WAVE-DOMAIN LOUDSPEAKER SIGNAL DECORRELATION FOR SYSTEM IDENTIFICATION IN MULTICHANNEL AUDIO REPRODUCTION SCENARIOS

Martin Schneider, Christian Huemmer, and Walter Kellermann

Multimedia Communications and Signal Processing*
University of Erlangen-Nuremberg
Cauerstr. 7, 91058 Erlangen, Germany
mail: {schneider,huemmer,wk}@LNT.de

ABSTRACT

For applications like acoustic echo cancellation (AEC) or listening room equalization (LRE), a loudspeaker-enclosure-microphone system (LEMS) must be identified. When using a large number of reproduction channels, as, e. g., for wave field synthesis (WFS) or Higher-Order Ambisonics (HOA), the strong correlation of the loudspeaker signals will hamper a unique identification. A state-of-the-art remedy against this so-called nonuniqueness problem is a decorrelation of the loudspeaker signals, which facilitates a unique identification. However, most of the known approaches are not suitable for acoustic wave field reproduction schemes, as they would distort the reproduced wave field in an uncontrolled manner or degrade the audio quality. In this contribution, we propose a wave-domain time-varying filtering of the loudspeaker signals, so that the reproduced wave field is rotated within a perceptually acceptable range, while preserving its shape.

Index Terms— Nonuniqueness problem, acoustic echo cancellation, system identification, decorrelation, wave domain

1. INTRODUCTION

WFS and HOA provide a high-quality spatial impression to the listener overcoming the limitations of a sweet spot by utilizing a large number of reproduction channels [1, 2]. To obtain a more immersive user experience, WFS systems may be complemented by a spatial recording system to allow further applications, such as interactivity, or to improve the reproduction quality by means of an LRE system. The combination of the loudspeaker array, the enclosing room, and the microphone array is referred to as loudspeaker-enclosure-microphone system (LEMS) and must be identified for those applications by simultaneously observing the loudspeaker and microphone signals. In the following, we only consider the AEC as a representative task for the class of applications requiring system identification. Besides the well-known teleconferencing scenarios, this would be relevant, e. g., for interactive immersive simulators and gaming, or telecollaboration of musicians. There, the signals of a far-end party are emitted by loudspeakers in the near-end room while the near-end acoustical scene is recorded simultaneously. Consequently, the microphone signals contain a mixture of the local signals and the loudspeaker echoes. As the latter are annoying to the far-end party, they should be removed by means of an AEC, prior to transmission. It is already known from stereophonic AEC that the so-called nonuniqueness problem may occur, i. e., the typically strong cross-correlation of the loudspeaker signals may preclude a perfect system identification [3]. In that case, the result of the system identification is only one out of infinitely many solutions determined by the correlation properties of the loudspeaker signals. Thus, a change of these properties can invalidate the previously optimum solution and the behavior of systems relying on adaptive filters may in fact become uncontrollable [3, 4]. To increase robustness under such conditions, the loudspeaker signals are decorrelated such that the true LEMS can be uniquely identified. To this end, different options are already known: adding mutually independent noise signals to the loudspeaker signals [5, 6, 7] or different preprocessing for each loudspeaker channel, such as nonlinear preprocessing [3, 8], time-varying filtering [9, 10] or resampling [11, 12]. However, massive multichannel reproduction with WFS and HOA has to meet stricter requirements than conventional reproduction. The addition of noise to the loudspeaker signals is probably unacceptable for the listeners, as they expect a very high reproduction quality. The same holds for nonlinear preprocessing which was essentially proposed for speech and not for music. Different to [9], the phase-modulation in [10] already aims at high-quality reproduction using psychoacoustics to perceptually hide phase distortion. Resampling based approaches can potentially also preserve an acceptable reproduction quality. However, even the approaches with minimal quality degradation are still not suited for WFS or HOA. This is due to the fact that the loudspeaker signals are analytically determined and any preprocessing involving a time-variant phase change might significantly distort the reproduced wave field. Additionally, those techniques were not proposed for a large number of reproduction channels.

With this contribution, we propose to apply a time-varying prefiltering of the loudspeaker signals as in [9, 10]. Different to the latter, we introduce an acoustic model to facilitate a controlled rotation of the reproduced wave field, which is inherently formulated for multichannel audio reproduction. Using a suitable wave-domain representation, prefilters achieving an arbitrary wave field rotation in the (horizontal) plane containing the loudspeaker array may be straightforwardly determined. This time-varying rotation angle is to be chosen in a range which still ensures acceptable perceived audio quality. Since we only need to filter the loudspeaker signals independently from the rendering process, this technique may also be applied for the reproduction of recorded loudspeaker signals and in teleconference systems obtaining only loudspeaker signals from the far-end side. It should be noted that model-based prefiltering was also proposed in [13] in the context of multichannel AEC, although the goal there was no decorrelation of the loudspeaker signals but a direct attenuation of the loudspeaker echo.

*The work of M. Schneider and W. Kellermann was performed and funded within a joint research project with the Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany.

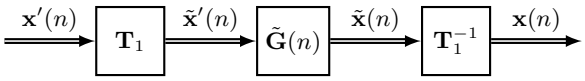

Fig. 1. Proposed prefilter structure. $\mathbf{T}_1$: transform to the wave domain, $\mathbf{T}_1^{-1}$: transform from the wave domain, $\tilde{\mathbf{G}}(n)$: wave domain prefilters

This paper is organized as follows: The acoustic model for the wave field rotation is explained in Section 2, while the implementation by digital filters is described in Section 3. The proposed approach is evaluated regarding the benefit for system identification and its influence on the perceived reproduction quality in the Section 4 and Section 5, respectively. The paper is concluded in Section 6.

2. ACOUSTICAL MODEL

In this section, the underlying acoustic model is explained. Given the loudspeaker positions and the driving signals, we may determine the wave field which would be produced under ideal free field conditions. Considering only a loudspeaker array in the horizontal plane, the sound pressure $P(\alpha, \varrho, j\omega)$ at any given point may be described in the continuous frequency domain as a function of the position in polar coordinates defined by the angle α and the radius ϱ. Equivalently, we may describe $P(\alpha, \varrho, j\omega)$ by a superposition of suitable basis functions, defined on a continuum. This representation is used by a technique called wave-domain adaptive filtering (WDAF) and is there referred to as free field description [14, 15, 16]. For the purpose considered here, the so-called circular harmonics [14] are used to describe the sound pressure in the free-field case. So we may describe the sound pressure by

$$P(\alpha, \varrho, j\omega) = \sum_{m=-\infty}^{\infty} \tilde{P}_m(j\omega) \mathcal{J}_m\left(k\varrho\right) e^{jm\alpha}, \tag{1}$$

where $\mathcal{J}_m\left(x\right)$ is the cylindrical Bessel function of first kind (defined by (10.2.1) in [17]) and order m, ω the angular frequency and $\tilde{P}_m(j\omega)$ represents the spectra of the superimposed incoming and outgoing waves with respect to the origin [16]. With c being the speed of sound and j being $\sqrt{-1}$, (1) represents the considered model of the ideally synthesized wave field. Circular harmonics were chosen, because they allow a straightforward description of the rotation of the wave field by

$$P(\alpha - \varphi(t), \varrho, j\omega) = \sum_{m=-\infty}^{\infty} \tilde{P}_m(j\omega) \mathcal{J}_m\left(k\varrho\right) e^{jm\alpha} e^{-jm\varphi(t)}, \tag{2}$$

which is just a multiplication of (1) by $e^{-jm\varphi(t)}$, where $\varphi(t)$ is the time-dependent rotation angle. For the decorrelation, we obtain the description of the loudspeaker signals according to (1), which is then rotated as described by (2). Finally, the rotated wave field is transformed back to the original domain of the loudspeaker signals.

3. WAVE FIELD ROTATION BY DIGITAL FILTERS

In this section, the realization of the proposed approach using digital filters is explained. The signal model of the prefilter is shown in Fig. 1. The block $\mathbf{T}_1$ is used to transform the loudspeaker signals $\mathbf{x}'(n)$ to a description according to (1) denoted by $\tilde{\mathbf{x}}'(n)$. The latter

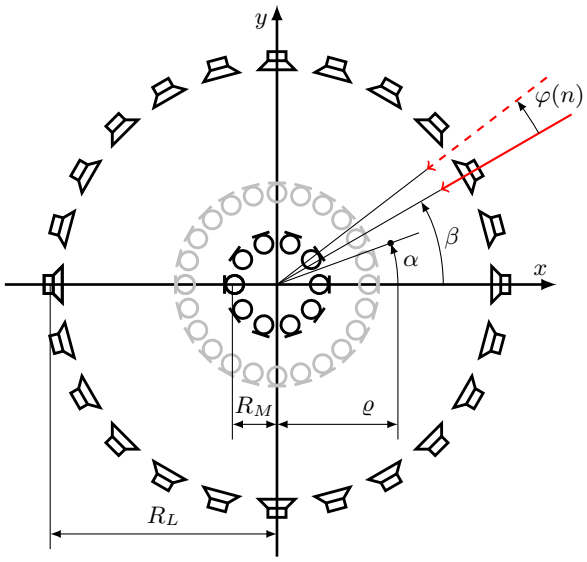

Fig. 2. Loudspeaker array, microphone array considered for the transforms (gray), and microphone array for the AEC (black)

is the discrete-time representation of $\tilde{P}_m(j\omega)$ and is rotated by $\tilde{\mathbf{G}}(n)$ from which we obtain $\tilde{\mathbf{x}}(n)$ as the rotated loudspeaker signals in the wave domain. Finally, we use $\mathbf{T}_1^{-1}$ to obtain the loudspeaker signals $\mathbf{x}(n)$ to synthesize the rotated wave field.

As the loudspeaker signal representation described by (1) is the same for WDAF with circular harmonics, we may actually use the same transforms as described in [18]. The notation of the transforms with $\mathbf{T}_1$ and $\mathbf{T}_1^{-1}$ was chosen accordingly. For the sake of brevity, we only consider a circular loudspeaker array as depicted in Fig. 2 in this work, although the transforms may be defined for an arbitrarily shaped loudspeaker array.

As the WDAF transforms aim at providing the free-field description in the vicinity of the microphone array (see [18]), we have to define a microphone array geometry prior to the definition of the transforms. As the processing involves $\mathbf{T}_1$ and its inverse $\mathbf{T}_1^{-1}$, the prefiltered loudspeaker signals do not depend on the chosen microphone array geometry except for a delay necessary to maintain causality. Consequently, the circular microphone array geometry considered for the transforms is not related to the microphone array geometry used later for AEC.

For a concise description, we define the vector $\mathbf{x}'(n)$ containing the unfiltered loudspeaker signals as follows:

$$\mathbf{x}'(n) = [\mathbf{x}_1'^T(n), \mathbf{x}_2'^T(n), \ldots, \mathbf{x}_{N_L}'^T(n)]^T, \tag{3}$$

$$\mathbf{x}_l'(n) = [x_l(nL_F - L_X + 1), \ldots, x_l(nL_F)]^T, \tag{4}$$

where $x_l(k)$ is a time-sample of the signal of loudspeaker l at time-instant k. To describe a block processing, we use L_X time-samples $x_l(k)$ for each block indexed by n, where L_F is the frame shift for the block processing. The N_L loudspeaker signals are transformed to the wave domain according to [18], using

$$\tilde{\mathbf{x}}'(n) = \mathbf{T}_1 \mathbf{x}'(n). \tag{5}$$

In general, the considered transforms can be realized using MIMO FIR filters [18]. However, for the considered special case of concentrically located uniform circular arrays for loudspeakers and microphones, we may actually use the DFT with respect to the loudspeaker

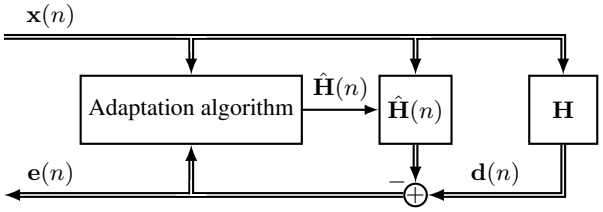

Fig. 3. Signal model of an AEC. $\mathbf{x}(n)$: loudspeaker signals, $\mathbf{d}(n)$: microphone signals, $\mathbf{e}(n)$: error signals, $\mathbf{H}$: LEMS, $\hat{\mathbf{H}}(n)$: identified LEMS.

indices l as transform $\mathbf{T}_1$ [16]. So we may write

$$\mathbf{T}_1 = \mathbf{F}_{N_L} \otimes \mathbf{I}_{L_X}, \quad \mathbf{T}_1^{-1} = \mathbf{T}_1^H, \tag{6}$$

where $\mathbf{F}_{N_L}$ is the $N_L \times N_L$ unitary DFT matrix, $\mathbf{I}_{L_X}$ is the $L_X \times L_X$ identity matrix and $\otimes$ denotes the Kronecker product.

Consequently, $\tilde{\mathbf{x}}'(n)$ shares the structure with $\mathbf{x}'(n)$, where we index the N_L components of $\tilde{\mathbf{x}}'(n)$ with m. For a rotation of $\tilde{\mathbf{x}}'(n)$ according to (2) we use

$$\tilde{\mathbf{x}}(n) = \tilde{\mathbf{G}}(n)\tilde{\mathbf{x}}'(n) \tag{7}$$

with a matrix

$$\tilde{\mathbf{G}}(n) = \begin{bmatrix} e^{-j0\varphi(n)} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & e^{-j1\varphi(n)} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & e^{-j(N_L-1)\varphi(n)} \end{bmatrix} \otimes \mathbf{I}_{L_X}, \tag{8}$$

where $\mathbf{0}$ is a matrix with zero-valued entries. The prefiltered wave-domain loudspeaker signals are then transformed back to their original domain by

$$\mathbf{x}(n) = \mathbf{T}_1^{-1}\tilde{\mathbf{x}}(n), \tag{9}$$

where $\mathbf{x}(n)$ exhibits again the same structure as $\mathbf{x}'(n)$. These signals are then fed to the loudspeakers. When considering a non-circular loudspeaker array, (6) is no longer valid and the matrices $\mathbf{T}_1$ and $\mathbf{T}_1^{-1}$ must be chosen such that they describe a convolution of the respective signals with the impulse responses realizing the transforms according to [18]. In that case, the number of considered time samples in $\mathbf{x}'(n)$, $\tilde{\mathbf{x}}'(n)$, and $\tilde{\mathbf{x}}(n)$ must be chosen accordingly.

Among many different functions which may be used for the time-variant wave-field rotation, a natural and intuitive choice seems to be a sine function as defined by

$$\varphi(n) = \varphi_\mathrm{a} \cdot \sin\left(2\pi \frac{n}{L_P}\right), \tag{10}$$

where we chose a maximum rotation angle φ_a and a period length L_P. We also investigated the use of a periodically repeated triangular function, a function proportional to $\mathrm{sign}(\varphi(n))\sqrt{|\varphi(n)|}$ and the addition of a random noise term to (10). None of the mentioned functions showed a larger benefit for the system identification than (10).

4. EVALUATION OF SYSTEM IDENTIFICATION

In this section, the benefit of the proposed approach for system identification is evaluated considering AEC as an exemplary system identification task. To this end, a transducer array setup as depicted in

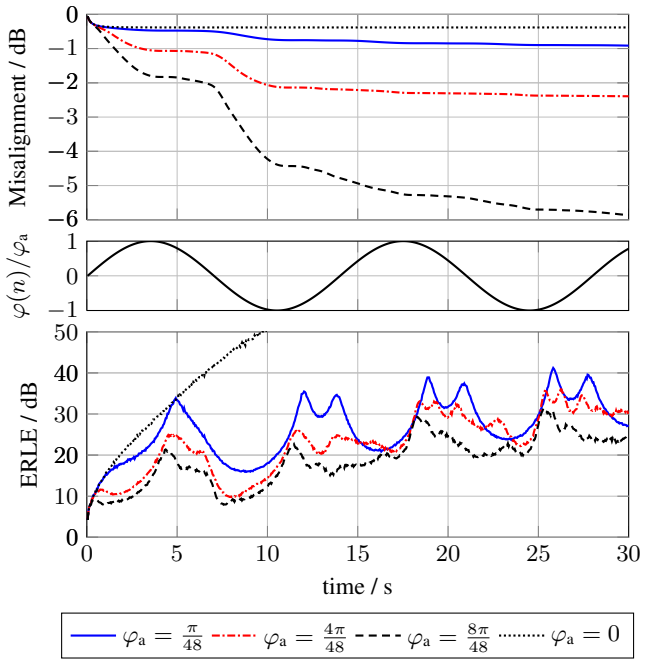

Fig. 4. Normalized misalignment and ERLE for different φ_a

Fig.2 with $N_L = 48$ loudspeakers (resulting in an angular spacing of $2\pi/48$) and $N_M = 10$ microphones was considered. The circular loudspeaker and microphone arrays had the radii $R_L = 1.5$m and $R_M = 0.05$m, respectively, and were located in a room with a reverberation time T_{60} of approximately 0.3s. The impulse responses were measured using a sampling frequency of 44.1kHz, converted to a sampling rate of 11025Hz and truncated to 1024 samples, which is also the length of the adaptive filters used for the AEC. The LEMS was simulated by convolution with these impulse responses with no noise on the microphone signal nor local sound sources within the LEMS. These ideal laboratory conditions were chosen to separate the influence of the proposed technique from other influences on the convergence of the adaptation algorithm, although other experiments with modeled near-end noise led to equivalent results.

The signal model of the AEC is shown in Fig.3. There, the prefiltered loudspeaker signals $\mathbf{x}(n)$ are fed to the LEMS $\mathbf{H}$, which is then identified by $\hat{\mathbf{H}}(n)$ based on observations of $\mathbf{x}(n)$ and the microphone signals $\mathbf{d}(n)$. The error signals $\mathbf{e}(n)$ capture the residual echo. We use the generalized frequency-domain adaptive filtering algorithm [4] for the AEC using an exponential forgetting factor $\lambda = 0.95$, a stepsize $\mu = 0.5$ (with $0 \leq \mu \leq 1$), and a frame shift $L_F = 512$. The time-varying rotation angle is described by (10) with a period length L_P of 301 blocks (ca. 14 secs), while the value for φ_a was varied.

The original loudspeaker signals are determined according to the theory of WFS [1] for synthesizing four planes waves simultaneously with the incidence angles β equal to 0, $\pi/2$, π, and $3\pi/2$. For the source signals we used mutually uncorrelated white noise signals such that all 48 loudspeakers are driven with the same average power. Although noise signals are rarely relevant in practice, this scenario allows a clear and concise evaluation of the influence of φ_a. Considering that there are only four independent signal sources, but 48 loudspeakers, the task of system identification is severely underdetermined and a high normalized system misalignment can be

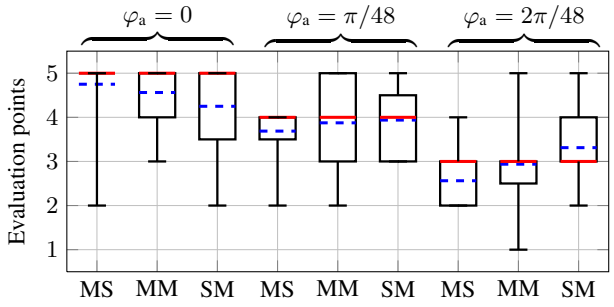

Fig. 5. Assessment of noticeability of rotation. MS=mono speech signal, MM=mono music signal, SM=stereo music signal

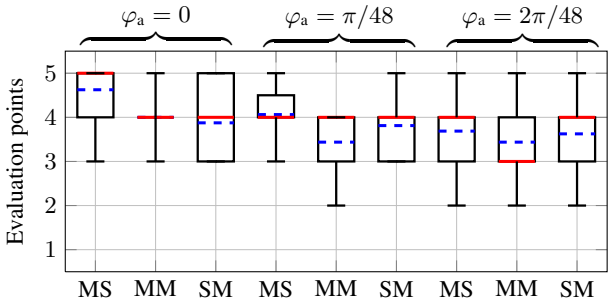

Fig. 6. Assessment of general quality, MS=mono speech signal, MM=mono music signal, SM=stereo music signal

expected. The latter is defined by

$$\Delta_h(n) = 10\log_{10}\left(\|\hat{\mathbf{H}}(n) - \mathbf{H}\|_{\mathrm{F}}^2 / \|\mathbf{H}\|_{\mathrm{F}}^2\right) \mathrm{dB}, \qquad (11)$$

where the Frobenius norm is denoted by $\|\cdot\|_{\mathrm{F}}$. The results regarding $\Delta_h(n)$ are shown in the upper plot of Fig.4. Without prefiltering (i. e. $\varphi_a = 0$), the normalized misalignment shown there is only reduced to -0.38dB, which confirms that the system identification for the considered scenario is extremely challenging (this difficulty might be one reason why scenarios with such a large number of channels are only barely investigated in the literature). When rotating the wave field with an amplitude of $\varphi_a = \pi/48, 4\pi/48$, and $8\pi/48$ (i. e., 3.75°, 15.0°, and 30°), a respective misalignment of -0.91, -2.4 and -5.9 dB can be achieved. Considering the curves over time, it can be noted that the misalignment is reduced faster the larger the slope of $\varphi(n)$ is.

However, as long as the normalized misalignment is not very low, a better system identification does not imply a better instantaneous AEC performance. The latter is measured by the Echo Return Loss Enhancement (ERLE) defined as

$$\mathrm{ERLE}(n) = 10\log_{10}\left(\|\mathbf{d}(n)\|_2^2 / \|\mathbf{e}(n)\|_2^2\right) \mathrm{dB}. \qquad (12)$$

Considering the lower plot of Fig.4, we can see that a time-varying rotation of the wave field does significantly influence the ERLE, most notably after steep slopes of $\varphi(n)$. Under the ideal simulation conditions, the AEC without prefiltering achieves an unrealistically high ERLE value such as 61dB after 30 seconds, while the AEC with decorrelated loudspeaker signals only achieves values reaching from 30 to 40dB. However, experiments showed that for this large-scale multichannel scenario an ERLE around 40dB is already the maximum which can be achieved in practice, so this degradation of the ERLE will be less noticeable in practice.

5. EVALUATION OF PERCEIVED SOUND QUALITY

We conducted an informal listening test to asses the influence of the wave field rotation on the perceived audio quality. There were 16 normal hearing subjects (including 4 expert listeners) aged between 20 and 30 years participating in the listening test. Considering an array setup as described in the previous section, we removed the microphones and placed the listeners in the center of the loudspeaker array. The loudspeaker signals were determined according to WFS such that plane waves were synthesized. For the mono signals, the incidence angle of the plane wave was 0°, for the stereo signals we used plane waves with incidence angles of $+30°$ and $-30°$ to mimic a perfect stereo setup.

The listening examples comprised a mono and a stereo recording of piano music as well as a mono recording of speech, each having a length of 30 seconds. The participants could repeatedly listen to 3 repetitions of each example obtained using the three values 0, $\pi/48$ and $2\pi/48$ (i. e., 0°, 3.75°, and 7.5°) of φ_a in a random order unknown to the listener. The values $L_P = 301$ and $L_F = 512$ were chosen as in the previous section, whereby we used a sampling rate of 44.1kHz ($L_P = 301$ equals ca. 3.5 sec). We evaluated two categories named "spatial stationarity" and "general quality". The first one is used to assess if a source is perceived as being reproduced from a fixed direction and was quantized to the five ratings: no noticeable rotation (5 points), barely noticeable rotation (4 points), rotation is noticeable but not disturbing (3 points), rotation is clearly noticeable and disturbing (2 points), and not acceptable (1 point). The second category named "general quality" described the overall perceived quality of the reproduced example along the MOS scale [19] as excellent (5 points), good (4 points), fair (3 points), poor (2 points), and bad (1 point).

The results in Figures 5 and 6 are shown using box plots, where the whiskers show the range of given evaluation points, the box spans from the lower to the upper quartile, the solid red line shows the median and the dashed blue line the mean value.

Considering the results for "spatial stationarity" in Fig.5, we see that a rotation of $\pi/48$ can only be barely noticed by most participants, while a rotation of $2\pi/48$ seems to be noticeable for most and disturbing for some listeners. Judging on the results for no rotation, the speech signal could be most accurately localized. Consequently its rotation was easier to notice compared to music. Regarding the latter, a stereo signal reproduced as two plane waves seemed to slightly conceal the wave field rotation.

However, the results for "spatial stationarity" may be a very critical measure for the proposed method, as it is directly focused on the systematic alteration of the loudspeaker signals. The listeners might instead be more interested in the perceived overall quality as shown in Fig.6. For this measure the perceived degradation due to the prefiltering is appreciably lower than for "spatial stationarity", showing that $\varphi_a = \pi/48$ might be acceptable in practice.

6. CONCLUSIONS

In this paper we proposed a new approach for the decorrelation of WFS loudspeaker signals to facilitate the task of system identification. Evaluation results show that the LEMS can be more accurately identified using this approach, while the parameters must be carefully chosen in order to retain an acceptable perceived reproduction quality. For further development, we propose a generalization of this approach to spherical harmonics to allow an additional variation of elevation angles, i. e., a tilt of the wave field.

7. REFERENCES

[1] A.J. Berkhout, D. De Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, vol. 93, pp. 2764 – 2778, May 1993.

[2] J. Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a variable, new ambisonic format," in *23rd International Conference of the Audio Eng. Soc.*, 2003.

[3] 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. Speech Audio Process.*, vol. 6, no. 2, pp. 156 – 165, March 1998.

[4] H. Buchner, J. Benesty, and W. Kellermann, "Multichannel frequency-domain adaptive algorithms with application to acoustic echo cancellation," in *Adaptive Signal Processing: Application to Real-World Problems*, J. Benesty and Y. Huang, Eds. Springer, Berlin, 2003.

[5] M.M. Sondhi, D.R. Morgan, and J.L. Hall, "Stereophonic acoustic echo cancellation – an overview of the fundamental problem," *IEEE Signal Process. Lett.*, vol. 2, no. 8, pp. 148 –151, August 1995.

[6] A. Gilloire and V. Turbin, "Using auditory properties to improve the behaviour of stereophonic acoustic echo cancellers," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Seattle, WA, May 1998, vol. 6, pp. 3681 – 3684.

[7] T. Gänsler and P. Eneroth, "Influence of audio coding on stereophonic acoustic echo cancellation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Seattle, WA, May 1998, vol. 6, pp. 3649 – 3652.

[8] D.R. Morgan, J.L. Hall, and J. Benesty, "Investigation of several types of nonlinearities for use in stereo acoustic echo cancellation," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 6, pp. 686 – 696, September 2001.

[9] M. Ali, "Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Seattle, WA, May 1998, vol. 6, pp. 3689 – 3692.

[10] J. Herre, H. Buchner, and W. Kellermann, "Acoustic echo cancellation for surround sound using perceptually motivated convergence enhancement," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Honolulu, Hawaii, April 2007, vol. 1, pp. I–17 – I–20.

[11] T. S. Wada and B.H. Juang, "Multi-channel acoustic echo cancellation based on residual echo enhancement with effective channel decorrelation via resampling," *Acoustic Signal Enhancement; Proceedings of IWAENC 2010; International Workshop on*, pp. 1 – 4, Sep. 2010.

[12] J. Wung, T. S. Wada, and B. H. Juang, "Inter-channel decorrelation by sub-band resampling in frequency domain," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, Kyoto, Japan, March 2012, pp. 29 – 32.

[13] K. Helwani, S. Spors, and H. Buchner, "Spatio-temporal signal preprocessing for multichannel acoustic echo cancellation," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, Prague, Czech Republic, May 2011, pp. 93 –96.

[14] S. Spors, H. Buchner, R. Rabenstein, and W. Herbordt, "Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering," *J. Acoust. Soc. Am.*, vol. 122, no. 1, pp. 354 – 369, Juli 2007.

[15] 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)*, Montreal, Canada, May 2004, vol. 4, pp. IV–117 – IV–120.

[16] M. Schneider and W. Kellermann, "A wave-domain model for acoustic MIMO systems with reduced complexity," in *Third Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, Edinburgh, UK, May 2011.

[17] Frank W. Olver, Daniel W. Lozier, Ronald F. Boisvert, and Charles W. Clark, *NIST Handbook of Mathematical Functions*, Cambridge University Press, New York, NY, 1st edition, 2010.

[18] M. Schneider and W. Kellermann, "A direct derivation of transforms for wave-domain adaptive filtering based on circular harmonics," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, Bucharest, Romania, August 2012, European Association for Signal Processing (EURASIP), pp. 1034 – 1038.

[19] ITU-T, *Recommendation P.85. Telephone transmission quality subjective opinion tests. A method for subjective performance assessment of the quality of speech voice output devices*, 1994.