Headphone Reproduction via Loudspeakers using Inverse HRTF-Filters

T. Weissgerber[1], K. Laumann[1,2], G. Theile[2], H. Fastl[1]

[1] AG Technische Akustik, MMK, TU München, Germany, Email: tobias.weissgerber@kgu.de

[2] Institut für Rundfunktechnik GmbH, München, Germany

Introduction

Binaural signals like dummy-head recordings usually can only sound correct by headphone reproduction. When audio signals are reproduced via loudspeakers, their sound fields overlap. Each ear of a listener receives signals from all loudspeakers. But with appropriate filter processing, loudspeakers can serve as a "*virtual headphone*" and headphones can simulate loudspeaker signals [1]. The *Binaural Sky* [2], an experimental system (pilot study) at the Institut für Rundfunktechnik, Munich, Germany, can serve as such a virtual headphone. The positions of the sources to generate the virtual headphone have been chosen as a result of some listening tests so far and without detailed measurements. Hence, in comprehensive studies the quality of a virtual headphone was evaluated for different speaker topologies in the horizontal plane.

Binaural Sky

The Binaural Sky is a circular loudspeaker array, mounted above the listener's head. 22 speakers are built in a wooden panel without enclosure and a single driver for low frequencies is placed in the centre. An electromagnetic motion tracking system continuously captures head position and orientation of the listener. Based on this data the individual input signals for all 23 array speakers are calculated in real-time. Focussed *wave field synthesis* sources around the listener's head serve as sources to realise a virtual headphone. Further technical details of the laboratory system are described in [2].

Virtual Headphone

Headphone reproduction via loudspeaker consists of two tasks: first, crosstalk has to be cancelled and, second, the naturally occurring head-related transfer functions (HRTFs) between source and ear has to be compensated. The transfer functions to the contralateral ears have to equal zero while the transfer functions for the direct transfer paths have to equal one. Bauck and Cooper generalized this topic in 1996, named by "Transaural Stereo" [3].

This principle can be formulated in matrix-vector notation. The speaker output signals y_{left} and y_{right} for the creation of virtual headphones can be calculated from the input signals x_{left} and x_{right} as described in equation 1.

$$\begin{pmatrix} y_{left} \\ y_{right} \end{pmatrix} = H^{-1} \begin{pmatrix} x_{left} \\ x_{right} \end{pmatrix} \tag{1}$$

The matrix H^{-1} is to be calculated as the inverse of matrix H containing all HRTFs (equation 2).

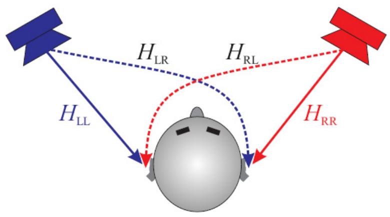

Figure 1: Transfer functions for loudspeaker reproduction

$$H = \begin{bmatrix} H_{LL} & H_{RL} \\ H_{LR} & H_{RR} \end{bmatrix} \tag{2}$$

This network H^{-1}, applied to the speaker system, is in literature often referred to as cross-talk cancellation network. Because the processing not only eliminates crosstalk but also compensates the HRTF of the direct path, the expression *inverse HRTF-Filter* will be used in this paper instead. In a static case these filters are only correct for one fixed head-orientation; if head movements are allowed, the filter network has to be updated with every head movement of the listener.

Measurement of HRTFs in the Near-Field

For the measurement of HRTFs in the near-field some additional considerations have to be taken. Brungart and Rabinowitz define the transition of the acoustic near-field into the far-field at a distance of one meter to the middle of the human head [4]. Inside this area, distance dependent changes in localisation cues occur. Especially the interaural level differences increase with decreasing distance between a sound source and the human head. Lentz describes noticeable changes in HRTF for distances less than 1.5 meters [5]. In this area it is necessary to measure HRTFs for every desired head distance instead of only measure one HRTF and adjusting the level for a certain distance.

Measurements were taken in the anechoic chamber of the Institut für Rundfunktechnik with a Philips PSS100 loudspeaker (diameter of 3.5 cm) in a plastic enclosure driven by a Terratec Sine PA 460 power amp as playback device and recorded with a Neumann KU100 dummy-head with diffuse-field equalization. The used measurement method was *Maximum-Length-Sequence* (MLS), which is robust against interfering signals. Furthermore the desired transfer functions are easy to retrieve out of the measurement. The MLS-Sequence was generated and recorded with 2048 frequency steps by the measurement device Audio Precision System One.

To have a database of near-field-HRTFs available for further investigations, measurements were taken in the horizontal

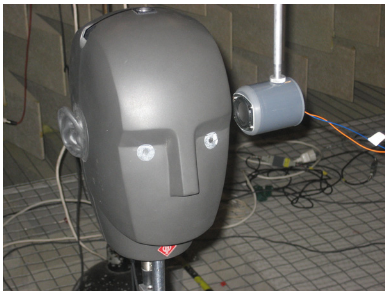

Figure 2: HRTF measurements in the near-field, here at a distance of 6cm between the loudspeaker membrane and the entrance to the ear canal.

Figure 3: Circular loudspeaker array with 8 speakers; the distance between the speaker membrane and the ear canal entrance measures 15 centimetres.

plane as well as with speaker elevation at distances between 6 cm and to 30 cm, (from the membrane of the loudspeaker and to the entrance of the ear canal of the dummy head), and elevation from 0 degree to 90 degree in a step size of 15 degrees. The resolution in azimuth was 10 degrees.

Experimental Setup

To realise the inverse HRTF-Filter playback, two circular wooden plates with two different diameters were built. Every 10 degree are drills to insert rods with a length of 50 cm to mount loudspeakers. The wooden plate is suspended above the listener's head and furnished with foam to reduce reflections. For the first array the distance between loudspeakers and the entrance of the human ear canal is 15 cm, for the second array a distance of 30 cm was chosen, emanated from a length of the interaural axis of 18 centimetres. One of the two speaker arrays is illustrated in Figure 3.

Evaluation

In preliminary measurements the advantage of the array with smaller diameter was proven, because of less crosstalk between the ears as a result of head shadowing and smaller influences of the playback room. Therefore further evaluation was only performed on the smaller loudspeaker array.

Eleven different loudspeaker setups in the horizontal plane at height of the listener's ears were tested. Setups with two, four, six and eight active loudspeakers were taken into account. The different loudspeaker setups are described in Table 1 and Figure 4. For all setups except *4-3* and *O* an inverse HRTF-Filter was calculated for playback. At setup *4-3* only the direct path of the speaker signal to the ipsilateral ear was compensated while setup *O* serving as a control setup without any filter processing. The loudspeakers are identical and the same as used for the near-field-HRTF measurement and driven by a Sonance SONAMP 1230 hifi amplifier with 12 channels. The frequency response of the playback speakers was flattened for every setup with respect to the listener position?

Measurements

The measurements took place in a laboratory with acoustic treatment at the Institut für Rundfunktechnik . Measurement method was MLS with 2048 frequency steps, generated and recorded with an AP System One via the Neumann KU100 dummy-head. The input signals for the virtual headphone were convolved in real-time with the inverse HRTF-Filters using a standard personal computer with Linux operating system and convolution software BruteFIR and played via RME Multiface II.

The transfer functions of the virtual left, virtual right and also both channels together, driven by the same signal, were measured and smoothed by a filter with a bandwidth of one third octave. In Figure 5 some results are exemplarily illustrated. The upper diagram shows the *crosstalk attenuation* between the virtual right channel at the right ear and the crosstalk at left ear. The attenuation is always greater than 10 dB and ranges up to more than 20 dB. This holds for all speaker setups except *4-3* and *O*, where crosstalk attenuation is worse. In the middle and lower diagram the transfer functions of virtual left and virtual right headphone channel are plotted for diotic playback with all speaker setups. The frequency response is almost flat up to 10 kHz for every speaker setup. Only the setups *O*, *4-3* and *8-2* (highlighted) have irregularities in frequency response. For all other setups the headphone reproduction should work without major signal coloration.

Setup identifier	Speaker positions (azimuth in degree)
2-1	90° - 270°
2-2	20° - 340°
4-1	0° - 90° - 180° - 270°
4-2	60° - 120° - 240° - 300°
6-1	0° - 60° - 120° - 180° - 240° - 300°
6-2	60° - 90° - 120° - 240° - 270° - 300°
6-3	20° - 70° - 110° - 160° - 200° - 250° - 290°
8-1	0° - 60° - 90° - 120° - 180° - 240° - 270° - 300°
8-2	20° - 70° - 160° - 200° - 250° - 290° - 340°
4-3	70° - 110° - 250° - 290°
O	90° - 270°

Table 1: List of evaluated loudspeaker array setups; 0 degree azimuth corresponds to the middle postion in front of the listener.

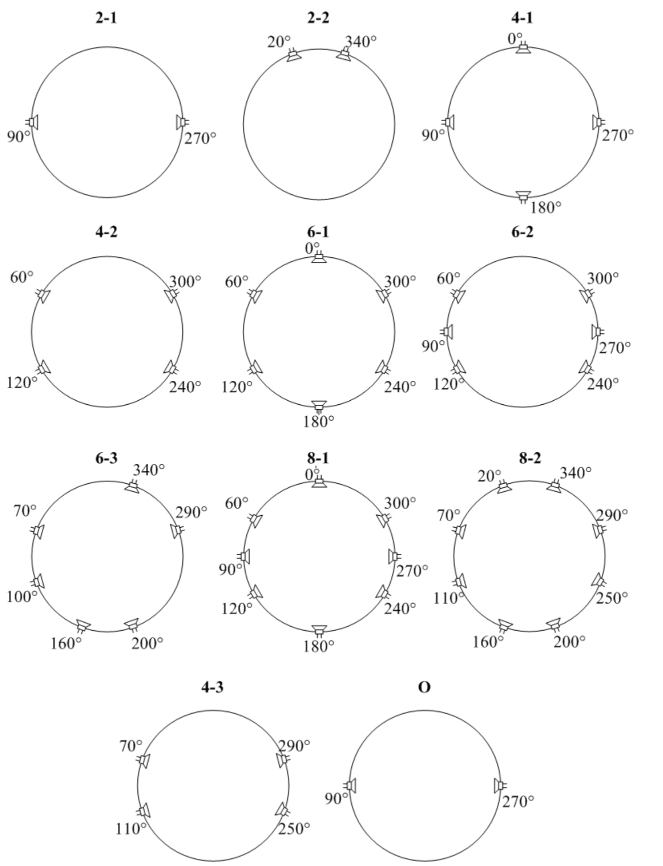

Figure 4: Evaluated loudspeaker setups

Psychoacoustic Experiments

Methods

To quantify the subjective perception of the headphone simulation via different setups, psychoacoustical experiments were necessary. 10 subjects with a mean age of 30.4 years participated in the hearing experiments, which were focussed on assessment of the position of sounds in the horizontal plane.

In a first experiment the subjects had to estimate the position of the presented stimuli on the inter-aural axis. If the inverse HRTF-Filter performs well for a certain position, the auditory event should take place inside the head. Such lateralisation experiments are often found in literature, cf [6,7]. The task in the second part was to determine the position of the sound source on the forward/backward axis through the middle of the listener's head. To quote the position of the stimuli, a category scaling from -25 to 25 was used, whereas the maxima label the position of the playback speakers. In a range from -10 to 10 the auditory event was perceived inside the head.

Stimuli

Two different stimuli were tested. Pulsed pink noise with a duration of 300 ms was presented 12 times in intervals of 300 ms and a sound pressure level of 65 dB(A). The duration of such a test sequence was about 10 seconds. As a second stimulus we choose an anechoic speech signal (male narrator) with 15 seconds duration, presented at a L_{eq} of 65 dB(A), measured over 30 seconds.

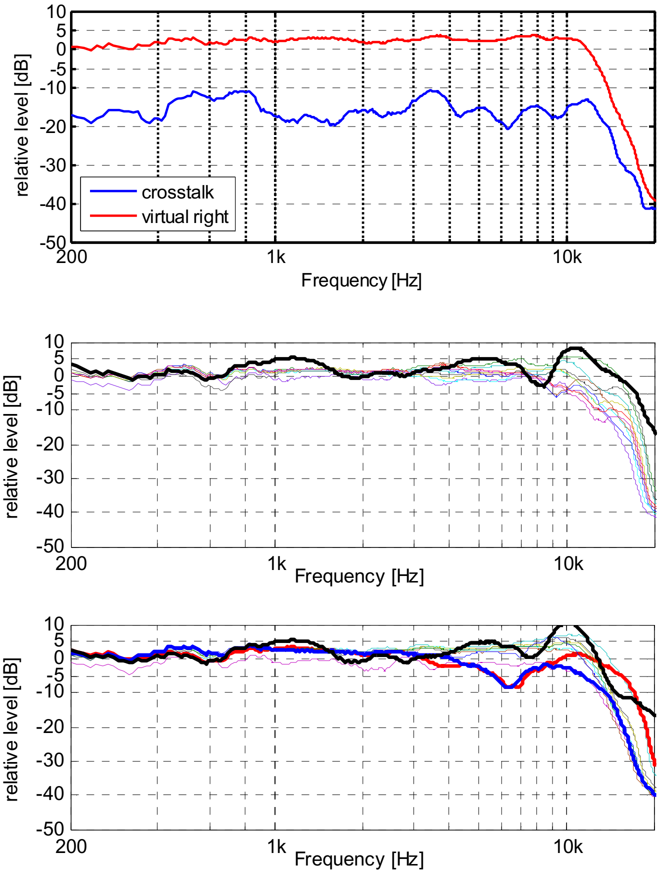

Figure 5: Measurements of the transfer function of the virtual right channel and crosstalk (upper) and of the virtual headphone channels for all speaker setups. The middle diagram depicts the virtual left channel, the lower diagram depicts the virtual right channel. All setups perform quite similarly, except the highlighted ones (black: O, blue: 4-3, red: 8-2).

Results

1. Lateralisation

The results of the lateralisation experiment are illustrated in Figure 6. For a diotic presentation on the virtual headphone the sound position was perceived close to the middle of the head for all loudspeaker setups (upper graph). Setups *4-1*, *6-1* and *8-2* yielded the best results for both types of stimuli with also a small interquartile range, even though not significant.

For an input signal on only the left virtual channel the auditory event should occur at the left ear (scale position -10). Only setup *4-3* and *O* are significantly worse than the others (kruskal-wallis test, $p<0.01$), what is obvious because of missing or only partial filter processing. All results are very similar for both stimuli.

2. Front/back localisation

Figure 7 depicts the assessment of the different speaker setups regarding localisation on the front/back axis. For a diotic input signal of the virtual headphone (upper graph) only setup *2-2* is significantly worse than all other ones ($p<0.01$). In this setup prominent front/back-irritations occur for the presented male speech signal while the pink noise

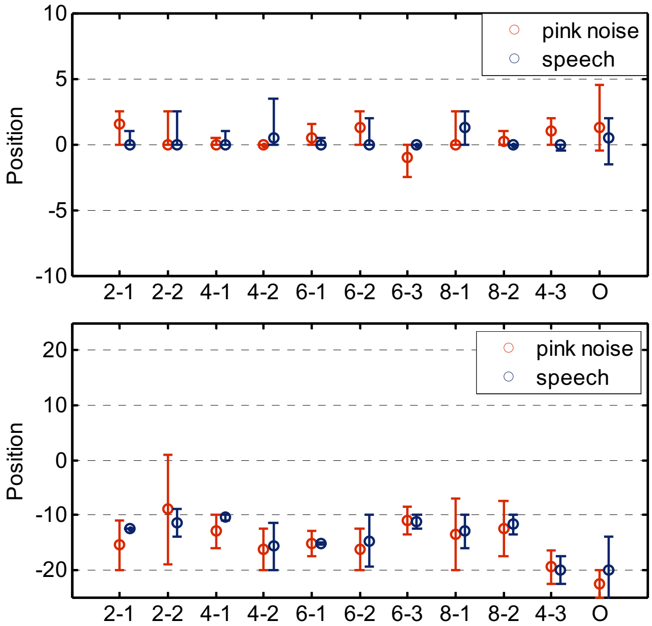

Figure 6: Results of the lateralisation experiment: Median and interquartile range for pink noise (red circles) and speech (blue circles), presented at both virtual headphone channels (upper diagram) and at left virtual channel (lower diagram). Position 0 corresponds to the middle of the head, ±10 to the ears and ±25 to the speaker positions on the ear axis.

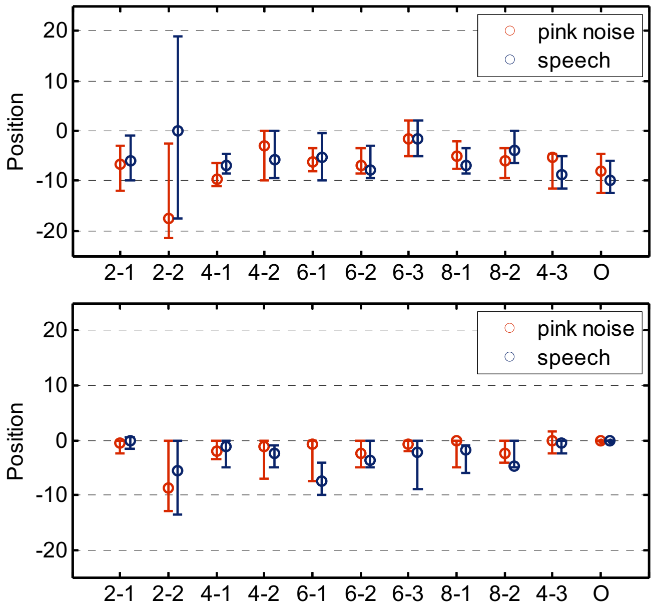

Figure 7: Results of the localisation experiment: Median and interquartile range for pink noise (red circles) and speech (blue circles), presented at both virtual headphone channels (upper diagram) and at left virtual channel (lower diagram). Position 0 corresponds to the middle of the head, ±10 to the head surface and ±25 to the speaker positions in front/back of the listener.

signal is often perceived outside the head. As in the lower graph of figure 7 depicted, only slight differences are perceived for an input signal on only the left virtual headphone channel.

Summary

The measurements as well as the psychoacoustic experiments revealed that for all speaker setups, except for the control setups *4-3* and *O*, from 2 up to 8 active loudspeakers the sensation of headphone listening could be achieved. The "nearly stereo setup" 2-2 with speaker positions at 20° and 340° often led to front/back-irritations while the other setup with only 2 loudspeakers on the interaural axis (90° and 270°) is very limited in stability for small head movements. For all other setups with more than 2 loudspeakers the physical and psychoacoustical quality of the virtual headphone seems to be very similar. Therefore, in practical usage, setups with 4 loudspeakers are completely sufficient to reproduce headphone signals via loudspeakers with static inverse HRTF-filters; a higher speaker amount only induces more processing power without significant quality improvements.

References

[1] G. Theile., "Zur Theorie der optimalen Wiedergabe von stereofonen Signalen über Lautsprecher und Kopfhörer“, *Rundfunktechn. Mitteilungen* Jahrg. 25, H. 4 (1981), S. 155–170

[2] D. Menzel, H. Wittek, G. Theile, H. Fastl, "The Binaural Sky: A Virtual Headphone for Binaural Room Synthesis", *Tonmeistersymposium*, Hohenkammer, Germany (2005)

[3] J. Bauck, D.H. Cooper, "Generalized Transaural Stereo and Applications", *J. Audio Eng. Soc*. 44, No. 9 (1996), S. 683–705

[4] D.S. Brungart, W.M. Rabinowitz, "Auditory Localization of Nearby Sources. Head-related transfer functions“, *J. Acoust. Soc. Am*. Volume 106, No. 3, Pt. 1 (1999)

[5] T. Lentz, "Near-field HRTFs", *Fortschritte der Akustik*, DAGA 2007, Dresden (2007).

[6] J. Blauert, "Spatial Hearing: The Psych ophysics of Human Sound Localization", MIT Press: Cambridge, MA (1997)

[7] H. Fastl, E. Zwicker, "Psychoacoutics – Facts and Models", 3rd ed., Springer: Berlin (2007)