A NOVEL DECORRELATION APPROACH FOR AN ADVANCED MULTICHANNEL ACOUSTIC ECHO CANCELLATION SYSTEM

Laura Romoli⋆, *Stefania Cecchi*⋆, *Danilo Comminiello*†, *Francesco Piazza*⋆, *Aurelio Uncini*†

⋆ DII - Università Politecnica delle Marche - Italy
† DIET, "Sapienza" University of Rome - Italy

ABSTRACT

A multichannel sound reproduction system aims at offering an immersive experience exploiting multiple microphones and loudspeakers. In the case of multichannel acoustic echo cancellation, a suitable solutions for overcoming the well-known non-uniqueness problem and an appropriate choice of the adaptive algorithm become essential to improve the audio reproduction quality. In this paper, an advanced system is proposed based on the introduction of a multichannel decorrelation solution exploiting the missing-fundamental phenomenon and a combined multiple-input multiple-output architecture updated by using the multichannel affine projection algorithm. Experimental results proved the effectiveness of the presented framework in terms of objective and subjective measures, providing a suitable solution for echo cancellation.

***Index Terms*—** Multichannel Acoustic Echo Cancellation, Channel Decorrelation, Adaptive Combination of Filters

1. INTRODUCTION

A multichannel sound reproduction system aims at offering an immersive experience to the listener by using multiple microphones and loudspeakers for capturing and reproducing the desired signals while focusing on preserving the audio quality perceived by users [1]. Adaptive signal processing systems are often introduced to overcome quality degradation due to interfering sources and reverberation, thus resulting in intelligent acoustic interfaces [2]. Multichannel acoustic echo caused by the multiple coupling between microphones and loudspeakers is one of the main problems of immersive scenarios. Differently from the single channel scenario, the linear relationship existing among the loudspeaker signals worsens the performance and a method to reduce interchannel coherence must be introduced [3].

Several efforts have been done in the field of stereophonic signal decorrelation as summarized in [3, 4]. The generalization to multichannel signal may not be much straightforward. Some approaches have been applied to the multichannel scenario, e.g, the phase modulation approach designed for surround sound systems like 5.1 [5] and the half-wave rectifier distortion producing a high impact on the audio quality [6]. A novel approach for multichannel audio decorrelation has been presented in [4], starting from the results obtained for stereophonic signals [3, 7, 8], based on the method derived from the missing fundamental theory combined with the introduction of time-varying allpass filters.

Then, the effectiveness of a multichannel acoustic echo canceller (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. A large number of coefficients has to be adapted, therefore an appropriate choice of the adaptive algorithm becomes essential [1, 9]. Several adaptive algorithms have been studied in the literature. The time-domain first-order adaptive algorithms as the least mean squares (LMS) are very attractive due to their simplicity and low computational cost. Hessian-based algorithms, such as the recursive least squares (RLS), improve convergence abilities but entails a larger computational cost and, moreover, it may perform worse than first-order algorithms in adverse environments [10]. The affine projection algorithm (APA) could be adopted to adapt MIMO filters, since it can be seen as a generalization of the normalized LMS (NLMS) algorithm involving cross-correlations of the input signals [11]. Moreover, adaptive combination of filters is capable of automatically switching between constituents according to the best performing filter, thus providing the best possible performance [12]. Recently, combination of filters was successfully applied to multiple-input single-output (MISO) systems for noise reduction in adaptive beamforming [13–15] and to MIMO systems for echo cancellation [1].

In this paper, an advanced multichannel acoustic echo cancellation system is presented based on a novel multichannel decorrelation approach for reducing the interchannel coherence and on convex combination of MIMO filters for guaranteeing optimal filtering performance.

The paper is organized as follows. The advanced multichannel acoustic echo cancellation system is presented in Section 2 focusing on both the multichannel decorrelation procedure (Section 2.1) and the convex combination of MIMO filters (Section 2.2). Then, the validation of the proposed solution is discussed in Section 3 and some concluding remarks are reported in Section 4.

2. PROPOSED ALGORITHM

An advanced multichannel acoustic echo cancellation system is presented based on a novel multichannel decorrelation approach exploiting the missing-fundamental theory and on convex combination of MIMO filters for improving the performance of AIRs identification. A block diagram of the proposed framework is reported in Fig. 1. At the n-th time instant, P far-end speech signals, denoted as $u_p[n]$, with $p = 1, \ldots, P$, arrive at the near-end side where they are processed by the proposed multichannel decorrelator. The decorrelated signals $x_p[n]$, with $p = 1, \ldots, P$, are reproduced by P loudspeakers and then acquired by Q microphones. The acoustic coupling between microphones and loudspeakers is characterized by $P \cdot Q$ AIRs, 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 with the addition of background noise. At the same time, the decorrelated signals $x_p[n]$ are processed by the combined MAEC (CMAEC) in order to estimate the AIRs between microphones and loudspeakers. In the following, the main aspects of the aforementioned issues are discussed.

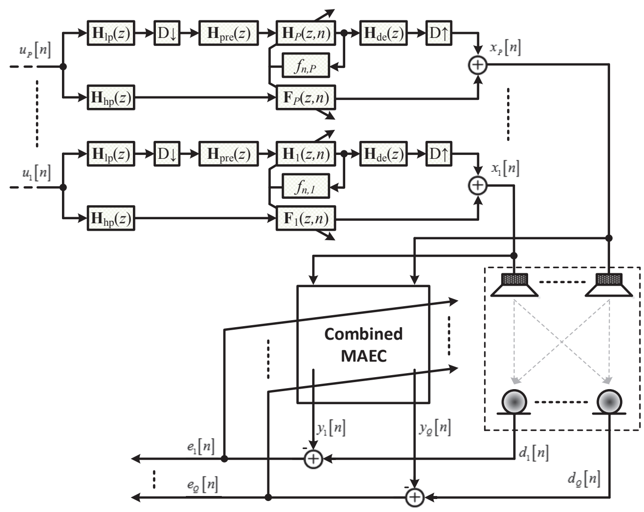


Fig. 1. Proposed advanced multichannel system.

2.1. Multichannel Decorrelation

A novel channel decorrelation solution for multichannel reproduction systems has been discussed in [4] based on the combination of the missing-fundamental theory with time-varying allpass filters. In this paper, the approach is applied to multichannel acoustic echo cancellation systems for the first time. Each far-end signal $u_p[n]$, with $p = 1, \ldots, P$, is divided into two subbands using the low-pass filter $\mathbf{H}_{\mathrm{lp}}(z)$ and the high-pass filter $\mathbf{H}_{\mathrm{hp}}(z)$, respectively. Then, an adaptive notch filter $\mathbf{H}_p(z, n)$ is applied in the low-frequency range for estimating and removing the fundamental frequency while in the high-frequency range the signal phase is altered through a second-order time-varying allpass filter $\mathbf{F}_p(z, n)$ varying with the estimated fundamental frequency $f_{n,p}$.

In the low-frequency band, the “missing-fundamental” phenomenon is exploited [4]. While in the two-channel scenario the adaptive notch filter was applied only on one channel of the stereo signal, in the multichannel scenario the fundamental frequency is estimated and removed from all the $p = 1, \cdots, P$ input channels. The p-th second-order lattice form notch filter is described as follows:

$$\mathbf{H}_p(z, n) = \frac{1 + 2k_{n,p}z^{-1} + z^{-2}}{1 + k_{n,p}(1 + a_{n,p})z^{-1} + a_{n,p}z^{-2}}, \quad (1)$$

This function is described by the adaptive coefficient $k_{n,p}$, related to the tracked frequency $f_{n,p}$, and the pole-zero contraction factor $a_{n,p}$ controlling the bandwidth of the filter [16]. Besides removing the fundamental frequency, the notch filter can change its cut-off frequency at each new sample to track the time-varying fundamental frequency. In this way, decorrelation is provided in the whole low-frequency band acting as happens with time-varying all-pass filters [17]. The disparity among channels is guaranteed also when the channels are characterized by the same fundamental frequency since a different value of the contraction factor $a_{n,p}$ is adopted. In particular, the vector $\boldsymbol{a}_{n,p} = [a_{n,1}, \cdots, a_{n,P}]$ is updated as follows:

$$\boldsymbol{a}_{n,p} = \begin{cases} s(\boldsymbol{a}_{n-1,p}, 1) & \text{if } \left(\mathrm{n} - \mathrm{L}\left\lfloor \frac{\mathrm{n}}{\mathrm{L}} \right\rfloor\right) = 0 \\ \boldsymbol{a}_{n-1,p} & \text{otherwise,} \end{cases} \quad (2)$$

where $s(\cdot, 1)$ is a right circular shift of one sample performed every L samples, being L the block length.
The coefficient $k_{n,p}$ is bounded in the range $(-1, 1)$ to prevent the filter from diverging using the following function:

$$k_{n,p} = \frac{2}{1 + e^{-g_{n,p}}} - 1, \quad (3)$$

where $g_{n,p} \in R$ is the parameter to be minimized in the cost function computed as fully described in [7]. Finally, given the estimated coefficient $k_{n,p}$, the sampling frequency f_{s}, and D the down-sampling factor, the estimated frequency $f_{n,p}$ is computed as follows

$$f_{n,p} = \frac{f_{\mathrm{s}}}{D} \cdot \frac{1}{2\pi} \cos^{-1}(-k_{n,p}). \quad (4)$$

The estimation and tracking performance of the approach, especially when the low-frequency range includes also some harmonics, are improved using a pre-emphasis stage to emphasize the low-frequency band and a de-emphasis stage to undo its effect, as follows:

$$\mathbf{H}_{\mathrm{pre}}(z) = \frac{1}{1-\nu z^{-1}} \tag{5}$$

$$\mathbf{H}_{\mathrm{de}}(z) = 1-\nu z^{-1}, \tag{6}$$

where $0 < \nu < 1$. Regarding the high-frequency range, P second-order time-varying all-pass filters are applied in order to alter the phase of the input channels without affecting the spatial perception [17]. Thus, the p-th allpass filter is characterized by a pole with multiplicity 2 related to the adaptive coefficient $k_{n,p}$ of Eq. (3) and it has the following transfer function [4]:

$$\mathbf{F}_p(z,n) = \frac{k_{n,p}^2 - 2k_{n,p}z^{-1} + z^{-2}}{1-2k_{n,p}z^{-1} + k_{n,p}^2 z^{-2}}. \tag{7}$$

The restriction $|k_{n,p}| < 1$ ensures the causality and stability of the filter. Moreover, considering the group delay of Eq. (7) and $|k_{n,p}| < 1$, the maximum change in the time of arrival of each frequency results limited to within about $40\,\mu$s, thus satisfying also the restriction derived from the known "just noticeable inter-aural delay" [18]. This delay represents the minimum change in the inter-aural time delay between the two ears that causes a noticeable change in the perception of the direction of sound. The introduced delay must vary between $30\,\mu$s and $200\,\mu$s [18].

2.2. Multichannel Acoustic Echo Cancellation

The decorrelated input signals $x_p[n]$ are collected in an input data matrix $\mathbf{X}_n \in \mathbb{R}^{K\times MP}$, 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. The input matrix is processed by the CMAEC, which is composed of 2 MIMO filters, each one represented by a coefficient matrix $\mathbf{W}_n^{(j)} \in \mathbb{R}^{MP\times Q}$, with $j = 1, 2$, which contains all the individual filters $\mathbf{w}_{n,pq} \in \mathbb{R}^{M\times 1}$. The output of each MIMO filter $\mathbf{Y}_n^{(j)} \in \mathbb{R}^{K\times Q} = \mathbf{X}_n^{(j)}\mathbf{W}_{n-1}^{(j)}$ can be seen as a concatenation of Q individual output vectors, i.e., $\mathbf{Y}_n^{(j)} = \begin{bmatrix} \mathbf{y}_{n,1}^{(j)} & \mathbf{y}_{n,2}^{(j)} & \dots & \mathbf{y}_{n,Q}^{(j)} \end{bmatrix}^T$, for $j = 1, 2$. For each MIMO filter, the error matrix is achieved as:

$$\mathbf{E}_n^{(j)} = \mathbf{D}_n - \mathbf{Y}_n^{(j)}, \tag{8}$$

where $\mathbf{D}_n \in \mathbb{R}^{K\times Q}$ is the matrix containing the Q desired signals. Also the error matrices can be seen as a concatenation of Q individual error vectors, i.e., $\mathbf{E}_n^{(j)} = \begin{bmatrix} \mathbf{e}_{n,1}^{(j)} & \mathbf{e}_{n,2}^{(j)} & \dots & \mathbf{e}_{n,Q}^{(j)} \end{bmatrix}^T$. The MIMO filters are individually updated according to the multichannel affine projection algorithm [11]:

$$\mathbf{W}_n^{(j)} = \mathbf{W}_{n-1}^{(j)} + \mu_j \mathbf{X}_n^T \left(\delta_j + \mathbf{X}_n\mathbf{X}_n^T\right)^{-1} \mathbf{E}_n^{(j)}, \tag{9}$$

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.

The outputs of the two MIMO filters are adaptively combined according to a system-by-system combination scheme [13]. In particular, the q-th output channel of the CMAEC is achieved by combining convexly the q-th individual output of the two MIMO filters related to the current projection, i.e.:

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

where $\lambda_q[n]$ is the q-th mixing parameter, which is constrained to remain in the range $[0, 1]$ [12]. The overall error signal related to the q-th output channel can be easily 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, $b_q[n]$, related to $\lambda_q[n]$ by the expression:

$$\lambda_q[n] = \beta\left(\frac{1}{1+e^{-b_q[n]}} - \alpha\right), \tag{12}$$

where

$$\alpha = \frac{1}{(1+e^4)}, \qquad \beta = \frac{1}{1-2\alpha}. \tag{13}$$

The update rule of the auxiliary parameter can be defined according to [13]:

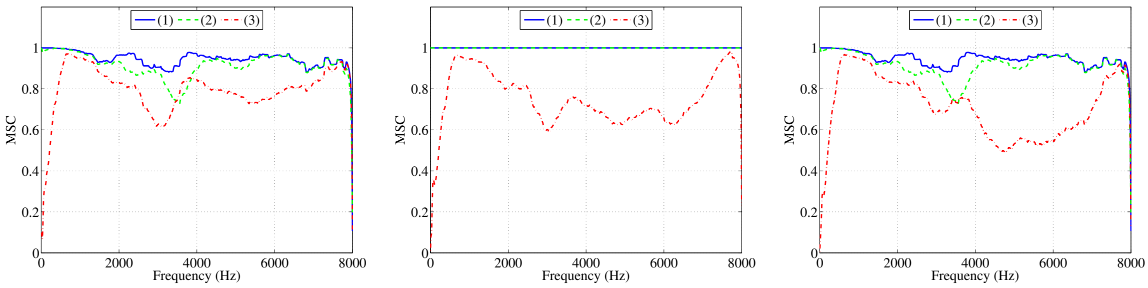
$$b_q[n+1] = b_q[n] + \frac{\mu_{\mathrm{c}}}{\beta r_q[n]} e_q[n]\left(y_q^{(1)}[n] - y_q^{(2)}[n]\right) \cdot (\lambda_q[n] + \alpha\beta)(\beta - \alpha\beta - \lambda_q[n]), \tag{14}$$

where μ_{c} is the step size of the combination, $r_q[n] = \gamma r_q[n-1] + (1-\gamma)\left(y_q^{(1)}[n] - y_q^{(2)}[n]\right)^2$ is the estimated power of $\left(y_q^{(1)}[n] - y_q^{(2)}[n]\right)$, and γ is a smoothing factor. By using the adaptation rule of (14), the combination scheme is able to adaptively recognize and select the best performing filters, thus improving the overall performance of the CMAEC.

3. ALGORITHM VALIDATION

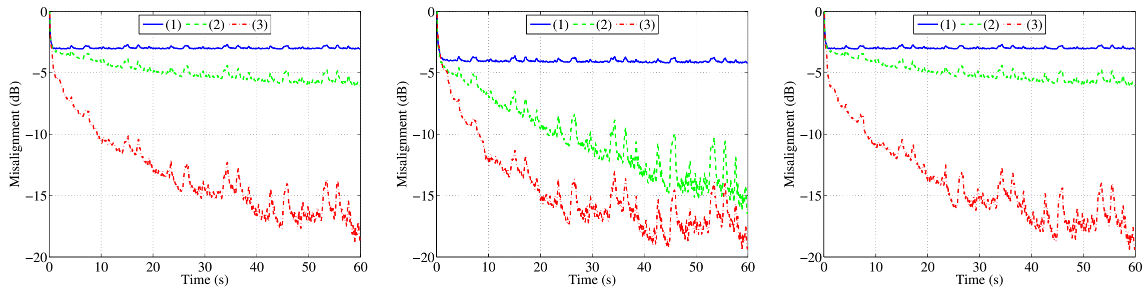
In this section, the effectiveness of the proposed approach in improving the performance achievable in estimating an unknown multichannel system is presented showing how the interchannel coherence affects the correct estimation of the actual AIRs at the near-end. Tests have been carried out considering the online system identification without any decorrelation algorithm as the reference scenario. A comparison with a well-known approach in the literature, i.e., the half-wave rectifier nonlinearity with nonlinear parameter $\gamma = 0.5$ [6], has been performed to underline the obtained improvement.

A speech signal (SQAM disk [19]) sampled at 16 kHz and AIRs of length $M = 512$ samples, which is also the length



(a) (b) (c)

Fig. 2. MSC obtained considering a speech signal (1) without decorrelation, (2) with the half-wave rectifier techinque [6], and (3) with the proposed approach. (a) Channel 1 vs channel 2. (b) Channel 1 vs channel 3. (c) Channel 2 vs channel 3.



(a) (b) (c)

Fig. 3. Misalignment obtained (1) without decorrelation, (2) with the half-wave rectifier techinque [6], and (3) with the proposed approach. (a) Microphone 1. (b) Microphone 2. (c) Microphone 3.

of the adaptive filters, simulated using the image method as shown in [4], have been considered. The scenario involves an unknown multichannel system with $P = 3$ loudspeakers and $Q = 3$ microphones. The room has fixed dimensions of $3\,\mathrm{m}\times 2\,\mathrm{m}\times 2.5\,\mathrm{m}$ with a reverberation time of about 0.1 s.

The proposed decorrelation method has been implemented using $D = 32$ to increase spectral resolution and $\nu = 0.5$. The vector $\boldsymbol{a} = [0.95, 0.75, 0.55]$ includes uniformly distributed values in the range from 0.55 to 0.95 frame-by-frame shifted every 0.1 s to provide a time-varying bandwidth and depth of the notch filter and thus to obtain the disparity among channels. All these parameters have been fixed to minimize the average misalignment but optimizing the trade-off between channel decorrelation and audio quality preservation. The proposed combined MIMO filter, which is composed of two MIMO filters, one using a small step size $\mu_1 = 0.1$ and the other one using a larger value $\mu_2 = 1$, has been adopted to identify the $P \cdot Q = 9$ unknown AIRs. Each MIMO filter is individually updated according to the multichannel APA with the same projection order $K = 4$ and regularization factor $\delta = 0.001$.

The influence of the correlation among the loudspeaker signals on the convergence performance is underlined showing the interchannel coherence in terms of magnitude-squared coherence (MSC) described by the following equation as a function of the discrete frequency k:

$$\mathrm{MSC}(k) = \frac{\left|P_{x_j x_h}(k)\right|^2}{P_{x_j x_j}(k) P_{x_h x_h}(k)}, \tag{15}$$

where $P_{x_j x_j}$ and $P_{x_j x_h}$ are the auto power spectra and cross power spectra between channels j and h ($j \neq h$), respectively. Then, results are presented in terms of the misalignment $\varepsilon_q[n]$ at microphone q computed for each time instant n between the vectors $\mathbf{h}_{pq}$ and $\mathbf{w}_{n,pq}$ of the actual and estimated AIRs (being p the loudspeaker index) as follows [5]:

$$\varepsilon_q[n] = \frac{\sum_{p=1}^{P} \left\|\mathbf{h}_{pq} - \mathbf{w}_{n,pq}\right\|^2}{\sum_{p=1}^{P} \left\|\mathbf{h}_{pq}\right\|^2}. \tag{16}$$

Fig. 2 shows the performance of the decorrelation methods under comparison in terms of MSC proving that a higher interchannel coherence reduction is provided by the proposed approach. As expected, the correlation among channels affects the performance of the system identification as reported in Fig. 3. The obtained misalignment shows that the better convergence performance is provided by the proposed approach corresponding to lower values of the MSC (Fig. 2). Finally, informal listening tests were carried out to obtain preliminary results. Involved subjects were asked to judge the global perceived sound quality reporting positive comments on the proposed approach, thus confirming its validity.

4. CONCLUSIONS

An advanced multichannel acoustic echo cancellation system has been presented based on a novel multichannel decorrelation approach and convex combination of MIMO filters. The missing-fundamental theory has been combined with time-varying allpass filters to reduce the interchannel coeherence among the loudspeakers signals still preserving audio quality. Then, convex combination of MIMO filters with different adaptation settings has been exploited to improve the estimation performance by automatically switching between the best performing filter. Experimental results proved the effectiveness of the presented framework providing a suitable solution for multichannel echo cancellation.

REFERENCES

[1] D. Comminiello, S. Scardapane, M. Scarpiniti, R. Parisi, and A. Uncini, "Convex combination of MIMO filters for multichannel acoustic echo cancellation," in *Image and Signal Processing and Analysis (ISPA), 2013 8th International Symposium on*, Sept 2013, pp. 778–782.

[2] D. Comminiello, M. Scarpiniti, R. Parisi, and A. Uncini, "Intelligent acoustic interfaces for immersive audio," in *Proc. 134th Audio Engineering Society Convention*, Rome, Italy, May 2013.

[3] S. Cecchi, L. Romoli, P. Peretti, and F. Piazza, "Low-complexity implementation of a real-time decorrelation algorithm for stereophonic acoustic echo cancellation," *Signal Processing*, vol. 92, no. 11, pp. 2668–2675, Nov. 2012.

[4] L. Romoli, S. Cecchi, and F. Piazza, "A novel decorrelation approach for multichannel system identification," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Florence, Italy, May 2014.

[5] J. Herre, H. Buchner, and W. Kellermann, "Acoustic echo cancellation for surround sound using perceptually motivated convergence enhancement," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Honolulu, HI, USA, Apr. 2007, vol. 1, pp. 17–20.

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

[7] 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 and Language Processing*, vol. 20, no. 2, pp. 690–698, Feb. 2012.

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

[9] A. Uncini, *Fundamentals of Adaptive Signal Processing*, Springer, 2014, ISBN 978-3-319-02806-4.

[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. 29372944, Mar. 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] D. Comminiello, M. Scarpiniti, R. Parisi, and A. Uncini, "Combined adaptive beamforming schemes for nonstationary interfering noise reduction," *Signal Process.*, vol. 93, no. 12, pp. 3306–3318, Dec. 2013.

[14] D. Comminiello, M. Scarpiniti, R. Parisi, and A. Uncini, "Combined adaptive beamforming techniques for noise reduction in changing environments," in *Proc. IEEE Int. Conf. Telecom. Signal Process.*, Rome, Italy, Jul. 2013, pp. 690–694.

[15] 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.*, Florence, Italy, Sep. 2011, pp. 64–67.

[16] J. Lee, E. Song, Y. Park, and D. Youn, "Effective bass enhancement using second-order adaptive notch filter," *IEEE Trans. Consum. Electron.*, vol. 54, pp. 663–668, May 2008.

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

[18] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, Springer-Verlag, 1990.

[19] "Sound Quality Assessment Material, recordings for subjective tests," Technical Review 3253-E, Apr. 1988.