides. Given the unequal spatial resolution tested in the experiment, this trend is interesting. In choosing larger zones at the sides of the head in the experimental method, one would expect accuracy performance to be relatively consistent across zones. Not only do we see poorer performance in zone 3/9, but also a significant deterioration in localization performance when looking at the data outside the zones directly in front of and directly behind the head. The error distribution graphs for zone 1/11 and 5/7, the zones adjacent to the full front position and its correspondent at the back of the head, suggest that subjects were more prone to pull their answers either towards the front and back of the head or further out to the sides when presented with stimuli in these zones. This is a possible indication of a lateralization trend, not uncommon for inexperienced subjects [1].

Overall, the performance data shows high variability between renderers. The performance of the renderers at the sides of the head heavily contributed to the differences in localization accuracy. Looking specifically at zone 3/9, the two highest performing renderers, renderers 02 and 05, show much better accuracy in that zone than the other renderers. In addition, within this zone, a sizable percentage of responses identified the source location as two zones away. The increase in the spread of the distribution as one moves from the front or back of the head towards the side of the head is consistent with literature about the non-linearity of the MAA and localization blur [1, 4, 10]. With this in mind, the distribution of error for zone 6 indicates a general lower accuracy in the back regions for all renderers. While zone 6 is most similar to zone 0 in terms of accuracy percentage (0 zones off), zone 6 has a significant number of two-zones-off errors which are nearly absent at zone 0. Greater number of errors in the rear regions have been consistently found [1, 4, 19]. A possible explanation for this could be that: since localization accuracy at the rear is heavily influenced by spectral cues, generalized HRTFs would make it difficult to precisely localize sounds behind the head [7].

5 CONCLUSION

A large multi-phase subjective study on the performance of commercially available binaural renderers was conducted with seventy-seven participants. This paper presented the results of the regional localization accuracy task within phase I. The main concern of the analysis was to investigate the horizontal localization performance of the tested renderers irrespective of front/back confusions. The GLMM procedure confirmed that even using a coarse-grained approach, renderers can still be discriminated between. In fact, significant effects on localization accuracy due to the renderers were found, with and without reversal errors corrected. On the other hand, no other factors - stimulus, and observation time, nor any interactions between all three dependent variables - were found to be significant. Generally, localization accuracy was found to be greater at the front and back of the head than at the sides. Significant localization blur at the sides of the head was found for most renderers, with the exceptions of renderers 02 (direct virtualization) and 05 (HOA), whose performance in those zones stand out from that of the other renderers. The FOA renderers (03 and 04) perform poorest even after reversals corrections. These two renderers also performed strongest on overall regional localization accuracy on the horizontal plane.

The results found and presented in this work are a small piece of a larger study on spatial audio perception of binaurally rendered content. Insights on the subjective appraisal of immersive audio content can be gained through comprehensive evaluation of the performance of commercially available binaural renderers.

6 ACKNOWLEDGEMENTS

The authors would like to thank THX Ltd for their support on this research. Special thanks to Dr. Johanna Devaney for her statistics guidance.

References

[1] Begault, D. R. and Trejo, L. J., *3-D sound for virtual reality and multimedia*, NASA, 2000.

[2] Reardon, G., Calle, J. S., Genovese, A., Zalles, G., Olko, M., Jerez, C., Flanagan, P., and Roginska, A., "Evaluation of Binaural Renderers: A Methodology," in *Audio Engineering Society Convention 143*, Audio Engineering Society, 2017.

[3] Reardon, G., Zalles, G., Genovese, A., Flanagan, P., and Roginska, A., "Evaluation of Binaural Renderers: Externalization, Front/Back and Up/Down Confusions," in *Audio Engineering Society Convention 144*, Audio Engineering Society, 2018.

[4] Middlebrooks, J. C. and Green, D. M., "Sound localization by human listeners," *Annual review of psychology*, 42(1), pp. 135–159, 1991.

[5] Wightman, F. L. and Kistler, D. J., "The dominant role of low-frequency interaural time differences in sound localization," *The Journal of the Acoustical Society of America*, 91(3), pp. 1648–1661, 1992.

[6] Pernaux, J.-M., Emerit, M., and Nicol, R., "Perceptual evaluation of binaural sound synthesis: the problem of reporting localization judgments," in *Audio Engineering Society Convention 114*, Audio Engineering Society, 2003.

[7] Blauert, J., *Spatial hearing: the psychophysics of human sound localization*, MIT press, 1997.

[8] Wightman, F. L. and Kistler, D. J., "Headphone simulation of free-field listening. II: Psychophysical validation," *The Journal of the Acoustical Society of America*, 85(2), pp. 868–878, 1989.

[9] Perrott, D. R. and Saberi, K., "Minimum audible angle thresholds for sources varying in both elevation and azimuth," *The Journal of the Acoustical Society of America*, 87(4), pp. 1728–1731, 1990.

[10] Mills, A. W., "On the minimum audible angle," *The Journal of the Acoustical Society of America*, 30(4), pp. 237–246, 1958.

[11] Martin, R. L., McAnally, K. I., and Senova, M. A., "Free-field equivalent localization of virtual audio," *Journal of the Audio Engineering Society*, 49(1/2), pp. 14–22, 2001.

[12] Møller, H., Hammershøi, D., Johnson, C. B., and Sørensen, M. F., "Evaluation of artificial heads in listening tests," *Journal of the Audio Engineering Society*, 47(3), pp. 83–100, 1999.

[13] Poulsen, T., "Hörvergleich unterschiedlicher kunstkopfsysteme," *Rundfunktechnisches Mitteilungen*, 22(4), pp. 211–214, 1978.

[14] Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L., "Localization using nonindividualized head-related transfer functions," *The Journal of the Acoustical Society of America*, 94(1), pp. 111–123, 1993.

[15] Daniel, J., Moreau, S., and Nicol, R., "Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging," in *Audio Engineering Society Convention 114*, Audio Engineering Society, 2003.

[16] Bertet, S., Daniel, J., Gros, L., Parizet, E., and Warusfel, O., "Investigation of the perceived spatial resolution of higher order Ambisonics sound fields: A subjective evaluation involving virtual and real 3D microphones," in *Audio Engineering Society Conference: 30th International Conference: Intelligent Audio Environments*, Audio Engineering Society, 2007.

[17] Pulkki, V. and Hirvonen, T., "Localization of virtual sources in multichannel audio reproduction," *IEEE Transactions on Speech and Audio Processing*, 13(1), pp. 105–119, 2005.

[18] Braun, S. and Frank, M., "Localization of 3D ambisonic recordings and ambisonic virtual sources," in *1st International Conference on Spatial Audio,(Detmold)*, 2011.

[19] Oldfield, S. R. and Parker, S. P., "Acuity of sound localisation: a topography of auditory space. I. Normal hearing conditions," *Perception*, 13(5), pp. 581–600, 1984.