Design of Crosstalk Cancellation Filters: Combining Inverse Filtering and Optimal Control

Tobias Kabzinski, Florian Hilgemann, Peter Jax
Institute of Communication Systems (IKS), RWTH Aachen University, Aachen, Germany
{kabzinski,hilgemann,jax}@iks.rwth-aachen.de

***Abstract*—Crosstalk cancellation (CTC) systems are crucial for loudspeaker-based reproduction of binaural signals. Two common paradigms for designing CTC filters exist. The first paradigm is based on inverse filtering, which utilizes convex optimization to obtain FIR CTC filters. The second paradigm is based on methods from optimal control, such as H-2 or H-infinity synthesis, to obtain optimal IIR filters in state-space form. In this contribution, we study similarities and differences of both paradigms. Based on our findings, we show how to state FIR CTC filter optimization problems advantageously to address the inherent trade-off between channel separation and equalization in an enhanced way.**

***Index Terms*—Control system design, crosstalk cancellation, inverse filtering, optimal control, weighted least squares**

I. Introduction

Binaural audio has gained increasingly more attention in the past few years. Many applications of binaural audio focus on headphone-based playback. Yet, loudspeaker-based playback of binaural audio can be preferred in certain applications [1]. Loudspeaker-based playback, though, causes crosstalk that alters binaural cues, which enable sound source localization, and crosstalk ultimately degrades the listening experience. As a remedy, CTC filters have been proposed to eliminate the crosstalk.

A large variety of crosstalk filter design methods has been proposed in the past decades. Most design methods yield finite impulse response (FIR) filters as a result of solving a convex minimax or least-squares problem. The problem is either solved in the time domain or in the frequency domain [2]–[7]. Alternatively, it has been proposed to utilize state-space methods that are common in control engineering [8], [9]. In general, these methods yield IIR CTC filters. Moreover, the CTC filter design can either be conducted separately for each individual input channel using a single-input-multiple-output (SIMO) topology, or jointly for all input channels using a multiple-input-multiple-output (MIMO) topology.

Design goals often have frequency-dependent requirements. This can be considered in the control framework by means of so-called plant augmentation using shaping filters. Here, plant refers to the acoustic system. These shaping filters allow, e.g., to apply a frequency weighting to error signals and/or to penalize control effort. The former is encountered in the minimax FIR design method in the simple form of scalar weightings in [3]. The latter is commonly referred to as regularization in the least-squares FIR CTC filter design methods, e.g., in [4].

Previous works have not addressed the links between inverse FIR filter design and optimal control-based IIR filter synthesis for CTC applications. The aim of this paper is three-fold. Firstly, we highlight similarities between both paradigms and point out under which circumstances they coincide. Secondly, we study the impact of the MIMO and SIMO topology on the CTC performance. Thirdly, inspired by mixed-sensitivity controller synthesis, we extend FIR CTC filter design methods by introducing frequency-dependent weights. As a result, the commonly used least-squares or minimax FIR CTC filter design methods now comprise frequency-weighted cost functions, or filter gain limitations. We demonstrate that the more precise control offered by the SIMO topology can improve the CTC performance when using suitable shaping filters.

II. System Models

The CTC system model consists of T loudspeakers, $\mathcal{R}$ input channels, and R receivers (microphones). In contrast to sound field reproduction, where $\mathcal{R} \neq R$, typical CTC setups reproduce one binaural signal for one listener such that $\mathcal{R} = R = 2$.

A. FIR System Model

The CTC system and the filter design problem can be described by representing impulse responses (IRs) of FIR filters as vectors and convolution matrices, as in [5]. The acoustic propagation model, e.g., including head-related transfer functions (HRTFs), from transmitter t to receiver r shall be given by an FIR filter of length N_h as $\mathbf{h}_{rt} = [h_{rt}(0), \ldots, h_{rt}(N_h-1)]^{\mathrm{T}}$. The corresponding $(N_h + N_c - 1) \times N_c$ convolution matrix shall be denoted as $\underline{\mathbf{h}}_{rt}$, and the MIMO convolution matrix is

$$\underline{\underline{\mathbf{h}}} = \begin{bmatrix} \underline{\mathbf{h}}_{11} & \cdots & \underline{\mathbf{h}}_{1T} \\ \vdots & \ddots & \vdots \\ \underline{\mathbf{h}}_{R1} & \cdots & \underline{\mathbf{h}}_{RT} \end{bmatrix}. \tag{1}$$

The N_c-tap FIR CTC filter, in Atal-Schroeder CTC structure [10], from input channel ρ to transmitter t is $\mathbf{c}_{t\rho} = [c_{t\rho}(0), \ldots, c_{t\rho}(N_c-1)]^{\mathrm{T}}$. If $\underline{\underline{\mathbf{h}}}$ is multiplied from the right with the stacked vector of all those CTC filters which receive the ρ-th input signal, $\mathbf{c}_\rho = \left[\mathbf{c}_{1\rho}^{\mathrm{T}}, \ldots, \mathbf{c}_{T\rho}^{\mathrm{T}}\right]^{\mathrm{T}}$, the cascade IRs from input ρ to all receivers are obtained in stacked form. The desired cascade (or target) IR vectors for input ρ, stacked similarly as $\mathbf{d}_\rho = \left[\mathbf{d}_{1\rho}^{\mathrm{T}}, \ldots, \mathbf{d}_{R\rho}^{\mathrm{T}}\right]^{\mathrm{T}}$, include the desired IRs $\mathbf{d}_{r\rho} = [d_{r\rho}(0), \ldots, d_{r\rho}(N_d-1)]^{\mathrm{T}}$ from input ρ to receiver r of length $N_d = N_h + N_c - 1$. To penalize high amplifications in the CTC filters, a regularization term can be introduced into least-squares formulations [4]–[7]. Therefore, we define an FIR regularization filter $\mathbf{b} = [b(0), \ldots, b(N_b-1)]^{\mathrm{T}}$ of length N_b.

B. MIMO State-Space System Model

A MIMO system model without explicit cross-path outputs, similar to [9], is shown in Fig. 1. The quantities of the MIMO topology are indicated by tildes, and we consider z-transforms of discrete-time signals. The listener perceives the R-dimensional output signal $\tilde{\mathbf{Y}}(z) = \left[\tilde{Y}_1(z), \ldots, \tilde{Y}_R(z)\right]^{\mathrm{T}} = \mathbf{H}(z)\,\tilde{\mathbf{U}}(z)$. Here, $\mathbf{H}(z)$ denotes the transfer function matrix (TFM) of the acoustic system. Similarly, the TFM of the CTC filters $\tilde{\mathbf{C}}(z)$ relates the $\mathcal{R}$-dimensional input signal $\mathbf{S}(z) = \left[S_1(z), \ldots, S_{\mathcal{R}}(z)\right]^{\mathrm{T}}$ and the T-dimensional loudspeaker signals $\tilde{\mathbf{U}}(z) = \left[\tilde{U}_1(z), \ldots, \tilde{U}_T(z)\right]^{\mathrm{T}}$, given by $\tilde{\mathbf{U}}(z) = \tilde{\mathbf{C}}(z)\,\mathbf{S}(z)$. The R-dimensional desired cascade output signal is $\tilde{\mathbf{D}}(z) = \tilde{\mathbf{A}}(z)\,\mathbf{S}(z)$. To suppress the crosstalk, $\tilde{\mathbf{A}}$ is chosen to be a diagonal matrix and includes the desired cascade transfer functions (TFs), including a modeling delay [5], on the diagonal entries. The error signal between the desired and the actual signal is $\tilde{\mathbf{E}}(z) = \tilde{\mathbf{D}}(z) - \tilde{\mathbf{Y}}(z)$. In addition, a weighting $\tilde{\mathbf{W}}(z)$ is introduced to enable frequency-selective optimization of the errors, i.e., $\tilde{\mathbf{E}}_w(z) = \tilde{\mathbf{W}}(z)\,\tilde{\mathbf{E}}(z)$, as in [8]. Here, $\tilde{\mathbf{W}}(z) = \mathbf{I}_R \otimes \tilde{W}(z)$ is a diagonal matrix as the same frequency weighting shall be applied to each error signal. This is because the MIMO topology outputs one error signal for each input signal. The plant is augmented by regularization signals $\tilde{\mathbf{E}}_b(z) = \tilde{\mathbf{B}}(z)\,\tilde{\mathbf{U}}(z)$ with a diagonal TFM weighting filter $\tilde{\mathbf{B}}(z)$ to enable control over the filter gain similar to [5]. Note that $\tilde{\mathbf{W}}(z)$ and $\tilde{\mathbf{B}}(z)$ correspond to two of three shaping filters in the widely-used mixed-sensitivity approach [11].

Applying $\mathcal{H}_2$ or $\mathcal{H}_\infty$ control system synthesis methods requires the block diagram in Fig. 1 to be cast into a linear fractional transformation (LFT) form [9], [11], as shown in Fig. 3a. While this might appear to be a feedback scheme at first, it is not. It holds that $\tilde{\mathbf{Q}}(z) = \mathbf{S}(z)$ as the model in Fig. 1 is a feedforward scheme. In the LFT form, the output $\tilde{\mathbf{G}}(z) = \left[\tilde{\mathbf{E}}_w^{\mathrm{T}}(z), \tilde{\mathbf{E}}_b^{\mathrm{T}}(z)\right]^{\mathrm{T}}$ is to be minimized, and the output $\tilde{\mathbf{Q}}(z)$ is fed to the controller $\tilde{\mathbf{K}}(z)$. Note that the term controller stems from control theory, but it essentially corresponds to the CTC filters. The plant $\tilde{\mathbf{P}}(z)$ can be represented as a standard state-space system with $(R+T+\mathcal{R})$-dimensional output $\tilde{\boldsymbol{\gamma}}(k)$ and $(\mathcal{R}+T)$-dimensional input $\tilde{\boldsymbol{\nu}}(k)$, with discrete time index k. We define an $N_{\tilde{x}}$-dimensional state-vector $\tilde{\boldsymbol{x}}$, which essentially contains delayed versions of input samples to $\tilde{\mathbf{P}}(z)$. Then, the summing and filtering in Fig. 1, and the additional interconnections to and from the controller $\tilde{\mathbf{K}}(z)$ in Fig. 3a can be implemented with appropriate state-space matrices [12], i.e., with a state transition matrix $\tilde{\boldsymbol{\mathcal{A}}} \in \mathbb{R}^{N_{\tilde{x}} \times N_{\tilde{x}}}$, an input-to-state matrix $\tilde{\boldsymbol{\mathcal{B}}} \in \mathbb{R}^{N_{\tilde{x}} \times (\mathcal{R}+T)}$, a state-to-output matrix $\tilde{\boldsymbol{\mathcal{C}}} \in \mathbb{R}^{(R+T+\mathcal{R}) \times N_{\tilde{x}}}$, and a feedthrough matrix $\tilde{\boldsymbol{\mathcal{D}}} \in \mathbb{R}^{(R+T+\mathcal{R}) \times (\mathcal{R}+T)}$. The state-space representation can be formalized by

$$\tilde{\boldsymbol{x}}(k+1) = \tilde{\boldsymbol{\mathcal{A}}}\tilde{\boldsymbol{x}}(k) + \tilde{\boldsymbol{\mathcal{B}}}\tilde{\boldsymbol{\nu}}(k)\,, \tag{2a}$$

$$\tilde{\boldsymbol{\gamma}}(k) = \tilde{\boldsymbol{\mathcal{C}}}\tilde{\boldsymbol{x}}(k) + \tilde{\boldsymbol{\mathcal{D}}}\tilde{\boldsymbol{\nu}}(k)\,. \tag{2b}$$

The overall TF from the input signal $\mathbf{S}(z)$ to the LFT model output signal $\tilde{\mathbf{G}}(z)$ is described by the matrix $\tilde{\mathbf{T}}_{\mathbf{S}\tilde{\mathbf{G}}}(z)$.

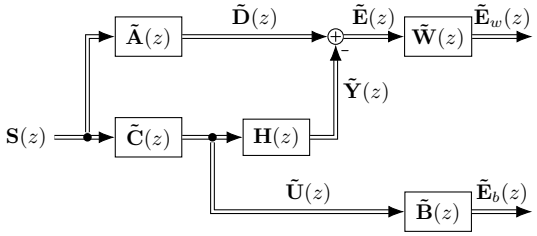


Fig. 1: Augmented MIMO system model

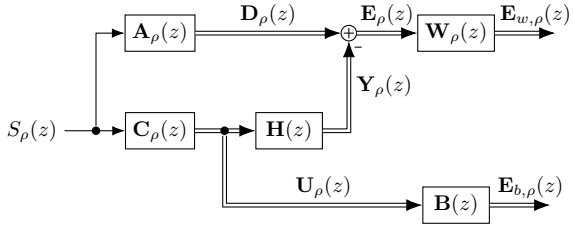


Fig. 2: Augmented SIMO system model for input ρ

C. SIMO State-Space System Model

The error terms in $\tilde{\mathbf{G}}(z)$, which is to be minimized in the MIMO topology, from Sec. II-B appear in aggregated form, i.e., the individual error contributions due to equalization mismatch and crosstalk are indistinguishable. For a better control over the error contributions, we propose to instead consider $\mathcal{R}$ SIMO systems as depicted in Fig. 2. This topology is comparable to the one found in the FIR CTC filter design methods, e.g., in [3], [5]. The desired output for all channels $r \neq \rho$ is zero except for output channel $r = \rho$ such that the columns of the matrix $\tilde{\mathbf{A}}(z)$ are considered as separate vectors $\mathbf{A}_\rho(z)$ instead. This separation allows to explicitly access the crosstalk signals and hence choose different weighting functions on the diagonal of $\mathbf{W}_\rho(z)$ for the errors of the desired direct-path and undesired cross-path signals. This can be exploited to increase the channel separation, as in [3]. In general, the filters $\mathbf{C}(z)$ and $\tilde{\mathbf{C}}(z)$ are not identical (cf. Sec. IV).

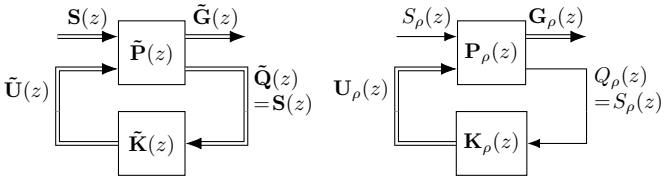
As above, the SIMO systems can be represented in the LFT form as shown in Fig. 3b. The output to be minimized is $\mathbf{G}_\rho(z) = \left[\mathbf{E}_{w,\rho}^{\mathrm{T}}(z), \mathbf{E}_{b,\rho}^{\mathrm{T}}(z)\right]^{\mathrm{T}}$. Each SIMO system defines a separate state-space system with N_x states as

$$\boldsymbol{x}_\rho(k+1) = \boldsymbol{\mathcal{A}}\boldsymbol{x}_\rho(k) + \boldsymbol{\mathcal{B}}\boldsymbol{\nu}_\rho(k)\,, \tag{3a}$$

$$\boldsymbol{\gamma}_\rho(k) = \boldsymbol{\mathcal{C}}_\rho\boldsymbol{x}_\rho(k) + \boldsymbol{\mathcal{D}}_\rho\boldsymbol{\nu}_\rho(k) \tag{3b}$$

with $\boldsymbol{\mathcal{A}}_\rho \in \mathbb{R}^{N_x \times N_x}$, $\boldsymbol{\mathcal{B}}_\rho \in \mathbb{R}^{N_x \times (1+T)}$, $\boldsymbol{\mathcal{C}}_\rho \in \mathbb{R}^{(R+T+1) \times N_x}$, and $\boldsymbol{\mathcal{D}}_\rho \in \mathbb{R}^{(R+T+1) \times (1+T)}$. Here, the overall TF $\mathbf{T}_{S_\rho\mathbf{G}_\rho}(z)$ is vector-valued as it relates a single channel to one direct path, $R-1$ cross paths and T augmented regularization outputs. Note that we refer to this system as SIMO because the controller is a SIMO system.

A noteworthy advantage of the SIMO topology is that the computational complexity of the design method can be reduced as designing $\mathcal{R}$ controllers for state-space systems of order N_x can be computationally cheaper than designing one controller for a state-space system of larger order $N_{\tilde{x}}$.



(a) MIMO controller model (b) SIMO controller model

Fig. 3: LFT models

III. Extensions of FIR CTC Filter Design Methods

We now present extensions to FIR CTC filter design methods found in the literature.

A. Least-Squares Formulations

Inspired by the error weighting commonly applied in the augmented plant in the mixed-sensitivity approach [11], we extend the regularized time-domain least-squares method from [5] to include separate FIR weighting filters $w_\mathrm{d}(k)$ and $w_\mathrm{c}(k)$ of length N_w for the direct path and the cross paths, respectively. This yields different problems for each input channel ρ, namely,

$$\min_{\mathbf{c}_\rho} \left\| \underline{\underline{\mathbf{w}}}_\rho \left(\mathbf{d}_\rho - \underline{\underline{\mathbf{h}}} \mathbf{c}_\rho \right) \right\|_2^2 + \| (\mathbf{I}_T \otimes \underline{\mathbf{b}}) \mathbf{c}_\rho \|_2^2 \tag{4}$$

with $\underline{\mathbf{b}}$ corresponding to the convolution matrix for the regularization filter $\mathbf{b}$ and the block diagonal weighting matrix $\underline{\underline{\mathbf{w}}}_\rho$ containing the convolution matrices corresponding to the weighting filters. E.g., for $T = R = \mathcal{R} = 2$ and $\rho = 1$:

$$\underline{\underline{\mathbf{w}}}_1 = \begin{bmatrix} \underline{\mathbf{w}}_\mathrm{d} & \mathbf{0}_{(N_d+N_w-1)\times N_d} \\ \mathbf{0}_{(N_d+N_w-1)\times N_d} & \underline{\mathbf{w}}_\mathrm{c} \end{bmatrix}. \tag{5}$$

The solution to (4) is given by

$$\mathbf{c}_\rho = \left(\underline{\underline{\mathbf{h}}}^\mathrm{T} \underline{\underline{\mathbf{w}}}_\rho^\mathrm{T} \underline{\underline{\mathbf{w}}}_\rho \underline{\underline{\mathbf{h}}} + \mathbf{I}_T \otimes \left(\underline{\mathbf{b}}^\mathrm{T} \underline{\mathbf{b}} \right) \right)^{-1} \underline{\underline{\mathbf{h}}}^\mathrm{T} \underline{\underline{\mathbf{w}}}_\rho^\mathrm{T} \underline{\underline{\mathbf{w}}}_\rho \mathbf{d}_\rho. \tag{6}$$

In contrast to [5], a different system of equations (6) needs to be solved for each input channel ρ now, which slightly increases the computational complexity. While this change due to the weightings might appear insignificant, the impact on the channel separation can be significant (cf. Sec. V). Scalar weightings, i.e., $N_b = 1$ as in [3], are included in (4) as a special case, and then (6) becomes the standard weighted least-squares solution [13]. The problem (4) with scalar weightings is equivalent to the personal sound zones filter design problem in [14] for exactly one receiver in the bright zone and $R-1$ receivers in the dark zone.

The frequency-dependent weightings can straightforwardly be included into the frequency-domain least-squares CTC filter design problem in its various forms [4], [6], [7]. Due to the potential real-time capability of these methods, they are especially attractive for dynamic scenarios, in which the CTC filters need to be updated multiple times per second.

B. Minimax Formulations

In [3], it is suggested to solve a convex minimax problem in the time domain to obtain CTC filters in the Atal-Schroeder structure. Scalar weightings for the errors in the direct path and cross path are included there. Contrastingly, it is suggested in [2] to solve a problem that minimizes a maximum frequency-domain error, to obtain filters for the single-filter structure. This structure, however, is restricted to symmetrical CTC setups. Both problems can be conceptually combined to obtain CTC filters in the Atal-Schroeder structure. This leads to a worst-case frequency-domain error minimization, similar to the $\mathcal{H}_\infty$ synthesis cost function (cf. IV-B). With $\bar{\mathbf{F}}$ denoting the $M \times M$ unitary Discrete Fourier Transform (DFT) matrix, a matrix that extracts a symmetric half of the DFT spectrum from a zero-padded input signal of length $N_\mathrm{in} \leq M$ is given by

$$\mathbf{F}_{M\times N_\mathrm{in}} = \begin{bmatrix} \mathbf{I}_{\lfloor \frac{M}{2}+1 \rfloor} & \mathbf{0}_{M-\lfloor \frac{M}{2}-1 \rfloor} \end{bmatrix} \bar{\mathbf{F}}_M \begin{bmatrix} \mathbf{I}_{N_\mathrm{in}} \\ \mathbf{0}_{(M-N_\mathrm{in})\times N_\mathrm{in}} \end{bmatrix},$$

where $\lfloor \cdot \rfloor$ is the floor function. This allows to write an extended frequency-domain minimax problem as follows:

$$\begin{aligned} &\min_{\mathbf{c}_\rho} \left\| \left(\mathbf{I}_R \otimes \mathbf{F}_{M\times(N_d+N_w-1)} \right) \underline{\underline{\mathbf{w}}}_\rho \left(\mathbf{d}_\rho - \underline{\underline{\mathbf{h}}} \mathbf{c}_\rho \right) \right\|_\infty \\ &\text{subject to } \left| \left(\mathbf{I}_R \otimes \mathbf{F}_{M\times(N_d+N_w-1)} \right) \mathbf{c}_\rho \right| \leq \boldsymbol{\mathcal{G}}. \end{aligned} \tag{7}$$

Here, $|\cdot|$ denotes the elementwise absolute value, and $\boldsymbol{\mathcal{G}} = \left[\boldsymbol{\mathcal{G}}_1^\mathrm{T}, \ldots, \boldsymbol{\mathcal{G}}_T^\mathrm{T} \right]^\mathrm{T} \in \mathbb{R}^{T\lfloor \frac{M}{2}+1 \rfloor}$ includes the frequency-dependent gain limits $\mathcal{G}_t(\mu)$ at frequency index μ for each transmitter t. The gain limits could be chosen similarly to those resulting from regularization filters in the least-squares formulation, where the gain at each frequency is approximately limited to $1/(2|B(\mu)|)$. Defining the gain limit as a hard constraint increases the computational complexity but also guarantees the gain limit rather than the "soft" gain constraint induced by the regularization in the least-squares formulation.

Without the frequency-domain transform term $\left(\mathbf{I}_R \otimes \mathbf{F}_{M\times(N_d+N_w-1)} \right)$ in (7), our formulation becomes a generalization of the minimax problem in [3]—now with frequency-dependent weighting filters.

IV. Relations of CTC Filter Design Methods

We now relate inverse filtering-based FIR to optimal control-based IIR CTC filter design methods.

A. $\mathcal{H}_2$ Synthesis and Time-Domain Least-Squares

A common tool for least-squares controller design is the $\mathcal{H}_2$ synthesis [11], [15]. It is based on the solution of two algebraic Riccati equations. For the MIMO systems above, the $\mathcal{H}_2$ synthesis aims to find a state-space controller $\tilde{\mathbf{K}}(z)$ that minimizes

$$\tilde{\mathcal{J}}^{(2)} = \left\| \tilde{\mathbf{T}}_{\mathbf{S}\tilde{\mathbf{G}}}(\Omega) \right\|_2^2 = \frac{1}{2\pi} \int_0^{2\pi} \mathrm{tr} \left\{ \tilde{\mathbf{T}}_{\mathbf{S}\tilde{\mathbf{G}}}(\Omega) \, \tilde{\mathbf{T}}_{\mathbf{S}\tilde{\mathbf{G}}}^\mathrm{H}(\Omega) \right\} \mathrm{d}\Omega.$$

Using the system model in Fig. 1 and neglecting the weighting $\tilde{\mathbf{W}}(z)$, it can be shown this is

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\sum_{r=1}^{R}\sum_{\rho=1}^{\mathcal{R}}\left|\left[\mathbf{H}\tilde{\mathbf{C}}-\tilde{\mathbf{A}}\right]_{r\rho}\right|^2+\sum_{t=1}^{T}\sum_{\rho=1}^{\mathcal{R}}\left|\left[\tilde{\mathbf{B}}\tilde{\mathbf{C}}\right]_{t\rho}\right|^2\right)\mathrm{d}\omega,$$

where we discarded the dependency on the normalized angular frequency $0 \leq \Omega < 2\pi$, defined as $\Omega = 2\pi f/f_\mathrm{s}$ with sampling frequency f_s, for brevity, and $[\cdot]_{ij}$ denotes the i-th row and j-th column of a matrix.

For the SIMO case, the $\mathcal{H}_2$ synthesis minimizes

$$\begin{aligned}\mathcal{J}_\rho^{(2)} &= \left\|\mathbf{T}_{S_\rho\mathbf{G}_\rho}(\Omega)\right\|_2^2 \\ &= \frac{1}{2\pi}\int_0^{2\pi}\left(\sum_{r=1}^{R}\left|\left[\mathbf{H}\mathbf{C}_\rho-\mathbf{A}_\rho\right]_{r1}\right|^2+\sum_{t=1}^{T}\left|\left[\mathbf{B}\mathbf{C}_\rho\right]_{t1}\right|^2\right)\mathrm{d}\Omega.\end{aligned} \tag{8}$$

From the second line of (8) it can be seen that $\tilde{\mathcal{J}}^{(2)} = \sum_\rho \mathcal{J}_\rho^{(2)}$. This implies that the MIMO and SIMO topology yield the same results for the $\mathcal{H}_2$ synthesis, i.e., $\tilde{\mathbf{C}} = \left[\mathbf{C}_1, \ldots, \mathbf{C}_\mathcal{R}\right]$. Through the $\mathrm{tr}\left\{(\cdot)\,(\cdot)^\mathrm{H}\right\}$ operation each element of the overall TF is considered in the optimization, even though the output signals only contain sums. Note that as soon as a weighting function is introduced that depends on ρ, the equality is violated.

The $\mathcal{H}_2$ synthesis minimizes the sum of all energies of the overall TF from input signal to the error outputs, which can be expressed equivalently [15] in the time domain as $\mathcal{J}_\rho^{(2)} = \sum_{k=0}^{\infty}\mathbf{t}_{S_\rho\mathbf{G}_\rho}^2(k)$, where $\mathbf{t}_{S_\rho\mathbf{G}_\rho}(k)$ is the vector of the corresponding overall IRs. As the time-domain least-squares formulation considers just the energies up to $k = N_d$, or $k = N_d + N_w - 1$ for the case with error weightings, and up to $k = N_b + N_c - 1$ for the regularization outputs, its cost functions converges to the one of the $\mathcal{H}_2$ synthesis for $N_c \to \infty$. Hence, the time-domain least-squares CTC filters approximate those of the $\mathcal{H}_2$ synthesis for sufficiently long filters.

B. $\mathcal{H}_\infty$ Synthesis and Minimax Optimization

The discrete-time $\mathcal{H}_\infty$ synthesis [11], [15] minimizes

$$\tilde{\mathcal{J}}^{(\infty)} = \left\|\tilde{\mathbf{T}}_{\mathbf{S}\tilde{\mathbf{G}}}(\Omega)\right\|_\infty = \max_\Omega \sigma\left(\tilde{\mathbf{T}}_{\mathbf{S}\tilde{\mathbf{G}}}(\Omega)\right), \tag{9}$$

where $\sigma(\cdot)$ denotes the singular values of a matrix. For the SIMO topology for channel ρ, the $\mathcal{H}_\infty$ synthesis minimizes

$$\mathcal{J}_\rho^{(\infty)} = \left\|\mathbf{T}_{S_\rho\mathbf{G}_\rho}(\Omega)\right\|_\infty. \tag{10}$$

As the single singular value of an $N \times 1$ matrix is the square root of the sum of the squared absolute value entries, the SIMO cost function is different from MIMO cost function.

The difference between the SIMO and MIMO topologies can be leveraged as follows: If $\mathcal{J}_1^{(\infty)}$ attains its maximum error at Ω_0 and $\mathcal{J}_2^{(\infty)}$ attains its larger maximum error at the same frequency Ω_0, the two SIMO solutions would find two different minimal errors. However, the MIMO formulation would only minimize one maximum error so that the system for $\rho = 1$ with the smaller error could have been further improved.

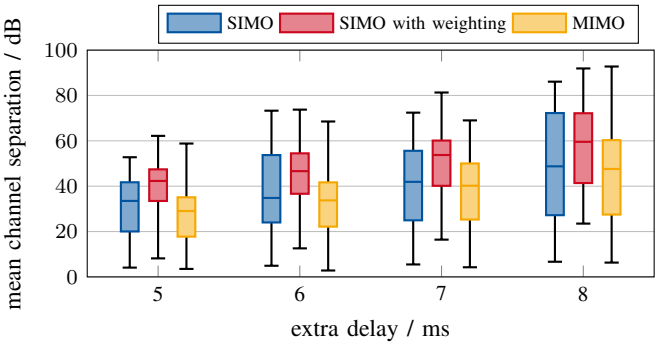


Fig. 4: Comparison of MIMO and SIMO topology with and without cross-path error weighting for $\mathcal{H}_\infty$ synthesis

For a sufficiently large number M of discrete frequency points, the DFT approximates the discrete-time Fourier transform (DTFT). If a large enough filter length N_c is chosen in this case, the solution of (7) without gain limitation approximates the $\mathcal{H}_\infty$ solution for the SIMO topology if the crossterm can go to zero, i.e., for a sufficiently long modeling delay. Therefore, both (7) and (10) minimize the maximum frequency-domain error if only one nonzero term occurs in $\mathbf{T}_{S_\rho\mathbf{G}_\rho}(\Omega)$.

The minimax formulation with the explicit gain limit appears to be more convenient as it requires less manual tuning than required in adjusting the weighting filters in the $\mathcal{H}_\infty$ synthesis.

V. Evaluation

To evaluate our findings, we conducted both simulations and acoustic measurements.

A. Simulation: $\mathcal{H}_\infty$ Synthesis for MIMO and SIMO Systems

We simulated CTC setups with two loudspeakers placed at $\pm 45°$ in the horizontal plane for all 96 subjects of the HUTUBS HRTF database [16]. The measured HRTFs were resampled to a sampling rate of 32 kHz, and the desired cascade response was a 4th-order Butterworth bandpass with a passband from 150 Hz to 15 kHz. Frequencies below 150 Hz are commonly reproduced monaurally. We applied the $\mathcal{H}_\infty$ synthesis to both the MIMO and SIMO topologies for different modeling delays, adding extra delay to the Butterworth IRs. No regularization was used. A weighting function was used that attenuates errors above and below the passband edges of the desired response by 20 dB and has unit magnitude elsewhere. To demonstrate the impact of the weighting function, we simulated the SIMO variant again and increased the error weighting in the passband of the cross path by 20 dB. Note that this weighting can only be applied in the SIMO topology. To simplify the comparison, we evaluated the frequency-averaged channel separations, as defined in [3], in the frequency range from 300 Hz to 14 kHz, where the desired cascade response is mostly flat.

Fig. 4 depicts the mean channel separation percentiles for 0%, 25%, 50%, 75%, and 100%. The channel separations tend to increase for the SIMO topology. With the weighting, the values are further enhanced as reducing the errors in the cross paths is incentivized. For shorter delays, the inversion becomes more difficult and hence the performance decreases. In these

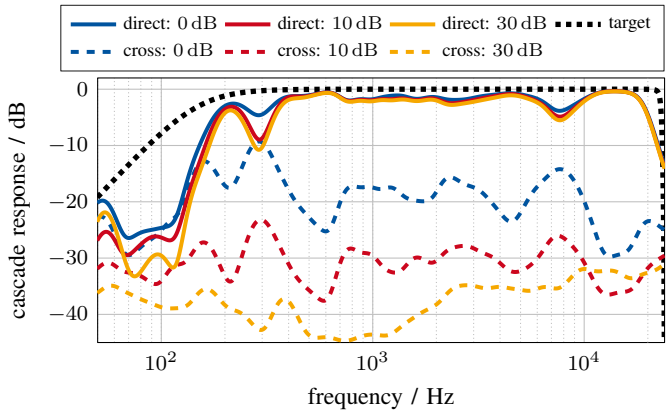


Fig. 5: Measured direct-path and cross-path magnitudes at left microphone for different cross-path error weightings

cases, it might be more likely that $\mathcal{J}_1^{(\infty)}$ and $\mathcal{J}_2^{(\infty)}$ significantly differ, which could be exloited in the SIMO topology.

B. Real-World Measurements

We implemented a CTC system and measured the performance to evaluate the impact of cross-path error weighting functions not only on the channel separation but also on the distortion of the desired direct-path responses. Binaural room impulse responses (BRIRs) from two loudspeakers (Neumann KH120A) to a dummy head (Neumann KU100) were measured. The loudspeakers spanned an angle of about $40°$ in the horizontal plane, and the distance was about $1.2\,\mathrm{m}$. All measurements were conducted in a soundproof booth (Studiobox Premium) of dimension $1.8\,\mathrm{m} \times 2.2\,\mathrm{m} \times 3\,\mathrm{m}$ with a reverberation time of about $150\,\mathrm{ms}$. The $150\,\mathrm{ms}$-long measured IRs were used to design CTC filters of length $N_c = 4096$ at a sampling rate of $48\,\mathrm{kHz}$ according to (6). A regularization gain limit was applied, and the modeling delay was chosen to be about $6.5\,\mathrm{ms}$ longer than the delay of first peak in the BRIRs. The desired response was a 4th-order Butterworth bandpass filter ($150\,\mathrm{Hz}$ to $23\,\mathrm{kHz}$). The error weighting outside the passband was $-20\,\mathrm{dB}$ for the direct path and the cross path.

Fig. 5 shows the measured cascade TF magnitudes at the left dummy head microphone for different cross-path error weightings in the passband. For better readability, a one-third octave band smoothing was applied. Above $400\,\mathrm{Hz}$ the differences in the direct paths do not exceed $2\,\mathrm{dB}$ while the crosstalk decreases by up to $20\,\mathrm{dB}$. The magnitude response dip at about $300\,\mathrm{Hz}$ is due to a BRIR notch that is difficult to invert [17]. The equalization becomes $5\,\mathrm{dB}$ to $10\,\mathrm{dB}$ worse for increased cross-path error weightings as relatively less emphasis is put on achieving a good equalization of the direct path compared to the crosstalk suppression. According to [18], the minimum audible channel separation is $15\,\mathrm{dB}$ to $20\,\mathrm{dB}$. Hence, the weighting could help to push the channel separation in a real-world CTC system above this threshold.

VI. Summary

We have analyzed that the regularized time-domain least-squares CTC FIR filter design method with flat error weighting is virtually identical to filters resulting from the $\mathcal{H}_2$ synthesis for both SIMO and MIMO topologies if the FIR filters are sufficiently long. Further, we have shown that a joint filter optimization for all channels, as pursued by conventional MIMO approaches, can be improved upon in the $\mathcal{H}_\infty$ case by opting for the SIMO topology, which is commonly used in the FIR CTC filter design methods. The channel separation can be further increased by amplifying the weighting of the cross-path error which is explicitly accessible in the SIMO topology. Inspired by mixed-sensitivity loop shaping, we have proposed to incorporate frequency-selective weighting filters into the optimization of FIR CTC filters. Based on simulations and measurements of a prototype CTC system, we have shown that this can lead to greater channel separation.

References

[1] A. Roginska and P. Geluso, *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*. New York, NY, USA: Taylor & Francis, 2017.

[2] H. I. Rao, V. J. Mathews, and Y.-C. Park, "Inverse filter design using minimax approximation techniques for 3-D audio," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.*, vol. 5, 2006, pp. 1–5.

[3] ——, "A minimax approach for the joint design of acoustic crosstalk cancellation filters," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 15, no. 8, pp. 2287–2298, 2007.

[4] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization," *IEEE Trans. Speech and Audio Process.*, vol. 6, no. 2, pp. 189–194, 1998.

[5] O. Kirkeby and P. A. Nelson, "Digital filter design for inversion problems in sound reproduction," *J. Audio Eng. Soc.*, vol. 47, no. 7/8, pp. 583–595, 1999.

[6] B. Masiero and M. Vorländer, "A framework for the calculation of dynamic crosstalk cancellation filters," *IEEE/ACM Trans. Audio, Speech, and Lang. Process.*, vol. 22, no. 9, pp. 1345–1354, 2014.

[7] T. Kabzinski and P. Jax, "A causality-constrained frequency-domain least-squares filter design method for crosstalk cancellation," *IEEE/ACM Trans. Audio, Speech, and Lang. Process.*, vol. 29, pp. 2942–2956, 2021.

[8] T. Samejima, I. Taniguchi, and H. Kitajima, "Application of H-infinity control theory to sound reproduction systems," in *Proc. Int. Conf. Acoust. 2004*, pp. 1943–1946.

[9] S. Mori, M. Inoue, K. Matsui, and S. Adachi, "Stable inversion design for binaural reproduction over loudspeakers," in *IEEE Conf. Control Technol. and Appl.*, 2017, pp. 1213–1217.

[10] B. S. Atal and M. R. Schroeder, "Apparent sound source translator," Feb. 1966, U.S. Patent 3 236 949.

[11] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design*. Hoboken, NJ, USA: John Wiley & Sons, 2005.

[12] E. L. Duke, "Combining and connecting linear, multi-input, multi-output subsystem models," Ames Research Center, Dryden Flight Research Facility, Edwards, CA, USA, Tech. Rep., 1986.

[13] G. Strang, *Introduction to Linear Algebra*, 4th ed. Wellesley, MA, USA: Wellesley-Cambridge Press, 2009.

[14] M. F. S. Gálvez, S. J. Elliott, and J. Cheer, "Time domain optimization of filters used in a loudspeaker array for personal audio," *IEEE/ACM Trans. Audio, Speech, and Lang. Process.*, vol. 23, no. 11, pp. 1869–1878, 2015.

[15] W. Gawronski, "Discrete-time norms of flexible structures," *J. Sound and Vib.*, vol. 264, no. 5, pp. 983–1004, 2003.

[16] F. Brinkmann, M. Dinakaran, R. Pelzer, P. Grosche, D. Voss, and S. Weinzierl, "A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses," *J. Audio Eng. Soc.*, vol. 67, no. 9, pp. 705–718, 2019.

[17] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 36, no. 2, pp. 145–152, 1988.

[18] Y. Lacouture-Parodi and P. Rubak, "A subjective evaluation of the minimum channel separation for reproducing binaural signals over loudspeakers," *J. Audio Eng. Soc.*, vol. 59, no. 7/8, pp. 487–497, 2011.