Towards Wave-Domain Adaptive Filtering for Multichannel Acoustic Echo Cancellation in Higher-Order Ambisonics Systems

Marcel Nophut, Robert Hupke, Stephan Preihs and Jürgen Peissig
Institute of Communications Technology
Leibniz University Hannover, Hannover, Germany
Email: {nophut,hupke,preihs,peissig}@ikt.uni-hannover.de

***Abstract*—The concept of wave-domain adaptive filtering (WDAF) is a powerful technique in the field of massive-multichannel acoustic system identification. By using fundamental solutions of the wave-equation for signal representation the wave-domain system model exhibits desirable properties in suitable setups. This allows efficient approximations of the model or can improve the convergence of an adaptive filter. The applicability of the WDAF technique to higher-order Ambisonics systems and three-dimensional wave fields was often mentioned in literature, but conducted experiments were always limited to two-dimensional setups.**

This paper investigates the WDAF technique in a new context of practical Ambisonics systems, aiming at acoustic echo cancellation for immersive telepresence systems. Experiments are conducted with simulated and measured room impulse responses in ideal and non-ideal loudspeaker setups. Furthermore different Ambisonics decoding methods are investigated and compared. The transferability of previously proposed WDAF approaches is examined and the results show the general applicability of these approaches, but with some limitations for real-world problems.

***Index Terms*—Wave-domain adaptive filtering, HOA, AEC, immersive telepresence**

I. Introduction

Modern telepresence systems aim to provide the user with an immersive experience, giving the natural impression of being present in a remote environment. Such systems shall enable natural speech communication and even immersive networked music performances between remote places [1]–[3]. The mentioned immersion commonly includes the capturing and reproduction of spatial audio. A widely used approach for that is the well-known higher-order Ambisonics (HOA) technique [4], [5]. Spatial sound fields can be captured with spherical microphone arrays by transforming the microphone signals to Ambisonics signals. These signals carry the spatial information of the sound field independently from the reproduction setup. Transferred to another place, they can be decoded to almost arbitrary multichannel loudspeaker setups reproducing the original sound field. However, a symmetrical implementation of this setup, as it is needed for the mentioned applications, causes echo-loops that can significantly deteriorate the user experience. A common approach to address this problem is acoustic echo cancellation (AEC) realized through adaptive filtering (e.g. [6]). But the estimation of an multiple-input/multiple-output (MIMO) loudspeaker-enclosure-microphone system (LEMS) with correlated loudspeaker signals is a mathematically underdetermined problem ("non-uniqueness problem") [2], [6] and means a considerable computational effort for a large numbers of channels. Wave-domain adaptive filtering (WDAF) as proposed by researchers in the past [2], [7], [8] addresses these difficulties and seems to be a promising approach and a natural choice in combination with Ambisonics systems.

The concept of WDAF was originally proposed in 2004 by Buchner et al. for AEC applications in wave field synthesis systems [8]. Later it was also used for room equalization in massive multichannel loudspeaker setups [9], [10]. In this approach, fundamental solutions of the wave equation, e.g. plane waves or circular and spherical harmonics, are used as basis functions for the representation of the involved signals that are passed to the adaptive filter. Thus, the LEMS modelled by the adaptive filter does not describe point-to-point acoustic paths from each loudspeaker to each microphone, but it describes the difference between the ideal wave field emitted by the loudspeakers under free-field conditions and the true wave field measured by the microphones. In past studies, the LEMS in wave-domain – in contrast to the LEMS in the original (i.e. point-to-point) domain – showed desirable properties such as few dominant couplings in a rather sparse coupling matrix (e.g. on the main diagonal) [2], [7], [11]. To exploit these properties, different approaches have been proposed, such as using approximated models to reduce computational complexity [7] or introducing additional constraints to increase robustness against non-uniqueness [2]. The applicability of the WDAF technique to higher-order Ambisonics systems and three-dimensional wave fields was often mentioned in literature, but the conducted experiments were always limited to two-dimensional setups [2], [7], [8], [11], [12]. This paper extends the WDAF technique to three-dimensional sound fields and practical Ambisonics setups using transforms based on spherical harmonics.

II. Wave-Domain model for Ambisonics systems

A conventional MIMO LEMS model considering point-to-point couplings from loudspeakers to microphones can be

This research was funded by the German Federal Ministry of Economic Affairs and Energy under the Grant Number 01MD18010G.

expressed as

$$P_{\mu}^{(d)}(j\omega) = \sum_{\lambda=0}^{N_{\lambda}-1} P_{\lambda}^{(x)}(j\omega) H_{\lambda,\mu}(j\omega) \tag{1}$$

where $P_{\lambda}^{(x)}(j\omega)$ ($\lambda = 1, ..., N_{\lambda}$) and $P_{\mu}^{(d)}(j\omega)$ ($\mu = 1, ..., N_{\mu}$) denote the spectra of N_{λ} loudspeaker and N_{μ} microphone signals, respectively, and $H_{\lambda,\mu}(j\omega)$ contains the transfer functions modelling the point-to-point MIMO LEMS. Alternatively, a wave-domain model in Ambisonics-domain describes the couplings from the spectra of N_p Ambisonics signals at the input $\tilde{P}_{p}^{(x)}(j\omega)$ to the spectra of N_q Ambisonics signals at the output $\tilde{P}_{q}^{(d)}(j\omega)$, which can be expressed as

$$\tilde{P}_{q}^{(d)}(j\omega) = \sum_{p=0}^{N_p-1} \tilde{P}_{p}^{(x)}(j\omega) \tilde{H}_{p,q}(j\omega) \tag{2}$$

with $\tilde{H}_{p,q}(j\omega)$ containing the transfer functions describing the couplings. The indices $p = 0, ..., N_p - 1$ and $q = 0, ..., N_q - 1$ denote the Ambisonics channel number ACN (cf. section III-A). The properties of the wave-domain model will be discussed in the following sections.

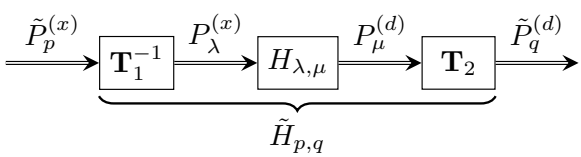

Fig. 1. Block diagram of transforms.

In related literature, the transforms to obtain the wave-domain signal representations from the loudspeaker and microphone signals are commonly denoted $\mathbf{T}_1$ and $\mathbf{T}_2$, respectively [8], [10]. It is given from the application of an immersive telepresence system based on HOA that the signals are passed from one site to the other in wave-domain representation as Ambisonics signals. Therefore, we need the inverse of $\mathbf{T}_1$, i.e. $\mathbf{T}_1^{-1}$, to obtain the loudspeaker signals from the received Ambisonics signals and $\mathbf{T}_2$ to transform the microphone signals to Ambisonics-domain (cf. Fig. 1). These transforms correspond to the Ambisonics decoder for the present loudspeaker setup ($\mathbf{T}_1^{-1}$) and the spherical microphone array to Ambisonics converter ($\mathbf{T}_2$). Since these are already part of the spatial audio encoding and decoding, WDAF does not need any additional transforms in this setup.

III. Transforms for WDAF in Ambisonics Domain

A. Sound field representation in spherical coordinates

The Ambisonics technique uses a sound field representation in spherical coordinates $(r, \boldsymbol{\theta})$, with r denoting the radius and $\boldsymbol{\theta} = (\vartheta, \varphi)$ denoting polar and azimuth angle. This appears to be a natural choice for a spherical and concentrical setup. In spherical coordinates, the sound field quantities (sound pressure and particle velocity) are expressed in terms of spherical Bessel and Hankel functions and spherical harmonics (SH). A detailed derivation of the following and the related fundamentals can be found in [13]–[15]. The mentioned SH $Y_n^m(\boldsymbol{\theta})$ are defined as

$$Y_n^m(\boldsymbol{\theta}) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\vartheta) e^{im\varphi} \tag{3}$$

where n and m denote function order and function degree, respectively, and $P_n^m(\cdot)$ are the associated Legendre functions [14]. However, in the context of Ambisonics the SH Y_n^m are commonly defined as a real-valued quantity as

$$Y_n^m(\boldsymbol{\theta}) = N_n^{|m|} P_n^{|m|}(\cos\vartheta) \begin{cases} \sin(|m|\varphi), & \text{for } m < 0, \\ \cos(|m|\varphi), & \text{for } m \geq 0. \end{cases} \tag{4}$$

According to the ambiX-Format [16] the SN3D normalization $N_n^{|m|} = \sqrt{\frac{2-\delta_m}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}}$ is used and the channels corresponding to the SH are ordered by the Ambisonics channel number $\mathrm{ACN} = n^2 + n + m$. This format is used in the following.

B. Spherical array to Ambisonics encoding

For deriving $\mathbf{T}_2$ that transforms the array's microphone signals to the wave-domain representation we describe the sound field around that microphone array by means of the SH. A rigid spherical microphone array works as a spherical scatterer in the sound field. Thus, the resulting spectrum of the sound pressure $P(r, \boldsymbol{\theta}, k)$, evaluated at radius r and the spatial frequency $k = \frac{\omega}{c}$ with c being the speed of sound, is a superposition of the incident and scattered sound waves. The rigid sphere of radius R_M implies another boundary condition of zero radial particle velocity at the microphone's surface ($u_r(R_M, \boldsymbol{\theta}, k) = 0$) [14]. Considering this, the spectrum of the sound pressure at the surface of the microphone array can be described as

$$P(R_M, \boldsymbol{\theta}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \Psi_{nm}(R_M, k) Y_n^m(\boldsymbol{\theta}) \tag{5}$$

with

$$\Psi_{nm}(R_M, k) = \frac{4\pi \mathrm{i}^{n+1}}{(kR_M)^2 h_n^{(2)\prime}(kR_M)} \tilde{P}_{nm}(k), \tag{6}$$

where $h_n^{(2)\prime}(kR_M)$ denotes the first derivative of the spherical Hankel function of the second kind with respect to (kR_M). The desired wave-domain representation $\tilde{P}_{nm}$ are the Ambisonics signals or, in other words, the spherical harmonic coefficients of a plane-wave amplitude density [14] assuming that the sound field consists of a sum of incident plane waves. In the following, we use $\tilde{P}_q^{(d)}$ as symbol for the wave-domain representation of the microphone signals by converting the indices n and m to the Ambisonics channel number q.

With a finite number of microphone capsules N_{μ} on the spherical surface the spherical harmonic coefficients can be obtained up to a certain Ambisonics order N. For $N_{\mu} > (N+1)^2$, Eq. (5) can be inverted and be written in matrix notation as

$$\boldsymbol{\Psi}_N(k) = (\boldsymbol{Y}_N^T)^{\dagger} \boldsymbol{P}^{(d)}(k) \tag{7}$$

where $(\cdot)^\dagger$ denotes the pseudo-inverse, $\boldsymbol{Y}_N^T = [\boldsymbol{y}_N(\boldsymbol{\theta}_\mu)^T]_{\mu,q}$ with $\boldsymbol{y}_N(\boldsymbol{\theta}_\mu) = \left[Y_0^0(\boldsymbol{\theta}_\mu), Y_1^{-1}(\boldsymbol{\theta}_\mu), ..., Y_N^N(\boldsymbol{\theta}_\mu)\right]^T$ and the column vectors $\boldsymbol{P}^{(d)}(k) = [P_\mu^{(d)}(k)]_\mu$ and $\boldsymbol{\Psi}_N(k) = [\Psi_q(k)]_q$ [15]. To obtain $\tilde{P}_q^{(d)}(k)$ we take the inverse of (6). Since this inverse exhibits a n-fold pole at 0 Hz a Tikhonov regularization is applied [5].

C. Mode-matching Ambisonics decoder

The idea of the *mode-matching decoder* (MMD) approach [13] is to synthesize a plane wave from an arbitrary direction using N_λ plane wave sources assuming the N_λ loudspeakers are at a sufficiently large distance from the listener. By solving the corresponding matching equation and for $N_L \geq (N+1)^2$ the decoding matrix is obtained as

$$\boldsymbol{D} = \sqrt{\frac{N_L}{4\pi}} \boldsymbol{Y}_N^T (\boldsymbol{Y}_N \boldsymbol{Y}_N^T)^{-1} \tag{8}$$

with $\boldsymbol{Y}_N = [\boldsymbol{y}_N(\boldsymbol{\theta}_1), ..., \boldsymbol{y}_N(\boldsymbol{\theta}_{N_L})]$. The loudspeaker signals can then be obtained as

$$\boldsymbol{P}^{(x)}(k) = \boldsymbol{D}\tilde{\boldsymbol{P}}^{(x)}(k). \tag{9}$$

This decoding method is non-optimum for irregular loudspeaker layouts since it produces strong loudness increases for poorly sampled panning directions.

D. Energy-preserving Ambisonics decoder

The *energy-preserving Ambisonics decoder* (EPAD) [17] is a more perceptually-motivated decoding method and it is designed to provide a panning-invariant loudness for non-uniform loudspeaker layouts. The singular value decomposition of

$$\boldsymbol{Y}_N^T = \boldsymbol{U}[\mathrm{diag}(\boldsymbol{s}), \boldsymbol{0}]^T \boldsymbol{V}^T \tag{10}$$

is used to obtain the decoding matrix

$$\boldsymbol{D} = \boldsymbol{U}[\boldsymbol{I}, \boldsymbol{0}]^T \boldsymbol{V}^T. \tag{11}$$

IV. Experiments

The considered setup consists of an "Eigenmike® em32", a spherical microphone array with 32 microphone capsules on a rigid sphere, and two different spherical loudspeaker layouts in concentric setups. The first layout is an ideal spherical setup of 20 uniformly distributed loudspeakers (i.e. platonic icosahedron) "*Uni20*" and the second layout is a practically motivated 32 loudspeaker layout of the Immersive Media Lab (IML) at our institute "*IML32*", which is depicted in Fig. 2. The speakers of "*IML32*" are not arranged on a sphere, but the channels are equalized regarding broadband gain and delay, and thus the setup can be considered a practically spherical loudspeaker array. Experiments are conducted with LEMS transfer functions obtained from a simulated anechoic environment "*SimAnechoic*" as well as from measurements at the IML "*MeasIML*". On the loudspeaker side two different decoding methods, the theoretically motivated MMD [13] and the more perceptually motivated EPAD [17], are investigated.

For evaluating the properties of the wave-domain representations the transfer functions of the wave-domain (WD) model $\tilde{H}_{p,q}(j\omega)$ were computed from the transfer functions of the point-to-point (PTP) model $H_{\lambda,\mu}(j\omega)$ using the transforms described above. This was computed both for the MMD (WD-MMD) and for the EPAD (WD-EPAD). The energy of each coupling $E_{i,j}(l)$ was computed as the squared magnitude of the l-th DFT bin ($E_{\lambda,\mu}(l) = |H_{\lambda,\mu}(l)|^2$ and $E_{p,q}(l) = |\tilde{H}_{p,q}(l)|^2$). For the frequency dependent evaluation (i.e. $(E_{i,j})_{3rdOct}$) the coupling energy over all DFT-bins of one-third octave bands was averaged and for the broadband evaluation the total energy of the couplings $(E_{i,j})_{tot}$ was computed as the sum of all DFT-bins up to 9 kHz, i.e. the upper frequency boundary for spherical harmonic decomposition with the Eigenmike without spatial aliasing [18]. All energies are illustrated in logarithmic scale. The Ambisonics order at input and output was set to $N = 3$, which determines the number of channels to be $N_p = N_q = 16$, assuming this reduction of order below the maximum possible ($N = 4$) means only a minor degradation in spatial resolution.

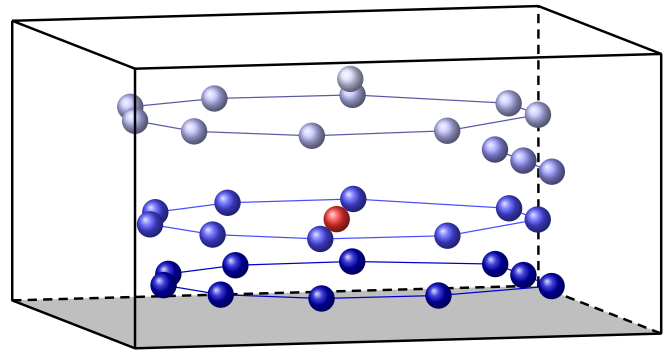

Fig. 2. Loudspeaker layout "*IML32*". For illustration purposes the speakers at different height layers are individually colored and interconnected, the position of the Eigenmike is depicted in red.

Figure 3 shows the coupling energies between the channels in the regular and in the wave domain for the "Uni20" loudspeaker layout in the "SimAnechoic" environment. The two decoder methods show a very similar behavior as it was expected for a uniform layout in an ideal anechoic environment. The attenuation of higher orders at lower frequencies is also an expected behavior of the encoder caused by the regularization. Most coupling energy lays on the main diagonal which allows an efficient approximation in the sense of former WDAF proposals [7].

Figure 4 shows the coupling energies for the "IML32" loudspeaker layout in the "SimAnechoic" environment. Here, the WD-EPAD exhibits more side couplings (i.e. coupling energy beside the main diagonal), even at lower frequencies, whereas both decoder methods produce side couplings at higher frequencies, which also affects the broadband behavior. An approximation of this model is possible, but with the limitation of using a more complex coupling pattern or accepting a less efficient approximation. However, the significant couplings on the main diagonal still may allow an improved adaptive filtering considering this property as prior knowledge and accounting for it with an additional constraint [2].

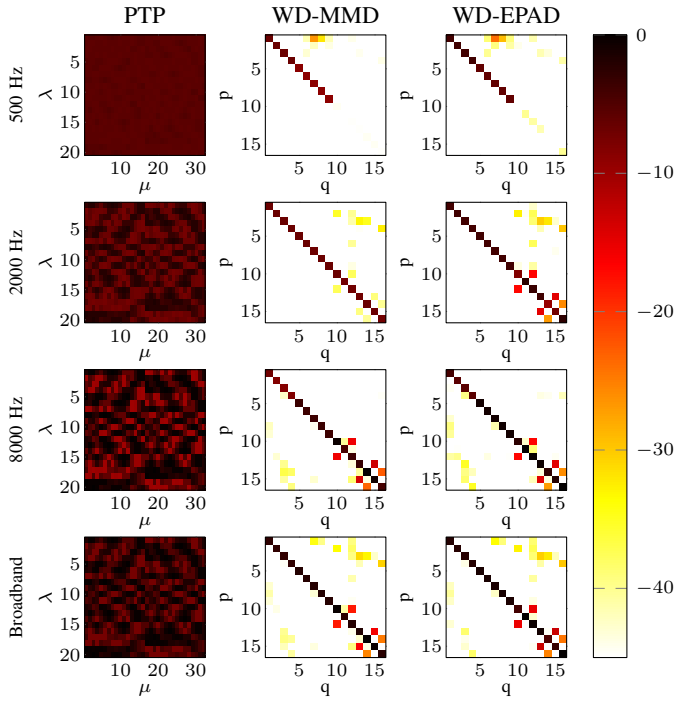

Fig. 3. Coupling energy matrices of frequency-dependent $(E_{i,j})_{3rdOct}$ and broadband $(E_{i,j})_{tot}$ energy (in dB) for layout "Uni20" and environment "SimAnechoic".

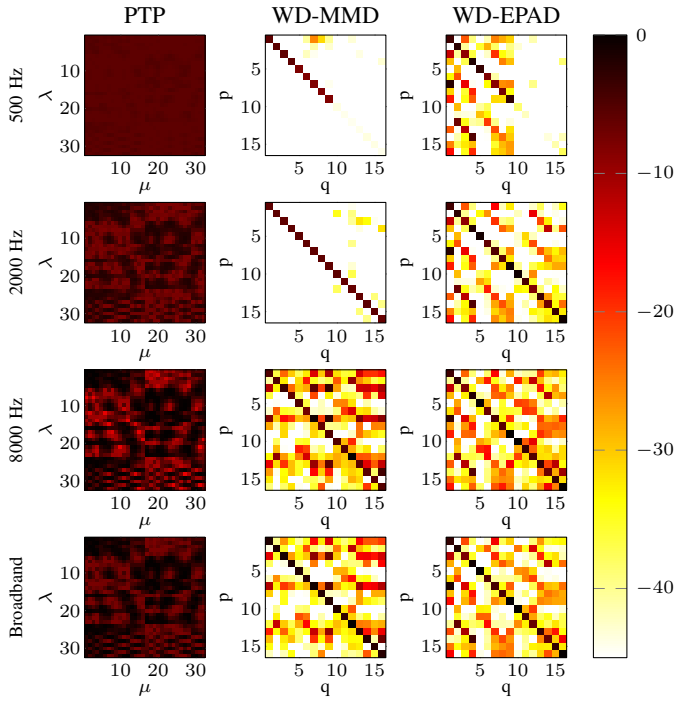

Fig. 4. Coupling energy matrices of frequency-dependent $(E_{i,j})_{3rdOct}$ and broadband $(E_{i,j})_{tot}$ energy (in dB) for layout "IML32" and environment "SimAnechoic".

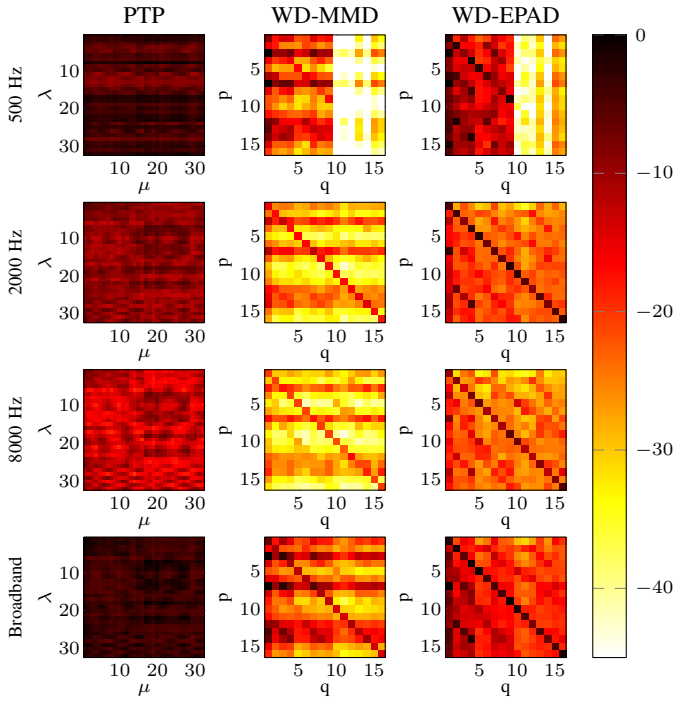

Fig. 5. Coupling energy matrices of frequency-dependent $(E_{i,j})_{3rdOct}$ and broadband $(E_{i,j})_{tot}$ energy (in dB) for layout "IML32" and environment "MeasIML".

Figure 5 shows the coupling energies for the "IML32" loudspeaker layout in the "MeasIML" environment, i.e. the most realistic case considered here. For both decoder methods the coupling matrix is not sparse anymore. The WD-MMD exhibits significant side couplings as horizontal lines, that are even more dominant than the main diagonal. The WD-EPAD exhibits dominant main diagonal couplings, but the energy of the side couplings are considerable and not negligible. So the model cannot be efficiently approximated considering only a few couplings (c.f. Fig. 6). However the property of the dominant main diagonal might still be exploited in the sense of a prior knowledge.

Figure 6 shows the approximation error relative to the total coupling energy in the case the wave-domain LEMS is only modelled by the n most dominant couplings of the broadband coupling matrix calculated as

$$e_{approx}(n) = 1 - \frac{\sum \max_{p,q}(E_{p,q}; n)}{\sum_{p,q} E_{p,q}} \tag{12}$$

where $\max_{p,q}(E_{p,q}; n)$ are the n most dominant elements of $E_{p,q}$.

It is obvious that it is much easier to approximate the wave-domain LEMS in the anechoic case with less channel couplings considered. However, in the measured real-world scenario (i.e. "IML32" and "MeasIML") 90 % of the coupling energy (error of $-10\,\mathrm{dB}$) could be modeled with 78 and 103 couplings for MMD and EPAD, respectively. The complete WD model comprises 256 couplings. But the most dominant couplings would have to be known in advance, which would

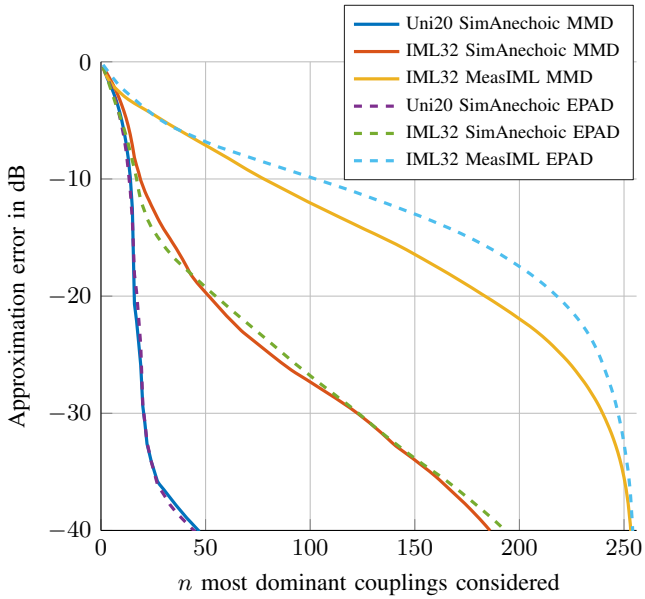

Fig. 6. Approximation Error (relative to total coupling energy) vs. the n most dominant couplings considered for approximation (broadband case).

be possible to track even for changing acoustic environments, but it would mean additional computational load. The WD-EPAD model seems to be harder to approximate in this case, but it could still be preferable (in the sense of [2]) due to its more regular structure of the coupling matrix (c.f. Fig. 5). Moreover, EPAD may be a preferable choice for Ambisonics reproduction in non-uniform loudspeaker layouts anyway.

The identified sources of error are various. The non-uniform speaker layout added many side couplings for both decoding methods and also the room reflections with the measured data had a considerable effect. The effect of the regularization method, other common techniques as max-r_E, more decoding methods or the effect of non-concentric positioning of the microphone were not evaluated and are subject of future research.

V. Conclusion and Future Work

This paper extended the WDAF technique to a three-dimensional wave-field and presented a practical integration into a bi-directional Ambisonics setup. Chances and limitations of the applicability of previously proposed WDAF approaches and error sources were identified. Our experiments show the EPAD decoder could be a suitable choice for the usage in WDAF-AEC with a previously proposed WDAF approach exploiting prior knowledge [2]. Our future research will investigate the performance of the technique in an actual AEC application, the influence of spatial aliasing above 9 kHz and a more perceptual view on the couplings which could be weighted depending on the perceived importance. Furthermore, an even more practical setup with an eccentrical microphone position could be investigated.

References

[1] H. Khalilian, I. V. Bajic, and R. G. Vaughan, "A glimpse of 3d acoustics for immersive communication," in *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. Piscataway, NJ: IEEE, 2016, pp. 1–4.

[2] M. Schneider and W. Kellermann, "Multichannel acoustic echo cancellation in the wave domain with increased robustness to nonuniqueness," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 3, pp. 518–529, 2016.

[3] M. Nophut, R. Hupke, S. Preihs, and J. Peissig, "Multichannel acoustic echo cancellation for ambisonics-based immersive distributed performances," in *Audio Engineering Society Convention 148*, Audio Engineering Society, Ed., 2020. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=20798

[4] J. Daniel, S. Moreau, and R. Nicol, "Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging," in *Audio Engineering Society Convention 114*, Audio Engineering Society, Ed., 2003.

[5] S. Moreau, J. Daniel, and S. Bertet, "3d sound field recording with higher order ambisonics - objective measurements and validation of spherical microphone," in *Audio Engineering Society Convention 120*, Audio Engineering Society, Ed., 2006. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=13661

[6] 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 Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 156–165, 1998.

[7] M. Schneider and W. Kellermann, "A wave-domain model for acoustic mimo systems with reduced complexity," in *Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA), 2011*. Piscataway, NJ: IEEE, 2011, pp. 133–138.

[8] H. Buchner, S. Spors, and W. Kellermann, "Wave-domain adaptive filtering: acoustic echo cancellation for full-duplex systems based on wave-field synthesis," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Piscataway, N.J: IEEE, 2004, pp. iv–117–iv–120.

[9] S. Spors, H. Buchner, R. Rabenstein, and W. Herbordt, "Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering," *The Journal of the Acoustical Society of America*, vol. 122, no. 1, pp. 354–369, 2007.

[10] M. Schneider and W. Kellermann, "Adaptive listening room equalization using a scalable filtering structure in thewave domain," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012*. Piscataway, NJ: IEEE, 2012, pp. 13–16.

[11] S. Emura, Y. Hiwasaki, and H. Ohmuro, "Wave-domain echo-path model with aliasing for echo cancellation," in *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. Piscataway, NJ: IEEE, 2013, pp. 1–4.

[12] M. Schneider and W. Kellermann, "A direct derivation of transforms for wave-domain adaptive filtering based on circular harmonics," in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO), 2012*. Piscataway, NJ: IEEE, 2012.

[13] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc*, vol. 53, no. 11, pp. 1004–1025, 2005. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=13396

[14] B. Rafaely, *Fundamentals of Spherical Array Processing*. Cham: Springer International Publishing, 2019, vol. 16.

[15] F. Zotter and M. Frank, *Ambisonics*. Cham: Springer International Publishing, 2019, vol. 19.

[16] C. Nachbar, F. Zotter, E. Deleflie, and A. Sontacchi, "ambix - a suggested ambisonics format," *AMBISONICS SYMPOSIUM 2011*, 2011.

[17] F. Zotter, H. Pomberger, and M. Noisternig, "Energy-preserving ambisonic decoding," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 37–47, 2012.

[18] "Beamformer datasheet version 2 rev. a: Specification for eigenmike software beamformer." [Online]. Available: www.mhacoustics.com