Convex Combination of MIMO Filters for Multichannel Acoustic Echo Cancellation

Danilo Comminiello, Simone Scardapane, Michele Scarpiniti, Raffaele Parisi, and Aurelio Uncini
Department of Information Engineering, Electronics and Telecommunications (DIET)
"Sapienza" University of Rome
Via Eudossiana 18, 00184 Rome, Italy
Email: danilo.comminiello@uniroma1.it

***Abstract*—This paper introduces a new framework for multichannel adaptive filtering, aiming at improving performance of an overall filtering system. The proposed architecture relies on the properties of the adaptive combination of filters which exploits the capabilities of different constituents, thus adaptively providing at least the behaviour of the best performing filter. Applying this concept to multichannel filtering systems, we define a scheme for the combination of *multiple-input multiple-output* (MIMO) filters. More precisely the proposed structure involves the combination of two different *multiple-input single-output* (MISO) systems for each MIMO output. We propose such framework with application to the *multichannel acoustic echo cancellation* (MAEC) with the goal of giving robustness to the system against impulsive background noise, and thus improving overall cancelling performance. Experimental results show the effectiveness of the proposed combined MAEC in the presence of adverse environmental conditions.**

I. Introduction

The growing availability of resources for data transmission and processing is encouraging the spread of hands-free speech communication systems which offer users an *immersive experience* [1]–[3]. An immersive communication scenario is characterized by the use of multiple microphones and loudspeakers, which aim at capturing and reproducing desired signals while focusing on preserving the audio quality perceived by users. In such context, the presence of interfering sources and reverberation may cause a quality degradation of speech communications. In order to address this problems, microphones and loudspeakers are usually connected with signal processing systems, thus resulting in *intelligent acoustic interfaces* [3], [4]. One of the main problems of immersive scenarios is the *multichannel acoustic echo cancellation* (MAEC), which is caused by the multiple coupling between microphones and loudspeakers. MAEC may be seen as a straightforward generalization of the monophonic acoustic echo cancellation (AEC), but it entails more problems to tackle, such as nonuniqueness and slow convergence due to inter-channel correlation, and the poor ability to react to changes in the environmental conditions [5]. Such issues have roused remarkable interest over the years [6]–[9].

The effectiveness of an MAEC strictly relies on the design of a *multiple-input multiple-output* (MIMO) filtering system, whose main task is to estimate several acoustic impulse responses (AIRs), depending on the number of microphones and loudspeakers. It results in a large number of coefficients to adapt, therefore an appropriate choice of the adaptive algorithm becomes essential. Generally, in the time-domain, first-order adaptive algorithms, such as the *least mean squares* (LMS), are very attractive due to their simplicity and low computational cost. However, LMS-based algorithms do not take into account the cross-correlation statistics of the input signals, thus resulting in poor convergence performance [5]. Hessian-based algorithms, such as the *recursive least squares* (RLS), improve convergence abilities since they consider inter-channel correlations. Unfortunately, the RLS entails a larger computational cost and, moreover, it may perform worse than first-order algorithms in adverse environments [10]. This is the reason why here we consider the *affine projection algorithm* (APA) to adapt the MIMO filters, since it can be seen as a generalization of the *normalized LMS* (NLMS) algorithm, in which cross-correlations of the input signals are involved [11].

However, even a multichannel APA may suffer from adverse conditions, especially in the presence of impulsive noise or in changing environments, which may make the adaptation process unstable and reduce performance. In order to tackle this problem, we introduce a new adaptation framework which relies on the adaptive combination of MIMO filtering systems. Adaptive combination of filters is capable of automatically switching between constituents according to the best performing filter, thus always providing the best possible performance [12]. Combined adaptive schemes are usually adopted with filters of the same family and complementary properties, but also with filters of different families or using different updating rules [13]–[15]. Recently, combination of filters was successfully applied to *multiple-input single-output* (MISO) systems [16]–[18] with application to adaptive beamforming for noise reduction.

In this paper we exploit the properties of adaptive combination of filters to perform a combination of MIMO filtering systems. In particular, we combine two MIMO systems, each one having different adaptation settings. We describe in detail a possible combining architecture for MIMO filters and we apply such technique to MAEC. As a consequence, an improvement of the overall performance of the canceller is expected, especially in the presence of any impulsive noise that may cause a change in the environment.

This paper is organized as follows: Section II describes the MAEC framework using the combined MIMO architecture updated by using the multichannel APA. The combination of adaptive MIMO filters is proposed in Section III. Section IV contains the evaluation of the effectiveness of the proposed framework with application to MAEC, and some conclusion are finally drawn in Section V.

II. Combined Multichannel Acoustic Echo Cancellation

The MAEC framework including the combination of MIMO filtering systems is depicted in Fig. 1 and it is referred to as *combined MAEC* (CMAEC). At n-th time instant, P speech signals coming from the remote environment, also known as *far-end*, and denoted as $x_p[n]$, with $p = 1, \ldots, P$, arrive at the other side of the full-duplex communication, or also said *near-end*. The far-end signals are reproduced by P loudspeakers and then acquired by Q microphones. The acoustic coupling between microphones and loudspeakers is characterized by $P \cdot Q$ acoustic paths, which also contain information about environment reverberations. The *desired signals* $d_q[n]$ $(q = 1, \ldots, Q)$ acquired by the microphones represent the echo signals, which may be possibly superimposed on any near-end contribution, containing the near-end speech signal $s[n]$ with the addition of background noise $v[n]$. At the same time, the far-end signals $x_p[n]$ are processed by the CMAEC in order to estimate the AIRs between microphones and loudspeakers. In particular, each input signal $x_p[n]$ is collected in an input data matrix $\mathbf{X}_{n,p} \in \mathbb{R}^{K \times M}$:

$$\mathbf{X}_{n,p} = \begin{bmatrix} \mathbf{x}_{n,p}^T \\ \mathbf{x}_{n-1,p}^T \\ \vdots \\ \mathbf{x}_{n-K+1,p}^T \end{bmatrix} \tag{1}$$

$$= \begin{bmatrix} x_p[n] & \cdots & x_p[n-M+1] \\ x_p[n-1] & \cdots & x_p[n-M-2] \\ \vdots & \ddots & \vdots \\ x_p[n-K+1] & \cdots & x_p[n-M-K-2] \end{bmatrix}$$

where M is the length of the adaptive filters and K denotes the number of previous entries to keep in memory, i.e. the *projection order*. Data matrices can be represented by a unique input data matrix $\mathbf{X}_n \in \mathbb{R}^{K \times MP}$:

$$\mathbf{X}_n = [\ \mathbf{X}_{n,1} \quad \mathbf{X}_{n,2} \quad \ldots \quad \mathbf{X}_{n,P}\ ]^T. \tag{2}$$

The input matrix is processed by the CMAEC, which is composed of J MIMO filtering systems. In this paper we take into account the combination of $J = 2$ MIMO filters, but it can be also generalized to the case of $J > 2$. Each MIMO filter is represented by the following matrix $\mathbf{W}_n^{(j)} \in \mathbb{R}^{MP \times Q}$, with $j = 1, 2$:

$$\mathbf{W}_n^{(j)} = \begin{bmatrix} \mathbf{w}_{n,11}^{(j)} & \mathbf{w}_{n,12}^{(j)} & \cdots & \mathbf{w}_{n,1Q}^{(j)} \\ \mathbf{w}_{n,21}^{(j)} & \mathbf{w}_{n,22}^{(j)} & \cdots & \mathbf{w}_{n,2Q}^{(j)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{w}_{n,P1}^{(j)} & \mathbf{w}_{n,P2}^{(j)} & \cdots & \mathbf{w}_{n,PQ}^{(j)} \end{bmatrix} \tag{3}$$

where inter-channel paths are also taken into account. Each individual adaptive filter $\mathbf{w}_{n,pq} \in \mathbb{R}^{M \times 1}$ included in (3) can be expressed as:

$$\mathbf{w}_{n,pq}^{(j)} = \begin{bmatrix} w_{pq,0}^{(j)}[n] & w_{pq,1}^{(j)}[n] & \ldots & w_{pq,M-1}^{(j)}[n] \end{bmatrix}^T. \tag{4}$$

The output of each MIMO filter $\mathbf{Y}_n^{(j)} \in \mathbb{R}^{K \times Q}$ can be easily achieved:

$$\mathbf{Y}_n^{(j)} = \mathbf{X}_n^{(j)} \mathbf{W}_{n-1}^{(j)} \tag{5}$$

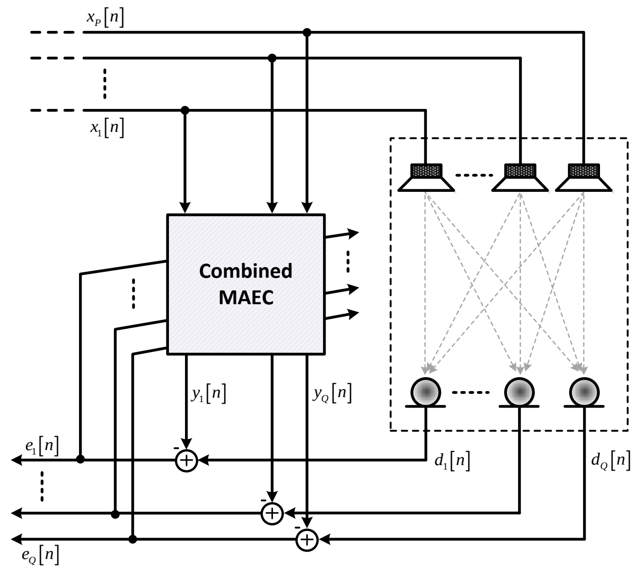

Fig. 1. Combined MAEC system.

that can be also seen as a concatenation of Q individual output vectors, i.e. $\mathbf{Y}_n^{(j)} = \begin{bmatrix} \mathbf{y}_{n,1}^{(j)} & \mathbf{y}_{n,2}^{(j)} & \ldots & \mathbf{y}_{n,Q}^{(j)} \end{bmatrix}^T$, for $j = 1, 2$. Taking into account the matrix of the desired signals $\mathbf{D}_n \in \mathbb{R}^{K \times Q}$, i.e:

$$\mathbf{D}_n = \begin{bmatrix} d_1[n] & \cdots & d_Q[n] \\ d_1[n-1] & \cdots & d_Q[n-1] \\ \vdots & \ddots & \vdots \\ d_1[n-K+1] & \cdots & d_Q[n-K+1] \end{bmatrix}, \tag{6}$$

for each MIMO filter it is possible to achieve the error signals, which are contained in the following error matrix:

$$\mathbf{E}_n^{(j)} = \mathbf{D}_n - \mathbf{Y}_n^{(j)}. \tag{7}$$

Similarly to (5), $\mathbf{E}_n^{(j)}$ can be seen as a concatenation of Q individual error vectors, i.e. $\mathbf{E}_n^{(j)} = \begin{bmatrix} \mathbf{e}_{n,1}^{(j)} & \mathbf{e}_{n,2}^{(j)} & \ldots & \mathbf{e}_{n,Q}^{(j)} \end{bmatrix}^T$.

Each MIMO filter is individually updated according to the *multichannel affine projection algorithm* [11]:

$$\mathbf{W}_n^{(j)} = \mathbf{W}_{n-1}^{(j)} + \mu_j \mathbf{X}_n^T \left(\delta_j + \mathbf{X}_n \mathbf{X}_n^T\right)^{-1} \mathbf{E}_n^{(j)} \tag{8}$$

where μ_j and δ_j are respectively the *step size* parameter and the *regularization factor*, which are the same for all the filters of the j-th MIMO system.

In the next section, a possible adaptive scheme is derived for the combination of the two MIMO systems, thus achieving the overall error signals for the CMAEC.

III. A Convex Combination Scheme for MIMO Filters

The trademark of the proposed system lies in the CMAEC, which involves the combination of the two MIMO filters. These systems can be differentiated in several ways according to their update rule. In this paper we distinguish the two MIMO filters by choosing different values for the step size parameter, a small value for $j = 1$ and a large one for $j = 2$, according to [12], [13]. As regards the kind of adaptive combination, in this work we focus on the convex constrained combination using a sigmoid nonlinearity on the output stage, since it introduces less gradient noise compared to unconstrained and affine constrained combinations.

The chosen method for combining MIMO filters is based on the *system-by-system* (SS) combination scheme [16] and it is represented in Fig. 2. An MIMO system can be seen as a parallel of Q MISO systems, each receiving the same P inputs. The idea which underpins the SS scheme is to combine the outputs of each pair of MISO systems, for all the Q output channels. The output of each MISO system for the q-th MIMO output channel, that we denote as $y_q^{(j)}[n]$, is achieved by summing each filter output of the corresponding MISO system, i.e.:

$$y_q^{(j)}[n] = \sum_{p=1}^{P} \mathbf{x}_{n,p}^T \mathbf{w}_{n-1,pq}^{(j)}. \tag{9}$$

The two MISO system outputs related to the q-th output channel are then convexly combined generating the q-th MIMO output:

$$y_q[n] = \lambda_q[n] y_q^{(1)}[n] + (1 - \lambda_q[n]) y_q^{(2)}[n] \tag{10}$$

where $\lambda_q[n]$ is the q-th *mixing parameter*, which is constrained to remain in the range $[0, 1]$ [12]. It is worth noting that, in order to not further increase the computational complexity, in (10) we do not consider the last K output samples but only the current sample. The q error signal $e_q[n]$ after the combination is achieved as:

$$e_q[n] = d[n] - y_q[n]. \tag{11}$$

The mixing parameters in (10) are usually updated by using an auxiliary parameter, $a_q[n]$, related to $\lambda_q[n]$ by the expression:

$$\lambda_q[n] = \mathrm{sgm}(a_q[n]) = \frac{1}{1 + e^{-a[n]}}. \tag{12}$$

The update rule of the auxiliary parameter can be defined according to [12], [19]:

$$a_q[n+1] = a_q[n] + \frac{\mu_a}{r_q[n]} e_q[n] \Delta e_q[n] \lambda_q[n] (1 - \lambda_q[n]) \tag{13}$$

where

$$\Delta e_q[n] = e_q^{(2)}[n] - e_q^{(1)}[n], \tag{14}$$

and

$$r_q[n] = \eta r_q[n-1] + (1 - \eta) \Delta e_q^2[n] \tag{15}$$

is the estimated power of $\Delta e_q[n]$; η is a smoothing factor. Therefore, $\mu_a / r_q[n]$ in (13) represents a normalized step size. By using the adaptation rule (13), the combination scheme is able to adaptively recognize and select the best performing filters, thus improving the overall performance of the CMAEC.

The computational complexity of the CMAEC is subjected to several parameters, but most of all it depends on the adopted

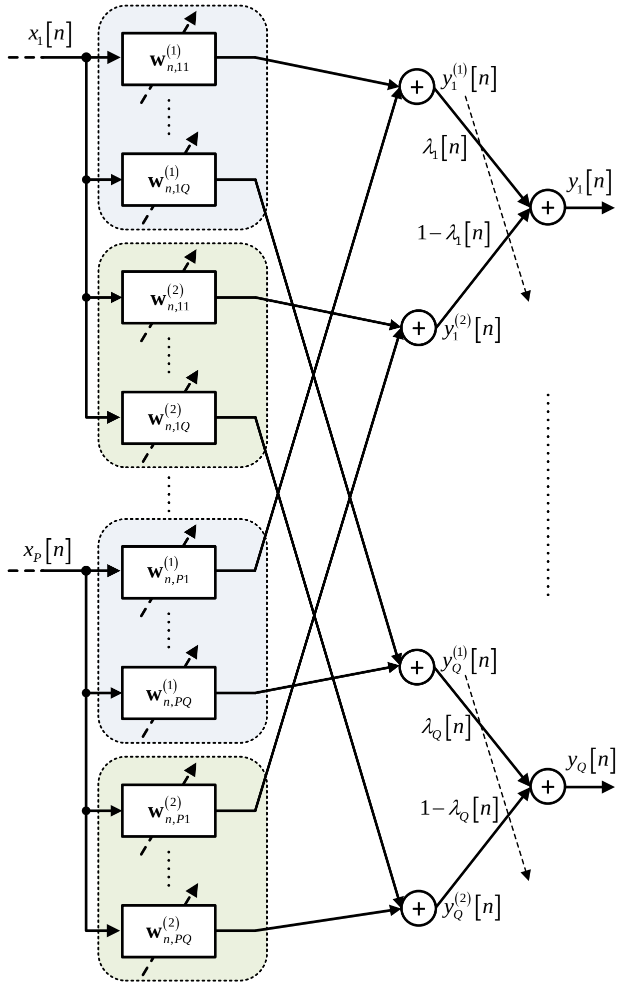

Fig. 2. System-by-system scheme of the combined MAEC.

adaptive algorithm and the number of channels. In particular, the overall cost of a CMAEC is exactly the double of a MAEC with the addition of an overhead due to the convex combinations. Such overhead consists of $8Q$ multiplications, $3Q$ additions and Q function evaluations. The reduction of the computational complexity of the CMAEC can be taken into account in future developments, in particular in the choice of the adaptive algorithm, which represents the most burdensome component.

IV. Experimental Results

In this section we present two sets of experiments aiming at assessing the performance of the proposed combined MIMO system. In order to have an exhaustive view of the capabilities of the system, we first analyse its convergence performance and then its effectiveness in an MAEC application.

A. Convergence Performance of the Combined MIMO Filter

In the first set of experiments we assess the convergence behaviour of the proposed combined MIMO filter in terms of the *excess mean square error* (EMSE), defined as the

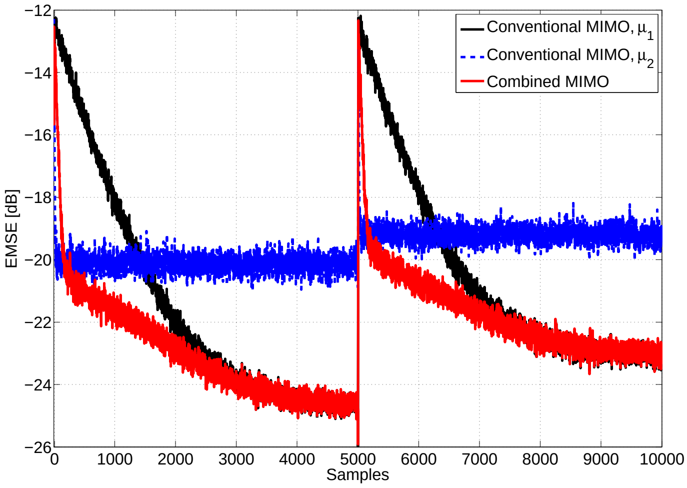

Fig. 3. Convergence performance in terms of the EMSE.

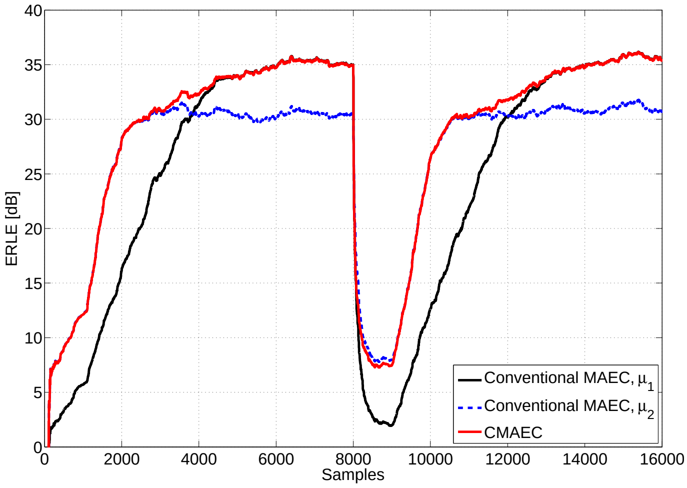

Fig. 4. Performance evaluation in terms of the ERLE.

excess over the minimum mean square error. The scenario involves an unknown multichannel system to identify, which is composed of $P = 4$ input channels and $Q = 4$ output channels. Each inter-channel impulse response is formed by $M = 7$ independent random values between -1 and 1. The input signal is generated by means of a first-order autoregressive model, whose transfer function is $\sqrt{1-\alpha^2}/\left(1-\alpha z^{-1}\right)$, with $\alpha = 0.8$, fed with an iid Gaussian random process. The length of the input signal is $L = 10000$ samples. In order to evaluate the ability of the proposed architecture to react to possible alterations in the environment, the impulse responses to estimate are randomly reassigned at the time instant $n = L/2$. An additive iid noise signal $v[n]$ with variance $\sigma_v^2 = 0.01$ is added at the output of each channel.

In order to identify the unknown $PQ = 16$ impulse responses we adopt the proposed combined MIMO filter, which is composed of two MIMO filters, one using a small step size $\mu_1 = 0.001$ and the other one using a larger value $\mu_2 = 0.1$. Convergence performance is evaluated in terms of the EMSE, which can be defined (in dB) for the multichannel case as:

$$\mathrm{EMSE}[n] = 10\log\left(\mathrm{E}\left\{\left(\frac{1}{Q}\sum_{q=1}^{Q} e_q[n] - v[n]\right)^2\right\}\right) \quad (16)$$

where the operator $\mathrm{E}\{\cdot\}$ is the mathematical expectation; indeed the EMSE is evaluated over 1000 independent runs. Results for the combined MIMO filter are compared with those of individual MIMO filters using respectively μ_1 and μ_2. The combined MIMO filter uses a step size value $\mu_a = 0.5$ and the following initial setting for the adaptation of the auxiliary parameter: $a_q[0] = 0$, $r_q[0]$; the smoothing factor in (15) is chosen as $\eta = 0.9$. All the systems use the same projection order $K = 4$ and the same regularization factor $\delta = 0.001$.

Convergence results are depicted in Fig. 3, where it is worth noting that the MIMO filter with μ_1 shows a slow convergence rate but a good precision at steady state, while the MIMO filter with μ_2 provides faster convergence rate but lower precision. The combined MIMO filter is capable of exploiting the advantages of both the MIMO filters, thus showing fast convergence rate and good precision at steady state. Moreover, it can be notice that, in proximity of the intersection between the two conventional systems, the proposed combined MIMO filter shows even a performance gain with respect to the conventional filters.

B. *Evaluation of the Combined MAEC*

In the second set of experiments we use the prosed combined architecture for an MAEC application. The experimental scenario is that of a reverberant room with size $6 \times 5 \times 3,3$ m where multiple loudspeakers and microphones are placed. In particular, a *uniform linear array* (ULA) of $P = 8$ loudspeakers is adopted with a spacing of 20 cm. In front of the loudspeaker array, at a distance of $4,5$ m, is located a ULA of $Q = 8$ microphones, in which the spacing is 5 cm. The $PQ = 64$ acoustic impulse responses (AIRs) are measured by an image source method, using a reflection factor of the walls of $\rho = 0.78$ with a sampling rate of 8 kHz. Each AIR is truncated after $M = 280$ samples. As far-end input is used a coloured signal generated by using the same autoregressive model described in the previous set of experiments. The length of the input signal is $L = 16000$ samples. An additive iid white Gaussian noise signal $v[n]$ is added to each microphone signal in order to provide 20 dB of *signal to noise ratio* (SNR). In order to introduce an abrupt change in the acoustic environment we shift the AIRs circularly to the right by 30 samples, at the time instant $n = L/2$.

Similarly to the previous experiment, we compare the performance of two conventional MIMO filters with the combination of them, i.e. the CMAEC. Again we differentiate the two MIMO filters according to their step size value. In particular, we choose a small value for the first MIMO filter, $\mu_1 = 0.1$, and a large value for the second one, $\mu_2 = 1$. The rest of the parameter setting for all the systems is the same of the previous setup. In this case, the performance is evaluated in terms of the *echo return loss enhancement* (ERLE), which can be defined (in dB) for the multichannel case as:

$$\begin{aligned}\mathrm{ERLE}\left[n\right] &= \frac{1}{Q}\sum_{q=1}^{Q}\mathrm{ERLE}_q\left[n\right] \\ &= 10\log\left(\frac{1}{Q}\sum_{q=1}^{Q}\frac{\mathrm{E}\left\{d_q^2\left[n\right]\right\}}{\mathrm{E}\left\{e_q^2\left[n\right]\right\}}\right)\end{aligned} \tag{17}$$

Results are depicted in Fig. 4, in which it is possible to notice the behaviour of the proposed CMAEC which always follows the best performing MIMO filter and outperforms the conventional systems in correspondence of their performance intersection. This proves the effectiveness of the proposed combined MIMO architecture for MAEC applications.

V. Conclusion

This paper introduces a new framework for MIMO filtering in MAEC applications. The proposed approach is based on the adaptive combination of MIMO filters, each one having different settings and, hence, different capabilities. The combined MIMO is capable of exploiting the best properties of each MIMO filter, thus providing superior performance. The combination can be achieved according to a system-by-system scheme, which guarantees robustness to the MAEC against abrupt changes in the environment. Experimental results have shown the effectiveness of the system in multichannel scenarios. The proposed framework paves the way for future works, since results could be further improved by involving other kinds of adaptive algorithms for each filter, e.g. in the frequency-domain, or other combination schemes, e.g. the filter-by-filter one, or also other constrained algorithms for the adaptation of the mixing parameters.

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] C. Kyriakakis, "Fundamental and technological limitations of immersive audio systems," *Proc. of the IEEE*, vol. 86, no. 5, pp. 941–951, 1998.

[3] D. Comminiello, M. Scarpiniti, R. Parisi, and A. Uncini, "Intelligent acoustic interfaces for immersive audio," in *134th AES Convention*, Rome, May 2013.

[4] D. Comminiello, "Adaptive algorithms for intelligent acoustic interfaces," Ph.D. dissertation, 'Sapienza' University of Rome, Dec. 2011.

[5] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing*. Berlin: Springer-Verlag, 2006.

[6] J. Benesty, D. Morgan, and M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 6, no. 2, pp. 156–165, Mar. 1998.

[7] H. Buchner, J. Benesty, and W. Kellermann, "Generalized multichannel frequency-domain adaptive filtering: Efficient realization and application to hands-free speech communication," *Signal Process.*, vol. 85, no. 3, pp. 549–570, Mar. 2005.

[8] S. Cecchi, L. Romoli, P. Peretti, and F. Piazza, "A combined psychoacoustic approach for stereo acoustic echo cancelllation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 6, pp. 1530–1539, Aug. 2011.

[9] D.-Q. Nguyen, W.-S. Gan, and A. W. H. Khong, "Time-reversal approach to the stereophonic acoustic echo cancellation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 2, pp. 358–395, Feb. 2011.

[10] E. Eweda, "Comparison of RLS, LMS and sign algorithms for tracking randomly time-varying channels," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 2937–2944, Nov. 1994.

[11] J. Benesty, P. Duhamel, and Y. Grenier, "A multichannel affine projection algorithm with applications to multichannel acoustic echo cancellation," *IEEE Signal Process. Lett.*, vol. 3, no. 2, pp. 35–37, Feb. 1996.

[12] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1078–1090, Mar. 2006.

[13] J. Arenas-García, M. Martínez-Ramón, A. Navia-Vázquez, and A. R. Figueiras-Vidal, "Plant identification via adaptive combination of transversal filters," *Signal Process.*, vol. 86, no. 9, pp. 2430–2438, Sep. 2006.

[14] N. J. Bershad, J. C. M. Bermudez, and J. Tourneret, "An affine combination of two LMS adaptive filters – Transient mean-square analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 853–1864, May 2008.

[15] M. T. M. Silva and V. H. Nascimento, "Improving the tracking capability of adaptive filters via convex combination," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3137–3149, Jul. 2008.

[16] D. Comminiello, M. Scarpiniti, R. Parisi, and A. Uncini, "Combined adaptive beamforming schemes for nonstationary interfering noise reduction," *Signal Process.*, 2013, 10.1016/j.sigpro.2013.05.014.

[17] ——, "Combined adaptive beamforming techniques for noise reduction in changing environments," in *Proc. IEEE Int. Conf. Telecom. Signal Process. (TSP '13)*, Rome, Jul. 2-4 2013.

[18] D. Comminiello, M. Scarpiniti, R. Parisi, A. Cirillo, M. Falcone, and A. Uncini, "Multi-stage collaborative microphone array beamforming in presence of nonstationary interfering signals," in *Proc. Int. Works. Machine List. Multisource Environ. (CHiME '11)*, Florence, Italy, Sep. 1 2011, pp. 64–67.

[19] L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, and J. Arenas-García, "A normalized adaptation scheme for the convex combination of two adaptive filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '08)*, Las Vegas, NV, USA, Mar. 2008, pp. 3301–3304.