University of Erlangen-Nuremberg

Chair of Multimedia Communication and Signal Processing

Prof. Dr.-Ing. Walter Kellermann

SIM PROJECT

Wave-Domain Decorrelation of Loudspeaker Signals for Wave Field Synthesis

Christian Hümmer, B.Sc.

July 2012

Professor: Prof. Dr.-Ing. Walter Kellermann

Supervisor: Dipl.-Ing. Martin Schneider

Contents

1 Abstract

The identification of a loudspeaker-enclosure-microphone system (LEMS) is necessary in a variety of applications like acoustic echo cancellation (AEC) or adaptive listening room equalization (LRE). This task emerges as being very challenging especially in scenarios with many reproduction channels. Due to the so called "non-uniqueness" problem, a strong cross-correlation between the loudspeaker signals prevents an unambiguous system identification. As a result of this, the decorrelation emerged as important issue. Various solutions for have been proposed, whereas these concepts show decisive disadvantages like a significant decrease in hearing quality.
In this thesis we introduce a new approach to decorrelate the loudspeaker signals in a multichannel AEC scenario. The wave field synthesis (WFS) is used with a concentric circular loudspeaker array of 48 elements. The number of reproduction channels of this LEMS is build with a recording system of 10 microphones. We follow the idea to decrease the loudspeaker cross-correlations with a perceptually acceptable wave field rotation. The corresponding model based approach is build on a set of basis function that are used for a transformation into a spatial continuum. In this so called free-field description for wave-domain adaptive filtering the loudspeaker signals are weighted with a time-varying complex exponent. Subsequent to the re-transformation into the original domain, the cross-correlation between the channels is decreased before the rotated acoustic wave field is generated. The evaluation of the system identification is realized with the generalized frequency-domain adaptive filtering (GFDAF) algorithm and quantized with the two parameters of normalized misalignment (NMA) and echo return loss enhancement (ERLE).
This thesis is structured into five parts. In the second chapter, the "non-uniqueness" problem is introduced for a stereophonic AEC scenario to give an intuition for the challenges of a multichannel system identification. The third part describes the wave-domain time-varying filtering with the derivation of the corresponding transformation equations. Furthermore we introduce the GFDAF algorithm and its update equation in Chapter 4. With the implementation of the wave-domain time-varying filtering simulation results for various wave field rotations are included in the fifth part. Afterwards a hearing test is invented and the outcomes for 16 listeners are evaluated in Chapter 6 to determine the parameters of a perceptually acceptable filter. Finally we make a conclusion for the system identification and summarize the results.
In the appendix mathematical preliminaries and notations are listed as well as an overview of all integrated Matlab files and the questionnaire of the hearing test.

2 Motivation

In this section challenges and some previously proposed solutions of a system identification in the case of stereophonic AEC are presented.
We consider the scenario shown in Figure 2.1 with linear and time-invariant finite impulse response (FIR) filters [1]. A single speaker in the transmission room generates the input signal. The acoustic paths from this source to the two microphones are captured by the impulse response vectors

$$\mathbf{g}_i = \left[g_{i,0}, g_{i,1}, ..., g_{i,M-1}\right]^T \qquad \text{with} \qquad i = 1, 2. \tag{2.1}$$

The corresponding coefficients are denoted as $g_{i,\kappa}$ with time instance $\kappa = 0, ..., M-1$. Both impulse response vectors are of length M.

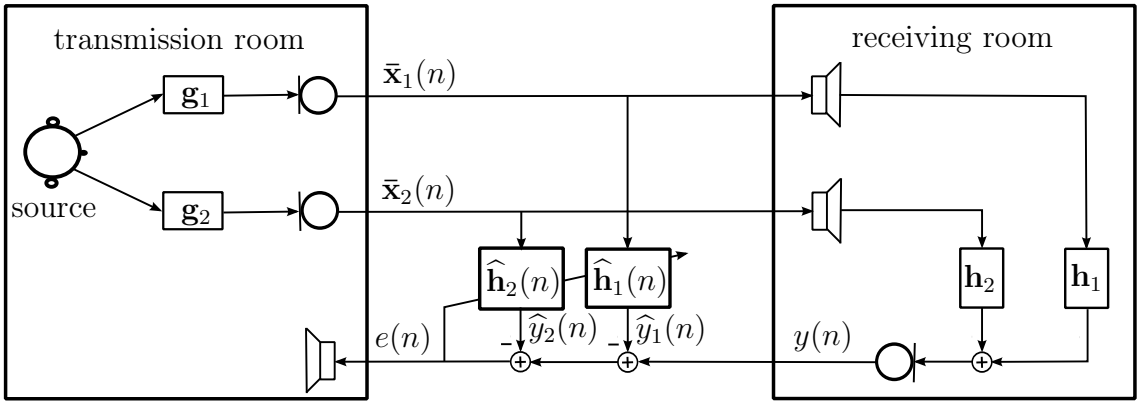


Figure 2.1: Schematic scenario for stereophonic AEC

The loudspeaker signals in the receiving room are described as vectors of length N:

$$\bar{\mathbf{x}}_i(n) = \left[\bar{x}_i(n), \bar{x}_i(n-1), ..., \bar{x}_i(n-N+1)\right]^T \qquad \text{with} \qquad i = 1, 2, \tag{2.2}$$

whereas $\bar{x}_i(k)$ denotes one sample at the time instance k. Superscript T indicates the transposition of a vector. The notation with an additional bar emphasizes the difference between the definition of the loudspeaker signals at this point and in the subsequent chapters. In the receiving room the acoustic paths from the loudspeakers to the microphone are captured by the impulse response vectors

$$\mathbf{h}_i = \left[h_{i,0}, h_{i,1}, ..., h_{i,N-1}\right]^T \qquad \text{with} \qquad i = 1, 2, \tag{2.3}$$

which are composed of N coefficients $h_{i,\kappa}$ with time instances $\kappa = 0, ..., N-1$. The primary goal of the stereophonic AEC is to estimate these impulse response vectors with two adaptive filters of length L:

$$\widehat{\mathbf{h}}_i(n) = \left[\widehat{h}_{i,0}(n), \widehat{h}_{i,1}(n), ..., \widehat{h}_{i,L-1}(n)\right]^T \qquad \text{with} \qquad i = 1, 2. \tag{2.4}$$

The estimated coefficients with time instances $\kappa = 0, ..., L-1$ are denoted as $\widehat{h}_{i,\kappa}(n)$ and will be dynamically adjusted with every block index n. The weighted least squares criterion is defined as

$$J(n) = \sum_{p=1}^{n} \lambda_s^{n-p} e^2(p) \qquad \text{with} \qquad 0 < \lambda_s < 1, \tag{2.5}$$

whereas λ_s is used as exponential forgetting factor and

$$e(n) = y(n) - \widehat{y}_1(n) - \widehat{y}_2(n) \tag{2.6}$$

equals the error signal at time instance n between the microphone output sample

$$y(n) = \mathbf{h}_1^T(n)\bar{\mathbf{x}}_1(n) - \mathbf{h}_2^T(n)\bar{\mathbf{x}}_2(n) \tag{2.7}$$

and its estimation

$$\widehat{y}(n) = \widehat{\mathbf{h}}_1^T(n)\bar{\mathbf{x}}_1(n) - \widehat{\mathbf{h}}_2^T(n)\bar{\mathbf{x}}_2(n). \tag{2.8}$$

The minimization of the weighted least squares criterion (see Equation 2.5) leads to the normal equation [1],[2]

$$\mathbf{R}(n) \begin{bmatrix} \widehat{\mathbf{h}}_1(n) \\ \widehat{\mathbf{h}}_2(n) \end{bmatrix} = \mathbf{r}(n), \tag{2.9}$$

with an estimate of the input signal covariance matrix

$$\mathbf{R}(n) = \sum_{p=1}^{n} \lambda_s^{n-p} \begin{bmatrix} \bar{\mathbf{x}}_1(n) \\ \bar{\mathbf{x}}_2(n) \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}_1^T(n) & \bar{\mathbf{x}}_2^T(n) \end{bmatrix}, \tag{2.10}$$

and an estimate of the cross-correlation vector between input and output signals

$$\mathbf{r}(n) = \sum_{p=1}^{n} \lambda_s^{n-p} y(p) \begin{bmatrix} \bar{\mathbf{x}}_1(n) \\ \bar{\mathbf{x}}_2(n) \end{bmatrix}. \tag{2.11}$$

The updates of the adaptive filters depend on the cross-correlation of the loudspeaker signals in the receiving room (see Equation 2.10). In the desired scenario with one common source (see Figure 2.1), this statement results in the so called "non-uniqueness" problem: The normal equation gets very ill-conditioned due to the strong cross-correlation of the loudspeaker signals . As a result of this, the weighted least squares criterion converges to a solution for the adaptive filters that does not unambiguously match system identification [1],[2],[3]. Furthermore the adjusted filter coefficients are determined by the estimated cross-correlation of the loudspeaker signals which can change abruptly (for example the speaker changes position). As a consequence all adaptive filter coefficients have to be recalculated [3]. These problems are tried to be solved with a reduction of the cross-correlations of all loudspeaker signals. Various solutions have been proposed like the addition of non-linear distortions or noise, the use of complementary comb filtering and the

implementation of time-varying phase shifts [3],[4]. These approaches show decisive disadvantages like the creation of artifacts and distortions that disturb the human perception.
In this thesis, a new approach is introduced and evaluated for a multichannel scenario with 48 loudspeakers and 10 microphones. This ill-conditioned optimization problem is discussed in the sense of AEC.

3 Wave-domain time-variant filtering

As described in the previous chapter, the identification of a LEMS is challenging due to the so-called "non-uniqueness problem". For an application like WFS many reproduction channels have to be estimated. This results in a severely ill-conditioned optimization problem. With the goal to decrease the cross-correlation of the loudspeaker signals in the receiving room (compared to Figure 2.1), a wave field rotation with a perceptually acceptable distortion will be introduced. The corresponding model based approach is build on a set of basis function that are used for a transformation into a spatial continuum. In the so called free-field description for wave-domain adaptive filtering the loudspeaker signals are weighted with an additional time-varying complex exponent [5]. After retransformation into the original domain, the cross-correlation between the channels can be decreased in front of the wave field generation.

3.1 Wave-domain transformation

The wave-domain transformation is based on the acoustic wave equation. Typical assumptions are a homogeneous, quiescent medium which can be characterized as an ideal gas in a steady state [6]. Furthermore only adiabatic processes are considered, so that state changes can be described without the release of heat in the system. In addition to this, the pressure of wave propagation is small compared to the static pressure in the medium. In this thesis, the homogeneous wave equation

$$\Delta p(\vec{u},t) - \frac{1}{c^2}\frac{\partial^2 p(\vec{u},t)}{\partial t^2} = 0 \tag{3.1}$$

is used in the cylindrical coordinate system (see Appendix A). Equation 3.1 includes the Laplacian operator Δ, the sound pressure p, the speed of sound c and $\vec{u}$ as position vector which is conventionally characterized with an arrow. The homogeneous wave equation is solved with the separation of variables. Furthermore it can be described with the circular harmonics expansion at the plane $z = 0$ as [7]:

$$P(\alpha,\varrho,j\omega) = \sum_{m=-\infty}^{+\infty} \left(\tilde{P}_m^{(1)}(j\omega)\mathcal{H}_m^{(1)}\left(\frac{\omega}{c}\varrho\right) + \tilde{P}_m^{(2)}(j\omega)\mathcal{H}_m^{(2)}\left(\frac{\omega}{c}\varrho\right)\right)\cdot e^{jm\alpha}, \tag{3.2}$$

whereas $P(\alpha,\varrho,j\omega)$ is the spectrum of the sound pressure. Moreover $\mathcal{H}_m^{(1)}(x)$ and $\mathcal{H}_m^{(2)}(x)$ are Hankel functions of the first and second kind of order m. In addition

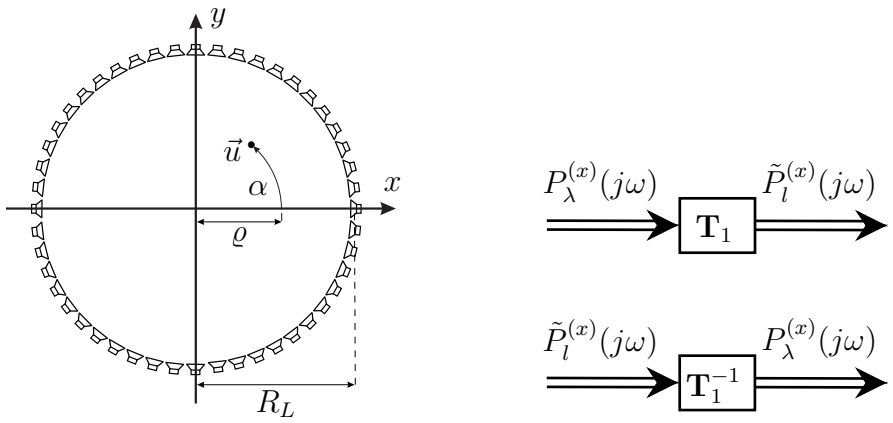
to this ω denotes the angular frequency and j is used as the imaginary unit. The spectra of the incoming and outgoing waves relative to the origin may be described by the components $\tilde{P}_m^{(1)}(j\omega)$ and $\tilde{P}_m^{(2)}(j\omega)$. With the assumption of ideal free field conditions we can describe their superposition as a standing wave quantity

$$\tilde{P}_m^{(x)}(j\omega)\mathcal{J}_m\Big(\frac{\omega}{c}\varrho\Big) = \tilde{P}_m^{(1)}(j\omega)\mathcal{H}_m^{(1)}\Big(\frac{\omega}{c}\varrho\Big) + \tilde{P}_m^{(2)}(j\omega)\mathcal{H}_m^{(2)}\Big(\frac{\omega}{c}\varrho\Big), \tag{3.3}$$

with the Bessel function $\mathcal{J}_m(x)$ of order m. Consequently, the inversion of Equation 3.2 can be obtained by a Fourier series expansion:

$$\tilde{P}_m^{(x)}(j\omega)\mathcal{J}_m\Big(\frac{\omega}{c}\varrho\Big) = \frac{1}{2\pi}\int_0^{2\pi} P(\alpha, \varrho, j\omega)e^{-jm\alpha}d\alpha \tag{3.4}$$

For the explicit definition of the transformation we consider a concentric circular loudspeaker array of Figure 3.1(a), whereas this derivation is also proved to be valid for other geometries [5].



(a) Loudspeaker array geometry (b) Denotations

Figure 3.1: Wave-domain transformation

With azimuth angle α and distance R_L to the origin, the loudspeaker positions can be written as

$$\vec{u}_\lambda = [\alpha_\lambda, \varrho_\lambda]^T \quad \text{with} \quad \alpha_\lambda = \lambda \cdot \frac{2\pi}{N_L} \quad \text{and} \quad \varrho_\lambda = R_L. \tag{3.5}$$

The index $\lambda = 0, ..., N_L - 1$ is introduced due to the number of N_L loudspeakers. It should be mentioned that the recording system is not considered in the derivation of the transformation equations. The microphone array is also concentrically positioned, but the distance to the origin is much smaller than R_L. As a result of this, the attenuation along the microphone array can be neglected and the distance between the loudspeakers and the recording elements is approximated as ϱ_λ. Figure 3.1(b) illustrates the corresponding denotations. The N_L spectra of the loudspeaker signals $P_\lambda^{(x)}(j\omega)$ are transformed into the wave-domain free-field description $\widetilde{P}_l^{(x)}(j\omega)$ with $l = -(N_L/2 - 1), ..., N_L/2$. We use the three-dimensional Green's function [5]

$$G\left(\vec{0}|\vec{u}_\lambda, j\omega\right) = \frac{e^{-j\varrho_\lambda k}}{\varrho_\lambda}, \tag{3.6}$$

to approximate the spectrum of a plane wave at the origin with an incident angle α_λ:

$$\tilde{P}_\lambda^{(p)}(j\omega) \approx P_\lambda^{(x)}(j\omega) \cdot G\left(\vec{0}|\vec{u}_\lambda, j\omega\right). \tag{3.7}$$

A superscript p marks the property of being a plane wave quantity. The sound pressure at a position $\vec{u} = [\alpha, \varrho]^T$ close to the origin ($\varrho << R_L$) can be calculated by the superposition of all loudspeaker signals (see Appendix A):

$$P(\alpha, \varrho, j\omega) = \sum_{\lambda=0}^{N_L-1} \tilde{P}_\lambda^{(p)}(j\omega) \cdot e^{j\varrho\cos(\alpha-\alpha_\lambda)k} \tag{3.8}$$

The combination with Equation 3.4 as well as the use of the Jakob-Anger expansion define the transformation $\mathbf{T}_1$ in the spectral representation as

$$\tilde{P}_l^{(x)}(j\omega) := j^l \sum_{\lambda=0}^{N_L-1} P_\lambda^{(x)}(j\omega) \frac{e^{-j\varrho_\lambda k}}{\varrho_\lambda} e^{-jl\alpha_\lambda}. \tag{3.9}$$

Equation 3.9 is structured like a DFT with respect to the loudspeaker indices [8]. The retransformation $\mathbf{T}_1^{-1}$ in spectral representation is defined as

$$P_\lambda^{(x)}(j\omega) := \frac{1}{N_L} \sum_{l=-N_L/2+1}^{N_L/2} \tilde{P}_l^{(x)}(j\omega) j^{-l} e^{jl\alpha_\lambda} \varrho_\lambda e^{j(\varrho_\lambda - \max\{\varrho_\lambda\})k}. \tag{3.10}$$

With the use of fractional delay filters [5], discrete-time impulse responses can be defined for the time-domain representations of the transformation:

$$h_{l,\lambda,i}^{(T1)} = j^l \frac{h_{d,i}(\varrho_\lambda f_s/c)}{\varrho_\lambda} e^{-jl\alpha_\lambda}. \tag{3.11}$$

The number of L_T impulse response coefficients $h_{l,\lambda,i}^{(T1)}$ are indexed with $i = 0, ..., L_T - 1$. Furthermore the sampling frequency f_s is used as input argument of an FIR filter

$$h_{d,i}(y) = \begin{cases} \frac{\sin(\pi(i-y-(L_T-1)/2))}{\pi(i-y-(L_T-1)/2)} & \text{for } 0 \le i \le L_T - 1, \\ 0 & \text{elsewhere,} \end{cases} \tag{3.12}$$

of odd length L_T and with non-integer delay y as argument. The retransformation can be defined with the use of Equation 3.12 as

$$h_{l,\lambda,i}^{(T1\text{inv})} = j^{-l} h_{d,i}\left((\max\{\varrho_\lambda\} - \varrho_\lambda) f_s/c\right) \varrho_\lambda \frac{e^{jl\alpha_\lambda}}{N_L}, \tag{3.13}$$

which equally describes a FIR filter of odd length L_T. The operator $\max\{\varrho_\lambda\}$ takes the maximum out of all ϱ_λ.

3.2 Time-variant filtering

The implementation of a field rotation is based on Equation 3.2 and Equation 3.3. It can be realized by the multiplication of $\tilde{P}_m^{(x)}(j\omega)$ with $e^{-jm\varphi(n)}$. A dynamically adjusted rotation angle $\varphi(n)$ is used for the time-variant filtering. The variable n is indexing the block number. Figure 3.2 illustrates the proposed prefilter structure.

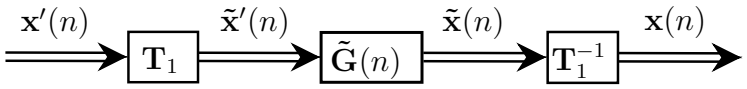


Figure 3.2: Entire prefilter structure

The loudspeaker signals are defined as

$$\mathbf{x}'(n) = \left[\mathbf{x}'^T_0(n), \mathbf{x}'^T_1(n), ..., \mathbf{x}'^T_{N_L-1}(n)\right]^T, \tag{3.14}$$

$$\mathbf{x}'_\lambda(n) = \left[x'_\lambda(nL'_F - L'_X + 1), x'_\lambda(nL'_F - L'_X + 2), ..., x'_\lambda(nL'_F)\right]^T, \tag{3.15}$$

with time samples $x'_\lambda(k)$, block length L'_X, frame shift L'_F and $\lambda = 0, ..., N_L - 1$ as index for the loudspeaker signals. The transformation into the wave-domain (see Figure 3.2) is realized according to:

$$\tilde{\mathbf{x}}'(n) = \mathbf{T}_1 \mathbf{x}'(n). \tag{3.16}$$

This representation in the spatial continuum also consists of N_L components, which are indexed with $l = -(N_L/2 - 1), ..., N_L/2$:

$$\tilde{\mathbf{x}}'_l(n) = \left[\tilde{x}'_l(n\tilde{L}'_F - \tilde{L}'_X + 1), \tilde{x}'_l(n\tilde{L}'_F - \tilde{L}'_X + 2), ..., \tilde{x}'_l(n\tilde{L}'_F)\right]^T. \tag{3.17}$$

The time samples are denoted as $\tilde{x}'_l(k)$. Furthermore $\tilde{L}'_X$ describes the block length and $\tilde{L}'_F$ equals the frame shift. The structure of $\tilde{\mathbf{x}}'(n)$ is identical to the one of Equation 3.14, whereas the index λ has to be replaced by l.
For the transformation introduced in Equation 3.16 a matrix $\mathbf{T}_1$ can be defined as

$$\mathbf{T1} = \begin{bmatrix} \mathbf{H}^{(T1)}_{-N_L/2+1,0} & \mathbf{H}^{(T1)}_{-N_L/2+1,1} & \cdots & \mathbf{H}^{(T1)}_{-N_L/2+1,N_L-1} \\ \mathbf{H}^{(T1)}_{-N_L/2+2,0} & \mathbf{H}^{(T1)}_{-N_L/2+2,1} & \cdots & \mathbf{H}^{(T1)}_{-N_L/2+2,N_L-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}^{(T1)}_{N_L/2,0} & \mathbf{H}^{(T1)}_{N_L/2,1} & \cdots & \mathbf{H}^{(T1)}_{N_L/2,N_L-1} \end{bmatrix}, \tag{3.18}$$

which consists of $N_L \cdot N_L$ submatrices. Every single component $\mathbf{H}^{(T1)}_{l,\lambda}$ decribes the transformation from $\mathbf{x}'_\lambda(n)$ to $\tilde{\mathbf{x}}'_l(n)$:

$$\mathbf{H}^{(T1)}_{l,\lambda} = \begin{bmatrix} h^{(T1)}_{l,\lambda,L_T-1} & h^{(T1)}_{l,\lambda,L_T-2} & \cdots & h^{(T1)}_{l,\lambda,0} & 0 & \cdots & 0 \\ 0 & h^{(T1)}_{l,\lambda,L_T-1} & \cdots & h^{(T1)}_{l,\lambda,1} & h^{(T1)}_{l,\lambda,0} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & h^{(T1)}_{l,\lambda,L_T-1} & \cdots & h^{(T1)}_{l,\lambda,0} \end{bmatrix}. \tag{3.19}$$

The individual components $h_{l,\lambda,i}^{(T1)}$ of this Sylvester matrix were introduced in Equation 3.11. As a result of this framework, the block length of the wave-domain representation $\tilde{\mathbf{x}}_l'(n)$ is determined by $\tilde{L}_X' = L_X' - L_T + 1$. The field rotation of the reproduced wave field is realized with

$$\tilde{\mathbf{x}}(n) = \tilde{\mathbf{G}}(n)\tilde{\mathbf{x}}'(n), \tag{3.20}$$

whereas $\tilde{\mathbf{G}}(n)$ is defined as a matrix with the purpose to individually weight the wave-domain representations $\tilde{\mathbf{x}}_l'(n)$ with a complex exponent that includes the time-varying rotation angle $\varphi(n)$:

$$\tilde{\mathbf{G}}(n) = \begin{bmatrix} e^{-j(-N_L/2+1)\varphi(n)} \cdot \mathbf{I}_{\tilde{L}_X'} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & e^{-j(-N_L/2+2)\varphi(n)} \cdot \mathbf{I}_{\tilde{L}_X'} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & e^{-j(N_L/2)\varphi(n)} \cdot \mathbf{I}_{\tilde{L}_X'} \end{bmatrix}. \tag{3.21}$$

The $\tilde{L}_X' \times \tilde{L}_X'$ - identity matrix is termed $\mathbf{I}_{\tilde{L}_X'}$. Furthermore $\mathbf{0}$ defines a matrix of equal dimensions with zero-valued entries. The weighted wave-domain signal $\tilde{\mathbf{x}}(n)$ is identically structured as $\tilde{\mathbf{x}}'(n)$ and can be retransformed by $\mathbf{T}_1^{-1}$ into the original domain:

$$\mathbf{x}(n) = \mathbf{T}_1^{-1}\tilde{\mathbf{x}}(n). \tag{3.22}$$

The structure of $\mathbf{T}_1^{-1}$ equals the one of Equation 3.18, whereas the components $h_{l,\lambda,i}^{(T1)}$ of Equation 3.19 have to be replaced by $h_{l,\lambda,i}^{(T1\text{inv})}$ of Equation 3.13. The loudspeaker signals $\mathbf{x}(n)$ are consequently equally structured as the original time representation $\mathbf{x}'(n)$ with the block length $L_X = \tilde{L}_X' - L_T + 1$.
The following examples was implemented to prove the wave-domain filtering with a constant rotation angle of $\varphi(n) = \pi/4$. Chapter B.1 includes an overview of all integrated files. The corresponding wave number vectors $\vec{k}_0$ and $\vec{k}_{\text{rot}}$ are shown in Figure 3.3. They characterize the incident angle of the original and rotated wave field.

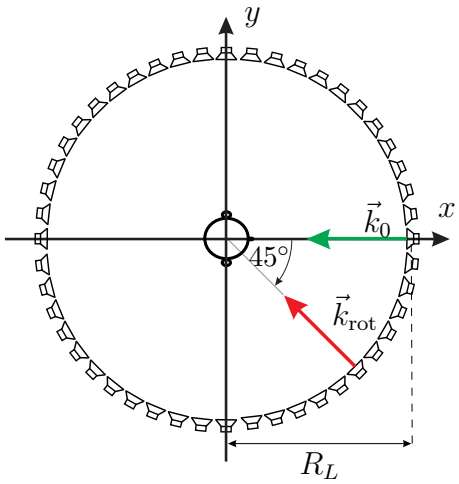
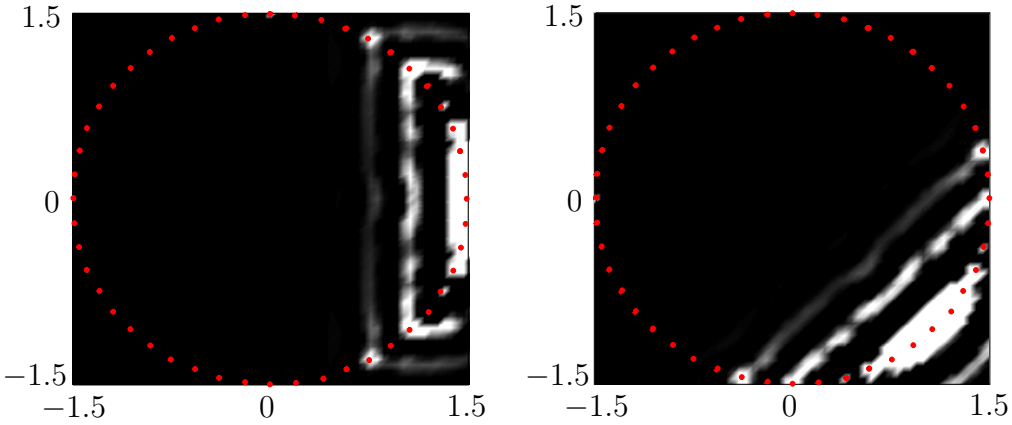


Figure 3.3: Wave number vectors for a field rotation with $\varphi(n) = \pi/4$

The sign of the argument in the complex exponent (see Equation 3.21) results in a rotation in negative angular direction. We implemented the wave-domain representation of the transformation equations for a loudspeaker array of height 1.4 meter, radius $R_L = 1.5$ meter and with $N_L = 48$ elements. The WFS algorithm is used to create a plane wave whose incident angle should be rotated. To improve the visualization, we chose a sampling rate of 2450 bits per second, a frame shift $L_F = 75$ and an overlap of 15 samples per block. The loudspeaker positions were marked as red points.



(a) Without field rotation (b) With field rotation

Figure 3.4: Wave field at 32 iterations (1 second)

Beside the characteristic of plane waves, the implemented field rotation is also visible in Figure 3.4.

4 Generalized frequency-domain adaptive filtering

As described in the previous chapter, the principle of wave-domain time-variant filtering should be used to improve the identification of a LEMS. In the following we choose a multichannel AEC scenario to evaluate the introduced wave field rotation. The corresponding signal processing is realized with the so called GFDAF algorithm [9],[10]. To start with the system model we consider Figure 4.1.

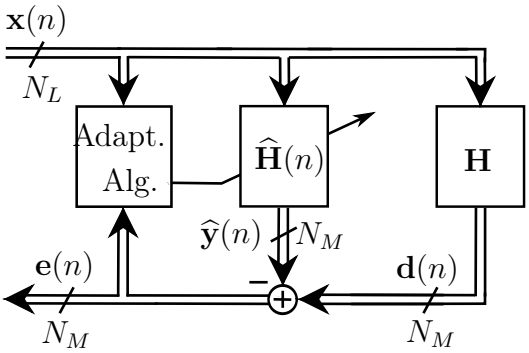


Figure 4.1: Signal model of a multichannel AEC scenario

The loudspeaker signals are defined as

$$\mathbf{x}(n) = \left[\mathbf{x}_0^T(n), \mathbf{x}_1^T(n), ..., \mathbf{x}_{N_L-1}^T(n)\right]^T, \tag{4.1}$$

$$\mathbf{x}_\lambda(n) = \left[x_\lambda(nL_F - 2L_H + 1), x_\lambda(nL_F - 2L_H + 2), ..., x_\lambda(nL_F)\right]^T, \tag{4.2}$$

with one time sample denoted as $x_\lambda(k)$. In this multichannel scenario we use N_L loudspeaker signals $\mathbf{x}_\lambda(n)$ indexed with $\lambda = 0, ..., N_L - 1$. The block length is equal to $2L_H$, whereas the frame shift L_F is also included in the definition of the microphone signals:

$$\mathbf{d}(n) = \left[\mathbf{d}_0^T(n), \mathbf{d}_1^T(n), ..., \mathbf{d}_{N_M-1}^T(n)\right]^T, \tag{4.3}$$

$$\mathbf{d}_\mu(n) = \left[d_\mu(nL_F - L_H + 1), d_\mu(nL_F - L_H + 2), ..., d_\mu(nL_F)\right]^T. \tag{4.4}$$

We denote the time samples of the N_M microphones $d_\mu(k)$ with $\mu = 0, ..., N_M - 1$. In addition to this each block indexed by n consists of L_H samples. Furthermore the path from loudspeaker λ to microphone μ is described with the impulse response coefficients $h_{\mu,\lambda}(k)$:

$$d_\mu(k) = \sum_{\lambda=0}^{N_L-1} \sum_{\kappa=0}^{L_H-1} x_\lambda(k - \kappa) h_{\mu,\lambda}(k), \tag{4.5}$$

which are included in the matrix $\mathbf{H}$. The corresponding structure is equivalent to the one of the transformation matrix from Equation 3.18. As a result of this, $N_L \cdot N_M$ submatrices $\mathbf{H}_{\lambda,\mu}$ (with $\lambda = 0, ..., N_L - 1$ and $\mu = 0, ..., N_M - 1$) have to be introduced just like in Equation 3.19. This entire structure is also valid for the estimated and dynamically adjusted impulse response matrix $\widehat{\mathbf{H}}(n)$.

4.1 Optimization criterion and update equation

As illustrated in Figure 4.1, the error signal vector can be calculated as the difference between the microphone signals and the outputs of the adaptive filter:

$$\mathbf{e}(n) = \mathbf{d}(n) - \widehat{\mathbf{y}}(n). \tag{4.6}$$

The vectors $\mathbf{e}(n)$ and $\widehat{\mathbf{y}}(n)$ are identically structured as $\mathbf{d}(n)$, so that they can equally be decomposed into their individual partitions indexed with μ. For the sake of brevity, we consider a separate optimization with the single components $\mathbf{e}_\mu(n)$, $\widehat{\mathbf{y}}_\mu(n)$ and $\mathbf{d}_\mu(n)$. A minimization of the corresponding weighted squared error

$$J_\mu(n) = (1 - \lambda_a) \sum_{i=0}^{n} \lambda_a^{n-i} \mathbf{e}_\mu^{\mathrm{H}}(i) \mathbf{e}_\mu(i) \qquad \text{with} \qquad 0 < \lambda_a < 1, \tag{4.7}$$

includes the exponential forgetting factor λ_a. Superscript H indicates the hermitian (i.e. conjugate transpose) of a vector. To derive the algorithm in the frequency-domain, error vector and loudspeaker signals are transformed with a DFT:

$$\underline{\mathbf{e}}_\mu(n) = \mathbf{F}\left[\mathbf{0}, \mathbf{I}_{L_H}\right]^T \mathbf{e}_\mu(n), \tag{4.8}$$

$$\underline{\mathbf{X}}(n) = \Big[\mathrm{Diag}\{\mathbf{F}\mathbf{x}_0(n)\}, \mathrm{Diag}\{\mathbf{F}\mathbf{x}_1(n)\}, ..., \mathrm{Diag}\{\mathbf{F}\mathbf{x}_{N_L-1}(n)\}\Big]. \tag{4.9}$$

Here, the matrix $\mathbf{F}$ is the unitary DFT-matrix with the dimensions $2L_H \times 2L_H$ [8]. Furthermore $\mathrm{Diag}\{\mathbf{x}\}$ creates a diagonal matrix with the vector $\mathbf{x}$ on its main diagonal. With this introduction of the DFT-representations and the solution of Equation 4.7 in the time domain, the update equation is obtained as

$$\hat{\underline{\mathbf{h}}}_\mu(n) = \hat{\underline{\mathbf{h}}}_\mu(n-1) + \mu_a(1-\lambda_a)\mathbf{F} \begin{bmatrix} \mathbf{I}_{L_H} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}^T \mathbf{F}^H \underbrace{\underline{\mathbf{S}}^{-1}(n)\underline{\mathbf{X}}^H(n)}_{\underline{\mathbf{K}}(n)} \underline{\mathbf{e}}_\mu(n), \tag{4.10}$$

with the so called Kalman gain $\underline{\mathbf{K}}(n)$, the step size $0 < \mu_a < 1$ and the sparse matrix

$$\underline{\mathbf{S}}(n) = \lambda_a \underline{\mathbf{S}}(n-1) + (1-\lambda_a)\frac{1}{2}\left(\underline{\mathbf{X}}^H(n)\underline{\mathbf{X}}(n) + \frac{\delta_{reg}}{N_L}\mathbf{I}_{N_L} \otimes \underline{\mathbf{X}}(n)\underline{\mathbf{X}}^H(n)\right) \tag{4.11}$$

which enables a bin-wise inversion. An additional Tychonov regularization is included, which can be removed by setting the weighting factor δ_{reg} to zero. Besides this, the Kronecker product is denoted by $\otimes$. We have to consider that the vector $\hat{\underline{\mathbf{h}}}_\mu(n)$ describes the DFT-representations of the estimated impulse response vectors for microphone μ:

$$\hat{\underline{\mathbf{h}}}_\mu(n) = \mathbf{F}\left[\mathbf{I}_{L_H}, \mathbf{0}\right]^T \hat{\mathbf{h}}_\mu(n), \tag{4.12}$$

which are denoted as

$$\hat{\mathbf{h}}_{\mu}(n) = \left[\hat{\mathbf{h}}_{0,\mu}^{T}(n), \hat{\mathbf{h}}_{1,\mu}^{T}(n), ..., \hat{\mathbf{h}}_{N_L-1,\mu}^{T}(n)\right]^{T}. \tag{4.13}$$

As a result of this, the estimated coefficients $\hat{h}_{\lambda,\mu,i}(k)$ from loudspeaker λ to microphone μ are included in

$$\hat{\mathbf{h}}_{\lambda,\mu}(n) = \left[\hat{h}_{\lambda,\mu,0}(n), \hat{h}_{\lambda,\mu,1}(n), ..., \hat{h}_{\lambda,\mu,L_H-1}(n)\right]^{T}, \tag{4.14}$$

whereas $i = 0, ..., L_H - 1$.

4.2 Evaluation parameters for system identification

The convergence behavior of the system identification will be shown both in terms [9] of NMA

$$\Delta h(n) = 10 \log_{10} \left(\frac{||\hat{\mathbf{H}}(n) - \mathbf{H}||_F^2}{||\mathbf{H}||_F^2} \right) \tag{4.15}$$

and ERLE

$$\mathrm{ERLE}(n) = 10 \log_{10} \left(\frac{||\mathbf{d}(n)||_2^2}{||\mathbf{e}(n)||_2^2} \right), \tag{4.16}$$

whereas $||\cdot||_2^F$ equals the Frobenius norm and $||\cdot||_2^2$ describes the Euclidean norm. In Equation 4.15 we calculate the normalized misalignment between estimated and measured impulse response coefficients. With the echo return loss enhancement, the Euclidean distances between the time samples of microphone signals and error vectors are concerned.

5 Identification of a LEMS

5.1 Multichannel AEC scenario

In this section we consider the multichannel AEC scenario shown in Figure 5.1.

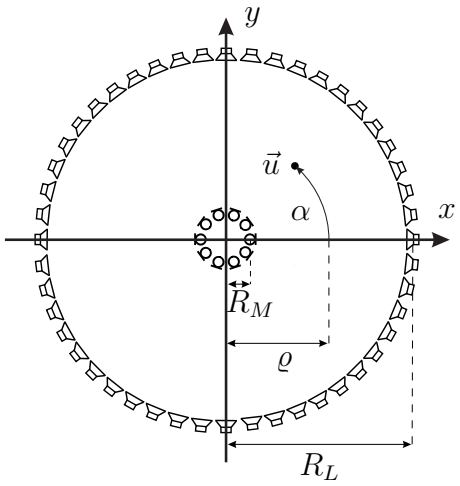
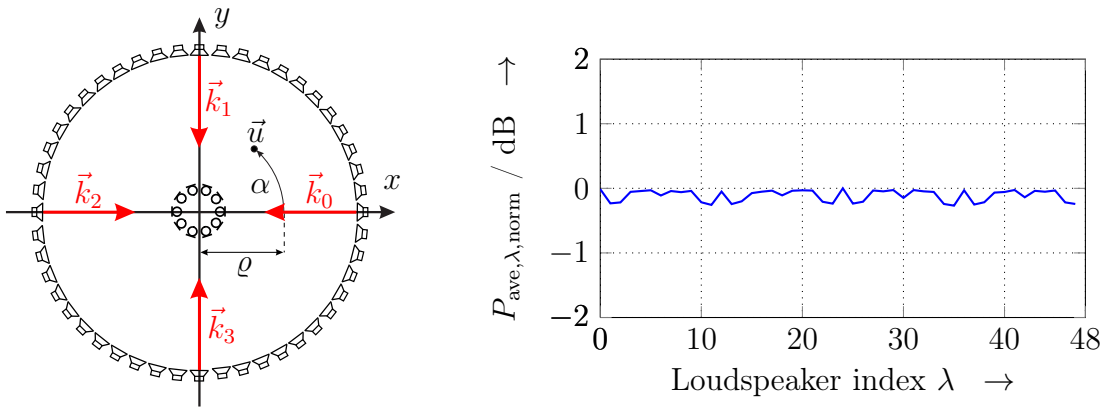


Figure 5.1: Multichannel AEC scenario

A concentric circular loudspeaker array with $N_L = 48$ elements, a height of 1.4 meter and a radius of $R_L = 1.5$ meter generates the acoustic wave field. The corresponding positions $\vec{u}$ in the cylindrical coordinate system were introduced in Chapter 3.1. The recording system is realized with a concentric circular microphone array of radius $R_M = 0.05$ meter and height 1.4 meter consisting of $N_M = 10$ elements.
For the identification of this LEMS, the GFDAF algorithm is used with the corresponding evaluation parameters ERLE and NMA (see Chapter 4.2). Measured impulse responses in a room with a reverberation time T_{60} of approximately 0.3 seconds are resampled to the sampling rate of 11025 bits per second and truncated to $L_H = 1024$ samples. The corresponding coefficients are integrated in the matrix $\mathbf{H}$ (see Chapter 4.1). For the parameters of the GFDAF algorithm we use an exponential forgetting factor $\lambda_a = 0.95$, a step size $\mu_a = 0.5$, a weighting factor $\delta_{reg} = 0.5$ and a frame shift $L_F = 512$. The initial values for the matrix $\mathbf{S}(n)$ (see Equation 4.11) are chosen as $\mathbf{S}(0) = 12.4 \cdot 10^4 \cdot \mathbf{I}_{2L_H \cdot N_L}$ with $\mathbf{I}_{2L_H \cdot N_L}$ being the identity matrix with the dimensions $2L_H \cdot N_L$.
The wave-domain transformation of Chapter 3.1 is implemented in the spectral representation with an overlap of $L_X - L_F = 75$ samples per block.

5.2 Input signals based on WFS

Plane waves can be generated with the WFS implementation . We want to supply every loudspeaker equally in the average sense. For this purpose, four plane waves sources with an incident angle of $\alpha_\nu = \nu \cdot \pi/2$ in radians are chosen with $\nu = 0, ..., 3$. The corresponding wave number vector directions $\vec{k}_\nu$ are indicated in Figure 5.2(a).



(a) Direction of the input signals

(b) Normalized average power of loudspeaker signals

Figure 5.2: Four inputs signals as plane waves with orthogonal incident angles

The input signals of the four plane waves are chosen to be normally distributed and statistically independent, so that the average power

$$P_{\text{ave},\lambda} = 10 \cdot \log_{10} \left[\frac{1}{N \cdot L_X} \sum_{k=0}^{N \cdot L_X - 1} |x_\lambda(k)|^2 \right] \tag{5.1}$$

at every loudspeaker with $\lambda = 0, ..., N_L - 1$ is approximately the same. The simulation results with $N = 500$ iteration steps are shown in Figure 5.2(b), whereas $P_{\text{ave},\lambda,\text{norm}}$ describes the averaged power of Equation 5.1 normalized to the largest value along λ. It can be noticed, that this choice of the input signals create a power distribution which makes an approximately uniform excitation of all loudspeakers in the average sense.

5.3 System identification without field rotation

We apply the concept that was introduced in the previous section with four plane waves created by statistically independent input signals (see Figure 5.2(a)). The system identification is calculated without the use of a wave-domain time-varying filtering. We can analyse the result for the NMA shown in Figure 5.3. The system identification improves during approximately 4 seconds and subsequently converges to a value of about -0.38 dB.
Considering the ERLE in Figure 5.3, a fast increase above 60 dB can be noticed (due to the ideal laboratory conditions). The combination of both quantities implies the

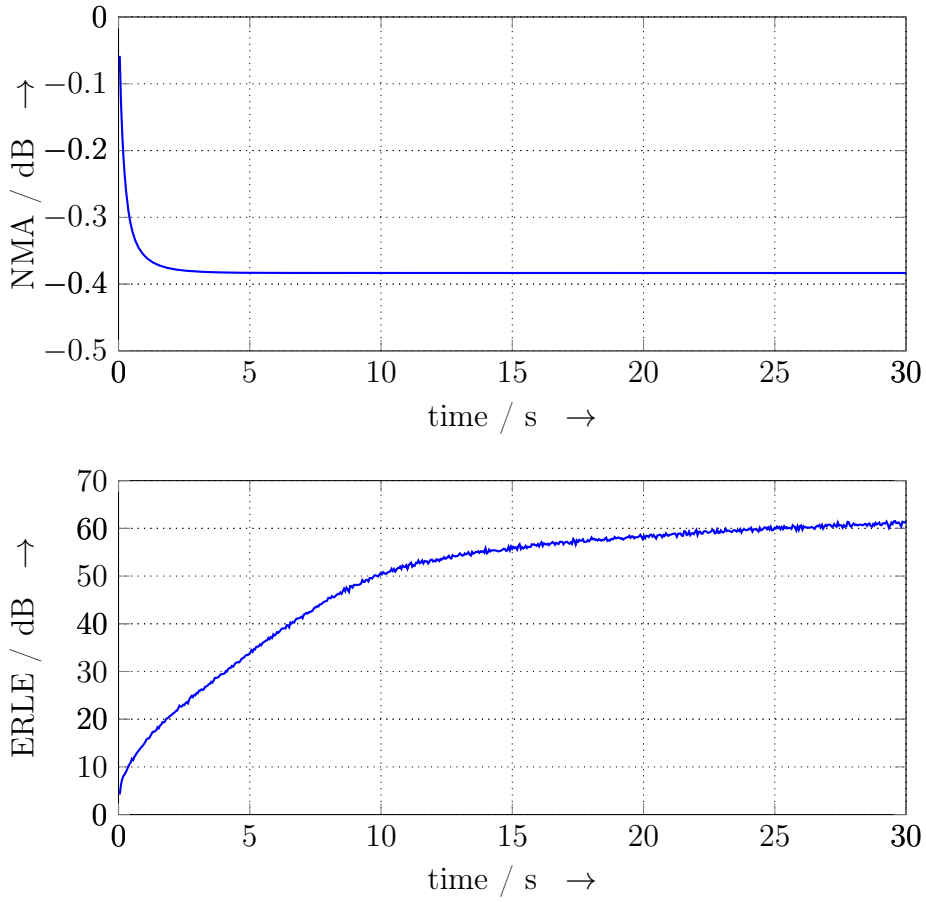


Figure 5.3: NMA and ERLE without wave-domain time-varying filtering

reduction of the error signal (ERLE, see Equation 4.16), while the misalignment between estimated and measured impulse response coefficients (NMA, see Equation 4.15) could not be decreased. This statement reflects the ill-conditioned optimization problem of this multichannel scenario that was introduced in Chapter 2 for the stereophonic case.

5.4 System identification with sinusoidal field rotation

The sinusoidal rotation is defined as

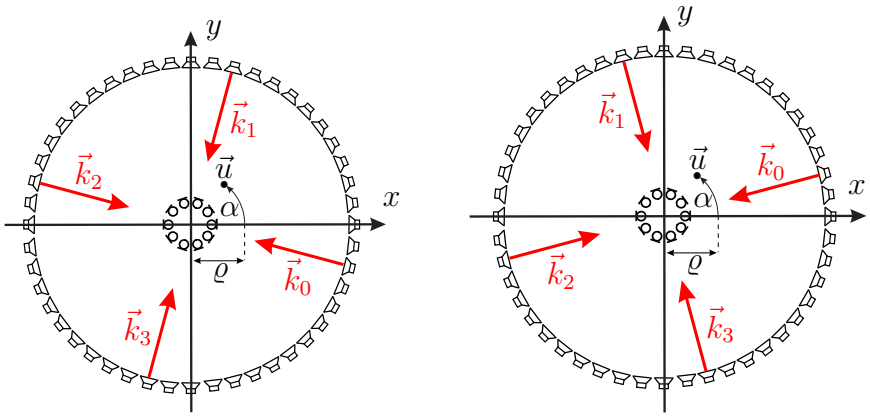
$$\varphi(n,\nu) = \varphi_{\max} \sin\left(2\pi \frac{\mathrm{mod}(n, L_P)}{L_P}\right) - \nu \cdot \pi/2 \qquad \text{with} \qquad \nu = 0, ..., 3, \tag{5.2}$$

whereas $\varphi_{\max}$ is the maximum amplitude, L_P describes the number of samples per period and $\mathrm{mod}(x)$ equals the modulo operator. It should be mentioned that the negative sign of ν is introduced as result from Equation 3.21 to create the wave number vectors $\vec{k}_\nu$ shown in Figure 5.2. In the following, the rotation angle will be

specified in radians as multiples of $\pi/48$. As an examples, we choose $\varphi_{\max} = 4\pi/48$ and $L_P = 100$. The extreme values of the incident angles (see Equation 5.2) during the first period are given as

$$\varphi(25, \nu) = 4\pi/48 - \nu \cdot \pi/2, \tag{5.3}$$
$$\varphi(75, \nu) = -4\pi/48 - \nu \cdot \pi/2. \tag{5.4}$$

The corresponding incident angles are represented by the wave number vectors shown in Figure 5.4.



(a) $\varphi(25, \nu) = 4\pi/48 - \nu \cdot \pi/2$ (b) $\varphi(75, \nu) = -4\pi/48 - \nu \cdot \pi/2$

Figure 5.4: Extreme values of sinusoidal field rotation

Due to the angular positions of the array elements (see Equation 3.5), $\varphi_{\max} = 4\pi/48$ can be interpreted as if the plane wave sources have been rotated by two loudspeakers in negative (see Figure 5.4(a)) or positive (see Figure 5.4(b)) angular direction relative to the intial one.

Influence of the maximum rotation angle

In this section we use Equation 5.2 to analyze the influence of $\varphi_{\max}$ on the system identification in a scenario with $L_P = 301$ samples (14 seconds) per period. The rotation angles of the four generated plane waves are shown in Figure 5.5 for three values of $\varphi_{\max}$.
Three essential statements can be made considering the NMA shown in Figure 5.5. Firstly, the system identification can be improved by increasing the maximum rotation angle. This is also confirmed with the decreased numerical values at 30 seconds compared to the initial one of -0.38 dB (see Chapter 5.3). Secondly, the greatest slope of the NMA occurs at the same time as the highest gradient of the rotation angle function. In combination with the first statement, the difference between the adjacent angular positions play a significant role. Thirdly, the NMA changes dominantly at the rotation angles that are used the first time. Consequently we can notice that innovative angular positions have a great impact on the system identification. This is confirmed with the behavior during the first and third quarter of

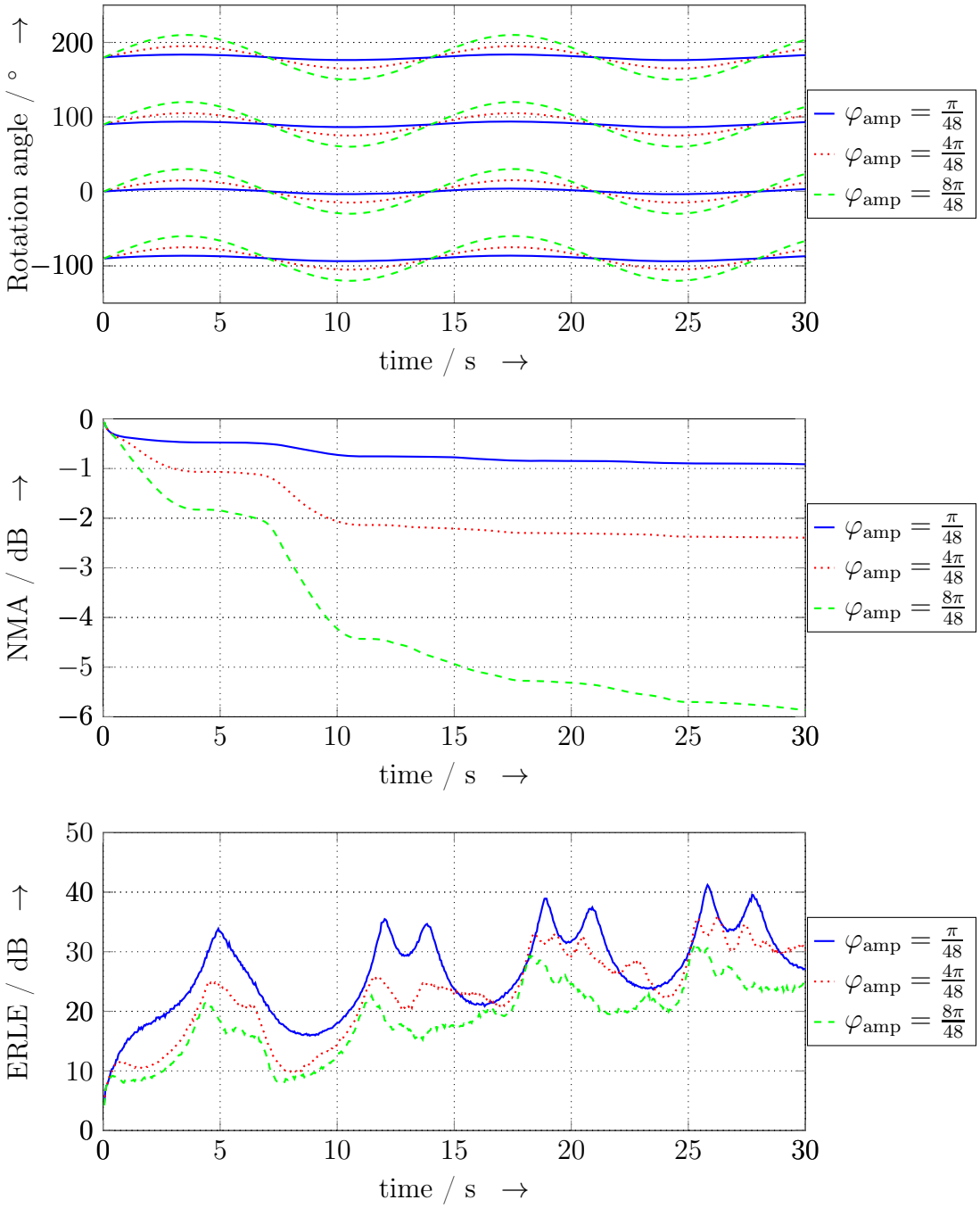


Figure 5.5: NMA and ERLE with time-varying filtering for different φ_{amp}

the the first period.
The improved system identification is achieved on the cost of reduced ERLE (see Figure 5.5).

Influence of the number of samples per period

In this section we use Equation 5.2 and determine the maximum rotation angle to $\varphi_{\max} = 8\pi/48$. The simulation results for three different values of L_P samples per period are shown in Figure 5.6.

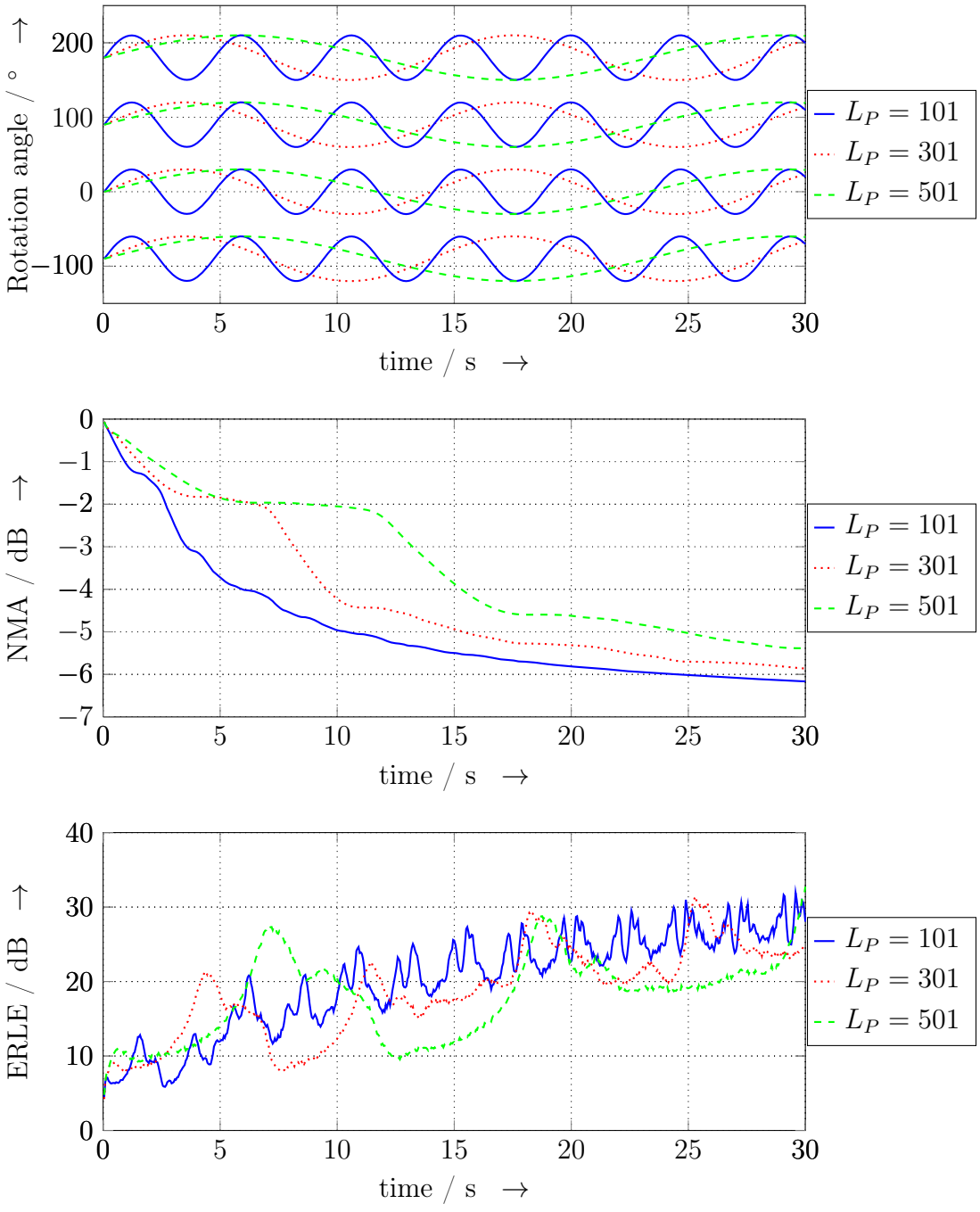


Figure 5.6: NMA and ERLE with time-varying filtering for different L_P

As described in the previous chapter, the fastest decrease of the NMA is at the same time as the highest gradient of the rotation angle function. A shorter period

length consequently results in a smaller value of the NMA during the first seconds. Furthermore it can be noticed that a faster change of the rotation angle function also influences the stability of the simulation. As shown in Figure 5.6, the ERLE includes an additional oscillation.
Compared with increasing the maximum rotation angle (see Chapter 5.4) we can resume that the reduction of the number of samples per period is a less powerful tool.
The comparison between an odd or an even number for the value of L_P is also of interest. As an example we choose $L_P = 8$ and $\nu = 0$, so that Equation 5.2 results to the values $\varphi(1) = \varphi(3) = \varphi_{\max}/\sqrt{2}$. This implies that every rotation angle is used twice during one period. The number of rotation angles can be increased by choosing an odd number. If we use $L_P = 7$, we can see that the value of $\varphi(1) = \varphi_{\max} \cdot 2\pi/7$ only appears once in a period.

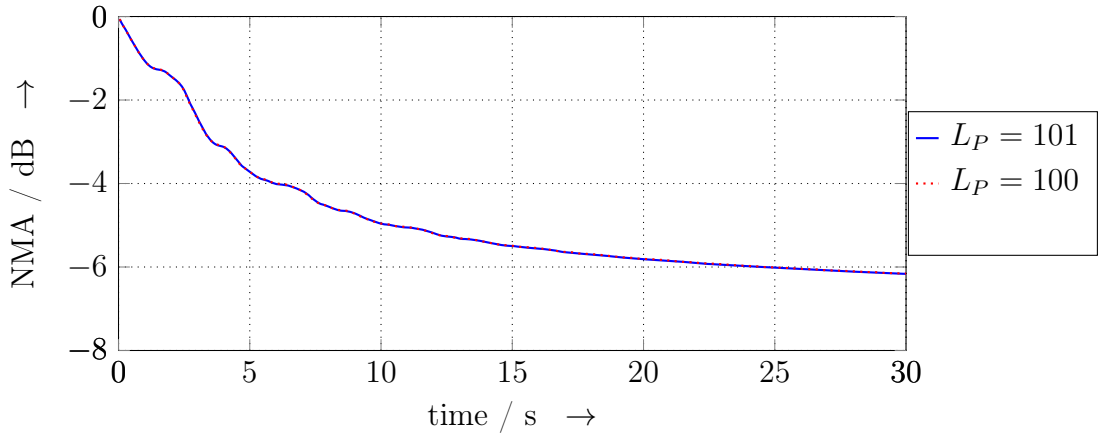


Figure 5.7: NMA with time-varying filtering for an even an odd value of L_P

In terms of the NMA, there is a negligible difference between and odd and even number of samples per period which appears in the second decimal place in dB (see Figure 5.7).

5.5 System identification with further rotation functions

In addition to the previous chapters, three further types of rotation angle functions for an even number of $L_P = 100$ samples per period will be examined. For this purpose we index the sinusoidal rotation that was introduced in Equation 5.2 by $\varphi_1(n, \nu)$. With the use of the signum function

$$\operatorname{sign}(x) = \begin{cases} 1 & \text{for } x > 0, \\ 0 & \text{for } x = 0, \\ -1 & \text{for } x < 0, \end{cases} \tag{5.5}$$

a rotation angle function with steeper slopes as $\varphi_1(n,\nu)$ can be defined as

$$\varphi_2(n,\nu) = \varphi_{\max} \cdot \operatorname{sign}\left(\varphi_1(n,\nu) + \nu \cdot \pi/2\right) \cdot \sqrt{\left|\frac{\varphi_1(n,\nu) + \nu \cdot \pi/2}{\varphi_{\max}}\right|} - \nu \cdot \pi/2, \quad (5.6)$$

whereas $\nu = 0, ..., 3$. As a second waveform we choose one with constant slopes:

$$\frac{\varphi_3(n,\nu)}{\varphi_{\max}} = \begin{cases} \left(\frac{4 \cdot \operatorname{mod}(n, L_P)}{L_P}\right) - \nu \cdot \pi/2 & \text{for} \quad \operatorname{mod}(n, L_P) < L_P/4, \\ \left(1 - \frac{4 \cdot \operatorname{mod}(n - L_P/4, L_P)}{L_P}\right) - \nu \cdot \pi/2 & \text{for} \quad L_P/4 \leq \operatorname{mod}(n, L_P) < 3L_P/4, \\ \left(-1 + \frac{4 \cdot \operatorname{mod}(n - 3L_P/4, L_P)}{L_P}\right) - \nu \cdot \pi/2 & \text{for} \quad 3L_P/4 \leq \operatorname{mod}(n, L_P). \end{cases} \quad (5.7)$$

As third function we use the original sinusoidal one of Equation 5.2 and modify the block index n by an addition with an uniformly distributed pseudorandom number g_r:

$$\varphi_4(n,\nu) = \varphi_1(n + g_r, \nu) \qquad \text{with} \qquad 0 < g_r < 1. \quad (5.8)$$

The difference between the three introduced functions is shown in Figure 5.8 for $\varphi_{\max} = 8\pi/48$ and $\nu = 0$.

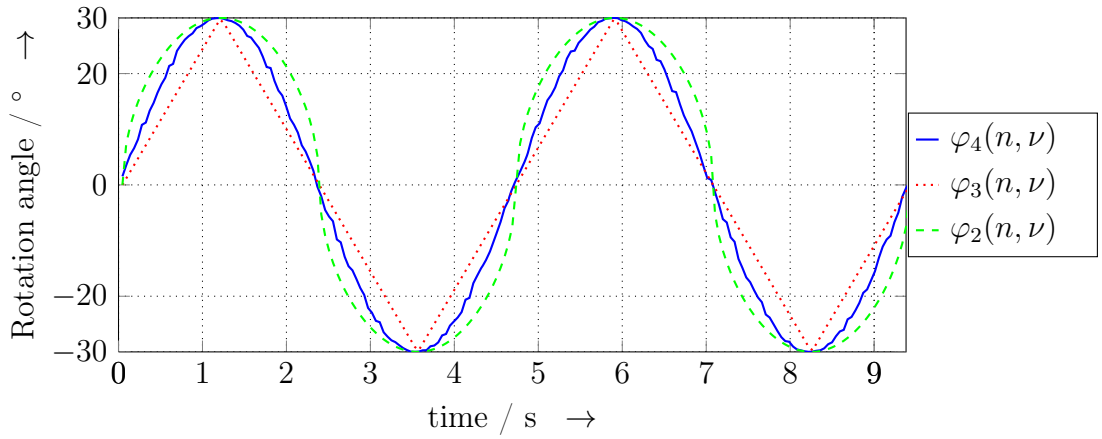


Figure 5.8: Rotation angle with time-varying filtering for different rotation functions

To evaluate the system identification for these three cases, we consider the initial scenario with four plane waves as input signals. The results of the NMA and the ERLE are shown in Figure 5.9.
In comparison to Figure 5.6, there is no improvement to the scenario with a sinusoidal rotation angle function. The introduced quantities $\varphi_2(n,\nu)$ and $\varphi_3(n,\nu)$ achieve a higher misalignment, while $\varphi_4(n,\nu)$ delivers approximately the same results as in Chapter 5.4. Concerning the ERLE, the absolute value of oscillation is also smallest for $\varphi_4(n,\nu)$. The functions $\varphi_2(n,\nu)$ and $\varphi_3(n,\nu)$ are consequently not improving the system identification.
As a result of this we use the sinusoidal rotation $\varphi_1(n,\nu)$ that is now proved to be most promising regarding ERLE and NMA.

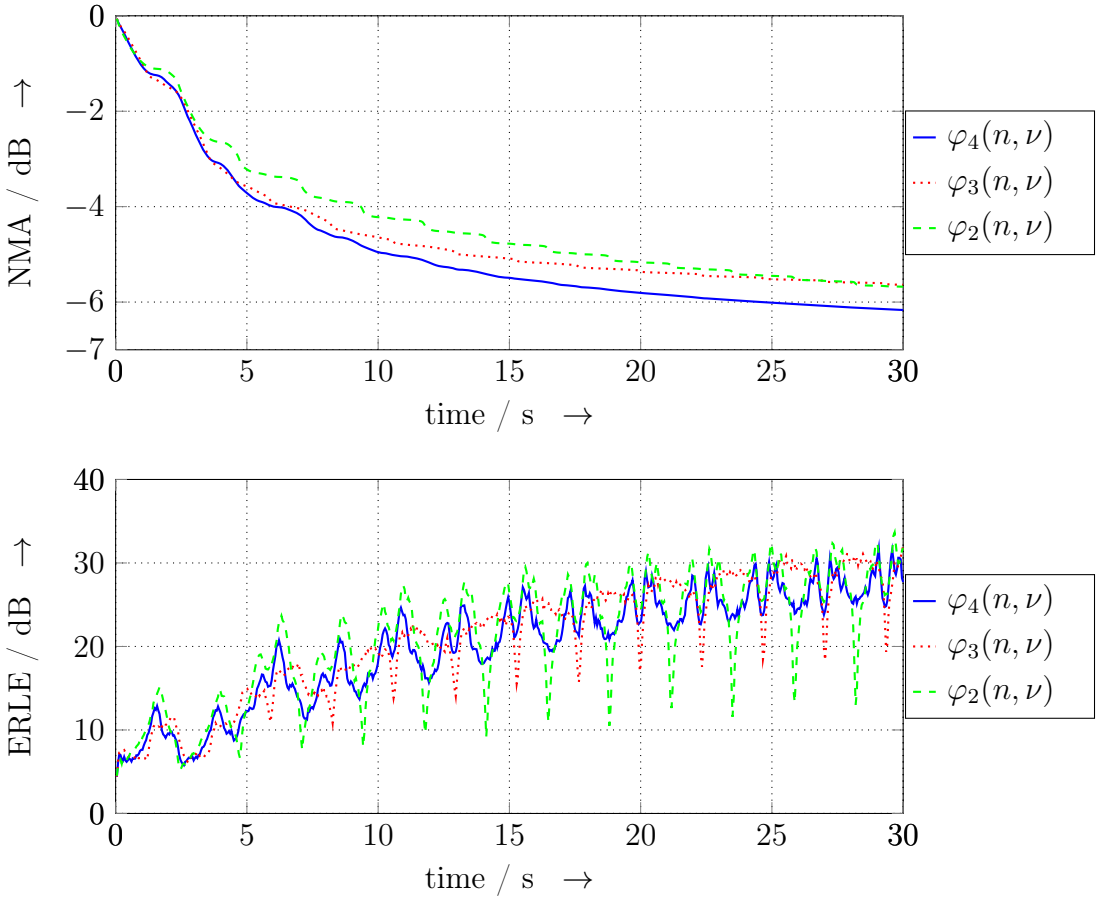
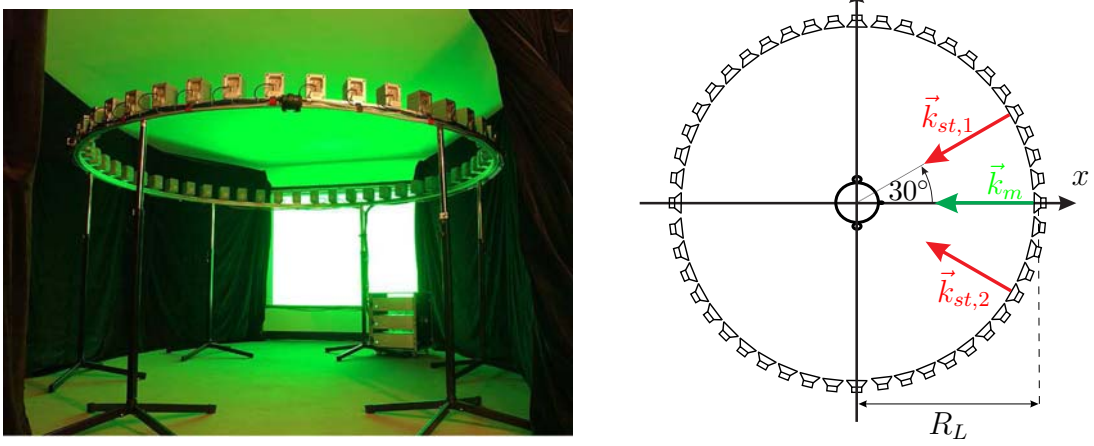


Figure 5.9: NMA and ERLE with time-varying filtering for different rotation functions

6 Hearing test

As introduced in the previous chapters we use a sinusoidal acoustic wave field rotation to improve the system identification of a LEMS. A hearing test has been invented to evaluate the subjective perception of this wave-domain time-varying filtering. We use the WFS algorithm to generate plane waves with a determined incident angle. Just like the previous chapters, the concentric circular loudspeaker array consist of 48 elements, a height of 1.4 meter and a radius of $R_L = 1.5$ meter (see Figure 6.1(a)). The subject is positioned at the center of the loudspeaker array which equals the origin of the coordinate system (see Figure 6.1(b)).



(a) Loudspeaker array

(b) Spatial position of the listener

Figure 6.1: Scenario for the hearing test

In the previous chapter, four plane waves of incident angle $\alpha_\nu = \nu \cdot \pi/2$ with $\nu = 0, ..., 3$ were generated by normally distributed and statistically independent input signals. At this point we use two monaural and three stereophonic sound files, whereas the corresponding wave number vectors for the mono ($\vec{k}_m$) and stereo cases ($\vec{k}_{st,1}$,$\vec{k}_{st,1}$) are shown in Figure 6.1(b). It should be mentioned that these inputs signal do not accomplish an uniform excitation of all loudspeakers in the average sense. The rotation angle function is adapted as

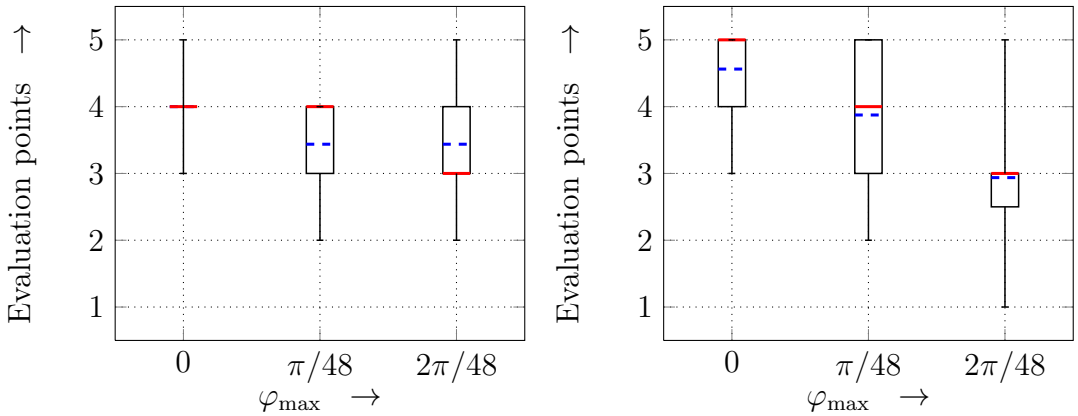
$$\varphi_{ht}(n, \xi) = \varphi_{\max} \sin\left(2\pi \frac{\mathrm{mod}(n, L_P)}{L_P}\right) - \xi \cdot \pi/3 \tag{6.1}$$

with $\xi = 0$ for the mono and $\xi = \pm 1$ for the stereo cases. Every input file is cut to a time sequence of 30 seconds. To evaluate the subjective perception we use two wave-domain time-varying filters with $L_P = 301$ samples per period and maximum

rotation angles of $\varphi_{\max,1} = \pi/48$ and $\varphi_{\max,2} = 2\pi/48$. The hearing test starts with the first input file, whereas original and filtered versions of the time sequence appear in an arbitrary manner with a pause of 5 seconds between. Before the second input file is played, a break can be used to evaluate the three parameters "Spatial stationarity", "General quality" and "Artifacts". The first parameter indicates the property of a sound source to be perceived as being located at a fixed position. Furthermore "Artifacts" corresponds to additional noise or audible jitter. The evaluation is oriented on the "Mean Opinion Score" with a scale from 1 to 5. With increasing number of points the quality is marked to be improving. As a result of this, 1 point corresponds to a bad (not acceptable) quality, while 5 represent excellent characteristics (no noticeable distortions). The entire questionnaire is included in Chapter D.

6.1 Results for three evaluation parameters

A total number of 16 persons aged between 20 and 30 years participated the introduced hearing test. The evaluation parameter "Artifacts" is not included in the following. This is because no significant difference between original file and filtered versions were noticed by the listeners. The evaluation parameters "General quality" and "Spatial stationarity" are marked with the mean value as dashed blue and the median as red line. We display the upper and lower quartiles as bottom and top of a box. Furthermore maximum and minimum value are also are marked as points linked to the boxes.
As first sound file we use a composition of the pianist Franz Liszt. As described in the previous section, one original and two filtered versions were played in an arbitrary manner. The results for the evaluation parameters of this monaural sound file are shown in Figure 6.2. As we can see in Figure 6.2(a), the "General quality"

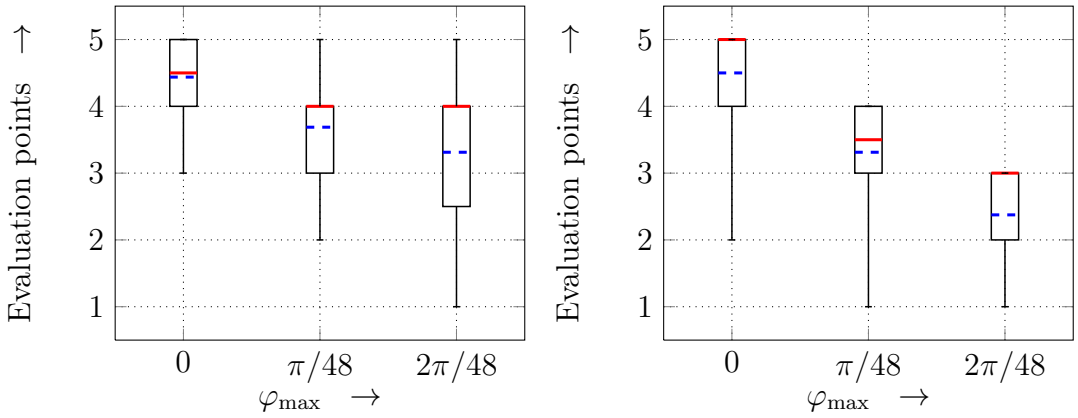


(a) General quality (b) Spatial stationarity

Figure 6.2: Hearing test results for the sound file "Liszt mono"

of both filtered versions is evaluated as being slightly worse. The original sound file has equal mean, median and quartiles, whereas the minimum value are 3 points (fair quality, not disturbing). In comparison to this we consider the two filtered

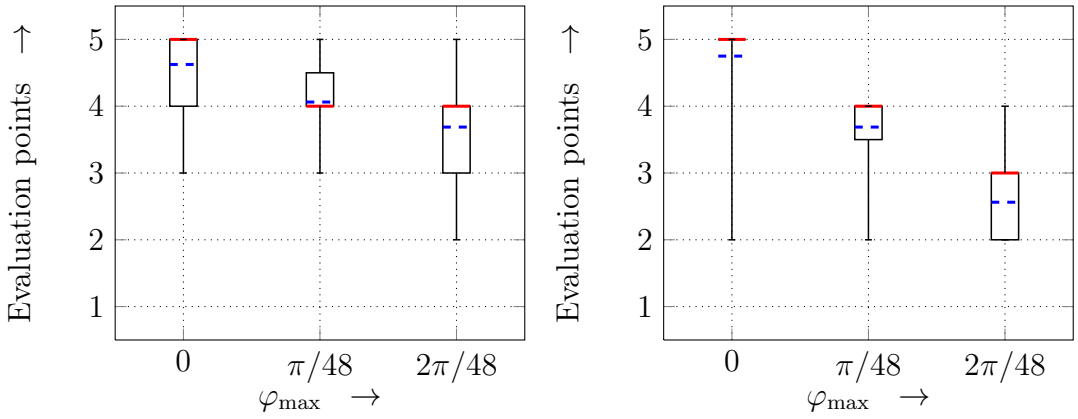
versions: For $\varphi_{\max,1} = \pi/48$ median, and upper quartile remain the same while mean, lower quartile and minimum/maximum value are decreased by 1 point at most. The second version with a higher rotation angle $\varphi_{\max,2} = 2\pi/48$ displays identical behavior, whereas the median is at 3 points and the maximum value equals 5 points. The results for the "Spatial stationarity" are shown in Figure 6.2(b). Here we can see that the wave-domain time-varying filtering creates a perceptible field rotation. The values for median and mean decrease with increasing maximum rotation angle. For $\varphi_{\max,1} = \pi/48$ this two quantities are approximately equal to 4 points (good quality, barely noticeable), whereas the lower quartile and minimum value are dropped by one point compared to the original version. The maximum rotation angle $\varphi_{\max,2} = 2\pi/48$ confirms a more serious subjective perception of the field rotation. The values for mean, median and upper quartile are approximately equal to 3 points (Fair quality, not disturbing). For this monaural sound file we can summarize that the wave-domain time-varying filtering with an maximum rotation angle of $\varphi_{\max,1} = \pi/48$ was noticeable on average but not disturbing. With the increase of $\varphi_{\max}$ the difference to the original version increases. Consequently the evaluation parameters get worse and the distortion becomes serverly disturbing.
As second file we choose the Norwegian band named "Ulver", whereas the music is part of the so called "Ambient" genre. The hearing test results for both evaluation



(a) General quality (b) Spatial stationarity

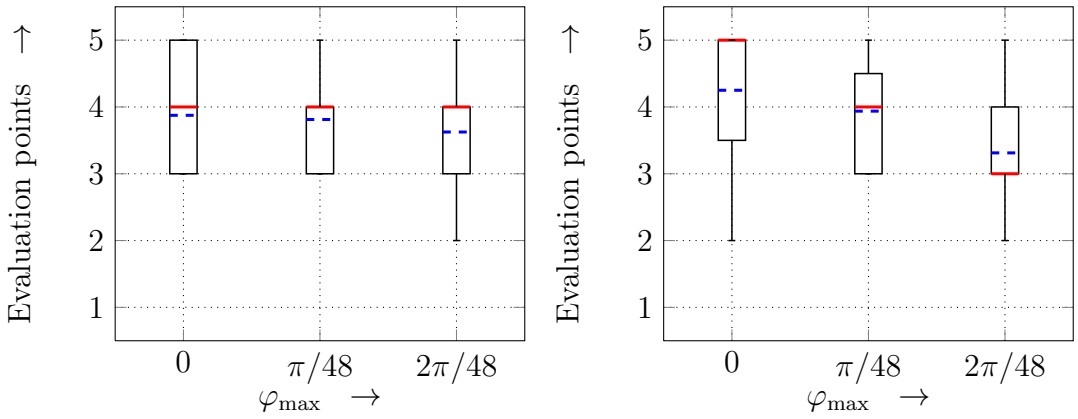
Figure 6.3: Hearing test results for the sound file "Ulver mono"

parameters are shown in Figure 6.3. Compared to the previous sound file we can notice the same tendencies. This can also be verified with a file containing a speech signal. The corresponding result are shown in Figure 6.4. In summary, monaural sound files include perceptible distortions that are essentially reflected with the parameter "Spatial stationarity". With a maximum rotation angle $\varphi_{\max,1} = \pi/48$ the wave-domain time-varying filtering is mainly characterized as noticeable, but not disturbing. Increasing this value is not recommendable due to the increasing perception of the field rotation.
In addition to the three monaural examples we consider two stereophonic sound files. The first one equals the previously used composition of the pianist Franz Liszt. Both evaluation parameters are shown in Figure 6.5. The "General quality" of the original



(a) General quality (b) Spatial stationarity

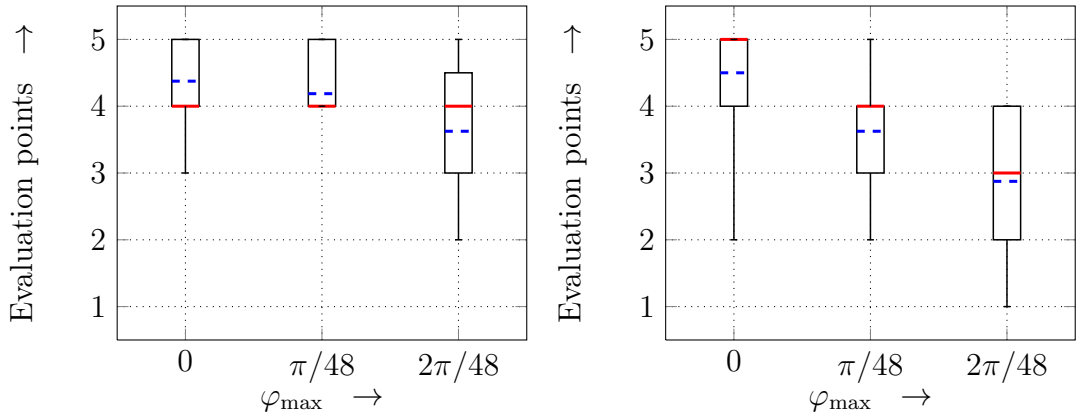
Figure 6.4: Hearing test results for the sound file “Speech Mono”



(a) General quality (b) Spatial stationarity

Figure 6.5: Hearing test results for the sound file “Liszt stereo”

file is characterized by the median of 4 points and the mean of 3.875 points. We can recognize a maximum value equally to the upper quartile at 5 points, whereas the minimum value is identical to the lower quartile at 3 points. In comparison to the original file, the filtered versions have the same value for median, maximum value and lower quartile. Furthermore the mean value slightly decreases and the upper quartile is reduced to a value of 4 points. In summary we can characterize the “General quality” of both filtered version to be acceptable for the subjective perception. The evaluation parameter “Spatial stationarity” is shown in Figure 6.5(b). Comparing the three version, upper quartile, median and mean value decrease with increasing maximum rotation angle. This behavior equals the monaural sound files, whereas the reduction of the evaluation points is less critical. We can notice the mean and median of the filtered version with $\varphi_{\max,1} = \pi/48$ as being equal to 4 points, what corresponds to the statement of a barely noticeable wave field rotation. To verify these tendencies, the previously introduced sound file of the band “Ulver” is used as second stereophonic input signal. The results shown in Figure 6.6 indicate



(a) General quality (b) Spatial stationarity

Figure 6.6: Hearing test results for the sound file “Ulver stereo”

a similar trend as introduced in the previous example.
We can summarize that the monaural sound files imply a more critical influence on the human perception. This seem intuitive because the “Spatial stationarity” indicates the property of a source to be perceived as being located at a fixed position.

6.2 Influence on the system identification

In this section we use the previously made experiences and combine these with the system identification procedure introduced in Chapter 5. Consquently, four

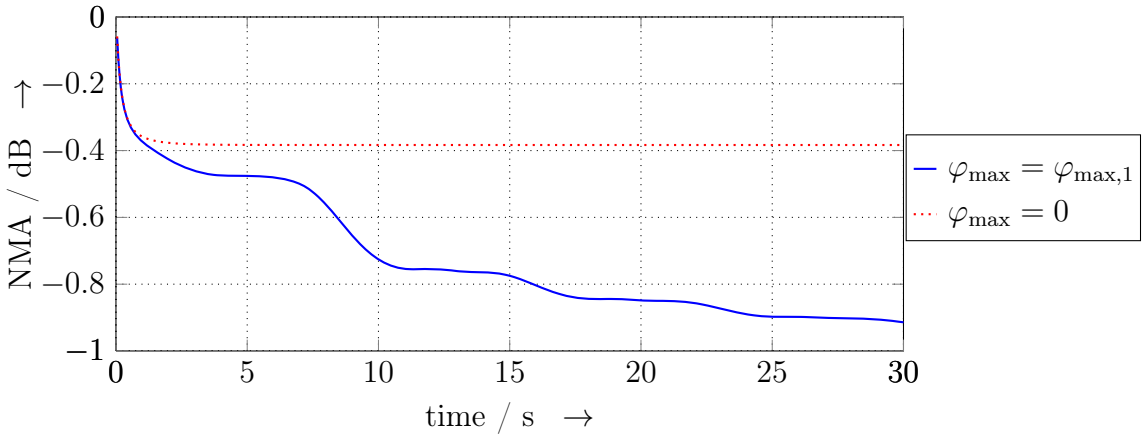


Figure 6.7: NMA with the proposed time-varying filtering

plane waves with normally distributed and statistically independent input signals are generated with the WFS implementation (see Equation 5.2). With this uniform excitation of all loudspeaker signals in the average sense we get the following result for the NMA shown in Figure 6.7 . The wave-domain time-varying filtering enables the improvement of the NMA from initial −0.38 dB to −0.91 dB at the instance

of 30 seconds. This result is combined with an acceptable influence on the human perception which was examined in the previous section.

7 Summary

In this thesis we introduced a new approach to decorrelate the loudspeaker signals in a multichannel AEC scenario. The appertaining challenges due to the so called "non-uniqueness" problem were introduced for the stereophonic case in Chapter 2. With the purpose to improve the system identification of a LEMS, a wave-domain time-varying filtering was proposed and implemented, whereas the wave field was generated applying WFS. The corresponding transformation equations and the realization of the wave field rotation were introduced in Chapter3. For a scenario with two concentric circular arrays, one with 48 loudspeakers and the other with 10 microphones, we used the GFDAF algorithm and the parameters NMA and ERLE for performing the evaluation (see Chapter 4). As simulation results we can summarize a sinusoidal rotation angle function as being most effective (see Chapter 5). Furthermore the system identification improves with increasing maximum rotation angle and decreasing period length. A hearing test was invented in Chapter 6, whereas 16 listeners could evaluate five sound files with each times two filtered and one original version in an arbitrary manner. We oriented on the "Mean Opinion Score" with a scale from 1 to 5, so that 1 point corresponds to a bad (not acceptable) quality. Consequently, 5 points represent excellent characteristics (no noticeable distortions). The listeners could evaluate "General quality", "Artifacts" and "Spatial stationarity". We quantized the results of the hearing test with mean, median, minimum and maximum value as well as lower and upper quartile. To sum up the conclusions, one filtered version with a maximum rotation angle of $\pi/48$ was classified as containing an acceptable deviation from the original file. Furthermore the corresponding absolute values for the evaluation parameters were rated as being justifiable. With this wave-domain time-variant filter we could decrease the NMA of the multichannel AEC scenario from -0.38 dB to -0.91 dB at a time instance of 30 seconds.

A Mathematical preliminaries

Cylindrical coordinate system

The cylindrical coordinate system is illustrated in Figure A.1.

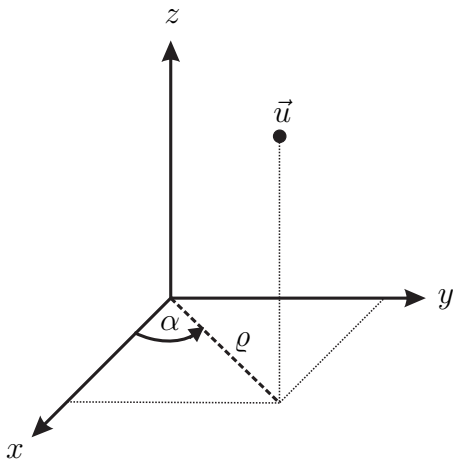


Figure A.1: Cylindrical coordinate system

As transformation from the cartesian coordinate system, the x-y-plane is represented by the distance to the origin ϱ and the azimuth angle α:

$$x = \varrho\cos\alpha, \quad y = \varrho\sin\alpha \quad \text{with} \quad 0 < \alpha < 2\pi, \quad \varrho \geq 0. \tag{A.1}$$

The position vector $\vec{u}$ is constructed with the unit vectors $\vec{e}_\varrho$ and $\vec{e}_z$:

$$\vec{u} = [\varrho, z]^T = \varrho\vec{e}_\varrho + z\vec{e}_z = \varrho\big(\vec{e}_x\cos\alpha + \vec{e}_y\sin\alpha\big) + z\vec{e}_z, \tag{A.2}$$

whereas the wavenumber vector $\vec{k}$ consists of two components:

$$\vec{k} = [k_\varrho, k_z]^T = k_\varrho\vec{e}_\varrho + k_z\vec{e}_z \quad \text{with} \quad |\vec{k}|^2 = k_z^2 + k_\varrho^2 = k^2. \tag{A.3}$$

The Laplace operator in cylindrical coordinates is built as follows:

$$\Delta = \nabla^2 = \frac{\partial^2}{\partial\varrho^2} + \frac{1}{\varrho}\frac{\partial}{\partial\varrho} + \frac{1}{\varrho^2}\frac{\partial^2}{\partial\alpha^2} + \frac{\partial^2}{\partial z^2}. \tag{A.4}$$

Superposition of plane waves

A position vector in the cylindrical coordinate system is defined as in equation A.2.

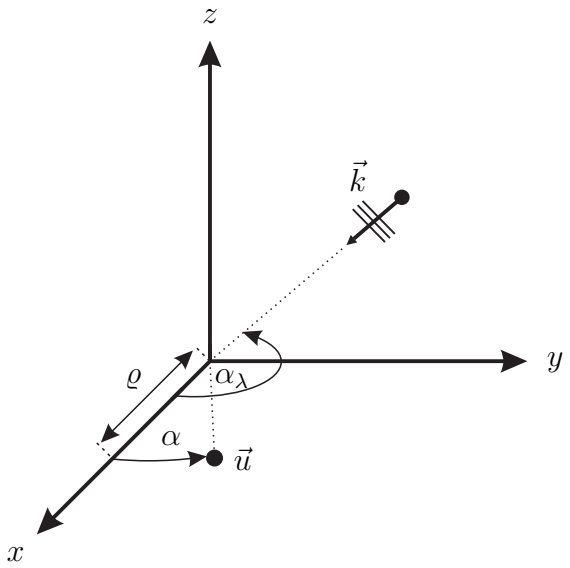


Figure A.2: Superposition of plane waves

The propagation of an incoming plane wave (see figure A.2) is characterized by its wavenumber vector $\vec{k}$, so that we get the following relation for a polar coordinate system at $z = 0$:

$$\vec{u} = \varrho\big(\vec{e}_x \cos\alpha + \vec{e}_y \sin\alpha\big), \tag{A.5}$$

$$\vec{k} = k\big(\vec{e}_x \cos\alpha_\lambda + \vec{e}_y \sin\alpha_\lambda\big). \tag{A.6}$$

The sound pressure at the position $\vec{u} = (\alpha, \varrho)^T$ is now calculated with the use of the far-field propagation of a plane wave:

$$P(\alpha, \varrho, j\omega) = \sum_{\lambda=0}^{N_L-1} \tilde{P}_\lambda^{(p)}(j\omega) \cdot e^{j(\vec{k}\bullet\vec{u})} = \sum_{\lambda=0}^{N_L-1} \tilde{P}_\lambda^{(p)}(j\omega) \cdot e^{j\varrho\cos(\alpha-\alpha_\lambda)k} \tag{A.7}$$

B Overview of the integrated Matlab-files

B.1 Visualization of a field rotation

Main file	**WFSrotation.m**
Get parameters	**get_reproduction_parameters.m** –> get_circ_array_struct.m **get_recording_parameters.m** –> get_circ_array_struct.m **get_room_parameters.m** **get_par.m** **get_source_parameters.m**
Get impulse responses	**get_wfs_impres.m** –> get_physical_frequency.m –> get_greens_function_3D.m –> get_max_diff_2D.m **get_free_field_impulse_response.m** –> get_greens_function_3D.m –> get_max_diff_2D.m
Initialization	**T1.m , T4.m** –> get_physical_frequency.m –> td_cut_3rd_order_tensor_new.m –> td_windowing_3rd_order_tensor_new.m
Transformation, Rotation	**matrix_tensor_conv.m** **T1.m , T4.m** **RotFilter.m**

B.2 Evaluation of the system identification

Main file	**WFSrotation.m**
Get parameters	**get_reproduction_parameters.m** –> get_circ_array_struct.m **get_par.m** **get_source_parameters.m**
Get impulse responses	**get_wfs_impres.m** –> get_physical_frequency.m –> get_greens_function_3D.m –> get_max_diff_2D.m **mmroom2_closed_01_00cm.mat**
Initialization	**T1.m , T4.m** –> get_physical_frequency.m –> td_cut_3rd_order_tensor_new.m –> td_windowing_3rd_order_tensor_new.m **adapt_FDAF.m**
Transformation, Rotation	**matrix_tensor_conv.m** **T1.m , T4.m** **RotFilter.m** **adapt_FDAF.m**

C Notations

C.1 Conventions and abbreviations

In this thesis we use lower case boldface for vectors which can include time samples or filter coefficients. Matrices are represented as upper case boldface. The time instance k of samples is denoted in the argument $x(k)$. This is contrary to the representation of filter coefficients with the time instance κ as subscript h_κ. Integer variables are generally described with their borders: $\mu = 0, ..., 1$. Position vectors used as directional quantities $\vec{u}$ are marked with an arrow. After the transformation into the wave-domain representation an additional superscript $\tilde{x}(k)$ is used to differentiate from the time-domain.
The following abbreviations are introduced:

AEC	acoustic echo cancellation
DFT	discrete Fourier transform
ERLE	echo return loss enhancement
FIR	finite impulse response
GFDAF	general frequency-domain adaptive filtering
LEMS	loudspeaker-enclosure-microphone system
LRS	listening room equalization
NMA	normalized misalignment
WFS	wave field synthesis

C.2 Mathematical Symbols

$(\cdot)^T$	transpose of $(\cdot)$
$(\cdot)^H$	hermitian, i.e. conjugate transpose of $(\cdot)$
$\lvert \cdot \rvert$	absolute value of $\lvert \cdot \rvert$
$\lvert\lvert \cdot \rvert\rvert_2^2$	Euclidean norm of $(\cdot)$ (vector or matrix)
$\lvert\lvert \cdot \rvert\rvert_F^2$	Frobenius norm of $(\cdot)$ (vector or matrix)
Diag$\{\mathbf{x}\}$	generating a diagonal matrix with the vector $\mathbf{x}$ on its main diagonal
Δ	Laplacian operator
$\frac{\partial}{\partial x}$	derivation with respect to x
$\max\{x_i\}$	takes the maximum out of all x_i
$\otimes$	denotes the Kronecker product

D Hearing test questionnaire

Date: _____ Name: _____

This evaluation considers the deterioration of sound quality with a point scale from 1 to 5, whereas more points correspond to a better quality:
5 points: Excellent (not noticeable)
4 points: Good (barely noticeable)
3 points: Fair (noticeable, but not disturbing)
2 points: Poor (clearly noticeable, disturbing)
1 point: Bad (not acceptable)
Five different sound signals with three versions each are taken into account, while the modified versions and the original file appear in an arbitrary sequence. The quantities "general quality", "artifacts" and "spatial stationarity" should be evaluated.

Liszt-mono	General quality					Artifacts					Spatial stationarity				
File 1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1
File 2	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1
File 3	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1

Liszt-stereo	General quality					Artifacts					Spatial stationarity				
File 1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1
File 2	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1
File 3	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1

Ulver-mono	General quality					Artifacts					Spatial stationarity				
File 1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1
File 2	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1
File 3	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1

Ulver-stereo	General quality					Artifacts					Spatial stationarity				
File 1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1
File 2	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1
File 3	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1

Speech-mono	General quality					Artifacts					Spatial stationarity				
File 1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1
File 2	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1
File 3	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1

List of Figures

Bibliography

[1] Dennis R. Morgan Jacob Benesty and Man Mohan Sondhi. A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation. In *IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL.6, NO. 2*, 1998.

[2] Shoji Makino. Stereophonic acoustic echo cancellation: An overview and recent solutions. In *Acoust. Sci. & Tech. 22, 5*, 2001.

[3] Walter Kellermann Jürgen Herre, Herbert Buchner. Acoustic echo cancellation for surround sound using perceptually motivated convergence enhancement. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, 2007.

[4] Dennis R. Morgan M. Mohan Sondhi and Joseph L. Hall. Stereophonic acoustic echo cancellation - an overview of the fundamental problem. In *IEEE SIGNAL PROCESSING LETTERS, VOL. 2, NO. 8*, 1995.

[5] M. Schneider and W. Kellermann. A direct derivation of transforms for wave-domain adaptive filtering based on circular harmonics. In *European Signal Processing Conf. (EUSIPCO)*, August 2012.

[6] Deutsche Gesellschaft für Akustik e.V. Empfehlung 101, 2006.

[7] Sascha Spors. *Active Listening Room Compensation for Spatial Sound Reproduction Systems*. PhD thesis, Friedrich-Alexander Universität Erlangen-Nürnberg, 2005.

[8] M. Schneider and W. Kellermann. A wave-domain model for acoustic mimo systems with reduced complexity. In *Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, Edinburgh, UK, May 2011.

[9] Walter Kellermann Herbert Buchner, Jacob Benesty. Generalized multichannel frequency-domain adaptive filtering: efficient realization and application to hands-free speech communication. In *Signal Processing 85*, September 2005.

[10] M. Schneider, F. Schuh, and W. Kellermann. The generalized frequency-domain adaptive filtering algorithm implemented on a gpu for large-scale multichannel acoustic echo cancellation. In *10. ITG-Fachtagung Sprachkommunikation in Braunschweig*, September 2012.