The Binaural Sky: A Virtual Headphone for Binaural Room Synthesis

Daniel Menzel[1,2], Helmut Wittek[3,1], Günther Theile[1], and Hugo Fastl[2]

[1]Institut für Rundfunktechnik, München
[2]AG Technische Akustik, MMK, Technische Universität München
[3]Schalltechnik Dr.-Ing. Schoeps GmbH (since 2005)

A novel system for the reproduction of virtual acoustics is presented in theory and practice. The system combines wave field synthesis, binaural techniques and transaural audio. Stable localisation of virtual sources is achieved for listeners that are allowed to turn around and rotate their heads. Focused sources are used for transaural signal reproduction. The position of the focused sources is kept constant relative to the ears of the listener for every head direction by means of a head tracking system. Therefore, there is no need to change the crosstalk cancellation filters during head rotations, so that audible artefacts and instabilities are avoided. The system is sufficiently stable allowing for loudspeaker installations even above the listener. As a result, binaural (e.g. BRS) reproduction can be enjoyed without wearing a headphone and without any loudspeakers in the listener's field of vision.

Motivation

Binaural Room Synthesis (BRS), a technique developed at the IRT in the early 90s, provides a virtual listening environment via headphones. A primary example where BRS can be used is to aid sound engineers in sonically difficult situations, like OB-vans [5]. The signal to be reproduced is convolved with measured binaural room impulse responses (BRIR) and played back via high quality diffuse field equalized headphones. For every virtual source that is to be synthesized, a real loudspeaker has to be measured using a dummy head. The dummy head is put on a rotation table, so that the BRIR of every head direction can be stored. To avoid front-back inversions and to achieve a stable localisation of the virtual sources, a head tracker is used during playback which dynamically changes the set of BRIR corresponding to the current azimuth of the listener's head. As a result, the virtual sources are perceived at their real distance and their location stays constant regardless of the head direction of the listener.

However, a drawback of the BRS system is the need to wear headphones. If it is not possible to wear headphones, e.g. in a car or in a situation where real and virtual sources have to be mixed ("augmented reality"), a binaural reproduction without headphones would be desirable.

Concept

The standard way to play binaural signals via loudspeakers is to use transaural stereo, which uses crosstalk cancellation (XTC) filters to eliminate the unwanted signal paths between the speakers and the ears [1][2]. To obtain the XTC filters, the head related transfer functions (HRTF) of the path between the loudspeakers and the ears of the listener have to be measured or calculated with a mathematical model (e.g. [3]). However, the need for head-tracking results in the use of a whole set of XTC filters which have to be updated with every head rotation. This can result in audible artefacts.

In the best case, the loudspeakers used to reproduce the transaural signal ought to move with the head rotation of the listener, so that the relative positions between the sound sources and the ears stay constant and only one set of cancellation filters is needed. This is shown in figure 1.

Of course this concept isn't feasible with normal loudspeakers. Therefore, in the *Binaural Sky* the speakers are replaced by focused sources produced with techniques known from wave field synthesis (WFS). These focused sources act as the transaural loudspeakers, but they can easily be moved around by adjusting the driving functions (i.e. the delay times and attenuations) of the array loudspeakers.

By synthesizing focused sources at a close distance to the listener's head and using these as the

From: Intern. Tonmeistersymposium, Hohenkammer

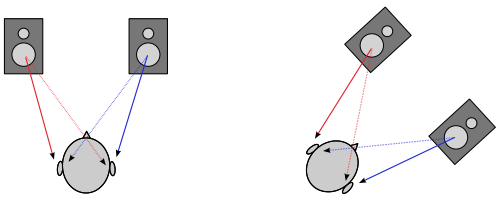

Figure 1: If the speakers of a transaural system move with the head rotation, only one set of XTC filters is needed because the HRTFs stay constant.

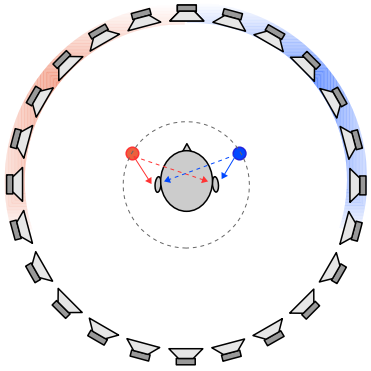

Figure 2: The circular array synthesizes focused sources which act as transaural sources.

transaural loudspeakers, a stable reproduction can be achieved without the need for adaptive XTC filters. The actual configuration of the focused sources (i.e. number and position) is flexible and can be changed as needed for an optimal transaural performance. Significant practical work was done at the IRT to optimise these parameters.

The Loudspeaker Array

Instead of a standard linear WFS array however, a circular design was chosen to ensure constant distances between the ears, the focused sources, and the array speakers. This leads to a constant aliasing frequency and greatly reduces audible sound colourations during head rotations. Figure 2 shows a schematic view of the circular array with two focused sources (red and blue dots). The dashed circle indicates the path on which the focused sources move during a full head rotation.

The real array has a diameter of 1m and consists of 22 broadband speakers (Ø 8 cm) with a single low frequency driver (Ø 20 cm) in the middle to reproduce frequencies below 120 Hz. The speakers are mounted in a baffle without an enclosure. The whole setup is suspended above the listener at a distance of 40 cm from the head (see figure 3). Note that there are no loudspeakers in the listener's field of vision so as not to obstruct the view on e.g. computer displays or TV monitors.

Signal Processing

All real time processing is done on a standard Linux PC with a 24 channel RME Hammerfall sound card. BruteFIR is used as the central processing engine, as it provides the necessary capabilities with its high throughput partitioned convolution algorithms. Figure 4 shows the signal path used for real time processing.

The N_{ch} input signals (e.g. $N_{ch} = 5$ for 5.1 surround playback) are fed into a first instance of the BruteFIR software. This module is responsible for performing the BRS related convolutions resulting in a binaural signal. The BRIR are changed according to the azimuth delivered by the head-tracker, so that at any time the correct BRIR is active and used for convolution. The binaural signal could already be played over headphones at this point, but the goal of the *Binaural Sky* is loudspeaker reproduction, so another processing stage is necessary. A second instance of BruteFIR convolves the binaural signal with pre-calculated filters and distributes it to the 22 array loudspeakers and the low frequency driver. This second stage acts as a "virtual headphone" - and also other non-binaural signals may be reproduced through this stage. Using this stage only, the inherent properties of pure headphone listening can be simulated, e.g. "In-Head-Localisation".

The pre-calculated filters consist of the crosstalk cancellation filters combined with the wave field driving functions. The XTC filters are obtained by taking the matrix of HRTFs of the focused sources $\mathbf{A}$ and computing its pseudo inverse $\mathbf{A}^{+}$, as described in [1] and [4]. The WFS filters can then be calculated from the array geometry and combined with the XTC filters. Note that only the wave field filters depend on the azimuth of the listener's head - the transaural part is constant.

Objective Evaluation

The performance of the complete system was evaluated using simulations and measurements. A central performance measure is the crosstalk attenuation, i.e. the level of a signal from the left binaural channel at the right ear and vice versa. In an ideal XTC system, the direct path transfer functions (left channel to left ear and right channel to right ear) would be one, and the crosstalk transfer functions would be zero. Figure 5 shows a simulation (left) and a measurement (right) of the crosstalk cancellation performance at the right ear. The red curves

Figure 3: Left: Measurement of the array with a dummy head. Right: The array in a typical listening position in the IRT studio.

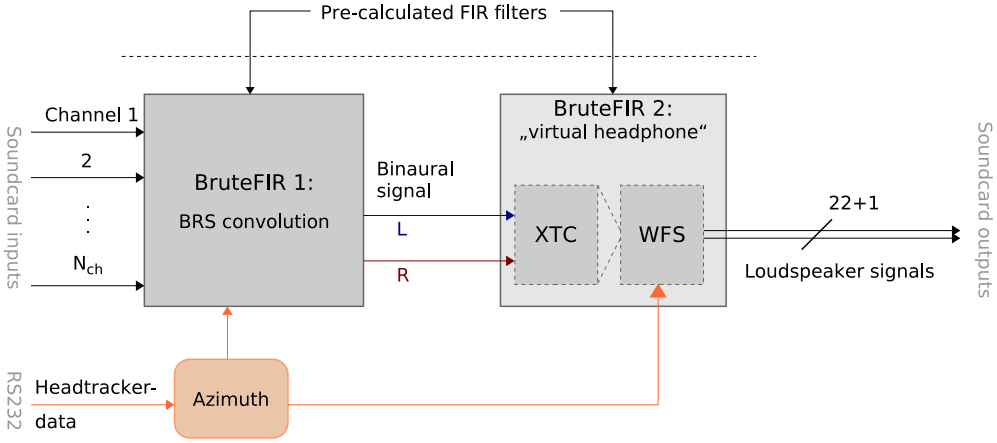

Figure 4: Real time signal processing necessary to produce the loudspeaker signals.

show that the direct path transfer functions are independent of frequency up to about 7 kHz, from where on spatial aliasing starts to have a negative influence on the XTC process. This can also be seen in the attenuation of the crosstalk, indicated by the blue curves, which stays at a level of about -20 dB up to the aliasing frequency.

At the moment, only head rotations are tracked and used for synthesis of virtual sources. Head movements out of the central listening area under the array (the "sweet spot") cause changes in the HRTF of the listener and therefore have negative effects on the success of the crosstalk cancellation. This can be seen in figure 6. The red curves show the measured transfer functions with the dummy head in the correct central position. The dummy head was then moved to the right in steps of 2 cm, which is depicted by the grey curves: the lighter the colour, the greater the lateral displacement. It is clearly visible in the left diagram that the crosstalk attenuation decreases and that (in the right diagram) the direct path transfer function at the right ear is distorted, which results in severe sound colourations and doesn't allow the perception of virtual sources. Subjective tests showed that these effects are tolerable up to a displacement of about 8 to 10 cm.

To minimize the negative effects of the limited sweet spot, the sound is switched off if the listener moves away from the central point by more than 5 cm to the sides or 10 cm to the front or back. Thereby it is assured that the listener only listens to a correct binaural signal.

In future applications it is well possible to include head movements into the dynamic process as well. This would lead to the need for delay time adjustments of the driving functions.

Subjective Evaluation

The performance of the synthesized virtual sources with respect to localisation and sound colouration

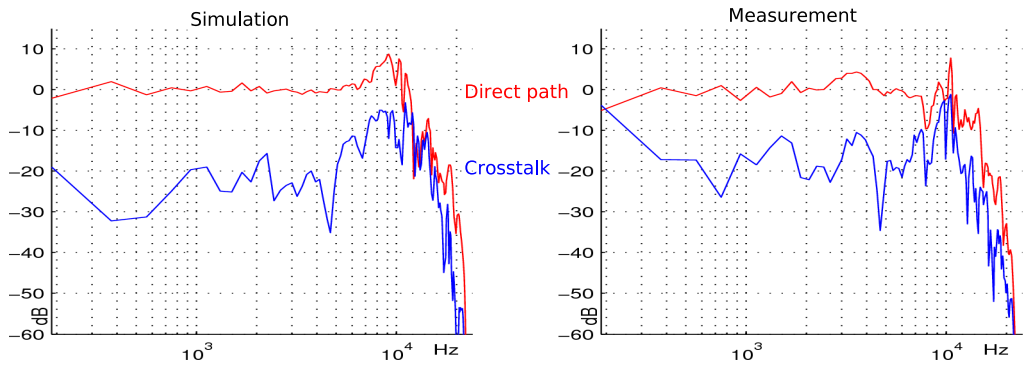

Figure 5: Simulation and measurement of the XTC performance at the right ear at an azimuth of 0°.

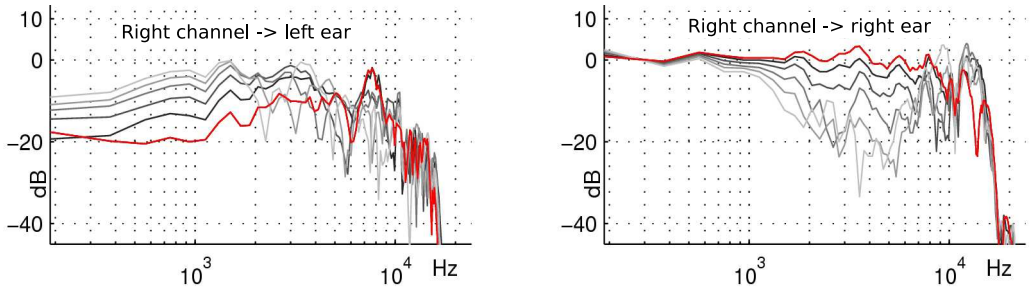

Figure 6: Measured effects of lateral displacement of the head on the crosstalk cancellation performance. Here, the dummy head was facing forward (azimuth of 0°, red curves), then moved to the right in steps of 2 cm up to 10 cm (gray curves).

was evaluated by psychoacoustical experiments.

Localization

Six real sound sources (small loudspeakers positioned around the subject) were compared to six virtual sources, which were synthesized at the same locations. All loudspeakers were hidden behind a sonically transparent curtain. The azimuths of the sources were 0°, 60°, 135°, 180°, 250° and 330°, the elevations were 0°, +25°, -25°, 0°, +10° and -10°, at a distance of 1m (figure 7). 15 subjects participated in this experiment and marked the perceived azimuth and elevation of the auditory events in a template on a piece of paper. Pink noise bursts with a sound pressure level of 60 dB(A) were used as a stimulus.

The horizontal localisation of virtual sources achievable with the *Binaural Sky* is comparable to real sound sources, which can be seen in figure 8. All virtual sources are perceived with good accuracy and only small variations, very similar to real sources and headphone reproduction [6].

A correct perception of elevation seems to be more difficult to achieve, as figure 9 suggests. Here the actual (green) and perceived (blue) elevations of the sources are plotted against the source number.

Figure 7: Setup of the localisation experiment with three of the six real sound sources visible. The array is hidden above the horizontal curtain.

It can be seen that the real sources are recognized quite accurately, but with greater variations than in the case of azimuth perception. The virtual sources show even more variation and a tendency of being perceived too high by about 10°. Virtual sources with negative elevation apparently cannot be synthesized with meaningful results, as can be seen from sources 3 and 6. These results however are

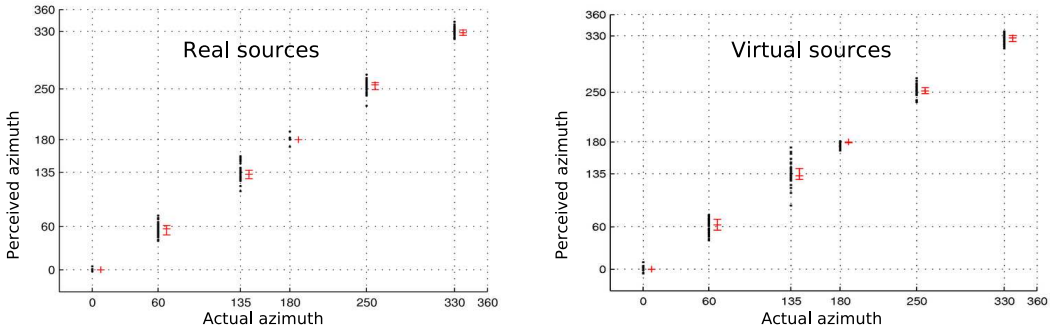

Figure 8: Results of the localization experiment: azimuth perception of real and virtual sources. The black dots show all answers given by the subjects, the median and interquartile range are drawn in red.

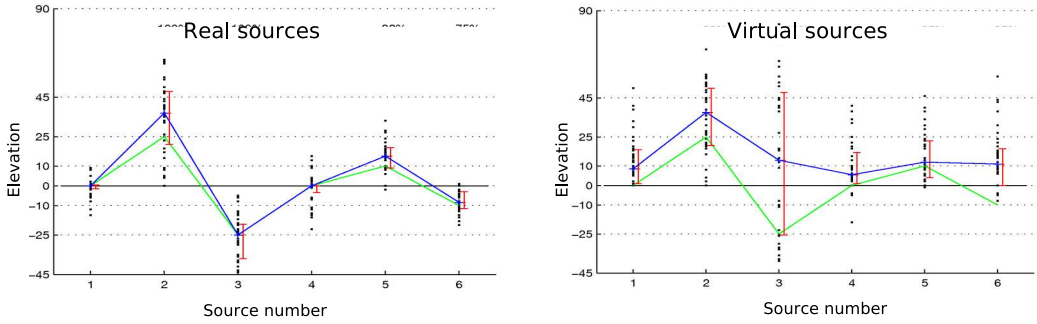

Figure 9: Results of the localization experiment: elevation perception of real and virtual sources. The black dots show all answers given by the subjects, the median and interquartile range are drawn in red. The green curve indicates the actual elevations, the blue curve shows the perceived elevations.

also known from headphone reproduction of binaural signals [6].

Sound coloration

In another experiment, possible sound colouration due to different source positions was investigated. Four virtual sources in front of the subject had to be compared regarding their differences in sound colour, using a standard 5-grade ITU scale (ranging from "very annoying" to "not perceptible"). Again, pink noise bursts were chosen as the stimulus. Three reproduction methods were compared: the *Binaural Sky* array, headphones and, as a reference, four individual loudspeakers of the circular array. The results can be seen in figure 10.

The differences in case of the *Binaural Sky* where rated "perceptible, but not annoying", just under the rating for headphone reproduction. The results of the individual array loudspeakers suggest that there are already considerable differences between the small broadband speakers used in the array, so that better drivers could further improve the whole system.

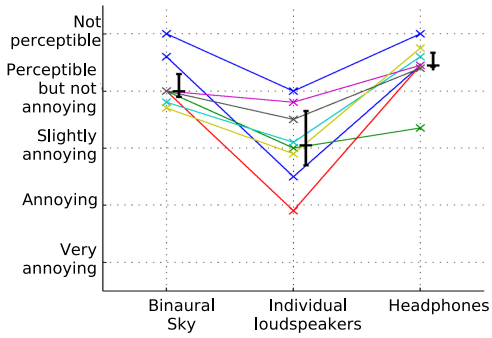

Figure 10: Variations in sound color between four sources with different positions. The colored curves show the results of the eight subjects, the median and interquartile range is drawn in black.

Conclusion

The *Binaural Sky* [7] can act as a virtual headphone for Binaural Room Synthesis, using wave field synthesis and transaural techniques. The usage of a head-tracker allows for stable virtual sources and avoids In-Head-Localisation. Psychoacoustical experiments showed a very good horizontal localisa-

tion of virtual sources, comparable to BRS playback via headphones or even real sound sources. However, there is no meaningful elevation perception for virtual sources below ear level, and sources at or above ear level frequently are heard about 10° too high. Variations in sound colour between different virtual sources are rather uncritical.

References

[1] Bauck, J. ; Cooper, D. H.: Generalized Transaural Stereo and Applications. In: *J. Audio Eng. Soc.* 44 (1996), September, Nr. 9, S. 683–705

[2] Damaske, P. : Head related two channel stereophony with loudspeaker reproduction. In: *JASA* 50 (1971), S. 1109–1115

[3] Gardner, W. G.: *3-D Audio Using Loudspeakers*, Massachusetts Institute of Technology, PhD thesis, September 1997

[4] Hokari, H. ; Furumi, Y. ; Shimada, S. : A study on Loudspeaker Arrangement in Multi-Channel Transaural System for Sound Image Localization. In: *AES 19th Int. Conference on Surround Sound, Elmau*, 2001

[5] Horbach, U. ; Pellegrini, R. ; Felderhof, U. ; Theile, G. : Ein virtueller Surround Sound Abhörraum im Ü-Wagen. In: *20. Tonmeistertagung, Karlsruhe*, 1998

[6] Spikofski, G. ; Fruhmann, M. : Optimisation of Binaural Room Scanning (BRS): Considering inter-individual HRTF-characteristics. In: *AES 19th Int.Conference on Surround Sound, Elmau*, 2001

[7] Wittek, H. . *http://www.hauptmikrofon.de/binauralsky.htm*