Article

The Influence of Binaural Room Impulse Responses on Externalization in Virtual Reality Scenarios

Song Li *, Roman Schlieper, Aly Tobbala and Jürgen Peissig

Institute of Communications Technology, Gottfried Wilhelm Leibniz Universität Hannover,
30167 Hannover, Germany; schlieper@ikt.uni-hannover.de (R.S.); alytobbala@gmail.com (A.T.);
peissig@ikt.uni-hannover.de (J.P.)
* Correspondence: song.li@ikt.uni-hannover.de

Abstract: A headphone-based virtual sound image can not be perceived as perfectly externalized if the acoustic of the synthesized room does not match that of the real listening environment. This effect has been well explored and is known as the room divergence effect (RDE). The RDE is important for perceived externalization of virtual sounds if listeners are aware of the room-related auditory information provided by the listening environment. In the case of virtual reality (VR) applications, users get a visual impression of the virtual room, but may not be aware of the auditory information of this room. It is unknown whether the acoustic congruence between the synthesized (binaurally rendered) room and the visual-only virtual listening environment is important for externalization. VR-based psychoacoustic experiments were performed and the results reveal that perceived externalization of virtual sounds depends on listeners' expectations of the acoustic of the visual-only virtual room. The virtual sound images can be perceived as externalized, although there is an acoustic divergence between the binaurally synthesized room and the visual-only virtual listening environment. However, the "correct" room information in binaural sounds may lead to degraded externalization if the acoustic properties of the room do not match listeners' expectations.

Keywords: room divergence effect; virtual reality; perceived externalization; binaural room impulse response; head-related transfer function; headphone-reproduced virtual sound images

1. Introduction

Headphone-based three-dimensional (3D) audio is becoming increasingly important due to the growing market for mobile devices, virtual/augmented/mixed reality (VR/AR/MR) applications, etc. [1]. A general way to create headphone-based virtual sounds is by convolving an input signal with a pair of binaural room impulse responses (BRIRs) [2]. Binaural room transfer functions (BRTFs) are the frequency domain representation of BRIRs. A BRIR describes an acoustic impulse response between a sound source and listener's eardrums in a room, and can be divided temporally into direct- and reverberant sound parts (early reflections and late reverberation). The direct sound part can be characterized by a head-related impulse response (HRIR, the time-domain representation of head-related transfer function (HRTF)), which contains relevant acoustic cues for sound localization [3].

Perceived externalization, i.e., out of the head [4], is one of the most important acoustic attributes related to the overall quality of headphone-based virtual sounds [1,5,6]. A vital factor for the perception of externalized sound images is the acoustic congruence between the binaurally synthesized room and the real listening environment. The degree of externalization decreases as the acoustic divergence between them increases, which is known as the room divergence effect (RDE) [7–9].

A typical way to demonstrate the RDE is to record test stimuli (or synthesize virtual sounds with BRIRs) in one room and play them back in a different room through headphones [7–9]. Gil-Carvajal et al. [10] investigated the effect of incongruent room-related

auditory and visual information in externalization of headphone-based virtual sounds. In their study, listeners were divided into two groups to assess externalization of stimuli. One group of subjects was blindfolded, but could hear sounds from a speaker in the playback room. In contrast, subjects in the second group could see the room but could not hear sounds in the room (except for test stimuli played back over headphones). In this way, the first and second groups received only auditory and visual information about the playback room, respectively. Their experimental results reveal that the congruent auditory information is more important to perceived externalization than the visual impression of the listening room. According to the localization model proposed by Plenge [7], the room-related auditory information is stored in human short-term memory and adapts quickly when the listening environment changes. The adaptation of this information is actually the familiarization phase to the listening environment.

The results of the above-mentioned studies [7–10] are important for binaural listening in AR/MR applications because listeners are aware of the acoustics of their listening environments. In order to perceive well-externalized sound images, the auditory information provided by the binaurally synthesized room should match that of the local listening room. However, for VR applications, listeners are introduced to a virtual environment that may differ (both acoustically and visually) from their local environment. VR users get a visual impression of the room, but may not be aware of its auditory information. It is unknown whether the acoustic congruence between the binaurally synthesized room and the visual-only virtual listening environment is important for externalization.

The present study aimed to investigate the influence of BRIRs with congruent and divergent auditory information on externalization in VR scenarios. The remainder of this paper is organized as follows. Section 2 introduces the experimental setups. The subjective evaluation results are presented and discussed in Sections 3 and 4, respectively. Finally, Section 5 concludes this study.

2. Experimental Setups

Two listening experiments were designed to evaluate the influence of BRIRs with different room-related auditory information on externalization in real (desktop-based experiment) and visual-only virtual (VR-based experiment) listening environments, respectively. The experimental setups were chosen to simulate common use cases of binaural listening. In this study, Oculus Rift was used as the VR headset for the VR-based experiment.

2.1. BRIR Measurements

Two pairs of BRIRs were measured with a dummy head KU-100 at azimuth angles of $0°$ (front) and $45°$ (front right) in a listening room (ITU-R BS.1116, $6.7\,\mathrm{m} \times 4.8\,\mathrm{m} \times 3.2\,\mathrm{m}$) and a computer room ($5.9\,\mathrm{m} \times 6.5\,\mathrm{m} \times 2.7\,\mathrm{m}$). The broadband reverberation time (RT_{60}) for these two rooms is about $0.25\,\mathrm{s}$ (listening room) and $0.6\,\mathrm{s}$ (computer room), respectively. In each room, two Neumann KH 120A loudspeakers located at these two incidence angles relative to the dummy head, were used as sound sources for the recording of non-individual BRIRs. The distance between loudspeakers and the dummy head was $2\,\mathrm{m}$. A $5\,\mathrm{s}$ long exponential sweep from $20\,\mathrm{Hz}$ to $20\,\mathrm{kHz}$ was used as an excitation signal [11], and the measurement was repeated five times. The BRIRs recorded were then windowed to $1\,\mathrm{s}$ with a $5\,\mathrm{ms}$ long half raised-cosine fall time. Non-individual BRIRs are often used in commercially available binaural rendering systems due to the difficulty in obtaining individual BRIRs for headphone users [11]. Since this paper aimed to investigate externalization of binaural sounds in common use cases, generic BRIRs were used in the experiments. Compared to the use of individual BRIRs, the localization errors of binaural sounds might be increased [12].

The main differences between these impulse responses were the reverberant components. Slight differences (e.g., minor differences in magnitude spectra at high frequencies) were observed in direct parts of BRIRs measured in the listening room and in the computer room due to the misalignment of the measurement setups. To avoid the differences in the direct parts, the direct parts of BRIRs measured in the computer room were replaced by

those measured in the listening room, and such modified BRIRs were used in the experiments. In this way, the potential influence of differences in direct sound components on externalization can be excluded. The difference in reverberant parts of BRIRs was the main factor that could lead to different externalization results.

The extraction of the direct parts was performed by windowing the BRIRs to 2.5 ms (after the onset delay) with a half raised-cosine fall time of 0.5 ms. No audible difference was found between the original and modified BRIRs through informal listening tests. It is well know that the direct parts of BRIRs can be approximated as HRIRs, which do not contain room information. In this study, the extracted direct sound components were treated as BRIRs/HRIRs measured in an anechoic chamber (an approximation of the free-field).

For convenience, the listening room, the computer room, and the anechoic chamber are referred to as room A, B, and C, respectively. In addition, their corresponding impulse responses are denoted as BRIR-A, BRIR-B, and BRIR-C, respectively.

2.2. Listeners and Test Stimuli

Ten normal-hearing subjects (2 females and 8 males, aged between 25 and 34) participated in both experiments. A 3 s long speech sentence taken from the European Broadcasting Union (EBU) Sound Quality Assessment Material (SQAM) [13] was used as the test stimulus. The obtained BRIRs (BRIR-A, BRIR-B, and BRIR-C) were convolved with the speech signal to simulate binaural recordings in the corresponding rooms (room A, B and C). In this way, three binaural signals were generated for each source direction (0° or 45°). In addition, a diotically played back (mono) speech sentence was used as an anchor signal, and subjects were informed that such stimulus should act as an "internalized" sound. This anchor was also included in the test stimuli to be evaluated (hidden anchor). For each source direction, four stimuli were presented in randomized order and should be evaluated in terms of externalization. The playback level of the binaurally rendered anechoic (BRIR-C) speech signal was 56 dBA (calibrated with the dummy head).

A speech sentence was used as a test stimulus in both experiments because it resembled a real-life sound source. Leclère et al. [14] revealed that stimulus type had a slight but did not have a significant influence on externalization. It can therefore be assumed that our results do not change noticeably when other stimuli such as noise and music are used.

2.3. Listening Environments

Two experiments were designed to evaluate externalization of presented stimuli in real (Experiment I) and visual-only virtual (Experiment II) listening environments, and both of them were performed physically in the listening room (room A).

In Experiment I, subjects could see the real listening environment (room A). Two real loudspeakers were placed at the positions where BRIRs were previously measured, and they were used as reference positions for the externalization assessment (see Section 2.4). As shown in the top panel of Figure 1, a subject listened to the stimuli through a pair of headphones and rated externalization using the graphical user interface (GUI) of the listening experiment displayed on the laptop screen (desktop-based experiment).

For Experiment II (VR-based experiment), a computer room and an anechoic chamber were virtually modeled as shown in Figure 2. The virtual computer room was the same size as the real one in our institute, and the size of the virtual anechoic chamber was chosen to be the same as that of the computer room. In addition, a pair of virtual loudspeakers was modeled and placed at the same locations where the BRIRs were measured. Subjects performed the listening experiment in these two virtual environments using the VR headset (see bottom panel of Figure 1).

Figure 1. Photographs of the experimental setups for performing desktop (**top**)- and VR (**bottom**)-based listening experiments.
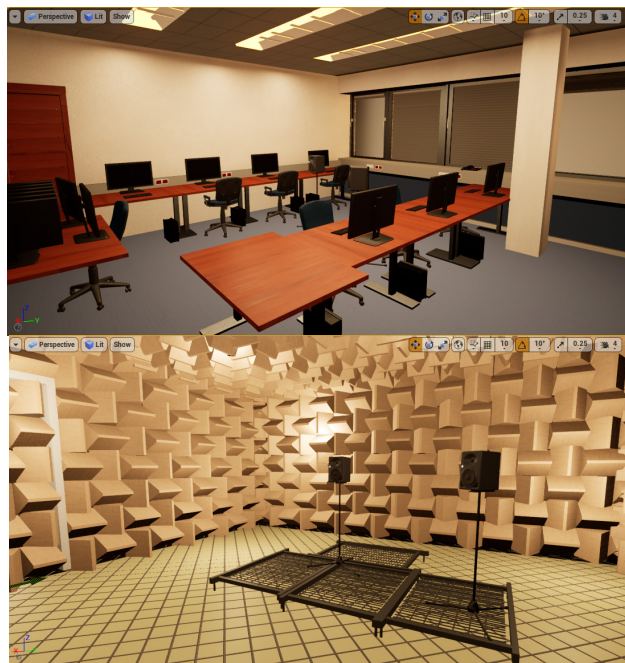

Figure 2. Simulated virtual rooms (**top**: computer room; **bottom**: anechoic chamber).

2.4. Externalization Evaluation

As shown in Table 1, a four-point discrete rating scale was used to assess the degree of externalization, similar to that used in [4,15].

Table 1. Rating scale for perceived externalization.

Degree	Meaning of the Degree
3	The sound is well externalized and at the speaker's position.
2	The sound is externalized but not at the speaker's position.
1	The sound is not well externalized. It is very close to me.
0	The sound is in my head.

According to this rating scale, a listening experiment was designed for evaluating externalization of test stimuli. In one listening session, three binaurally rendered signals from one source direction (0° or 45°), and the anchor signal were presented. In Experiment I, the GUI of the listening test was displayed on the laptop screen (implemented in the software Max/MSP), and subjects were asked to rate each stimulus using a slider with a mouse (top panel of Figure 3). In Experiment II, the GUI of the test was shown in the VR display (implemented in the software Unreal Engine), and subjects could rate the test stimuli using the trigger of the Oculus Rift controller (bottom panel of Figure 3). With the consideration of the clarity and conciseness, the descriptions of externalization levels in GUIs have been slightly modified. Subjects were informed about the meaning of the levels in the rating scale.
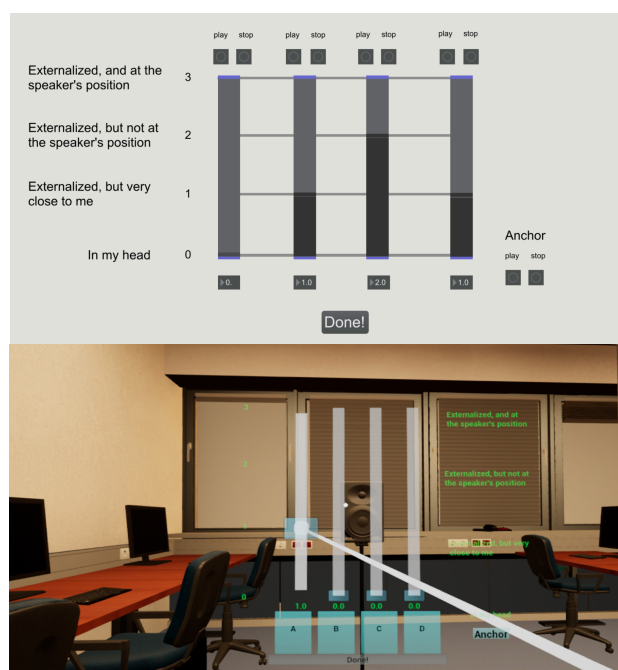

Figure 3. GUI of the externalization experiment (**top**: desktop-based listening test (Experiment I); **bottom**: VR-based listening test (Experiment II)).

2.5. Experiment I: Desktop-Based Experiment

In Experiment I, subjects listened to the test stimuli presented over a pair of Sennheiser HD800 headphones. The influence of the headphone transfer function (HpTF) to the binaural signals was compensated by calculating its inverse filter. The HpTF was measured on the dummy head with ten repetitions (the headphone was repositioned by each measurement), and the compensation filter was calculated by using the least-squares inversion approach [16]. Subjects were asked to evaluate externalization of the binaurally generated signals and the hidden anchor. They were able to repeat each test stimulus and listened to the stimuli in arbitrary order. The experiment was tested twice, resulting in 16 stimuli (4 stimuli/session × 2 directions × 2 presentations/subject) to be evaluated. The externalization rating for each subject was taken as the average of these two scores. Subjects were

asked to rate externalization of each stimulus and to ignore audible artefacts that do not affect their externalization perception [17]. Before starting the experiment, subjects listened to speech signals reproduced over loudspeakers for about 3 min to become familiar with the acoustic of the listening room. In addition, they could walk around the room. This study focused on externalization of virtual sounds in static situations (no head tracking), i.e., the position of virtual sound sources changed with listeners' head movements. Several studies have reported that externalization of static sound sources can be reduced when listeners rotate their heads without head tracking [18–20]. Hence, subjects were not allowed to rotate their heads during the listening experiment (monitored by the supervisor through a transparent window). When a large head movement was observed, the listener was asked to stop the experiment immediately and restart it.

2.6. Experiment II: VR-Based Experiment

In Experiment II, subjects physically sat on a chair in the listening room, viewed one of the virtual rooms through the VR headset (see Section 2.3) and listened to the test stimuli presented over headphones. They were asked to perform listening test for each virtual room, and the order of the presented room was randomized for each subject. As in Experiment I, subjects (without wearing the VR headset) were asked to listen to speech signals played back over loudspeakers for 3 min prior to the start of this experiment. In this way, the acoustic information of the listening room remained in the listeners' short-term memory [15]. Further, before starting the listening experiment in each session (with the VR headset worn), subjects were encouraged to rotate their heads to see their virtual listening environment (without reproducing the stimuli). During the test, they were not allowed to rotate their heads. In total, 32 stimuli (2 virtual rooms × 4 stimuli/session × 2 directions × 2 presentations/subject) had to be evaluated in terms of externalization. All other settings were the same as those in Experiment I.

3. Experimental Results

This section shows the experimental results for desktop- and VR-based listening experiments. The anchor signal could be unambiguously identified by each subject (externalization score = 0), therefore its rating is not shown in the following figures. A Shapiro–Wilk test revealed that the data for each condition (BRIR type) did not all have a normal distribution [21]. As a result, non-parametric statistical tests were performed for analyzing the experimental results. A Friedman test was used to analyze the significant effect of BRIR types on externalizing ratings [21]. Furthermore, Wilcoxon tests were performed to analyze significant differences in externalization results between BRIR types, and p-values were adjusted using the Bonferroni correction [21].

3.1. Results for Experiment I

Figure 4 shows the externalization results for the desktop-based experiment using box plots. Each panel shows the results for a different source direction (left: 0°; right: 45°). It can be seen that the externalization ratings of lateral sources are generally higher than those of frontal sources. For each source direction, the results affected significantly by different BRIR types (Friedman test: $p \ll 0.05$). The ratings for the condition "BRIR-C" are significantly lower than for the other two conditions (a series of Wilcoxon tests: $p < 0.04$). Overall, subjects rated the externalization higher for the "BRIR-A" condition than for the other two conditions, consistent with the theory of the RDE [7–9]. However, only a weak enhancement of externalization is observed for the condition "BRIR-A" compared to the condition "BRIR-B" (Wilcoxon tests: $p \approx 0.06$ for both directions). Nevertheless, this result suggests that matching the room acoustics between the binaural signals and the real listening environment is important for externalization.
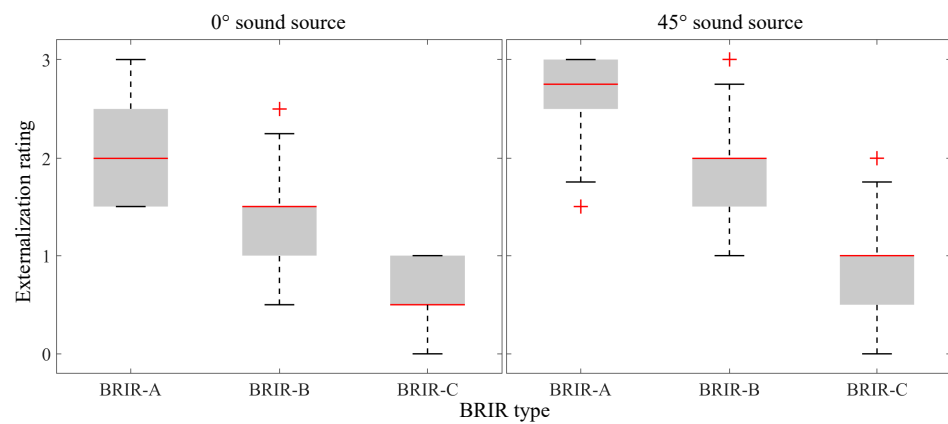
Figure 4. Box plot of externalization ratings for stimuli generated with different BRIR types in Experiment I. The left and right panels show the externalization results for 0° and 45° sound sources, respectively. The median externalization ratings (red lines), 25th and 75th percentiles (bottom and top edges of the boxes), whiskers (1.5 × interquartile range) and outliers (red crosses) are shown in the figure.

3.2. Results for Experiment II

Figure 5 shows the externalization ratings for the VR-based experiment with box plots. Each column shows the results for a different source direction (left: 0°; right: 45°). The top and bottom rows denote the externalization results for virtual rooms B and C, respectively. Similar to the results obtained in Experiment I, the Friedman test confirmed the significant effect of BRIR types on externalization for each source direction and virtual room ($p \ll 0.05$). Additionally, the results for the condition "BRIR-C" are significantly lower than for the other two conditions (a series of Wilcoxon tests: $p < 0.02$). Although the median values of externalization are slightly higher for the condition "BRIR-A" than for the condition "BRIR-B", the difference becomes smaller compared to that in Experiment I (Wilcoxon tests: $p > 0.5$ for both directions and virtual rooms). When listening in these two visual-only virtual environments, the externalization ratings of binaural sounds rendered by BRIR-B increase, but the sounds generated by HRIRs (BRIR-C) can still not be perceived as well externalized. It seems that the auditory information provided by the local listening environment does not play a dominant role for determining externalization in VR scenarios, as ratings for the conditions "BRIR-A" and "BRIR-B" are both high. Furthermore, regardless of the visual impression of the virtual rooms, subjects consistently gave higher externalization ratings for reverberant sounds (BRIR-A and BRIR-B) than for anechoic sounds.
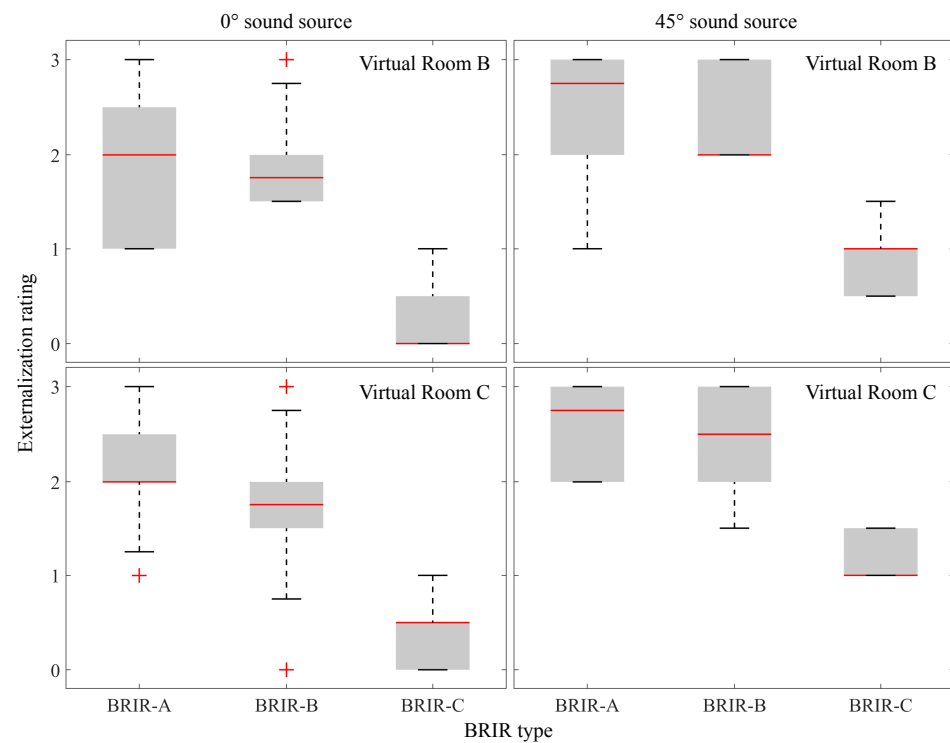
Figure 5. Box plot of externalization ratings for stimuli generated with different BRIR types in Experiment II. The left and right columns show the externalization results for 0° and 45° sound sources, respectively. The top and bottom rows denote the results for virtual rooms B and C, respectively. All other conventions are as in Figure 4.

4. Discussion

Experiment I and II were conducted in real and visual-only virtual listening environments, respectively. The first experiment was designed to confirm the influence of RDE on externalization and further to compare with the results obtained in visual-only VR listening environments. The second experiment was used to investigate the externalization of binaural sounds in common use cases of VR scenarios. In the following these experimental results are discussed.

4.1. The Influence of BRIRs on Externalization in a Real Listening Environment

Experiment I can be considered as a common use case for the application of binaural listening, where subjects could see the real listening environment and hear virtual sounds over headphones (AR/MR scenarios). In addition, they were familiar with the acoustics (e.g., reverberation time) of the listening room.

As expected, the median externalization ratings of frontal and lateral virtual sounds for the congruent condition (BRIR-A) are the highest compared to divergent conditions (BRIR-B and BRIR-C). This phenomenon is well explained by the RDE, where the externalization ratings are highly dependent on the acoustic congruence between the binaurally synthesized and the real room [7–10]. Maybe due to the use of non-individual BRIRs, only a weak significant difference of externalization between conditions BRIR-A and BRIR-B is observed ($p \approx 0.06$). A similar observation can be found in [15].

For the two divergent conditions, externalization results of reverberant sounds (BRIR-B) are significantly higher than those of anechoic sounds (BRIR-C). This outcome can be explained by the fact that listeners are more familiar with reverberant sounds due to their daily life. It should be noted that anechoic sounds can be perceived as well externalized if the experiment is conducted in an anechoic chamber (sufficient familiarization with the listening environment is required) [4,22]. The results of this experiment are consistent with the findings of previous studies [7–10], and reveal the importance of RDE for externalization when listening to binaural sounds in AR/MR scenarios.

4.2. The Influence of BRIRs on Externalization in a Visual-Only Virtual Listening Environment

In the case of VR applications (e.g., gaming), users are usually introduced to virtual rooms that are different from the rooms in which they are physically present. Compared to the AR/MR scenarios in Experiment I, only the visual information provided to the listeners changes.

When listening binaural sounds in visual-only virtual rooms B and C, the externalization ratings are high not only for the BRIR-A but also for the BRIR-B condition. Although the median externalization results differ slightly, there is no significant difference between them ($p > 0.5$). Eight out of ten subjects have visited or performed activities in anechoic chambers and know the room acoustics of such special rooms, but the ratings for anechoic sounds (BRIR-C) when listening in the virtual room C are still low. The experimental results can be explained by the fact that the listeners were not aware of the acoustic information of these two visual-only virtual rooms during the experiment, and they evaluated the sounds based on their expectations or experiences with the acoustics of these two virtual rooms. For the virtual room C, listeners are familiar with the reverberant sounds in daily life and such experience is stored in human short-term memory, and they expected to hear reflections in the virtual room. Therefore, they rated the anechoic sounds as low based on their expectations and experiences [23].

In AR/MR scenarios, listeners could learn the room-related auditory information of the local listening environment through any kind of acoustic feedback from the room, e.g., their own voice, footsteps, etc. To perceive well-externalized virtual sounds, the room-related auditory cues contained in binaural sounds should match to that provided by the room. As reported in Experiment I, the use of BRIRs (e.g., matched reverberation) for generating well-externalized binaural sounds is therefore restricted.

According to the results in Experiment II, it seems that the use of BRIRs is not restricted in VR scenarios compared to that in AR scenarios when listening in typical virtual rooms. When no acoustic feedback provided by the virtual room, listeners' expectations are generated based on their experiences and the visual impression of the room, e.g., the size and decoration of the room. If the room-related acoustic information contained in binaural sounds meets listeners' expectations, the sounds can be perceived as externalized. However, the "correct" room information contained in binaural sounds may degrade perceived externalization when listening in some special virtual rooms. This is the case when listening to anechoic sounds in the virtual room C.

4.3. Limitations of This Study

One limitation is the use of non-individual BRIRs in both experiments. Based on the results of previous studies, the influence of RDE on externalization maybe more pronounced when using individual BRIRs, especially for frontal sound sources [9,15]. However, various studies have investigated the effect of individualization on externalization, the results are still inconsistent [4,5,14,24,25]. Since most commercially available binaural rendering systems are based on non-individual HRTFs/BRIRs [11], the experimental results are valuable for the externalization study in common use cases of binaural listening.

Furthermore, only two virtual rooms were tested in the experiment. An anechoic chamber is chosen as a special room for testing listeners' expectations of externalized binaural sounds. To validate the results of this study, other rooms with different sizes and acoustic properties should be further used.

Finally, this study investigated perceived externalization of static binaural sounds, and the dynamic scenarios with head movements were not taken into account. Several studies have reported that the dynamic cues caused by head movements (with head tracking) play an important role in externalization, especially for frontal sound sources [18–20]. It can be hypothesized that the overall externalization rating of frontal sound sources should be enhanced when using a binaural rendering system coupled with a head tracking device. Hence, to generalize the outcomes of this study, future experiments should consider the dynamic conditions.

5. Conclusions

This study investigated externalization of binaural sounds in VR applications. When listening in a virtual room and there is no acoustic feedback from the room, the use of BRIRs needs to match listeners' expectations based on the visual impression of the room. In comparison with the binaural listening in common scenarios of AR/MR applications, the use of BRIRs is not as restricted as in VR scenarios. However, the "correct" room information in binaural sounds may degrade perceived externalization if the acoustic properties of the room do not match the listener's expectations.

Future work includes using individual BRIRs, designing more virtual rooms, and allowing head movements for externalization experiments. In addition, the influence of acoustic feedback from virtual rooms on externalization will be investigated.

References

1. Sunder, K.; He, J.J.; Tan, E.L.; Gan, W.S. Natural Sound Rendering for Headphones: Integration of signal processing techniques. *IEEE Signal Process. Mag.* **2015**, *32*, 100–113. [CrossRef]
2. Vorländer, M. *Auralization*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2008. [CrossRef]
3. Wightman, F.L.; Kistler, D.J. Headphone simulation of free-field listening. I: Stimulus synthesis. *J. Acoust. Soc. Am.* **1989**, *85*, 858–867. [CrossRef] [PubMed]
4. Hartmann, W.M.; Wittenberg, A. On the externalization of sound images. *J. Acoust. Soc. Am.* **1996**, *99*, 3678–3688. [CrossRef] [PubMed]
5. Best, V.; Baumgartner, R.; Lavandier, M.; Majdak, P.; Kopčo, N. Sound Externalization: A Review of Recent Research. *Trends Hear.* **2020**, *24*, 1–14. [CrossRef]
6. Sakamoto, N.; Gotoh, T.; Kimura, Y. On "Out-of-Head Localization" in Headphone Listening. *J. Audio Eng. Soc.* **1976**, *24*, 710–716.
7. Plenge, G. On the problem of "In Head Localization". *Acta Acust. United Acust.* **1972**, *26*, 241–252.
8. Udesen, J.; Piechowiak, T.; Gran, F. The Effect of Vision on Psychoacoustic Testing with Headphone-Based Virtual Sound. *J. Audio Eng. Soc.* **2015**, *63*, 552–561. [CrossRef]
9. Werner, S.; Klein, F.; Mayenfels, T.; Brandenburg, K. A summary on acoustic room divergence and its effect on externalization of auditory events. In Proceedings of the 8th International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 6–8 June 2016. [CrossRef]
10. Gil-Carvajal, J.C.; Cubick, J.; Santurette, S.; Dau, T. Spatial Hearing with Incongruent Visual or Auditory Room Cues. *Sci. Rep.* **2016**, *6*, 1–10. [CrossRef] [PubMed]
11. Li, S.; Peissig, J. Measurement of Head-Related Transfer Functions: A Review. *Appl. Sci.* **2020**, *10*, 5014. [CrossRef]
12. Ben-Hur, Z.; Alon, D.; Robinson, P.W.; Mehra, R. Localization of virtual sounds in dynamic listening using sparse HRTFs. In Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality, Online, 17–19 August 2020.
13. EBU. 3253: Sound Quality Assessment Material Recordings for Subjective Tests, 2008. Available online: https://tech.ebu.ch/publications/sqamcd (accessed on 1 October 2021).
14. Leclère, T.; Lavandier, M.; Perrin, F. On the externalization of sound sources with headphones without reference to a real source. *J. Acoust. Soc. Am.* **2019**, *146*, 2309. [CrossRef] [PubMed]
15. Wendt, F.; Höldrich, R.; Marschall, M. How binaural room impulse responses influence the externalization of speech. In Proceedings of the Tagungsband Fortschritte der Akustik-DAGA, Rostock, Germany, 18–21 March 2019.
16. Schärer, Z.; Lindau, A. Evaluation of Equalization Methods for Binaural Signals. In Proceedings of the 126th AES Convention, Munich, Germany, 7–10 May 2009.
17. Lindau, A.; Erbes, V.; Lepa, S.; Maempel, H.J.; Brinkman, F.; Weinzierl, S. A Spatial Audio Quality Inventory (SAQI). *Acta Acust. United Acust.* **2014**, *100*, 984–994. [CrossRef]

18. Brimijoin, W.O.; Boyd, A.W.; Akeroyd, M.A. The Contribution of Head Movement to the Externalization and Internalization of Sounds. *PLoS ONE* **2013**, *8*, e83068. [CrossRef] [PubMed]
19. Hendrickx, E.; Stitt, P.; Messonnier, J.C.; Lyzwa, J.M.; Katz, B.F.G.; Boishéraud, C.D. Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis. *J. Acoust. Soc. Am.* **2017**, *141*, 2011–2023. [CrossRef] [PubMed]
20. Li, S.; Schlieper, R.; Peissig, J. The Impact of Head Movement on Perceived Externalization of a Virtual Sound Source with Different BRIR Lengths. In Proceedings of the AES International Conference on Immersive and Interactive Audio, York, UK, 27–29 March 2019.
21. Bech, S.; Zacharov, N. *Perceptual Audio Evaluation–Theory, Method and Application*; Wiley: Hoboken, NJ, USA, 2007.
22. Li, S.; Baumgartner, R.; Peissig, J. Modeling perceived externalization of a static, lateral sound image. *Acta Acust.* **2020**, *4*, 21. [CrossRef]
23. Barron, M. The subjective effects of first reflections in concert halls-The need for lateral reflections. *J. Sound Vib.* **1971**, *15*, 475–494. [CrossRef]
24. Kim, S.M.; Choi, W. On the externalization of virtual sound images in headphone reproduction: a Wiener filter approach. *J. Acoust. Soc. Am.* **2005**, *117*, 3657–3665. [CrossRef] [PubMed]
25. Begault, D.R.; Wenzel, E.M.; 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. *J. Audio Eng. Soc.* **2001**, *49*, 904–916. [PubMed]