A virtual headphone based on wave field synthesis

Klaus Laumann[a,b], Günther Theile[a] and Hugo Fastl[b]

[a]Institut für Rundfunktechnik GmbH, Floriansmühlstraße 60, 80939 München, Germany
[b]AG Technische Akustik, MMK, TU München, Arcisstr. 21, 80333 München, Germany
laumann@irt.de

The term "virtual headphone" refers to specially designed loudspeaker systems aiming for transmission characteristics equal to real headphones. Particularly of interest is the exact pre-filtering of the speaker signals to compensate the effect of head-related transfer functions (HRTFs) between loudspeakers and ear canal. These so-called "HRTF inverse filters" are dependent on geometrical conditions and so they have to be updated with every head movement.

In order to avoid problematic adaptive HRTF inverse filtering, the real loudspeakers are replaced by focussed sources generated according to the principles of Wave Field Synthesis (WFS). Head tracking controlled adjustment of driving functions allows easy source movement and thus fixed source positions in relation to the listener's ears, providing stable virtual headphone reproduction. A single static HRTF inverse filter network can be used. It is designed to ensure precise headphone equalization according to ITU-R BS.708 and offers accurate reproduction of e.g. binaural signals.

A pilot study with a circular WFS array built in a panel above the listener's head has verified the functional capability of this concept.

1   Introduction

Essentially, two electro-acoustical reproduction devices have to be distinguished: headphones and loudspeakers. Although the transducer principles are similar, their placement causes two different systems. For practical purposes it is sometimes necessary to transform each system into its counterpart, and suitable filters have been proposed in the past.

On a simplified example the general aspects will be introduced: When audio signals are reproduced on two (or more) loudspeakers, naturally their sound fields overlap. Each ear of a listener receives signals from both (all) loudspeakers. Assuming a simple system with two speakers and two ears in free-field conditions, 4 transfer functions can be specified. These transfer functions are generally known as *head-related transfer functions*, defined as follows:

> A head-related transfer function (HRTF) is a transfer function that, for a certain angle of incidence, describes the sound transmission from a free field (plane wave) to a point in the ear canal of a human subject.[1]

This means, a pair of (complex-valued) HRTFs defines the transfer path between one sound source and both eardrums, including diffraction, reflection, absorption, and shadowing of sound waves by pinna, head, and body causing frequency dependent filtering, as well as interaural time and level differences. All these are considered to be essential cues for sound source *localization*, i.e. determination of the sound source's position. For details on this topic refer to [2], an elaborate treatise on HRTFs and their measurement can be found in [1].

Headphones, in contrast, feature a high channel separation. In practice the direct transmission lines between ear cup and ear (H_{LL}, H_{RR}) will be characterized by a certain *headphone transfer function*, but to simplify matters, we suppose that they are equal to 1. As stated in [4], this holds true for headphones equalized to a diffuse field according to [5]. Transfer functions to the opposite ear (H_{LR}, H_{RL}), in terms of communications engineering commonly referred to as cross-talk, equal zero (see Figure 1).

1.1   Headphones → Loudspeakers

From a signal-theoretic point of view, the simulation of loudspeaker reproduction via headphones is quite sim-
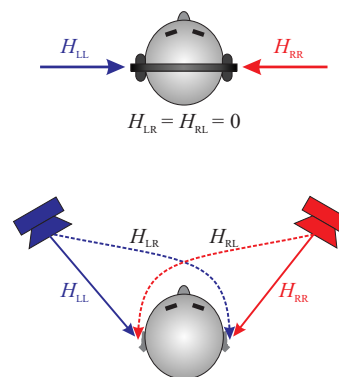

Figure 1: Transfer functions in case of headphone (upper part) and loudspeaker reproduction (lower part)

ple. The input signal of the speaker to be simulated goes into each headphone channel and is filtered with the corresponding HRTF. For stable room-related localization of this **virtual speaker** the HRTFs have to be readjusted when the listener turns or moves the head. Supposing linear, time-invariant systems, further sources can be added according to the superposition principle. Practical methods are summed up as *binaural technique* or *binaural technology*:

> The definition of *binaural technique* may be extended more generally to *binaural technology* as follows: Binaural technology is a body of methods that involves the acoustic input signals to both ears of the listener for achieving practical purposes, for example, by recording, analyzing, synthesizing, processing, presenting, and evaluating such signals.[3]

The compensation of headphone transfer functions by splitting up the transmission chain into a recording (e.g. dummy head) and a reproduction (e.g. headphones) part was already addressed by Blauert. A possible solution, equalization of the interfaces to a *diffuse field*, was proposed by Theile [4] and is also discussed in [2, 3] and specified for headphones in [5]. Successfully implemented systems based on binaural technology were reported, for example, by Technische Universität Berlin, Germany [6], or Institut für Rundfunktechnik, Munich, Germany, together with Studer Professional Audio AG, Zurich, Switzerland, and Ruhr-Universität, Bochum, Germany [7].

1.2 Loudspeakers → Headphones

The creation of headphone transmission characteristics on loudspeaker systems can be subdivided into two tasks: first, cross-talk has to be cancelled and, second, the naturally occurring HRTF between source and ear has to be compensated. This means, ideally, transfer functions to the opposite ears have to equal zero, and equal one for the direct transfer paths. Bauck and Cooper generalized this topic in 1996 and coined the term "Transaural Stereo" [8].

The problem is often formulated in matrix-vector notation. In our simplified case, 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 follows:

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

The matrix H^{-1} is to be calculated as the inverse of matrix H containing all HRTFs:

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

The network H^{-1} applied to the speaker system is also called cross-talk cancellation network. As this term is somewhat misleading, because the procedure involves not only cross-talk elimination but also HRTF compensation, the expression *HRTF inverse filter* — according to the mathematical concept of inverting the matrix H — will be used in this paper instead.

Numerous publications deal with problems of this concept. As HRTFs are dependent on the angle of incidence, a cross-talk cancellation network has to be updated with every head movement of the listener. In practice this adaption is often causing problems, however, to give an example of application, the installation of a dynamic cross-talk cancellation system in a virtual reality environment at RWTH Aachen, Germany, has been described recently [9].

2 Concept

As stated before, cross-talk cancellation can be problematic in practice. A network is valid for just one geometrical configuration of speakers and listener position. Every modification, i.e. movement of the listener or the speakers, causes altered HRTFs and so the filter has to be recalculated by updating and inverting the matrix of HRTFs H. An obvious solution is, depending on the range the listener is allowed to move, to either record necessary HRTFs into a database or model them in real-time.

In addition, head rotations are limited to the angle area spanned by the speakers. This means, additional speakers are necessary for compensating full head rotations [9].

This could be avoided only if the speakers somehow remain at a constant position in relation to the ears. Literally speaking, the following solution could be considered: the loudspeakers must be able to follow all head movements.
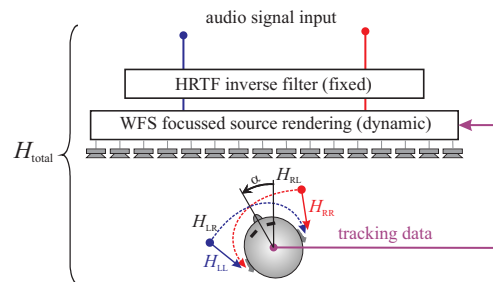

Figure 2: Schematic of the virtual headphone concept: A wave field synthesis renderer generates focussed sources at a constant position in relation to the head. The input signals are pre-filtered with a fixed HRTF inverse network.

Indeed, this idea can be realized by means of sound field reconstruction techniques. Building on physical models, appropriate superposition of the fields radiated by a number of secondary sources allows for simulating the sound field of e.g. a monopole source. One of the established methods — *wave field synthesis* — enables the generation of so-called *focussed sources*, virtual monopole sources generated in front of an array of loudspeakers [10]. They can be placed anywhere between listener and array, and they even can be moved by simply adjusting the computation of the focus points.

These virtual monopole sources can be seen as a representation of real loudspeakers that are able to keep a fixed position in relation to the ears of a moving head. Building on that, a static network for HRTF inverse filtering is sufficient, an adaptive system is not necessary. As the number of focussed sources is not restricted, several of them might be placed around the listener at optimum positions to improve stability and balance of the presentation. Dependent on the array size it is also possible to generate focussed sources for more than one listener, offering an own virtual headphone for every person.

Instead of adaptive HRTF inverse filtering, movements of the listener are compensated by dynamic sound field reconstruction. Figure 2 outlines this concept: Two audio channels, left and right of the virtual headphone, are pre-filtered by the HRTF inverse network. A wave field synthesis renderer is updated with the current head position and orientation delivered by a head tracking device and generates focussed sources at constant relative positions. Ideally the transfer function H_{total} equals one for the direct paths from left and right input to left and right ear, respectively, and zero for the cross-talk paths from the left input to the right ear and vice versa.

3 Prototype

A pilot study was carried out at the Institut für Rundfunktechnik, Munich, Germany, to proof the practicability of the concept. For practical reasons it was decided to design an array that can be mounted above the head. In this case no speaker will obstruct the listener's field of vision.

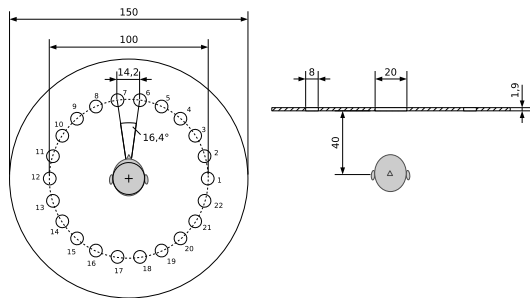For the reason of symmetry a circular loudspeaker array was designed. Arranged in a circle of 1.0 m in diameter,

Figure 3: Topview and section of the circular prototype array. The 22 speakers generate virtual monopole sources that remain at a close and constant position in relation to the listener's ears.
(All measures are given in cm)

22 speakers (ø 8 cm) are built in a wooden panel without enclosure. The whole panel measures 1.5 m in diameter to avoid acoustic short circuit effects. Additionally, a single driver (ø 20 cm) for low frequencies is placed in the centre. The distance between panel and head does not need to be clearly specified; for the experimental setup the array was usually suspended at about 40 cm above the listener and thus it is currently optimized for this configuration (see Figure 3).

A separate channel including D/A-conversion and amplification exists for each speaker. Three optical ADAT lines connect the array to the processing unit.

An electromagnetic motion tracking system (Polhemus FASTRAK) 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 realtime on a standard personal computer running with Linux operating system and a multi-channel digital sound card (RME Hammerfall DSP 9652).

This prototype is capable of compensating full head rotations, but there are restrictions in lateral movements: focussed sources cannot be generated anymore when one of the listener's ears leaves the area below the circle and therefore movement is limited to a radius of about 40 cm. Further technical details of the laboratory system are also described in [11].

4 Evaluation of the Prototype

4.1 Measurements

Measurements proved that the HRTF of a focussed source generated by the circular array is similar to that of a real source. Figure 4 exemplifies amplitude responses of head-related transfer functions for a real and a focussed source. For this plot, a small closed loudspeaker (ø 3 cm) was placed at a short distance to a dummy head (Neumann KU 100) near the left ear and compared to a focussed source generated at the same position. Although technical restrictions of wave field synthesis have a slight influence, it can be stated that, in general, a focussed source can replace a real loudspeaker: the theoretical concept is verified.

Figure 5 illustrates the amplitude response of the transfer functions H_{total} for both ears measured by means of the mentioned dummy head. It indicates a good perfor-
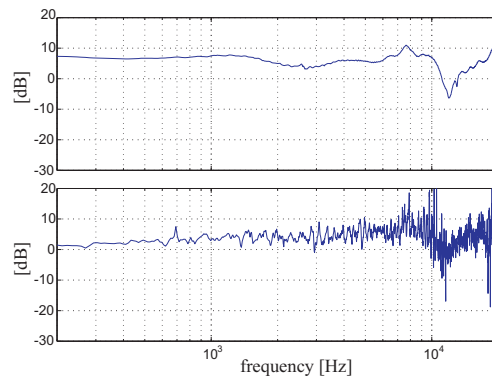

Figure 4: Magnitude plots of head-related transfer functions of a *real source* (top) and a *focussed source* (bottom), both placed at close distance to the head: the graphs show a similar characteristic.
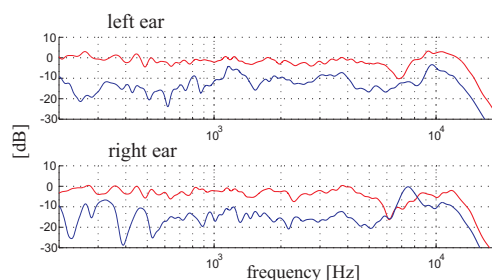

Figure 5: Amplitude responses H_{total} between audio signal input and both ears. In each diagram the upper curve represents the direct path, the lower one the unwanted cross-talk, which is considerably attenuated.

mance of the HRTF inverse filter: both diagrams show an attenuation of the unwanted signal (lower curves) of 10 to 20 dB in a frequency range up to about 7 kHz in comparison to the direct path (upper curves). Additionally, as the utilized dummy head is equalized to a diffuse field and the direct transmission paths are more or less linear in the frequency domain, it can be inferred that the virtual headphone is also equalized to a diffuse field, according to [4, 5].

4.2 Psychoacoustic Experiments

A very characteristic attribute of headphone reproduction — often seen as undesiered side effect — is *inside-the-head-locatedness* (IHL) or intracranial locatedness and corresponding effects of *lateralization* [2].

The psychoacoustic experiment shortly outlined in the following, is based on the idea that these attributes can be considered as an indicator for headphone reproduction. An experiment was designed to examine whether the virtual headphone is able to produce similar auditory events according to reported experiments on lateralization [2, 12].

Four different groups of stimuli were tested: a pure tone of 1 kHz, pink noise bandpassed between 200 and 10.000 Hz, one critical band wide noise around 1 kHz, and a second one around 7 kHz. The stimuli were presented at seven different inter-aural level differences and rated by 17 expert listeners (♀=4, ♂=13) without hearing disabilities, aged between 24 and 63 (ø=34).

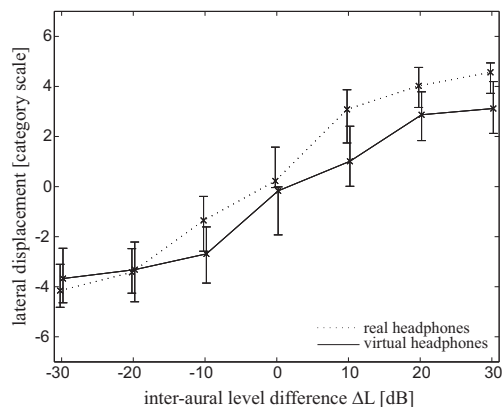The result is shown in Figure 6: negative values on the

Figure 6: Effect of lateralization as a function of the inter-aural level difference for a real and the virtual headphone (median and interquartile range): results are similar, but the lateralization curve of the virtual headphone is less steep.

x-axis denote that the left channel was louder; the category scale on the y-axis for the displacement of the auditory event ranged from "outside the left ear" (-6) over "at the left ear" (-5) up to "outside the right ear" $(+6)$. As expected, the resulting graph is very similar to those reported in literature [2, 12]. However, the lateralization curve of the virtual headphone is less steep than the curve for a real headphone. The auditory event cannot be shifted completely to one ear, very likely because of too much cross-talk (cp. Figure 5).

But the experiment suggests that the virtual headphone can reproduce the effect of inter-aural level differences on lateral shifts of perceived sources in a similar way as real headphones.

5 Summary

From a signal-theoretic point of view, the two different electro-acoustical reproduction principles of headphones and loudspeaker can be transformed into the respective counterpart. Methods to simulate a loudspeaker setup via headphones are summed up under the term binaural technology, and techniques for emulating the transmission characteristics of headphones by (at least two) loudspeakers are often referred to as virtual headphones.

The following concept for such virtual headphones is proposed: instead of an adaptive HRTF inverse filter, as it is usually applied, the loudspeakers are replaced by virtual monopole sources. These virtual monopoles can be generated by a number of secondary sources according to techniques of sound field reconstruction like wave field synthesis. The main advantage is that these sources are able to follow the movements of a listener and to keep a constant position in relation to the head. Thus a static HRTF inverse filter and an optimized source configuration can be used.

A prototype system with the secondary sources placed above the head of a listener could verify the functional capability of the concept.

As claimed in [11], examples of application for such a system can be seen especially in combination with binaural technology. It allows simulation of three-dimensional auditory scenes, which are not disturbed by headphones or visible speakers.

Acknowledgments

References

[1] H. Møller, M. F. Sørensen, D. Hammershøi, C. B. Jensen, "Head-Related Transfer Functions of Human Subjects", *J. Audio Eng. Soc.*, vol. 43, no. 5, pp. 300-321 (1995)

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

[3] D. Hammershøi, H. Møller, "Binaural Technique – Basic Methods for Recording, Synthesis, and Reproduction", *Communication Acoustics*, J. Blauert (Ed.), Springer: Berlin (2005)

[4] G. Theile, "On the Standardization of the Frequency Response of High-Quality Studio Headphones", *J. Audio Eng. Soc.*, vol. 34, no. 12, pp. 956-969 (1986)

[5] Recommendation ITU-R BS.708, "Determination of the electro-acoustical properties of studio monitor headphones", ITU, Geneva, Switzerland (1990)

[6] A. Lindau, T. Hohn, S. Weinzierl, "Binaural resynthesis for comparative studies of acoustical environments", Convention Paper 7032, 122nd *AES Convention*, Vienna, Austria (2007)

[7] U. Horbach, A. Karamustafaoglu, R. Pellegrini, P. Mackensen, G. Theile, "Design and Applications of a Data-based Auralization System for Surround Sound", Convention Paper 4976, 106th *AES Convention*, Munich, Germany (1999)

[8] J. Bauck, D. H. Cooper, "Generalized Transaural Stereo and Applications", *J. Audio Eng. Soc.*, vol. 44, no. 9, pp. 638-705 (1996)

[9] T. Lentz, "Dynamic Crosstalk Cancellation for Binaural Synthesis in Virtual Reality Environments", *J. Audio Eng. Soc.*, vol. 54, no. 4, pp. 283-294 (2006)

[10] M. M. Boone, E. N. G. Verheijen, and G. Jansen: "Virtual Reality by Sound Reproduction based on Wave Field Synthesis", Convention Paper 4145, 100th *AES Convention*, Copenhagen, Denmark, (1996)

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

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