A Framework for the Calculation of Dynamic Crosstalk Cancellation Filters

Bruno Masiero, *Member, IEEE,* Michael Vorländer

***Abstract*—Dynamic crosstalk cancellation (CTC) systems commonly find use in immersive virtual reality (VR) applications. Such dynamic setups require extremely high filter update rates, so filter calculation is usually performed in the frequency-domain for higher efficiency. This paper proposes a general framework for the calculation of dynamic CTC filters to be used in immersive VR applications. Within this framework, we introduce a causality constraint to the frequency-domain calculation to avoid undesirable wrap-around effects and echo artefacts. Furthermore, when regularization is applied to the CTC filter calculation, in order to limit the output levels at the loudspeakers, noncausal artefacts appear at the CTC filters and the resulting ear signals. We propose a global minimum-phase regularization to convert these anti-causal ringing artefacts into causal artefacts. Finally, an aspect that is especially critical for dynamic CTC systems is the filter switch between active loudspeakers distributed in a surround audio-visual display system with 360° of freedom of operator orientation. Within this framework we apply a weighted filter calculation to control the filter switch, which allows the loudspeakers' contribution to be windowed in space, resulting in a smooth filter transition.**

***Index Terms*—binaural technique, dynamic crosstalk cancellation, causal implementation, minimum-phase regularization.**

I. Introduction

SPATIAL cues contained in a binaural signal may be severely degraded if these signals are reproduced directly through loudspeakers. To reestablish these cues, *crosstalk cancellation* (CTC) filters are used to generate (from the input binaural signal) the transaural signals to be fed to the loudspeakers. These signals will interact to reestablish the binaural signal at the listener's ears with sufficient channel separation—i.e. sufficient cancellation of the crosstalk path without severely altering the direct path.

The common scenario envisioned for binaural reproduction is that when the listener is seated and makes no large movement with his/her head during the presentation of the binaural scene. In this situation, a static set of CTC filters is used. Ideally, these filters should be robust to small head movements, which is equivalent to say that the CTC filters should generate a large reproduction area, also known as sweet spot. As these filters are static, large delays are acceptable to ensure filter causality.

Contrary to the static situation, in a virtual reality (VR) environment the user should be allowed to freely move inside the reproduction space [1]. To compensate for the listener movement, a dynamic system is necessary where the CTC filters are constantly updated to compensate for the movement of the listener [2]. This paper focuses on CTC filters for dynamic systems.

To allow high update rate, the CTC filter calculation should be conducted as fast as possible. Digital CTC filters can be calculated either in the time- or in the frequency-domain. Frequency-domain calculations are computationally more efficient then their time-domain counterpart. This is even emphasized by the fact that most parts of the VR signal processing such as fast convolution of audio with binaural room impulse responses is done in the frequency domain anyway. However, it has the drawback that the resulting CTC filters might present a noncausal behavior that can lead to audible artefacts—while time-domain calculation produces strictly causal filters [3]. We show that the use of a time window to eliminate such noncausal components can reduce the obtained channel separation. The calculation of the causal response obtained by the time-domain method can be approximated in the frequency-domain by using the Wiener-Hopf decomposition [4]. In this article the time-domain deconvolution method presented in [5] is substituted by a causality constrained MIMO frequency-domain calculation.

Crosstalk cancellation is achieved by means of constructive and destructive wave interference and at some critical frequencies, loudspeakers are required to produce very high sound pressure only to be later canceled at the listener's ears [6]. To avoid clipping and distortion at these frequencies, the overall gain of the CTC filter has to be reduced, causing the dynamic range of the reproduced binaural signal to shrink. These frequencies with extreme high energy also result in a poorly damped ringing behavior in the time-domain, which can also be understood a very narrow sweet spot in the spatial domain [7].

Kirkeby proposed the use of Tikhonov regularization to control undesirably large peaks in the frequency response of the CTC filters [8]. Regularization not only limits the gain in frequency-domain but also reduces the length of the CTC filters [9]. One side-effect of regularization is the appearance of unwanted noncausal artefacts in both the CTC filters and the resulting ear signals [10], [11]. The regularization procedure can be altered so that pre-ringing components are converted into post-ringing in the resulting ear signal [11]. For the multichannel case, the method presented in [12] has the drawback that the minimum-phase correction has to be made for each channel individually. This procedure can generate an interaural phase difference [13], which in turn can compromise the quality of the reproduced binaural scene.

The authors are with the Institute of Technical Acoustics, RWTH Aachen University, 52066 Aachen, Germany. e-mail: bma@akustik.rwth-aachen.de.

B. Masiero was financed by CNPq, Brazil.

Manuscript received xx-xx-xx; revised xx-xx-xx.

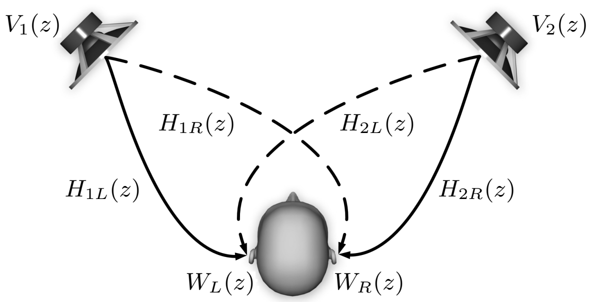

Fig. 1. Diagram of binaural reproduction through loudspeakers. The direct (solid line) and crosstalk acoustic paths (dashed line) are displayed.

In this paper we propose a method to approximate a global minimum-phase regularization, avoiding any such interaural phase discrepancies.

Another important aspect of the dynamic CTC systems is the switching of the CTC filters between active loudspeakers. CTC works well when the listener's head is pointing within the angle spanned between (at least) two active loudspeakers. Outside of this area, the filters can become unstable [2]. For these applications, a reduced number (at least three) of loudspeakers surrounding the reproduction space should be used. Lentz designed a setup with four loudspeakers from which only two loudspeakers are active at one time, depending on the orientation of the listener's head [2]. Cross-fading is used to switch between active loudspeaker pairs. During the cross-fading operation three loudspeakers will be active and the switching might lead to audible artefacts.

Within the described framework, a larger number of loudspeakers could be used for the binaural reproduction and a different solution for the filter switching strategy is described, where fading is incorporated into the filter design stage. In this way all active loudspeakers are simultaneously considered during the filter design process. It can be expected that filters calculated in this way will result in a system with flatter response and sufficient channel separation, thus reproducing the binaural signals with higher fidelity.

This paper starts by reviewing the principles of the crosstalk cancellation technique in section II. Section III discusses a general framework for the calculation of digital CTC filters in the frequency-domain. The presented filter calculation framework can cope with all the above mentioned effects. Cyclic aliasing, and thus noncausal CTC filters, are dealt with by applying a causality constraint in the frequency-domain calculation. The effects of regularization, like noncausal artefacts in the resulting ear signal, are reduced by conducting a global minimum-phase regularization that adds no interaural phase discrepancies to the resulting ear signal. Further, to avoid artefacts when switching between active loudspeakers, a simple weighted matrix inversion is used to provide a smooth filter update. Simulation results are presented within IV. The paper concludes with a discussion of the presented results in section V.

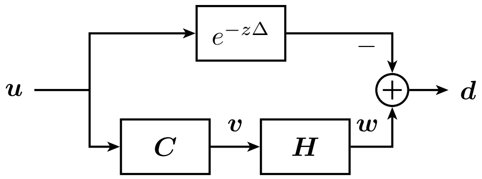

Fig. 2. The crosstalk cancellation problem in block diagram form.

II. Binaural Reproduction through Loudspeakers

When reproducing binaural signals through loudspeakers, a set of filters has to be used to achieve the required cancellation of the crosstalk paths (depicted with a dashed line in fig. 1). These filters are called crosstalk cancellation (CTC) filters.

Based on fig. 1 the binaural reproduction with two loudspeakers can be described in the frequency-domain as ($z = e^{-j\omega}$)

$$W_L(z) = H_{1L}(z)V_1(z) + H_{2L}(z)V_2(z) \overset{!}{=} U_L(z) \tag{1a}$$

$$W_R(z) = H_{1R}(z)V_1(z) + H_{2R}(z)V_2(z) \overset{!}{=} U_R(z), \tag{1b}$$

where $W_L(z)$ and $W_R(z)$ are the actual binaural signal arriving at the listener's left and right ears respectively, the so-called *ear signals*, whereas $U_L(z)$ and $U_R(z)$ are the desired left and right *binaural signals*. $V_1(z)$ and $V_2(z)$ are the *loudspeaker signals* (here feeding the left and right loudspeakers respectively). $H_{\mathrm{ls,ear}}$ is the head-related transfer function (HRTF) from a given loudspeaker ls to a given ear, also called *playback* HRTF.

Using the notation that a bold lower case-letter represents a frequency vector and a bold upper-case letter represents a frequency matrix, we write

$$\boldsymbol{u} = \begin{bmatrix} U_L(z) & U_R(z) \end{bmatrix}^T,$$

$$\boldsymbol{v} = \begin{bmatrix} V_1(z) & V_2(z) \end{bmatrix}^T,$$

$$\boldsymbol{w} = \begin{bmatrix} W_L(z) & W_R(z) \end{bmatrix}^T,$$

and

$$\boldsymbol{H} = \begin{bmatrix} H_{1L}(z) & H_{2L}(z) \\ H_{1R}(z) & H_{2R}(z) \end{bmatrix},$$

where $\boldsymbol{H}$ is named the *acoustic transfer matrix*. Equation (1) can be written in matrix notation as

$$\boldsymbol{w} = \boldsymbol{H}\boldsymbol{v}. \tag{2}$$

The crosstalk path can be canceled out (or, at least, be considerably attenuated) with the use of an adequate filter structure. This filter structure should always be placed between the input binaural signal and the loudspeakers, and can be represented as matrix $\boldsymbol{C}$, the so-called *crosstalk cancellation matrix*, such that

$$\boldsymbol{v} = \boldsymbol{C}\boldsymbol{u}, \tag{3}$$

resulting in the complete transmission path

$$\boldsymbol{w} = \boldsymbol{H}\boldsymbol{C}\boldsymbol{u}, \tag{4}$$

shown in fig. 2 in the form of a block diagram.

The reproduction error is defined as

$$\boldsymbol{d} = \left(\boldsymbol{w} - \boldsymbol{u} \cdot e^{-z\Delta}\right), \tag{5}$$

where Δ is a time delay required to guarantee the resulting filters are causal [7].

For the two loudspeaker setup presented in fig. 1, the reproduction error can be canceled in a straightforward manner by setting

$$\boldsymbol{C} = \boldsymbol{H}^{-1} e^{-z\Delta}, \tag{6}$$

as long as $\boldsymbol{H}$ is invertible.

If, instead of two, N loudspeakers are now used for the binaural reproduction—as discussed in [14]—$\boldsymbol{H}$ expands to

$$\boldsymbol{H} = \begin{bmatrix} H_{1L}(z) & H_{2L}(z) & \cdots & H_{NL}(z) \\ H_{1R}(z) & H_{2R}(z) & \cdots & H_{NR}(z) \end{bmatrix},$$

where $H_{nL}(z)$ and $H_{nR}(z)$ represent the acoustic path from the n^{th} loudspeaker to the left and right ears, respectively. $\boldsymbol{H}$ now describes an underdetermined system—with an infinite set of CTC filter combinations that can drive the energy of $\boldsymbol{d}$ to zero—and is no longer invertible. In this case, besides the minimization of the error energy, the *control effort*, i.e. the energy of the loudspeakers' signals, is also minimized. This extra constraint added to the cost function leads to a single optimal solution to this minimization problem. Such minimization requirements can be cast as a constrained optimization problem using Lagrange multipliers [15]. The cost function

$$J(z) = \boldsymbol{v}^H \boldsymbol{v} - \boldsymbol{\lambda}^H \boldsymbol{d} - \boldsymbol{d}^H \boldsymbol{\lambda} \tag{7}$$

should be minimized, where $\boldsymbol{\lambda}$ is a vector of Lagrange multipliers and $(\cdot)^H$ represents the Hermitian transpose. The solution to this least-norm minimization problem is given by deriving eq. (7) in relation to $\boldsymbol{v}$ and $\boldsymbol{\lambda}$, equating both terms to zero and combining them, which results in

$$\boldsymbol{v} = \boldsymbol{H}^H \left(\boldsymbol{H}\boldsymbol{H}^H\right)^{-1} \boldsymbol{u} \cdot e^{-z\Delta}. \tag{8}$$

According to eq. (3), $\boldsymbol{v}$ is optimal for any choice of $\boldsymbol{u}$ if

$$\boldsymbol{C} = \boldsymbol{H}^H \left(\boldsymbol{H}\boldsymbol{H}^H\right)^{-1} e^{-z\Delta}. \tag{9}$$

The solution to eq. (9) can often be ill-conditioned, resulting in CTC filters with very high gains at certain frequencies, which causes not only a loss of dynamic range, but also generates the so-called "ringing frequencies" [7]. Regularization is often used to limit the energy of the loudspeaker signals and consequently reduce loudspeakers' fatigue as well as nonlinear behavior [16]. It is obtained by relaxing the constraint that the energy of $\boldsymbol{d}$ must be zero while still minimizing the energy of the CTC filters and the reproduction error [17], which can be obtained by minimizing the cost function

$$J(z) = \boldsymbol{d}^H \boldsymbol{d} + \beta(z) \boldsymbol{v}^H \boldsymbol{v}, \tag{10}$$

where $\beta(z)$ is a frequency dependent regularization parameter with real values in the range $0 \leq \beta(z) \leq \infty$. $\beta(z)$ acts as a trade-off factor between the amount of cancellation present in the contralateral channel—and thus the channel separation—described on the left term of eq. (10) and the amount of gain in the CTC filters—and thus the resulting loss of dynamic range [6]—described on the right term of eq. (10).

The optimum filters that satisfy these constraints are given by [16]

$$\boldsymbol{C} = \left(\boldsymbol{H}^H \boldsymbol{H} + \beta(z)\boldsymbol{I}\right)^{-1} \boldsymbol{H}^H e^{-z\Delta}, \tag{11}$$

which can be shown to be equivalent to

$$\boldsymbol{C} = \boldsymbol{H}^H \left(\boldsymbol{H}\boldsymbol{H}^H + \beta(z)\boldsymbol{I}\right)^{-1} e^{-z\Delta}. \tag{12}$$

III. Filter Design

The calculation of CTC filters can be carried out either in the time or in the frequency-domain. The frequency-domain solution given in the previous section can be recast in the time-domain as the concatenation of the convolution matrices of each HRTF in $\boldsymbol{H}$ [18], [19].

If the available head-related impulse responses are L samples long and the desired CTC filters are M samples long, then, for the two loudspeakers configuration, the resulting matrix to be inverted will be a $2(M + L - 1) \times 2M$ matrix. Meanwhile, in the frequency-domain, assuming that the filter length for both $\boldsymbol{H}$ and $\boldsymbol{C}$ is $M + L - 1$, the calculation of the CTC filters will require $\lceil (M + L - 1)/2 \rceil$ times the inversion of a mere 2×2 complex matrix.[1] As the inversion of a $n \times n$ matrix has a computational complexity of approximately $O(n^3)$, inversion in the frequency-domain has the clear advantage that its computational requirements are considerably lower than for computation in the time-domain. This is a major advantage for dynamic CTC systems, which require constant filter update.

A. Causality Constraint

The minimization problem described by eq. (10) will deliver the best possible channel separation for a certain listener-loudspeaker setup and regularization parameter. These filters, however, usually contain anti-causal components, even after the acoustic lag between loudspeakers and the listener has been compensated for, as can be seen in the example in fig. 3. These anti-causal components may wrap around in time—because of the cyclic behavior of the discrete Fourier transform (DFT) used in the calculations—and lead to audible artefacts. The common approach to deal with these artefacts is to apply extra delay and time window the CTC filters. Even though this is a very practical method, it increases the overall latency and can lead to a reduction in channel separation.

A more reliable method to avoid these artefacts would be to calculate the CTC filters in time-domain, thus guaranteeing strictly causal filters. As calculation in time-domain is too computationally expensive, in this manuscript we suggest applying a causality constraint to the frequency-domain calculation. This requires little extra computation effort and guarantees that only the anti-causal components of the filters

[1] Please note that $\boldsymbol{H}$ and $\boldsymbol{C}$ are a three-dimensional tensor, while $\boldsymbol{e}$, $\boldsymbol{v}$, and $\boldsymbol{b}$ are two-dimensional tensors. As in the frequency dimension the addition and multiplication operations can be applied independently for each frequency, the three-dimensional tensors can be considered as matrices and the two-dimensional tensors can be considered as vectors.

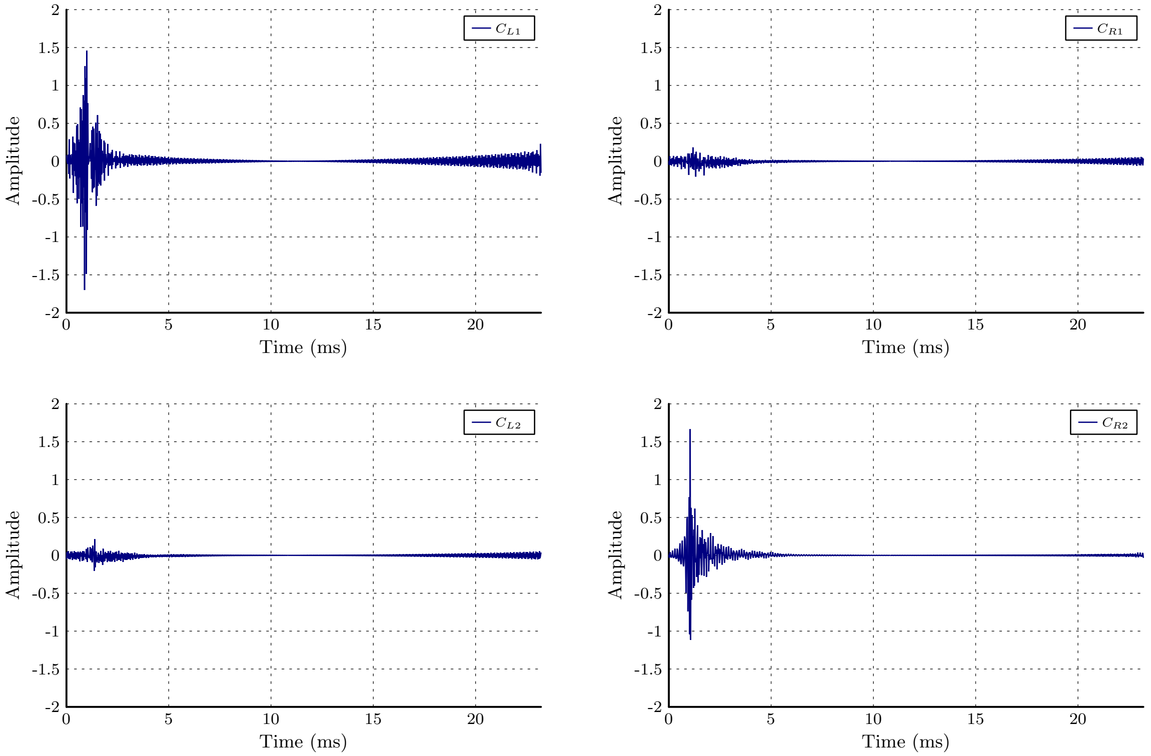

Fig. 3. Time response of $\boldsymbol{C}$ for two loudspeakers placed at $\phi = \pm 45°$ calculated with eq. (11) using $\mu = 0.005$ for all frequencies and $\Delta = 3.4\,\text{ms}$. $\boldsymbol{H}$ was obtained from the HRTF data set described in section IV. Noncausal oscillations are clearly visible in all four filters, even though a time delay larger than the acoustic lag between loudspeakers and the head was used.

will be windowed out, reducing the eventual loss in channel separation.

Using the identity

$$(\cdot)^{-1} = \text{adj}(\cdot)/\det(\cdot), \tag{13}$$

where $\text{adj}(\cdot)$ is the adjugate of a matrix[2] and $\det(\cdot)$ its determinant, we can rewrite equation eq. (12) as

$$\boldsymbol{Y} = \boldsymbol{L} - \boldsymbol{C}D(z) = 0, \tag{14}$$

where $\boldsymbol{L} = \boldsymbol{H}^H \text{adj}(\boldsymbol{H}\boldsymbol{H}^H + \beta(z)\boldsymbol{I})e^{-z\Delta}$ is a matrix and $D(z) = \det(\boldsymbol{H}\boldsymbol{H}^H + \beta(z)\boldsymbol{I})$ a vector.

As there is no matrix multiplication in eq. (14), each element of $\boldsymbol{Y}$ is an independent linear equation and the causal constraint solution based on the spectral factorization discussed in [20, pg. 340] can be independently calculated for each of these elements, resulting in

$$C(z)'_{ij} = \frac{1}{D(z)^+}\left[\frac{L(z)_{ij}}{D(z)^-}\right]_+, \tag{15}$$

where $(\cdot)^+$ and $(\cdot)^-$ are, respectively, the minimum causal stable and minimum anti-causal stable parts resulting from the Wiener-Hopf decomposition and $[\,\cdot\,]_+$ denotes the causal part of an impulse response.

Within the proposed framework we implement the spectral factorization in the cepstral domain. As $D(z)$ is the determinant of a Hermitian matrix, it is a real valued spectrum and consequently possesses a symmetric cepstrum. The spectral factorization is calculated allocating the first half of the cepstrum to the causal stable part and its second half to the anti-causal stable part. Additionally, we estimate the causal part by simply multiplying the impulse response of the term inside brackets with an asymmetric time window that sets the second half of the impulse response to zero. It is important to notice that the same window must be applied to each element of $\boldsymbol{Y}$ to avoid any interaural phase difference to be created at this step of the calculation procedure.

Equation (15) can be rewritten in matrix form as

$$\begin{aligned}\boldsymbol{C}' = &\frac{1}{\det\left(\boldsymbol{H}\boldsymbol{H}^H + \beta(z)\boldsymbol{I}\right)^+} \times \\ &\times \left[\frac{\boldsymbol{H}^H \,\text{adj}\left(\boldsymbol{H}\boldsymbol{H}^H + \beta(z)\boldsymbol{I}\right) e^{-z\Delta}}{\det\left(\boldsymbol{H}\boldsymbol{H}^H + \beta(z)\boldsymbol{I}\right)^-}\right]_+ .\end{aligned} \tag{16}$$

B. Minimum-Phase Regularization

Regularization introduces a pre-ringing in the resulting ear signals [10], [12]. This pre-ringing can result in audible artefacts if the filters are heavily regularized at certain frequencies. Since the human auditory system has a much longer post-masking behavior than pre-masking [21], for some applications, like binaural reproduction, it might be desirable to alter the regularization procedure so that pre-ringing is converted into post-ringing in the resulting ear signal.

[2] Note that for the special case of a 2×2 matrix the adjugate can be obtained without further calculation.

In [11] a technique was presented to control the pre-ringing of a single channel regularized inverse filter. In the single channel case, the regularized inverse $D(z)$ of a system described by $B(z)$ is calculated from

$$D(z) = \left[B(z)^H B(z) + \beta(z)\right]^{-1} B(z)^H. \tag{17}$$

We define the regularization shape-factor $A(z)$ and rewrite eq. (17) as

$$D(z) = A(z)B(z)^{-1}. \tag{18}$$

Comparing eqs. (17) and (18) we obtain that the regularization shape-factor is given by

$$A(z) = \frac{1}{1 + \beta(z)/\left|B(z)\right|^2}, \tag{19}$$

where $|\cdot|$ is the absolute value operator.

It is easy to verify that $A(z)$ is a real vector if $\beta(z)$ is also a real vector, so $A(z)$ will exhibit a symmetric and noncausal associated impulse response. To avoid pre-ringing caused by filtering the inverse of $B(z)$ with $A(z)$, Norcross [12] suggests substituting $A(z)$ by its minimum-phase equivalent $A(z)_{\mathrm{mp}}$, thus guarantying frequency regularization without any pre-ringing artefacts caused by the regularization itself.

For the multichannel case, the method presented in [12] has the drawback that the minimum-phase equivalent regularization shape-factor has to be calculated for each channel individually. This approach can introduce interaural phase differences that can compromise the quality of the reproduced binaural scene. Therefore, we propose a new method that approximates a global minimum-phase regularization, applying the same minimum-phase regularization shape-vector to all channels, thus avoiding any such interaural phase discrepancies.

Applying eq. (13) to eq. (12) gives us

$$\boldsymbol{C} = \frac{\boldsymbol{H}^H \operatorname{adj}\left(\boldsymbol{H}\boldsymbol{H}^H + \beta(z)\boldsymbol{I}\right)}{\det\left(\boldsymbol{H}\boldsymbol{H}^H + \beta(z)\boldsymbol{I}\right)} e^{-z\Delta}. \tag{20}$$

We now assume we can describe the effect of regularization in all channels with a single regularization shape-factor $A(z)$, so that

$$\boldsymbol{C} = \left(\frac{A(z)}{\det(\boldsymbol{H}\boldsymbol{H}^H)}\right) \boldsymbol{H}^H \operatorname{adj}\left(\boldsymbol{H}\boldsymbol{H}^H\right) e^{-z\Delta}. \tag{21}$$

The first term of eq. (21) emphasizes the effect of the regularization shape-factor $A(z)$ in limiting large values that may be produced by the inversion of $\det(\boldsymbol{H}\boldsymbol{H}^H)$.

We also assume that for small values of $\beta(z)$

$$\operatorname{adj}\left(\boldsymbol{H}\boldsymbol{H}^H + \beta(z)\boldsymbol{I}\right) \approx \operatorname{adj}\left(\boldsymbol{H}\boldsymbol{H}^H\right) \tag{22}$$

as no element inversion is conducted in the calculation of the adjugate. Under this assumption, equating eqs. (20) and (21) results in

$$A(z) = \frac{\det\left(\boldsymbol{H}\boldsymbol{H}^H\right)}{\det\left(\boldsymbol{H}\boldsymbol{H}^H + \beta(z)\boldsymbol{I}\right)}. \tag{23}$$

Again, as the determinant of a Hermitian matrix is real, the numerator and the denominator of eq. (23) are also real and, consequently, $A(z)$ is also real, thus exhibiting a symmetric and noncausal associated impulse response responsible for the pre-ringing present in filters designed with regularization.

We can now substitute $A(z)$ in eq. (21) by its minimum-phase equivalent $A(z)_{mp}$ to obtain a minimum-phase regularized CTC filter, which presents approximately the same amplitude response as eq. (11) but with all noncausal ringing effects produced by regularization in the ear signal converted in its causal equivalent. It is also possible to combine the zero-phase with the minimum-phase equivalents of $A(z)$, as described in [12].

According to section III-A, to obtain a causal CTC filter we should apply the Wiener-Hopf decomposition to the fraction term in eq. (21) and estimate the causal part of the non-minimum causal stable components of the signal, resulting in

$$\begin{aligned}\boldsymbol{C}' = \left(\frac{A(z)^+}{\det\left(\boldsymbol{H}\boldsymbol{H}^H\right)^+}\right) \Bigg[\left(\frac{A(z)^-}{\det\left(\boldsymbol{H}\boldsymbol{H}^H\right)^-}\right) \times \\ \times \boldsymbol{H}^H \operatorname{adj}\left(\boldsymbol{H}\boldsymbol{H}^H\right) e^{-z\Delta}\Bigg]_+ .\end{aligned} \tag{24}$$

When the minimum-phase equivalent of $A(z)$ is used we have $A(z)_{mp}^+ \cdot A(z)_{mp}^- = A(z)_{mp} \cdot 1(z)$ and in this case the regularization shape-factor is applied only to the minimum causal stable component of $\det(\boldsymbol{H}\boldsymbol{H}^H)$. The lack of regularization applied to the anti-causal component may result, after its inversion, in large spectral values and, consequently, in a ringing behavior in the time-domain that can compromise the estimation of its causal part. This issue can be solved using the decomposition $A(z)_{mp} = \sqrt{A(z)}_{mp} \cdot \sqrt{A(z)}_{mp}$ instead, where $\sqrt{A(z)}_{mp}$ is the minimum-phase version of the square-root of $A(z)$. The causal CTC filters, which will provide an ear signal that is causal and free of pre-ringing, are then given by

$$\begin{aligned}\boldsymbol{C}'_{mp} = \left(\frac{\sqrt{A(z)}_{mp}}{\det\left(\boldsymbol{H}\boldsymbol{H}^H\right)^+}\right) \Bigg[\left(\frac{\sqrt{A(z)}_{mp}}{\det\left(\boldsymbol{H}\boldsymbol{H}^H\right)^-}\right) \times \\ \times \boldsymbol{H}^H \operatorname{adj}\left(\boldsymbol{H}\boldsymbol{H}^H\right) e^{-z\Delta}\Bigg]_+ .\end{aligned} \tag{25}$$

The CTC system performance should not be compromised by the minimum-phase regularization, as long as $\beta(z)$ is sufficiently small for the approximation in eq. (22) to be valid. Only a phase variation, identical to both binaural channels, will be present.

C. Weighting

When designing a CTC reproduction system for a dynamic system, two loudspeakers will not be sufficient to allow the listener to rotate his/her head freely. If the listener's head points in a direction outside of the arc spanned by both loudspeakers, the CTC system will become unstable [2]. To meet the requirements of an immersive VR environment, Lentz designed a system with four loudspeakers [2]. However, as he employed the truncated CTC filter calculation algorithm [22], only two loudspeakers could be used to reproduce the binaural signals. Thus, the active pair of loudspeakers had to be exchanged according to the orientation of the listener's head. The switching between each pair of active loudspeakers was made by a soft fading between the filters.

Many CTC filter design strategies can handle only two loudspeaker setups, e.g. generic CTC [23] or the iterative CTC approach [2]. In these cases, cross-fading is used to switch between active loudspeaker pairs, i.e. the CTC filters are calculated to two different pairs of loudspeakers and their active region are superposed in a small angular region where both pairs are active. The switching, however, might lead to audible artefacts. For example if the two sets of filters are not perfectly aligned in time, their superposition can lead to a comb filter artefact. Furthermore, compromises are made while calculating each pair of filters independently. Combining the two sets of non-ideal filters can lead to variations in the resulting frequency response, i.e. to coloration of the reproduced signal, and to reduction in channel separation.

The proposed framework allow all loudspeakers to be used simultaneously. However, measurements show that "two-channel configurations result in wider controlled area and are more robust to head rotation and frontal displacement than the four-channel configurations" [24], suggesting that fewer loudspeakers will provide a more robust CTC system. Thus, it is reasonable to reduce the number of active loudspeakers to two or three, depending on the geometry,[3] but with an improved filter fading strategy.

We thus propose the use of a weighted matrix inversion to control the filter switch. A different set of weights can be applied to each loudspeaker according to the direction in which the listener's head is pointing, allowing for a seamless filter update.

The *weighted* ℓ_2 norm is given by

$$\|\boldsymbol{x}\|_{\boldsymbol{W}}^2 = \boldsymbol{x}^H \boldsymbol{W} \boldsymbol{x}, \tag{26}$$

where $\boldsymbol{W}$ is a diagonal weighting matrix containing positive weights for each element of $\boldsymbol{x}$.

The cost function presented in eq. (7) can be reformulate as

$$J(z) = \boldsymbol{v}^H \boldsymbol{W} \boldsymbol{v} - \boldsymbol{\lambda}^H \boldsymbol{d} - \boldsymbol{d}^H \boldsymbol{\lambda}. \tag{27}$$

The optimum set of filters that minimizes eq. (27) for any choice of $\boldsymbol{u}$ is

$$\boldsymbol{C} = \boldsymbol{W}^{-1} \boldsymbol{H}^H \left(\boldsymbol{H} \boldsymbol{W}^{-1} \boldsymbol{H}^H \right)^{-1} e^{-z\Delta}. \tag{28}$$

The larger the weight w_{ii} applied to loudspeaker i compared to the other weights, the higher the effort made by the algorithm to minimize the energy of this loudspeaker's output and thus the smallest the energy of the filters related to this loudspeaker. This filter calculation method allow a flexible choice of the number and position of the active loudspeakers used in the dynamic CTC system, thus yielding fading between pairs of loudspeakers, as described in [2], obsolete.

As discussed in section II, also the weighted inversion problem can be ill-conditioned and regularization is applied to avoid problems with singularity. The new cost function to be minimized is now

$$J(z) = \boldsymbol{v}^H \boldsymbol{W} \boldsymbol{v} - \boldsymbol{\lambda}^H \boldsymbol{d} - \boldsymbol{d}^H \boldsymbol{\lambda} - \beta(z) \boldsymbol{\lambda}^H \boldsymbol{\lambda} \tag{29}$$

and the optimum set of filters that minimizes eq. (29) (see Appendix) for any choice of $\boldsymbol{u}$ is given by

$$\boldsymbol{C} = \boldsymbol{W}^{-1} \boldsymbol{H}^H \left(\boldsymbol{H} \boldsymbol{W}^{-1} \boldsymbol{H}^H + \beta(z) \boldsymbol{I} \right)^{-1} e^{-z\Delta}. \tag{30}$$

To make the choice of the weights more straightforward, we substituted $\boldsymbol{W}$ by $\boldsymbol{Z} = \boldsymbol{W}^{-1}$. In this way, the smaller the weight z_{ii} applied at a loudspeaker, the lower the sound pressure that this loudspeaker is supposed to generate, up to the point that if the weight $z_{ii} = 0$ is applied, the respective loudspeaker will be left completely inactive.

Applying the causality constraint and causal regularization to eq. (30) yields

$$\begin{aligned} \boldsymbol{C}'_{mp} = & \frac{\sqrt{A(z)}_{mp}}{\det\left(\boldsymbol{H}\boldsymbol{Z}\boldsymbol{H}^H\right)^+} \left[\left(\frac{\sqrt{A(z)}_{mp}}{\det\left(\boldsymbol{H}\boldsymbol{Z}\boldsymbol{H}^H\right)^-} \right) \times \right. \\ & \left. \times \boldsymbol{Z}\boldsymbol{H}^H \operatorname{adj}\left(\boldsymbol{H}\boldsymbol{Z}\boldsymbol{H}^H\right) e^{-z\Delta} \right]_+ , \end{aligned} \tag{31}$$

where $A(z) = \det(\boldsymbol{H}\boldsymbol{Z}\boldsymbol{H}^H) / \det(\boldsymbol{H}\boldsymbol{Z}\boldsymbol{H}^H + \beta(z)\boldsymbol{I})$.

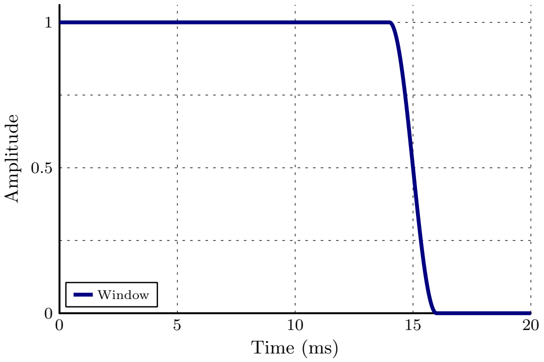

Fig. 4. Asymmetric Tukey window starting at 14 ms and ending at 16 ms.

IV. Results

We use two numerical examples to analyze the performance of the presented framework. All filter are calculated from a transfer matrix $\boldsymbol{H}$ constructed with HRTFs measured with an asymmetric dummy head. To verify the frequency response of the complete transfer-path between the binaural signals and the ear signals an individualized but mismatched set of HRTFs was used for the calculation [26]. In the examples where regularization was applied, we used $\beta = 0.005$ for all frequencies. For all examples a time delay of $\Delta = 3.4$ ms was used.

First we evaluate the causality constraint and minimum-phase regularization aspects of the proposed framework. As previously discussed, the common approach to deal with pre-ringing artefacts is to apply an extra time delay and to time window the CTC filters. We used this method as a reference to compare the proposed framework with.

[3] Simulation results suggest that for certain geometries the use of three loudspeakers will increase the robustness of the system [25].

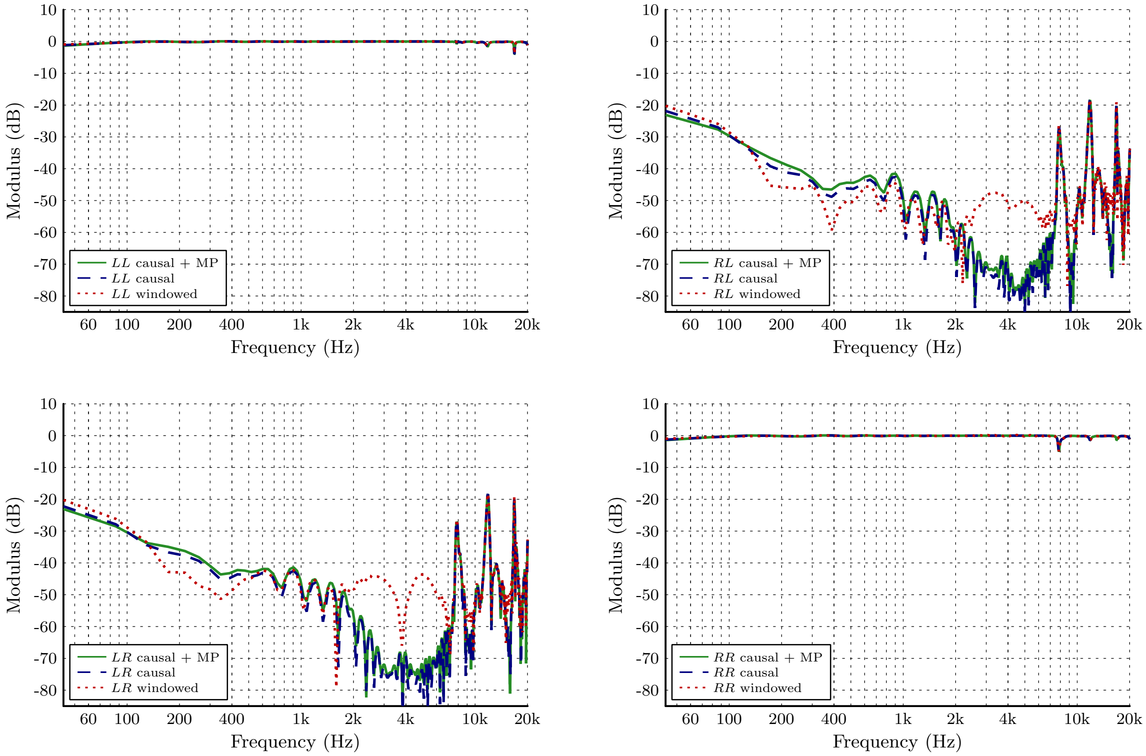

Fig. 5. Frequency response of the complete transfer-path between the binaural signals and the ear signals for three different CTC filters: *windowed* is calculated with eq. (12) and the result is time windowed,*causal* is calculated with the causal constraint contained in eq. (16);and *causal + MP* is calculated with the causal constraint and the minimum-phase regularization contained in eq. (25).

The *reference* filters are obtained using eq. (12), time shifting the result by Δ and applying an asymmetric Tukey window starting at 14 ms and ending at 16 ms as depicted in fig. 4. A filter with *causal* constraints is calculated using eq. (16)—the causal estimation is done by applying the same asymmetric Tukey window from the previous method. The effect of the *minimum phase* regularization is analyzed from a set of CTC filters calculated using eq. (25) with the same regularization, time delay and time window as the previous method.

A CTC system composed of two loudspeakers placed 2 m away from the center of the listener's head and $\pm 45^\circ$ to its line of sight is used in this example. The effect of causality constraint and minimum phase regularization in the frequency response can be observed in fig. 5, where the plots in the main diagonal should ideally have value 0 dB and the ones in the off-diagonal should ideally have value $-\infty$ dB. As regularization was used in all three cases, the example shows small deviations in the diagonal elements and a considerable reduction in cancellation, observable in the off-diagonal plots. Figure 5 exemplify how the use of the new framework provide filters with similar behavior to the reference filters and can even provide better channel separation at certain frequencies as only the noncausal components of the filters are time windowed. It also shows how the use of minimum-phase regularization will not alter the CTC system's magnitude response.

The major claims of the new framework is that the obtained CTC filters are causal and that the system response will not exhibit any pre-ringing. Figure 6 shows the CTC filters obtained with eq. (25) in time domain. This example shows how all resulting filters are strictly causal.

Even though the CTC filters are causal, the resulting ear signals may contain pre-ringing artefacts stemming from the regularization process. Figure 7 shows the resulting ear signals for the filters shown in fig. 6 and exemplifies how the minimum-phase regularization can control the pre-ringing artefacts. The effect of minimum-phase regularization can be observed in the impulse responses of the diagonal elements, as the impulse responses have a sharp onset—note that the oscillations prior to the impulse response are caused by the use of individualized but mismatched HRTFs.

We now evaluate the third aspect of the presented framework: the weighted filter design. Four loudspeakers are placed on a circumference of radius 2 m at angular positions $45^\circ, 135^\circ, 225^\circ$ and 315°. The same set of HRTFs described in the previous example is again used.

The filters obtained with eq. (31), using $z_{ii} = 1$ for the three active loudspeakers and $z_{ii} = 0$ for the loudspeaker behind the listener, are compared to the filters obtained with the system described in [2]. The listener's head is oriented towards 22°, in a position where the system described in [2] is fading between two active loudspeaker pairs. The resulting ear signals obtained with both methods for this example are presented in fig. 8.

In this example the fading strategy leads to an uneven frequency response of the direct path. It is interesting to observe that the interaction of the two active pairs of loudspeakers leads to a low-pass behavior of the frequency response when

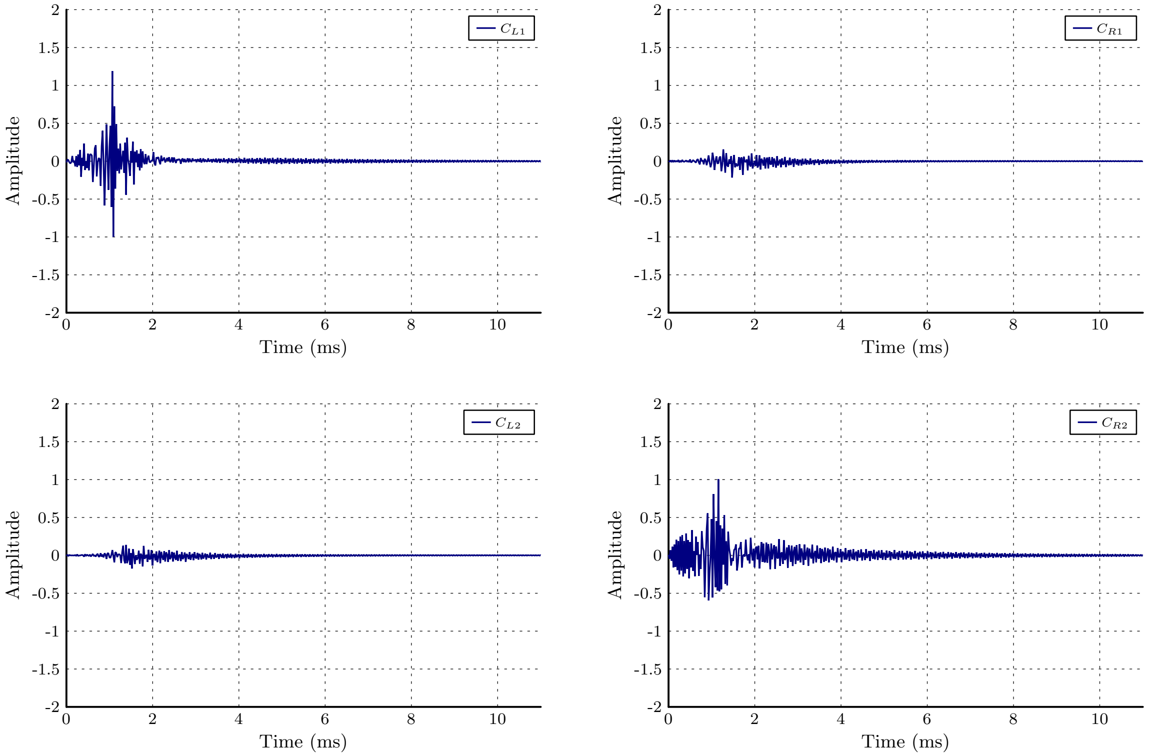

Fig. 6. Time response of $\boldsymbol{C}$ for two loudspeakers placed at $\phi = \pm 45°$ calculated with eq. (25). Note how all resulting filters are strictly causal.

compared to the frequency response of each pair alone. The weighted inversion method provides a flat frequency response for the direct path with improved channel separation. In this particular example, the obtained frequency response is even flatter than the one obtained with the two loudspeaker setup presented from the previous example, as now three instead of two loudspeakers are active.

Note that the absolute value of the weights is not relevant when no regularization is used, as the presence of the weight matrix both inside and outside the parenthesis in eq. (30) will result in a normalization of the weights. The relative value between the weights, however, define how much energy the signal from each loudspeaker will have.

If regularization is used, the weight matrix will change the amplitude of the element of $\boldsymbol{HZH}^H$ in relation to $\beta(z)\boldsymbol{I}$. Here two situations arise. Either the weights increase the magnitude of the element of $\boldsymbol{HZH}^H$, which is equivalent to decreasing the amplitude of β and equivalently to reducing the effect of regularization while increasing pre-ringing and the amplitude of the filters; or the weights decrease the magnitude of the element of $\boldsymbol{HZH}^H$, which is equivalent to increasing the amplitude of β and equivalently to strengthening the effect of regularization, which provoke a decrease in the system's performance. If, for example, we use a rectangular window with value 1 over half of the circumference or a Hann window over the full circumference for the weights, we can observe very similar results as the active loudspeakers will receive weights close or equal to 1 and regularization will be applied evenly to all channels. Other more narrow windows, e.g. a Tukey window with 75% taper ratio over half of the circumference, will already show reduction in channel separation performance as some weights applied to the active loudspeakers can often have values close to the values of β.

V. Discussion

This paper presents a general framework for the calculation of dynamic crosstalk cancellation (CTC) filters to be applied to binaural reproduction in immersive virtual reality environments using a dynamic CTC setup with multiple loudspeakers. Such setups require high filter update rates, suggesting that filter calculations should be performed in the frequency-domain for higher efficiency.

Since a direct calculation in frequency-domain might yield noncausal artefacts, a causality constraint in the frequency-domain calculation is introduced by applying the spectral factorization method, which avoids undesirable wrap-around effects and echo artefacts. Regularization is commonly applied to the CTC filter calculation in order to limit the output levels at the loudspeakers, increasing the dynamic range at the same time that it decreases the channel separation. This leads to a more compact CTC filter, with the side effect that noncausal artefacts appears in the resulting ear signal. These artefacts can be controlled using the proposed minimum-phase regularization. Even though extra calculation steps are added, it was verified that the calculation time required by this framework is one order of magnitude faster than an equivalent calculation in time-domain for a two loudspeakers setup having CTC filters with 512 taps. Moreover, the advantage of frequency calculation tends to increase for larger filters.

Another aspect that is especially critical for dynamic CTC systems is the switch between active loudspeakers in the

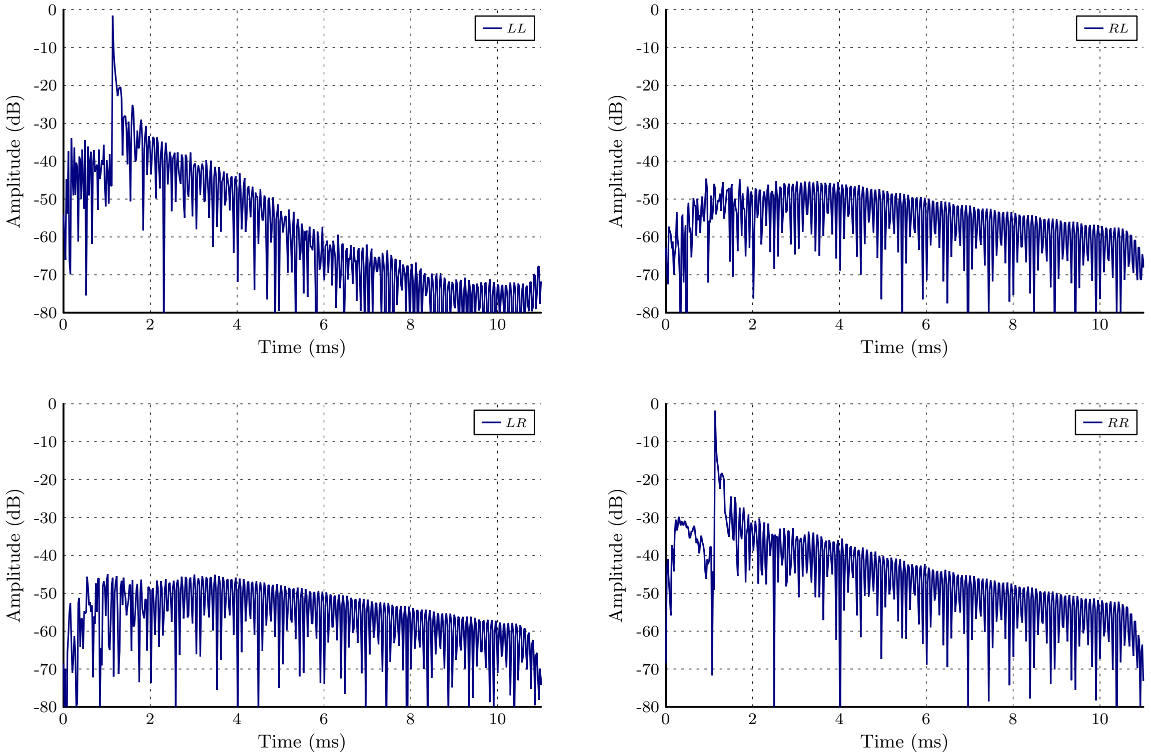

Fig. 7. Time response of the complete transfer-path between the binaural signals and the ear signals for the filters shown in fig. 6. The sharp onset observed in the diagonal elements is caused by the minimum-phase regularization.

setup. The use of a weighted filter calculation allows the loudspeakers' contribution to be windowed in space, resulting in a smooth filter transition with improved channel separation and frequency response. Weights could be made frequency-dependent, allowing for a frequency-dependent choice of active loudspeakers cf. [6]. Nevertheless, optimal weights distribution remains to be investigated.

All filter calculation described so far assumed *a priori* knowledge of the transmission matrix to be equalized by the CTC system. However, realistic CTC systems will not deliver a channel separation that is as high as the one obtained using an ideal CTC system [26]. Especially at high frequencies, the obtained channel separation is often lower than the channel separation naturally resulting from the head shadow. Gardner already verified this deficiency of mismatched CTC systems and suggested that CTC should be used only at low and middle frequencies and that the binaural signal should be played directly via two loudspeakers at high frequencies [27, pp. 65–77]. He achieved this by by-passing the CTC filters at high frequencies, only equalizing the direct path between loudspeaker and ipsilateral ear at these frequencies. The presented framework could be expanded to include a panning function for high frequencies, allowing the binaural signal to be smoothly panned between the active loudspeakers.

Appendix

The cost function eq. (29) is minimized by deriving $J(z)$ in regards to $\boldsymbol{v}$ and $\boldsymbol{\lambda}$, yielding

$$\partial J/\partial \boldsymbol{v} = 2\boldsymbol{W}\boldsymbol{v} - 2\boldsymbol{H}^H\boldsymbol{\lambda} \tag{32}$$

and

$$\partial J/\partial \boldsymbol{\lambda} = -2\left(\boldsymbol{H}\boldsymbol{v} - \boldsymbol{u}e^{-z\Delta}\right) - 2\beta(z)\boldsymbol{\lambda}, \tag{33}$$

and setting both functions to zero. Assuming that no weight element w_{ii} is equal to zero, than the diagonal weight matrix $\boldsymbol{W}$ is invertible and we have

$$\boldsymbol{v} = \boldsymbol{W}^{-1}\boldsymbol{H}^H\boldsymbol{\lambda}, \tag{34}$$

which we substitute in eq. (33) to obtain

$$\left(\boldsymbol{H}\boldsymbol{W}^{-1}\boldsymbol{H}^H + \beta(z)\boldsymbol{I}\right)\boldsymbol{\lambda} = \boldsymbol{u}e^{-z\Delta}. \tag{35}$$

For $\beta(z) > 0$ the term inside the parentheses in eq. (35) is a strictly positive-definite matrix and, thus, invertible. To conclude, we substitute eq. (35) back into eq. (34), obtaining

$$\boldsymbol{v} = \boldsymbol{W}^{-1}\boldsymbol{H}^H\left(\boldsymbol{H}\boldsymbol{W}^{-1}\boldsymbol{H}^H + \beta(z)\boldsymbol{I}\right)^{-1}\boldsymbol{u}e^{-z\Delta}. \tag{36}$$

Acknowledgment

The authors would like to thank Prof. Dr. V. Nascimento and Dr. F. Fazi for the fruitful discussion and suggestions and Prof. Dr. Craig Jin for providing the work environment for the revision of this manuscript. The algorithms described in this paper are available within the ITA-Toolbox at www.ita-toolbox.org.

References

[1] T. Lentz, "Binaural technology for virtual reality," Ph.D. dissertation, Institut für Technische Akustik, RWTH-Aachen, 2007.

[2] ——, "Dynamic crosstalk cancellation for binaural synthesis in virtual reality environments," *J. Audio Eng. Soc.*, vol. 54, no. 4, pp. 283–294, 2006.

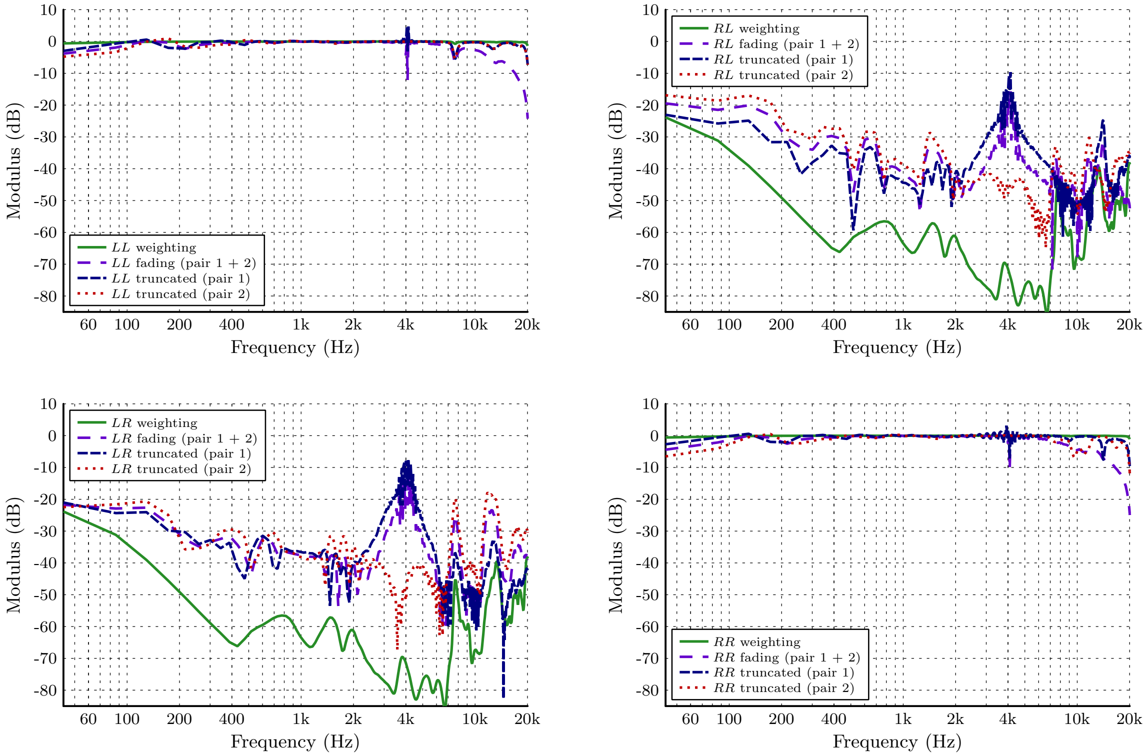

Fig. 8. Frequency response of the complete transfer-path between the binaural signals and the ear signals. *weighting* is calculated using the proposed weighting strategy. *fading* is calculated using the cross-fading strategy described in [2] with the truncated CTC filter calculation algorithm. Both *truncated* curves are the transfer functions obtained with the truncated CTC filter calculation algorithm for each active pair of loudspeakers independently.

[3] O. Kirkeby, P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, “Local sound field reproduction using digital signal processing,” *The Journal of the Acoustical Society of America*, vol. 100, no. 3, pp. 1584–1593, 1996.

[4] M. Brennan and S.-M. Kim, “Feedforward and Feedback Control of Sound and Vibration – a Wiener Filter Approach,” *Journal of Sound and Vibration*, vol. 246, no. 2, pp. 281–296, 2001.

[5] S.-M. Kim and S. Wang, “A Wiener filter approach to the binaural reproduction of stereo sound,” *J. Acoust. Soc. Am.*, vol. 114, no. 6, pp. 3179–3188, 2003.

[6] T. Takeuchi and P. A. Nelson, “Subjective and objective evaluation of the optimal source distribution for virtual acoustic imaging,” *J. Audio Eng. Soc.*, vol. 55, no. 11, pp. 981–997, 2007.

[7] P. A. Nelson and J. F. W. Rose, “The time domain response of some systems for sound reproduction,” *Journal of Sound and Vibration*, vol. 296, no. 3, pp. 461–493, 2006.

[8] O. Kirkeby, P. A. Nelson, and H. Hamada, “The ’stereo dipole’ a virtual source imaging system using two closely spaced loudspeakers,” *J. Audio Eng. Soc.*, vol. 46, no. 5, pp. 387–395, 1998.

[9] O. Kirkeby, P. Rubak, and A. Farina, “Analysis of ill-conditioning of multi-channel deconvolution problems,” in *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA’99 (Cat. No.99TH8452)*. IEEE, 1999, pp. 155–158.

[10] L. D. Fielder, “Analysis of traditional and reverberation-reducing methods of room equalization,” *J. Audio Eng. Soc.*, vol. 51, no. 1/2, pp. 3–26, 2003.

[11] M. Bouchard, S. G. Norcross, and G. Soulodre, “Inverse Filtering Design Using a Minimal-Phase Target Function from Regularization,” in *121st AES Convention*, San Francisco, USA, 2006.

[12] S. G. Norcross and M. Bouchard, “Multichannel Inverse Filtering with Minimal-Phase Regularization,” in *123rd AES Convention*, 2007, pp. 1–8.

[13] W. A. Yost, “Discriminations of interaural phase differences,” *The Journal of the Acoustical Society of America*, vol. 55, no. 6, pp. 1299–1303, 1974.

[14] J. L. Bauck and D. H. Cooper, “Generalized Transaural Stereo,” in *93rd AES Convention*, San Francisco, USA, 1992.

[15] P. A. Nelson and S. J. Elliott, *Active Control of Sound*, 3rd ed. San Diego, CA: Academic Press, 1995.

[16] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-bustamante, “Fast deconvolution of multichannel systems using regularization,” *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 189–194, 1998.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization Theory*. New York: Cambridge University Press, 2004.

[18] 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.

[19] Y. L. Parodi, “A Systematic Study of Binaural Reproduction Systems Through Loudspeakers: A Multiple Stereo-Dipole Approach,” Ph.D., Aalborg University, 2010.

[20] A. Papoulis, *Signal Analysis*. McGraw-Hill, 1977.

[21] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models*. Springer, 2007.

[22] J. Köring and A. Schmitz, “Simplifying Cancellation of Cross-Talk for Playback of Head-Related Recordings in a Two-Speaker System,” *Acustica*, vol. 79, pp. 221–232, 1993.

[23] Y. Lacouture Parodi and P. Rubak, “Analysis of Design Parameters for Crosstalk Cancellation Filters Applied to Different Loudspeaker Configurations,” *J. Audio Eng. Soc.*, vol. 59, no. 5, pp. 304–320, 2011.

[24] Y. L. Parodi and P. Rubak, “Objective evaluation of the sweet spot size in spatial sound reproduction using elevated loudspeakers.” *J. Acoust. Soc. Am.*, vol. 128, no. 3, pp. 1045–55, 2010.

[25] J. Yang, W. Gan, and S.-E. Tan, “Improved sound separation using three loudspeakers,” *Acoustic Research Letters*, vol. 4, no. April, pp. 47–52, 2003.

[26] P. Majdak, B. Masiero, and J. Fels, “Sound localization in individualized and non-individualized crosstalk cancellation systems,” *J. Acoust. Soc. Am.*, vol. 133, no. 4, pp. 2055–2068, 2013.

[27] W. G. Gardner, “3-D audio using loudspeakers,” Ph.D., Massachusetts Institute of Technology, 1997.

Bruno Masiero (M'10) was born in São Paulo, Brazil. He received the B.S. and M.S. degrees in electrical engineering from the University of São Paulo, Brazil, in 2005 and 2007, respectively. In 2012 he received his doctoral degree from the RWTH Aachen University, Germany.

From 2007 to 2012, he was a Research Assistant with the Institute of Technical Acoustics, RWTH Aachen University, Germany. His research interests includes acoustic signal processing, spatial sound reproduction, and psychoacoustics.

Michael Vorländer graduated in physics in 1984, gained a doctor degree in 1989 at RWTH Aachen University, Germany, and a habilitation degree at Technical University Dresden, Germany, in 1995. He is now Professor at RWTH Aachen University, Germany, and the Director of the Institute of Technical Acoustics.

His book "Auralization" (Berlin, Germany: Springer 2008) is a reference on the field of Acoustic Virtual Reality. His current research interest is auralization including simulation techniques and signal processing.

Prof. Vorländer was president of the European Acoustics Association, EAA, in the term 2004–2006 and president of the International Commission for Acoustics, ICA, in the term 2010–2013. He was awarded with the RWB Stephens medal (IoA United Kingdom) in 2005 and Fellow of the Acoustical Society of America in 2006.