Improved localization in the median plane with cue-preserving headphones

Hannes Pomberger[1], Alois Sontacchi[1], Matthias Frank[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, sontacchi@iem.at, frank@iem.at, thomas.gmeiner@usound.com, michele.lucchi@usound.com

Introduction

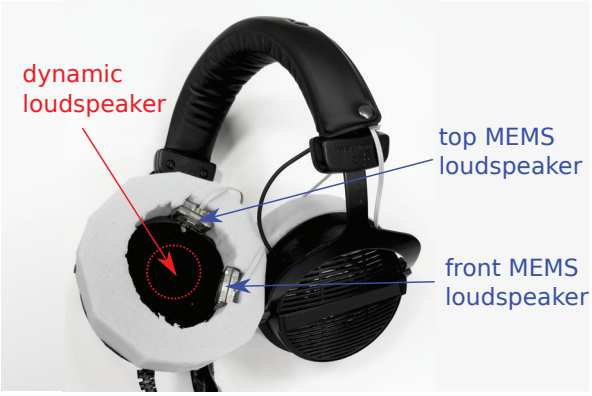
In a real-world scenario, sound sources from directions in the median plane cause neither interaural time nor level differences. The only perceivable directional cue is the spectral coloration, which is caused by the interaction of the direct acoustical propagation path with various reflected and diffracted paths at pinna, head, and torso. In [1] the influence of torso reflections on head related impulse responses (HRIR) is examined. There, it is shown that only for a small number of source directions the influence of the torso will exhibit an energetic notable impact and hence cause perceptible spectral colorations. Further, from a geometrical consideration these interactions are mainly relevant in the frequency range between 1 and 2 kHz. Therefore, it might be obvious that spectral colorations due to different source directions will be mainly caused by the pinna of the listener (complementary see also [2]). However, this coloration is highly individual and depends on the physiognomy, i.e. the actual size and shape of a listener's ear.

Binaural reproduction of sound sources located in the median plane with non-individualized head-related transfer functions (HRTFs) are rarely perceived from the intended direction, cf. [3], [4]. This phenomenon still remains when integrating dynamical cues although front-back confusions are basically reduced, cf. [5]–[7]. Typically, the usage of individual HRTFs improves the conformance of perceived and the intended sound source direction [8]. However, individual HRTFs for every user have to be determined in advance, either by time-consuming measurements or by complex models, based on anthropometric data, which are still limited in accuracy. Therefore, Sunder et al. [9] presented a sound source projection approach, activating individual directional cues by discrete loudspeaker placements.

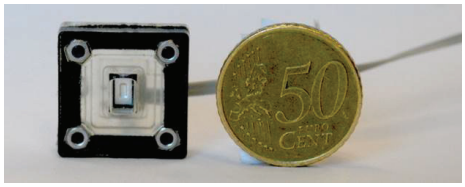
This proposal forms the concept of the individual cue-preserving headphone using two additional tiny loudspeakers distributed within each ear cup. This work examines the localization improvement by activating individual directional cues with these tiny loudspeakers in an informal listening experiment. The following sections describes the hardware setup of the prototype, the used signal processing and the experimental setup, as well as the results of the listening experiment.

Prototype

The investigated prototype is based on an open circumaural headphone with an electrodynamic transducer, which



(a) Prototype



(b) MEMS loudspeaker

Figure 1: Prototype headphone: modified conventional open circumaural headphone with two additional MEMS loudspeakers at each side.

is commercially available. This headphone has been extended by placing two additional tiny loudspeaker in front and above the ear at each ear cup, in the closest possible sagittal plane. The employed tiny loudspeakers are based on micro-electromechanical systems (MEMS) technology. Figure 1 shows (a) picture of the prototype headphone and (b) a picture of the used MEMS loudspeaker.

Our concept aims on profiting from both, the conventional dynamic loudspeakers to provide low frequencies and the MEMS loudspeakers to activate individual directional cues at frequencies related to Blauert's directional bands as shown in fig. 2, cf. [10, Fig. 8]. Due to the technical capabilities of the MEMS loudspeakers, their operating range has been set to a frequency range of 1.6 to 10 kHz, agreeing quite well with the pinna-related spectral cue regions. Whereas the signal components below 1.6kHz and above 10kHz are played back by the dynamic loudspeakers. The schematic operating ranges are shown in fig. 3.

To account for the frequency characteristics of the MEMS loudspeakers, their transfer functions was measured under free field conditions in a distance of 3cm using a professional measuring chain[1]. Based on this measure-

[1] G.R.A.S. 46AE 1/2" Free-field Set (cf. `https://www.gras.`

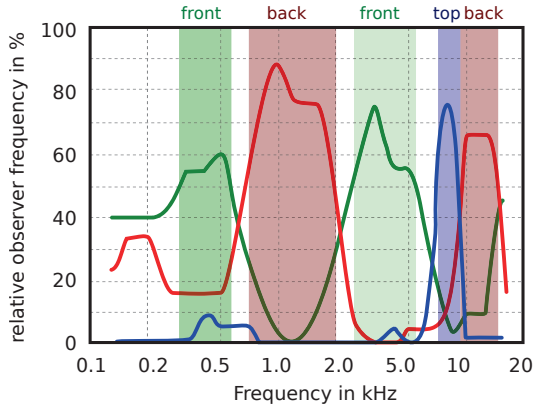


Figure 2: Directional bands according to Blauert.

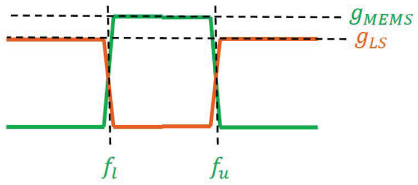


Figure 3: Schematic operation range for conventional dynamic loudspeaker (gain factor over frequency g_{LS}) and MEMS loudspeaker (gain factor over frequency g_{MEMS}).

ments, and equalization filter is realized by a minimum phase filter with perception-related smoothing using a Kautz-filter-design as described in [11]. The measured magnitude of the MEMS loudspeaker, as well as the corresponding equalization filter, and their conjunction are shown in fig. 4.

The prototype is intended to reproduce sources from in front ("front"), above ("top") and behind ("back") the listener. Figure 5 illustrates the applied signal processing to achieve these 3 directions. The source signal, lowered by 6dB within 1.6 to 10kHz by a "bathtub"-shaped filter, is convolved with a conventional anechoic HRTF from the

dk/46ae.html) amplified with Bruel & Kjaer Nexus 2690 (cf. https://www.bksv.com/en/products/transducers/conditioning/microphone/2690-A-0F2) Microphone Conditioner

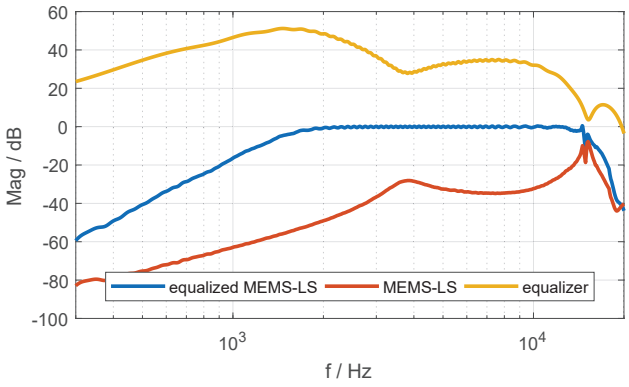


Figure 4: Magnitude response of a MEMS loudspeaker (orange), the used equalization filter (yellow) and the equalized MEMS loudspeaker (blue).

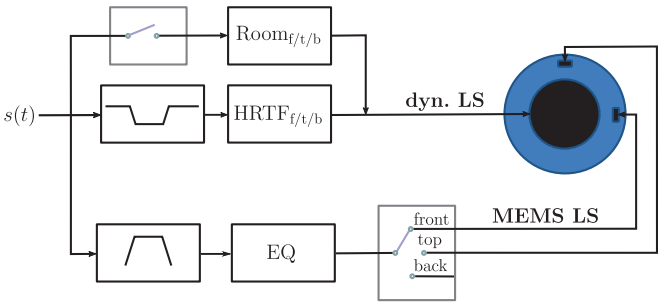


Figure 5: Block diagram of the signal processing for the prototype headphone.

intended source direction and played back by the conventional dynamic speaker of the prototype headphone. The used HRTFs have been taken from an database measured with a dummy head, cf. [12]. For the directions "front" and "top", within 1.6 to 10kHz the source signal is played back by the equalized MEMS loudspeaker corresponding to the intended source direction to provide individual directional cues to the listener. The direction "back" is realized similar to the approach presented in [13] simply not activating any of the MEMS speakers, what results in a level reduction by 6dB from 1.6 to 10 kHz in the HRTF due to the "bathtub"-filter.

Furthermore, additional room information is optionally played back by the dynamic speaker to support the impression of externalization. This room information is generated by convolving the source signal with the non-direct part of a binaural room impulse response corresponding to the intended source direction.

Experimental Setup

In an informal listening experiment the specific hardware setup and signal processing has been examined. The listening experiment was split into two parts. Within the first part externalization and sound coloration and within the second part the localization were examined. Nine (male) experienced listeners participated the experiment and executed both parts in about 1 hour on average. In both parts three different audio materials (male speech, a thin instrumented jazz piece, and an extended orchestral-instrumented film music) have been presented for the three different intended directions ("front", "top", "back").

In part one, the subjects rated externalization and sound coloration in an absolute magnitude estimation. For both attributes a set of seven signal conditions has been evaluated, which are listed in table 1. Mono playback over the dynamic loudspeaker constitutes the lower anchor (condition 1). Condition 2 represents a standard binaural playback using non-individual HRTFs without any room information via the dynamic loudspeakers, only. Condition 3 represents the reproduction with the additional MEMS loudspeakers as described in the previous section and illustrated in fig. 5 without any room information. In order to examine the increase of perceived externalization, conditions 4 to 7 are formed by the settings in condition 2 respectively 3 plus additional room information repro-

condition	label
1. mono playback via dynamic LS - *anchor*	MONO
2. binaural playback via dynamic LS using anechoic HRTFs	HRTF
3. equal to 2. but with "bathtub"–filter plus MEMS LS from the intended direction	MEMS
4. equal to 2. but with minor room information	$\mathrm{HRTF_{sr}}$
5. equal to 3. but with minor room information	$\mathrm{MEMS_{sr}}$
6. equal to 2. but with enriched room information	$\mathrm{HRTF_{mr}}$
7. equal to 3. but with enriched room information	$\mathrm{MEMS_{mr}}$

Table 1: Experimental conditions of the 1st part of the listening experiment.

condition	label
1. binaural playback via dynamic LS using anechoic HRTFs	HRTF
2. equal to 1. but with "bathtub"–filter plus MEMS LS from the intended direction	MEMS
3. equal to 2. but with enriched room information	$\mathrm{MEMS_{mr}}$

Table 2: Experimental conditions of the 2nd part of the listening experiment.

duced by the dynamic loudspeakers. Thereby the room information for conditions 4 and 5 are obtained from a simple synthetic room model (shoe box) including first order reflections only ("minor room information"), whereas for conditions 6 and 7 binaural room impulse responses measured in a studio room were applied ("enriched room information").

Evaluating the attribute externalization, subjects were asked to assess the degree of perceived externalization on a continuous scale within the end marks "im Kopf" (translated "inside the head") at the left side and "außer Kopf" (translated "out of head") at the right side, respectively. Evaluating the attribute sound coloration, subjects were asked to score the perceived coloration on a continuous scale with the given markers "sehr stark" (translated "most intensive"), "mittel" (trans. "medium"), and "natürlich" (trans. "natural") at the left side, in the middle, and at the right side, respectively.

In the second part of the listening experiments the localization of the presented stimuli has been examined. As before, only sound sources in the median plane, i.e. at frontal, top and back position, have been evaluated. Subjects were asked to identify and judge the perceived direction on a continuous scale with the given three markers indicating front, top, and back position, respectively.

For each direction a set of three signal conditions has been evaluated, which are listed in table 2. Condition 1 in part 2 is equal to condition 2 in part 1 and represents a standard binaural playback using non-individual HRTFs without any room information. Reproduction with the additional MEMS loudspeakers according to fig. 5 without any room information forms condition 2, which is equal to condition 3 in part 1. Condition 3 is formed by the setting in condition 2 plus the measured studio room impulse responses played back by the dynamic loudspeakers.

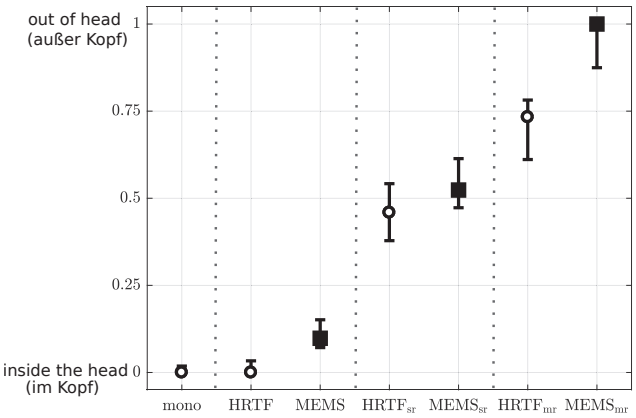


Figure 6: Perceived externalization (all subjects, all stimuli, all directions aggregated).

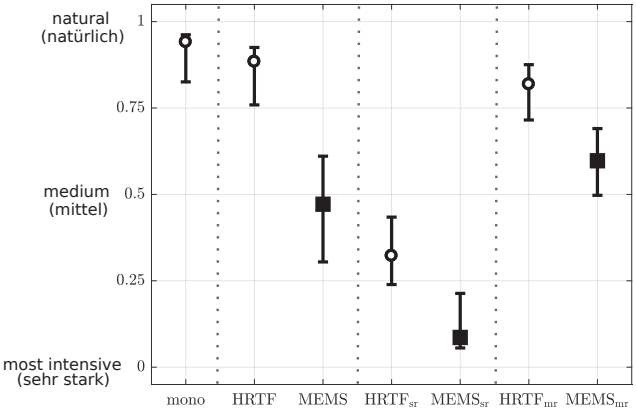


Figure 7: Rated sound coloration (all subjects, all stimuli, all directions aggregated).

Results

In the following the listing experiment results of the 9 subjects are summarized. Figures 6 to 8 show the condensed data without differentiating the audio material further, the examined directions in case of the rated externalization and coloration are aggregated, too.

Therefore, each plot represents 81 responses (9 subjects, 3 signal contents, and 3 sound source directions) in figs. 6 and 7. In fig. 8, summarizing the judgments according to the perceived sound source directions 27 responses (9 subjects and 3 signal contents) are forming the database for each plot.

The used evaluation representation is deduced from the notched box-and-whisker diagram and provides a compact data location and data distribution information. The resulting median response of the examined judgments is depicted by circle and square markers. Moreover, the 95-percent confidence interval (CI) of the median is indicated by vertical lines, which is typically provided by the notch around the median line in a box-and-whisker diagram.

Therefore, the statistically discrimination of the examined conditions can be graphically deduced. If the 95-percent confidence intervals of two compared conditions do not overlap then their related median values will be different at a significance level of lower than 5 percent.

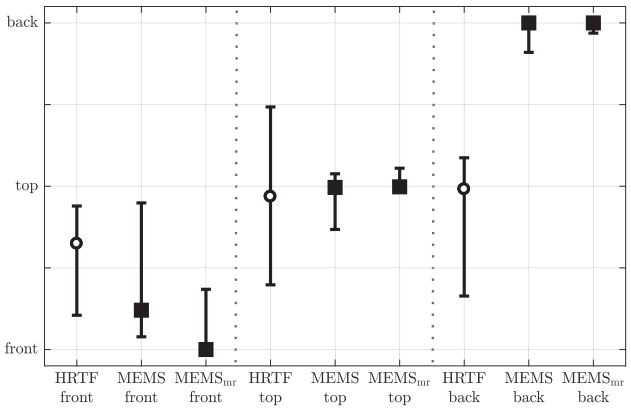


Figure 8: Identified direction for frontal, top and back position (all subjects, all stimuli aggregated).

Caused by the fact that the received rating data do not agree with the underlying hypothesis of normal distributed data, the subsequent presented statistical results are obtained using the Kruskal-Wallis H test and the Mann–Whitney U test, respectively. Observed significant differences are qualified with the calculated p-value.

Externalization, see fig. 6: All neighboring conditions yield significantly different externalization ($p < 0.031$), except for mono/HRTF ($p = 0.27$) and HRTF$_{sr}$/MEMS$_{sr}$ ($p = 0.17$). The measured room yields best externalization results, followed by the simple room. Compared to pure HRTF rendering, the additional MEMS loudspeakers improve externalization in all room conditions, except for the simple room.

Coloration, see fig. 7: There is no significant difference between the coloration of mono and HRTF ($p = 0.31$). HRTF playback with the measured room does not significantly differ from dry HRTF playback ($p = 0.22$), but is significantly worse than mono ($p = 0.013$). In all room conditions, the MEMS are significantly worse compared to the respective HRTF playback ($p < 0.0125$). In comparison to the measured room, coloration is worse in the simple room ($p << 0.001$).

Localization, see fig. 8: There is no significant difference in between the frontal directions ($p > 0.081$) and the top directions ($p > 0.15$), however for the back direction, the two MEMS conditions are not distinguishable ($p = 0.31$) but both significantly different to the HRTF ($p << 0.001$). For the HRFTs the three directions are not distinguishable ($p > 0.34$), whereas the back direction of both MEMS with and without room yield significantly different perceived directions than all other directions ($p << 0.001$). The dry MEMS from the front was perceived elevated and is thus not distinguishable from the top direction ($p = 0.1723$). This does not hold for the MEMS with the measured room ($p << 0.001$).

Conclusion

Distributed loudspeakers, according to the proposal made, have a (great) impact and potential to support the activation of individual cues without elaborated signal preprocessing. The externalization of critical sound source positions in the median plane improve significantly compared to standard binaural pre-processing with HRTFs (at least at the examined directions front and top). Moreover, the localization of the spatial arranged sound sources is (dramatically) enhanced. Furthermore, adding room information (from measured rooms) will boost the perceived externalization, reduce deviations in the sound coloration, and last but not least manifest further the localization accuracy. Sound coloration in case of using MEMS loudspeakers is an issue. However, in this case the direct comparison to the mono reproduction (obvious benchmark) might be misleading, as this will also deviate from a real sound source representation in space.

References

[1] M. Guldenschuh *et al.*, "HRTF modelling in due consideration variable torso reflections", in *Acoustics 08 Paris*, 2008.

[2] M. B. Gardner and R. S. Gardner, "Problem of localization in the median plane: Effect of pinnae cavity occlusion", *Journal of the Acoustical Society of America*, 1973.

[3] A. D. Musicant and R. A. Butler, "The influence of pinnae-based spectral cues on sound localization", *Journal of the Acoustical Society of America*, 1984.

[4] S. G. Weinrich, "Improved externalization and frontal perception of headphone signals", in *92nd AES Conv.*, 1992.

[5] H. Wallach, "On sound localization", *Journal of the Acoustical Society of America*, 1939.

[6] ——, "The role of head movements and vestibular and visual cues in sound localization.", *Journal of Experimental Psychology*, 1940.

[7] P. Mackensen *et al.*, "Einfluß der spontanen Kopfdrehungen auf die Lokalisation beim binauralen Hören", in *Bericht 20. Tonmeistertagung*, 1998.

[8] H. Møller *et al.*, "Binaural technique: Do we need individual recordings?", *Journal of the AES*, 1996.

[9] K. Sunder *et al.*, "Individualization of binaural synthesis using frontal projection headphones", *Journal of the AES*, 2013.

[10] J. Blauert, "Sound localization in the median plane", *Acta Acustica united with Acustica*, 1969.

[11] M. Karjalainen and T. Paatero, "Equalization of audio systems using kautz filters with log-like frequency resolution", in *120th AES Conv.*, 2006.

[12] B. Bernschütz, "A spherical far field HRIR/HRTF compilation of the Neumann KU 100", in *Fortschritte der Akustik, AIA-DAGA*, 2013.

[13] M. Frank and F. Zotter, "Simple reduction of front-back confusion in static binaural rendering", in *Fortschritte der Akustik, DAGA*, 2018.