Audio Engineering Society
Convention Paper
Presented at the 116th Convention
2004 May 8–11 Berlin, Germany

Full-Duplex Systems for Sound Field Recording and Auralization Based on Wave Field Synthesis

Herbert Buchner, Sascha Spors, and Walter Kellermann[1]

[1] *University of Erlangen-Nuremberg, Cauerstr. 7, 91058 Erlangen, Germany*

Correspondence should be addressed to Herbert Buchner (buchner@LNT.de)

ABSTRACT

For high-quality multimedia communication systems such as telecollaboration or virtual reality applications, both multichannel sound reproduction and full-duplex capability are highly desirable. Full 3D sound spatialization over a large listening area is offered by wave field synthesis, where arrays of loudspeakers generate a prespecified sound field. However, before this new technique can be utilized for full-duplex systems with microphone arrays and loudspeaker arrays, an efficient solution to the problem of multichannel acoustic echo cancellation (MC AEC) has to be found in order to avoid acoustic feedback. This paper presents a novel approach that extends the current state of the art of MC AEC and transform-domain adaptive filtering by reconciling the flexibility of adaptive filtering and the underlying physics of acoustic waves in a systematic and efficient way. Our new framework of wave-domain adaptive filtering (WDAF) explicitly takes into account the spatial dimensions of loudspeaker arrays and microphone arrays with closely spaced transducers. Experimental results with a 32-channel AEC verify the concept for both, simulated and measured room acoustics.

1. INTRODUCTION

Multichannel techniques for reproduction and acquisition of speech and audio signals at the acoustic human-machine interface offer spatial selectivity and diversity as additional degrees of freedom over single-channel systems.

Multichannel sound reproduction enhances sound realism in virtual reality and multimedia communication systems, such as teleconferencing or teleteaching (especially of music), and aims at creating a three-dimensional illusion of sound sources positioned in a virtual acoustical environment. However, advanced loudspeaker-based approaches, like the 3/2-surround format still rely on a restricted listening area ('sweet spot'). A volume solution for a large listening space is offered by the Wave Field Synthesis (WFS) method [1] which is based on wave physics. In WFS, arrays of a large number P of individually driven loudspeakers generate a prespecified sound field. P may lie between 20 and several hundred.

On the recording side of advanced acoustic human-machine interfaces, the use of microphone arrays [2], where the number Q of microphones may reach up to 500 [3], is an effective approach to separate desired and undesired sources in the listening environment, and to cope with reverberation in the recorded signal. Figure 1 shows an example for a general multichannel communication setup.

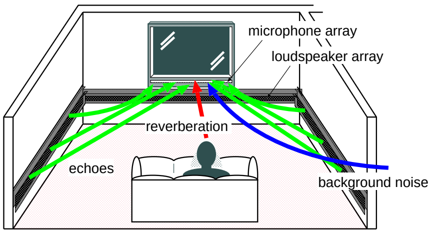

Fig. 1: Exemplary setup for multichannel communication.

In this paper, we consider full-duplex systems based on such massive multichannel techniques for high-quality recording and reproduction. A major challenge to fully exploit the potential of array processing in such applications lies in the development of adaptive MIMO (multiple-input and multiple-output) systems that are suitable for the large number of channels in this environment. The point-to-point optimization in adaptive MIMO systems often suffers from convergence problems and high computational complexity so that some applications are beyond reach with current techniques. In particular, before full-duplex communication in two-way systems can be deployed, acoustic echo cancellation (AEC) needs to be implemented for the resulting $P \cdot Q$ echo paths which seems to be out of reach for current multichannel AEC [4, 9] in conjunction with large loudspeaker arrays for spatial audio. Similar problems arise in other building blocks of the acoustic interface, e.g., for acoustic room compensation (ARC) on the reproduction side, where a system of suitable prefilters takes into account the actual room acoustics prior to sound reproduction by WFS, and also for adaptive interference cancellation on the recording side [2].

To address the specific problems of adaptive array processing for acoustic human-machine interfaces, we present in this paper a novel framework for spatio-temporal transform-domain adaptive filtering, called wave-domain adaptive filtering (WDAF). This concept is reconciling the flexibility of adaptive filtering and the underlying physics described by the acoustic wave equation. It is suitable for spatial audio reproduction systems like wave field synthesis with an arbitrarily high number of reproduction channels. Although we refer here to two-dimensional wave fields and WFS, the proposed technique can also be applied to Ambisonics and extended to 3D fields. We illustrate the concept by means of a full-duplex acoustic interface consisting of an AEC, beamforming for signal acquisition, and acoustic room compensation for high-quality sound reproduction.

2. WAVE FIELD SYNTHESIS AND ANALYSIS

Sound reproduction by wave field synthesis (WFS) using loudspeaker arrays is based on Huygens principle. It states that any point of a wave front of a propagating sound pressure wave $p(\mathbf{r}, t)$ at any instant of time conforms to the envelope of spherical waves emanating from every point on the wave front at the prior instant. This principle can be used to synthesize acoustical wavefronts of arbitrary shape. Due to the reciprocity of wave propagation it also applies to wave field analysis (WFA) on the recording side. Its mathematical formulation is given by the Kirchhoff-Helmholtz integrals (e.g., [1, 10]) which can be derived from the acoustic wave equation (given here for lossless media) and Newton's second law,

$$\nabla^2 p(\mathbf{r},t) - \frac{1}{c^2}\frac{\partial^2 p(\mathbf{r},t)}{\partial t^2} = 0, \tag{1}$$

$$-\nabla p(\mathbf{r},t) = \rho\frac{\partial \mathbf{v}(\mathbf{r},t)}{\partial t}, \tag{2}$$

respectively, where c denotes the velocity of sound, ρ is the density of the medium, and $\mathbf{v}(\mathbf{r},t)$ is the particle velocity. Since we assume two-dimensional wavefields, we choose polar coordinates (r,θ) throughout this paper. Using the second theorem of Green, applied to a contour C enclosing a region $\mathcal{S}$, we obtain from (1) and (2) the 2D forward Kirchhoff-Helmholtz integral

$$\begin{aligned}\underline{p}^{(2)}(\mathbf{r},\omega) = \frac{-jk}{4}\oint_C \Big\{&\underline{p}(\mathbf{r}',\omega)\cos\varphi H_1^{(2)}(k\Delta r)\\ &+ j\rho c\underline{v}_{\mathrm{n}}(\mathbf{r}',\omega)H_0^{(2)}(k\Delta r)\Big\}\,d\ell\end{aligned} \tag{3}$$

and the 2D inverse Kirchhoff-Helmholtz integral

$$\begin{aligned}\underline{p}^{(1)}(\mathbf{r},\omega) = \frac{-jk}{4}\oint_C \Big\{&\underline{p}(\mathbf{r}',\omega)\cos\varphi H_1^{(1)}(k\Delta r)\\ &+ j\rho c\underline{v}_{\mathrm{n}}(\mathbf{r}',\omega)H_0^{(1)}(k\Delta r)\Big\}\,d\ell,\end{aligned} \tag{4}$$

where $\Delta r = ||\mathbf{r}-\mathbf{r}'||$ and $k = \omega/c$. $H_n^{(1)}$ and $H_n^{(2)}$ are the Hankel functions of the first and second kind, respectively, which are the fundamental solutions of the wave equation in polar coordinates. All quantities in the temporal frequency domain are underlined. $\underline{v}_{\mathrm{n}}$ denotes the frequency-domain version of the radial component of $\mathbf{v}$. The total wave field is then given by the sum of the incoming and outgoing contributions w.r.t. $\mathcal{S}$:

$$\underline{p}(\mathbf{r},\omega) = \underline{p}^{(1)}(\mathbf{r},\omega) + \underline{p}^{(2)}(\mathbf{r},\omega). \tag{5}$$

The 2D Kirchhoff-Helmholtz integrals (3) and (4) state that at any listening point within the source-free listening area the sound pressure can be calculated if both the sound pressure and its gradient are known on the contour C enclosing this area. For practical implementations in 2D sound fields the acoustic sources on the closed contour are realized by loudspeakers on discrete positions. Note that (3) and (4) can analogously be applied for wave field analysis using a microphone array consisting of pressure and pressure gradient microphones. The spatial sampling along the contour C defines the aliasing frequencies. While microphone spacings are usually designed for a wide frequency range, lower aliasing frequencies may be tolerated for reproduction as the human auditory system seems not to be very sensitive to spatial aliasing artifacts above approximately 1.5kHz. Thus, without loss of generality, for higher frequencies, a practical system could be easily complemented by other existing methods, e.g., 5.1 systems.

3. CONVENTIONAL ADAPTIVE MULTI-CHANNEL PROCESSING

3.1. Multichannel Acoustic Echo Cancellation

Classical AEC applications are hands-free telephony or teleconference systems, where most of them are still based on monaural sound reproduction. Only recently, first stereophonic prototypes appeared [11], [12], and lately, it has become possible to extend the system to the multichannel case (for 5-channel surround sound see, e.g., [13]). In this paper, the concept of this frequency-domain framework will be extended for WFS in Sect. 4.

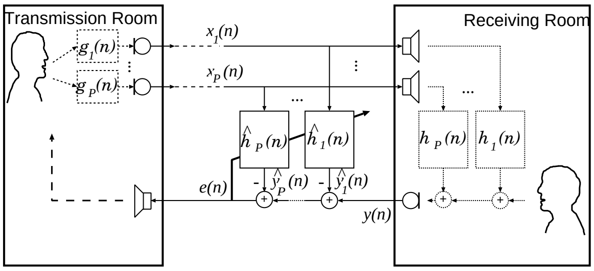

Fig. 2: Basic MC AEC structure

The fundamental idea of any P-channel AEC structure (Fig. 2) is to use adaptive FIR filters of length L with impulse response vectors $\hat{\mathbf{h}}_i(n)$, $i = 1,\ldots,P$ that continuously identify the truncated (generally time-varying) echo path impulse responses $\mathbf{h}_i(n)$ whenever the sources in the receiving room are inactive. The filters $\hat{\mathbf{h}}_i(n)$ are stimulated by the loudspeaker signals $x_i(n)$ and, then, the resulting echo estimates $\hat{y}_i(n)$ are subtracted from the microphone signal $y(n)$ to cancel the echoes. For multiple microphones, each of them is considered separately in this way. The filter length L may be on the order of several thousand.

The specific problems of MC AEC include all those known for mono AEC, but in addition to that, MC AEC often has to cope with high correlation of the different loudspeaker signals [7, 9]. The correlation results from the fact that the signals are almost always derived from common sound sources in the transmission room, as shown in Fig. 2. The optimization problem therefore often leads to a severely ill-conditioned normal equation to be solved for the $P \cdot L$ filter coefficients. Therefore, sophisticated adaptation algorithms taking the cross-correlation into account are necessary for MC AEC [9] (see Sect. 3.3).

3.2. A Conventional Approach to System Integration with WFS

Figure 3 shows a multichannel loudspeaker-enclosure-microphone (LEM) setup which acts as transmission and receiving room simultaneously. In general, the loudspeaker signals are generated in a two-step procedure: auralization of a transmission room or an arbitrary virtual room, and compensation of the acoustics in the receiving room. Auralization using WFS is performed by convolution of source signals $x''(n)$ with a - generally time-varying - matrix $\mathbf{A}(n)$ of impulse responses which may be computed according to the WFS theory as shown above [1]. Matrix $\mathbf{G}(n)$ stands for an adaptive MIMO system for acoustic room compensation (ARC). Similar to AEC for array processing, ARC is still a challenging research topic, as it requires measurement and control of the wave field in the entire listening area which is hardly possible with current methods. (The impulse response matrix from the WFS array to the possible listener positions is given by $\mathbf{H}_L(n)$, while the corresponding matrix from the WFS array to the microphone array is given by $\mathbf{H}(n)$.) However, application of the new concept presented in Sect. 4, to ARC shows promising results [6].

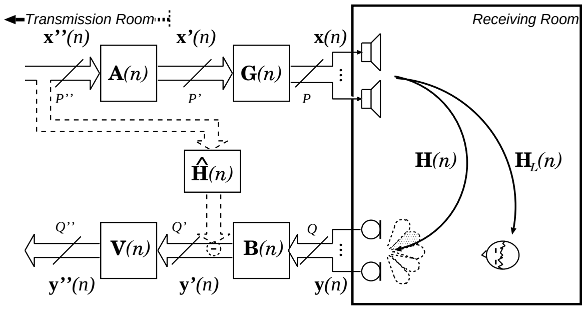

Fig. 3: Building blocks of conventional structure.

Processing on the recording side using fixed or time-varying (adaptive) beamformers (BF) can generally be described by another MIMO system $\mathbf{B}(n)$ in Fig. 3. Using $\mathbf{B}(n)$, beams of increased sensitivity can be directed at the active talker(s), so that interfering sources, background noise, and reverberation are attenuated at the output of $\mathbf{B}(n)$.

To facilitate the integration of AEC into the microphone path, a decomposition of $\mathbf{B}(n)$ may be carried out, e.g., as shown in [2, 12]. At first, a set of Q' fixed beams is generated from the Q microphone signals. These fixed beams cover all potential sources of interest and correspond to a time-invariant impulse response matrix $\mathbf{B}(n)$. The fixed beamformer is followed by a time-variant stage $\mathbf{V}(n)$ ('voting').

The advantage of this decomposition is twofold. At first, it allows integration of AEC as explained below. Secondly, automatic beam steering towards sources of interest is possible, whereby external information on the positions via audio, video, or multimodal object localization can be easily incorporated.

When placing the AEC between the two branches in Fig. 3, ideally, it is desirable that the number of impulse responses to be identified is minimized and the echo paths are time-invariant or very slowly time-varying. In [4] it has been concluded that the most practical solution is placing the AEC between $\mathbf{x}''$ and $\mathbf{y}'$ as shown in Fig. 3, since placing the AEC in parallel to the room echoes $\mathbf{H}(n)$ (i.e., between $\mathbf{x}$ and $\mathbf{y}$) is prohibitive due to the high number of $P \cdot Q$ impulse responses. On the other hand, positioning the AEC between $\mathbf{x}''$ and $\mathbf{y}''$ ($P'' \cdot Q''$ impulse responses) would include the time-variant matrix $\mathbf{V}(n)$ into the LEM model. However, a major drawback of this system is that the wave field rendering system $\mathbf{A}(n)$ is not allowed to be time-varying which limits the applicability to render only fixed virtual sources. The new approach in Sect. 4 does not exhibit this limitation.

3.3. Multichannel Adaptive Filtering

For various ill-conditioned optimization problems in adaptive signal processing, such as MC AEC, the recursive least-squares (RLS) algorithm is known to be the optimum choice in terms of convergence speed as - in contrast to other algorithms - it exhibits properties that are independent of the eigenvalue spread, i.e., the condition number, of the input correlation matrix [14]. The update equation of the multichannel RLS (MC RLS) algorithm reads for one output channel

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mathbf{R}_{\mathbf{xx}}^{-1}(n)\mathbf{x}(n)e(n), \qquad (6)$$

where $\hat{\mathbf{h}}(n)$ is the multichannel coefficient vector obtained by concatenating the length-L impulse response vectors $\hat{\mathbf{h}}_i(n)$ of all input channels, $e(n) = y(n) - \hat{y}(n)$ is the current residual error between the echoes and the echo replicas. The length-PL vector $\mathbf{x}(n)$ is a concatenation of the input signal vectors containing the L most recent input samples of each channel. The correlation matrix $\mathbf{R}_{\mathbf{xx}}$ takes all auto-correlations within, and - most importantly for multichannel processing - all cross-correlations between the input channels into account (see upper left corner of Fig. 4). However, the major problems of RLS algorithms are the very high computational complexity (mainly due to the large matrix inversion) and potential numerical instabilities which often limit the actual performance in practice.

An efficient and popular alternative to time-domain algorithms are transform-domain adaptive filtering algorithms [15], and in particular algorithms working in the DFT-domain, called frequency-domain adaptive filtering (FDAF) algorithms [16]. In FDAF, the adaptive filters are updated in a block-by-block fashion, using the fast Fourier transform (FFT) as a powerful vehicle. Recently, the FDAF approach has been extended to the multichannel case (MC FDAF) by a mathematically rigorous derivation based on a weighted least-squares criterion [13, 17]. It has been shown that there is a generic wideband frequency-domain algorithm which is equivalent to the RLS algorithm. As a result of this approach, the arithmetic complexity of multichannel algorithms can be significantly reduced compared to time-domain adaptive algorithms while the desirable RLS-like properties and the basic structure of (6) are maintained by an inherent approximate block-diagonalization of the correlation matrix as shown in the second column of Fig. 4. This allows to perform the matrix inversion in (6) in a frequency-bin selective way using only small and better conditioned $P \times P$ matrices $\mathbf{S}_{\mathrm{xx}}^{(\nu)}$ in the bins $\nu = 0, \ldots, 2L-1$. Note that all cross-correlations between different input channels are still fully taken into account by this approach.

4. THE NOVEL APPROACH: WAVE-DOMAIN ADAPTIVE FILTERING

With the dramatically increased number of highly correlated loudspeaker channels in WFS-based systems, even the matrices $\mathbf{S}_{\mathrm{xx}}^{(\nu)}$ become large and ill-conditioned so that current algorithms cannot be used. In this section we extend the conventional concept of MC FDAF by a more detailed consideration of the spatial dimensions and by exploitation of wave physics foundations shown in Sect. 2.

4.1. Basic Concept

From a physical point of view, the nice properties of FDAF result from the orthogonality property of the DFT basis functions, i.e., the complex exponentials. Obviously, these exponentials also separate the temporal dimension of the wave equation (1). Therefore, it is desirable to find a suitable spatio-temporal transform domain based on orthogonal basis functions that allow not only an approximate decomposition among the temporal frequencies as in MC FDAF, but also an approximate spatial decomposition with basis functions fulfilling (1) as illustrated by the third column of Fig. 4. In the next subsection we will introduce a suitable transform domain. Performing the adaptive filtering in a spatio-temporal transform domain requires spatial sampling on both, the input and the output of the system. Then, in contrast to conventional MC FDAF, not only all loudspeaker signals, but also all microphone signals must simultaneously be taken

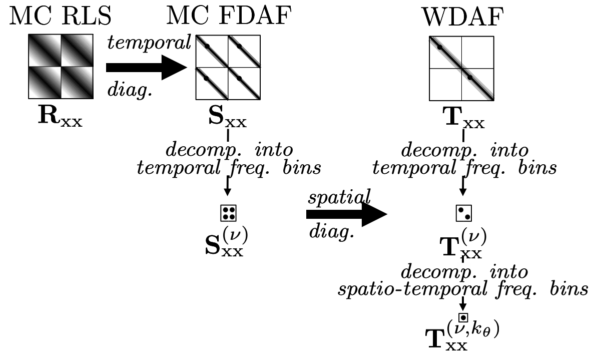
MC RLS
MC FDAF
WDAF
temporal
diag.
$\mathbf{R}_{\mathrm{xx}}$
$\mathbf{S}_{\mathrm{xx}}$
$\mathbf{T}_{\mathrm{xx}}$
decomp. into
temporal freq. bins
decomp. into
temporal freq. bins
spatial
diag.
$\mathbf{S}_{\mathrm{xx}}^{(\nu)}$
$\mathbf{T}_{\mathrm{xx}}^{(\nu)}$
decomp. into
spatio-temporal freq. bins
$\mathbf{T}_{\mathrm{xx}}^{(\nu,k_\theta)}$

Fig. 4: Illustration of the WDAF concept and its relation to conventional algorithms.

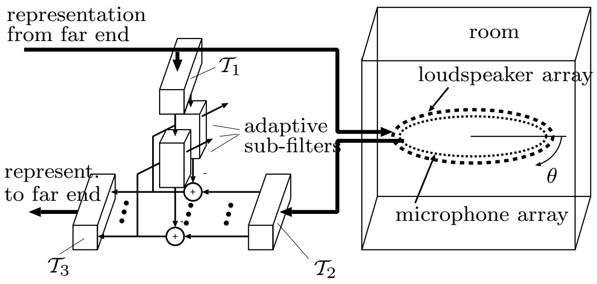

Fig. 5: Setup for proposed AEC in the wave domain.

into account for the adaptive processing. Moreover, with the given orthogonality between the spatial components in the transform domain, most cross-channels in the transform domain can be completely neglected, so that in practice only the main diagonal (see Sect. 6), and possibly (depending on the application) the first off-diagonals of the filter coefficient matrix need to be adapted. This leads to the general setup of WDAF-based acoustic interface processing incorporating spatial filtering (analogously to Fig. 3) and AEC, as shown in Fig. 5. Due to the decoupling of the channels, not only the convergence properties are improved but also the computational complexity is reduced dramatically. Let us assume $Q = P$ microphone channels. In the simplest case, instead of P^2 filters in the conventional approach, we only need to adapt P channels in the transform domain. By additionally taking into account the symmetry property of spatial frequency components, this number is further reduced to $P/2$. Thus, for a typical system with $P = 48$, the number of channels is reduced from 2304 to 24 (or, e.g., 70 if we also include the first off-diagonals).

4.2. Transformations and Adaptive Filtering

In this section we introduce suitable transformations $\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{T}_3$ shown in Fig. 5. Note that in general there are many possible spatial transformations depending on the choice of the coordinate system. A first approach to obtain the desired decoupling would be to simply perform spatial Fourier transforms analogously to the temporal dimension. This corresponds to a decomposition into plane waves [10] which is known to be a flexible format for auralization purposes [18]. However, in this case we would need loudspeakers and microphones at each point of the listening area which is not practicable. Therefore, plane wave decompositions taking into account the Kirchhoff-Helmholtz integrals are desirable. These transformations depend on the array geometries and have been derived for various configurations [18]. Circular arrays are known to show a particularly good performance in wave field analysis [18], and lead to an efficient WDAF solution. A cylindrical coordinate system is used (see Fig. 5 for the definition of the angle θ.). For the realization, temporal and spatial sampling are implemented according to the desired spatial aliasing frequency. For transform $\mathcal{T}_1$ we obtain [18] the following plane wave decomposition of the wave field to be emitted by the loudspeaker array with radius R:

$$\begin{aligned}\underline{\tilde{x}}^{(1)}(k_\theta,\omega) &= \frac{j^{1-k_\theta}}{D_R(k_\theta,\omega)}\Big\{H_{k_\theta}^{(2)'}(kR)\underline{\tilde{p}}_{\mathrm{x}}(k_\theta,\omega)\\ &\qquad -H_{k_\theta}^{(2)}(kR)j\rho c\underline{\tilde{v}}_{\mathrm{x,n}}(k_\theta,\omega)\Big\}, \qquad (7)\\ \underline{\tilde{x}}^{(2)}(k_\theta,\omega) &= \frac{-j^{1+k_\theta}}{D_R(k_\theta,\omega)}\Big\{H_{k_\theta}^{(1)'}(kR)\underline{\tilde{p}}_{\mathrm{x}}(k_\theta,\omega)\\ &\qquad -H_{k_\theta}^{(1)}(kR)j\rho c\underline{\tilde{v}}_{\mathrm{x,n}}(k_\theta,\omega)\Big\}, \qquad (8)\end{aligned}$$

$$D_R(k_\theta,\omega) = H_{k_\theta}^{(1)}(kR)H_{k_\theta}^{(2)'}(kR) - H_{k_\theta}^{(2)}(kR)H_{k_\theta}^{(1)'}(kR). \qquad (9)$$

$H_{k_\theta}^{(\cdot)'}$ denotes the derivative of the respective Hankel function with the angular wave number k_θ, and $k = \omega/c$ as in Sect. 2. Underlined quantities with a tilde denote spatio-temporal frequency components, e.g.,

$$\underline{\tilde{p}}_{\mathrm{x}}(k_\theta,\omega) = \frac{1}{2\pi}\int_0^{2\pi} \underline{p}_{\mathrm{x}}(\theta,\omega)e^{-jk_\theta\theta}d\theta. \qquad (10)$$

Analogously to (7) and (8) the plane wave components $\underline{\tilde{y}}^{(1)}(k_\theta,\omega)$ and $\underline{\tilde{y}}^{(2)}(k_\theta,\omega)$ of the recorded signals in the receiving room are obtained by transform $\mathcal{T}_2$ with

$$\begin{aligned}\underline{\tilde{y}}^{(1)}(k_\theta,\omega) &= \frac{j^{1-k_\theta}}{D_R(k_\theta,\omega)}\Big\{H_{k_\theta}^{(2)'}(kR)\underline{\tilde{p}}_{\mathrm{y}}(k_\theta,\omega)\\ &\qquad -H_{k_\theta}^{(2)}(kR)j\rho c\underline{\tilde{v}}_{\mathrm{y,n}}(k_\theta,\omega)\Big\}, \quad (11)\\ \underline{\tilde{y}}^{(2)}(k_\theta,\omega) &= \frac{-j^{1+k_\theta}}{D_R(k_\theta,\omega)}\Big\{H_{k_\theta}^{(1)'}(kR)\underline{\tilde{p}}_{\mathrm{y}}(k_\theta,\omega)\\ &\qquad -H_{k_\theta}^{(1)}(kR)j\rho c\underline{\tilde{v}}_{\mathrm{y,n}}(k_\theta,\omega)\Big\} \quad (12)\end{aligned}$$

from $\underline{\tilde{p}}_{\mathrm{y}}(k_\theta,\omega)$ and $\underline{\tilde{v}}_{\mathrm{y,n}}(k_\theta,\omega)$ using the pressure and pressure gradient microphone elements. On the loudspeaker side an additional spatial extrapolation assuming free field propagation of each loudspeaker signal to the microphone positions is necessary within $\mathcal{T}_1$ prior to using (7) and (8) in order to obtain p_{x} and $v_{\mathrm{x,n}}$ of the incident waves on the microphone positions.

Adaptive filtering is then carried out for each spatio-temporal frequency bin. Note that conventional single-channel FDAF algorithms realizing FIR filtering can directly be applied to each subfilter in Fig. 5. These sub-filters already contain the temporal part of the transformation into the spatio-temporal frequency domain. In practice, both, the spatial transformation, and the temporal transformation are realized by DFTs. However, while in the temporal component, we have to ensure linear convolutions by certain constraints within FDAF [13, 16], this is not necessary for the spatial (angular) component, as it is inherently circulant.

Since the plane wave representation after the AEC is independent of the array geometries, the plane wave components $\underline{\tilde{e}}^{(\cdot)}(k_\theta,\omega) = \underline{\tilde{y}}^{(\cdot)}(k_\theta,\omega) - \hat{\underline{\tilde{y}}}^{(\cdot)}(k_\theta,\omega)$ can either be sent to the far end directly, or they can be used to synthesize the total spatio-temporal wave field using an extrapolation $\mathcal{T}_3$ of the wave field [10]

$$\begin{aligned}\underline{p}_{\mathrm{e}}^{(1)}(r,\theta,\omega) &= \int_0^{2\pi} \underline{e}^{(1)}(\theta',\omega)e^{-jkr\cos(\theta-\theta')}d\theta',\\ \underline{p}_{\mathrm{e}}^{(2)}(r,\theta,\omega) &= \int_0^{2\pi} \underline{e}^{(2)}(\theta',\omega)e^{jkr\cos(\theta-\theta')}d\theta'\end{aligned}$$

which corresponds to inverse spatial Fourier transforms in polar coordinates. Due to the independence from the array geometries, the plane-wave representation is very suitable for direct transmission. Moreover, application of linear prediction techniques on this representation is attractive for source coding of acoustic wavefields.

5. SYSTEM INTEGRATION

As in Sect. 3.2, we now study how to integrate the proposed AEC into a multichannel acoustic human-machine interface. In contrast to the conventional structure in Fig. 3, the WDAF-based AEC can now be applied after auralization and ARC. Moreover, the concept of WDAF can also efficiently be applied to ARC, as shown in [6]. Fig. 6 shows the structure of the WDAF-based ARC. It can easily be verified that the adaptations of ARC and AEC in the integrated solution after Fig. 7 are then mutually fully separable from each other, so that there are no repercussions between them.

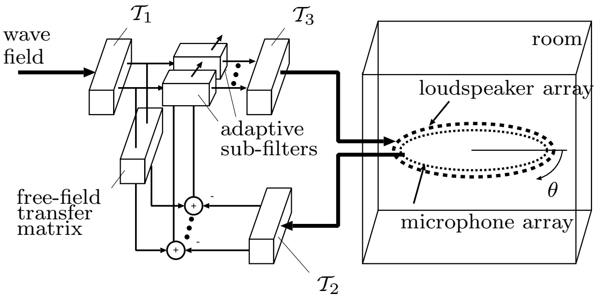

Fig. 6: WDAF-based ARC after [6].

The new WDAF structure offers another interesting aspect: since the plane wave decomposition can be interpreted as a special set of spatial filters, the set of beamformers for acquisition (as in Fig. 3) is inherently integrated in a natural way. Thus, the spatial filter **B** and the transformation $\mathcal{T}_2$ may be simply merged, and could be implemented as a masking in the Θ-domain. 'Voting', as in the conventional setup in Sect. 3.2 is obtained by additional time-varying weighting of the (already available) spatial components.

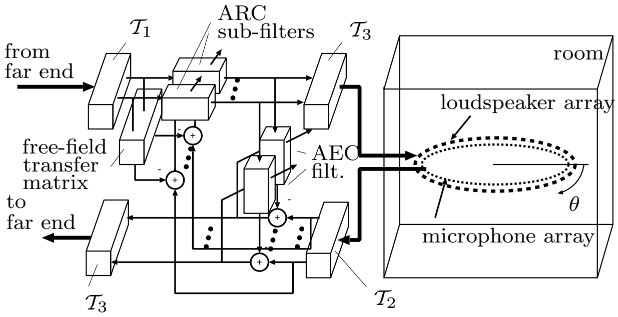

Fig. 7: Integrated system in the wave domain.

6. EVALUATION OF THE AEC

We verify the proposed concept for the AEC application. For the simulations using measured data from a real room, we used two concentric circular arrays of 48 loudspeakers and 48 microphones, respectively (the recording was done by one rotating sound field microphone mounted on a stepper motor), as shown in Fig. 8. The radius of the loudspeaker array is 142cm (spacing 19cm), and the radius of the microphone array is 75cm (spacing 9.8cm). The reverberation time T_{60} of the room is approximately 500ms. A virtual point source (music signal) was placed by WFS at 3m distance from the array center. All signals were downsampled to the aliasing frequency of the microphone array of $f_{\mathrm{al}} \approx 1.7$kHz (as discussed in Sect. 2). For the adaptation of the parameters, wavenumber-selective FDAF algorithms (filter length 1024 each) with an overlap factor 256 after [13] were applied. Figure 9 shows the so-called echo return loss enhancement ($ERLE$), i.e., the attenuation of the echoes (note that the usual fluctuations in any $ERLE$ curve result from the source signal statistics as $ERLE$ is a signal-dependent measure.). While conventional AEC techniques cannot be applied in this case ($48 \times 48 = 2304$ filters would have to be adapted, giving a total of 2359296 FIR filter taps for this extremely ill-conditioned least-squares problem), the WDAF approach allows stable adaptation and sufficient attenuation levels. The convergence speed is well comparable to conventional single-channel AECs. However, a high overlap factor for FDAF is necessary due to the low sampling rate [5] (note that efficient realizations exploiting high overlap factors exist [13]). In [5] it is shown that the performance of WDAF-based AEC is also relatively robust against time-varying scenarios in the transmission room. This robustness is a very important indicator of the quality of the estimated room parameters [9].

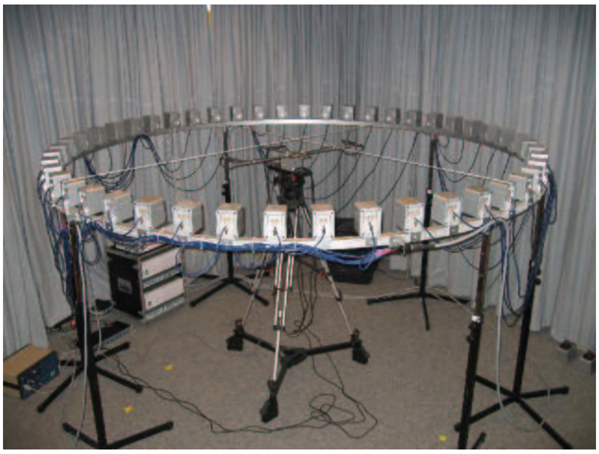

Fig. 8: Setup for measurements.

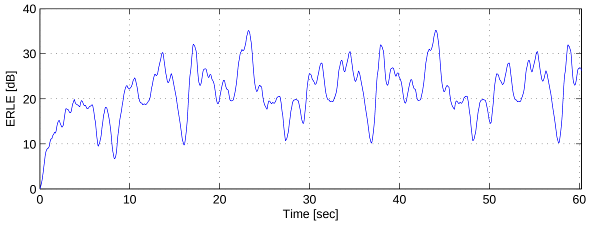

Fig. 9: ERLE convergence of WDAF-based 48×48-channel AEC.

7. CONCLUSIONS

A novel concept for efficient adaptive MIMO filtering in the wave domain has been proposed in the context of acoustic human-machine interfaces based on wavefield analysis and synthesis using loudspeaker arrays and microphone arrays. The illustration by means of acoustic echo cancellation shows promising results.

8. REFERENCES

[1] A.J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *Journal of the Acoustic Society of America*, vol. 93, no. 5, pp. 2764–2778, May 1993.

[2] M.S. Brandstein and D.B. Ward, *Microphone Arrays,* Springer, 2001.

[3] H.F. Silverman, W.R. Patterson, J.L. Flanagan, and D. Rabinkin, "A digital system for source location and sound capture by large microphone arrays," in *Proc. IEEE ICASSP*, 1997.

[4] H. Buchner, S. Spors, W. Kellermann, and R. Rabenstein, "Full-Duplex Communication Systems with Loudspeaker Arrays and Microphone Arrays," *Proc. IEEE Int. Conference on Multimedia and Expo (ICME),* Lausanne, Switzerland, Aug. 2002.

[5] H. Buchner, S. Spors, and W. Kellermann, "Wave-domain adaptive filtering: Acoustic echo cancellation for full-duplex systems based on wave-field synthesis," in *Proc. IEEE ICASSP*, 2004.

[6] S. Spors, H. Buchner, and R. Rabenstein,"An Efficient Approach to Active Listening Room Compensation for Wave Field Synthesis," *116th Convention of the Audio Engineering Society (AES)*, May 2004.

[7] M. M. Sondhi and D. R. Morgan, "Stereophonic Acoustic Echo Cancellation - An Overview of the Fundamental Problem," *IEEE SP Lett.*, Vol.2, No.8, Aug. 1995, pp. 148-151.

[8] S. Shimauchi and S. Makino, "Stereo Projection Echo Canceller with True Echo Path Estimation," in *Proc. IEEE ICASSP*, 1995, pp. 3059-3062.

[9] J. Benesty, D.R. Morgan, and M.M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Trans. on Speech and Audio Processing*, vol. 6, no.2, March 1998.

[10] A.J. Berkhout, *Applied Seismic Wave Theory,* Elsevier, 1987.

[11] V. Fischer et al., "A Software Stereo Acoustic Echo Canceler for Microsoft Windows," in *Proc. IWAENC*, Darmstadt, Germany, pp. 87-90, Sept. 2001.

[12] H. Buchner, W. Herbordt, and W. Kellermann, "An Efficient Combination of Multichannel Acoustic Echo Cancellation With a Beamforming Microphone Array," in *Proc. Int. Workshop on Hands-Free Speech Communication*, Kyoto, Japan, pp. 55-58, April 2001.

[13] H. Buchner, J. Benesty, and W. Kellermann, "Multichannel Frequency-Domain Adaptive Algorithms with Application to Acoustic Echo Cancellation," in J.Benesty and Y.Huang (eds.), *Adaptive signal processing: Application to real-world problems*, Springer-Verlag, Berlin/Heidelberg, Jan. 2003.

[14] S. Haykin, *Adaptive Filter Theory*, 3rd ed., Prentice Hall Inc., Englewood Cliffs, NJ, 1996.

[15] S.S. Narayan, A.M. Peterson, and M.J. Narasimha, "Transform Domain LMS Algorithm," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-31, no.3, June 1983.

[16] J.J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE SP Magazine,* pp. 14-37, Jan. 1992

[17] J. Benesty, A. Gilloire, and Y. Grenier, "A frequency-domain stereophonic acoustic echo canceler exploiting the coherence between the channels," *J. Acoust. Soc. Am.*, vol. 106, pp. L30-L35, Sept. 1999.

[18] E. Hulsebos, D. de Vries, and E. Bourdillat, "Improved microphone array configurations for auralization of sound fields by Wave Field Synthesis," *110th Convention of the Audio Engineering Society (AES)*, May 2001.