Experiment on externalization in binaural directional-source auralization

Bartłomiej Mróz[1], Franz Zotter[2], Florian Wendt[2], Markus Zaunschirm[2], Matthias Frank[2]
[1] *Multimedia Systems Department, Faculty of Electronics, Telecommunications, and Informatics*
Gdańsk University of Technology, Narutowicza 11/12 street, 80-233 Gdańsk, Poland
e-mail: bartlomiej.mroz@pg.edu.pl
[2] *Institut für Elektronische Musik und Akustik, Inffeldgasse 10/3, 8010 Graz, Austria*
e-mail: {zotter,wendt,zaunschirm,frank}@iem.at

Introduction

Real-world sources are typically perceived to be located outside the head. Dynamic binaural rendering, however, sometimes suffers from a collapse of externalization so that sources are perceived inside the listeners head [7]. Parameters influencing the externalization of binaurally rendered sources include (i) individualized head-related transfer functions (HRTFs) [10], (ii) head tracking [3], (iii) reverberation [4], (iv) listening room (e.g. room divergence effect [15, 14]) and (v) individual headphone equalization [2, 5]. Supplementing what is known from previous studies, e.g. [1, 12], we aim at covering and revisiting most of the recent and relevant aspects in the experimental study presented below.

Experimental design

It has been shown in [13] that a source with variable directivity can be used to influence the perceived distance of a source in a real room by controlling the direct-to-diffuse ratio. The configuration yielding the most distant impression is used for the experiment here. Further, the room is modeled via a source and receiver directional (SRD) RIR in 2D Ambisonics as proposed in [16]. The source order is set to $N_S = 1$ and the receiver order to $N_R = 11$.

Room Simulation

The pizza-box room with dimensions 11m $\times$ 12m is simulated using a simple 2D image source model up to order 3. The corresponding direction of arrival (DOA), time of arrival (TOA), and level for each propagation path is listed in Tab. 1. The diffuse reverberation tail is constructed from normally distributed white noise generated separately for each of the SRDRIR channels. To involve frequency dependency of the diffuse tail, it was divided into octave bands (see Tab. 2) in each of which a decaying envelope was applied matching the desired reverberation time (T_{60}). The generation of the modeled SRDRIR is depicted in Fig. 1 for a single channel, where $q = 0.0011$, and the diffuse part is faded in with 20ms starting at the mixing time [9].

Binaural synthesis, HRTFs, and Equalization

Resulting from the SRDRIR, 2D Ambisonics signals up to the order 11 are decoded to a equiangular 24 channel virtual loudspeaker array via a pseudo-inverse decoder. The head-rotation of listener is tracked via a compact

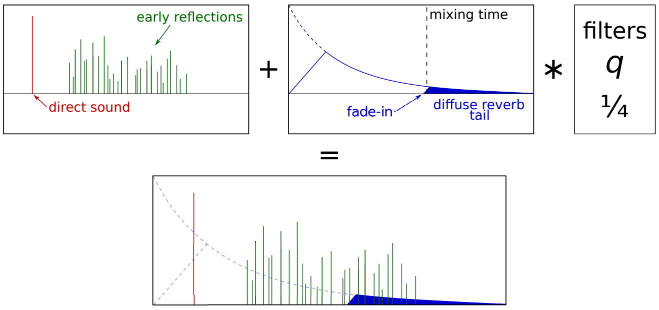

Figure 1: Schematic RIR modeling using a 2D image source model of order 3 and a diffuse modeled in octave-bands.

headphone-mounted device [11] and the entire scene is rotated via a simple frequency-independent matrix multiplication prior to decoding. The decoded binaural signals are obtained by a convolution of the loudspeaker signals with the HRTFs (artificial head or individualized) for the corresponding direction (nearest neighbor selection).
The individual far-field HRTFs used in experiment were measured at the Acoustics Research Institute of Austrian Academy of Sciences in Vienna for 1550 directions and the non-individual HRTFs were selected from KEMAR database.
Instead of measured headphone equalization, we had all participants follow Griesinger's equalization procedure [6] for the last condition, to account for the colorations introduced by the headphones of the experiment.

Listening Experiment

The overall 6 male listeners (average age 30 years) with experience in spatial audio were asked to rate the quality of externalization on a continuous scale from *poor* to *very good* in a multi-stimulus task. In the experiment we varied

- the complexity of the modeled room acoustics: only the direct sound, direct sound with early reflections; direct sound with early reflections and diffuse reverberation tail;
- the dynamic behavior: head tracking on and off;
- the HRTFs: individual vs. non-individual;
- the headphone equalization: on vs. off.

Additionally, two anchor conditions were involved that model the time delays simulating the ITD only (with and without headtracking). All experimental conditions are listed in Table 3.

order	DOA (deg)	TOA (ms)	LVL (dB)	sign	order	DOA (deg)	TOA (ms)	LVL (dB)	sign	order	DOA (deg)	TOA (ms)	LVL (dB)	sign	order	DOA (deg)	TOA (ms)	LVL (dB)	sign
0	0	5	0.0	-															
1	-9	12	3.7	+	2	-100	59	-16.0	+	3	-94	60	-13.6	+	3	-14	83	-12.7	+
1	53	13	-7.4	+	2	153	60	-23.9	-	3	64	63	-20.0	+	3	-150	83	-37.2	+
1	-99	48	-24.9	+	2	70	62	-26.3	+	3	154	67	-10.7	+	3	118	85	-28.0	-
1	164	58	-21.7	-	2	164	66	-10.8	+	3	71	73	-17.7	+	3	-48	89	-25.7	-
2	31	17	0.2	+	2	-14	76	-23.9	-	3	-6	77	-25.5	-	3	-103	109	-30.2	+
2	-91	50	-17.1	+	2	-156	76	-25.5	-	3	-159	81	-14.8	+	3	164	129	-28.4	-

Table 1: DOA, TOA, and level of each simulated propagation path up to an image source order of 3.

Type	T_{60}	Lower cut-off	Upper cut-off
Lowpass	700 ms		250 Hz
Bandpass	500 ms	250 Hz	500 Hz
Bandpass	500 ms	500 Hz	1000 Hz
Bandpass	500 ms	1000 Hz	2000 Hz
Bandpass	415 ms	2000 Hz	4000 Hz
Bandpass	335 ms	4000 Hz	8000 Hz
Highpass	250 ms	8000 Hz	

Table 2: Filter bank design parameters for the diffuse reverberation tail

Condition	room impulse response length		
1,2	direct sound only (dry)		
3,6,9,12	direct		
4,7,10,13	direct+early		
5,8,11,14	direct+early+diffuse		
Condition	**HRTF**	**HeadTracking**	**EQ**
1	delay time	✗	✗
2	delay time	✓	✗
3-5	KEMAR	✗	✗
6-8	KEMAR	✓	✗
9-11	individual	✓	✗
12-14	individual	✓	✓

Table 3: List of conditions evaluated in the experiment.

Each listener had to complete the multi-stimulus rating task perform the listing experiment in an anechoic chamber and in a lecture room (randomized order). In each room, the multi-stimulus task had to be completed twice, each time with with randomized presentation order of the 14 conditions. The source signal was male speech (EBU SQAM CD, track ID 50) and resulting stimuli were played back in a loop via Stax Omega II (SR-007) headphones. Listeners were allowed to seamlessly switch between the stimuli and listen to them as often as desired.

Results

In the following section we discuss the results of all ratings shown as median and 95% confidence intervals in Figs. 2 and 3. Independent of the listening room, the anchor conditions (delay time stereophony) were rated to be inside the head (poor externalization). As assumed, increasing the complexity of the room model also increases externalization. The addition of early reflections significantly improves externalization, while further addition a diffuse part gives comparable results to the direct plus early reflections simulation.
Apparently, head-tracking and the use of individualized HRTFS can further improve the externalization, while the headphone equalization shows no significant improvement. Note that similar results are found in [7, 3, 8].

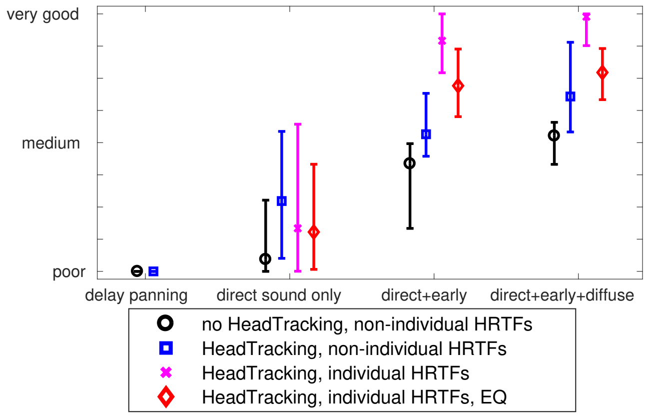

Figure 2: Results for externalisation rating trial in the lecture room – headphone playback, speech signal. The results were scaled and shifted for every individual to fill the entire scale.

Room divergence effect

As stated in [15], the externalization of a binaurally rendered sound scene decreases as the acoustic impression of the listening room diverges from the auralized room. This hypothesis is partly supported by the results obtained here. While rendering of the direct part only yields almost poor externalization in the lecture room, some listeners perceived a well-externalized source in the anechoic chamber. This may be explained by the similarity of the presented acoustic scene with the expected room acoustics of the listening room. For the lecture room an increase of complexity in the modeled room also improves externalization while no significant improvement is found for the anechoic chamber.

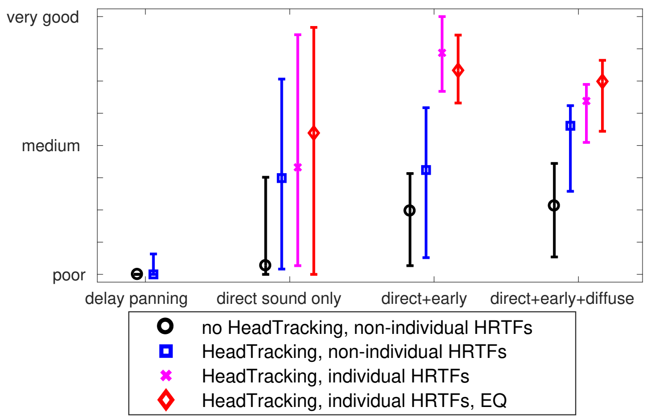

Figure 3: Results for externalisation rating trial in the anechoic chamber – headphone playback, speech signal. The results were scaled and shifted for every individual to fill the entire scale.

Conclusions

As expected, the anchor conditions (no room, no HRTFs) were perceived to be located inside the head, and externalization improves as the complexity of the room simulation increases. Whereas the addition of early reflections improves externalization, the addition of the diffuse part may not be as important. This complies with literature that suggests that the early 30 ms of the room response is beneficial for externalization [4]. In addition, our results also confirmed that dynamic binaural synthesis (head tracking) improves externalization, and rendering with individualized HRTFs outperformed the results obtained for rendering with the HRTFs of an artificial head (KEMAR data set). Finally, our results also confirm that avoiding room divergence is beneficial, as externalization for listening in the anechoic environment suffers more from adding diffuse reverberation.
The benefit of the particular procedure employed for third-octave equalization between headphones and a frontal loudspeaker was negligible, or slightly counterproductive.

Acknowledgments

We would like to thank all listeners taking part in the experiment and Michael Mihocic from the Acoustics Research Institute of Austrian Academy of Sciences in Vienna for the HRTFs measurement.

References

[1] Durant R. Begault, Elisabeth M. Wenzel, and Mark R. Anderson. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *Journ. Audio Eng. Soc.*, 49(10):904–916, 2001.

[2] Zamir Ben-Hur, Fabian Brinkmann, Jonathan Sheaffer, Stefan Weinzierl, and Boaz Rafaely. Spectral equalization in binaural signals represented by order-truncated spherical harmonics. *The Journal of the Acoustical Society of America*, 141(6):4087–4096, 2017.

[3] W. Owen Brimijoin, Alan W. Boyd, and Michael A. Akeroyd. The contribution of head movement to the externalization and internalization of sounds. *PLoS ONE*, 8(12):1–12, 2013.

[4] Ryan Crawford-Emery and Hyunkook Lee. The subjective effect of brir length perceived headphone sound externalisation and tonal colouration. In *prepr. 9044, 136th AES Conv.*, Berlin, 2014.

[5] David Griesinger. Accurate timbre and frontal localization without head tracking through individual eardrum equalization of headphones. In *Audio Engineering Society Convention 141*, 09 2016.

[6] David Griesinger. Laboratory reproduction of binaural concert hall measurements through individual headphone equalization at the eardrum. *Audio Engineering Society Convention Paper*, (9691), 05 2017.

[7] William M. Hartmann and Andrew Wittenberg. On the externalization of sound images. *The Journal of the Acoustical Society of America*, 99(6):3678–3688, 1996.

[8] Etienne Hendrickx, Peter Stitt, Jean-Christophe Messonnier, Jean-Marc Lyzwa, Brian FG Katz, and Catherine de Boishéraud. Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis. *The Journal of the Acoustical Society of America*, 141(3):2011–2023, 2017.

[9] Alexander Lindau, Linda Kosanke, and Stefan Weinzierl. Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses. *Audio Engineering Society Convention Paper*, (8089), 05 2010.

[10] Josefa Oberem, Bruno Masiero, and Janina Fels. Experiments on the authenticity and plausibility of binaural reproduction via headphones employing different recording methods. *Applied Acoustics*, 114:71–78, 2016.

[11] Michael Romanov, Paul Berghold, Matthias Frank, Daniel Rudrich, Markus Zaunschirm, and Franz Zotter. Implementation and evaluation of a low-cost headtracker for binaural synthesis. *Audio Engineering Society Convention Paper*, (9689), 05 2017.

[12] Florian Völk. Externalization in data-based binaural synthesis. In *Fortschritte der Akustik - NAG/DAGA*, Rotterdam, 2009.

[13] F. Wendt, M. Frank, F. Zotter, and R. Höldrich. Directivity patterns controlling the auditory distance, 2016.

[14] Stephan Werner, Georg Götz, and Florian Klein. 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*, May 2017.

[15] Stephan Werner, F. Klein, T. Mayenfels, and K. Brandenburg. A summary on acoustic room divergence and its effect on externalization of auditory events. In *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6, 06 2016.

[16] Markus Zaunschirm, Christoph Baumgartner, Christian Schörkhuber, Matthias Frank, and Franz Zotter. An Efficient Source-and-Receiver-Directional RIR Measurement Method. In *Fortschritte der Akustik AIA-DAGA 2017*, pages 1343–1346, Kiel, 2017.