IMPROVED ONLINE IDENTIFICATION OF ACOUSTIC MISO SYSTEMS BASED ON SEPARATED INPUT SIGNAL COMPONENTS

Philipp Thüne and Gerald Enzner

Institute of Communication Acoustics (IKA), Ruhr-Universität Bochum, Germany
{philipp.thuene | gerald.enzner}@rub.de

ABSTRACT

Creating an immersive listening experience and providing the audience with improved spatial realism is the goal of many adaptive audio reproduction techniques such as room equalization or crosstalk cancellation. The majority of these approaches currently relies on acoustic impulse responses (AIRs) that have been measured prior to the actual audio reproduction. In order to maintain a high degree of adaptivity, however, the AIRs need to be estimated online during the reproduction process, which turns out to be a severely ill-conditioned problem due to the high inter-channel correlation of the loudspeaker signals. In this paper, we present a novel approach to MISO system identification with realistically correlated excitation. Based on the idea of separate treatment of correlated and uncorrelated signal components, we propose two extended filter structures for gradient-descent-based adaptive system identification and provide theoretical analysis and experimental validation of their effectiveness.

Index Terms— room acoustics, multichannel acoustic system identification, nonuniqueness, gradient descent algorithm.

1. INTRODUCTION

One of the most ambitious goals in modern audio reproduction processing is to provide the audience with an immersive listening environment and to create an improved spatial realism in the displayed acoustical scene [1]. To this end, several techniques have been proposed to modify room acoustics by means of room equalization [2, 3, 4] or listening room compensation [5, 6]. Other works investigate crosstalk cancellation to present binaural recordings or binaurally rendered signals directly to the listener's ears without the cumbersome use of headphones [7, 8, 9]. Combined solutions [10] and approaches aiming at physical sound field reproduction [11, 12] also provide promising results.

The majority of these approaches, however, relies on a linear description of the acoustical transmission in terms of point-to-point acoustic impulse responses (AIRs). Typically, these AIRs are measured prior to the audio reproduction process. Although most of the proposed reproduction techniques are adaptive by themselves, a high degree of adaptivity can only be maintained if the AIRs are also estimated online by means of multichannel system identification. This implies that no probe or measurement signals can be displayed via loudspeakers but only the loudspeaker signals provided by the reproduction application are available for estimation. These signals are, however, highly mutually correlated since they contain spatial cues for the listener. Thus, the adaptive acquisition of AIRs during sound reproduction is in fact a supervised multichannel system identification problem with highly correlated system excitation.

A very similar problem is encountered in stereophonic acoustic echo cancellation (SAEC) and in this context it has been shown that the problem is severely ill-conditioned in a mathematical sense [13, 14]. In the respective SAEC literature, this issue is primarily tackled by decorrelating the system excitation, i.e., altering the loudspeaker signals in order to break up their linear dependency and thereby improving the conditioning of the underlying mathematical problem. Several approaches have been investigated in this field, ranging from classical nonlinear [13, 15] and time-variant [16, 17] preprocessing to recent resampling [18], time-reversal [19], or psychoacoustically motivated techniques [20]. A disadvantage common to all decorrelation preprocessing, however, is a degradation of signal quality, e.g., spatial perception of the stereo signal. Although the distortions introduced by state-of-the-art decorrelation may be tolerable in speech communication applications, these techniques may not be suitable for high-fidelity audio reproduction setups.

In this contribution, we assume that the loudspeaker signals of reproduction applications can be decomposed into correlated and uncorrelated components. On this basis, we propose a novel processing strategy for the identification of acoustic multiple-input single-output (MISO) systems by explicitly exploiting the characteristics of the isolated excitation components. In particular, we present two extended filter structures that can be implemented with state-of-the-art system identification techniques and allow for improved system identification performance compared to the conventional scheme.

This paper is structured as follows. Sec. 2 introduces the signal model and problem formulation. Then, Sec. 3 describes the general idea of separated input signal components and presents two filter structures treating these components as individual excitations. An analysis of these structures in the context of gradient-descent-based system identification is given in Sec. 4. Experimental results are presented and discussed in Sec. 5. Sec. 6 concludes this contribution.

Throughout this paper, we will use the following notation. Bold lower and upper case letters refer to vectors and matrices, respectively, while non-bold letters denote scalar quantities. Matrices $\mathbf{I}$ and $\mathbf{0}$ are appropriately sized identity and zero matrices, respectively, and superscript $\cdot^T$ denotes transposition. $E\{\cdot\}$ is statistical expectation, ∇ is the gradient operator with respect to the quantity specified by a subscript, and $\|\cdot\|$ denotes the l_2 vector norm.

2. SIGNAL MODEL

A general multiple-input multiple-output (MIMO) system identification task in reproduction applications can be decomposed into several P-channel MISO system identification problems as illustrated in Fig. 1. Loudspeaker signals $x_i(k)$ for $i = 1 \dots P$ at discrete time k are convolved with the AIRs $\mathbf{h}_i = [h_i(0) \ \cdots \ h_i(L_h - 1)]^T$ of length L_h to form an audio signal $d(k)$. Adding independent observation noise $n(k)$ to $d(k)$ then yields signal $y(k)$ that is observed at a reference microphone. By defining excitation vectors

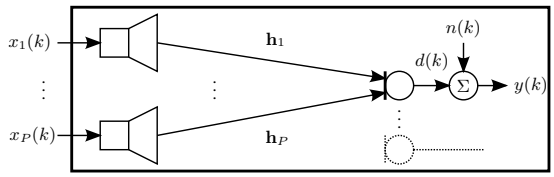

Fig. 1. MISO reproduction setup.

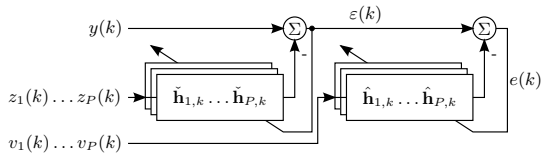

Fig. 2. Proposed serial filter structure.

$\mathbf{x}'_{i,k} = [x_i(k) \ \cdots \ x_i(k - L_h + 1)]^T$, $y(k)$ can be expressed as

$$\begin{aligned} y(k) &= \sum_{i=1}^{P} \mathbf{x}_{i,k}'^T \mathbf{h}_i + n(k) \\ &= \mathbf{x}_k'^T \mathbf{h} + n(k) \\ &= d(k) + n(k)\,, \end{aligned} \tag{1}$$

where $\mathbf{h} = \left[\mathbf{h}_1^T \ \cdots \ \mathbf{h}_P^T\right]^T$ and $\mathbf{x}_k'^T = \left[\mathbf{x}_{1,k}'^T \ \cdots \ \mathbf{x}_{P,k}'^T\right]^T$ are stacked versions of the AIRs and excitation vectors, respectively.

We further assume that excitations $x_i(k)$ can be decomposed as

$$x_i(k) = z_i(k) + v_i(k)\,, \tag{2}$$

where $z_i(k)$ are correlated signal components and $v_i(k)$ denote uncorrelated components, i.e., $\exists \kappa, i \neq j \ E\{z_i(k+\kappa) z_j(k)\} \neq 0$, $E\{v_i(k+\kappa) v_j(k)\} = 0 \ \forall \kappa, i \neq j$, and $E\{z_i(k+\kappa) v_j(k)\} = 0 \ \forall \kappa, i, j$. In the following, the correlated and uncorrelated signal components will therefore shortly be referred to as z- and v-components, respectively. The process of isolating the excitation components itself is not considered in this contribution. There are, however, cases in which the v-components are directly accessible, especially if they are artificially introduced, e.g., [13, 15].

The decomposition in (2) allows the description of the problem conditioning via the power ratio of the respective components, i.e.,

$$\eta_i = 10 \log_{10} \left(\frac{E\{z_i^2(k)\}}{E\{v_i^2(k)\}} \right) \ \forall i \tag{3}$$

which is defined similar to a signal-to-noise ratio. With η_i, we can roughly classify the possible situations in system identification.

Nonunique ($\eta_i \to \infty$): With strictly correlated excitation only, the problem is mathematically singular and decorrelation must be applied [13]. Note though that such strict linear dependency is rarely encountered in practical applications.

Ill-conditioned ($0 < \eta_i < \infty$): Small v-components are inherently present in the system excitation or they are generated by a decorrelation method applied to the nonunique case.

Well-conditioned ($-\infty < \eta_i < 0$): For low values of η_i the problem conditioning improves. Measuring with uncorrelated probe signals, i.e., $\eta_i \to -\infty$, renders the identification task rather trivial.

In this paper, we consider the ill-conditioned case only, since the nonunique situation must be cast into an ill-conditioned problem by decorrelation anyway, while the well-conditioned case is not very challenging from a system identification perspective for $\eta_i \ll 0$.

3. PROPOSED SOLUTION

3.1. From Conventional MISO System Identification to Separate Processing of Excitation Components

Conventional approaches to MISO system identification aim to mimic the structure of the reproduction setup illustrated in Fig. 1. Therefore, a conventional structure employs P adaptive filters $\hat{\mathbf{h}}_{i,k}$ to model the first $L \leq L_h$ coefficients of each AIR $\mathbf{h}_i$. The adaptation process is driven by the observed microphone signal $y(k)$ and the known loudspeaker signals $x_i(k)$, i.e., the additive mixture of z- and v-components. This procedure only results in satisfactory performance for rather low values of η_i, i.e., a large amount of decorrelation is required.

An early idea of a dedicated treatment of the z- and v-components has been presented over a decade ago. In [21, 22], the authors propose to amplify the v-components originating from a nonlinear half-wave rectifier by introducing so-called enhanced input signal vectors to improve the convergence of a modified version of the normalized least-mean-square algorithm (NLMS).

Assuming that we can isolate the z- and v-components, however, enables us to go even further and abandon the conventional paradigm of employing P adaptive filters for a P-channel system. Instead, we consider the isolated components as individual input signals and propose extended filter structures to adapt twice the number of physically available filters with individual purposes.

It is important to note that our processing approach operates without predistortion if the excitation inherently contains sufficiently strong v-components or it follows after a necessary decorrelation stage to achieve better convergence at the same level of distortion. Therefore, our approach is not meant as an alternative, but rather as an addition to already existing decorrelation techniques.

3.2. Proposed Serial Filter Structure

The first filter structure we propose is a serial combination of two P-channel adaptive filters as illustrated in Fig. 2. The first stage estimates filters $\check{\mathbf{h}}_{i,k}$ of length L based on the z-components and the microphone signal $y(k)$ only. As the excitation of the first stage is strictly correlated, the identification problem is singular and the filters $\check{\mathbf{h}}_{i,k}$ will not be unique. Nonetheless, the adaptation algorithm will minimize the output error $\varepsilon(k)$ of this stage. The error signal $\varepsilon(k)$ only contains, in the ideal case, the observation noise $n(k)$ and those signal components of $y(k)$ that cannot be modeled by $z_i(k)$ since they originate from $v_i(k)$. This fact is exploited in the second stage in which $\varepsilon(k)$ acts as the desired signal and $v_i(k)$ are used as input signals. The adaptive filters $\hat{\mathbf{h}}_{i,k}$ will converge to the true AIRs since their excitation is uncorrelated. Unfortunately, the first stage must be converged before the second stage can operate successfully.

3.3. Proposed Parallel Filter Structure

Our second filter structure that treats the input signal components as separate system excitations is the parallel $2P$-channel structure shown in Fig. 3. The idea behind this structure is very similar to the one of the serial structure. In fact, the parallel structure can be derived from the serial structure by using $e(k)$ instead of $\varepsilon(k)$ to adapt $\check{\mathbf{h}}_{i,k}$. This eliminates the two-stage convergence behavior of the serial structure, but requires the adaptive algorithm to adapt twice as many channels from a single microphone signal observation. A

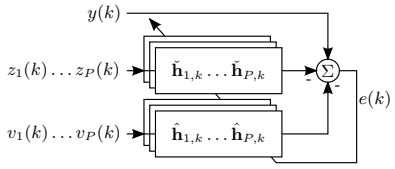

Fig. 3. Proposed parallel filter structure.

similar structure has been proposed in [23] for SAEC applications based on a modified nonlinear half-wave rectifier.

Although serial and parallel filter structures adapt $2P$ filters in total, only filters $\hat{\mathbf{h}}_{i,k}$ are considered as estimates of the true AIRs $\mathbf{h}_i$, whereas $\check{\mathbf{h}}_{i,k}$ are internal quantities that are discarded eventually.

4. THEORETICAL ANALYSIS

4.1. Gradient Descent Review for Conventional Identification

The extended filter structures introduced in Sec. 3 can basically be implemented with any gradient-descent-based algorithm. As a reference, we first illustrate how an ideal gradient descent approach with known correlation quantities operates in the conventional P-channel filter structure without considering observation noise.

System identification based on the gradient descent approach estimates the unknown system responses recursively by correcting the current estimate with the gradient of a cost function $J(\hat{\mathbf{h}}_k)$ [24], i.e.,

$$\hat{\mathbf{h}}_{k+1} = \hat{\mathbf{h}}_k - \frac{1}{2}\beta_k \nabla_{\hat{\mathbf{h}}_k} J(\hat{\mathbf{h}}_k)\,, \tag{4}$$

where $\hat{\mathbf{h}}_k$ is a stacked version of the individual estimates $\hat{\mathbf{h}}_{i,k}$ and β_k is a scalar time-variant step-size parameter. For the conventional approach the estimation error is $e(k) = d(k) - \hat{\mathbf{h}}_k^T \mathbf{x}_k$, where $\mathbf{x}_k$ is defined analogously to $\mathbf{x}'_k$ but only the most recent L samples per channel are considered. The mean square of $e(k)$ is our cost function, which can, along with its gradient, be expressed as [24]

$$\begin{aligned} J(\hat{\mathbf{h}}_k) &= E\{e^2(k)\} \\ &= \sigma_d^2 - 2\hat{\mathbf{h}}_k^T \mathbf{r}_{zd} - 2\hat{\mathbf{h}}_k^T \mathbf{r}_{vd} + \hat{\mathbf{h}}_k^T \mathbf{R}_{zz} \hat{\mathbf{h}}_k + \hat{\mathbf{h}}_k^T \mathbf{R}_{vv} \hat{\mathbf{h}}_k\,, \\ \nabla_{\hat{\mathbf{h}}_k} J(\hat{\mathbf{h}}_k) &= -2\mathbf{r}_{zd} - 2\mathbf{r}_{vd} + 2\mathbf{R}_{zz}\hat{\mathbf{h}}_k + 2\mathbf{R}_{vv}\hat{\mathbf{h}}_k\,. \end{aligned} \tag{5}$$

Here, $\mathbf{r}_{zd} = E\{\mathbf{z}_k d(k)\}$, $\mathbf{r}_{vd} = E\{\mathbf{v}_k d(k)\}$, $\mathbf{R}_{zz} = E\{\mathbf{z}_k \mathbf{z}_k^T\}$, and $\mathbf{R}_{vv} = E\{\mathbf{v}_k \mathbf{v}_k^T\}$ are correlation vectors and matrices with $\mathbf{x}_k = \mathbf{z}_k + \mathbf{v}_k$ in accordance with (2) and $\sigma_d^2 = E\{d^2(k)\}$.

Forming a vector $\mathbf{h}_0$ by considering the first L coefficients of the true AIRs, we can define a coefficient error $\Delta\hat{\mathbf{h}}_k = \mathbf{h}_0 - \hat{\mathbf{h}}_k$ that will vanish once the system is identified correctly. For sufficiently large L we can approximate $d(k) \approx \mathbf{x}_k^T \mathbf{h}_0 = \mathbf{z}_k^T \mathbf{h}_0 + \mathbf{v}_k^T \mathbf{h}_0$, thus $\mathbf{r}_{zd} = E\{\mathbf{z}_k d(k)\} \approx E\{\mathbf{z}_k \mathbf{z}_k^T \mathbf{h}_0 + \mathbf{z}_k \mathbf{v}_k^T \mathbf{h}_0\} = \mathbf{R}_{zz}\mathbf{h}_0$ and similarly $\mathbf{r}_{vd} \approx \mathbf{R}_{vv}\mathbf{h}_0$. Using these approximations along with (5), we can turn (4) into a difference equation for $\Delta\hat{\mathbf{h}}_k$, i.e.,

$$\Delta\hat{\mathbf{h}}_{k+1} = (\mathbf{I} - \beta_k(\mathbf{R}_{zz} + \mathbf{R}_{vv}))\,\Delta\hat{\mathbf{h}}_k\,. \tag{6}$$

This difference equation will only converge to zero coefficient error safely if all eigenvalues λ of $\beta_k(\mathbf{R}_{zz} + \mathbf{R}_{vv})$ well satisfy $0 < \lambda < 2$ [24]. Since $\mathbf{R}_{zz}$ is ill-conditioned, i.e., it has a large eigenvalue spread due to the mutual correlation of the z-components, this criterion cannot be met simply by adjusting β_k. Even though $\mathbf{R}_{vv}$ is added to $\mathbf{R}_{zz}$, the conditioning will only improve slightly because $E\{z_i^2(k)\} \gg E\{v_i^2(k)\}$. The goal of our extended structures must therefore be to eliminate $\mathbf{R}_{zz}$ from the update equation for $\Delta\hat{\mathbf{h}}_k$.

4.2. Gradient Descent for the Serial Filter Structure

For the serial structure, the estimation error of the first stage is $\varepsilon(k) = d(k) - \check{\mathbf{h}}_k^T \mathbf{z}_k$ according to Fig. 2 with $\sigma_\varepsilon^2 = E\{\varepsilon^2(k)\}$. In our analysis, we consider the second stage only, assuming that the first stage successfully minimizes $\varepsilon(k)$ but not the coefficient error $\Delta\check{\mathbf{h}}_k = \mathbf{h}_0 - \check{\mathbf{h}}_k$. With the estimation error $e(k) = \varepsilon(k) - \hat{\mathbf{h}}_k^T \mathbf{v}_k$ of the second stage, we can write its cost function as

$$\begin{aligned} J(\hat{\mathbf{h}}_k) &= E\{e^2(k)\} \\ &= \sigma_\varepsilon^2 - 2\check{\mathbf{h}}_k^T \mathbf{r}_{zd} - 2\hat{\mathbf{h}}_k^T \mathbf{r}_{vd} + \check{\mathbf{h}}_k^T \mathbf{R}_{zz} \check{\mathbf{h}}_k + \hat{\mathbf{h}}_k^T \mathbf{R}_{vv} \hat{\mathbf{h}}_k\,. \end{aligned}$$

As the second stage can only adapt filters $\hat{\mathbf{h}}_k$, the gradient is

$$\nabla_{\hat{\mathbf{h}}_k} J(\hat{\mathbf{h}}_k) = -2\mathbf{r}_{vd} + 2\mathbf{R}_{vv}\hat{\mathbf{h}}_k\,. \tag{7}$$

With (4), (7), and by approximating $\mathbf{r}_{vd} \approx \mathbf{R}_{vv}\mathbf{h}_0$ as before, we can derive the difference equation for the coefficient error,

$$\Delta\hat{\mathbf{h}}_{k+1} = (\mathbf{I} - \beta_k \mathbf{R}_{vv})\,\Delta\hat{\mathbf{h}}_k\,, \tag{8}$$

which will eventually converge to zero for an appropriate step-size β_k, since the eigenvalue spread of $\mathbf{R}_{vv}$ is low.

4.3. Gradient Descent for the Parallel Filter Structure

Introducing composite vectors $\tilde{\mathbf{h}}_k = [\check{\mathbf{h}}_k^T\ \hat{\mathbf{h}}_k^T]^T$ and $\tilde{\mathbf{x}}_k = [\mathbf{z}_k^T\ \mathbf{v}_k^T]^T$ of length $2PL$, the error signal of the parallel structure becomes $e(k) = d(k) - \tilde{\mathbf{h}}_k^T \tilde{\mathbf{x}}_k$ (cf. Fig. 3). Applying the gradient descent approach to the filter vector $\tilde{\mathbf{h}}_k$ yields

$$\tilde{\mathbf{h}}_{k+1} = \tilde{\mathbf{h}}_k - \frac{1}{2}\beta_k \nabla_{\tilde{\mathbf{h}}_k} J(\tilde{\mathbf{h}}_k)\,. \tag{9}$$

The cost function is now a function of the composite vector $\tilde{\mathbf{h}}_k$ instead of just $\hat{\mathbf{h}}_k$ and can be expressed, along with its gradient, as

$$\begin{aligned} J(\tilde{\mathbf{h}}_k) &= E\{e^2(k)\} \\ &= \sigma_d^2 - 2\tilde{\mathbf{h}}_k^T \mathbf{r}_{\tilde{x}d} + \tilde{\mathbf{h}}_k^T \mathbf{R}_{\tilde{x}\tilde{x}} \tilde{\mathbf{h}}_k\,, \\ \nabla_{\tilde{\mathbf{h}}_k} J(\tilde{\mathbf{h}}_k) &= -2\mathbf{r}_{\tilde{x}d} + 2\mathbf{R}_{\tilde{x}\tilde{x}}\tilde{\mathbf{h}}_k \end{aligned} \tag{10}$$

with $\mathbf{r}_{\tilde{x}d} = [\mathbf{r}_{zd}{}^T\ \mathbf{r}_{vd}{}^T]^T$ and $\mathbf{R}_{\tilde{x}\tilde{x}} = \begin{pmatrix} \mathbf{R}_{zz} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{vv} \end{pmatrix}$.

By combining (9) and (10), utilizing the same approximation as in the conventional case, i.e., $\mathbf{r}_{\tilde{x}d} \approx [(\mathbf{R}_{zz}\mathbf{h}_0)^T\ (\mathbf{R}_{vv}\mathbf{h}_0)^T]^T$, and exploiting the block-diagonal structure of $\mathbf{R}_{\tilde{x}\tilde{x}}$, we get two decoupled difference equations for the upper and lower half of $\tilde{\mathbf{h}}_k$, i.e.,

$$\Delta\check{\mathbf{h}}_{k+1} = (\mathbf{I} - \beta_k \mathbf{R}_{zz})\,\Delta\check{\mathbf{h}}_k\,, \tag{11}$$

$$\Delta\hat{\mathbf{h}}_{k+1} = (\mathbf{I} - \beta_k \mathbf{R}_{vv})\,\Delta\hat{\mathbf{h}}_k\,. \tag{12}$$

The difference equation for $\Delta\hat{\mathbf{h}}_k$ in (12) is identical to the one in (8) and will again converge to zero for an appropriate step-size β_k.

5. EXPERIMENTAL ANALYSIS

5.1. Performance of the Proposed Structures

Since the theoretical analysis in Sec. 4 is based on known correlation matrices and does not take noise or the imperfectness of practical algorithms into account, we will support the effectiveness of the proposed filter structures by simulations with data-driven gradient-descent type algorithms. We set up a virtual room of size $4\,\mathrm{m} \times 5\,\mathrm{m} \times$

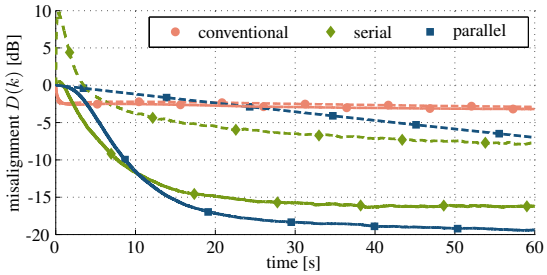

Fig. 4. Convergence of conventional and proposed filter structures for NLMS (dashed) and MCSSFDAF (solid).

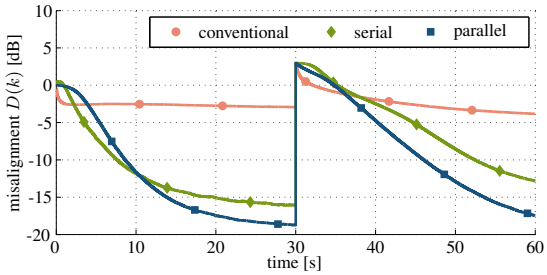

Fig. 5. Reconvergence of MCSSFDAF filter structures with a sudden change of the true MISO system after 30 s.

3 m with a reverberation time $T_{60} = 0.25$ s and place $P = 5$ loudspeakers according to ITU-R BS.775 inside the room. At the center, we place a microphone and calculate P AIRs of length $L_h = 4096$ at 48 kHz using the image source method [25]. All z-components of the loudspeaker signals are generated by filtering a single white Gaussian test signal using filters of length 4096 that represent the AIRs of a different room. The v-components are created as independent white Gaussian noises at $\eta_i = 20$ dB and both components are added to form the loudspeaker signals according to (2). The microphone signal is finally obtained by using (1) with white observation noise $n(k)$ to yield a microphone SNR of 30 dB.

In a first experiment, we compare the performance of the conventional, serial, and parallel filter structures. For system identification, we use a simple multichannel NLMS algorithm [24] with normalized step-size $\mu = 0.95$ (conventional and parallel case) or $\mu = 0.2$ (serial case), and the multichannel state-space frequency domain adaptive filter (MCSSFDAF) with transition coefficient $A = 0.9997$, frame-size 4096, and frame-shift 2048, which can be understood as a recent gradient descent algorithm with optimal step-size [26]. Both algorithms estimate filters with $L = 2048$ coefficients.

The results of this experiment are shown in Fig. 4 in terms of misalignment $D(k) = 10\log_{10}(\|\Delta\hat{\mathbf{h}}_k\|^2 / \|\mathbf{h}_0\|^2)$. They clearly show that $\eta_i = 20$ dB is not suitable for the conventional approach to converge to a satisfactory level for both algorithms, i.e., more decorrelation would be required. The proposed structures, however, achieve very good identification performance implemented with MCSSFDAF. The NLMS curves also converge better for the proposed structures than for conventional identification, but NLMS still faces problems: In the serial structure, the first stage must have converged before the second one can operate correctly. As a result, the second stage even diverges during the first 5 s for NLMS. In the parallel structure, NLMS identification is just too slow due to the normalization across all components, which is more sophisticatedly handled as a channel-wise normalization by MCSSFDAF.

5.2. Robustness of the Proposed Structures

We run two more experiments with the same setup in order to evaluate the robustness of the extended filter structures to practical conditions. Since the NLMS identification has turned out to be less effective than MCSSFDAF, we proceed with the latter only.

In order to evaluate the reconvergence behavior of the proposed approaches, we introduce a rapid change of the true AIRs $\mathbf{h}_i$ after 30 s. The results for this experiment are depicted in Fig. 5 and show a reconvergence rate similar to the initial convergence for both of the proposed structures. The results support the idea that the proposed schemes could be used to track changes in the AIRs during audio reproduction processing.

In the final experiment, we introduce burst noise into the microphone signal. Three times, the SNR at the microphone will decrease to -25 dB simulating, e.g., an independent talker within the reproduction room. The identification performance shown in Fig. 6 reveals that both filter structures behave very robust during the noise bursts. For the serial structure, we observe a slight delay before the effect becomes evident in the second stage, since the bursts affect the first stage first. The results demonstrate that noise robustness known from MCSSFDAF is preserved in the proposed filter structures.

6. CONCLUSION AND RELATION TO PRIOR WORKS

In this paper, we proposed a novel processing scheme for online MISO acoustic system identification with correlated input signals as they are encountered in numerous audio reproduction applications, such as room equalization [2, 3, 4] or crosstalk cancellation [7, 9, 8, 10]. Unlike prior works on MISO system identification in the context of SAEC, we did not present a new decorrelation method. Instead, we introduced extended filter structures that treat the existing correlated and uncorrelated excitation components as separate input signals. The uncorrelated components required in this approach can be naturally contained in the excitation or artificially introduced by a predistortion technique, e.g., [15, 18, 20]. We proved our processing scheme in an idealized general gradient descent framework and demonstrated its effectiveness and robustness in experiments with classical and state-of-the-art system identification algorithms. Once a suitable separation method for the excitation components has been found, the proposed scheme could be used to reduce the amount of predistortion required, or, to allow for system identification without decorrelation if sufficiently uncorrelated components are already present in the original input signals.

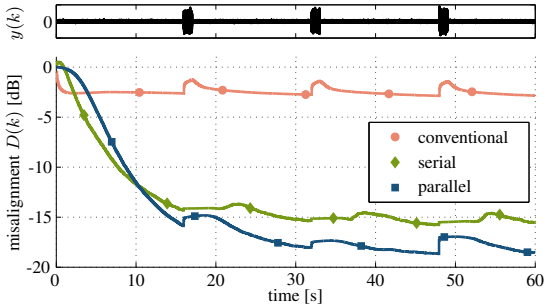

Fig. 6. Robustness of MCSSFDAF filter structures with 1 s noise bursts at -25 dB SNR in the microphone signal $y(k)$.

7. REFERENCES

[1] Y. Huang, J. Chen, and J. Benesty, "Immersive audio schemes," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 20–32, Jan. 2011.

[2] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 2, pp. 145–152, Feb. 1988.

[3] M. Kolundžija, C. Faller, and M. Vetterli, "Multi-channel low-frequency room equalization using perceptually motivated constrained optimization," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012.

[4] M. Schneider and W. Kellermann, "Adaptive listening room equalization using a scalable filtering structure in the wave domain," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012.

[5] 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, July 2007.

[6] M. Kolundžija, C. Faller, and M. Vetterli, "Reproducing sound fields using MIMO acoustic channel inversion," *J. Audio Eng. Soc.*, vol. 59, no. 10, pp. 721–734, Oct. 2011.

[7] Y. Huang, J. Benesty, and J. Chen, "On crosstalk cancellation and equalization with multiple loudspeakers for 3-d sound reproduction," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 649–652, Oct. 2007.

[8] Y. Lacouture-Parodi and E.A.P. Habets, "Crosstalk cancellation system using a head tracker based on interaural time differences," in *Proc. Int. Workshop Acoustic Signal Enhancement (IWAENC)*, Aachen, Germany, Sept. 2012.

[9] T. Betlehem, P.D. Teal, and Y. Hioka, "Efficient crosstalk canceler design with impulse response shortening filters," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012.

[10] J.O. Jungmann, R. Mazur, M. Kallinger, T. Mei, and A. Mertins, "Combined acoustic MIMO channel crosstalk cancellation and room impulse response reshaping," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 6, pp. 1829–1842, Aug. 2012.

[11] S. Yon, M. Tanter, and M. Fink, "Sound focusing in rooms: The time-reversal approach," *J. Acoust. Soc. Am.*, vol. 113, no. 3, pp. 1533–1543, Mar. 2003.

[12] S. Yon, M. Tanter, and M. Fink, "Sound focusing in rooms. II. The spatio-temporal inverse filter," *J. Acoust. Soc. Am.*, vol. 114, no. 6, pp. 3044–3052, Dec. 2003.

[13] 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, Mar. 1998.

[14] T. Gänsler and J. Benesty, "Multichannel acoustic echo cancellation: What's new?," in *Proc. Int. Workshop Acoustic Echo and Noise Control (IWAENC)*, Darmstadt, Germany, Sept. 2001.

[15] 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, Sept. 2001.

[16] M. Ali, "Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Seattle, WA, USA, May 1998.

[17] A. Sugiyama, Y. Joncour, and A. Hirano, "A stereo echo canceler with correct echo-path identification based on an input-sliding technique," *IEEE Trans. Signal Process.*, vol. 49, no. 11, pp. 2577–2587, Nov. 2001.

[18] J. Wung, T.S. Wada, and B.-H. Juang, "Inter-channel decorrelation by sub-band resampling in frequency domain," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012.

[19] D.-Q. Nguyen, W.-S. Gan, and A.W.H. Khong, "Time-reversal approach to the stereophonic acoustic echo cancellation problem," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 2, pp. 385–395, Feb. 2011.

[20] L. Romoli, S. Cecchi, P. Peretti, and F. Piazza, "A mixed decorrelation approach for stereo acoustic echo cancellation based on the estimation of the fundamental frequency," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 2, pp. 690–698, Feb. 2012.

[21] S. Emura, Y. Haneda, and S. Makino, "Adaptive filtering algorithm enhancing decorrelated additive signals for stereo echo cancellation," in *Proc. Int. Workshop Acoustic Echo and Noise Control (IWAENC)*, Darmstadt, Germany, Sept. 2001.

[22] S. Emura, Y. Haneda, A. Kataoka, and S. Makino, "Stereo echo cancellation algorithm using adaptive update on the basis of enhanced input-signal vector," *Signal Process.*, vol. 86, no. 6, pp. 1157–1167, June 2006.

[23] S. Shimauchi, Y. Haneda, S. Makino, and Y. Kaneda, "New configuration for a stereo echo canceller with nonlinear preprocessing," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Seattle, WA, USA, May 1998.

[24] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle Hill, NJ, USA, 2002.

[25] J.B. Allen and D.A. Berkley, "Image method for efficiently simulating small room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[26] S. Malik and G. Enzner, "Recursive Bayesian control of multichannel acoustic echo cancellation," *IEEE Signal Process. Lett.*, vol. 18, no. 11, pp. 619–622, Nov. 2011.