Robust Blind Multichannel Identification based on a Phase Constraint and Different ℓ_p-norm Constraints

Byeongho Jo, Paul Calamia
Facebook Reality Labs
Redmond, WA, USA
{bhjo, pcalamia}@fb.com

***Abstract*—Blind multichannel identification has played a crucial role as a prerequisite for channel equalization, speech dereverberation, and time-delay estimation for decades. Algorithms based on cross-relation (CR) errors have been widely studied, however, these algorithms have been reported to be fundamentally vulnerable to additive noise. Consequently, there have been many studies to improve robustness. Among them, the ℓ_p-robust normalized multichannel frequency-domain least-mean-square (ℓ_p-RNMCFLMS) algorithm was developed recently to blindly identify a single-input-multiple-output system in a reverberant environment. However, the additional penalty functions of conventional algorithms including ℓ_p-RNMCFLMS are based only on the magnitude spectra of the microphone channels. In this work, we propose a new penalty function applied to the phase spectra of the system. Furthermore, we develop an extended ℓ_p-RNMCFLMS to improve the steady-state performance by discriminating p values on impulse responses separated into significant and insignificant components. Numerical simulation results show that the proposed algorithm outperforms conventional algorithms even in low-SNR conditions.**

***Index Terms*—Blind multichannel identification, adaptive algorithms, phase spectrum**

I. Introduction

The problem of blind multichannel identification (BMCI), which estimates channel impulse responses of an unknown system only from output signals, has been extensively researched. BMCI algorithms have the advantage that they can be combined and utilized with many applications such as time-delay estimation of direct sound and early reflections [1, 2], speech dereverberation [3, 4], and multichannel equalization [5, 6]. In the past few decades, various BMCI algorithms have been developed such as subspace (SS) algorithms [7, 8], cross-relation based algorithms [9, 10], least-squares (LS) algorithms [11, 12], maximum-likelihood (ML) algorithms [13], and normalized multichannel frequency-domain least-mean-square (NMCFLMS) algorithms [14].

In particular, NMCFLMS is considered to be an effective and efficient algorithm since it can decrease the computational complexity relative to other options by using a fast Fourier transform (FFT). However, like other cross-relation (CR) based algorithms [10, 15], NMCFLMS is hard to apply directly in a practical situation due to its vulnerability to additive noise [16]. To improve the robustness against additive noise, various approaches have been proposed. Specifically to overcome the mis-convergence problem due to corruption from additive noise, LMS-based algorithms using variable or optimal step sizes have been proposed [17–19].

On the other hand, many researchers have tried to improve the robustness by introducing additional constraints to the original cost function of NMCFLMS such as a direct-path component related constraint [20–22], a spectral energy constraint [16, 23] and spectral flatness of impulse responses [24–26]. Most recently, He *et al.* proposed an ℓ_p-norm constraint for a practical acoustic environment with different degrees of reverberation [27].

Those works, however, focused on additional constraints related only to the magnitude spectra of impulse responses. In this work, we introduce an additional constraint related to the phase spectra of impulse responses that can improve the identification performance in reverberant and noisy environments. Furthermore, we propose the extended ℓ_p-RNMCFLMS algorithm that divides estimated impulse responses into two parts and introduce the use of different p values for the two parts.

II. Background of BMCI

Consider a single acoustic source measured by M microphones. The measured signal from i-th microphone can be characterized by a finite impulse response (FIR) system as

$$x_i(n) = h_i(n) * s(n) + v_i(n), \tag{1}$$

where $s(n)$ is the sound source signal, $h_i(n)$ is the channel impulse response between the source and the i-th microphone [1], $v_i(n)$ is the additive noise of i-th microphone that is assumed to be zero mean and uncorrelated to each channel and the source signal, and $(*)$ denotes linear convolution. We assume that the system satisfies the identifiability conditions from [10]: 1) the polynomials formed from h_i are co-prime; 2) the autocorrelation matrix of the source signal is of full rank.

LMS-type BMCI algorithms [24–27] employ a CR based cost function. The CR between two different microphone channels is defined as

$$e_{ij}(n) = x_i(n) * \hat{h}_j(n) - x_j(n) * \hat{h}_i(n), \tag{2}$$

where $\hat{h}_i(n)$ is the estimated impulse response of i-th channel.

[1] Here we assume all impulse responses have length L.

Among the many LMS-type algorithms, the NMCFLMS algorithm [14] is known to be computationally efficient as a result of computing the cost function in the frequency domain,

$$J_F(m) = \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \underline{\mathbf{e}}_{ij}^H(m)\underline{\mathbf{e}}_{ij}(m), \tag{3}$$

where $(\cdot)^H$ is a conjugate transpose. $\underline{\mathbf{e}}_{ij}(m)$ is a frequency-domain error signal for the m-th time-block between the i-th and j-th channels defined as

$$\underline{\mathbf{e}}_{ij}(m) = \mathbf{F}_L \mathbf{e}_{ij}(m), \tag{4}$$

where $\mathbf{F}$ denotes the discrete Fourier transform (DFT) matrix and L denotes the length of each channel's response.

The zoomed spectrum of the k-th channel at the m-th time-block, which is the frequency spectrum of the signal zero padded to length L (to alleviate the picket-fence effect when using the overlap-save technique), is updated using Newton's iterative method [14]:

$$\begin{aligned}\hat{\underline{\mathbf{h}}}_k^{10}(m+1) = \hat{\underline{\mathbf{h}}}_k^{10}(m) - \mu_f \left[\mathbf{P}_k(m) + \delta \mathbf{I}_{2L\times 2L}\right]^{-1} \\ \times \sum_{i=1}^{M} \mathbf{D}_i^*(m)\underline{\mathbf{e}}_{ik}^{01}(m),\ k = 1, 2, \cdots, M\end{aligned} \tag{5}$$

where

$$\begin{aligned}\hat{\underline{\mathbf{h}}}_k^{10}(m) &= \mathbf{F}_{2L}\left[\hat{\mathbf{h}}_k^T(m)\ \mathbf{0}\right]^T \\ &= \left[\hat{\underline{h}}_{k,1}^{10}(m),\ \cdots,\ \hat{\underline{h}}_{k,2L}^{10}(m)\right]^T,\end{aligned} \tag{6}$$

$$\mathbf{D}_i(m) = \mathrm{diag}\left[\mathbf{F}_{2L}\mathbf{x}_i(m)\right], \tag{7}$$

$$\begin{aligned}\mathbf{x}_i(m) = &[x_i(mL-L)\ \ x_i(mL-L+1) \\ &\cdots x_i(mL+L-1)]^T,\end{aligned} \tag{8}$$

$$\begin{aligned}\mathbf{P}_k(m) = &\lambda \mathbf{P}_k(m-1) \\ &+ (1-\lambda)\sum_{i=1, i\neq k}^{M} \mathbf{D}_i^*(m)\mathbf{D}_i(m),\end{aligned} \tag{9}$$

$$\underline{\mathbf{e}}_{ik}^{01}(m) = \mathbf{F}_{2L}\left[\mathbf{0}\ \ \mathbf{e}_{ik}^T(m)\right]^T. \tag{10}$$

Here, $\mathbf{x}_i(m)$ is the measured signal of the i-th channel at the m-th time block, diag$[\cdot]$ makes a diagonal matrix with a given vector and $(\cdot)^T$ is the transpose operator. μ_f is the step-size, δ is a small number to avoid singularities during matrix inversion, λ is a forgetting factor set as $\lambda = [1 - 1/(3L)]^L$ [14], and $\mathbf{0}$ is null matrix of size $1 \times L$.

The NMCFLMS algorithm is known to be vulnerable to additive noise [18, 23, 24, 28], and many algorithms have been proposed to address this robustness issue. Recently, He *et al.* proposed an improved version of the NMCFLMS algorithm utilizing an ℓ_p-norm penalty function. As described in [27], the ℓ_p-RNMCFLMS algorithm can effectively penalize the different reverberation levels of acoustic channels by using different p values $1 \leq p < 2$. The detailed updating procedure of ℓ_p-RNMCFLMS is given as

$$\begin{aligned}\hat{\underline{\mathbf{h}}}_k^{10}(m+1) =&\hat{\underline{\mathbf{h}}}_k^{10}(m) - \mu_f \nabla J_{\mathrm{F},k}(m) + \mu_f \eta(m)\nabla J_{\mathrm{FP},k}(m), \\ &k = 1, 2, \cdots, M\end{aligned} \tag{11}$$

where

$$\nabla J_{F,k}(m) = \frac{1}{2}\mathbf{S}_k^{-1}(m)\sum_{i=1}^{M} \mathbf{D}_i^*(m)\underline{\mathbf{e}}_{ik}(m), \tag{12}$$

$$\begin{aligned}\nabla J_{FP,k}(m) = &p\mathbf{S}_k^{-1}(m)|\hat{\underline{\mathbf{h}}}_k^{10}(m)|^{p-1} \\ &\odot \exp\{\jmath \arg(\hat{\underline{\mathbf{h}}}_k^{10}(m))\},\end{aligned} \tag{13}$$

$$\eta(m) = \left|\frac{[\nabla J_{FP}(m)]^H \nabla J_F(m)}{\|\nabla J_{FP}(m)\|^2}\right|, \tag{14}$$

$$\mathbf{S}_k(m) = \mathbf{P}_k(m) - \eta(m)p^2\mathbf{H}_k(m), \tag{15}$$

$$\mathbf{H}_k(m) = \mathrm{diag}\left[|\hat{\underline{h}}_{k,1}^{10}(m)|^{p-2}, \cdots, |\hat{\underline{h}}_{k,2L}^{10}(m)|^{p-2}\right], \tag{16}$$

where ∇J_{FP} and ∇J_F are stacked column vectors for all channels ($1 \leq k \leq M$), η is a Lagrangian multiplier, $\odot$ denotes element-wise multiplication, and $\jmath$ is the imaginary unit ($\jmath^2 = -1$). The identification performance can be evaluated using normalized projection misalignment (NPM) [29], which evaluates how well the algorithms estimate responses disregarding the scale. NPM is defined as:

$$\begin{aligned}\mathrm{NPM}(m) &= 20\log_{10}\|\epsilon(m)\|/\|\mathbf{h}\|\ [\mathrm{dB}] \\ \epsilon(m) &= \mathbf{h} - \left\{\mathbf{h}^T\hat{\mathbf{h}}(m)/\|\hat{\mathbf{h}}(m)\|^2\right\}\hat{\mathbf{h}}.\end{aligned} \tag{17}$$

In the next section, the details of two modifications of the ℓ_p-norm based algorithm [27] are explained.

III. Proposed Algorithm

A. Phase-based penalty function

The conventional LMS algorithms based on additional penalty functions only focused on the magnitude spectrum [20–22, 24–27]. In this work, we propose the additional phase-based penalty function to improve the identification performance from a phase perspective. The ideal anechoic channel has a linear phase, and correspondingly the second derivative of the phase is zero. However, in a more realistic case, the room impulse responses (RIRs) include reverberant components that make the phase not linear. As reverberant components of the RIRs increase, the second derivatives of the phases deviate from zero. In consideration of this, we propose a minimization penalty function based on the second derivatives of phases of estimated channel coefficients in the presence of additive noise. The modified cost function including the phase-based penalty function is written as

$$\tilde{J}(m) = J_{\mathrm{F}}(m) - \eta(m)J_{\mathrm{FP}}(m) + \eta_G(m)J_{\mathrm{FG}}(m), \tag{18}$$

$$\text{s.t. } \|\hat{\mathbf{h}}(m)\|^2 = 1 \tag{19}$$

where $\hat{\mathbf{h}}(m) = [\hat{\mathbf{h}}_1^T(m)\ \hat{\mathbf{h}}_2^T(m)\cdots\hat{\mathbf{h}}_M^T(m)]^T$, J_{FP} is the ℓ_p-norm penalty function defined in [27], and J_{FG} is a phase-related penalty function defined as

$$\begin{aligned}J_{FG,k}(m) = &\sum_{i=2}^{2L-1} |\ \arg(\hat{\underline{h}}_{k,i+1}^{10}(m)) \\ &- 2\arg(\hat{\underline{h}}_{k,i}^{10}(m)) + \arg(\hat{\underline{h}}_{k,i-1}^{10}(m))\ |^2,\end{aligned} \tag{20}$$

which is the squared summation of the normalized second-order numerical differentiation of the phase. Using this additional penalty function, the update equation for the ℓ_p-RNMCFLMS algorithm can be re-written as

$$\begin{aligned}\underline{\hat{\mathbf{h}}}_k^{10}(m+1) &= \underline{\hat{\mathbf{h}}}_k^{10}(m) - \mu_f \nabla J_{F,k}(m) \\ &+ \mu_f \left\{\eta(m)\nabla J_{FP,k}(m) - \eta_G(m)\nabla J_{FG,k}(m)\right\}\end{aligned} \tag{21}$$

where

$$\begin{bmatrix}\eta(m)\\ \eta_G(m)\end{bmatrix} = \left|\left[\nabla J_{FP}(m) \quad -\nabla J_{FG}(m)\right]^{\dagger} \nabla J_F(m)\right| \tag{22}$$

$$\nabla J_{FG,k}(m) = \mathbf{T}_k(m)\left\{\mathbf{Q}\arg\left(\underline{\hat{\mathbf{h}}}_k^{10}(m)\right)\right\}, \tag{23}$$

$$\begin{aligned}\mathbf{Q} = \Bigg(8\mathbf{I}_{2L} - 4\begin{bmatrix}\mathbf{0}_{(2L-1)\times 1} & \mathbf{I}_{2L-1}\\ 0 & \mathbf{0}_{1\times(2L-1)}\end{bmatrix} - \\ 4\begin{bmatrix}\mathbf{0}_{1\times(2L-1)} & 0\\ \mathbf{I}_{2L-1} & \mathbf{0}_{(2L-1)\times 1}\end{bmatrix}\Bigg) \odot \begin{bmatrix}\mathbf{0}_{1\times 2L}\\ \mathbf{1}_{(2L-2)\times 2L}\\ \mathbf{0}_{1\times 2L}\end{bmatrix},\end{aligned} \tag{24}$$

$$\mathbf{T}_k(m) = \mathrm{diag}\left[\jmath\frac{\underline{\hat{h}}_{k,1}^{10}(m)}{\left|\underline{\hat{h}}_{k,1}^{10}(m)\right|^2}, \;\cdots,\; \jmath\frac{\underline{\hat{h}}_{k,2L}^{10}(m)}{\left|\underline{\hat{h}}_{k,2L}^{10}(m)\right|^2}\right]. \tag{25}$$

Here, Re$\{\cdot\}$ is the operator computing the real part of the argument, and $\{\cdot\}^{\dagger}$ denotes the pseudo inverse operator. The detailed derivation of (23) is shown in the Appendix.

B. Extended ℓ_p-RNMCFLMS

The extended ℓ_p-RNMCFLMS aims to improve the identification performance of the steady-state region, which is inspired by [22]. Here, the steady-state represents the region where the NPM converges (Fig. 1). Ahmad *et al*. proposed a dual-filter structure to overcome the misconvergence problem of NMCFLMS by using an estimated critical point from the cost function $J(n) = \sum_{i=1}^{M-1}\sum_{j=i+1}^{M} e_{ij}^2(n)$ in the time domain. In this work, we do not compute parameters such as the critical time, flattening time, or the time interval defined in [22], but only focus on the convergence of $J(n)$ and NPM as updates progress. As shown in Fig. 1, the NPM of ℓ_p-RNMCFLMS algorithm converges well even with low SNR (SNR = 5 dB). We compute the smoothed cost function $\bar{J}(n)$ which is filtered by a $L/4$ tap rectangular window moving average filter (Fig. 1). The trend of $\bar{J}(n)$ converges to $J(n)$ near the point when the NPM reaches a steady-state value. From this observation, the transition time (t_0) is determined to be the first time index when the condition $\bar{J}(n) > J(n)$ is satisfied. After the transition time, the updating procedure of the proposed algorithm changes from ℓ_p-RNMCFLMS to extended ℓ_p-RNMCFLMS as shown in Fig. 2. Whenever the $\bar{J}(n) > J(n)$ condition is satisfied, the estimated RIRs $\hat{\mathbf{h}}_k(m)$ are divided into a significant part $\hat{\mathbf{h}}_k^1(m)$ and an insignificant part $\hat{\mathbf{h}}_k^2(m)$ (Fig. 2). The significant and insignificant parts are computed as

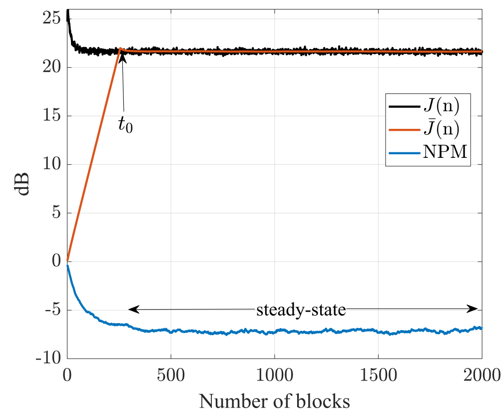


Fig. 1. NPM and cost function in the time domain at SNR = 5 dB (t_0 indicates a transition time, $M = 3$, $L = 1024$, $\mu_f = 0.5$, $p = 1.6$, WGN input.)

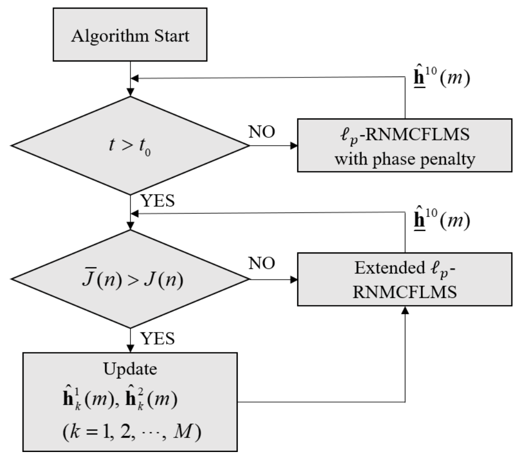


Fig. 2. Flow diagram of the proposed algorithm.

$$\begin{aligned}\hat{h}_{k,l}^1(m) &= \begin{cases}\hat{h}_{k,l}(m), & \left|\hat{h}_{k,l}(m)\right| > \rho_k(m)\\ 0, & \left|\hat{h}_{k,l}(m)\right| \le \rho_k(m)\end{cases}\\ \hat{h}_{k,l}^2(m) &= \begin{cases}0, & \left|\hat{h}_{k,l}(m)\right| > \rho_k(m)\\ \hat{h}_{k,l}(m), & \left|\hat{h}_{k,l}(m)\right| \le \rho_k(m)\end{cases}\\ &\text{for } l = 1,\cdots,2L,\; k = 1,\cdots,M\end{aligned} \tag{26}$$

where $\rho_k(m) = 10^{-1.5}\max\|\hat{\mathbf{h}}_k(m)\|$. The reference magnitude of each channel is chosen as -30 dB from the maximum magnitude of each estimated RIR. The example of the divided responses for $L = 1024$ are shown in Fig. 3. Then, the two parts in the time-domain are transformed to zoomed spectra as

$$\underline{\hat{\mathbf{h}}}_k^{10,\{1,2\}}(m) = \mathbf{F}_{2L}\left[\hat{\mathbf{h}}_k^{\{1,2\}}(m)^T \;\mathbf{0}\right]^T. \tag{27}$$

The extended ℓ_p-RNMFLMS re-computes the gradient (13) and Hessian matrix (15) with different p values ($p_1,\; p_2$) as

$$\begin{aligned}\nabla\mathcal{J}_{\mathrm{FP},k}(m) &= p_1\mathbf{S}_k^{-1}(m)|\underline{\hat{\mathbf{h}}}_k^{10,1}(m)|^{p_1-1}\\ &\odot\exp\{\jmath\arg(\underline{\hat{\mathbf{h}}}_k^{10,1}(m))\},\end{aligned} \tag{28}$$

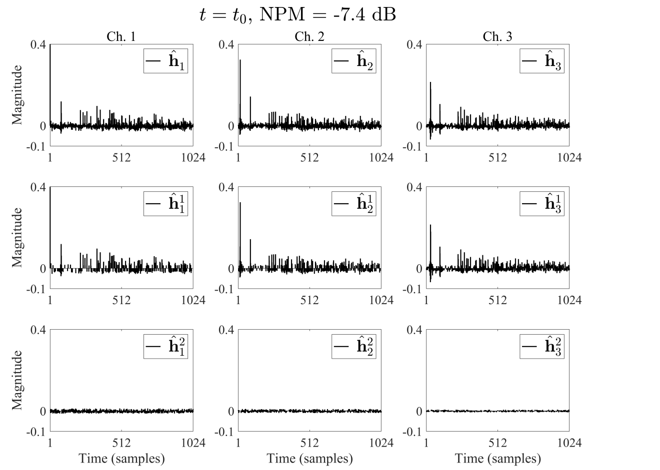


Fig. 3. Examples of estimated impulsed responses of ℓ_p-RNMCFLMS with phase penalty (top), its significant parts (middle), and insignificant parts (bottom) ($M = 3$, $L = 1024$, SNR = 5 dB, $\mu_f = 0.5$, $p = 1.6$, $T_{60} = 500$ ms).

$$\mathbf{S}_k(m) = \mathbf{P}_k(m) - \eta(m)\left\{p_1^2\mathbf{H}_k^1(m) + p_2^2\mathbf{H}_k^2(m)\right\}, \quad \text{for } 1 \le p_1,\, p_2 < 2, \tag{29}$$

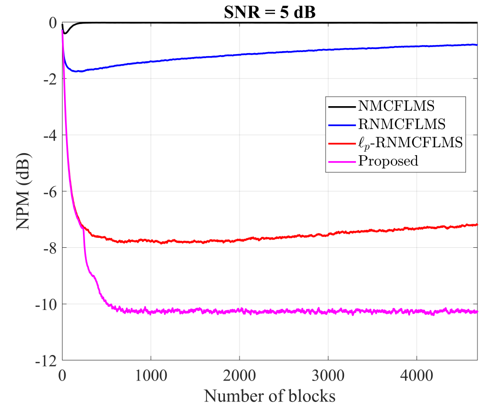
where

$$\mathbf{H}_k^{\{1,2\}}(m) = \operatorname{diag}\left[\left|\underline{\hat{h}}_{k,1}^{10,\{1,2\}}(m)\right|^{p_{\{1,2\}}-2}, \cdots, \left|\underline{\hat{h}}_{k,2L}^{10,\{1,2\}}(m)\right|^{p_{\{1,2\}}-2}\right], \tag{30}$$

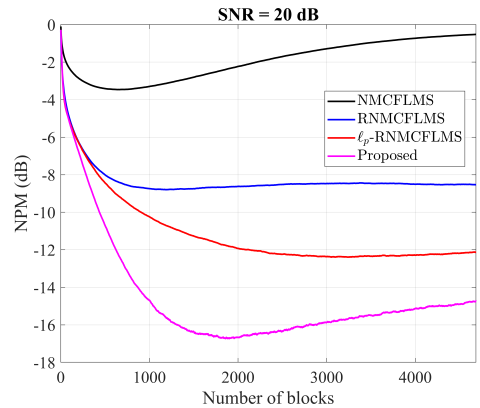
Exclusion of the insignificant parts in the gradient calculation (28) is found empirically from numerical simulations. In other words, we found that the ℓ_p-norm penalty function applied to insignificant parts does not help to reduce errors after the algorithm reaches steady-state ($t > t_0$).

IV. Performance Validation

The proposed algorithm has been evaluated with numerical simulations, and compared to NMCFLMS [14], RNMCFLMS [24] and ℓ_p-RNMCFLMS [27]. The robustness to additive noise is evaluated. The RIR for each microphone was generated by the image-source method [30], and truncated to 1024 samples, which was assumed to be the length of the true RIRs for the BMCI problem. The dimensions of the simulated room were $(9, 7, 5)$ m, and the reverberation time (T_{60}) was set to 500 ms. The omni-directional microphones are linearly positioned at $(5, 2.5, 1.2)$ m, $(5.5, 2.5, 1.2)$ m, and $(6, 2.5, 1.2)$ m. A loudspeaker is positioned at $(4, 4, 2)$ m. The sampling frequency is 16 kHz. The source signal is 5 minutes of white Gaussian noise (WGN), and the additive noise for each microphone is also uncorrelated WGN. Signal-to-noise ratio (SNR) is defined as SNR [dB] $= 10\log_{10}\sigma_s^2\|\mathbf{h}\|^2/M\sigma_n^2$, where σ_n^2 and σ_s^2 are the power of the signal and the noise, respectively [14]. The RIR for each microphone was initialized as $\hat{\mathbf{h}}_k = [1\ 0\ \cdots\ 0]^T, (k = 1, 2, \cdots, M)$ [14]. All results are averaged over 30 trials. For a fair comparison, the p value for ℓ_p-norm was chosen to be $p = 1.6$ which was reported in [27] to be optimal for non-anechoic environments. The step sizes (μ_f) for all algorithms are set to 0.5. After the convergence point, parameters for extended ℓ_p-RNMCFLMS chosen in this simulation were $p_1 = p$ and $p_2 = 1$.



(a)



(b)

Fig. 4. NPM profile for a $M = 3$ channel, $L = 1024$ coefficients of the simulated RIRs with WGN input at (a) SNR $= 5$ dB and (b) 20 dB.

The results for low SNR (5 dB) are shown in Fig. 4 (a). It can be observed that three algorithms except NMCFLMS overcome the mis-convergence problem. The NPMs of the proposed algorithm follow the same profile of ℓ_p-RNMCFLMS until t_0, and the proposed algorithm improves the steady-state performance of ℓ_p-RNMCFLMS after t_0. Likewise, even in a high SNR condition (SNR $= 20$ dB, Fig. 4 (b)), the trends are similar. The proposed algorithm outperforms the other algorithms, which indicates that the proposed algorithm is more robust to additive noise than even the state-of-the-art algorithm [27].

V. Conclusion

In this work, we have proposed the extended ℓ_p-RNMCFLMS algorithm for blind multichannel identification, using a phase-based penalty function and separating parts of RIRs. The squared sum of the phase spectrum and its gradient were derived. The transition time was introduced to improve the steady-state performance by separating the estimated RIRs into significant and insignificant parts after that time and applying different p values. Through numerical simulations, the proposed algorithm has been shown to be robust against additive noise and to improve the steady-state performance.

VI. Appendix

As computed in [27], the gradient of the additional penalty function (20) with respect to $(\hat{\underline{\mathbf{h}}}_k^{10}(m))^*$ is computed as

$$\begin{aligned}\nabla J_{FG,k}(m) &= \frac{2\partial}{\partial\left(\hat{\underline{\mathbf{h}}}_k^{10}(m)\right)^*} J_{FG,k}(m) \\ &= 2\frac{\partial \arg\left(\hat{\underline{\mathbf{h}}}_k^{10}(m)\right)}{\partial\left(\hat{\underline{\mathbf{h}}}_k^{10}(m)\right)^*} \frac{\partial}{\partial \arg\left(\hat{\underline{\mathbf{h}}}_k^{10}(m)\right)} J_{FG,k}(m),\end{aligned} \tag{31}$$

where the diagonal elements of the first term can be computed as

$$\begin{aligned}&\frac{\partial \arg\left(\hat{\underline{h}}_{k,l}^{10}(m)\right)}{\partial\left(\hat{\underline{h}}_{k,l}^{10}(m)\right)^*} = \frac{\partial}{\partial\left(\hat{\underline{h}}_{k,l}^{10}(m)\right)^*}\tan^{-1}\left(\frac{\hat{\underline{h}}_{k,l}^{10,im}(m)}{\hat{\underline{h}}_{k,l}^{10,re}(m)}\right) \\ &= \frac{1}{2}\left\{\frac{-\hat{\underline{h}}_{k,l}^{10,im}(m)}{\left|\hat{\underline{h}}_{k,l}^{10}(m)\right|^2} + j\frac{\hat{\underline{h}}_{k,l}^{10,re}(m)}{\left|\hat{\underline{h}}_{k,l}^{10}(m)\right|^2}\right\} = \frac{j}{2}\frac{\hat{\underline{h}}_{k,l}^{10}(m)}{\left|\hat{\underline{h}}_{k,l}^{10}(m)\right|^2}, \\ &l = 1, 2, \cdots, 2L.\end{aligned} \tag{32}$$

Here, $\hat{\underline{h}}_{k,l}^{10,re} = \mathrm{Re}\{\hat{\underline{h}}_{k,l}^{10*}\}$ and $\hat{\underline{h}}_{k,l}^{10,im} = \mathrm{Im}\{\hat{\underline{h}}_{k,l}^{10*}\}$. The element of the second term of (31) is computed using (20):

$$\begin{aligned}\left[\frac{\partial J_{FG,k}(m)}{\partial \arg\left(\hat{\underline{\mathbf{h}}}_k^{10}(m)\right)}\right]_{k,l} &= 8\arg\left(\underline{h}_{k,l}^{10}(m)\right) \\ &- 4\left\{\arg\left(\underline{h}_{k,l-1}^{10}(m)\right) - \arg\left(\underline{h}_{k,l+1}^{10}(m)\right)\right\},\end{aligned} \tag{33}$$

for $2 \leq l \leq 2L-1$.

References

[1] K. Kowalczyk, E. A. P. Habets, W. Kellermann, and P. A. Naylor, "Blind system identification using sparse learning for TDOA estimation of room reflections," *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 653–656, 2013.

[2] W. Xue, M. Brookes, and P. A. Naylor, "Under-modelled blind system identification for time delay estimation in reverberant environments," in *Proc. Intl. Workshop Acoust. Signal Enhance.*, Xi'an, China, Sep. 2016.

[3] Y. Huang, J. Benesty, and J. Chen, "A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment," *IEEE Trans. Speech, Audio Process.*, vol. 13, no. 5, pp. 882–895, 2005.

[4] X. Lin, A. W. Khong, and P. A. Naylor, "A forced spectral diversity algorithm for speech dereverberation in the presence of near-common zeros," *IEEE Trans. Audio, Speech, Lang. Process*, vol. 20, no. 3, pp. 888–889, 2011.

[5] R. Guidorzi, R. Diversi, and U. Soverini, "Blind identification and equalization of multichannel FIR systems in unbalanced noise environments," *Signal Process.*, vol. 87, no. 4, pp. 654–664, 2007.

[6] X. Li, S. Gannot, L. Girin, and R. Horaud, "Multichannel identification and nonnegative equalization for dereverberation and noise reduction based on convolutive transfer function," *IEEE/ACM Trans. Audio, Speech, Lang. Process*, vol. 26, no. 10, pp. 1755–1768, 2018.

[7] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Proc. 25th Asilomar Conf.*, Pacific Grove, CA, 1991, pp. 856–860.

[8] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516–525, 1995.

[9] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: a time domain approach," *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 340–349, 1994.

[10] H. H. Zeng and L. Tong, "Blind channel estimation using the second-order statistics: Asymptotic performance and limitations," *IEEE Trans. Signal Process.*, vol. 45, no. 8, pp. 2060–2071, 1997.

[11] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, 1995.

[12] Q. Zhao and L. Tong, "Adaptive blind channel estimation by least squares smoothing," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 3000–3012, 1999.

[13] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Process.*, vol. 44, no. 3, pp. 661–672, 1996.

[14] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 11–24, 2003.

[15] Y. Huang and J. Benesty, "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification," *Signal Process.*, vol. 82, no. 8, pp. 1127–1138, 2002.

[16] M. Hasan and P. A. Naylor, "Analyzing effect of noise on LMS-type approaches to blind estimation of SIMO channels: Robustness issue," in *Proc. Eur. Signal Process. Conf.*, Florence, Italy, Sep. 2006.

[17] Y. Huang, J. Benesty, and J. Chen, "Optimal step size of the adaptive multichannel LMS Algorithm for blind SIMO identification," *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 173–176, 2005.

[18] M. Haque and M. Hasan, "Variable step size frequency domain multichannel LMS algorithm for blind channel identification with noise," in *Proc. Comm. Systems, Networks and Digital Signal Process.*, Patras, Greece, Jul. 2006.

[19] M. Haque and M. Hasan, "Performance comparison of the blind multichannel frequency domain normalized LMS and variable step-size LMS with noise," in *Proc. Eur. Signal Process. Conf.*, Poznań, Poland, Sep. 2007.

[20] M. Hasan, J. Benesty, P. A. Naylor, and D. B. Ward, "Improving robustness of blind adaptive multichannel identification algorithms using constraints," in *Proc. Eur. Signal Process. Conf.*, Antalya, Turkey, Sep. 2005.

[21] R. Ahmad, A. W. Khong, and P. A. Naylor, "Proportionate frequency domain adaptive algorithms for blind channel identification," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, Toulouse, France, May. 2006.

[22] R. Ahmad, N. D. Gaubitch, and P. A. Naylor, "A noise-robust dual filter approach to multichannel blind system identification," in *Proc. Eur. Signal Process. Conf.*, Poznań, Poland, Sep. 2007.

[23] M. Haque, M. S. Al Bashar, P. A. Naylor, K. Hirose, and M. Hasan, "Energy constrained frequency-domain normalized LMS algorithm for blind channel identification," *Signal, Image and Video Process.*, vol. 1, no. 3, pp. 203–213, 2007.

[24] M. Haque and M. Hasan, "Noise robust multichannel frequency-domain LMS algorithms for blind channel identification," *IEEE Signal Process. Lett.*, vol. 15, pp. 305–308, 2008.

[25] M. Haque and M. Hasan, "Robust multichannel least mean square-type algorithms with fast decaying transient for blind identification of acoustic channels," *IET Signal Process.*, vol. 2, no. 4, pp. 431–441, 2008.

[26] M. Haque, T. Islam, and M. Hasan, "Robust speech dereverberation based on blind adaptive estimation of acoustic channels," *IEEE Trans. Audio, Speech, Lang. Process*, vol. 19, no. 4, pp. 775–787, 2010.

[27] H. He, J. Chen, J. Benesty, and T. Yang, "Noise robust frequency-domain adaptive blind multichannel identification with ℓ_p-Norm constraint," *IEEE/ACM Trans. Audio, Speech, Lang. Process*, vol. 26, no. 9, pp. 1608–1619, 2018.

[28] K. Hasan, "Damped variable step size multichannel Wiener LMS algorithm for blind channel identification with noise," in *Proc. Int. Symp. Commun. Syst., Netw., Digit. Signal Process.*, Patras, Greece, Jul. 2006.

[29] D. R. Morgan, J. Benesty, and M. Mohan Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 174–176, 1998.

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