# GENERALIZED WAVE-DOMAIN TRANSFORMS FOR LISTENING ROOM EQUALIZATION WITH AZIMUTHALLY IRREGULARLY SPACED LOUDSPEAKER ARRAYS

Christian Hofmann and Walter Kellermann

Friedrich-Alexander University Erlangen-Nürnberg (FAU), Multimedia Communications and Signal Processing, Cauerstr. 7, D-91058 Erlangen, Germany {christian.hofmann,walter.kellermann}@FAU.de

#### ABSTRACT

In reverberant environments, Listening Room Equalization (LRE) by pre-filtering of loudspeaker signals is highly desirable for premium sound reproduction systems with a high number of loudspeakers. In this contribution, the efficient concept of LRE by wave-domain adaptive filtering is extended by deriving generalized loudspeaker signal transforms for loudspeaker arrays with an irregular azimuthal spacing. In particular, a matrix approximation allows to restore the unitarity property of a transform matrix which lost the unitarity when applied to irregular arrays instead of regular ones. Simulations of adaptive LRE systems with irregular loudspeaker arrays confirm the efficacy of the proposed, novel, generalized transforms.

*Index Terms*— WDAF, Wave Domain, Azimuthally Irregular Arrays, Spatial Transforms, Equalization

## 1. INTRODUCTION

Premium sound reproduction systems employ a high number of loudspeakers in order to reproduce a virtual acoustic scene in a spatially extended listening area [1-11], or to produce multiple acoustically bright and dark zones, allowing to convey different acoustic scenes to multiple listeners in the same physical environment [12-15]. The reproduction performance of such systems severely degrades in typical listening environments with reflective walls and objects, as each reflection of the played-back signals can re-enter listening and quiet zones in an unpredictable way. As countermeasure, LRE by pre-filtering the loudspeaker signals is highly desirable. To this end, the time-varying acoustical properties of the actual listening environment are typically inferred employing adaptive filters, which identify Impulse Responses (IRs) between the loudspeakers and microphones at control points within or around the listening zones. Afterwards, a pre-equalizer can be determined from the estimated Room Impulse Responses (RIRs) of the Loudspeaker-Enclosure-Microphone System (LEMS). This is depicted schematically in Fig. 1 for the typical case of a combination with an Acoustic Echo Cancellation (AEC) unit to allow for a voice control of an entertainment system or for hands-free communication. As the system identification for LRE is structurally identical to the one of AEC, all challenges of multichannel AEC apply to LRE as well: the high computational complexity to identify the  $N_{\rm L} \cdot N_{\rm M}$ acoustic paths between the  $N_{\rm L}$  loudspeakers and each of the  $N_{\rm M}$ microphones, as well as the so-called non-uniqueness problem [16], which means that the typically highly correlated loudspeaker signals preclude a unique identification of the RIRs of the LEMS

Alleviating both problems, the concept of Wave-Domain Adaptive Filtering (WDAF) has been proposed for AEC and LRE by Buchner et al. [17] and Spors et al. [18] and has been refined and extended since then [19–24]. WDAF allows to exploit prior knowledge about typical couplings between spatially transformed loudspeaker



**Fig. 1**: Application scenario of LRE, where the reflections (green) of the loudspeaker signals (blue) are to be canceled in the listening area (orange) by exploiting reference microphones (red), which can also be employed for voice-control of the entertainment system or hands-free communication when being combined with an AEC unit.

and microphone signals to alleviate the non-uniqueness problem and to reduce the computational complexity by just considering a subset of couplings between the inputs and outputs of the involved transform-domain Multiple-Input/Multiple-Output (MIMO) systems [22].

Other approaches to maintain the rendering performance in reverberant environments require a dedicated calibration phase where the IRs from the actual acoustical environment are measured to allow for a dedicated design of the rendering system for the actual environment [5–7] and thus cannot follow the natural time-variance of the environment. Further approaches aim at reshaping RIRs in order to dereverberate only a perceptually relevant part of a known single-channel RIR [25]. Others aim at crosstalk cancellation for multiple but isolated points in space [26–28], some also in conjunction with the aforementioned impulse response reshaping [29]. Recently, an approach aiming at a correct synthesis of binaural cues while considering the desired acoustics of the recording room and the playback room jointly has been proposed in [30]—also an approach considering only isolated points in the listening zone.

As opposed to this, WDAF LRE is able to adaptively equalize entire listening zones of a Wave Field Synthesis (WFS) or Ambisonics (recently adopted for MPEG-H standard [31, 32]) systems. Therefore, this contribution extends the highly efficient concept of WDAF LRE [21–23] to loudspeaker arrays with irregular azimuthal spacing. To this end, the state-of-the-art wave-domain transforms are revisited in Sec. 2. Then, Sec. 3 explains WDAF LRE, investigates the impact of irregular azimuthal loudspeaker spacing, and introduces the novel, generalized transforms. These transforms are evaluated in Sec. 4, before Sec. 5 concludes the presented work.

# 2. STATE-OF-THE-ART WAVE-DOMAIN TRANSFORMS

This section summarizes the wave-domain transforms as key component for WDAF. These are linear transforms and perform a scalar product between a sound field (actually its spatially and temporally

The authors would like to thank the Fraunhofer Institute for Digital Media Technology (IDMT) in Ilmenau, Germany, for supporting this work.

sampled version) and a set of spatial basis functions which result from elementary solutions to the acoustic wave equation. Thus, the transform domain is termed 'wave domain' and the transformdomain signals are called wave-domain signals or, as they represent the intensity of particular modes within the analyzed sound field, they are often referred to as microphone and loudspeaker modes. The original domain, characterized by impulse responses or transfer functions between individual points in space, will be referred to as transducer domain.

The wave-domain transforms after [22, 24] decompose the wave fields into expansion coefficients with respect to circular harmonics [33], which form an orthogonal basis for the solutions of the acoustic wave equation in cylindrical coordinates. For these transforms, consider an  $N_{\rm L}$ -element loudspeaker array with the  $l^{\rm th}$  loudspeaker  $(l \in \{0, \ldots, N_{\rm L} - 1\})$  at radius  $\varrho_{{\rm L},l}$  and azimuth angle  $\varphi_{{\rm L},l}$ , summarized by the vector  $[\varrho_{\mathrm{L},l}, \varphi_{\mathrm{L},l}]^{\mathrm{T}}$ . Furthermore, consider an  $N_{\mathrm{M}}$ element circular microphone array with the  $m^{\text{th}}$  microphone  $(m \in \{0, \ldots, N_{\text{M}} - 1\})$  located at  $[\varrho_{\text{M},m}, \varphi_{\text{M},m}]^{\text{T}} = [R_{\text{M}}, \frac{2\pi}{N_{\text{M}}}m]^{\text{T}}$ . After wave-domain transforms, the signals can be indexed by the nonnegative wave-domain indices  $\tilde{l} \in [0, N_{\rm L} - 1]$  and  $\tilde{m} \in [0, N_{\rm M} - 1]$ , which are related to the mode numbers l and m of the spatial basis functions by  $\tilde{l} = (\tilde{l} \mod N_{\rm L})$  and  $\tilde{m} = (\tilde{m} \mod N_{\rm M})$ . Furthermore, let T2 denote the microphone signal transform and T1 denote the loudspeaker signal transform, each of which consists of a forward and a backward transform, denoted as transforms  $T_2^f$ ,  $T_2^b$ ,  $T_1^f$ , and  $T_1^b$ , respectively. In this notation, discrete-time representations of  $T_2^f$  and  $T_2^b$  after [22] are described by a MIMO system with the single-tap IRs

$$h_{\tilde{m},m}^{\text{T2f}}(k) = \delta(k) \frac{1}{\sqrt{N_{\text{M}}}} e^{-j\frac{2\pi}{N_{\text{M}}}\check{m}m}$$
(1)

$$h_{m,\tilde{m}}^{\text{T2b}}(k) = \delta(k) \frac{1}{\sqrt{N_{\text{M}}}} e^{j\frac{2\pi}{N_{\text{M}}}\tilde{m}m}$$
 (2)

with the discrete-time sample index k and the unit impulse  $\delta(k)$ , which obviously represents a Discrete Fourier Transform (DFT) of the  $N_{\rm M}$  microphone signals at a given time instant.

The corresponding loudspeaker signal transform of [22] requires fractional-delay filters [34] which model each loudspeaker signal's individual propagation to the microphone array origin. These fractional delay filters are obtained by sampling properly delayed and windowed si (·) functions, where si  $(k) = \frac{\sin(k)}{k}$ . With a window function  $w(k) = 0 \forall k \notin [0, L_{\text{TI}} - 1]$ , the forward and backward transform's scaled fractional delay filters of length  $L_{\text{TI}}$  can be expressed as

$$\begin{split} h_{\text{frac},l}^{\text{f}}\left(k\right) &= w(k) \cdot \frac{1}{\varrho_{l}} \cdot \operatorname{si}\left(k - \varrho_{l} \frac{f_{\text{S}}}{c_{\text{air}}}\right) \\ h_{\text{frac},l}^{\text{b}}\left(k\right) &= w(k) \cdot \varrho_{l} \cdot \operatorname{si}\left(k - \left(\max_{l}(\varrho_{l}) - \varrho_{l}\right) \frac{f_{\text{S}}}{c_{\text{air}}}\right), \end{split}$$

where  $c_{\text{air}}$  is the sound speed and  $\max_l(\varrho_l)$  introduces a delay for causality. The loudspeaker signal transform's IRs after [22] finally result in

$$h_{\tilde{l},l}^{\text{TIf}}(k) = h_{\text{frac},l}^{\text{f}}(k) \cdot \mathbf{j}^{\tilde{l}} \frac{1}{\sqrt{N_{\text{L}}}} \mathbf{e}^{-\mathbf{j}\tilde{l}\varphi_{l}}$$
(3)

$$h_{l,\tilde{l}}^{\text{TIb}}(k) = h_{\text{frac},l}^{\text{b}}(k) \cdot j^{-\tilde{l}} \frac{1}{\sqrt{N_{\text{L}}}} e^{j\tilde{l}\varphi_{l}}.$$
 (4)

When omitting the common discrete-time normalized angular frequency variable  $\Omega$  in the Discrete-Time Fourier Transform (DTFT) domain, the IRs of Eqs. (1) to (4) can be expressed as transfer function matrices  $\mathbf{T}_2^{f}$ ,  $\mathbf{T}_2^{b}$ ,  $\mathbf{T}_1^{f}$ , and  $\mathbf{T}_2^{h}$ , respectively. Analogously, **H** and  $\tilde{\mathbf{H}}$  denote the transfer function matrices of the LEMS and of the wave-domain LEMS, respectively. The equivalence of the resulting wave-domain representation and the transducer-domain representation of an LEMS is illustrated in Fig. 2. Therein, **x** and **d**, denote transducer-domain loudspeaker and microphone signal



**Fig. 2**: Relation between wave-domain LEMS  $\hat{\mathbf{H}}$  and transducerdomain LEMS  $\mathbf{H}$  and the roles of the wave-domain transforms  $\mathbf{T}_{1/2}^{t/b}$  with respect to the LEMS for perfectly invertible wave-domain transforms. Equivalent systems are highlighted by a commonly colored background and brackets.



**Fig. 3**: Efficient wave-domain LRE. The signals actually played back and recorded are marked by a loudspeaker and a microphone symbol, respectively.

vectors, respectively, and  $\tilde{\mathbf{x}}$  and  $\mathbf{d}$  denote wave-domain loudspeaker and microphone signal vectors, respectively. The wave-domain transforms allow a conversion between LEMS representations by

$$\mathbf{H} = \mathbf{T}_2^{\mathsf{b}} \mathbf{H} \mathbf{T}_1^{\mathsf{r}} \tag{5}$$

$$\dot{\mathbf{H}} = \mathbf{T}_2^{\mathrm{t}} \mathbf{H} \mathbf{T}_1^{\mathrm{b}}.$$
 (6)

Note that Eq. (5) implicitly defines the discrete-time wave-domain LEMS  $\tilde{\mathbf{H}}$  as the transforms are fixed, while Eq. (6) results from Eq. (5) by pre- and post-multiplication with transform matrices  $\mathbf{T}_2^f$  and  $\mathbf{T}_1^b$  under the assumption that forward and backward transforms invert each other. However, if the forward and backward transforms would not be strictly inverse up to a delay, Eq. (6) may only yield a filtered version of the ideal wave-domain representation of  $\tilde{\mathbf{H}}$ . Potentially different delays introduced by each of the transforms  $\mathbf{T}_1^f, \mathbf{T}_2^b$ , and  $\mathbf{T}_2^b$  can lead to a temporal shift of the system representation. In any case, the delays are given by the system design and could be compensated for if a time-aligned conversion between transducer and wave domain systems is needed. Note that  $N_M \neq N_L$  only changes the dimensions of the involved matrices, but not the validity of Eqs. (5) and (6), as all transform matrices remain square.

#### 3. WAVE-DOMAIN LRE AND LOUDSPEAKER ARRAYS WITH IRREGULAR AZIMUTHAL SPACING

The signal model for an adaptive wave-domain LRE system is depicted in Fig. 3. At first, the unprocessed loudspeaker signal  $\mathbf{x}$  is transformed to the wave domain by  $\mathbf{T}_1^f$ , filtered by the adaptive wave-domain equalizer  $\tilde{\mathbf{G}}(k)$ , and transformed back into equalized loudspeaker signals by  $\mathbf{T}_1^b$  before they can pass the physical LEMS **H**. After capturing the microphone signals, these are transformed into the wave domain by  $\mathbf{T}_2^f$ . As the block diagram of Fig. 3 results from the so-called filtered-X structure [35], the computation of the pre-equalizer is split into two parts: first, a filtered version of the LEMS containing the cascade of the forward and backward transform,

$$\tilde{\mathbf{H}}_{c} = \mathbf{T}_{2}^{f} \quad \mathbf{H} \quad \mathbf{T}_{1}^{b} \tag{7}$$

$$\stackrel{(5)}{=} \mathbf{T}_{2}^{\mathrm{f}} \mathbf{T}_{2}^{\mathrm{b}} \tilde{\mathbf{H}} \mathbf{T}_{1}^{\mathrm{f}} \mathbf{T}_{1}^{\mathrm{b}}, \tag{8}$$

has to be identified by employing the system identification error signal  $\tilde{\mathbf{e}}_{ss}$ . Afterwards,  $\widehat{\tilde{\mathbf{H}}}_{c}(k)$  is employed to compute the inverse filters



(a) Forward transform  $T_1^f$  of Eq. (3). (b) Backward transform  $T_1^b$  of Eq. (4).

Fig. 4: Decomposition of discrete-time loudspeaker signal transforms into  $N_{\rm L}$  SISO fractional delay systems and single-tap spatial mixing matrices.

by adaptively minimizing the equalization error signals  $\tilde{\mathbf{e}}_{\text{EQ}}$ . These error signals  $\tilde{\mathbf{e}}_{\text{EQ}}$  are obtained as difference between the actual and the desired wave-domain microphone signals, where the latter ones are determined with the desired wave-domain system  $\tilde{\mathbf{H}}_0$  as reference. Previous investigations [22] proved the applicability of approximate wave-domain models for the involved MIMO systems in order to reduce the computational complexity. In particular, a wavedomain LEMS model where only couplings with  $|\mathring{m} - \mathring{l}| < \frac{N_D}{2}$  are modeled will be denoted as model with  $N_D$  diagonals and has been shown to be effective with a low number of diagonals already in [21].

# 3.1. Impact of Azimuthally Irregularly Spaced Loudspeaker Arrays

As depicted in Fig. 4, both  $T_1^f$  and  $T_2^h$  can be decomposed into two subsystems. For the former (Fig. 4a), the first subsystem consists of the  $N_L$  fractional delay filters of Eq. (3) (without inter-loudspeaker coupling), stacked in a transfer function vector  $\mathbf{h}_{\text{frac}}^{\text{Tiff}}$ . The second subsystem is a spatial mixing matrix  $\mathbf{A}^f = \mathbf{A} = [a_{\tilde{l},l}]$  with the elements  $a_{\tilde{l},l} = 1/\sqrt{N_L} \cdot j^{\tilde{l}} e^{-j\tilde{l}\varphi_l}$ . With this, the conversion of the delayed loudspeaker signals to the wave-domain signals is then a sample-wise matrix multiplication of the former signals with  $\mathbf{A}^f$ . Consequently, the inverse transform (depicted in Fig. 4b) consists of a spatial de-mixing matrix  $\mathbf{A}^b = \mathbf{A}^H$  and a subsequent set of fractional delay filters in  $\mathbf{h}_{\text{frac}}^{\text{T1b}}$ . In case of uniform radii  $\varrho_l = R_L$ , the delay filters are identical for all loudspeakers and can therefore be neglected. For loudspeaker arrays with uniform azimuthal spacing,

$$\mathbf{A} = [a_{\tilde{l},l}] = \mathfrak{F}_{N_{\mathrm{L}}} \text{ with } a_{\tilde{l},l} = \frac{\mathrm{j}^{\tilde{l}}}{\sqrt{N_{\mathrm{L}}}} \mathrm{e}^{-\mathrm{j}\frac{2\pi}{N_{\mathrm{L}}}\tilde{l}l} \tag{9}$$

equals a unitary DFT matrix  $\mathfrak{F}_{N_{\rm L}}$  [36]<sup>1</sup>. In case of azimuthal irregularities, the complex harmonics constituting the columns of **A** are not sampled equidistantly, and are therefore not orthogonal anymore and thus do not correspond to a DFT matrix anymore (so that the approximate signs in Fig. 4 become necessary). This resembles a nonuniform DFT [37-39], where a uniform temporal (or spatial) sampling is possible, but the frequency-domain support points are nonuniformly distributed. Furthermore,  $\mathbf{A}^{b} = (\mathbf{A}^{f})^{H} \neq (\mathbf{A}^{f})^{-1}$ leads to  $\mathbf{T}_1^{f} \mathbf{T}_1^{b} \neq \mathbf{I}$ , which means that the backward transform of T1 according to Eq. (4) does not invert the forward transform according to Eq. (3). As the potentially fully coupled matrix  $\mathbf{T}_{1}^{t}\mathbf{T}_{1}^{b}$  is part of  $\mathbf{H}_c$  in Eq. (8),  $\mathbf{H}_c$  can be expected to have broader couplings than for a uniformly spaced loudspeaker setup. The lost inverse relationship also has to be considered when pre-processing loudspeaker signals in the wave domain, as the pre-processed loudspeaker signals have to be reconstructed from their wave-domain representation prior to playback. LRE is obviously an application which requires this (see Fig. 3). Along with the loss of orthogonality, the condition number [40] of A increases as well, which precludes the use of a numerically determined inverse  $\mathbf{A}^{-1}$  as it would severely amplify adaptation errors-with highly undesired effects for both users and loudspeakers of an LRE system. To give an example, the blue loudspeaker array of Fig. 5 leads to a condition number of  $C = 1.67 \times 10^3$ , while a corresponding uniform array has C = 1.

#### 3.2. Generalized Wave-Domain Transforms for Azimuthally Irregularly Spaced Transducers

In this section, the lost inverse relationship of  $T_1^f$  and  $T_D^h$  of the stateof-the-art loudspeaker signal transform, denoted as T1-A, will be tackled by introducing novel, generalized loudspeaker signal transforms T1-R, and T1-H. These improved transforms differ in terms of the spatial transform matrices  $A^{f/b}$  employed during the forward and backward transform (see Fig. 4). These generalized transforms aim at increased LRE performance by an improved reconstruction of loudspeaker signals from modes and more accurate approximations of  $\tilde{H}_c$  according to Eq. (8) with just a few diagonals  $N_D$  for azimuthally irregular loudspeaker arrays.

**T1-A: State-of-the-Art transform:** The transform matrices  $\mathbf{A}^{f} = \mathbf{A}$  and  $\mathbf{A}^{b} = \mathbf{A}^{H}$  resulting from Eqs. (3) and (4) will be referred to as  $\mathbf{A}_{A}^{f}$  and  $\mathbf{A}_{A}^{b}$ , respectively. Although both the forward and the backward transform have the same and potentially large condition number due to closely spaced loudspeakers, this does not lead to numerical problems for the inverse transform, but to inaccurate solutions with a low 2-norm, because low singular values in  $\mathbf{A}_{A}^{f}$  are not inverted by the backwards transform. These transform matrices  $\mathbf{A}_{A}^{f}$  and  $\mathbf{A}_{A}^{b}$  are employed as components of T1-A.

**T1-R: Re-orthogonalized transform:** Recall that uniformly spaced arrays result in a DFT matrix as spatial transform matrix, which has a condition number of C = 1. This desired property can be employed to formulate an optimization problem for an approximate loudspeaker signal transform. The spatial transform matrix can be approximated by a unitary matrix with Frobenius norm optimality, which can be written as optimization problem

$$\mathbf{A}_{R}^{f} = \underset{\mathbf{A}}{\operatorname{argmin}} \left\| \mathbf{A}_{A}^{f} - \mathbf{A} \right\|_{F}, \text{ s.t. } \mathbf{A}^{H} \mathbf{A} = \mathbf{I},$$
(10)

where  $\|\cdot\|_{F}$  denotes the Frobenius norm of a matrix. According to [41, p. 601], this is achieved by

$$\mathbf{A}_{\mathbf{R}}^{\mathrm{f}} = \mathbf{U}\mathbf{V}^{\mathrm{H}},\tag{11}$$

with unitary matrices U and V from the Singular Value Decomposition (SVD)  $\mathbf{A}_{A}^{f} = \mathbf{U}\Sigma\mathbf{V}^{H}$ , where  $\Sigma$  is the diagonal matrix of singular values. The product  $\mathbf{U}\mathbf{V}^{H}$  is also known as polar factor of the polar decomposition [40] of  $\mathbf{A}_{A}^{f}$  and describes a rotation—the component of  $\mathbf{A}_{A}^{f}$  describing a stretching is discarded. Hence, the vector spaces between which the mapping takes place are similar for T1-A and T1-R (they actually share the same basis vectors in U and  $\mathbf{V}^{H}$ ). Thus, the re-orthogonalized transform T1-R, can be expected to retain most of the spatial properties of T1-A. As  $\mathbf{A}_{R}^{f}$  is a unitary matrix, its inverse can be determined without numerical problems and is simply given by

$$\mathbf{A}_{R}^{b} = \left(\mathbf{A}_{R}^{f}\right)^{-1} = \left(\mathbf{A}_{R}^{f}\right)^{H} = \mathbf{V}\mathbf{U}^{H}.$$
 (13)

Still, the wave-domain coefficients determined by  $\mathbf{A}_{R}^{T}$  will have a larger deviation from the true loudspeaker modes than for  $\mathbf{A}_{A}^{f}$ . Thereby, this approach trades modeling accuracy for numerically stable inversion. Note that, while a subunit of a spatial transform is being approximated by a unitary transform here in order to obtain the signal-independent transform T1-R, a similar procedure has been applied in [42] for the approximation of filterbank transfer matrices, taking into account the single-channel signal's autocorrelation as well.

**T1-H: Hybrid transform:** Additionally, consider the transform pair with the spatial transform matrix combination

$$\mathbf{A}_{\mathrm{H}}^{\mathrm{r}} = \mathbf{A}_{\mathrm{A}}^{\mathrm{r}} \tag{14}$$

$$\mathbf{A}_{\mathrm{H}}^{\mathrm{b}} = \mathbf{A}_{\mathrm{R}}^{\mathrm{b}},\tag{15}$$

which is justified because  $\mathbf{A}_{R}^{b}\mathbf{A}_{A}^{f} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^{H}$  is even closer to an identity matrix than  $\mathbf{A}_{A}^{b}\mathbf{A}_{A}^{f} = \mathbf{V}\boldsymbol{\Sigma}^{2}\mathbf{V}^{H}$ . Therefore, a performance increase of T1-H with respect to T1-A can be expected. Additional approximation errors like for  $\mathbf{A}_{R}^{f}$  are prevented.

<sup>&</sup>lt;sup>1</sup>Equality holds up to row-wise multiplications with powers of j, which is merely a phase shift of the basis functions.



Fig. 5: Arrays in an image-source environment: microphones are arranged around the listening area as uniform circular array (red) and the loudspeakers are placed irregularly but concentrically on a circle (blue). The latter array resulted from the projection of an amorphous irregular array onto a circle.

#### 3.3. Performance Measures for LRE

In this section, objective performance measures for LRE systems will be specified. First of all, the performance of an LRE system can be measured by the difference between the actually reproduced sound field and the desired one, sampled by microphones at discrete spatial locations. To this end, the set of error microphones required for adapting the LRE system can be employed again. For WDAF, these error microphones are arranged as a ring of  $N_{\rm M}$  microphones with central radius  $R_{\rm M}$  (red microphones in Fig. 5). With these, the rendering-signal-dependent scene equalization errors can be defined in the DTFT domain as

$$e_{\rm M}(k) = 10 \log_{10} \left( \frac{\int_0^{2\pi} \|({\bf HG}(k) - {\bf H}_0) {\bf x}\|_2^2 \mathrm{d}\Omega}{e_0} \right) \mathrm{dB}, \qquad (16)$$

which describes the energy of the deviation from the desired signal after equalization with  $\mathbf{G}(k) = \mathbf{T}_1^{\mathrm{b}} \tilde{\mathbf{G}}(k) \mathbf{T}_1^{\mathrm{f}}$ , normalized to the respective energy  $e_0$  without adapted equalizer filters  $\mathbf{G}(k)$ . Similarly, a signal-independent measure, denoted as **room equalization error**, can be computed from

$$E_{\rm M}(k) = 10 \log_{10} \left( \frac{\int_0^{2\pi} \|\mathbf{H}\mathbf{G}(k) - \mathbf{H}_0\|_{\rm F}^2 \mathrm{d}\Omega}{E_0} \right) \mathrm{d}\mathbf{B}.$$
 (17)

It measures the logarithmic ratio between the Frobenius norm of the undesired signal components with equalization (numerator) and without equalization ( $E_0$  in denominator). While a low system equalization error  $E_M(k)$  (signal-independent measure) implies a low scene equalization error  $e_M(k)$  (signal-dependent measure) as well, a low scene equalization error does not ensure a low system equalization error. This means that the actual scene might be reproduced very well although the actual LEMS is not identified and equalized correctly. For evaluation, Eqs. (16) and (17) will be computed (employing Parseval's theorem) in the time domain and the time-dependency will be dropped, as the measures will be computed with the filters at the end of a simulation with  $k = k_{\text{final}}$ .

### 4. EXPERIMENTAL RESULTS

The system to be equalized in the following experiments is a 2<sup>nd</sup>order image source environment [43] in which a circular loudspeaker array and a circular microphone array with  $N_{\rm M} = N_{\rm L} = 48$  elements and radii of  $R_{\rm L} = 1.5$  m and  $R_{\rm M} = 0.5$  m are placed centered at the origin of the coordinate system, respectively, as depicted in Fig. 5. The image-source environment is composed of four plane walls with a reflection coefficient of  $r_{\rm wall} = 0.9$ , while floor and ceiling are modeled as fully absorptive. The simulations will be conducted at a sampling rate of  $f_{\rm S} = 2000$  Hz, for which the sound field can be controlled by the loudspeakers reasonably well (an exact spatial aliasing frequency cannot be given for the irregular array: instead



**Fig. 6**: Benefit for LRE by employing the re-orthogonalized T1-R ( $\diamond$ ) or T1-H ( $\times$ ) instead of T1-A (+) for an azimuthally irregular loudspeaker placement for  $N_{\rm S} = 3$  virtual sources. As reference, the corresponding experiment for a uniform loudspeaker array (o) is included as well.

of periodic repetitions, noise-like artifacts increase with frequency for irregular spatial sampling [44]). In the experiments, WFS [10] is employed to synthesize  $N_{\rm S} = 3$  virtual plane waves. The plane waves carry mutually independent white Gaussian noise signals and impinge with an angular spacing of  $\frac{2\pi}{N_S}$  onto the microphone array. To assess the suitability of approximate wave-domain models, the number of diagonals  $N_D$  (see Sec. 3) for the wave-domain models is varied between  $N_{\rm D}=1$  and  $\dot{N}_{\rm D}=48$ . The estimation of the LEMS and the adaptation of the equalizer filters has been performed with the modified Generalized Frequency-Domain Adaptive Filtering (GFDAF) algorithm according to [23]. Figure 6 depicts the performance measures for the state-of-the-art T1-A (+), the novel transforms T1-R ( $\diamond$ ) and T1-H ( $\times$ ), and for a uniform loudspeaker array (o) with the same  $R_{\rm L} = 1.5$  m. The latter has been included as reference, for which all three generalized transforms according to Sec. 3.2 are identical. Furthermore,  $e_0$  and  $E_0$  have been determined for the uniform loudspeaker setup and used for the irregular setups as well, which allows a direct assessment of the performance loss due to irregular geometries.

As expected, the best performance is achieved by a WDAF LRE with Uniform Circular Concentric Arrays (UCCAs) (o) -both in terms of scene equalization (upper plot in Fig. 6) and in terms of room equalization (bottom plot in Fig. 6). The increasing room equalization errors in case of  $N_{\rm D} \gg 3$  can be accounted to an overadaptation to the virtual scene (analogously to the non-uniqueness problem of multichannel AEC [16]). The other extremal case is given by the irregular loudspeaker array in combination with the state-of the art transform T1-A (+), which achieves a moderate equalization of the played-back virtual scene  $(e_{\rm M})$ , but which totally fails to equalize the actual impulse responses of the room in this experiment (positive error measures  $E_{\rm M}$ ). The novel transform T1-H  $(\times)$  with improved inverse transform clearly outperforms the state-of-the-art transform. Still, the fully re-orthogonalized T1-R ( $\diamond$ ) clearly performs best for the irregular loudspeaker array, although a moderate degradation with respect to the UCCAs has to be accepted.

#### 5. CONCLUSION

In this work, the wave-domain transforms based on circular harmonics have been extended for loudspeaker arrays with irregular azimuthal spacing. For this case, a new re-orthogonalized transform and a hybrid transform have been proposed which significantly improve LRE performance over the state-of-the-art transform for an irregular/nonuniform loudspeaker spacing and are identical to the state-of-the-art transform for a uniform spacing. Best performance for irregular arrays is obtained with the fully re-orthogonalized transform, which shows that the exact invertibility of this transform outweighs the additional approximation with respect to the hybrid transform.

#### 6. REFERENCES

- M. A. Gerzon, "Periphony: With-height sound reproduction," *Journal* of the Audio Engineering Society, vol. 21, no. 1, pp. 2–10, 1973.
- [2] M. A. Gerzon, "Ambisonics in multichannel broadcasting and video," *Journal of the Audio Engineering Society*, vol. 33, no. 11, pp. 859–871, 1985.
- [3] J. Ahrens and S. Spors, "Analytical driving functions for higher-order Ambisonics," in *IEEE International Conference on Acoustics, Speech* and Signal Processing (ICASSP), Las Vegas, USA, April 2008, pp. 373–376.
- [4] M. A. Poletti, "Robust two-dimensional surround sound reproduction for nonuniform loudspeaker layouts," *Journal of the Audio Engineering Society*, vol. 55, no. 7/8, pp. 598–610, 2007.
- [5] G. Lilis, D. Angelosante, and G. Giannakis, "Sound field reproduction using the lasso," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 8, pp. 1902–1912, 2010.
- [6] T. Betlehem and C. Withers, "Sound field reproduction with energy constraint on loudspeaker weights," *IEEE Transactions on Audio*, *Speech, and Language Processing*, vol. 20, no. 8, pp. 2388–2392, 2012.
- [7] M. A. Poletti, T. Betlehem, and T. D. Abhayapala, "Higher-order loudspeakers and active compensation for improved 2d sound field reproduction in rooms," *Journal of the Audio Engineering Society*, vol. 63, no. 1/2, pp. 31–45, 2015.
- [8] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *Journal of the Acoustical Society of America (JASA)*, vol. 93, no. 5, pp. 2764–2778, 1993.
- [9] E. N. G. Verheijen, "Sound reproduction by wave field synthesis," Ph.D. dissertation, Delft University of Technology (TU Delft), 1998.
- [10] S. Spors, R. Rabenstein, and J. Ahrens, "The theory of wave field synthesis revisited," in Audio Engineering Society Convention 124, 2008.
- [11] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, 1997.
- [12] M. Shin *et al.*, "Maximization of acoustic energy difference between two spaces," *Journal of the Acoustical Society of America (JASA)*, vol. 128, no. 1, pp. 121–131, 2010.
- [13] M. A. Poletti, "An investigation of 2-D multizone surround sound systems," in *Audio Engineering Society Convention 125*, San Francisco, USA, October 2008.
- [14] M. Buerger, C. Hofmann, and W. Kellermann, "The impact of loudspeaker array imperfections and reverberation on the performance of a multizone sound field synthesis system," in *3rd International Conference on Spatial Audio*, Graz, Austria, September 2015.
- [15] M. Buerger, R. Maas, H. W. Löllmann, and W. Kellermann, "Multizone sound field synthesis based on the joint optimization of the sound pressure and particle velocity vector on closed contours," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (WASPAA), October 2015.
- [16] J. Benesty, D. Morgan, and M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 156–165, March 1998.
- [17] H. Buchner, S. Spors, and W. Kellermann, "Wave-domain adaptive filtering: Acoustic echo cancellation for full-duplex systems based on wave-field synthesis," in *IEEE International Conference on Acoustics*, *Speech and Signal Processing (ICASSP)*, vol. 4, Montreal, Canada, May 2004, pp. iv–117 – iv–120.
- [18] S. Spors, H. Buchner, and R. Rabenstein, "A novel approach to active listening room compensation for wave field synthesis using wave-domain adaptive filtering," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, Montreal, Canada, May 2004, pp. iv–29 – iv–32.
- [19] S. Spors, H. Buchner, R. Rabenstein, and W. Herbordt, "Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering," *Journal of the Acoustical Society of America*, vol. 122, no. 1, pp. 354–369, July 2007.
- [20] 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, pp. 133–138.

- [21] M. Schneider and W. Kellermann, "Adaptive listening room equalization using a scalable filtering structure in the wave domain," in *IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), Kyoto, Japan, March 2012, pp. 13–16.
- [22] M. Schneider and W. Kellermann, "A direct derivation of transforms for wave-domain adaptive filtering based on circular harmonics," in *European Signal Processing Conference (EUSIPCO)*, Bucharest, Romania, August 2012, pp. 1034–1038.
- [23] M. Schneider and W. Kellermann, "Apparatus and method for listening room equalization using a scalable filtering structure in the wave domain," Patent PCT/EP2012/068 562, April 04, 2013.
- [24] C. Hofmann and W. Kellermann, "Wave-domain transforms for irregular loudspeaker array topologies," in 41. Jahrestagung für Akustik (DAGA 2015), Nürnberg, Germany, March 2015.
- [25] J. Jungmann, R. Mazur, and A. Mertins, "Joint time-domain reshaping and frequency-domain equalization of room impulse responses," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 6642–6646.
- [26] B. S. Atal and M. R. Schroeder, "Apparent sound source translator," US Patent US3 236 949A, February 22, 1966.
- [27] D. B. Ward and G. Elko, "Optimum loudspeaker spacing for robust crosstalk cancellation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 6, Seattle, USA, May 1998, pp. 3541–3544.
- [28] A. Sakurai and S. Trautmann, "Cross-talk cancellation," Patent US 2005/0254 660 A1, November, 2005.
- [29] L. Krishnan, P. D. Teal, and T. Betlehem, "A robust sparse approach to acoustic impulse response shaping," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brisbane, Australia, April 2015, pp. 738–742.
- [30] J. Grosse and S. van de Par, "Perceptually accurate reproduction of recorded sound fields in a reverberant room using spatially distributed loudspeakers," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 867–880, August 2015.
- [31] J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, "MPEG-H audio—the new standard for universal spatial/3D audio coding," *Journal of the Audio Engineering Society*, vol. 62, no. 12, pp. 821–830, 2015.
- [32] ISO/IEC JTC 1/SC 29/WG 11 N15066, "MPEG-H 3D audio progresses to international standard," Geneva, Switzerland, February 2015.
- [33] A. Kuntz, "Wave field analysis using virtual circular microphone arrays," Ph.D. dissertation, Friedrich-Alexander University Erlangen-Nürnberg (FAU), June 2008.
- [34] T. Laakso, V. Välimäki, M. Karjalainen, and U. Laine, "Splitting the unit delay," *IEEE Signal Processing Magazine*, vol. 13, pp. 30–60, January 1996.
- [35] E. Bjarnason, "Analysis of the filtered-X LMS algorithm," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 6, pp. 504–514, November 1995.
- [36] D. F. Elliott and K. R. Rao, Fast Transforms Algorithms, Analyses, Applications. Academine Press, 1982.
- [37] S. Bagchi and S. K. Mitra, *The Nonuniform Discrete Fourier Transform and Its Applications in Signal Processing*. Kluwer Academic Publishers, 1999.
- [38] A. Makur and S. Mitra, "Warped discrete-fourier transform: Theory and applications," *IEEE Transactions on Circuits and Systems*, vol. 48, no. 9, pp. 1086–1093, September 2001.
- [39] A. Makur, "Computational schemes for warped DFT and its inverse," *IEEE Transactions on Circuits and Systems*, vol. 55, no. 9, pp. 2686–2695, October 2008.
- [40] G. Strang, *Introduction to Linear Algebra*, 4th ed. Wellesley-Cambridge, 2009.
- [41] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Johns Hopkins University Press, 1996.
- [42] R. De Queiroz, "On unitary transform approximations," *IEEE Signal Processing Letters*, vol. 5, no. 2, pp. 46–47, February 1998.
- [43] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, April 1979.
- [44] G. Hennenfent and F. J. Herrmann, "Irregular sampling-from aliasing to noise," in 69th EAGE Conference & Exhibition, London, UK, June 2007.