Fabio Kaiser

Transaural Audio - The reproduction of binaural signals over loudspeakers

Toningenieur-Projekt

vorgelegt an der
Universität für Musik und darstellende Kunst Graz (KUG),
Institut für Elektronische Musik und Akustik (IEM)

durchgeführt in der
Equipe Espaces Acoustiques et Cognitifs
de l'Institut de Recherche et Coordination Acoustique/Musique
(IRCAM)
Paris – France

Supervisors:
Markus NOISTERNIG, Olivier WARUSFEL, Thibaut CARPANTIER

March 2011

Abstract

Binaural audio is a technique to create an aural perception of a three-dimensional sound field for a dedicated listener. The foundation is the knowledge that the ears of a human function as spatial filters. The idea is to reproduce the exact same sound pressure levels at the entrance of the ear canals, which would be prevalent in a real acoustic scene. As a consequence the real perception cannot be distinguished from the virtual one.

The challenge is on the one hand the correct measurement or synthesis of the sound pressure at the two ears and on the other hand the correct reproduction of these. Here we regard only the reproduction problem.

The use of headphones for reproduction is self evident. However, loudspeakers can be used as well. As the signal reproduced by one loudspeaker arrives at both ears (crosstalk) the question is how to get a signal into one ear only. The idea is to use crosstalk cancellation (CTC) before playback.

This project was dedicated to the implementation of known algorithms for CTC for two loudspeakers. The general condition of the problem of CTC for different loudspeaker spacings was investigated and further the performance of different algorithms in different setups was objectively evaluated. Finally, a preliminary listening test was undertaken in order to evaluate the performance of diverse loudspeaker spacings.

Contents

Chapter 1

Introduction

In the field of 3D audio, binaural technology states a powerful tool for simulating natural localization processes of human beings. It tries to deliver signals to the listener's ears containing all auditory cues corresponding to a sound source position in space. These cues depend on the physiology of head, pinna, and torso of the listener and the angle of incidence of a sound event. This information is all comprised in the so called head related transfer function (HRTF). HRTFs can be measured and used for the synthesis of binaural signals or microphones can be placed at the entrance of the ear canal, either of a real person or an artificial head, to record binaural signals.

During playback no further filtering through the listener's HRTFs is desired because these directional cues are already existent in the binaural signal. This leads to the use of headphones for binaural reproduction. Nevertheless, attempts have been made to deliver binaural recordings via loudspeakers. Cooper and Bauck called this process, transaural stereo [1]. In this thesis the more generalized name - transaural audio - will be employed.

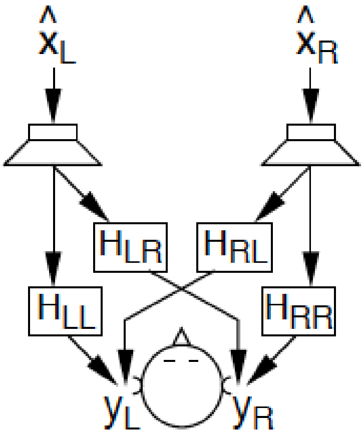


Fig. 1.1: Crosstalk occurring at loudspeaker reproduction [2]

The main difference in loudspeaker listening compared to headphone listening is the fact that the signal of one loudspeaker always reaches both ears. This is called crosstalk (Fig. 1.1).

The challenge now is to avoid this crosstalk and thus to create a *virtual headphone.* First attempts tried to process the binaural signals before playback through loudspeakers so that the crosstalk cancels. This process is called crosstalk cancellation (CTC) and was first mentioned by Bauer in 1961 [3] and the first implementation was proposed by Atal and Schroeder [4] in 1973.

The basic principle of CTC for a two-channel setup will be explained now. The following equations are expressed in frequency domain and the nomenclature is adopted from Gardner [5].

The idea is to cancel the crosstalk paths from the loudspeakers to the ears H_{LR} and H_{RL} shown in Fig. 1.1. The matrix $\boldsymbol{H}$ comprises the transfer functions inherent to the playback situation and its geometry, the so called plant. If we suppose a human is present in this plant the acoustical paths correspond to the human's head related transfer functions (HRTFs) and the two points of perfect CTC are the positions of the ears. $\boldsymbol{H}$ is further called the head transfer matrix.

The ear signals are given by

$$\boldsymbol{e} = \boldsymbol{H}\boldsymbol{y} \tag{1.1}$$

$$\boldsymbol{y} = \begin{bmatrix} y_L \\ y_R \end{bmatrix} \quad , \quad \boldsymbol{e} = \begin{bmatrix} e_L \\ e_R \end{bmatrix} \quad , \quad \boldsymbol{H} = \begin{bmatrix} H_{LL} & H_{RL} \\ H_{LR} & H_{RR} \end{bmatrix} \tag{1.2}$$

Where the vector $\boldsymbol{e}$ contains the signals entering the ears and $\boldsymbol{y}$ is the vector of the loudspeaker signals. The head transfer matrix $\boldsymbol{H}$ contains the HRTFs of the subject. It can be seen that this matrix not only consits of non-zero elements on the main diagonal but also on the antidiagonal, which is the crosstalk. This is valid only for a free-field situation where the head transfer matrix $\boldsymbol{H}$ doesn't contain the plant's room acoustics, the frequency response of the loudspeakers and the air transmission transfer function. For readability this is left out.

The cancellation is achieved by processing the input signals, *i.e.* placing a cancellation matrix between the input vector $\boldsymbol{x}$ and the output of the loudspeakers $\boldsymbol{y}$.

$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} \tag{1.3}$$

$$\boldsymbol{x} = \begin{bmatrix} x_L \\ x_R \end{bmatrix} \quad , \quad \boldsymbol{C} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \tag{1.4}$$

Inserting equation 1.3 in 1.1 yields

$$\boldsymbol{e} = \boldsymbol{H}\boldsymbol{C}\boldsymbol{x} \tag{1.5}$$

In order to achieve perfect reconstruction of $\boldsymbol{x}$ at the listener's ears the matrix has to be inverted.

$$\boldsymbol{C} = \boldsymbol{H}^{-1} \tag{1.6}$$

The inversion of the two channel setup results in

$$\boldsymbol{C} = \boldsymbol{H}^{-1} = \frac{1}{D} \begin{bmatrix} H_{RR} & -H_{RL} \\ -H_{LR} & H_{LL} \end{bmatrix} \tag{1.7}$$

where D is the determinant given by

$$D = H_{LL}H_{RR} - H_{LR}H_{RL} \tag{1.8}$$

As mentioned, for a perfect reconstruction of the input signals at the ears of the listener the cancellation matrix has to be the exact inverse of the head transfer matrix. In general, the HRTFs and the corresponding determinant are non minimum phase, and therefore direct inversion of the matrix is not feasible. Approximations have to be found. Hence, perfect crosstalk cancellation is not possible.

Since the first pioneers of transaural audio many researchers have worked on the topic. In 1989 Cooper and Bauck revived the ideas of Atal and Schroeder and suggested improvements to the implementations [1]. Further in 1996 they published a generalized theory of transaural audio, where any number of loudspeakers and listeners are possible [6]. William Gardner introduced in his PhD thesis the concept of the inter-aural transfer function (ITF) to improve implementations [5]. In 1998 Kirkeby et al. [7] suggested the *stereo dipole*, which put two loudspeakers close together and improves the robustness of the reproduction to movements of the listener. Takeuchi et al. [8] proposed a continuous source (loudspeaker) distribution in order to achieve a system with low sound coloration.

There are also acoustical alternatives to the CTC method. In [9], focused sound sources generated using wave field synthesis deliver binaural signals to the listeners ears. A similar approach tries to focus sound sources to the listener ear positions by using beam-forming techniques [10]. Usually a high amount of loudspeakers are necessary. Although these methods state very interesting approaches, this thesis concentrates on the problem of crosstalk cancellation using two-channel loudspeakers setups.

This report reviews in detail the above mentioned approaches for binaural sound reproduction over loudspeakers. The inversion of non minimum-phase filters is examined and different structures and implementations of CTC will be investigated. These structures are compared against each other using offline simulations, real-time processing, and preliminary listening tests.

The report is based on an internship at the *Equipe Espaces Acoustiques et Cognitifs* of the *Institut de la Recherche et Coordination Acoustique/Musique (IRCAM).* All computations have been implented for offline-rendering in Matlab; for real-time implementations of the different processing structures Max5 (Cycling'74) and IRCAM's Spat (Spatialisateur) spatial audio engine have been applied.

Chapter 2

Inversion in the context of transaural audio

Crosstalk cancellation in the context of transaural audio refers to system inversion and more specifically the inversion of a 2×2 matrix. A common tool for measuring the invertibility of a matrix is the condition number. Further the filters to be inverted, *i.e.* a HRTF or the determinant are of non minimum phase. Thus no direct inversion is possible.

In this chapter the condition number of a matrix is used as a measure and further techniques for the inversion of non minimum phase filters are presented.

2.1 Matrix conditioning

The condition number is a measures of how well numerically a problem is conditioned. The linear equation system of $b = \boldsymbol{A}x$, can compute be well-conditioned or ill-conditioned, which corresponds to a small condition number and a high condition number. It can also be interpreted as a measure of how strongly an error in b influences the result in x. If the problem is well-conditioned, a small error in b results in a small error in x and vice versa. So it is a measure of how accurate the solution will be.

The condition number is defined as the ratio of the norm of the relative error in Δx to the norm of the relative error in Δb.

$$\frac{||A^{-1}\Delta b||/||A^{-1}b||}{||\Delta b||/||b||} \tag{2.1}$$

Rewriting then yields

$$\kappa(A) = ||A^{-1}|| \cdot ||A|| \tag{2.2}$$

It can easily be seen that the condition number is dependent on the norm. A common norm is the L_2-norm. Hence, the condition number is the ratio of the biggest to the smallest singular value.

$$\kappa(A) = \frac{\sigma_{max}(A)}{\sigma_{min}(A)} \tag{2.3}$$

The analysis of the conditioning of the head transfer matrix can be found in literature [8][11][12]. Nelson et al. [12] have, for instance, shown that the condition number not only depends on frequency but also on the spacing of the loudspeakers. Takeuchi suggested [8] the “Optimal source distribution”. This system intends a continuous distribution of transducers in order to minimize the condition number for all frequencies. A practical solution comprises a loudspeaker system where three loudspeakers are positioned at ±90° for CTC at low frequencies, at ±16° for mid frequencies and at ±3.1° for high frequencies [13]. This system minimizes the condition number approximately over the whole frequency range which means that the inversion yields less errors.

An in-depth discussion of the conditioning of transaural systems is provided in Ch. 4.

2.2 Inverting a mixed phase FIR filter

A finite-length impulse response filter (FIR) can be described by its roots and zeros in the complex z-plane. The poles are all at $z = 0$ and the zeros can either be inside the unit circle (minimum phase), outside the unit cirlce (non-minumum phase) or on the unit circle. The inversion of a FIR filter results in an recursive IIR (infinite impulse response) filter. Is one of the FIR’s zeros outside the unit circle the resulting IIR filter has a pole outside the unit circle thus is unstable. Due to reflections in the pinna, which focus the sonic energy of very high frequencies to the ear canal, HRTFs cannot be seen as minimum-phase filters over all frequencies.

Fig. 2.1 depicts the zeros-pole plot of an diffuse-field equalized HRTF. It can be clearly seen that all the zeros are close to the unit circle and some are placed outside the unit circle.

An alternative to the use of measured HRTFs is to modelelize the transfer functions. The model could be simple using just a magnitude of one for the ipsilateral HRTF, and a delay, attenuation and a low-pass filter for the contralateral ear’s transfer function (see [2] and [5]). Further Cooper and Bauck used a rigid sphere model to calculate the HRTF database [1]. These approaches yield a set of plant transfer functions that has minimum phase properties.

However, the more accurate way is to use measured HRTFs and find approximations for the inverse. In an ideal scenario the impulse responses of the prevalent playback situation are measured. These then include also the room impulse response (RIR). In this research report the HRTF base to the following investigations is built by HRTFs measured on real persons in an anechoic chamber.

The first part of this chapter introduces techniques to solve the inversion problem and qulitative comparisons are presented. In the second part the conditioning of the inversion over frequency and over different loudspeakers spans is analyzed.

2.2.1 Inversion techniques

This section reviews methods for the inversion of non-minimum phase filters.

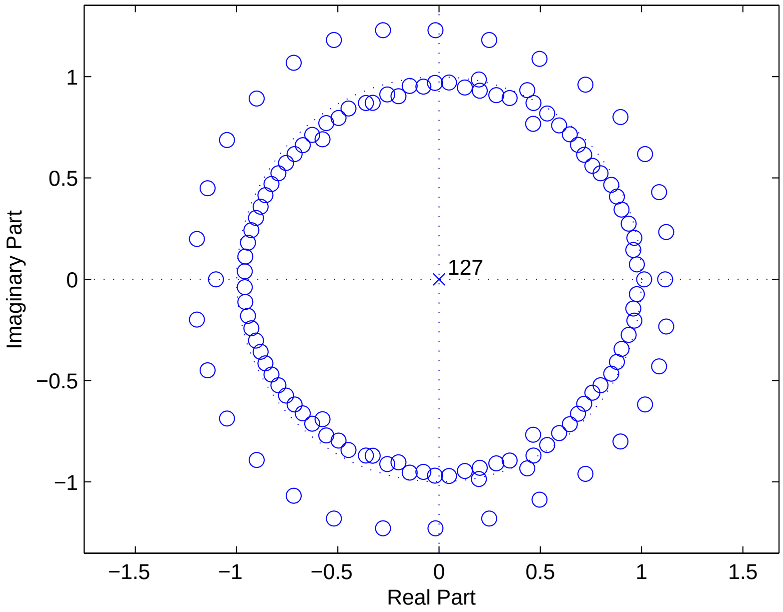


Fig. 2.1: Zero-pole plot of a diffuse-field equalized HRTF ($\phi = 30°, \theta = 0°$, length is 128 points at 44.1kHz)

The unilateral z-transform of a discrete-time signal h[n] is given by:

$$H(z) = \sum_{n=0}^{\infty} \frac{h[n]}{z^n} \tag{2.4}$$

and the inverse G(z) is

$$G(z) = \frac{1}{H(z)} \tag{2.5}$$

A mentioned before if any zero of $H(z)$ lies outside the unit circle it follows that the inverse $G(z)$ is unstable. In order to achieve a stable filter the region of convergence (ROC) of G(z) has to be chosen so that the unit circle is included. This yields an inverse g[n], that is stable but infinitely long and two-sided (it contains causal and anti-causal exponentials [5][14]). If the resulting infinite impulse response is windowed, a finite impulse response can be obtained (Eq. 2.6). This results in time aliasing but if the window includes the maximal energy of the inverse time response, the effect of time aliasing will be reduced.

$$\tilde{g}[n] = g[n] \cdot w[n], w[n] = \begin{cases} 1, L < n < M \\ 0, otherwise \end{cases} \tag{2.6}$$

where w[n] is a window function set to maximize the energy inside the window. The inverse filter $\tilde{g}[n]$ will be a finite length approximation of the true inverse. Important to note is that this approximation could be anti-causal depending on the time at which the window was set. If that is the case and a real-time implementation is desired, then $\tilde{g}[n]$ has to be delayed by m samples to make it causal.

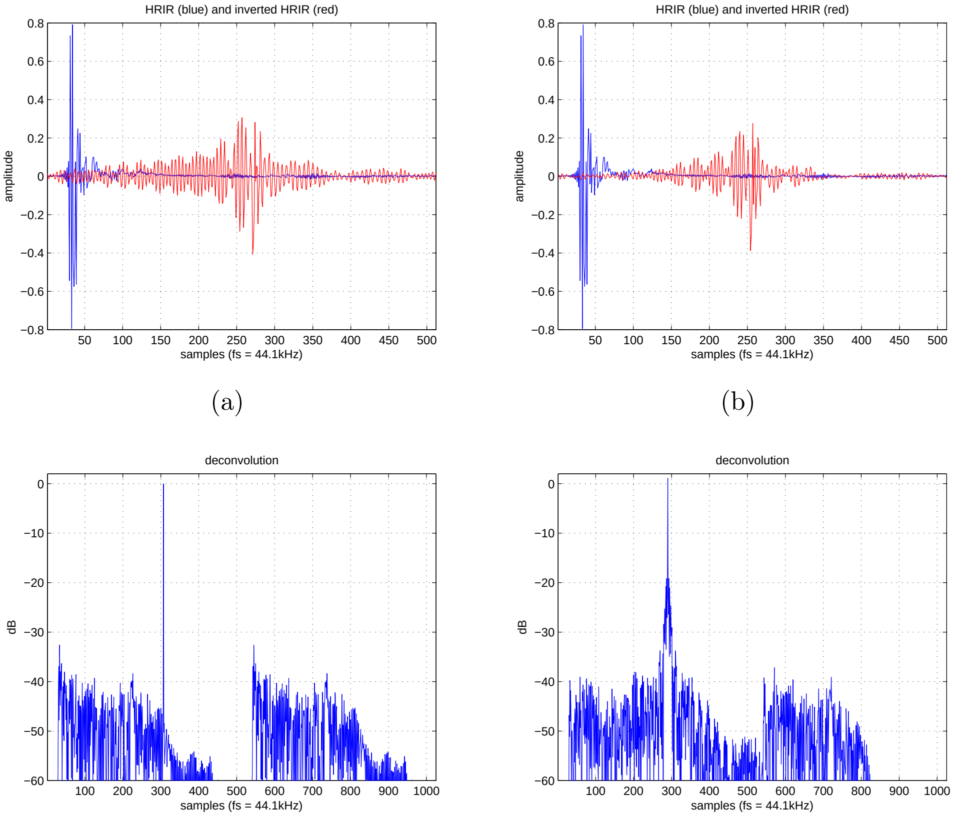
The deconvolution of h[n] and it's inverse results in

$$h[n] * \tilde{g}[n-m] = \delta[n-m] \tag{2.7}$$

Inversion using the DFT

In order to obtain a stable approximation of an non-minimum phase FIR filter, Gardner suggested to make use of the discrete Fourier transform (DFT) [5]. The concept is to compute the DFT of h[n], then invert it and limit the resulting magnitude response, then compute the inverse DFT and window the result. If necessary a delay of the time response has to be applied in order to achieve a causal filter.

The limitation is necessary to avoid excessive time aliasing after the inverse DFT operation. The windowing can be accomplished by using window functions different to the rectangular window (e.g.: Hanning, Kaiser). Finally, the longer the window is, the more accurate the approximated inverse will be (for a detailed comparison of different window lengths see [15]).



(a)

(b)

(c)

(d)

Fig. 2.2: Figures (a) and (b): HRIR (blue) and inverted HRIR (red) using the DFT inversion technique. Figures (c) and (d): the resulting deconvolution. The limitation for (a) and (c) is ±20dB and for (b) and (d) +5dB and -10dB. (HRIR:$\phi = 30°, \theta = 0°$, 512 points at 44.1kHz)

Fig. 2.2 shows a HRIR, its inverse and the deconvolutions. It can be seen that the deconvolution in Fig.2.2 (d), in which the magnitude is distinctly limited before inversion, the time aliasing effects are reduced compared to Fig. 2.2 (c), in which the limits are chosen too high.

Least square technique

The least squares techniques is a robust and efficient approximation method for solving overdetermined systems and hence is widely used [16] [15].

A filter $g[n]$ is the inverse of the filter $h[n]$. The error between an ideal inversion and the real inversion is

$$e[n] = \delta[n] - \tilde{\delta}[n] \tag{2.8}$$

where $\delta[n]$ is the ideal or desired output and $\tilde{\delta}[n]$ the actual output. The least squares technique now tries to minimize the squared error between a desired output and the real output [15].

$$J = \frac{1}{L}\sum_{n=0}^{L} e^2[n] = \frac{1}{L}\sum_{n=0}^{L}(\delta[n] - \tilde{\delta}[n])^2 \tag{2.9}$$

is the so called cost function, where N is the length of the input filter $h[n]$, M is the length of the inverse filter $\tilde{g}[n]$ and $L = N + M - 1$, the resulting deconvolution. This deterministic cost function is to be minimized.

As said before, the resulting inverse has to be delayed to make it causal. We can take the convolution of Eq.(2.7) and rewrite it in matrix form [15]

$$\begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} h[0] & 0 & \cdots & \cdots & 0 & 0 \\ h[1] & h[0] & 0 & & & \\ \vdots & h[1] & \ddots & & & \\ \vdots & \vdots & & \ddots & 0 & \\ h[N-1] & h[N-2] & \cdots & \cdots & h[0] & 0 \\ 0 & h[N-1] & \cdots & \cdots & h[1] & h[0] \\ \vdots & & \ddots & \cdots & \cdots & \vdots \\ \vdots & & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & h[N-1] & h[N-2] \end{pmatrix} \begin{pmatrix} \tilde{g}[0] \\ \tilde{g}[1] \\ \vdots \\ \vdots \\ \tilde{g}[M-1] \end{pmatrix} \tag{2.10}$$

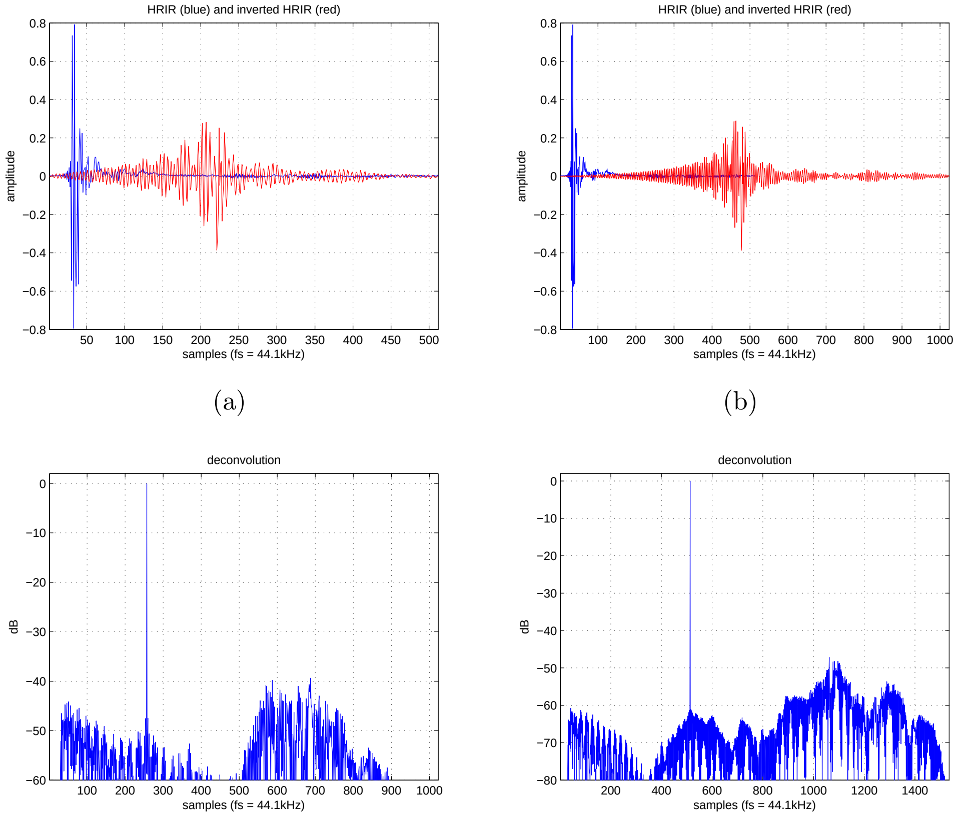
or more compact $\boldsymbol{d} = \boldsymbol{H}\tilde{\boldsymbol{g}}$, where d is the desired output vector with length L, $\boldsymbol{H}$ is a Toeplitz matrix comprising the input filter with the size of $L \times M$ and $\tilde{g}$ is the inverse filter vector with length M. Further we can build the error function

$$\boldsymbol{e} = \boldsymbol{d} - \boldsymbol{H}\tilde{\boldsymbol{g}} \tag{2.11}$$

and minimize the squared norm of the error function, $||e||^2 = e^T e$ by setting the gradient zero. We solve for $\tilde{g}[n]$ which leads to

$$\tilde{\boldsymbol{g}} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{d} \tag{2.12}$$

The expression $(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T$ is also called the pseudo-inverse of $\boldsymbol{H}$.



(a)

(b)

(c)

(d)

Fig. 2.3: Figures (a) and (b): HRIR (blue) and inverted HRIR (red) using the least-squares inversion technique. Figures (c) and (d): the resulting deconvolution. The length of the inverse filter is 512 samples in (a) and (c) and 1024 samples in (b) and (d). (HRIR: $\phi = 30°, \theta = 0°$, 512 points at 44.1kHz)

Fig.2.3 shows the results of the least-squares approximation. In Fig.2.3 (c) it can be seen that the deconvolution for a 512 samples long inverse is more accurate compared to the DFT inversion in Fig. 2.2. The peak is more pure and the artifacts are less in amplitude. Fig.2.3 (b) and (d) depict that the longer the inverse filter is, the better the deconvolution works.

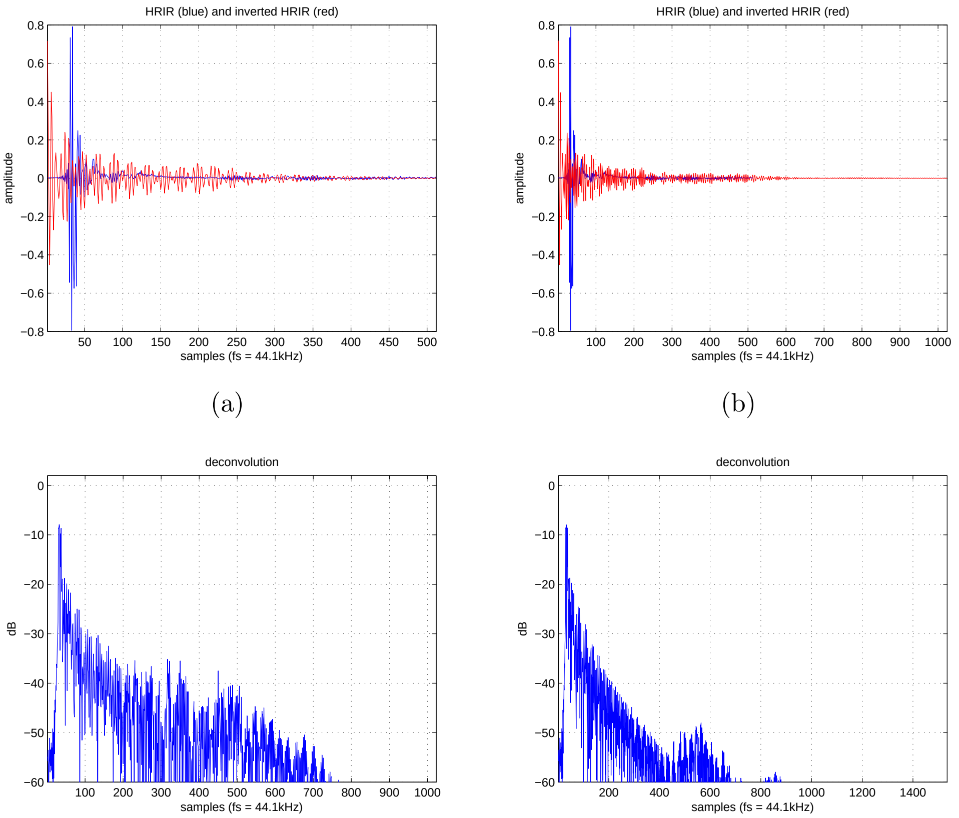
Minimum phase inversion

A HRTF is generally of minimum-phase except for a nearly frequency independent delay. For the inversion of HRTFs, Schroeder found the delay to be ignorable. The idea is to invert the magnitude response and compute the minimum phase response by the Hilbert transform of the log-magnitude spectrum [1].

If $H(j\omega)$ is the complex HRTF then the minimum phase response is

$$arg[H(j\omega)] = -\mathcal{H}\left\{log(|H(j\omega)|)\right\} \tag{2.13}$$

It can be seen that the deconvolution using the minimum-phase reconstruction of the inverse is of poor performance compared to the DFT inversion or the least-



(a)

(b)

(c)

(d)

Fig. 2.4: Figures (a) and (b): HRIR (blue) and inverted HRIR (red) using the minimum phase inversion technique. Figures (c) and (d): the resulting deconvolution. The length of the inverse filter is 512 samples in (a) and (c) and 1024 samples in (b) and (d). (HRIR: $\phi = 30°, \theta = 0°$, 512 points at 44.1kHz)

squares technique (Figs. 2.2, 2.3). An advantage though is that no delay has to be used.

Conclusions

This short comparison of inversion techniques for non minimum-phase filters outlined the differences in performance and effort. It was found that the least-squares techniques yields the best results and thus it is used in the further simulations.

It should be noted that more techniques exist. For a more detailed overview the reader is referred to [15].

Chapter 3

Implementation of CTC

This chapter reviews the different kinds of implementations of the inverse head transfer matrix and the methods used to solve the common problems.

Atal and Schroeder were the first to implement a transaural system. in order to reproduce the measured binaural room impulse responses (BRIR) of concert halls over loudspeakers, by which a more realistic impression should be achieved as by using headphones. [4][17] [18] [19]

For their implementaion they supposed a symmetric loudspeaker setup and so the head matrix and its inverse can be rewritten to

$$\boldsymbol{H} = \begin{bmatrix} S & A \\ A & S \end{bmatrix}, \boldsymbol{C} = \frac{1}{S^2 - A^2} \begin{bmatrix} S & -A \\ -A & S \end{bmatrix} \tag{3.1}$$

where S is the ipsilateral and A the contralateral HRTF in reference to the loudspeakers. Reordering leads to their implementation shown in Fig. 3.1

$$\boldsymbol{C} = \frac{\frac{1}{S}}{1 - (\frac{A}{S})^2} \begin{bmatrix} 1 & -\frac{A}{S} \\ -\frac{A}{S} & 1 \end{bmatrix}. \tag{3.2}$$

The filter $C = -\frac{A}{S}$ is said to be realizable because the delay in A is always bigger than the one in S and so the filter is causal and can be put into the feedback loop. The inversion of S was realized with sufficient zero padding.

The method proposed by Atal and Schroeder forms the basis for all the further investigations by researchers. In the following the basic methods are presented.

More theoretical considerations

Before presenting the existing methods in the following pages, a more detailed formulation of the CTC process is given.

If Eq. (1.5) is written out it yields

$$\begin{bmatrix} e_L \\ e_R \end{bmatrix} = \begin{bmatrix} H_{LL} & H_{RL} \\ H_{LR} & H_{RR} \end{bmatrix} \begin{bmatrix} \frac{H_{RR}}{D} & \frac{-H_{RL}}{D} \\ \frac{-H_{LR}}{D} & \frac{H_{LL}}{D} \end{bmatrix} \begin{bmatrix} x_L \\ x_R \end{bmatrix}. \tag{3.3}$$

The multiplication of the inverse head transfer matrix $\boldsymbol{C}$ and the input vector $\boldsymbol{x}$ result in the driving signals, the vector $\boldsymbol{y}$. In the here considered two-channel case the left driving signal is

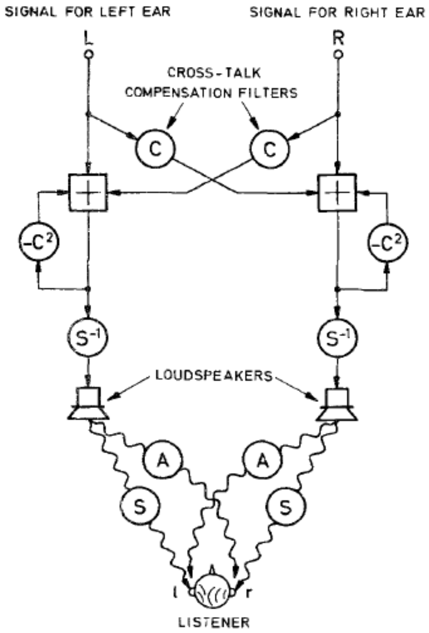


Fig. 3.1: The crosstalk canceler by Atal and Schroeder [4]. The filter C $= -\frac{A}{S}$ is not to be confused with the inverse head transfer matrix $\boldsymbol{C}$

$$y_L = \frac{x_L H_{RR} - x_R H_{RL}}{D} \tag{3.4}$$

and the right driving signal is

$$y_R = \frac{x_R H_{LL} - x_L H_{LR}}{D}. \tag{3.5}$$

Eq.(3.3) rewritten yields for the left ear

$$e_L = \frac{x_L(H_{LL}H_{RR} - H_{LR}H_{RL}) + x_R(H_{LL}H_{RL} - H_{LL}H_{RL})}{D} \tag{3.6}$$

and equivalent for the right ear.

It can be seen that the actual crosstalk cancellation is accomplished without the contribution of the determinant. It is sufficient to subtract the crosstalk path (in the case of transaural the contralateral HRTF) from the input signal of interest (x_L for left ear, x_R for right ear) to cancel the crosstalk. The determinant is necessary to cancel for the residual convolutions. It takes care of creating the *virtual headphone.*

To specify what the system filters are, we write

$$e_L = \frac{x_L(H_{LL}\tilde{H}_{RR} - \tilde{H}_{LR}H_{RL}) + x_R(H_{LL}\tilde{H}_{RL} - \tilde{H}_{LL}H_{RL})}{\tilde{H}_{LL}\tilde{H}_{RR} - \tilde{H}_{LR}\tilde{H}_{RL}} \tag{3.7}$$

where the tilde on the letters indicate the filters originating from the CTC system (approximated filters) and the ones without a tilde are the filters from the plant.

If the system filters now differ too much in terms of magnitude or phase (delay) response, the cancellation will be unsatisfying.

3.1 General topology

This topology is the direct implementation of Eq.(1.7) and is illustrated in Fig.3.2. The main problem is the inversion of the determinant, which can be treated by one of the techniques presented in chapter 2.

It can be seen that 6 convolutions are necessary. This can be reduced to 4 convolutions if the determinant is preconvolved with the HRTFs. The computational load depends on the filter type (FIR, IIR), the number of coefficients and on the implementation of the convolution (linear convolution, FFT convolution).

The general topology allows the use of symmetric speaker setups as well as asymmetric, which implies a possible use in a dynamic scenario when using head-tracking. The only constraint is that it's restricted for the use of two speakers and one listener.

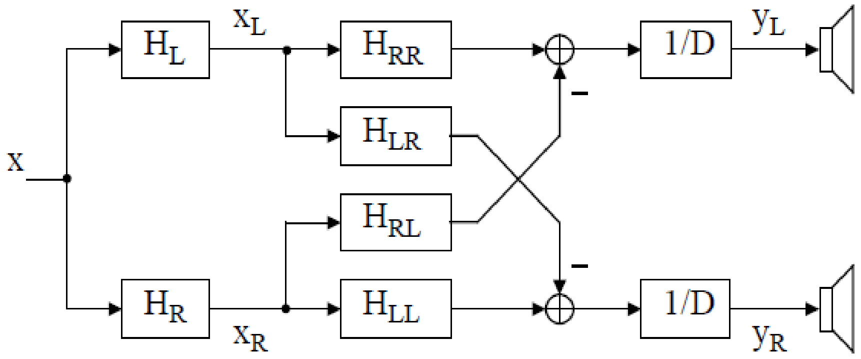


Fig. 3.2: Block diagram of the general topology including the binaural synthesis stage [5].

3.2 Shuffler topology

In 1989 Cooper ad Bauck suggested a simplification of the topology by Atal and Schroeder [1]. The original implementation can be interpreted as a lattice topology with the filters in the forward path being $\frac{S}{S^2-A^2}$ and the ones in the cross path $\frac{-A}{S^2-A^2}$. The idea is that the lattice topology is equivalent to the shuffler topology shown in Fig 3.3. The filters Σ and Δ then have to be the inverse of the sum and difference of the ipsi- and contralateral filters.

$$\Sigma = \frac{1}{S+A}, \Delta = \frac{1}{S-A} \tag{3.8}$$

This relation can also be derived from the knowledge that a symmetric square matrix is diagonalizable. A square matrix $\boldsymbol{A}$ is diagonalizable if it can be rewritten to

$$\boldsymbol{A} = \boldsymbol{PDP}^{-1} \tag{3.9}$$

where $\boldsymbol{D}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{P}$ is a nonsingular matrix containing the corresponding eigenvectors. Further the coloumns of $\boldsymbol{A}$ have to be linearly independent.

The inversion of Eq. 3.9, $\boldsymbol{P}^{-1}\boldsymbol{AP}$, results in a diagonal matrix. In our case, we suppose the head transfer matrix to be symmetric, thus it is diagonalizable.

$$\boldsymbol{H} = \boldsymbol{UEU}^{-1} \tag{3.10}$$

where $\boldsymbol{U}$ contains the eigenvectors of $\boldsymbol{H}$ in the columns and $\boldsymbol{E}$ is a diagonal matrix with the eigenvalues of $\boldsymbol{H}$ in the diagonal. The eigenvalues are the sum and difference of the ipsi- and contralateral filters and the eigenvectors can be chosen arbitrarily as long as they keep the norm equal to one. This results in a simplified expression for the inversion of $\boldsymbol{H}$:

$$\boldsymbol{H}^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} \frac{1}{S+A} & 1 \\ 1 & \frac{1}{S-A} \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{3.11}$$

The block scheme is shown in Fig. 3.3.

Further Cooper and Bauck introduced the concept of *joint minimum phase* [1] To be of joint minimum phase a set of filters has to have the same excess phase and this has to be a frequency independent delay. They found that the sum and difference filter more or less have the same excess phase. The consequence is that the excess phase has not to be considered anymore because it states just a common delay to both paths. Hence, it is sufficient to invert only the minimum phase part of the sum and difference filter.

The main advantage of this topology is that it is reduced to two convolutions but only symmetric loudspeaker setups can be used. Therefore a dynamic implementation is not possible.

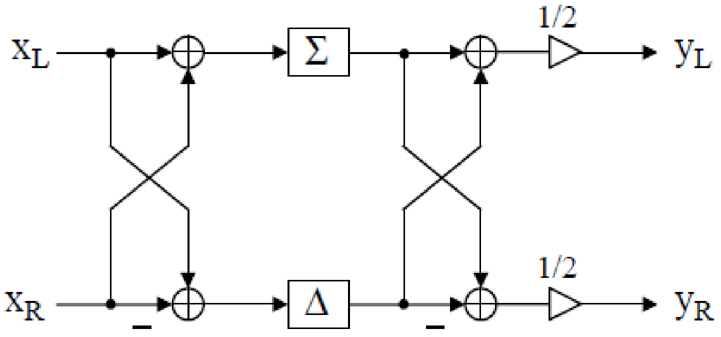


Fig. 3.3: Shuffler topology [5]. Here Σ denotes the sum filter and Δ the difference filter.

3.3 Asymmetric shuffler topology

Vandernoot suggested in his PhD thesis the asymmetric shuffler topology [15]. The structure is equvivalent to the shuffler topology, just the filters have to change. Fig. 3.4 shows the topology with the filters being

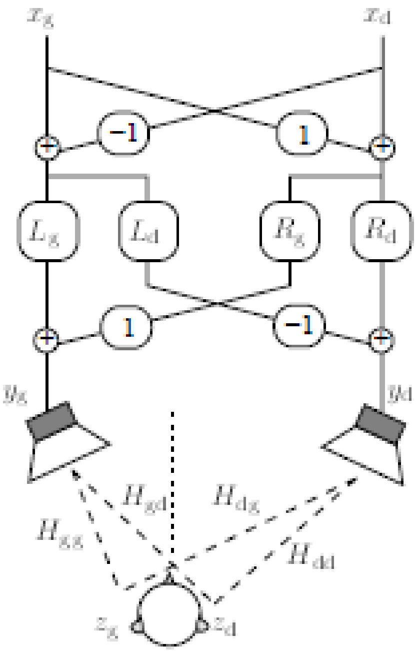


Fig. 3.4: Asymmetric shuffler topology [15].

$$\begin{aligned} H_{LL} &= \frac{H_{RR} + H_{RL}}{D} \\ H_{LR} &= \frac{H_{LR} + H_{LL}}{D} \\ H_{RL} &= \frac{H_{RR} - H_{RL}}{D} \\ H_{RR} &= \frac{-H_{LR} + H_{LL}}{D} \end{aligned} \tag{3.12}$$

and D is the determinant.

The principle of joint-minimum phase cannot be applied here. The inversion of the full phase has to be conducted.

This topology uses four convolutions and allows the use of asymmetric setups or movements of the listeners. It can be shown that the output of the structure, i.e. the speaker signals are exactly the same as the ones of the general topology.

3.4 Fast least-squares deconvolution

The principle of the least squares approach was already explained in chapter 2. Kirkeby et al. [20] proposed a fast deconvolution algorithm based on the least-squares approximation in the frequency domain using additional regularization. The block diagram of the deconvolution problem is shown in Fig. 3.5.

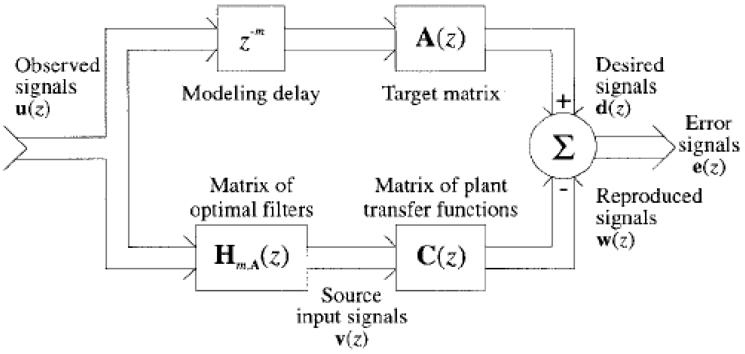


Fig. 3.5: Block diagram of the fast deconvolution problem [20]. Here $\boldsymbol{H}$ is the CTC matrix and $\boldsymbol{C}$ the head transfer matrix.

In the z-domain the given relationships are

$$\boldsymbol{y(z)} = \boldsymbol{x(z)C(z)} \tag{3.13}$$
$$\boldsymbol{w(z)} = \boldsymbol{H(z)y(z)} \tag{3.14}$$
$$\boldsymbol{d(z)} = z^{-m}\boldsymbol{A(z)x(z)} \tag{3.15}$$
$$\boldsymbol{e(z)} = \boldsymbol{d(z)} - \boldsymbol{w(z)} \tag{3.16}$$

The cost function J is built of a “performance erro” term, measuring the reproduction of the desired signals and an “effort penalty” term, which is dedicated to the total input power of the loudspeakers.

$$J(z) = e^H(z)e(z) + \beta v^H(z)v(z) \tag{3.17}$$

where β is a regularization parameter weighting the effort term [20] and H is the Hermetian transpose.

Finding the minimum of this cost function yields the CTC matrix $\boldsymbol{C}$, which minimizes the error in the frequency domain in the least squares sense.

$$\boldsymbol{C}(z) = \left[\boldsymbol{H}^T(z^{-1})\boldsymbol{H}(z) + \beta\boldsymbol{I}\right]^{-1}\boldsymbol{H}^T(z^{-1})\boldsymbol{A}(z) \tag{3.18}$$

where T denotes the transpose.

The implementation uses the FFT algorithm. Eq.(3.18) becomes

$$\boldsymbol{C}(k) = \left[\boldsymbol{H}^T(k)\boldsymbol{H}(k) + \beta\boldsymbol{I}\right]^{-1}\boldsymbol{H}^T(k)\boldsymbol{A}(k) \tag{3.19}$$

where k denotes the k-th frequency index. The regularization parameter β has to be set appropriately. In [20] a value of $\beta = 0.0001$ is suggested. Another suggestion is to set it in relation to the condition number, $\beta = \kappa(H) \cdot 10^{-2}$.

The algorithm can be summarized into a few steps:

- Calculate the N-point FFT of the relevant HRIRs
- Calculate $\boldsymbol{C}$ by applying Eq.(3.19)

- Calculate the N-point inverse FFT of $\boldsymbol{C}$
- Introduce a modeling delay by doing a cyclic shift of m samples of the resulted IRs. A good value for m is suggested by N/2.

It is important to note, that this method allows the use of any number of listeners and speakers in any (reasonably) possible geometry because the head transfer matrix is not restricted to be a square matrix. A generalized theory for that is given in [6] and [21].

In the two speaker, one listener scenario, four convolutions are necessary. If the algorithm is implemented dynamically, the speed of the calculation is determined by the FFT block size.

3.5 Recursive topology

The use of recursive topologies was already suggested by Iwahara and Mori in 1978 [22]. The topology can be seen in Fig. 3.6. The cross-coupled filters contain the interaural transfer functions (ITF), which is the ratio of the contralateral HRTF to the ipsilateral HRTF

$$ITF = \frac{H_c}{H_i}. \tag{3.20}$$

This topology can be derived by step-by-step analysis of the crosstalk process. A formulation of the iterative crosstalk process is shown in the table below (all in frequency domain). The parentheses in the table serve as visual support indicating time steps. Each column adds to the next one (sum sign is left out for visual reason). The time delay from the loudspeakers to the ears is ignored for simplification. Only the interaural time difference (ITD) constitutes temporal order.

$$\tau_N - \tau_{N-1} = ITD \tag{3.21}$$

where N states a time step in the process and it extends to infinity. In the further discussions the plant HRTFs are renamed for simplification

$$H_{LL} = A, \quad H_{LR} = B, \quad H_{RL} = C, \quad H_{RR} = D \tag{3.22}$$

The problem set is the following. We have a system with two inputs (x_L, x_R) and two outputs (e_L, e_R). The inputs are passed through a network of filters and result in y_L and y_R. This network is to be identified. In between the input and output there is the plant. The input x_L is carrying the signal of interest and is supposed to reach the left ear only. The input x_R is zero and so should be the right ear's output e_L.

In order to compensate for the direct path HRTF, in the first time step y_L is build by the division of x_L and H_{LL}. Therefore, the left ear, e_L, receives directly x_L and the right ear receives y_L a time step later and additionally convolved with the cross path HRTF H_{LR}.

To compensate for what the right ear receives, we add the opposite of the input to e_L by sending it out of y_R. Again, this compensating signal will also reach the left ear and this process is repeated infinitely.

	τ_0	τ_1	τ_2	τ_3	$\dots$
$y_L =$	$\frac{x_L}{A^{(0)}}$	0	$x_L \frac{B^{(1)}}{A^{(1)}} \frac{C^{(2)}}{D^{(2)}} \frac{1}{A^{(2)}}$	0	$\dots$
$y_R =$	0	$-x_L \frac{B^{(1)}}{A^{(1)}} \frac{1}{D^{(1)}}$	0	$-x_L \frac{B^{(1)}}{A^{(1)}} \frac{C^{(2)}}{C^{(2)}} \frac{B^{(3)}}{A^{(3)}} \frac{1}{D^{(3)}}$	$\dots$
$e_L =$	x_L	0	$-x_L \frac{B^{(1)}}{A^{(1)}} \frac{C^{(2)}}{D^{(1)}} + y_L^{(2)} A^{(2)}$	0	$\dots$
$e_R =$	0	$x_L \frac{B^{(1)}}{A^{(0)}} + y_R^{(1)} D^{(1)}$	0	$x_L \frac{B^{(1)}}{A^{(1)}} \frac{C^{(2)}}{D^{(2)}} \frac{B^{(3)}}{A^{(2)}} + y_R^{(3)} D^{(3)}$	$\dots$

It can be seen that the expressions, which have to be added to cancel the undesired crosstalk are the same in each time step ($\frac{B}{A}\frac{1}{D}$). Therefore the process is recursive. Substituting the ratio of contra- and ipsilateral HRTF by the ITF and returning to standard notation leads to the topology shown in Fig. 3.6.

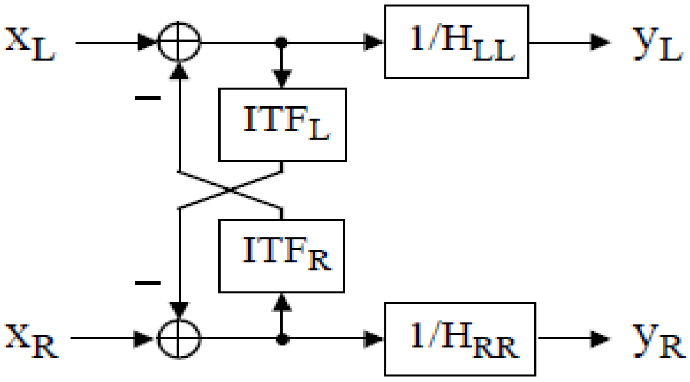


Fig. 3.6: Block diagram of the recursive topology [5].

This topology is very intuitive because the ITF can be seen as the prediction of crosstalk. If it is put into a cross coupled feedback loop also higher-order crosstalk is cancelled for.

The feedback loop of the CTC part can be written as a power series of ITFs

$$1 + ITF + ITF^2 + ITF^3 + \cdots = \frac{1}{1 - ITF} \tag{3.23}$$

This recursive structure makes clear that absolute values equal to one have to be avoided or the system will be instable. Due to numerical inaccuracies also absolute values close to one have to be avoided.

It has been found in this work, that this topology is instable for all speaker setups. Further approximations must be found.

A commercial system using a recursive topology is the *Ambiophonics* system [23]. The plant is modeled by simple attenuation and delay and the CTC process is calculated only in the frequencies between 250Hz-5kHz [24]. Therefore the instability problems are avoided.

Chapter 4

Objective evaluation

This chapter aims for the objective evaluation of transaural systems in two ways. First an analysis of the plants conditioning is made and second, free-field conditions are simulated and the amount of CTC as a function of frequency is calculated in a two speaker - two ear setup.

It has been found that binaural synthesis gives the best results if the set of HRTFs used for the synthesis match the one of the listener. This applies equally to transaural audio [25] and means that the following evaluation cannot be seen as representative for the whole population. The set of HRTFs used is taken out of IRCAMs Listen database [26]. One subject is chosen for most of the analysis and plots made (subject 1066, not online available); if not it is explicitly indicated. The impulse responses comprises 512 samples and they are diffuse-field equalized. Fig. 4.1 shows the magnitude spectra as a function of azimuth for the horizontal plane.

4.1 Conditioning of the head transfer matrix

The condition number is a measures of how well numerically a problem is conditioned. The linear equation system of $b = \boldsymbol{A}x$, can be well-conditioned or ill-conditioned, which corresponds to a small condition number and a high condition number. It can also be interpreted as a measure of how strongly an error in b influences the result in x. If the problem is well-conditioned, a small error in b results in a small error in x and vice versa. So it is a measure of how accurate the solution will be.

The condition number is defined as the ratio of the norm of the relative error in Δx to the norm of the relative error in Δb.

$$\frac{||A^{-1}\Delta b||/||A^{-1}b||}{||\Delta b||/||b||} \tag{4.1}$$

Rewriting then yields

$$\kappa(A) = ||A^{-1}|| \cdot ||A|| \tag{4.2}$$

It can easily be seen that the condition number is dependent on the norm. A common norm is the L_2-norm. Hence, the condition number is the ratio of the

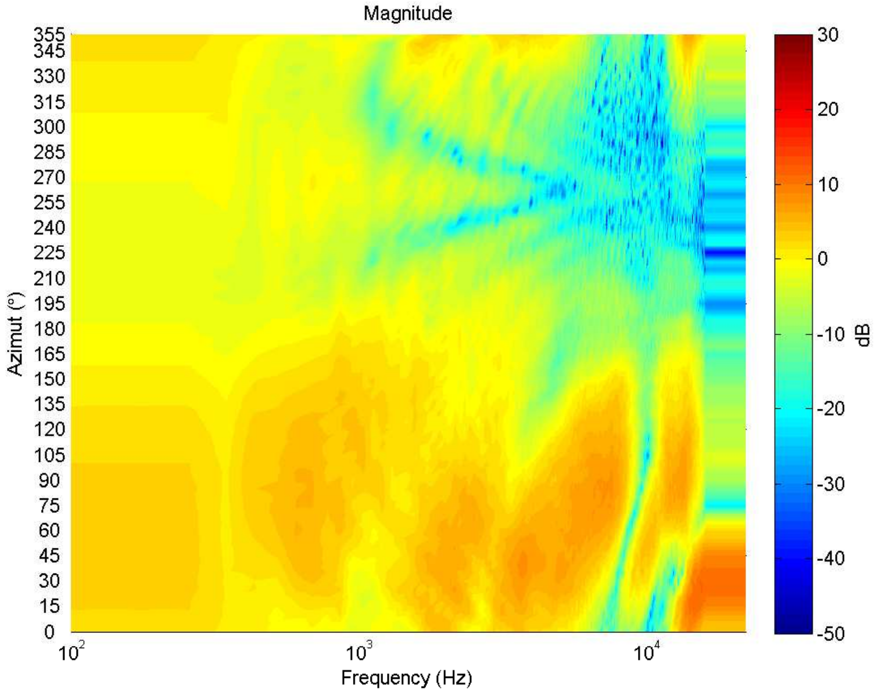


Fig. 4.1: Magnitude spectra as a function of horizontal azimuth angles (Subject 1066 of the Listen database).

biggest to the smallest singular value.

$$\kappa(A) = \frac{\sigma_{max}(A)}{\sigma_{min}(A)} \tag{4.3}$$

The analysis of the conditioning of the head transfer matrix can be found in literature [8][11][12]. Nelson et al. [12] have, for instance, shown that the condition number not only depends on frequency but also on the spacing of the loudspeakers. Takeuchi suggested [8] the “Optimal source distribution”. This system intends a continuous distribution of transducers in order to minimize the condition number for all frequencies. A practical solution comprises a loudspeaker system where three loudspeakers are positioned at ±90° for CTC at low frequencies, at ±16° for mid frequencies and at ±3.1° for high frequencies [13]. This system minimizes the condition number approximately over the whole frequency range which means that the inversion yields less errors.

4.1.1 Analysis

In Fig. 4.2 the condition numbers as a function of frequency and source span is plotted (only for symmetric setups). It can be seen that $\boldsymbol{H}$ is overall ill-conditioned for low frequencies. Moreover, no source span exists where the head transfer matrix is well-conditioned for all frequencies. As Takeuchi et al. stated in [27] “the optimal source span must vary as a function of frequency''. Hence, they proposed the

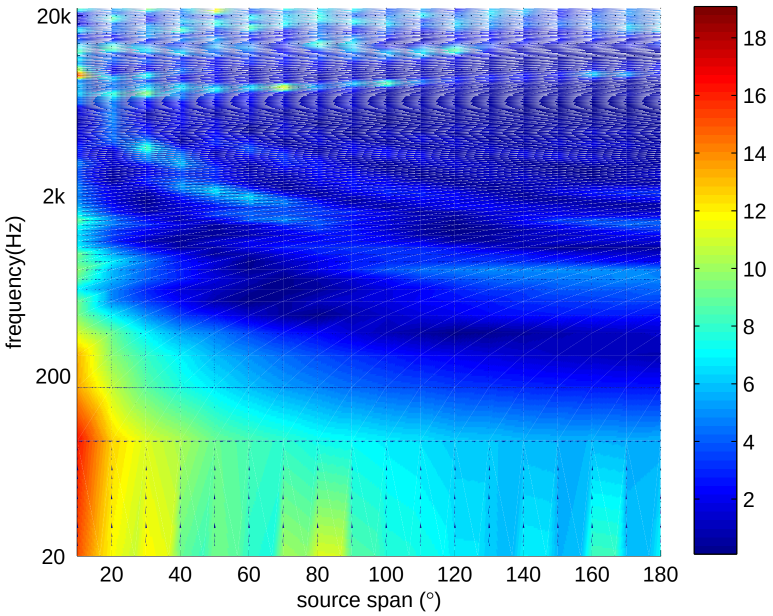


Fig. 4.2: Condition Number as a function of frequency (log) and source span (cp. [27]).

optimal source distribution system (OSD). However, this system requires transducers at different locations for different frequencies, based on a continuous function of frequency. A practical solution was found by using a 3-way system described in [13].

In this work an alternative interpretation of the above results is proposed. This leads to another system design method, which is based on an average best-fit. The idea is to compute the condition numbers averaged over frequency for each speaker span. The span providing the average best-fit is proposed to be of better performance than other source spans. It was found that the spacings between 130° and 180° provide the lowest condition numbers.

In Fig. 4.3 the frequency averages using the HRTFs of three different subjects is shown. It can be seen that the minima are different for different subjects and therefore are individual. Fig. 4.4 depicts that also the separation into 3 sub-bands, as suggested in [27], still leads to the above conclusion.

It should be stated that, in contrast to the OSD system, the proposed method does not yield a system that is well-conditioned over all frequencies. However, it potentially leads to a system that provides sufficient performance in practice and can be used by conventional transducer systems. A psychoacoustic evaluation is necessary.

4.2 Free-field validation

In order to evaluate the CTC of the topologies for different topologies, a free-field simulation was made. Therefore HRTFs measured in an anechoic chamber

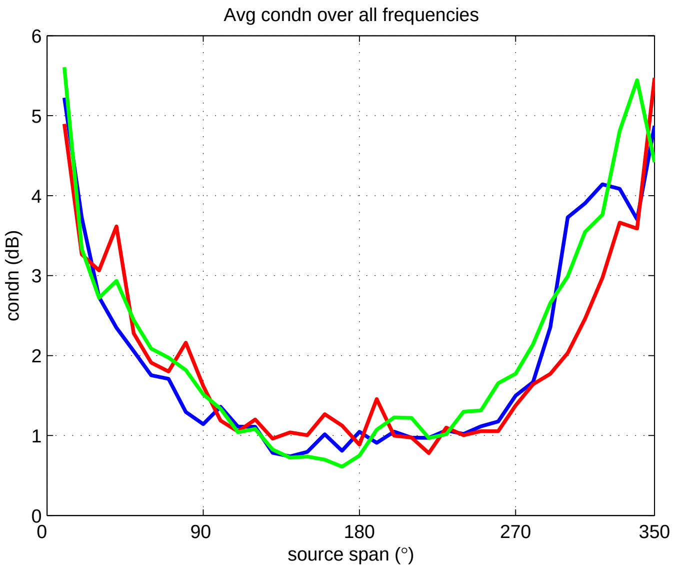


Fig. 4.3: Condition number averaged over frequency for three different head transfer matrices (subjects). Subject 1, blue; Subject 2, red; Subject 3, green.

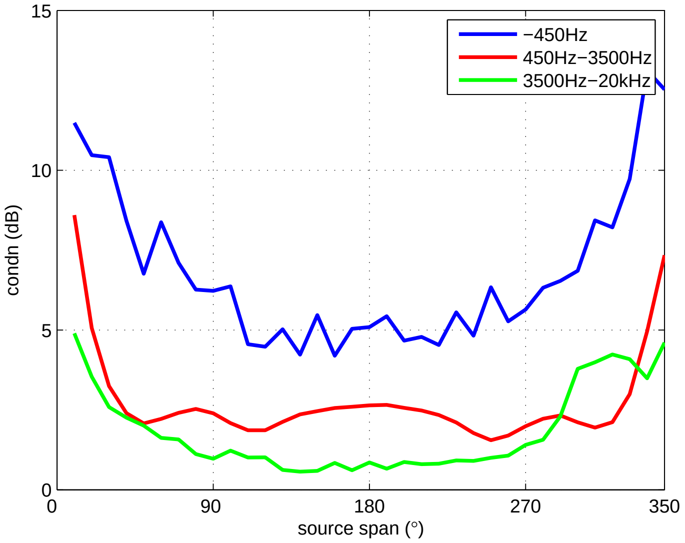


Fig. 4.4: Condition number averaged over three different frequency bands for one head transfer matrix (subject).

are used. Consequently the influences of loudspeakers and room acoustics are avoided. The evaluation has been limited to symmetric setups but can be easily extended to asymmetric setups. The input signal to the CTC network consists of a Dirac impulse on the left side and a null signal on the right side.

$$\begin{aligned} x_L(t) &= \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{otherwise} \end{cases} \\ x_R(t) &= 0 \end{aligned} \tag{4.4}$$

This yields the impulse response of the system which is further convolved with the plant HRIRs in order to compute the resulting ear signals. On the left ear a perfect reconstruction of the impulse (flat spectrum) is desired and on the right ear the signal should be cancelled.

Generally speaking CTC refers to a system inversion; the matrix of HRTFs prevalent in the plant has to be inverted. The different topologies try to achieve this in different ways (described in Ch. 3). In the case of the general topology and the asymmetric shuffler topology the determinant D of the matrix $\boldsymbol{H}$ has to be inverted. Because D is generally not of minimum-phase there is no exact solution. The inversion is achieved by a technique which yields the optimum in the sense of least-squares (see Ch. 2). The asymmetric shuffler topology has been excluded in this comparison because it can be shown that the output of the topology is equal to the output of the general topology.

The fast least-squares deconvolution also uses least-squares approximation. The difference is that it is computed in the frequency domain and additional regularization of the effort necessary for the inversion is used. It was found that a constant value for the regularization parameter for different positions of the loudspeakers is not the best choice. This has not been further investigated and $\beta = 0.0001$ was chosen as suggested by Kirkeby [20]. A parameter dependent on the condition number may be used.

In case of the symmetric shuffler topology the principle of *joint minimum-phase* is applied (described in Ch. 3), which says that only an inversion of the minimum-phase part of the summation and difference filter is necessary.

All filter computations are done in the frequency domain using 1024 samples. Only the least squares inversion is computed in time domain.

Requirements

In the best case the system achieves perfect CTC. However, contingent on the application the CTC does not have to be perfect. In the context of transaural audio binaural signals are reproduced via a CTC network and loudspeakers. In other words the minimum CTC necessary is determined by the dynamics of the binaural signal. Therefore, the system has to be able to reproduce the full dynamics of the binaural synthesis; otherwise, the performance is degraded.

More specifically if the interaural transfer funtions (ITF) are analyzed it can be seen how much CTC is necessary at different frequencies. Fig. 4.5 shows the magnitudes of the ITFs for several incident directions.

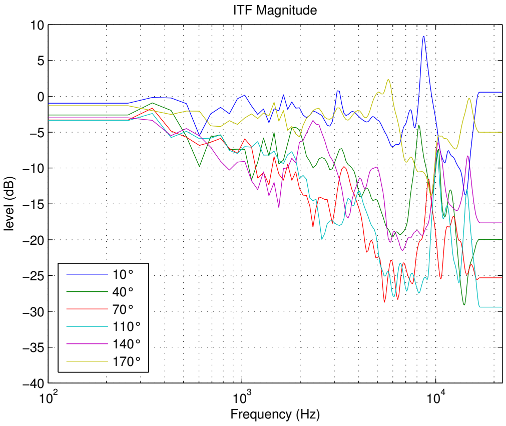


Fig. 4.5: Magnitudes of some ITFs.

It can be seen that the maximum CTC cancellation is necessary at high frequencies where the ITF has values of about -28dB. Hence it can be said the the crosstalk canceler has to be able to provide approximately 30dB CTC.

Here the idea of band-limitation proposed in literature becomes self-evident [5]. If the process of CTC is restricted to a certain frequency range the dynamics to reproduce are much less.

Recursive Topology

As mentioned the aim of the remainder is the objective comparison of the topologies for CTC described in the former chapter. The recursive topology is here excluded due to stability problems. The CTC in this topology is created by the infinite series of ITFs.

$$1 + ITF(\omega) + ITF(\omega)^2 + \ldots = \frac{1}{1 - ITF(\omega)} \tag{4.5}$$

Theoretically no value of $abs(ITF(\omega))$ must be 1 or even larger otherwise, the system becomes instable. If a value is close to one, instability can already occur due to accumulation errors and measurement errors.

In Fig. 4.6 the ITFs of all angles are plotted. Here we can find that there is a frequency range (8-9kHz) that is close to 1 or even larger over all angles. Also, the low frequency range is partly close to one.

A workaround is to again use band-limitation as described in [5]. Here the recursive topology is excluded in the further considerations.

4.2.1 Analysis

The following plots show the comparison of four different topologies concerning the amount of CTC and concerning different source setups (different plants). The colors represent the topologies as:

- General topology, blue

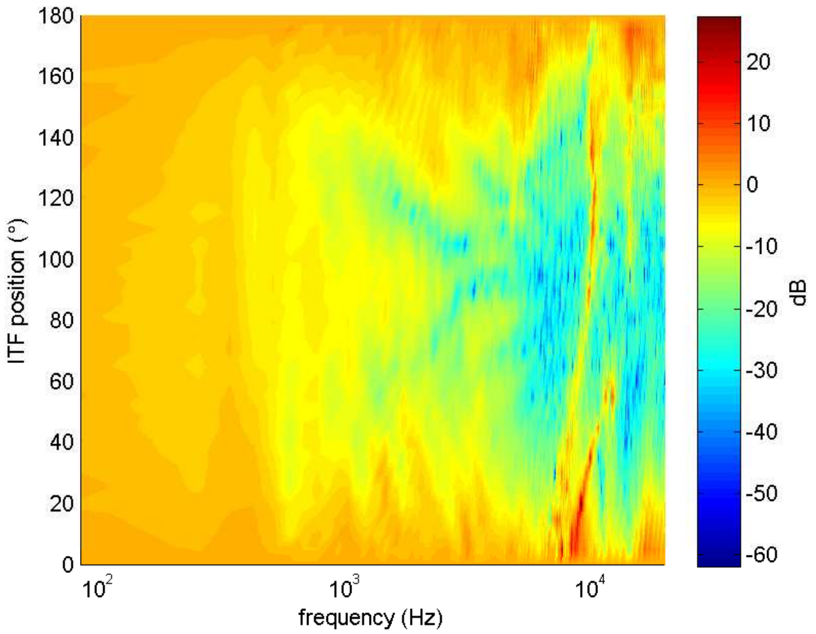


Fig. 4.6: Magnitudes of the inter-aural transfer functions over all angles.

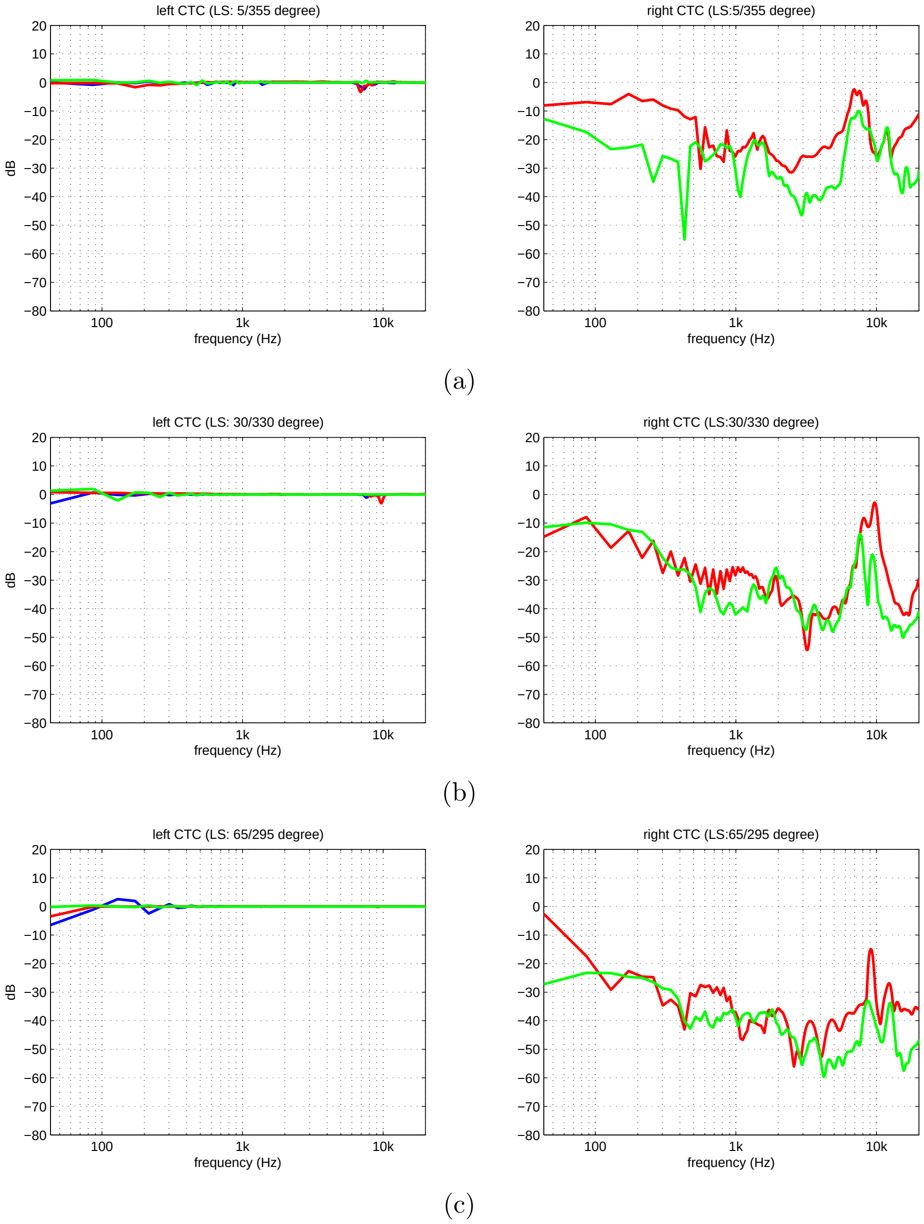
- Symmetric shuffler topology, red
- Fast least-squares deconvolution, green

It can clearly be seen that the general topology is missing on the right side of the plot. That is because it cancels crosstalk perfectly; hence the line is located beyond the scales.

Looking at the plot for the ±5° setup it can be seen that the direct path reconstruction is correct except for some around 7kHz where the cancelation fails for about 4dB. The CTC curves on the right side plot do not show a sufficient cancellation in the low and high frequency range. The shuffler topology has solely CTC in between 20dB and 30dB in the range from 600Hz-4kHz. Basically the fast least-squares technique has very high CTC in between 1.8kHz and 5.5kHz and reaches values of about 40dB.

The CTC plot of the ±30° setup shows a similar pattern. The direct paths again deviate from perfect flatness only around 9-10kHz, and the fast least-squares cancellation is nearly perfectly flat. The CTC curves show a quite similar shape. In between 300Hz-6kHz high CTC being in average 30dB with peaks up to 50dB can be seen. The CTC in the high and low frequency range is generally low again.

The last figure shows the curves for the ±65° setup. Now all the topologies have nearly perfect direct path reconstruction except for the very low frequencies. Both topologies show CTC above 30dB in between 400Hz and 7kHz.



(a)

(b)

(c)

Fig. 4.7: Magnitude spectra of the acoustic paths between left speaker and left ear (left) and left speaker and right ear (right). Figure (a) corresponds to a source span of ±5°, (b) to ±30° and (c) to ±65°.

4.2.2 Conclusions

In this work the performance of different CTC topologies was evaluated by simulating a free-field and by computing the CTC as a function of frequency and across different source spans.

It was found that there is a high dependence of the performance on the source span. The results of the analysis of the condition numbers, which suggests that wider source spans are better conditioned , thus providing a less erroneous inversion, could have been confirmed. It was shown that the CTC is generally higher for wide source spans than for narrow source spans. The only exception is that the direct path reconstruction of the general form in the very low frequencies gets worse the wider the source span is. This is unexpected because the condition number analysis showed that the low frequency conditioning for wider source spans is much better than for more narrow ones. This question was not solved here.

The general topology was found to cancel crosstalk perfectly over all source spans. This is because there is no modeling of the HRTFs used to compute the CTC filters. The performance of the fast least-squares deconvolution and the symmetric shuffler form are quite similar except for a narrow source span where the fast least-squares deconvolution yields better results. This is remarkable because the symmetric shuffler form uses only two filters.

It should be mentioned that this evaluation is insufficient in order to be able to state real objective results. A better evaluation would measure the reproduction error of a HRTF concerning magnitude and ITD (cp. [15]). Further a hearing model could be applied to evaluate the accuracy of the spatial perception. All this was beyond the scope of this work.

Chapter 5

Subjective evaluation

In the frame of this work, a preliminary listening test was conducted. 3 different setups were tested all using the general form implementation. The question was if the results of the analysis of the conditioning of different setups can be subjectively confirmed.

The test setup was comprised of 3 loudspeaker pairs, placed symmetricly with span angles of ±10°,±30° and ±70° in reference to the listener. In each step a direct comparison of two setups was made. The test signals were binaural signals comprised of white noise slowly rotated across 360°. The subjects were asked to indicate their preference concerning different aspects:

- general exernalization
- persistence of rotation
- sound coloration

4 subjects were tested using individual HRTFs for the binaural synthesis and the CTC and 2 subjects were tested using the KEMAR HRTFs.

No statistical validation will be presented here; the listening test states just a preliminary test.

It could be observed was that a clear preference in all aspects for the ±30° and the ±70° setup compared to the ±10° setup exists. The sound coloration for the ±10° is too high and the externalization is insufficient. The ±70° setup was declared as the best setup in view of the asked properties.

These results motivate further investigation.

Chapter 6

Conclusions

Transaural audio is a three-dimensional sound spatialization technique which is capable of reproducing binaural signals over loudspeakers. It is based on the cancellation of the acoustic paths occurring between loudspeakers and the listeners ears.

There are two categories of CTC. Static and dynamic CTC. In a static CTC scenario the transfer function between the listeners ears and the loudspeakers are not changing, whereas in a dynamic scenario this is the case. In this report only static scenarios using two loudspeakers are discussed. For the dynamic and/or adaptive techniques the reader is referred to [28][2][9][29][30][31].

A critical point in the design of a static transaural system is the position of the transducers. On the one side the *Stereo Dipole* was suggested [7] which puts loudspeakers closely together in front of the listener. The reproduced soundfield becomes less complicated and further results in a system being more robust to small movements of the listener. However, the CTC for low frequencies needs very high output levels and this leads to sound coloration effects.

On the other side the OSD system was proposed [8]. This system suggests a continuous distribution of sound source in space. Sources close together in front of the listener reproduce high frequencies and sound sources at $\pm 90^\circ$ with respect to the listener reproduce low frequencies. This source distribution provides a low condition number for all frequencies and the full dynamic range is preserved. In this work an alternative was proposed. It was found that a configuration of loudspeakers can be found minimizing the frequency average of the condition number. The advantage is that two conventional loudspeakers can be used and it still leads to a well conditioned system with less sound coloration. However, errors will occur especially at some high frequency ranges.

Further the basic algorithms for CTC where presented and evaluated. It was found that the general topology algorithm yields the best results in a free-field simulation. The fast least squares deconvolution and the shuffler topology gave generally similar results. This is remarkable because the shuffler topology uses only two filters compared to four filters in the case of the fast least-squares deconvolution. It seems that a constant regularization parameter is not the best choice and a frequency dependent regularization might perform better.

Alternative approaches to the presented algorithms exist and the interested reader is referred to [32][33][10]. Especially multichannel solutions are worth mentioning where a higher number of transducers than ears lead to exact solutions and the system allows for more than one listener [6][21].

The last point to mention in the context of transaural audio is the individualization of the system. In [25] it was found that the performance deteriorates when the HRTFs used for calculating the CTC filters mismatch the ones of the listener. In an ideal case the impulse responses of the prevalent plant are measured and then used to cancel crosstalk. This means that the room impulse response may be included in the measurement and therefore the inversion of filters will become more difficult but the performance would improve. Adaptive techniques are here of interest.

To summarize, the main challenges and issues in a static transaural system are

- sweet spot/robustness against listener movements
- sound coloration due to imperfect inversions
- imperfect loudspeaker frequency response
- distortions through room reflections

However, listening to a transaural systems was enjoyable and motivating for further work on the development of transaural audio.

Bibliography

[1] D. H. Cooper and J. L. Bauck, “Prospects for transaural recording,” *Journal of the Audio Engineering Society*, vol. 37, no. 1–2, pp. 3–19, 1989.

[2] W. Gardner, “Transaural 3-d audio,” 1995.

[3] B. B. Bauer, “Stereophonic earphones and binaural loudspeakers,” *J. Audio Eng. Soc*, vol. 9, no. 2, pp. 148–151, 1961. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=471

[4] M. R. Schroeder, “Computer models for concert hall acoustics,” *Am. J. Phys.*, vol. 41, pp. 461–471, 1973.

[5] W. Gardner, “3-d audio using loudspeakers(phd),” Ph.D. dissertation, MIT Media Lab, 1997.

[6] J. L. Bauck and D. H. Cooper, “Generalized transaural stereo and applications,” *Journal of the Audio Engineering Society*, 1996.

[7] O. Kirkeby and P. A. Nelson, “The ”stereo dipole” - a virtual source imaging system using two closely spaced loudspeakers,” *Journal of the Audio Engineering Society*, 1998.

[8] T. Takeuchi and P. A. Nelson, “Optimal source distribution for binaural synthesis over loudspeakers,” *J. Acoust. Soc. Am.*, 2000.

[9] D. Menzel, H. Wittek, G. Theile, and H. Fastl, “The binaural sky: A virtual headphone for binaural room synthesis,” *Tonmeistersymposium*, 2005.

[10] M. Guldenschuh and A. Sontacchi, “Transaural stereo in a beamforming approach,” *Proc. of the 12th Int. Conference on Digital Audio Effects (DAFx-09), Como, Italy*, 2009.

[11] D. B. Ward and G. W. Elko, “Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation,” *Signal Processing Letters, IEEE*, vol. 6, no. 5, pp. 106–108, 1999.

[12] P. A. Nelson and J. F. W. Rose, “Errors in two-point sound reproduction,” *J. Acoust. Soc. Am.*, vol. 118, no. 1, pp. 193–204, 2005.

[13] T. Takeuchi and P. A. Nelson, “Subjective and objective evaluation of the optimal source distribution for virtual acoustic imaging,” *Journal of the Audio Engineering Society*, vol. 55, no. 11, 2007.

[14] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time signal processing (2nd ed.).* Upper Saddle River, NJ, USA: Prentice-Hall, Inc, 1999.

[15] G. Vandernoot, "Caracterisation et optimisation de la restitution haute-fidelite en vehicule," Ph.D. dissertation, UNIVERSITE PARIS 6, 2001.

[16] J. N. Mourjopoulos, "Digital equalization of room acoustics," *J. Acoust. Soc. Am.*, vol. 42, no. 11, pp. 884–900, 1994.

[17] M. R. Schroeder and B. S. Atal, "Computer simulation of sound transmission in rooms," *Proceedings of the IEEE*, vol. 51, no. 3, pp. 536–537, 1963.

[18] M. R. Schroeder, "Digital simulation of sound transmission in reverberant spaces," *The Journal of the Acoustical Society of America*, vol. 45, no. 1, pp. 303–303, 1969. [Online]. Available: http://link.aip.org/link/?JAS/45/303/4

[19] ——, "Models of hearing," *Proceedings of the IEEE*, vol. 63, no. 9, pp. 1332–1350, 1975.

[20] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization," *IEEE Transactions on speech and audio processing*, vol. 6, no. 2, pp. 189–194, 1998.

[21] Y. Huang, J. Benesty, and J. Chen, "On crosstalk cancellation and equalization with multiple loudspeakers for 3-d sound reproduction," *Signal Processing Letters, IEEE*, vol. 14, no. 10, pp. 649–652, 2007.

[22] M. Iwahara and T. Mori, "Stereophonic sound reproduction system," *US Patent*, 1978.

[23] "www.ambiophonics.org."

[24] R. Glasgal, "360 ° localization via 4.x race processing," *AES Convention Paper*, vol. Presented at the 123rd Convention, 2007.

[25] M. A. Akeroyd, J. Chambers, D. Bullock, A. R. Palmer, A. Q. Summerfield, P. A. Nelson, and S. Gatehouse, "The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics," *The Journal of the Acoustical Society of America*, vol. 121, no. 2, pp. 1056–1069, 2007. [Online]. Available: http://link.aip.org/link/?JAS/121/1056/1

[26] O. Warusfel, "Listen hrtf database," 2002. [Online]. Available: http://recherche.ircam.fr/equipes/salles/listen/

[27] T. Takeuchi and P. A. Nelson, "Optimal source distribution for virtual acoustic imaging," *ISVR Technical Report*, no. 288, 2000.

[28] P. A. Nelson, H. Hamada, and S. J. Elliott, "Adaptive inverse filters for stereophonic sound reproduction," *Signal Processing, IEEE Transactions on*, vol. 40, no. 7, pp. 1621–1632, 1992.

[29] T. Lentz, "Dynamic crosstalk cancellation for binaural synthesis in virtual reality environments," *Journal of the Audio Engineering Society*, 2006.

[30] S. Cecchi, L. Palestini, P. Paolo, F. Piazza, and F. Bettarelli, "Sub-band adaptive crosstalk cancellation: A novel approach for immersive audio," *Journal of the Audio Engineering Society*, 2008.

[31] F. Völk, T. Musialik, and H. Fastl, "Crosstalk cancellation between phantom sources," *Journal of the Audio Engineering Society*, 2009.

[32] H. I. K. Rao and V. J. Mathews, "Inverse filter design using minimax approximation techniques for 3-d audio," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings 2006 IEEE International Conference on*, vol. 5, 2006, pp. V–V.

[33] J. Lee, Y.-C. Park, and D.-H. Youn, "Robust crosstalk cancellation based on energy-based control," *Journal of the Audio Engineering Society*, 2008.