Evaluation of Binaural Reproduction Systems from Behavioral Patterns in a Six-Degrees-of-Freedom Wayfinding Task

Olli Rummukainen, Sebastian Schlecht, Axel Plinge, and Emanuël A. P. Habets
International Audio Laboratories Erlangen*, Germany
Email: {olli.rummukainen; axel.plinge; emanuel.habets}@iis.fraunhofer.de, sebastian.schlecht@audiolabs-erlangen.de

***Abstract*—This paper proposes a new method for evaluating real-time binaural reproduction systems by means of a wayfinding task in six degrees of freedom. Participants physically walk to sound objects in a virtual reality created by a head-mounted display and binaural audio. We show how the localization accuracy of spatial audio rendering is reflected by objective measures of the participants' behavior. The method allows for comparative evaluation of different rendering systems as well as the subjective assessment of the quality of experience.**

***Keywords*—Spatial sound; Binaural; Six-degrees-of-freedom**

I. Introduction

Quality evaluation of spatial sound systems in virtual reality (VR) applications is challenging but vital for the success of VR. A common methodology applied today in audio quality evaluations is the multiple stimulus with hidden reference and anchor (MUSHRA) test setup [1], where multiple test items are evaluated at once, comparing them to a predefined reference along varying sets of unidimensional perceptual scales [2]. An important perceptual attribute in spatial sound quality evaluation is the localization accuracy, which is under investigation in this study.

With more degrees of freedom in movement, and with multimodal sensory input, it becomes increasingly difficult to separate the audio quality from the quality of the whole VR system, or the complete Quality of Experience (QoE). Moreover, quality evaluation during immersion in VR systems is a problem with the MUSHRA-like paradigms, where participants must be constantly aware of the quality judgment task. Recently, there have been attempts to create objective QoE metrics from heart rate and electrodermal activity [3], but interpreting these results requires further research.

Another method to objectively assess the quality of VR systems is observing the ease of wayfinding. Ruddle & Lessels [4] identify three distinct levels of behavioral metric for evaluating wayfinding: task performance, physical behavior, and cognitive rationale. In this study, the localization accuracy of three real-time binaural rendering systems is indirectly inferred from the task performance and physical behavior during a wayfinding task. We show the method to discriminate between the systems without direct evaluation done by the participant.

*A joint institution of the Friedrich-Alexander-University Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits (IIS).

QoMEX2017 – Erfurt, Germany; 978-1-5386-4024-1/17/$31.00 ©2017 IEEE

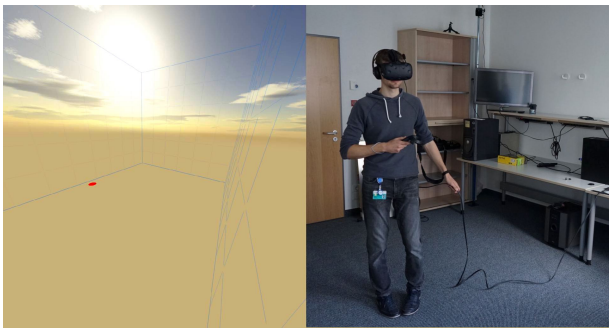

Fig. 1. Virtual and real worlds. Red circle shows the starting location and the blue grid boundaries of the search area.

II. Experimental setup and stimuli

The HTC Vive head-mounted display (HMD) with two tracking beacons [5] was used for estimating the position and orientation of the participant's head with 6 Degrees-of-Freedom (DoF), i.e, 3-axis translation and 3-axis rotation. The tracking information was sent to a Max/MSP patch locating the participant's point of view in a simple computer-generated imagery (CGI) world depicting an endless desert. The CGI world along with a view of the real world space are displayed in Fig. 1. The blue grid in the virtual world limits the search area, and prevents the participant from colliding with real world obstacles.

Audio was delivered via headphones (Beyerdynamic MMX2). The target sound was a looped 8-bit music sample, the Super Mario theme. Three sound rendering methods were under test: Binaural with 3-dimensional reverberation, binaural without reverberation, and intensity only monaural rendering. The binaural rendering combines high-resolution non-individualized head-related transfer functions (HRTF) with a distance-dependent gain, and the reverberation was created by a feedback delay network with early reflections matching the first and second order image sources. The sound objects were placed in six different locations at ear level right outside of the search area (3.4×2.6 m), as depicted in Fig. 2. The paths from the start location (black dot) to the targets were repeated once as mirrored versions along the horizontal axis. The paths were divided into three groups (i.e., long, mid, and short) based on the path lengths.

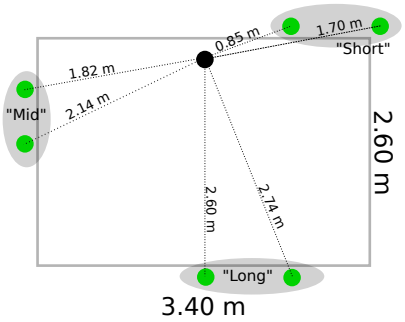

Fig. 2. Tracked area, starting location (black dot), and stimulus locations (green dots) with distances.

III. Method

A total of 9 people (2 female, 7 male) took part in the test. Average age of the participants is 29.7 years (SD=9.0). None reported any known hearing impairments and four participants are trained listeners. Two of the authors participated in the test.

The task is to locate an audio object and walk to its location as fast as possible. The participant begins from a starting location shown on the HMD. A target sound is then played that originates from a predefined location in space, and the participant walks to the target. The participant signals to have found the target by pressing a button on a controller. No feedback is given whether they are at correct location, instead the next starting location is visually presented and the same procedure is repeated.

There are in total 18 unique trials in one session (3 renderings $\times$ 6 paths). The trials are repeated once as mirrored versions resulting in 36 trials. The trials are presented in random order. There is a training session with two examples of each rendering method with an added visual cue to familiarize the participants with the stimuli. On average the test took 20 minutes to complete after the training.

IV. Results

Following Ruddle & Lessels [4], we computed three task performance metrics: time to complete a trial, path length traveled, and error at the end-point. These data are presented in the three first panels in Fig. 3. The unique trials are grouped into three path length groups as shown in Fig. 2. According to a two-way ANOVA, significant effects were found in path length and rendering method. Significant differences were found between intensity rendering compared to the two binaural rendering methods, but not between the binaural methods in all three metrics. Additionally, we evaluated physical behavior, i.e., what people were doing during the trial. We computed the accumulated head rotation angle in five normalized time bins, shown in the fourth panel in Fig. 3 collated across all path lengths. In all rendering conditions there is a decreasing tendency for the amount of rotation as the participants approach the target. Significant differences were again found with intensity rendering compared to the two binaural renderings.

Fig. 4 further explores the physical behavior during a trial by accumulating the time spent at normalized distances from the target. Intensity rendering stands out from the two other cases by displaying a more even distribution throughout the distances. Observable differences are found between the

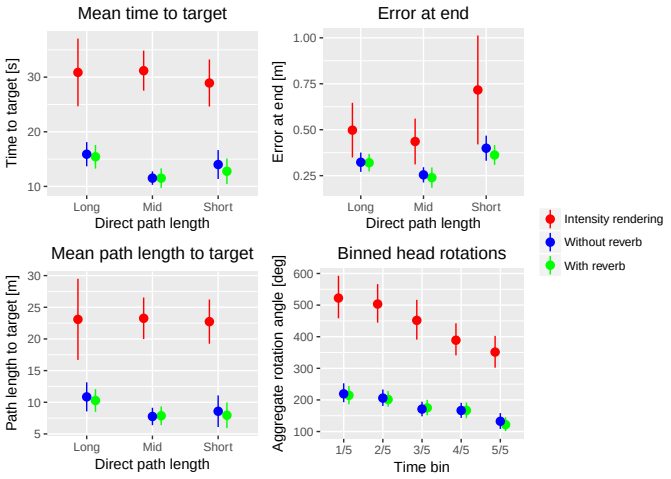

Fig. 3. Means for time to complete a trial, path length to target, error at end of trial, and aggregated head rotations in five time bins. Whiskers denote the 95 % confidence intervals of the mean.

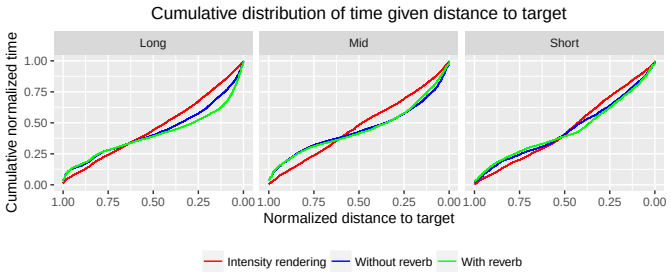

Fig. 4. Cumulative distribution function of time spent at different distances from the target. Distance to target is normalized from 1 at start to 0 at target.

binaural rendering methods in the time required to reach 0.25 of the normalized distance to target for the "mid" and "long" path groups. People seem to approach the target faster with reverberation, but are more indecisive about the final target location than without reverberation.

The cognitive rationales [4] were not systematically tested, but informal post-test discussions revealed the intensity rendering to be perceived as the most frustrating condition. Some mentioned the reverberant rendering having felt easier to complete, but the metrics do not lend support to this assertion. We can only speculate this to be due to the additional reverberation cue making it worthwhile to spend more time fine-tuning the final location, or vice-versa, there was something confusing in the near-field rendering.

V. Conclusions and Future Work

A method for indirectly evaluating localization accuracy of binaural reproduction systems from task performance and behavioral data was presented. Using the proposed method it was possible to discriminate between those reproduction systems without direct evaluation done by the participant.

Our next steps are to create a real-world reference with loudspeakers, to enlarge the tracked area for more freedom of movement, and to experiment with an extended set of behavioral metrics. Our future goals include extending the test suite with timbre related tests, and to approach a more general QoE evaluation based on behavioral data.

REFERENCES

[1] International Telecommunications Union, "Recommendation ITU-R BS.1534-3 Method for the subjective assessment of intermediate quality level of audio systems," Geneva, Switzerland, 2015.

[2] N. Zacharov and T. Holm Pedersen, "Spatial sound attributes - development of a common lexicon," in *AES 139th Convention*, New York, (NY), 2015, pp. 1–11.

[3] D. Egan, S. Brennan, J. Barrett, Y. Qiao, C. Timmerer, and N. Murray, "An evaluation of Heart Rate and Electrodermal Activity as an Objective QoE Evaluation Method for Immersive Virtual Reality Environments," in *Quality of Multimedia Experience (QoMEX)*, Lisbon, Portugal, 2016, pp. 1–6.

[4] R. A. Ruddle and S. Lessels, "Three Levels of Metric for Evaluating Wayfinding," *Presence: Teleoperators and Virtual Environments*, vol. 15, pp. 637–654, 2006.

[5] HTC, "HTC Vive Head-mounted Display." [Online]. Available: https://www.vive.com/