Audio Engineering Society

Convention Paper 6315

Presented at the 117th Convention
2004 October 28–31 San Francisco, CA, USA



# Dynamic Cross-Talk Cancellation for Binaural Synthesis in Virtual Reality Environments

Tobias Lentz[1], Gottfried Behler[2]

[1] Tobias Lentz, Institute of Technical Acoustics - RWTH Aachen University, Aachen, 52066 Germany
tobias.lentz@akustik.rwth-aachen.de

[2] Gottfried Behler, Institute of Technical Acoustics - RWTH Aachen University, Aachen, 52066 Germany
gkb@akustik.rwth-aachen.de

**ABSTRACT**

To create a Virtual Reality environment with true immersion a precise spatial audio reproduction system is required. Since the placement of large loudspeaker arrays which are needed for wave field synthesis systems may be impossible for some environments, alternative solutions must be found. One application of this kind, for instance, is a multi screen VR system where the stereoscopic video images envelope the user. In such a case the presented binaural approach has many advantages.

This paper describes the virtual sound source imaging by binaural synthesis and the reproduction over loudspeakers with a dynamic (tracked) cross-talk cancellation system which only needs three to four loudspeakers to cover all listening positions.

## 1. INTRODUCTION

Common research in the field of Virtual Reality (VR) considers acoustic stimulation as a highly important necessity for enhanced immersion into virtual scenes [8].

A very common way of placing the virtual sound sources into the scene is it to use the 5.1 technology which is sufficient for many applications. But for some applications a more sophisticated solution is necessary especially when a very exact sound source imaging concerning position and distance is required.

A well understood technique for the creation of an acoustic virtual scene is the wave field synthesis [9]. The main advantage is that more than one user can listen to the simulated sound field without tracking of the individuals position. This approach, however, needs a cost-intensive technical infrastructure, i.e. a large number of loudspeakers and amplifiers. For a proper 2D sound field synthesis with a reasonably high spatial aliasing frequency of about 5 kHz speakers must be placed nearly every 4 cm in a plane all around the listener. If true 3D scenes are required, the effort is even higher. Anyway, wave field synthesis is an excellent, maybe the best, solution for cinemas but hard to achieve in general especially, in virtual environments like a CAVE[1].

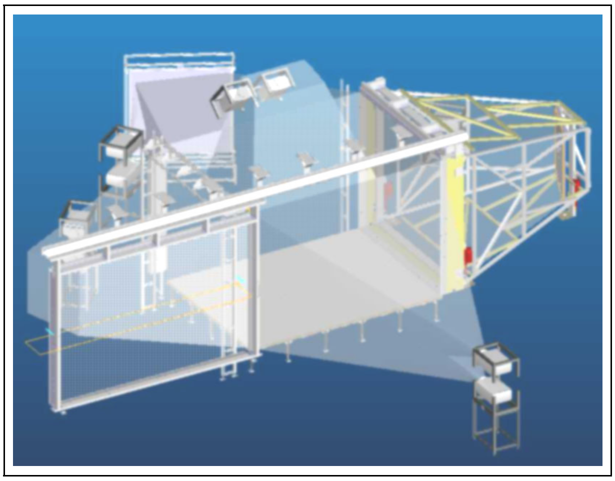

Figure 1. Schematic draft of the 5 screen (walls+bottom) CAVE build at the Center for Computing and Communication, RWTH Aachen

Binaural synthesis in combination with dynamic cross-talk cancellation (CTC), on the other hand, is an interesting way to allow spatial auditory representation by using few loudspeakers. In this case only a single person can act in the virtual scene because an exact reproduction of the binaural signals is only possible at designated points in space (i.e. the listener's ears). Furthermore, tracking of the listener's position is also needed to provide movements of the user. This solution may not be appropriate for any application but it does not introduce a drawback especially in VR environments like a CAVE or a Holo-Bench[2] where the stereoscopic video imaging technique already requires the use of a tracking system and, anyway, the creation of three-dimensional images is not possible for more than one user.

This paper describes a system that combines the virtual imaging by binaural synthesis and the reproduction over loudspeaker with a dynamic CTC for the listener who is interacting in a common visual VR application in real-time. The dynamic cross-talk cancellation procedure requires a system that provides a valid filter-set for each position. GARDNER [3] showed in his investigations the general applicability In the approach presented here, filters are calculated online using an HRTF-database with a resolution of 1°, both in azimuth and elevation. Further research has shown that a dynamic CTC using two speakers is stable only within the angle spanned by the loudspeakers. To provide a full 360° rotation for the listener the 2-speaker solution was expanded to four channels. With four loudspeakers eight combinations of a normal CTC system are possible; however, the validity areas of the two-channel CTC systems overlap. In these areas a cross-fade algorithm is implemented using two parallel working CTC systems, which are then partly superimposed. This way the current binaural audio signal is filtered with the correct cross-talk cancelling filter for the listener's present position.

Furthermore the described system provides a multi track binaural synthesis for generating virtual sources. The position of these sources can be either fixed in the room or driven by a stereoscopic video VR application to generate a congruent visual and acoustical scene which allows e.g. the research on interaction between visual and auditory human perception systems.

## 2. BINAURAL HEARING

Due to the fact that humans hear with two ears, a direction can be assigned to sound events [6]. This is similar to the processing of visual stimuli, where the brain compares pictures from both eyes to determine the objects' placing in a scene and with this information creates a three-dimensional cognitive representation that humans perceive as a three dimensional image. In straight analogy, acoustical stimuli that are present at the eardrums will be compared by the brain to determine the nature and the direction of a sound event. Depending on the horizontal angle of incidence, different time delays and levels between both ears appear. In addition, frequency characteristics dependent on the angle of incidence are influenced by the interference between the direct signal and the reflections of auricle, head, shoulder and other parts of the human body. The interaction of these three factors permits humans to assign a direction to acoustic events [7].

[1] CAVE: Video projection system with 3 to 6 screens.
[2] Holo Bench: Two screen video projection system

## 3. CROSS-TALK CANCELLATION

Since all information needed for a correct spatial impression is covered by a binaural signal, all one has to do is to reproduce these signals at the eardrums as correctly as possible [5]. Especially in a virtual reality environment a use of ear covering headphones is undesirably, so the only way is the use of loudspeakers. But the problem of loudspeaker reproduction is the cross-talk between the channels that destroys the three dimensional cues of the binaural signal. The essential requirement for a correct binaural presentation is that the right channel of the signal is audible only in the right ear and the left one is audible only in the left ear [1][2]. This problem can be solved by a cross-talk cancellation filter, which is shown in Figure 2. The four transfer functions are labelled $H_{LL}$, $H_{LR}$, $H_{RL}$ and $H_{RR}$. $H_{LR}$ and $H_{RL}$ indicate the cross-talk paths, which have to be cancelled by the system. The ear signals $Z_L$ and $Z_R$ can be described as:

$$Z_L = Y_L \cdot H_{LL} + Y_R \cdot H_{RL} = X_L \quad (1)$$

$$Z_R = Y_R \cdot H_{RR} + Y_L \cdot H_{LR} = X_R \quad (2)$$

The solution of equation (1) and (2) can be written as:

$$Y_L = \underbrace{\frac{H_{RR}}{H_{LL} \cdot H_{RR} - H_{LR} \cdot H_{RL}}}_{Ctc\_LL} \cdot X_L - \underbrace{\frac{H_{RL}}{H_{LL} \cdot H_{RR} - H_{LR} \cdot H_{RL}}}_{Ctc\_RL} \cdot X_R \quad (3)$$

$$Y_R = \underbrace{\frac{H_{LL}}{H_{LL} \cdot H_{RR} - H_{LR} \cdot H_{RL}}}_{Ctc\_RR} \cdot X_R - \underbrace{\frac{H_{LR}}{H_{LL} \cdot H_{RR} - H_{LR} \cdot H_{RL}}}_{Ctc\_LR} \cdot X_L \quad (4)$$

The parts labelled with brackets of equation (1) and (2) representing the four ctc-filters shown in Figure 2.

Using a head tracking system and a HRTF database it is possible to calculate a valid filter set for the listener's current position. The software is implemented on a PC architecture realising a cross-talk cancellation structure which is able to change its parameters in real time [4]. In this way the current binaural audio signal is filtered with the correct cross-talk cancellation filter for the specific position.

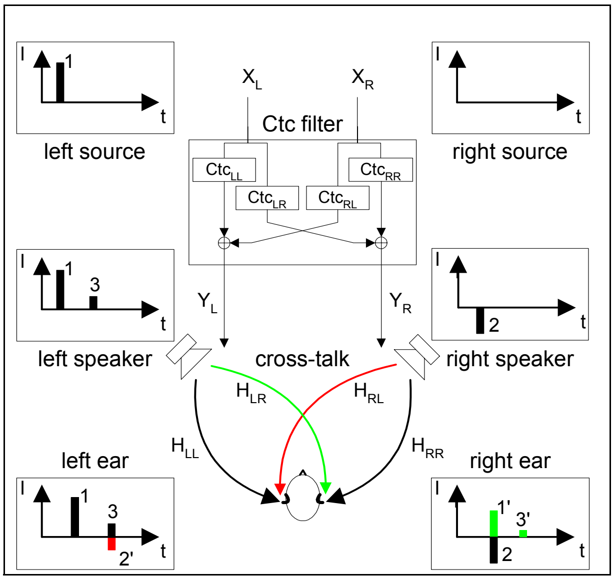


Figure 2. Principle of cross-talk cancellation

## 4. STABILITY

Tests showed that a dynamic cancellation works only in the angle spanned by the loudspeakers. This behaviour becomes clear looking at the denominator (5) of the filter equations (3) and (4).

$$D = H_{LL} \cdot H_{RR} - H_{LR} \cdot H_{RL} \quad (5)$$

If for any frequency $H_{LL} \cdot H_{RR}$ is equal to $H_{LR} \cdot H_{RL}$, the denominator becomes zero and the filter is not stable anymore. Even when the result is not exactly but almost zero, the resulting filter reaches a very high amplitude, which may result in ringing or even range overflow at the sound output device. Dividing the two parts of the denominator causes a relative scaling (equation (6) ) without the need to take the absolute amplitude of the transfer functions into account.

$$K = \frac{H_{LL} \cdot H_{RR}}{H_{LR} \cdot H_{RL}} \quad (6)$$

To detect possible singularities, $K$ was calculated for every head direction. Afterwards the algorithm searches the value closest to 1 in each filter and stores that value for the current head position. Note that $K$ is a frequency vector and all values have he below 1 to ensure the stability of the resulting filter set. Figure 3 shows the results searching critical values of *K* over a complete head rotation. On the left a ±45° speaker configuration was used, in the right plot it was ±90°. Plot a) reveals that a stable cross-talk cancellation is possible in a range of approximately ±40°, in plot b), the ±90° configuration, the valid area is about ±75°.

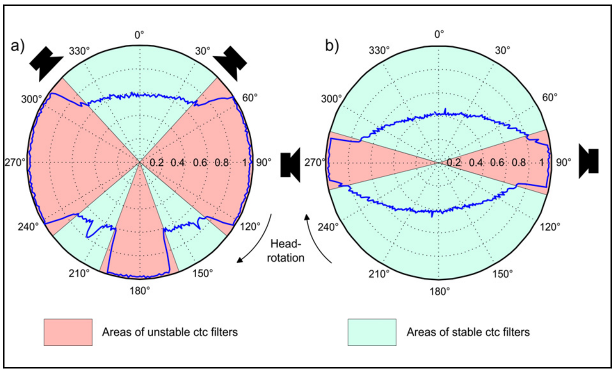


Figure 3. Plot of critical values of K for each direction. a) ±45°- and b) ±90°-speaker configuration. Both, speaker and ears are placed 1,7 m above the floor.

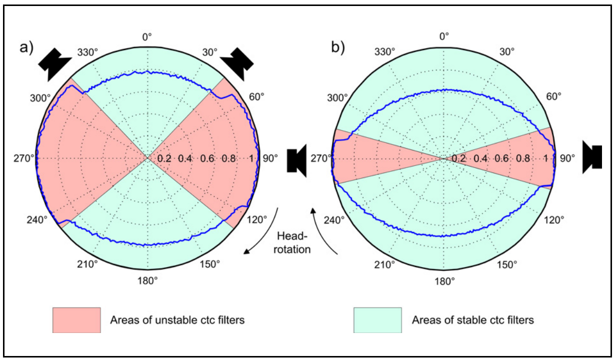


Figure 4. Plot of critical values of K for each direction. a) ±45°- and b) ±90°-speaker configuration. Speakers are placed 3 m, ears at 1.7 m above the floor

It is not necessary to place the speaker in the plane of the listener's ears; it is also possible to mound the speakers in an elevated plane. Choosing the adapted HRTFs from the database to calculate the compensation filters causes the same sound pressure at the eardrums of the listener. This fact is a strong argument for using this technique in a CAVE-like environment, where the placing of speakers is the main problem and for the most part only possible at the upper corners of the CAVE. The stability of the speaker configurations mounted in a height of 3 m is shown in Figure 4.

## 5. SWITCHING BETWEEN AREAS

To provide a complete head rotation, it is necessary to use a stable configuration at every possible viewpoint of the user and this leads to a four speaker environment, which makes it possible to combine the ±45° and ±90° configuration in the same set-up. Depending on the current viewpoint, the system automatically chooses the valid configuration. Measurements and listening tests showed that the cancellation achieves good results inside the particular areas, but switching between areas is still audible as a "click". Comparing CTC filters one step before and one step after switching the filters reveals some differences in particular at high frequencies. Inaccuracies in the speaker placement and the determination of the head position by the head tracker can cause a mismatch in the time alignment so that a sufficient consistent cancellation is not possible.

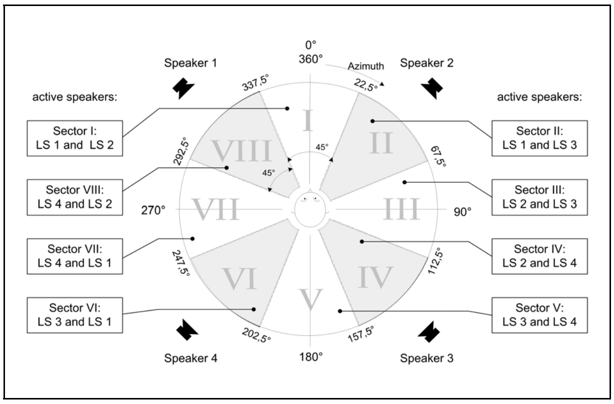


Figure 5. Grey: Areas for ±90° configuration, white areas for ±45° configuration

## 6. DUAL CROSS-TALK CANCELLATION

To reduce the interfering "clicks" a smoother changeover from one sector to the next is needed.

The whole four speaker system is described in equation (7) and (8).

$$Z_L = Y_1 \cdot H_{1L} + Y_2 \cdot H_{2L} + Y_3 \cdot H_{3L} + Y_4 \cdot H_{4L} \qquad (7)$$

$$Z_R = Y_1 \cdot H_{1R} + Y_2 \cdot H_{2R} + Y_3 \cdot H_{3R} + Y_4 \cdot H_{4R} \qquad (8)$$

This set contains four unknowns, but only two equations which makes a closed solution impossible [10]. It is only possible to give a numerical solution which is an approximation to an exact solution. Furthermore the calculation of the filters in that way is a CPU-time intensive method and not necessary stable in all situations e.g. when the listener faces exactly one speaker. In virtual reality environments focused in this work the cross-talk cancellation system has to react to the listener's movements very quickly and it must be stable in any situation.

Due to the fact that the cross-talk cancellation filter structure is a linear system, a linear superposition of two classical 2-speaker systems is possible. Based on the switching method described in chapter 5 an alternating use of the two (±45° and ±90°) speaker configurations with a superposition of both in the valid area in between can be developed. To describe the procedure the following indexing will be used:

We define an active area and a destination area where in each case only one speaker configuration has to be active to provide a sufficient and stable cross-talk cancellation. Speakers in the active area are labelled A and B, speakers in the destination area are labelled A and C. For example fading from sector *I* to *II* (Figure 5) speaker 1 and 2 (A, B) are active and after fading is complete, speaker 1 and 3 (A, C) are active. With this indexing a generally applicable system of equations can be established.

$$Z_L = Y_A \cdot H_{AL} + Y_B \cdot H_{BL} + Y_C \cdot H_{CL} \qquad (9)$$

$$Z_R = Y_A \cdot H_{AR} + Y_B \cdot H_{BR} + Y_C \cdot H_{CR} \qquad (10)$$

Labeling:

| | |
|---|---|
| $X_L, X_R$ | binaural input |
| $Y$ | speaker signal |
| $Z_L, Z_R$ | signals at the ears |
| $H_{Source\ Drain}$ | transfer function from source to drain |

The equation set (9), (10) again contains more unknowns than equations which makes a closed solution impossible. Taking account of the boundary conditions before and after fading, this set of equations can be split into two independent cancellation problems, which will be superimposed.

A complete channel separation and therefore a perfect cross-talk cancellation is true for:

$$Z_L = X_L \qquad Z_R = X_R \qquad (11)$$

In this case the signals at the listener's ears and the binaural input signal are identical. Because of the system's linearity it is permitted to split the input signal and one speaker signal each in two parts.

$$\begin{aligned} X_L &= X_L^{AB} + X_L^{AC} \\ X_R &= X_R^{AB} + X_R^{AC} \\ Y_A &= Y_A^{AB} + Y_A^{AC} \end{aligned} \qquad (12)$$

$X_L^{AB}$ is the part of the left input signal reproduced by speaker A and B. $Y_A^{AB}$ is the component of the signal at speaker A related to the signal at speaker B. This becomes clear by inserting the terms given in (12) into the equations (9) and (10):

$$X_L = Y_A^{AB} \cdot H_{AL} + Y_B \cdot H_{BL} \quad + \quad Y_A^{AC} \cdot H_{AL} + Y_C \cdot H_{CL} \qquad (13)$$

$$X_R = Y_A^{AB} \cdot H_{AR} + Y_B \cdot H_{BR} \quad + \quad Y_A^{AC} \cdot H_{AR} + Y_C \cdot H_{CR} \qquad (14)$$

It is now possible to separate the equations each in two parts and solve them separately.

System AB:

$$X_L^{AB} = Y_A^{AB} \cdot H_{AL} + Y_B \cdot H_{BL} \qquad (15)$$

$$X_R^{AB} = Y_A^{AB} \cdot H_{AR} + Y_B \cdot H_{BR} \qquad (16)$$

System AC:

$$X_L^{AC} = Y_A^{AC} \cdot H_{AL} + Y_C \cdot H_{CL} \qquad (17)$$

$$X_R^{AC} = Y_A^{AC} \cdot H_{AR} + Y_C \cdot H_{CR} \qquad (18)$$

The resulting systems represent two “classic” cross-talk cancellation structures. Both systems contain two unknowns and two equations, so a closed solution is possible.

System AB)

$$Y_A^{AB} = \frac{H_{BR}}{D^{AB}} \cdot X_L - \frac{H_{BL}}{D^{AB}} \cdot X_R \quad (19)$$

$$Y_B = \frac{H_{AL}}{D^{AB}} \cdot X_R - \frac{H_{AR}}{D^{AB}} \cdot X_L \quad (20)$$

With:

$$D^{AB} = H_{AL} \cdot H_{BR} - H_{AR} \cdot H_{BL} \quad (21)$$

System AC)

$$Y_A^{AC} = \frac{H_{CR}}{D^{AC}} \cdot X_L - \frac{H_{CL}}{D^{AC}} \cdot X_R \quad (22)$$

$$Y_C = \frac{H_{AL}}{D^{AC}} \cdot X_R - \frac{H_{AR}}{D^{AC}} \cdot X_L \quad (23)$$

With:

$$D^{AC} = H_{AL} \cdot H_{CR} - H_{AR} \cdot H_{CL} \quad (24)$$

To fade from one sector to the next a factor dependent on the actual orientation of the listener’s head will be defined. Each sector limit will be replaced by a small fading area. Again the active sector before entering the fading area is only driven by the AB system. At the beginning of the fading the actual head angle $\varphi$ is equal to $\varphi_0$ at the end it is $\varphi_1$.

Conditions at start of fading ( $\varphi = \varphi_0$ ):

$$X_L = X_L^{AB} \qquad X_R = X_R^{AB} \qquad Y_A = Y_A^{AB}$$

Conditions at end of fading ( $\varphi = \varphi_1$ ):

$$X_L = X_L^{AC} \qquad X_R = X_R^{AC} \qquad Y_A = Y_A^{AC}$$

The outcome of this is the definition of the head angle dependant weighting factor $a(\varphi)$.

$$a(\varphi) = 0 \qquad \varphi \le \varphi_0 \,,$$

$$a(\varphi) = \frac{\varphi - \varphi_0}{\varphi_1} \qquad \varphi_0 \le \varphi \le \varphi_1$$

$$a(\varphi) = 1 \qquad \varphi \ge \varphi_1$$

According to this the input signal is distributed to the two CTC systems.

$$X_L = (1 - a(\varphi))X_L^{AB} + a(\varphi)X_L^{AC} \quad (25)$$

$$X_R = (1 - a(\varphi))X_R^{AB} + a(\varphi)X_R^{AC} \quad (26)$$

Figure 6 shows the complete cross-talk cancellation filter structure of a three speaker solution. To provide full rotation the ±45° and ±90° configurations alternate, as well as practiced by the switching method. Only in between two sectors both configurations are active at the same time. Listening tests using this dual CTC showed that “clicks” are apparently not audible anymore. With this system an efficient cross-talk cancellation can be established in the full space around the listener and is especially suited to enhance virtual reality systems like a CAVE with spatial audio to combine visual and acoustical stimuli in an excellent way.

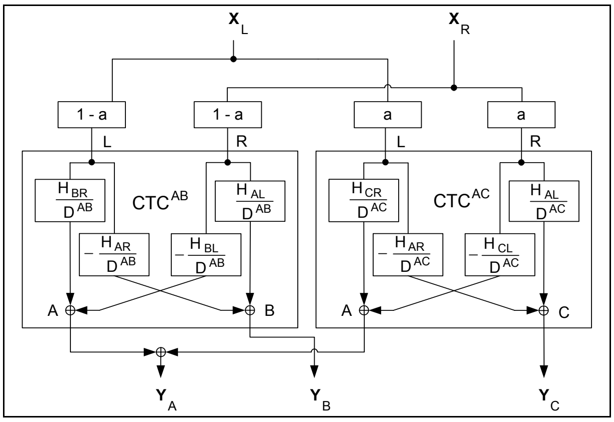


Figure 6. Dynamic three channel dual cross-talk cancellation filter structure with cross-fading

## 7. DYNAMIC BINAURAL SYNTESIS

The procedure of convolving a mono sound source with an appropriate pair of HRTFs in order to obtain a synthetic binaural signal called "binaural synthesis" is well known. The synthesized signals contain the direction information of the source, which is provided by the information in the HRTFs. The binaural synthesis transforms a sound source without position information into a virtual source related to the listener's head. Combining binaural synthesis and cross-talk cancellation into one system has considerable advantages. A recorded binaural signal is related to the listener's head and if the listener moves his head the source follows this movement. In a dynamic cancellation system the listener's position is always known and can also be used to realize a synthetic sound source with a fixed position corresponding to the room coordinate system. The software calculates the relative position and orientation of the listener's head to the imaginary point where the source should be localized. By knowing the relative position and orientation the appropriate HRTF can be chosen from the existing database of the cross-talk canceling system. It is also possible to realize many different sources and create a complex three-dimensional acoustic scenario.

Figure 7 shows the schematic implementation combining binaural synthesis and cross-talk cancellation. Future extension will be the implementation of an additional database which can hold long room impulse responses. With an effective low latency convolution algorithm it will be possible to auralise a room and represent the signal by a CTC loudspeaker system, without loosing the binaural cues.

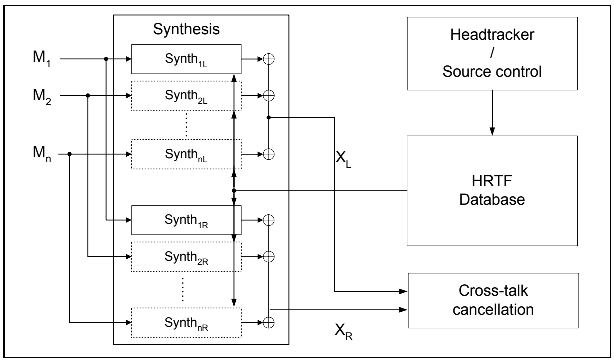


Figure 7. Combined dynamic binaural synthesis and cross-talk cancellation

## 8. COMPLETE SYSTEM - PERFORMANCE

The complete system's layout with all components is shown in Figure 8. The input section, connected to the head tracking device and to the HRTF database, accomplishes the multi track convolution and mixing of the mono sound files or sound device input channels. The rendering of a single source causes a CPU load between 3% and 5% on an 1800 MHz standard PC. The other main part of the system is the CTC unit including the sector control. It is also connected to the database and the head tracking. This part of the system causes 15% to 30% CPU load at the same 1800 MHz PC, depending on the listeners' movement velocity. A filter update will be performed when the weighted sum of the listeners' movement in all degrees of freedom is above 1. The threshold always refers to the value where the last exceeding occurred. The resulting hysteresis prevents a permanent switching between two filters as it may occur when a fixed spacing determines the boundaries between two filters and the tracking data jitter a little bit.

$$s = \left( \frac{|x_n - x_a|}{\Delta x} + \frac{|y_n - y_a|}{\Delta y} + \frac{|z_n - z_a|}{\Delta z} + \frac{|\varphi_n - \varphi_a|}{\Delta \varphi} + \frac{|\vartheta_n - \vartheta_a|}{\Delta \vartheta} + \frac{|\rho_n - \rho_a|}{\Delta \rho} \right) \geq 1$$

$n$ = new value, $a$ = old value

An other point is the latency of the system, the time elapsed between a real head movement and the point of time the output signal is generated with the updated filter functions. There are primarily three factors which influence the latency namely the head tracker, the filter calculation and the block convolution. In the test environment a "Flock of Birds" [11] was used as tracking device with a latency of about 20 ms. The block length of the convolution is 256 taps (5.8 ms at 44.1 kHz sampling rate) and the calculation of a new filter set takes 1.5 ms on a 1800 MHz PC.

In a worst case scenario the filter calculation just finishes after the sound output device fetched the next block, so it takes the time playing this block until the updated filter becomes active at the output. That would cause a latency of one block (5,8 ms). In such a case the overall latency accumulates to 27.3 ms which is sill fast enough for an interactive system. Tests showed that the listener mostly did not notice the latency. Beside very fast movements resulting in a short after image.

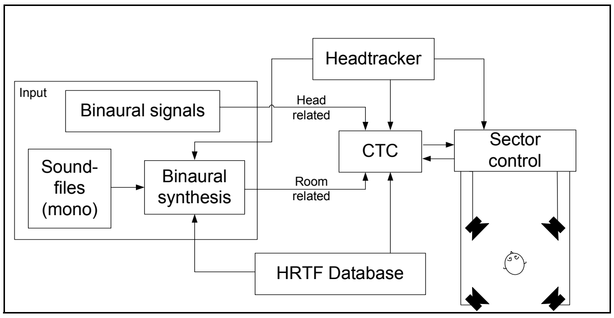


Figure 8. Complete system of dynamic binaural synthesis and cross-talk cancellation

## 9. INFLUENCE OF REFLECTIONS

The localisation results in an anechoic chamber, an ideal acoustical environment, are very good, but when using the cross-talk cancellation and the binaural synthesis together with a video VR-system the influence of reflections has to be examined to ensure the applicability. All screens for video projection also cause interfering reflections and may reduce the ability to localise the virtual sources at the correct position. A preliminary hearing-test was performed to investigate the difference of the localisation comparing an ideal environment to an environment with reflecting walls around the test person. The test was performed in an anechoic chamber with additional walls of plywood to simulate the video walls. The chosen material provides roughly the same acoustical properties as the fibreglass for the video projection. This workaround was chosen because the real CAVE environment was still under construction but all tests will be performed again after completition to prove the results. But this will be a topic for further research. Anyway this is still a good estimation of the ability of virtual image localisation in reflective environments.

The first test was performed only with a binaural signal presented over the dynamic cross-talk cancellation system. The dynamic binaural synthesis was not active. In this case all spatial information is related fixedly to the listeners head and turns just as well as the head turns (see Figure 9).

Furthermore the synthesis was performed online (dynamic synthesis) and the relative direction of the source was updated instantly. This generates a virtual source located at a fixed position in the room and the listener can turn his head to the imaginary point where he hears the source to examine and encourage his impression.

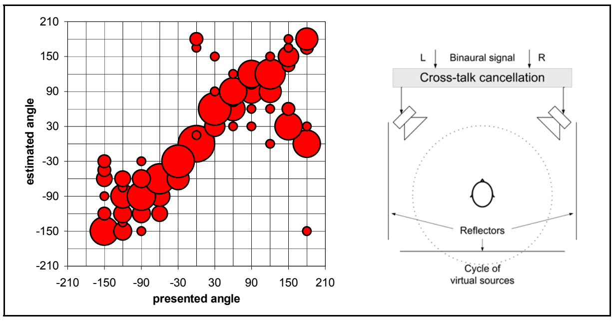


Figure 9. Localisation results by presenting a binaural signal over the dynamic CTC

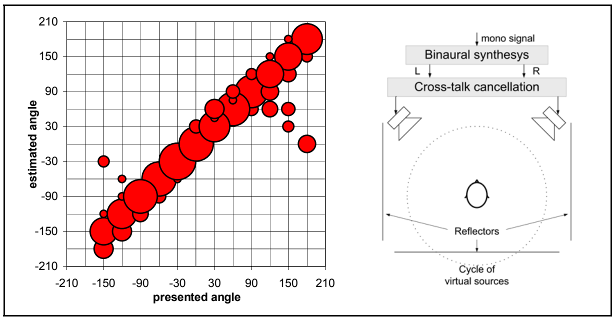


Figure 10. Localisation results by presenting a binaural signal fixed related to the room coordinate system over the dynamic CTC

The results shown in Figure 10 are very good despite of the reflections and reveal a significant improvement of correct localisation. Due to the fact that reflections arrive at the ears later than the direct signals, humans are still able to detect the right direction of the source. These tests show that cross-talk cancellation together with binaural synthesis is an appropriate technique for sound reproduction in virtual reality systems. A more detailed description of the influence of reflection influence is still a topic of further research.

## 10. OUTLOOK

There are still many things to do or to investigate. One interesting point is the implementation of a fast low latency convolution unit to extend the dynamic binaural synthesis to longer filters e.g. for room impulse responses. Furthermore source directivity is a fundamental attribute which makes the impression much more realistic, especially in an environment where a listener is able to interact with the scene. This feature is implemented already but not tested and studied jet.

To enhance the channel separation of the cross-talk canceller two possibilities will be checked in future. The prediction of the listeners position enables the system to calculate the filters more precisely for the position of the listener to the point of time the filter becomes active at the sound output device. The other step is to include the first few reflections of the walls into the cancellation process. This would cause a better performance, especially in reflective environments.

Ongoing work is also the coupling of the presented system with the VR software ViSTA [12]. Therewith it will be possible to combine three dimensional stereoscopic video images with true spatial audio for a enhanced immersion of a user into virtual scene.

## 11. ACKNOWLEDGEMENTS

This work was supported by Christian Renner programming parts of the CTC algorithm and Karl Slenczka doing the investigation on the influence of reflections.

## 12. REFERENCES

[1] MØLLER H.: Reproduction of artificial head recordings through loudspeakers., Journal of the Audio Engineering Society. 37 (1989).

[2] SCHMITZ A.: Naturgetreue Wiedergabe kopfbezogener Schallaufnahmen über zwei Lautsprecher mit Hilfe eines Übersprechkompensators., Dissertation, Institut für Technische Akustik, RWTH Aachen (1993).

[3] GARDNER W.G.: 3-D audio using loudspeakers., Ph.D. thesis, Massachusetts Institute of Technology, (1997)

[4] LENTZ T., SCHMITZ O.: Realisation of an adaptive cross-talk cancellation system for a moving listener, 21st AES Conference on Architectural Acoustics & Sound Reinforcement (2002)

[5] BAUER B.: Stereophonic earphones and binaural loudspeakers. Journal of the AES 9 (1963).

[6] BLAUERT J.: Spatial Hearing, Revised edition. Cambridge, Massachusetts: The MIT Press, 1997.

[7] MØLLER H.: Fundamentals of binaural technology. Applied Acoustics 36 (1992), 171–218.

[8] NAEF M., STAADT O., GROSS M.: Spatialized audio rendering for immersive virtual environments. In Proceedings of the ACM symposium on Virtual reality software and technology, Hong Kong, China (2002), pp. 65 – 72.

[9] THEILE: Potential wavefield synthesis applications in the multichannel stereophonic world. In 24th AES International Conference on Multichannel Audio (2003).

[10] BAUCK J., COOPER D.H.: Generalization Transaural Stereo and Applications, . Journal of the AES, Volume 44 (1996). pp. 683-705

[11] ASCENSION TECHNOLOGY: The Flock of Birds, Installation and operation guide, (1999)

[12] VAN REIMERSDAHL T., KUHLEN T., GERNDT A., HENRICHS J., BISCHOF C.: Vista: a multimodal, platform-independent vr-toolkit based on wtk, vtk and mpi. In Fourth International Immersive Projection Technology Workshop (IPT2000), Ames, Iowa (2000).