Audio Engineering Society

Convention Paper 9989

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

Evaluation of Binaural Renderers: Externalization, Front/Back and Up/Down Confusions

Gregory Reardon[1], Gabriel Zalles[1], Andrea Genovese[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 dynamic spatial audio over headphones and deliver immersive audio content. Six commercially available binaural renderers with different rendering methodologies were evaluated in a multi-phase subjective study. This paper presents and discusses the testing methodology, evaluation criteria, and main findings of the externalization, front/back discrimination and up/down discrimination tasks which are part of the first phase. A statistical analysis over a large number of subjects revealed that the choice of renderer has a significant effect on all three dependent measures. Further, ratings of perceived externalization for the renderers were found to be content-specific, while renderer reversal rates were much more robust to different stimuli.

1 INTRODUCTION

Recent interest in immersive audio has motivated a proliferation of binaural renderers for creating spatial audio content. These technologies render complex audio scenes into a binaural stereo output for reproduction over headphones. That is, they take a collection of audio waveforms with associated metadata describing the location, reverb characteristics, directivity, etc., of the waveform in virtual space, known together as an *audio object*, and, by leveraging psychophysical features of human hearing, reproduce a 3D sound image over headphones. The location and orientation of audio objects with respect to the user's head can be continuously updated by tracking the user. Audio objects in the virtual scene can then be made to appear as naturally occurring in the user's environment [1]. Within this context there are three main types of sound localization errors that are studied in psychoacoustic literature: *localization*, *externalization*, and *reversal errors*. This publication will examine the effects of renderer choice on the latter two type of errors.

This work presents results from a portion of a larger subjective experiment on the performance of commercially available binaural renderers. A three-phase methodology was presented in a previous work by the authors [2] and a large subjective test was carried out using this methodology. The first phase of the test was concerned with the measurement of sound localization errors of 3D audio reproduced over headphones. Externalization, front/back and up/down confusions, and localization were assessed individually. The second phase of the subjective experiment was concerned with evaluations of perceived spatial sound quality attributes, such as *naturalness*, *spaciousness*, and *clarity*. The third phase consisted of an overall preference assessment

and yielded a forced-choice ranking of renderers, the results of these phases are to follow in future publications. It is beyond the scope of this study to identify the specific renderers that were tested. Rather, the study is concerned with designing a methodology that can be used for evaluating binaural renderers and determining the impact these metrics have in creating an immersive binaural experience. This publication presents the results of the externalization and front/back and up/down confusions tests for the first phase of the experiment. The results from the localization task can be found in [3].

The goals of the subjective experiment at large are twofold. First, the authors want to better understand the variance in performance and take a survey of commercially available binaural renderers. Second, the authors seek to understand how the different individual metrics of binaural renderer performance are correlated with user preference. This will provide an understanding of where to focus improvements in the rendering procedure and how to ameliorate renderer performance. In order to perform both of the above tasks, each test must first be analyzed individually. This modularized approach will provide an opportunity to evaluate the methodology presented in [2] and indicate any improvements that can be made to the proposed methodology for a comprehensive evaluation of binaural renderers.

1.1 Externalization

The externalization of auditory images, such that spatial audio images appear indistinguishable from the auditory images produced by real-world stimuli, has been a large topic of research in psychoacoustics [4]. Often, subjects do not experienced the intended externalization. *Inside-the-head-locatedness* is a phenomenon of reproduced audio in which the auditory image appears internalized [5]. In static settings, externalization has been found to be easier to achieve on side locations than in front or rear locations [6]. Externalization lacks a standardized metric; measures of externalization have been proposed in the form of binary paradigm [4, 7] and with a discrete scale of distance from the head center as *externalization index* [6, 8].

A number of factors have been identified as affecting perceived externalization of 3D binaural sound reproduction over headphones. The aspect most commonly found to affect externalization is the presence of reverberation in the head-related transfer functions (HRTFs), or binaural filters. HRTFs gathered under non-anechoic conditions are known as binaural room impulse responses (BRIRs). Plausible or realistic reverberation, whether captured in a BRIR or synthesized to complement the HRTFs, has been consistently shown to significantly improve ratings of externalization [9]. Even in minimal form, added reflections were found to improve the externalization rate from 40% to 79% [7]. The presence of reflections does indeed interact with an internal sense of realism, which in turn is an important part of believing that virtual sound sources are situated outside the head. A related factor which has recently come under attention is the *room divergence effect*. This is defined as a divergence in audiovisual congruence by virtue of differences between synthesized scene and listening environment. Divergence between synthesized scene and listening environment has been found to decrease perceived externalization and the effect is more pronounced for frontal and rear sound sources than for sources located at the sides of the head [8].

Another discussed factor is the use of filtering in the form of generalized versus personalized HRTFs. HRTFs are composites of interaural differences and pinna cues, which depend on the anthropometric features of humans. As such, HRTFs are unique to individuals. Gathering personalized HRTFs is, in its current state, a labor-intensive process. Generalized HRTFs, in contrast, are gathered using dummy head microphones whose anthropometric features provide an approximation of a given individual's HRTFs and are typically employed in most binaural renderers. Literature addressing the effect of personalized HRTFs on the externalization of auditory images has been conflicting, at times showing improved ratings of perceived externalization and improved consistency among locations [6, 10], while other times proving ineffective [7, 11]. Ecologically viable stimuli choices also help to provide an implicit reference of real sounds used to judge externalization. In theory, the more a binaural signal resembles a real life source in its acoustical details, the more likely it is to be externalized [4].

Head tracking has also has also been reported to influence ratings of perceived externalization [12, 13]. Though some authors have reported no significant effect of head tracking on externalization [7], the unnatural experience of static binaural content leads to degradation in the spatial image quality. Externalization

errors have been found to persist even given dynamic personalized binaural reproduction.

1.2 Reversal Errors

A second type of sound localization error particularly endemic in 3D audio reproduced over headphones is that of front/back and up/down reversals. These errors occur along auditory *cones of confusion*. Each auditory cone of confusion is the set of all points on a sphere surrounding an individual's head for which the two main human localization cues, interaural time differences (ITDs) and interaural level differences (ILDs), are the same. Within a cone, two types of errors can occur: *front/back reversals* and *up/down reversals*. In such errors, the individual perceives an auditory illusion occurring at a location symmetric to the actual audio event over the frontal plane or transverse plane. Reversal errors over the frontal plane, often referred to as the interaural axis, are known as front-back and back-front confusions while the reversal errors over the transverse plane are known as up-down and down-up confusions [1].

A large contributor to reversal errors is the severity of spectral differences between the the generalized HRTFs used for binaural reproduction and the actual HRTFs of the subject. Generalized pinnae cues have been shown to increase the prevalence of front/back confusions when compared to individualized cues [14]. Spectral details can affect up/down reversals too. While unnatural ITDs were not found to have a significant effect, spectral distortions from non-individualized HRTFs can lead to poor vertical localization in static settings, especially at the below-the-ear elevation levels where frequency notch migrations happen more rapidly as the elevation decreases [15]. These HRTF spectral notches are considered to be salient cues for elevation - and vertical trajectory - discrimination, an inherently difficult task even for real sources [16].

Front-back reversals tend to be much more common than back-front reversals. Mean confusion rates over headphones with generalized HRTFs have been found to be around $25\% \pm 15\%$ for front sources, and $6\% \pm 5\%$ for rear sources, depending on factors such as listener's proficiency, type of stimuli used and choice of HRTFs [14, 17, 18]. This could be explained as a part of a primitive survival heuristic that, in the absence of a visual stimuli in front of the listener, defaults to perceiving the location of the acoustic event as occurring behind the individual [1].

In regards to vertical localization, the few studies addressing the issue have found no particular statistical bias towards either up/down reversal direction, neither on the median plane [14] nor other sagittal planes [15]. Confusion rates are reported to generally be around 15-30% [14, 18]. In terms of motion, the reported vertical Minimum Audible Movement Angle (MAMA) ranges from 10° to 16°, although levels as high as 45° have been reported for below ear-level starting point [15, 19, 20]. Within the proposed methodology, this figure is useful for understanding the range of movement a source should take for a change of elevation to be perceptually noticed. No particular azimuth dependency has been reported. One study in particular suggests that vertical motion cues are equally valid across sagittal planes (or azimuth locations) [15]. However, in [14] about half of the up/down confusions reported were found to be combined confusions, involving both up/down and front/back reversals.

Other factors known to influence reversal errors in virtual sound reproduction include the signal's frequency content and the use of head-tracking technology. Broadband stimuli with energy in bands above the threshold of 7 kHz can exploit the function of the pinnae in creating high-frequency distortions that help to discriminate the general quadrant of incidence, thus decreasing the rate of front/back confusions [18], especially when paired with personalized HRTFs [17]. "Real world" stimuli, as opposed to noise bursts or clicks, generally show higher reversal rates. One publication reported mean reversal error rates as high as 59% for static reproduction using speech stimuli [7]. The other important factor, head-tracking, is well-known to dramatically reduce the occurrence of reversal errors. Even subtle head movements can help to disambiguate sound localization cues [21], especially on the horizontal plane. However, it is reasonable to assume that most current cases of binaural audio reproduction happen in static settings. Moreover, to collect personalized HRTFs from users is a big technical challenge not currently addressed by the commercial renderer solutions available for this study. For these reasons, the most viable test case for this methodology was that of static generalized reproduction.

2 METHODOLOGY

2.1 Rendering Procedure, Stimuli, and Presentation

Six different commercial renderers were tested comparatively. The renderers are labeled from 00 - 05. Renderers 00, 01, and 05 render binaural audio using higher-order Ambisonics (HOA). Renderers 03 and 04 use first-order Ambisonics (FOA). Renderer 02 uses direct virtualization through HRTFs. Three different stimuli were used to assess externalization, front-back and up-down confusions for these different binaural renderers. These stimuli are labeled 0 - 2. The stimuli were two-second mono drum loops of different styles created in Pro Tools. These stimuli were output at 48 kHz sampling rate and 24 bit depth. A decision was made against testing those stimuli traditionally found in psychoacoustic literature, such as noise or infrapitch sound. This selection reflects a desire to understand how the renderers perform in commercial settings. And, the stimuli have broadband spectral energy which is required to exploit the full range of auditory cues [1, 18].

For each renderer, the stimuli were processed in the audio scene at a chosen distance of one meter from the listener at various azimuths and elevation angles. Though each of renderers supported headtracking in their native application, for the purposes of the experiment, the content was head-locked. Thus a discrete set of over one thousand stimuli was rendered as static binaural audio to be used in the first phase of the experiment. Each subject was presented a subset of these stimuli. In an attempt to evaluate the base rendering engine of each binaural renderer, all room information was turned off. This included both room reverb and early reflections, for those renderers that supported such a property. This also permitted more uniform comparison between renderers. All other export settings were set to their highest quality.

A total of seventy-nine subjects participated in the test. All seventy-nine subjects participated in the externalization test while only sixty-nine subjects participated in the front/back and up/down confusions test. In this work, because each test was analyzed individually, no data was excluded from analysis. Stimuli were presented over circumaural stereophonic headphones (Sennheiser HD-650) in a soundproof booth (NYU Dolan Isolation Booth). Custom software was used to administer the experiment and collect data. Subjects

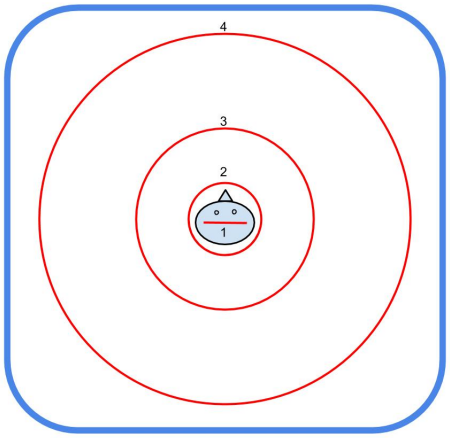

Fig. 1: Graphical representation of the discrete levels of externalization tested.

indicated their responses directly on a graphical user interface and had the option to replay the content for that trial before submitting a response and moving on to the next trial. For each trial, subjects were given a comment box to provide feedback on that specific trial. Defined here and for the rest of the paper, 0° azimuth refers to directly in front of the subject, with azimuth increasing in a clockwise direction.

2.2 Externalization Test

In the externalization portion of the test, each subject performed twenty-four trials - six training and eighteen testing. Over the eighteen test trials, each renderer and each stimulus was shown once. In this test all stimuli presented were located on the horizontal plane (0° elevation). The set of azimuths in this test were all locations at 10° azimuth increments (0°, 10°, ..., 340°, 350°). In each trial, four stimuli were presented. The first presentation was always a reference unprocessed stimuli. The following three were spatialized versions of that reference stimuli, drawn at random from the set of all possible azimuths. Subjects were asked to rate on a scale from one to four, where one was “inside the head,” and four was “far away from the head, externalized in space,” how far away the subset of spatialized stimuli appeared as a whole. A graphic accompanied this verbal description and is pictured in *Fig. 1*.

The goal of presenting multiple spatialized stimuli at different azimuths within each trial is to improve the

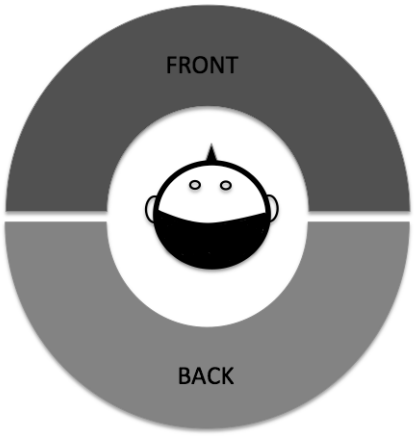

Fig. 2: Accompanying graphic used to assess front-back confusions.

consistency of externalization ratings. Given that ratings of externalization are dependent on the location of the stimuli with respect to the head [14], presenting individual stimulus would increase the variance of results.

2.3 Front/Back and Up/Down Confusion Tests

In the front/back test, subjects performed eighteen trials. Each renderer and each stimulus was shown once per subject. In each trial, a sagittal plane was first selected by choosing an available azimuth location. The associated azimuth pairs on the horizontal plane that were tested were as follows:

$$Pairs = [T, 180^\circ - T] \quad (1)$$

where $-60^\circ <= T <= -20^\circ$ and $20^\circ <= T <= 60^\circ$ and T is a multiple of 10. After an azimuth pair was selected, the order of presentation of the pair was randomized. The pair, or trajectory, was played three times and then subjects were instructed to identify the location of the second sound in the pair as either in front of or behind their head. Subjects indicated their responses by selecting a region on the graphic pictured in *Fig. 2*.

The up/down test was similar. Subjects performed eighteen trials, one for each renderer and each stimulus. In this case, reversals over the transverse plane are of interest. The motion trajectory ranged between +30° and -30° elevation angles, for a total motion of 60°, well above the MAMA indicated in [15]. In each trial, an azimuth was selected at random from the azimuths displayed in *Table 1*. The pair of spatialized stimuli located at that azimuth with elevation +30° and -30° were grabbed from the repository of stimuli, the order randomized, and the trajectory played three times. Subjects were asked to indicate the location of the second sound as either above the head or below the head. A similar, but appropriately altered, graphic as that found in *Fig. 2* accompanied this description.

Region	Azimuth				
Front-Right	20°	30°	40°	50°	60°
Back-Right	120°	130°	140°	150°	160°
Back-Left	200°	210°	220°	230°	240°
Front-Left	300°	310°	320°	330°	340°

Table 1: Azimuths tested in the up/down confusions test broken down into regions.

3 RESULTS

Different statistical models were used to analyze the data, depending on the nature of the dependent measure. Given the binary outcomes of the front/back and up/down confusions tests a generalized linear mixed model (GLMM) constructed as a repeated-measures logistic regression was used in lieu of a repeated-measures analysis of variance (ANOVA). A significance level of 0.05 ($\alpha < 0.05$) was used for all statistical tests.

3.1 Externalization

Seventy-nine subjects participated in the externalization test. Ratings of externalization were performed on a discrete 1-4 scale, with 1 being "inside the head" and 4 being "far away from the head, externalized in space." Six training trials were excluded from the analysis. A two-way 6 x 3 (6 renderers and 3 stimuli) univariate repeated-measures ANOVA was performed to analyze the data. The ANOVA indicated that "renderer" ($F(10,780)=32.170$, $p<0.001$*, Partial $ETA^2=0.292$), "stimulus" ($F(2,156)=0.793$, $p=0.001$*, Partial $ETA^2=0.093$), and the interaction term "renderer*stimulus" ($F(10,780)=3.706$, $p<0.001$*, Partial $ETA^2=0.045$) all had significant effects on ratings of externalization. Given that all three factors were significant, the estimated means for each are presented in *Figs. 3-5*.

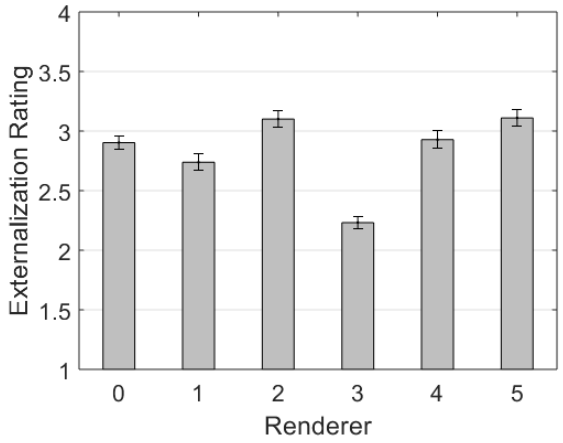

Fig. 3: Externalization - estimated marginal means and standard error bars for the renderers.

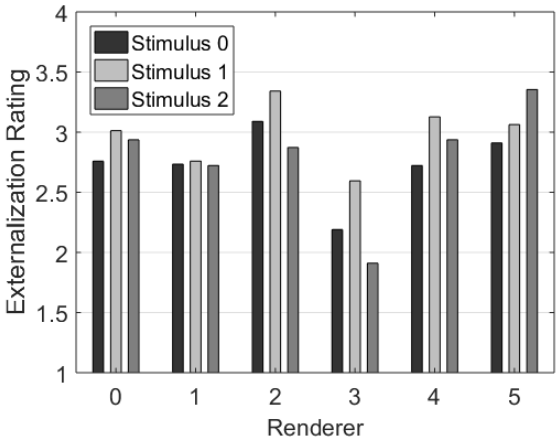

Fig. 5: Externalization - estimated means for each renderer and stimulus.

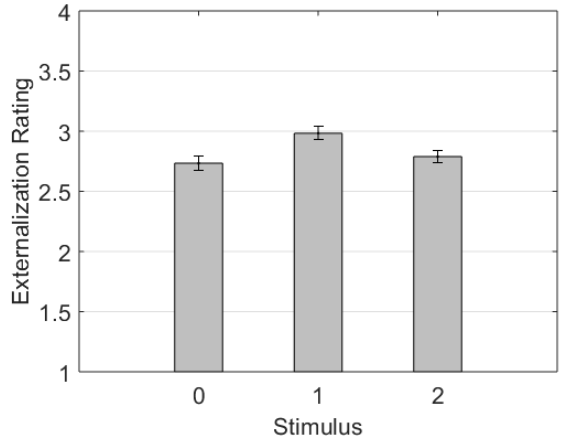

Fig. 4: Externalization - estimated marginal means and standard error bars for the stimuli.

3.2 Front/Back Confusions

Sixty-nine subjects participated in the front-back confusions test. Front-back confusions were assessed by having subjects indicate whether the second sound in the trajectory pair of stimuli was located as either "behind the head" or "in front of the head." Due to the binary nature of the outcomes, an ANOVA could not be used to analyze the data. Instead, a GLMM was used to establish significant factors. A 6 x 3 (6 renderers and 3 stimuli) repeated-measures structure with a logit link function was used in the model specification. Subject-specific effects were treated as a random effect. Three factors - "renderer," "stimulus," "renderer*stimulus" - were treated as fixed effects. An initial GLMM was run using a subset of the data (thirty-nine subjects). The analysis indicated that "renderer" ($F(5,684)=15.842$, $p<0.001$*) had a significant effect on front/back confusions. "Stimulus" and "renderer*stimulus" were not significant. A follow-up GLMM was then run with the whole sixty-nine person dataset. "Renderer" was still significant ($F(5,1224)=24.995$, $p<0.001$*). The later model was used to establish the estimated marginal means and standard errors presented in *Fig. 6*.

Fig. 7 presents a descriptive breakdown of the type of reversal errors for each renderer. The graph is categorized by frontal stimuli reversed to the rear (front-back reversals) and rear stimuli reversed to the front (back-front reversals). The mean front-back and back-front reversal rates are calculated out of the total number of trials possible to reverse in their respective directions.

The mean front-back reversal rate, agnostic to the direction of the reversal error, for each renderer as a function of azimuth is plotted in (*Fig. 8*). The grand mean front-back reversal rate as a function of azimuth is plotted separately in *Fig. 9*. In order to perform these calculations, the data was first folded symmetrically across the median plane, resulting in five aggregated position pairs, labeled in the graphs as 20°/160°, 30°/150°, 40°/140°, 50°/130°, and 60°/120°. Given that there

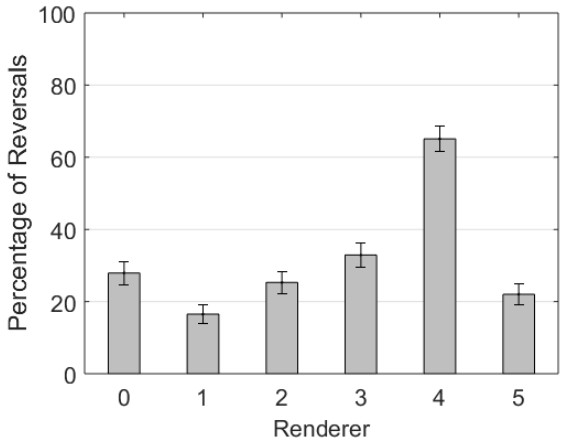

Fig. 6: Front/Back Confusions - estimated marginal mean reversal rates and standard errors for the renderers (includes both types of reversal errors).

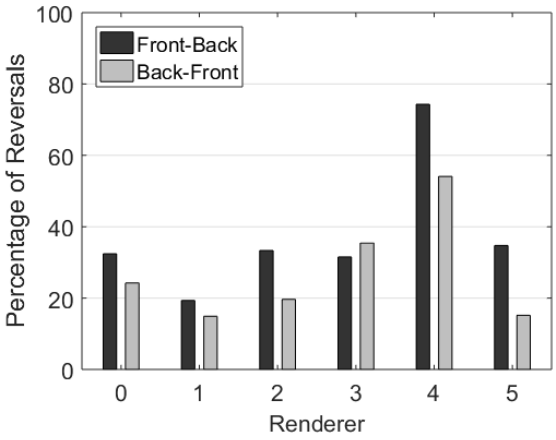

Fig. 7: Front/Back Confusions - descriptive mean reversal rate for each renderer by error type.

were unequal observations for each renderer at a given azimuth, the grand mean was then calculated by aggregating all responses at an azimuth pair.

3.3 Up/Down Confusions

Sixty-nine subjects participated in the up-down confusions test. Similar to the front-back test, subjects assessed whether the location of the second sound in a pair of stimuli was "above the head" or "below the head." A GLMM was also used to analyze this data. The model was identical to that used to analyze the front-back task: 6 x 3 repeated-measures structure, logit link function, one random effect (subject-specific effects), and three fixed effects ("renderer," "stimulus," and "renderer*stimulus"). The analysis indicated that "renderer" ($F(5,1224)=18.051$, $p<0.001$*) had a significant effect on up/down confusions. "Stimulus" and "renderer*stimulus" were not significant. The estimated marginal mean and standard error for each renderer is displayed in *Fig. 10*.

Fig. 11 breaks down the performance of the renderers in the up/down test by type of reversal error. The graph displays the descriptive mean up-down reversal rate and down-up reversal rate for each renderer as a percentage of the total number of trials possible to reverse in their respective directions.

In order to gain an understanding of the azimuth-dependency of the reversal ratings, the mean reversal rate for each renderer, along with a grand mean, as a function of azimuth is displayed in *Table 2*. Given the symmetric nature of the head, the data was folded over the median plane. Following from *Table 1*, there were ten aggregated azimuth positions. Further, nine subjects whose accuracy on all trials was below 50% were removed from this calculation.

4 DISCUSSION

It is possible to evaluate these results in isolation from the rest of the subjective experiment in a few ways. The most important result is that the performance of commercial renderers in terms of externalization, front/back and up/down confusions is, statistically speaking, different. For each of the individual tests, the main effect for renderer was significant. In each test, different renderers excelled, indicating that specific rendering techniques result in trade-offs in performance. Renderers 01 and 02 generally perform strongly on all metrics tested in this work. These renderers are a HOA renderer and a direct virtualization renderer, respectively. There does not appear to be defined clustering of results when grouped by spatialization method. But the FOA renderers, renderer 03 and 04, do, generally, perform poorly when compared to the other renderers. Specifically, the poor performance of renderer 03 in the externalization test (*Fig. 3*) and of renderer 04 in the front-back test (*Fig. 6*) stand out.

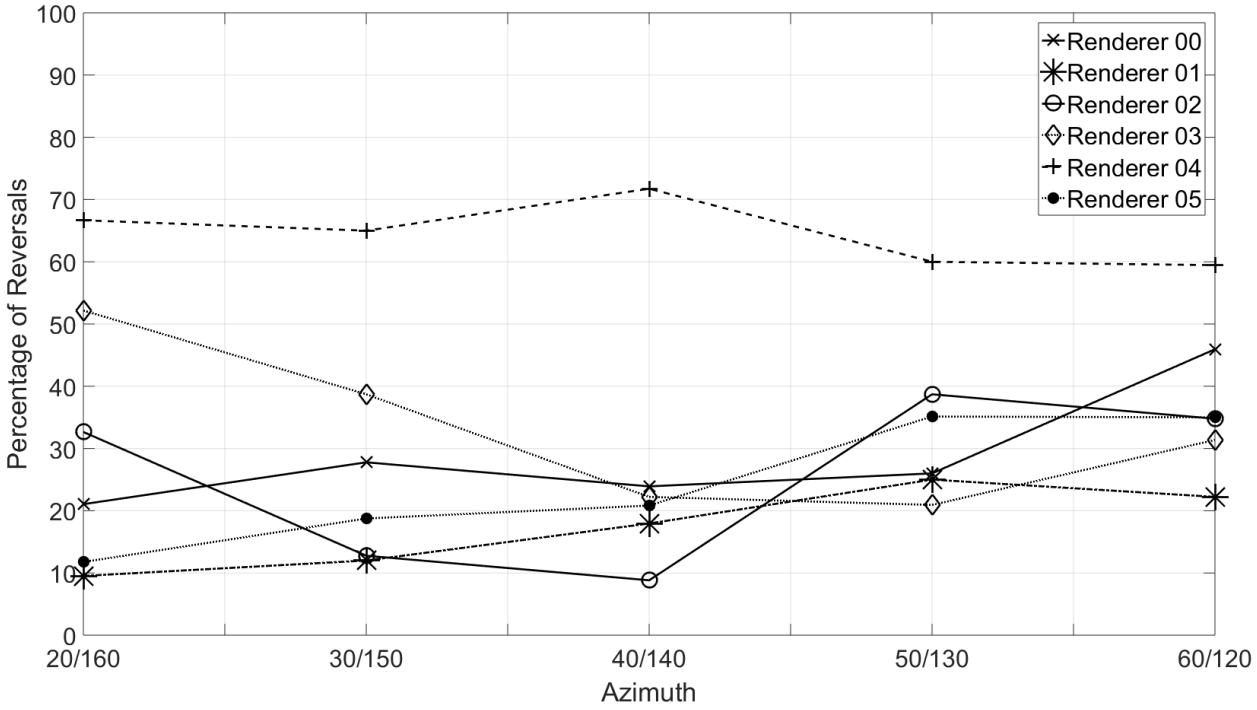

Fig. 8: Front/Back Confusions - mean reversal rate for each renderer as a function of azimuth (folded over the median plane).

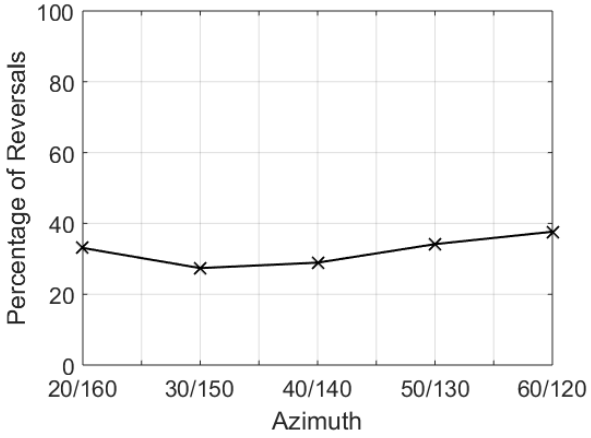

Fig. 9: Front/Back Confusions - grand mean reversal rate as a function of azimuth (folded over the median plane).

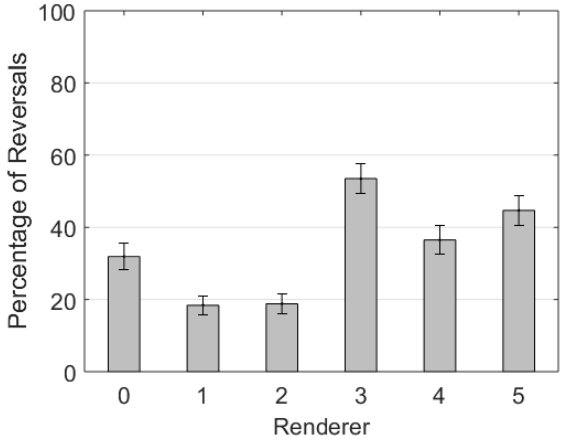

Fig. 10: Up/Down Confusions - estimated marginal mean reversal rates and standard errors for the renderers (includes both types of reversal errors).

	Azimuth									
	20°	30°	40°	50°	60°	120°	130°	140°	150°	160°
Grand Mean	39.09	43.40	32.32	30.84	22.52	19.44	24.53	21.93	30.84	38.39
Renderer 00	31.58	64.29	62.50	56.25	21.05	8.33	5.00	18.18	36.35	26.32
Renderer 01	17.65	16.67	27.27	20.00	5.56	9.09	5.26	0.00	6.25	11.76
Renderer 02	7.69	18.75	17.65	5.26	15.79	10.53	27.27	9.52	13.04	22.73
Renderer 03	54.55	60.87	16.67	33.33	38.10	61.54	54.55	40.00	76.92	52.63
Renderer 04	21.05	42.11	35.29	36.36	25.00	41.18	28.57	35.00	38.89	55.56
Renderer 05	85.00	56.25	33.33	40.00	27.78	0.00	25.00	20.00	30.77	64.71

Table 2: Up/Down Confusions - mean reversal rates in percent for each renderer as a function of azimuth (folded over median plane).

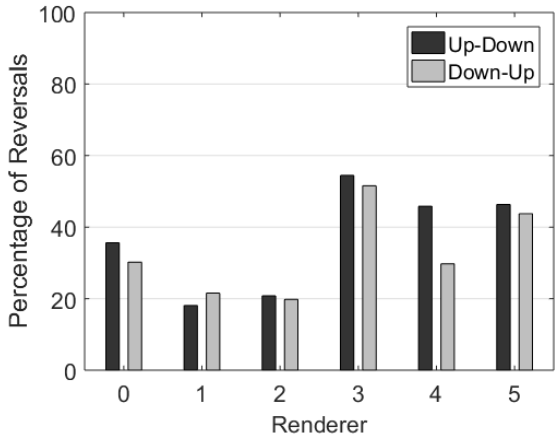

Fig. 11: Up/Down Confusions - descriptive mean reversal rate for each renderer by error type.

The results of the externalization test differ from those of the front/back and up/down confusions tests. "Stimulus" and "renderer*stimulus" both had a significant effect on ratings of perceived externalization. This is consistent with comments received during testing. Subjects noted that the unprocessed stimulus 1 appeared more externalized than the the other two unprocessed stimuli, which is mirrored in the results (*Fig. 4*). The content-dependence of perceived externalization is well-documented [7] and it makes generalizing any findings about externalization difficult, especially in this study where a small number of stimuli were tested. Further, the significant interaction term and *Fig. 5* suggest that the spectral distortions introduced by the binaural rendering process interact with the frequency content of the input signal, making evaluating the perceived externalization of a binaural renderer an even more difficult task and necessarily limited in scope to the specific content tested. On the other hand, this also suggests that the perceived externalization of any given renderer is content-dependent, and their use or purpose may affect overall performance.

Although no other studies in literature provide a direct comparison for the externalization metric used in this methodology, each single renderer - with maybe the exception of the FOA renderer 03 - resulted in a mean score between 2.5 and 3.5, which can be regarded as satisfactorily externalized. For a static setting without added reflections and using generalized HRTFs these are good average results. However, given the lack of studies for direct comparison it can't be guaranteed that the renderers produced well-externalized auditory images.

As seen in *Fig. 7* there does appear to be consistent bias for front-back reversals, as opposed to back-front reversals, for most of the renderers. This has been reported by other authors. Mean reversal rates mostly fall in the range of previous studies [1, 14]. Presenting either front-back or back-front trajectories with repetition was meant to improve the accuracy of localization judgements. Even using this methodology, FOA renderer 04 had a mean reversal rate $> 60\%$, slightly higher than the value reported in [1] for similar testing conditions. This suggests that subjects might be learning the incorrect association of the imposed binaural cues for that specific renderer. Back-front reversal rates are much higher than previously reported (perhaps barring renderers 01 and 02), although the previous figures were found over sixteen subjects only [14]. The grand mean reversal rate broken up by azimuth location (*Fig. 9*) indicates that the reversals were quite consistent over

the azimuths tested. When broken up for each renderer though (*Fig. 8*), it appears that the spread of mean reversal rates reaches a minimum at 40° azimuth.

Consistent with the findings in [14], the up/down confusions test did not reveal any particular bias for type of reversal error, either up-down or down-up (*Fig. 11*). With respect to mean reversal rates, only two renderers had $< 20\%$ reversal rates, while the rest presented higher reversals than previously reported [14, 18]. No obvious pattern can be discerned from azimuthal change in up/down confusion rates when looking at the individual renderers. This fact seems to agree with what has been found in other studies about elevation-dependent spectral distances being thought to be independent of the sagittal plane [15]. On the other hand, the grand mean as a function of azimuth (*Table 2*) appears to indicate a trend towards reduced up/down confusions as one moves from the front/back of the head (20°/160°) towards to the sides of the head (60°/120°). This was not found in previous literature. However, given the variability in the performance of each renderer, it is difficult to make a definitive statement about this trend. It is also possible that the up/down discrimination task was affected by front/back confusions, resulting in a combined reversal error, possibly diagonal, that could explain why the transversal confusion trend is higher at front and rear angles. Because the initial goal of the work was not to test azimuthal dependency of confusions, the experimental design was not set up to control for all factors (ie: equivalent representation of each azimuth for each renderer and each stimulus) in order to make such a conclusion.

5 CONCLUSIONS AND FUTURE WORK

A large multi-phase subjective study on the performance of commercially available binaural renderers presented under a static condition was conducted. The results from the externalization, front-back and up-down confusions tests of the study were presented. The renderers were found to have a significant effects on all three dependent. Renderer 01 and 02 performed well on all three tests, while the performance of renderers 03 and 04 (the FOA renderers) was generally poor. No renderer performed best in all three tests indicating that there are tradeoffs in baseline performance due to rendering methodology. The results of the externalization test indicated that the choice of stimulus and the interaction between the rendering procedure and the frequency content of the stimulus had a significant effect on perceived externalization. Measurements of front/back and up/down reversal rates were more robust to changes in stimuli. The confusions test revealed bias towards front-back confusions when compared to back-front confusions, but no clear bias for up-down versus down-up confusions.

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," 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., Genovese, A., Zalles, G., Flanagan, P., and Roginska, A., "Evaluation of Binaural Renderers: Localization," in *Audio Engineering Society Convention 144*, Audio Engineering Society, 2018.

[4] Hartmann, W. M. and Wittenberg, A., "On the externalization of sound images," *The Journal of the Acoustical Society of America*, 99(6), pp. 3678–3688, 1996.

[5] Mills, A. W., "Lateralization of High-Frequency Tones," *The Journal of the Acoustical Society of America*, 32(1), pp. 132–134, 1960.

[6] Kim, S.-M. and Choi, W., "On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach," *The Journal of the Acoustical Society of America*, 117(6), pp. 3657–3665, 2005.

[7] Begault, D. R., Wenzel, E. M., and Anderson, M. R., "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *Journal of the Audio Engineering Society*, 49(10), pp. 904–916, 2001.

[8] Werner, S. and Klein, F., "Influence of Context Dependent Quality Parameters on the Perception of Externalization and Direction of an Auditory Event," in *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*, Audio Engineering Society, 2014.

[9] Durlach, N. I., Rigopulos, A., Pang, X., Woods, W., Kulkarni, A., Colburn, H., and Wenzel, E., "On the externalization of auditory images," *Presence: Teleoperators & Virtual Environments*, 1(2), pp. 251–257, 1992.

[10] Werner, S., Klein, F., Mayenfels, T., and Brandenburg, K., "A summary on acoustic room divergence and its effect on externalization of auditory events," in *Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference on*, pp. 1–6, IEEE, 2016.

[11] Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D., "Binaural technique: Do we need individual recordings?" *Journal of the Audio Engineering Society*, 44(6), pp. 451–469, 1996.

[12] Brimijoin, W. O., Boyd, A. W., and Akeroyd, M. A., "The contribution of head movement to the externalization and internalization of sounds," *PloS one*, 8(12), p. e83068, 2013.

[13] Werner, S., Götz, G., and Klein, F., "Influence of Head Tracking on the Externalization of Auditory Events at Divergence between Synthesized and Listening Room Using a Binaural Headphone System," in *Audio Engineering Society Convention 142*, Audio Engineering Society, 2017.

[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] Benson, D. H., Martens, W. L., and Scavone, G. P., "Thresholds for discriminating upward from downward trajectories for smooth virtual source motion within a sagittal plane," in *Audio Engineering Society Convention 123*, Audio Engineering Society, 2007.

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

[17] Begault, D. R. and Wenzel, E. M., "Headphone localization of speech," *Human Factors*, 35(2), pp. 361–376, 1993.

[18] Bronkhorst, A. W., "Localization of real and virtual sound sources," *The Journal of the Acoustical Society of America*, 98(5), pp. 2542–2553, 1995.

[19] 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.

[20] Grantham, D. W., Hornsby, B. W., and Erpenbeck, E. A., "Auditory spatial resolution in horizontal, vertical, and diagonal planes," *The Journal of the Acoustical Society of America*, 114(2), pp. 1009–1022, 2003.

[21] Thurlow, W. R. and Runge, P. S., "Effect of induced head movements on localization of direction of sounds," *The Journal of the Acoustical Society of America*, 42(2), pp. 480–488, 1967.