Evaluating out-of-head localization of a dry source from the front reproduced with cue-preserving headphones

Hannes Pomberger[1], Alois Sontacchi[1], Matthias Frank[1],
Robert Höldrich[1],Thomas Gmeiner[2], Michele Lucchi[2]
[1] *Institut für Elektronische Musik und Akustik, Universität für Musik und Darstellende Kunst, Graz, Austria*
[2] *USound GmbH, Kratkystraße 2, 8020 Graz, Austria*
Email: pomberger@iem.at

Introduction

The binaural reproduction of sound sources from the frontal direction is particularly challenging. Their perceived image typically suffers from vertical mislocalization, in-head localization, an unnatural source widening or may even lead to a fuzzy cloud of several indistinct auditory events. In literature, different reasons for this problem can be found: the lack of individual spectral pinna-cues, see [1], [2], as well as missing room information, see [1], [3]. However, a real sound source in an anechoic environment is typically localized clearly outside of the head even for a frontal position, see [4]. As shown in [5], without any room information the perceived sound sources are located in the proximity of the head. Room information significantly supports distance perception and consequently externalization, however, it cannot provoke out-of-head localization by itself, c.f. [6].

Using a new headphone concept, we investigate dry reproduction of frontal sources as well as reproduction with little reverberation via headphones. This new concept includes additional tiny loudspeakers, which create similar individual spectral pinna-cues as a source from the frontal position [7], [8].

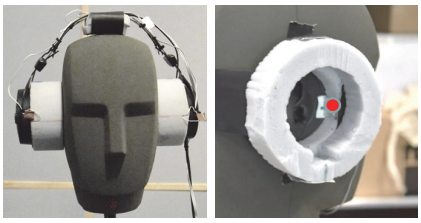
Supported by previous dummy head measurements, we claim that the actual position of the additional speakers as well as the enclosure formed by the earcups of the headphone have a strong effect on the quality of the resemblance of the spectral pinna-cues. Therefore, in addition to a first prototype, c.f. fig. 1(a), which is a slight modification of an existing product, a second prototype in a very preliminary stage, c.f. fig. 1(b), is examined in a listening experiment.

A listening experiment was conducted in which these prototypes are compared to existing binaural reproduction approaches in terms of similarity of the perceived source image to a real loudspeaker, which is located in an acoustically damped room. During this experiment, the subjects used a chin rest to ensure a well-defined distance and orientation of the reference speaker, since already slight shifts of the head position or minor head-rotations yield audible differences, c.f. [9].

In order to benchmark the results of the experiments, a commercial software[1] allowing for dry binaural reproduction was used. In brief, this benchmark, providing proprietary head-related transfer functions (HRTFs), should proof that the undertaken measurements and signal processing steps will deliver comparable or better results.

[1] Nx – Virtual Mix Room over Headphones plugin, Waves Audio Ltd.

(a) prototype A



(b) prototype B

Figure 1: Headphone prototypes, position of additional tiny loudspeakers indicated by the red dot.

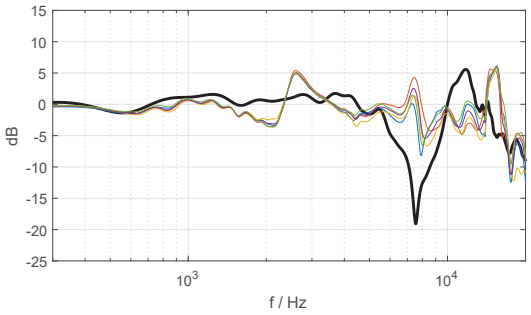
Prototypes

Figure 1 shows the examined prototypes, which we will refer to as *prototype A* and *prototype B*, hereafter. Both prototypes are based on an open circumaural headphone with an electrodynamic transducer, which is commercially available, and have in each earcup an additional tiny loudspeaker, positioned in front of the ear.

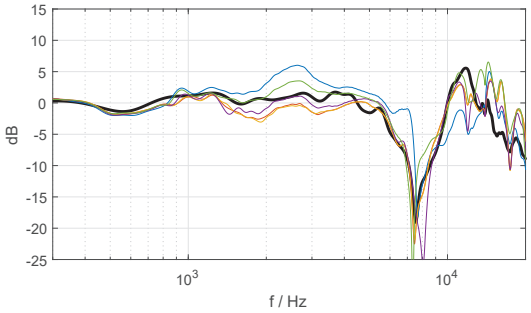
For prototype A, this additional speaker is mounted on the plate which carries also the conventional transducer. The position of the additional speaker is indicated by the red dot in fig. 1(a). Hence, only minor modifications of the original hardware are necessary. For prototype B the original ear pads have been replaced by a foam ring of 5cm thickness, which carries the additional speaker, see fig. 1(b).

Signal Processing

The binaural signal processing applied to the prototypes is designed to reproduce sources from the frontal direction. Therefore, prototypes are operated as a 2-way system. Frequencies below 1kHz are filtered with a HRTF from the frontal direction and are played back on the original headphone drivers. The employed HRTFs were measured

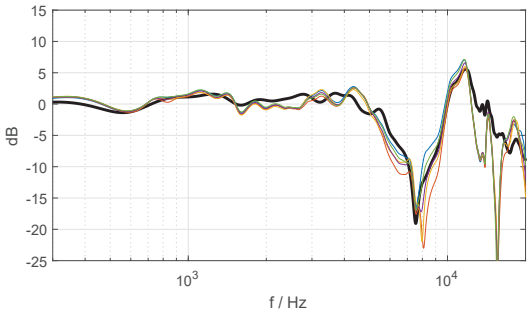


(a) prototype A

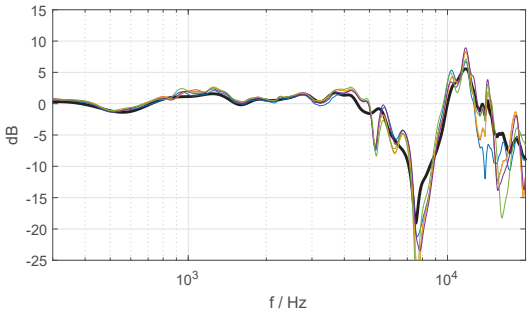


(b) prototype B

Figure 2: Magnitude response on the right ear of the dummy head for playback with the 2-way system, for prototype A and prototype B. The thin colored lines correspond to 5 repeated measurements on the dummy head, the black line shows the originally measured HRTF.



(a) prototype A



(b) prototype B

Figure 3: Magnitude response of the original headphone driver on the right ear of the dummy head, convolved with its equalization filter and the measured HRTF for the frontal position, for prototype A and prototype B. The thin colored lines correspond to 5 repeated measurements on the dummy head, the thick black line shows the originally measured HRTF.

Figure 4: Arrangement for measuring the frontal HRTFs of the dummy head.

in-situ with a dummy head, see below. Frequencies above 1kHz are played back on the additional tiny loudspeakers, directly, assuming that they resemble the HRTF from the front due to their relative position to the listener pinna.

The crossover-frequency is due to the lower frequency limit of the tiny speakers. The audio crossover is implemented by a low-pass and a high-pass filter of 4^{th} order. Furthermore, the signals of original headphone drivers and the tiny speakers are time aligned based on dummy head measurements.

The frequency characteristics of all drivers in both prototypes are equalized to a flat frequency response. Therefore, the original drivers were measured on a dummy head[2] and the additional tiny speakers were measured on a rigid sphere with two built-in microphones[3], which was used to resemble a head without pinna. All drivers were measured (repeatedly) 5 times with repositioning the headphone on the dummy head, respectively the sphere microphone, to account for the varying headphone position on a listener's head. Based on these measurements an average minimum-phase equalization filter was calculated for each transducer.

The HRTFs for the frontal position were determined by measuring binaural room impulse responses (BRIRs) with the dummy head in the same room where the listening experiment took place. Thereby, the dummy head was on the same position relative to the reference loudspeaker on which the head of the listener was located later in the listening experiment, see fig. 4. The first 118 samples ($@f_s = 44.1$kHz) within the measured BRIRs, corresponding to 2.7ms, are the HRTFs, i.e. the direct path, whereas the rest of the impulse response is considered as binaural reverberation.

Figure 2 shows the magnitude response on the right ear of the dummy head for playback with the 2-way system for prototype A and B, respectively. The thin colored lines correspond to 5 repeated measurements, whereby the headphone was repositioned on the dummy head for each repetition. The thick black line shows the magnitude response of the measured HRTF for comparison. Similarly, fig. 3 shows the magnitude response of the original headphone driver convolved with the measured HRTF and its equalization filter.

[2]KU100, Georg Neumann GmbH
[3]KFM 6, Schoeps GmbH

label	condition
① LS	playback on loudspeaker (*hidden reference*)
② mono	mono playback on headphone (*anchor*)
③ waves	encoded with Waves-Nx plugin without reverb (*benchmark*)
④ $\mathrm{HRTF_{dry}}$	HRTF
⑤ $\mathrm{HRTF_{BR}}$	HRTF + binaural reverb
⑥ $\mathrm{HRTF_{BR+T}}$	HRTF + binaural reverb + emulated torso reflection
⑦ $\mathrm{2\text{-}way_{dry}}$	2-way-system
⑧ $\mathrm{2\text{-}way_{BR}}$	2-way-system + binaural reverb
⑨ $\mathrm{2\text{-}way_{BR+T}}$	2-way-system + binaural reverb + emulated torso reflection

Table 1: Conditions of the experiment and related labels.

Experimental Setup

The perceived spatial image of a frontal source evoked by binaural playback on the headphone prototypes was examined in a listening experiment. This listening experiment took place in an acoustically optimized measurement room, lined with sound-absorbing material and non-parallel walls and ceiling. Its approximate dimensions are $5\mathrm{m}\times4\mathrm{m}\times3\mathrm{m}$ and its reverberation time is <50ms above 300Hz and <30ms above 1kHz.

The rating was conducted in a MUSHRA-like test paradigm. The participants had been instructed to rate the spatial aspects of the binaurally reproduced source compared to the reference (real loudspeaker) on a scale from "equal" to "strongly different", considering only the spatial aspects of the perceived source, i.e. its distance, direction and source width, but not any possible spectral coloration or level differences. The audio material used in the listening experiment was a male speech sample, which was high-pass filtered at 300Hz to avoid exciting the room at frequencies with longer reverberation time.

A total number of 9 different conditions, including hidden reference and anchor, were rated in the experiment. All conditions and the corresponding labels, as used in the diagrams and tables in the result section, are listed in table 1. A detailed description is given in the following paragraphs.

As reference condition ①, playback on a real loudspeaker at a distance of 1m in front of the listener was used. To assure a fixed position relative to the reference loudspeaker, the participants had to place their heads on a chin rest, as shown for the dummy head in fig. 4. Conditions ② to ⑥ use the equalized original headphone transducer only. Mono playback was used as anchor condition ②. In the benchmark condition ③, a frontal source without reverb is encoded with the commercially available plugin, and condition ④ uses the dummy head HRTFs measured at the listeners position. Condition ⑦ corresponds to the 2-way system using the additional tiny speakers, as described in the previous section.

In the initial design of the experiment, the loudspeaker playback was considered as reference for a "dry" frontal source. However, it showed, that even the small amount of reverberation in the measurement room significantly supports the perceived externalization. Therefore, further conditions were considered applying additional reverberation comparable to that of the reference loudspeaker in the reproduction room. Conditions ⑤, ⑧ are equal to conditions ④, ⑦, but in addition to the direct path, the binaural reverberation measured with the dummy head at the position of the listener is played back by the original headphone driver. Conditions ⑥, ⑨ are equal to conditions ⑤, ⑧, but with an and additional emulated torso reflection. The emulated torso reflection was achieved by a 0.57ms-delayed version of the direct path, which was attenuated by -6dB and low-pass filtered at 4kHz.

In total, the experiment consisted of 2 training parts and 3 rating parts. During the training parts, the participants were able to switch between the different conditions played back on prototype A and B. To facilitate comparison, the listeners were instructed to keep on the respective prototype during the whole training part, also when listening to the reference loudspeaker. Therefore, headphone reproduction was equalized to the reference by a filter accounting for the magnitude response of the reference loudspeaker and the damping by the worn headphone. The headphone damping was measured with the dummy head wearing the respective prototype.

The rating parts 1 and 2 were identical to the training parts 1 and 2, except that the listener was here asked to rate the conditions against the reference. The order of prototypes was random for each participant.

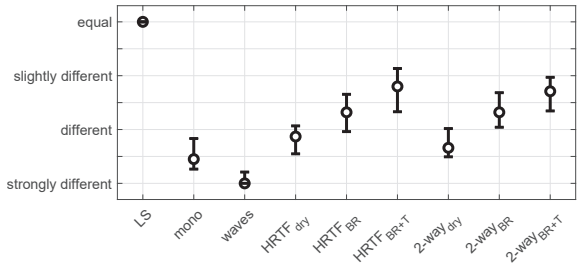
In rating part 3, conditions ①, ⑥, and ⑨ for each prototype were rated simultaneously against the reference, yielding a total number of 6 conditions. Within this part the listeners had to wear either prototype A or B, depending on which condition was played back. To this end, the prototypes were marked by 2 different colors, and the playback buttons in the user interface appeared in the corresponding colors. The listeners were instructed to wear no headphone while listening to the explicit reference. All conditions were equalized to the hidden reference for prototype A, i.e. the reference plays while wearing prototype A, as prototype A showed a higher damping of the reference.

Results

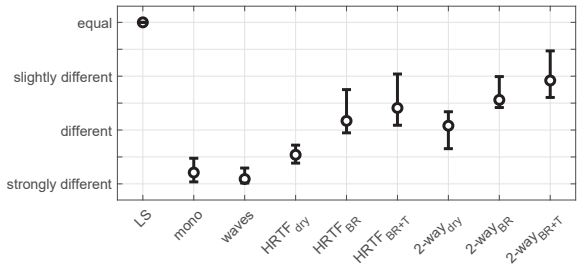
17 experienced subjects participated the experiment. On average, the test duration was 20min. Due to the training, the order of prototypes had no influence on the subjective ratings. The ratings obtained from the MUSHRA-test are depicted in fig. 5(a) to (c), showing the median values (small circles) and the confidence intervals thereof (bars).

The statistical differentiation between the tested conditions was evaluated by the Wilcoxon signed-rank test. In order to account for the multiple comparison problem, the Holm–Bonferroni method was applied. Instead of correcting significance level, the resulting p-values were adjusted, accordingly, c.f. [10]. These values are listed in table 2(a) to (c), whereby values below the significance level of 0.05 are printed bold.

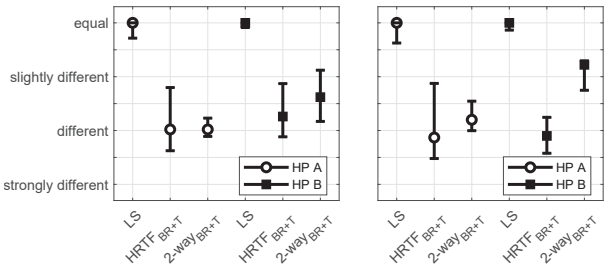
Examining the results of part 2 and 3, ignoring hidden reference, anchor and commercial benchmark, cluster analyses exhibit one group of correlated rating behavior with 11 subjects. From the 6 remaining subjects, two show different preferences in part 2 and 3, and four show deviating rating behavior compared to all other subjects



(a) part 1, prototype A



(b) part 2, prototype B



(c) part 3 **(d)** part 3

Figure 5: Median and confidence intervals of the results, considering all subjects (a) to (c) and (d) for part 3, considering the group homogeneous subjects, only.

		mono	waves	HRTF			2-way		
				dry	BR	BR+T	dry	BR	BR+T
	LS	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**
	mono	-	**0.01**	0.05	**0.01**	**0.01**	1.00	**0.02**	**0.01**
	waves	-	-	**0.01**	**0.01**	**0.01**	**0.03**	**0.01**	**0.01**
HRTF	dry	-	-	-	**0.03**	**0.04**	1.00	0.35	**0.04**
HRTF	BR	-	-	-	-	1.00	0.10	1.00	1.00
HRTF	BR+T	-	-	-	-	-	**0.04**	1.00	1.00
2-way	dry	-	-	-	-	-	-	**0.02**	**0.01**
2-way	BR	-	-	-	-	-	-	-	0.13

(a) part 1, prototype A

		mono	waves	HRTF			2-way		
				dry	BR	BR+T	dry	BR	BR+T
	LS	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**	**0.00**
	mono	-	0.79	0.26	**0.01**	**0.01**	0.05	**0.01**	**0.01**
	waves	-	-	0.05	**0.03**	**0.01**	0.09	0.05	**0.01**
HRTF	dry	-	-	-	**0.01**	**0.01**	0.64	0.05	**0.01**
HRTF	BR	-	-	-	-	0.74	0.77	0.77	0.11
HRTF	BR+T	-	-	-	-	-	0.27	0.79	0.43
2-way	dry	-	-	-	-	-	-	**0.03**	**0.01**
2-way	BR	-	-	-	-	-	-	-	0.26

(b) part 2, prototype B

		HP-A		*HP-B*		
		HRTF BR+T	2-way BR+T	LS	HRTF BR+T	2-way BR+T
HP-A	LS	**0.00**	**0.00**	0.74	**0.00**	**0.00**
HP-A	$HRTF_{BR+T}$	-	1.00	**0.00**	0.94	0.74
HP-A	2-way_{BR+T}	-	-	**0.00**	1.00	**0.01**
HP-B	LS	-	-	-	**0.00**	**0.00**
HP-B	$HRTF_{BR+T}$	-	-	-	-	1.00

(c) part 3

Table 2: p-values for part 1 to 3, considering all subjects.

		HP-A		*HP-B*		
		HRTF BR+T	2-way BR+T	LS	HRTF BR+T	2-way BR+T
HP-A	LS	**0.01**	**0.01**	0.75	**0.01**	**0.01**
HP-A	$HRTF_{BR+T}$	-	1.00	**0.01**	1.00	**0.02**
HP-A	2-way_{BR+T}	-	-	**0.01**	0.11	**0.01**
HP-B	LS	-	-	-	**0.01**	**0.01**
HP-B	$HRTF_{BR+T}$	-	-	-	-	**0.01**

Table 3: p-values for rating part 3, considering the group of homogeneous subjects, only.

in both cases. Figure 5(d) shows the rating results for the homogeneous subjects and the according corrected p-values are listed in table 3.

Conclusion

The evaluation shows that the HRTF playback on both prototypes results in comparable ratings, which are indistinguishable. Considering only ratings from the group of subjects who rated consistently and homogeneously, then the 2-way system on prototype B resembles the reference loudspeaker significantly better than the HRTF playback. Although the 2-way system achieves better results on prototype B than on prototype A, dummy head measurements have shown that the frequency response of the 2-way system on prototype B is sensitive to slight variations of the headphone position. This is presumably a reason for the inconsistent and deviating rating behavior of some of the subjects.

References

[1] N. I. Durlach *et al.*, "On the externalization of auditory images", *Pres.: Teleop. & Virt. Env.*, vol. 1, no. 2, 1992.

[2] W. M. Hartmann and A. Wittenberg, "On the externalization of sound images", *JASA*, vol. 99, no. 6, 1996.

[3] P. Zahorik *et al.*, "Auditory distance perception in humans: A summary of past and present research", *ACTA Acustica united with Acustica*, vol. 91, no. 3, 2005.

[4] H.-Y. Kim *et al.*, "Control of auditory distance perception based on the auditory parallax model", *Applied Acoustics*, vol. 62, no. 3, 2001.

[5] S.-M. Kim and W. Choi, "On the externalization of virtual sound images in headphone reproduction: A wiener filter approach", *JASA*, vol. 117, no. 6, 2005.

[6] H. G. Hassager *et al.*, "The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment", *JASA*, vol. 139, no. 5, 2016.

[7] K. Sunder *et al.*, "Individualization of binaural synthesis using frontal projection headphones", *J. Audio Eng. Soc.*, vol. 61, no. 12, 2013.

[8] H. Pomberger *et al.*, "Improved localization in the median plane with cue-preserving headphone", *Fortschritte der Akustik - DAGA*, 2018.

[9] F. Brinkmann *et al.*, "On the authenticity of individual dynamic binaural synthesis", *JASA*, vol. 142, no. 4, 2017.

[10] M. Aickin and H. Gensler, "Adjusting for multiple testing when reporting research results: The bonferroni vs holm methods.", *Am. J. Public Health*, vol. 86, no. 5, 1996.