CHANNEL ESTIMATION FOR CROSSTALK CANCELLATION IN WIRELESS ACOUSTIC NETWORKS

Gema Piñero[*]

Inst. Telecommunication and
Multimedia Applications (iTEAM)
Universitat Politècnica de València (Spain)
gpinyero@iteam.upv.es

Patrick A. Naylor

Department of Electrical and
Electronic Engineering
Imperial College London (United Kingdom)
p.naylor@imperial.ac.uk

ABSTRACT

In this paper we deal with the estimation of the room impulse response (RIR) between each loudspeaker and each microphone of a wireless acoustic network of two nodes when used to implement a crosstalk canceller. The nodes of the network are commercial devices connected via standard wireless links, presenting low computational requirements and non-ideal synchronization between them. Moreover, the nodes can exchange information, but they cannot share their signals due to the high throughput and perfect synchronism that would be required. The proposed scheme adaptively estimates the global impulse response between the source signals and the recorded signal at each node of the network, and afterwards estimates the corresponding RIRs between each loudspeaker and the node's microphone. This scheme does not need any additional synchronism between loudspeakers. Simulations show that proportionate-type affine projection algorithms obtain good performance for order $N = 4$, being their cost affordable in commercial devices.

Index Terms— Channel identification; crosstalk cancellation; wireless acoustic networks; adaptive algorithms

1. INTRODUCTION

Most of the applications regarding the enhancement of the quality of a sound signal recorded inside a room, or regarding the rendering of a sound signal to a particular location of a room, need to know the characteristics of the acoustic channels between the sound source and the recording position. Therefore, the estimation of the room impulse response (RIR) has been widely studied in the literature, resulting in a variety of techniques. The classical methods make use of white noise, chirp signals or maximum length sequences (MLS) [1, 2], but these techniques cannot be used when the RIRs have to be adaptively updated at the same time that the system is operating.

One approach is to characterize the RIR, or its acoustic transfer function, by a statistical model including a direct path and a reverberation part. In [3] this statistical model is used to equalize only the direct path and to measure the error that produces the other part. In [4] this statistical model is used to evaluate the performance of a multichannel Wiener filter for dereverberation. A second popular approach is to perform a blind estimation of the RIRs based on the minimization of the cross-relation between the signals recorded by multiple microphones [5]. A sparsity constraint may also be applied to the estimated RIR or any other property as its non-negativity [6].

Generally speaking, RIR estimation depends on the specific system and application. Focusing on the crosstalk canceller (CC) [7, 8] presented in this paper, it is implemented on a two-node wireless acoustic sensor network (WASN) [9], which usually presents certain constraints: non-perfect synchronism, low computation requirements, and low data exchange capacity between nodes [10, 11, 12]. Taking into account these constraints, we propose in this paper an adaptive system to estimate the RIRs involved in the WASN that is robust against any lack of synchronism due to the wireless connections, and that requires low computation. Simulations show that the proposed scheme obtains good results even for abrupt changes in the RIRs involved.

The outline of the paper is as follows: Section 2 states the basic formulation of a crosstalk canceller over a two-node WASN. Section 3 discusses two possible approaches in order to re-estimate the acoustic channels and proposes two different solutions based on the same adaptive system. Section 4 presents some simulation results and Section 5 summarizes the main conclusions of the paper.

2. MODEL FORMULATION

The block diagram of a CC of two microphones and two loudspeakers implemented over a WASN is shown in Fig. 1. This is the simplest example case of CC, but our method for channel estimation can be straightforwardly extended to $M > 2$ loudspeakers controlled by two or more nodes. The target application of the CC is to use the loudspeakers to render sound $s_1(n)$ at the location of the first node microphone, and at the same time, to render sound $s_2(n)$ at the position of the second node microphone. If we denote $x_1(n)$ and $x_2(n)$ of Fig. 1 the signals recorded by the microphones of the first and second node respectively. The relation between the microphones and the source signals are expressed by

$$\begin{aligned} x_1(n) &= [\mathbf{c}_{11} * \mathbf{h}_{11} + \mathbf{c}_{12} * \mathbf{h}_{21}] * s_1(n) \\ &+ [\mathbf{c}_{11} * \mathbf{h}_{12} + \mathbf{c}_{12} * \mathbf{h}_{22}] * s_2(n)\,. \end{aligned} \quad (1)$$

$$\begin{aligned} x_2(n) &= [\mathbf{c}_{21} * \mathbf{h}_{11} + \mathbf{c}_{22} * \mathbf{h}_{21}] * s_1(n) \\ &+ [\mathbf{c}_{21} * \mathbf{h}_{12} + \mathbf{c}_{22} * \mathbf{h}_{22}] * s_2(n)\,. \end{aligned} \quad (2)$$

where $*$ denotes discrete-time convolution, $\mathbf{c}_{ij}$ is the electro-acoustic path between the jth loudspeaker and the ith microphone and is modelled as the finite impulse response (FIR) filter:

$$\mathbf{c}_{ij} = \left[\begin{array}{cccc} c_{ij}(0) & c_{ij}(1) & \cdots & c_{ij}(L_c-1) \end{array}\right]^T,$$

[*]This work has been partially supported by EU together with Spanish Government through TEC2015-67387-C4-1-R (MINECO/FEDER) and PRX15/00548 (Ministry of Education), Generalitat Valenciana through PROMETEOII/2014/003 and UPV through PAID-00-15 grant.

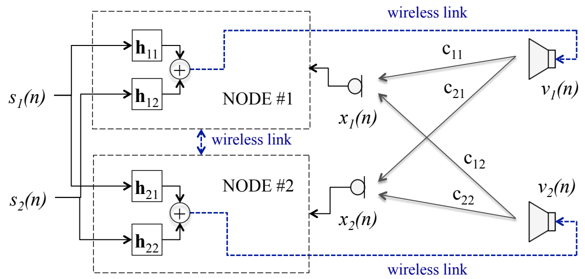

Fig. 1. Crosstalk canceller implemented in a two-node WASN.

where $(\cdot)^T$ stands for matrix transpose, $i, j = 1, 2$, and L_c is the maximum length in samples of all the paths involved. The filters $\mathbf{h}_{ji}$ are FIR filters of L_h coefficients defined as:

$$\mathbf{h}_{ji} = \begin{bmatrix} h_{ji}(0) & h_{ji}(1) & \cdots & h_{ji}(L_h - 1) \end{bmatrix}^T,$$

Due to the low computation requirement of the WASN, the filters $\mathbf{h}_{ji}$ of Fig. 1 are computed using the least squares (LS) solution in the frequency domain [13]. Denoting the kth frequency bin of the FFT of the acoustic channel $\mathbf{c}_{ij}$ as $C_{ij}(k)$, and the kth frequency bin of the FFT of the filter $\mathbf{h}_{ji}$ as $H_{ji}(k)$, we build matrices

$$\mathbf{C}(k) = \begin{bmatrix} C_{11}(k) & C_{12}(k) \\ C_{21}(k) & C_{22}(k) \end{bmatrix},$$

$$\mathbf{H}(k) = \begin{bmatrix} H_{11}(k) & H_{12}(k) \\ H_{21}(k) & H_{22}(k) \end{bmatrix},$$

and formulate the least squares problem as:

$$\mathbf{C}(k)\mathbf{H}(k) = \mathbf{I}, \tag{3}$$

and its LS solution as:

$$\mathbf{H}(k) = \mathbf{C}^H(k)\left(\mathbf{C}(k)\mathbf{C}^H(k) + \beta_{\mathrm{C}}\mathbf{I}\right)^{-1}, \tag{4}$$

where $(\cdot)^H$ indicates conjugate transpose, $\mathbf{I}$ is the identity matrix, and β_{C} is a regularization parameter [13]. For a system of M loudspeakers, dimensions of matrices $\mathbf{C}(k)$ and $\mathbf{H}(k)$ will be $[2 \times M]$ and $[M \times 2]$ respectively.

Regarding the calculation of matrix $\mathbf{H}(k)$ in (4), its first row contains the filter coefficients of the first node, whereas its second row contains those of node 2. As (4) is not separable in rows, both nodes must calculate the whole matrix $\mathbf{H}(k)$. For this purpose both nodes need to know the whole matrix $\mathbf{C}(k)$, that is, the acoustic channel $\mathbf{c}_{ij}$ must be previously estimated at the nodes. A detailed explanation on how to simultaneously estimate the four channels $\mathbf{c}_{ij}$ making use of two MLS can be found in [14]. Summarizing the main aspects of the estimation process in [14], node 1 estimates $\mathbf{c}_{11}$ and $\mathbf{c}_{12}$ at the same time that node 2 estimates $\mathbf{c}_{21}$ and $\mathbf{c}_{22}$. Afterwards they exchange their respective estimates to calculate the FFTs needed to form matrix $\mathbf{C}(k)$ and obtain (3). In order to exchange information between the nodes, they are interconnected via a wireless link and they run a specifically designed procedure to synchronize their clocks [15, 14].

3. ADAPTIVE IDENTIFICATION OF THE ACOUSTIC CHANNELS

One of the main advantages of a WASN is its flexibility in its deployment, so one loudspeaker or a microphone can be moved to a new position, thus the acoustic paths involved in Fig. 1 can vary in time and should therefore be adaptively re-estimated. Regarding the availability of signals at each node, node 1 of Fig. 1 has access to source signals $s_1(n)$, $s_2(n)$, to its own microphone signal, $x_1(n)$, and to its own loudspeaker signal $v_1(n)$, whereas the signals available at node 2 are $s_1(n)$, $s_2(n)$, $x_2(n)$ and $v_2(n)$. In order to design an adaptive scheme to re-estimate acoustic channels $\mathbf{c}_{ij}$, two different approaches are stated in the following.

3.1. Direct estimation

To discuss this approach, let us consider only node 1 of Fig. 1. The direct estimation of acoustic paths $\mathbf{c}_{11}$ and $\mathbf{c}_{12}$ should need loudspeakers signals $v_1(n)$ and $v_2(n)$ as the inputs to their respective adaptive filters, and the microphone signal $x_1(n)$ as the reference signal. Since all the filters involved in the CC have been computed at each node through equation (4), signal $v_2(n) = \mathbf{h}_{21} * s_1(n) + \mathbf{h}_{22} * s_2(n)$ can also be computed at node 1. Assuming the filtering is performed in the frequency domain through FFTs of N_{F} points, the extra computational cost to obtain $v_2(n)$ is due to the complex products $H_{21}(k)S_1(k)$ and $H_{22}(k)S_2(k)$ for $k = 0, \ldots, N_{\mathrm{F}}/2$ and to the inverse FFT needed to generate $v_2(n)$. Although the extra computational cost could be affordable, the main disadvantage of this method is the assumption that the signal $v_2(n)$ generated by node 1 is perfectly synchronized to the real signal $v_2(n)$ generated by node 2, which is not the case in WASNs [14].

3.2. Global Impulse Response (GIR) estimation

This second approach uses the source signals $s_1(n)$ and $s_2(n)$ as the inputs to the adaptive filters, and microphone signal $x_1(n)$ and $x_2(n)$ as the reference signals, and adaptively estimates the global impulse responses $\mathbf{a}_{11}$, $\mathbf{a}_{12}$, $\mathbf{a}_{21}$ and $\mathbf{a}_{22}$ based on:

$$x_1(n) = \mathbf{a}_{11} * s_1(n) + \mathbf{a}_{12} * s_2(n), \tag{5}$$

$$x_2(n) = \mathbf{a}_{21} * s_1(n) + \mathbf{a}_{22} * s_2(n), \tag{6}$$

where the GIRs are defined from (1)-(2) as:

$$\mathbf{a}_{11} = \mathbf{c}_{11} * \mathbf{h}_{11} + \mathbf{c}_{12} * \mathbf{h}_{21}, \tag{7}$$

$$\mathbf{a}_{12} = \mathbf{c}_{11} * \mathbf{h}_{12} + \mathbf{c}_{12} * \mathbf{h}_{22}, \tag{8}$$

$$\mathbf{a}_{21} = \mathbf{c}_{21} * \mathbf{h}_{11} + \mathbf{c}_{22} * \mathbf{h}_{21}, \tag{9}$$

$$\mathbf{a}_{22} = \mathbf{c}_{21} * \mathbf{h}_{12} + \mathbf{c}_{22} * \mathbf{h}_{22}. \tag{10}$$

The block diagrams to estimate the GIRs at each node are shown in Fig. 2. They are based on equations (5)-(6). The upper branch of node 1 in Fig. 2(a) tries to minimize the mean square error (MSE) of error signal $e_{11}(n)$ defined as:

$$e_{11}(n) = \mathbf{a}_{11} * s_1(n) - \hat{\mathbf{a}}_{11} * s_1(n). \tag{11}$$

Substituting $\mathbf{a}_{11} * s_1(n)$ by its expression in (5), we obtain:

$$e_{11}(n) = x_1(n) - \hat{\mathbf{a}}_{12} * s_2(n) - \hat{\mathbf{a}}_{11} * s_1(n), \tag{12}$$

where we have used the current estimate of $\mathbf{a}_{12}$ obtained in the lower branch of Fig. 2(a) and denoted by $\hat{\mathbf{a}}_{12}$. In the same way, the estimation of GIR $\mathbf{a}_{12}$ is obtained through the minimization of the MSE of $e_{12}(n) = \mathbf{a}_{12} * s_2(n) - \hat{\mathbf{a}}_{12} * s_2(n)$, substituting $\mathbf{a}_{12} * s_2(n)$ from its expression in (5), and using the current estimate of the direct GIR given by $\hat{\mathbf{a}}_{11}$:

$$e_{12}(n) = x_1(n) - \hat{\mathbf{a}}_{11} * s_1(n) - \hat{\mathbf{a}}_{12} * s_2(n). \tag{13}$$

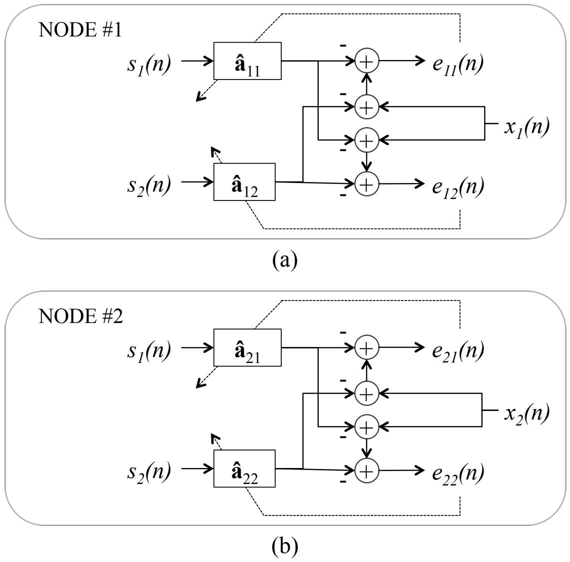

Fig. 2. Adaptive systems to identify (a) GIRs $\mathbf{a}_{1j}$ of node 1, and (b) GIRs $\mathbf{a}_{2j}$ of node 2.

It is straightforward to state from expressions (12) and (13) that both error signals are equal, and that a compact system can be proposed to jointly identify the GIRs of Fig. 2(a). Let us define the length of the GIRs as $L_a = L_c + L_h - 1$, and define vectors

$$\begin{aligned} \hat{\mathbf{a}}_1 &= \begin{bmatrix} \hat{\mathbf{a}}_{11}^T & \hat{\mathbf{a}}_{12}^T \end{bmatrix}^T , \\ \mathbf{s}(n) &= \begin{bmatrix} \mathbf{s}_1(n)^T & \mathbf{s}_2(n)^T \end{bmatrix}^T , \end{aligned}$$

as the concatenation of GIRs and the concatenation of the source signals samples needed to perform the filtering in Fig. 2(a) respectively, where $\mathbf{s}_i(n) = \begin{bmatrix} s_i(n) & s_i(n-1) & \cdots & s_i(n-L_a+1) \end{bmatrix}^T$ for $i = 1, 2$. Then, the estimation can be carried out minimizing the MSE of the same error signals but expressed as

$$e_{11}(n) = e_{12}(n) = x_1(n) - \hat{\mathbf{a}}_1^T \mathbf{s}(n) \,. \tag{14}$$

Finally, the same idea can be applied to adaptive system of node 2 in Fig. 2(b), obtaining its common error signal as

$$e_{21}(n) = e_{22}(n) = x_2(n) - \hat{\mathbf{a}}_2^T \mathbf{s}(n) \,, \tag{15}$$

where $\hat{\mathbf{a}}_2 = \begin{bmatrix} \hat{\mathbf{a}}_{21}^T & \hat{\mathbf{a}}_{22}^T \end{bmatrix}^T$.

3.2.1. RIR estimation

Once the adaptive filters have converged and the GIRs are identified either using the systems of Fig. 2 or their equivalent compact expression, the estimation of the acoustic channels $\mathbf{c}_{ij}$ is performed through the LS solution to (7)-(10) in the frequency domain. Defining row vectors $\hat{\mathbf{c}}_j(k) = [\hat{C}_{j1}(k), \hat{C}_{j2}(k)]$ and $\mathbf{g}_j(k) = [\hat{A}_{j1}(k), \hat{A}_{j2}(k)]$ for $j = 1, 2$, where $\hat{A}_{j1}(k)$ and $\hat{A}_{j2}(k)$ are the kth bin of the FFT of $\hat{\mathbf{a}}_{j1}$ and $\hat{\mathbf{a}}_{j2}$ respectively, then the acoustic channel estimate is obtained as the regularized LS solution to:

$$\mathbf{H}^T(k)\hat{\mathbf{c}}_j^T(k) = \mathbf{g}_j^T(k) \,, \; j = 1, 2 \,, \tag{16}$$

that is given by [16]:

$$\hat{\mathbf{c}}_j^T(k) = \left(\mathbf{H}(k)\mathbf{H}^H(k) + \beta_{\mathrm{H}}\mathbf{I}\right)^{-1} \mathbf{H}(k)\mathbf{g}_j^T(k), \tag{17}$$

where β_{H} is the regularization parameter. Regarding the synchronization between both nodes for this solution, any asynchronism between the nodes will be reflected in the estimation of the GIRs, providing a more robust solution compared to the direct estimation.

3.2.2. Adaptive algorithms

To estimate the GIRs we propose to investigate the use of the affine projection algorithm (APA) [17, 18], the improved proportionate NLMS (IPNLMS) [19] and the memory-improved proportionate APA (MIPAPA) [20]. The two last algorithms were originally proposed for channels with sparse impulse responses [21], presenting fast convergence and good performance estimation. A supporting reason to choose the proportionate-type algorithms is that the GIRs are expected to be sparse in time since they are estimated once the CC filters have been designed and are already operating. To design the filters in (4), the WASN performs an initial estimation of the RIRs $\mathbf{c}_{ij}$ using MLS [14], as it was explained in Section 2. Therefore the initial shape of the direct GIRs $\mathbf{a}_{11}$ and $\mathbf{a}_{22}$ is impulse-like and therefore is highly sparse. When the channel varies, their impulse-type RIRs also vary, but even so they maintain a considerable degree of sparseness.

4. SIMULATION RESULTS

The WASN used to implement the crosstalk canceller is formed by two tablet computers running the Android operating system and two wireless loudspeakers connected to the tablets via Bluetooth. The tablets include one built-in microphone each. Some pictures of the WASN deployed, together with experiment results of the CC system can be found in [22].

The channels to be estimated are real acoustic responses measured inside a listening room of 9.36m long by 4.78m wide by 2.63m high located at the Audio Processing Laboratory of the Polytechnic University of Valencia, and modelled as FIR filters of $L_c = 1200$ coefficients. The length of the CC filters is $L_h = L_c$ and the sampling frequency is 11025 kHz. Source signals $s_1(n)$ and $s_2(n)$ are uncorrelated white noises and have been scaled to a maximum amplitude of $|s_i(n)| \leq 1$. The microphone signals $x_1(n)$ and $x_2(n)$ are respectively corrupted by two independent white Gaussian noises resulting in a signal-to-noise ratio of 20 dB. All the algorithms use the same step-size $\mu = 0.2$, whereas the regularization constant is $\delta = 20\sigma_s^2/2L_a$ for the MIPAPA algorithms [20], $\delta = 0.5\sigma_s^2$ for the APA algorithms, and $\delta = (1-\alpha)\sigma_s^2/2L_a$ for the IPNLMS algorithm, being σ_s^2 the power of the corresponding input signal. The parameter $\alpha = -0.5$ for all the proportionate-type algorithms.

First, we obtain the normalized misalignment (in dB) defined as $20\log_{10}(\|\mathbf{a}_{ij} - \hat{\mathbf{a}}_{ij}(n)\|_2/\|\mathbf{a}_{ij}\|_2)$ as the performance measurement, where $\mathbf{a}_{ij}$ are the true GIRs obtained in (7)-(10). The results are averaged over 20 independent trials. In order to evaluate the speed of convergence, an abrupt change of the acoustic channels is introduced half-way through the experiment. During the first time period the adaptive system is operating with the CC filters designed in (4). In this first period, the acoustic channels used in the simulation are the same channels used to design the filters. At the time of the channel change, the acoustic channels change abruptly (they are real channels measured in the same room and with the same deployed WASN, but with different relative delays between them), whereas the CC filters are unchanged. Fig. 3 and Fig. 4 show the misalignment performance of the system for node 1 and node 2 respectively.

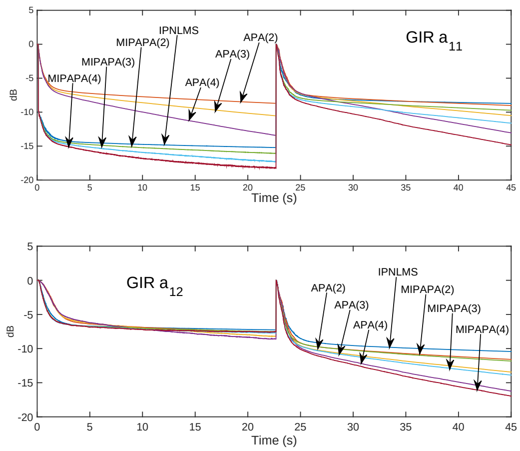

Fig. 3. Misalignment curves of the adaptive filters of node 1 in Fig. 2(a). Upper figure corresponds to $\hat{\mathbf{a}}_{11}$ and lower figure to $\hat{\mathbf{a}}_{12}$.

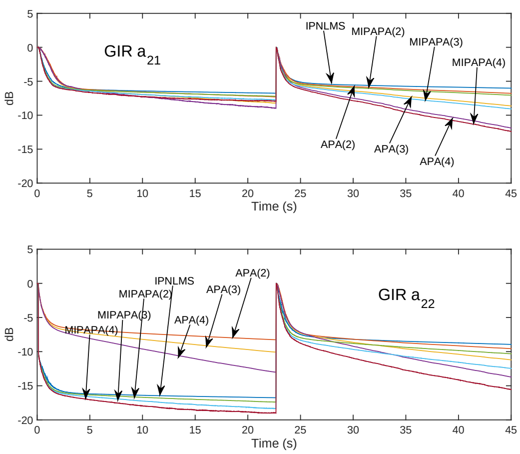

Fig. 4. Misalignment curves of the adaptive filters of node 2 in Fig. 2(b). Upper figure corresponds to $\hat{\mathbf{a}}_{21}$ and lower figure to $\hat{\mathbf{a}}_{22}$.

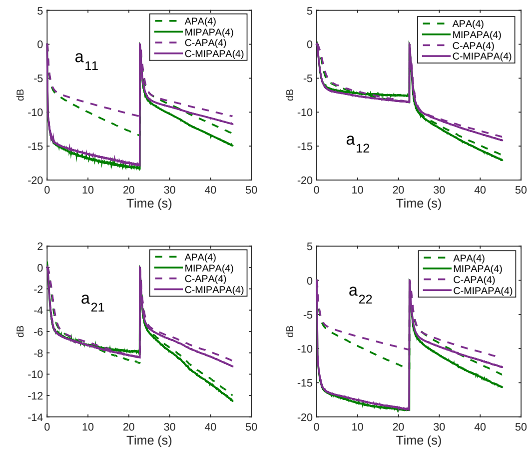

Fig. 5. Misalignment curves of the adaptive filters of Fig. 2 and their compact version expressed in (14)-(15).

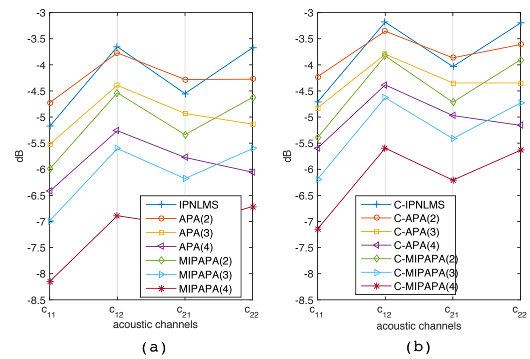

Fig. 6. MSE of the estimated acoustic channels using (17) for (a) adaptive systems of Fig. 2, and (b) compact approach of (14) -(15).

It can be observed that the performance of the proportionate-type algorithms is significantly better than the APA during the first half of the experiment for the direct GIRs $\mathbf{a}_{11}$ and $\mathbf{a}_{22}$, which is consistent with the good performance of proportionate algorithms when dealing with sparse impulse responses. Regarding the behaviour of the adaptive algorithms during the second half of the experiment, the MIPAPA of order N outperforms the corresponding APA of the same order, but the difference is very small. Moreover, APA and MIPAPA algorithms of order $N = 4$ present a decreasing misalignment in the second half of the the cross GIRs $\mathbf{a}_{12}$ in Fig. 3 and $\mathbf{a}_{21}$ in Fig. 4, improving the misalignment obtained for the same filters within the first time period.

In Fig. 5 the results obtained by the compact system of equations (14)-(15) are compared to those obtained by the adaptive system of Fig. 2, but only for the MIPAPA and APA algorithms of order $N = 4$. It can be seen that the performance of the compact system (denoted by the prefix "C-" in the name of the curves) is slightly worse than that of the original system for the second half of samples. A plausible explanation of this behaviour is to consider the influence of the covariance matrix of the source signals in the projection step of the APA algorithms. It seems more suitable to use separate projection steps for every filter $\mathbf{a}_{ij}$ such that each adaptive algorithm uses only the covariance matrix of its own source $s_i(n)$.

Finally, Fig. 6 shows the normalized mean square error (MSE) in dB between the true and the estimated RIRs once the algorithms have converged. The figure at the left shows the error for the original system proposed in Fig. 2 whereas the figure at the right shows the results obtained by the compact system. Both figures are represented in the same dB scale in order to be easily compared. It can be seen that the two-branch model of Fig. 2 obtains better performance in the final estimation of the RIRs $\mathbf{c}_{ij}$ as it did in the case of the GIRs. Regarding the comparison between MIPAPA and APA algorithms, the MSE of MIPAPA of order $N - 1$ is similar to that of APA of order N. Summarizing, the best choice in our experiments is to use MIPAPA of order $N = 4$ (or $N = 3$ if low computation is required) and the adaptive systems proposed in Fig. 2.

5. CONCLUSIONS

In this paper we have presented an adaptive scheme to estimate the RIRs involved in a WASN of two nodes when a crosstalk canceller is working. The proposed scheme is robust to the lack of synchronism in the wireless links, has a small computational cost, and obtains good performance when proportionate-type APA is used. The normalized MSE of the estimated RIRs are around 7.5 dB for the MIPAPA algorithm of order $N = 4$, and 6.5 dB for the same algorithm and order $N = 3$.

6. REFERENCES

[1] J. Vanderkooy, “Aspects of MLS measuring systems,” *J. Audio Eng. Soc*, vol. 42, no. 4, pp. 219–231, 1994.

[2] H. Tataria, P.D. Teal, M. Poletti, and T. Betlehem, “Simultaneous channel estimation and joint time-frequency domain crosstalk cancellation in multichannel personal audio systems,” in *2014 IEEE Workshop on Statistical Signal Processing (SSP)*, 2014, pp. 488–491.

[3] D.B. Ward, “On the performance of acoustic crosstalk cancellation in a reverberant environment,” *The Journal of the Acoustical Society of America*, vol. 110, no. 2, pp. 1195–1198, 2001.

[4] T.C. Lawin-Ore and S. Doclo, “Analysis of the average performance of the multi-channel Wiener filter for distributed microphone arrays using statistical room acoustics,” *Signal Processing*, vol. 107, pp. 96–108, 2015.

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

[6] M. Crocco and A. Del Bue, “Room impulse response estimation by iterative weighted L1-norm,” in *2015 23rd European Signal Processing Conference (EUSIPCO)*, 2015, pp. 1895–1899.

[7] M. Miyoshi and Y. Kaneda, “Inverse filtering of room acoustics,” *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 36, no. 2, pp. 145–152, 1988.

[8] S. Elliott, C. House, J. Cheer, and M. Simon-Galvez, “Crosstalk cancellation for headrest sound reproduction,” in *2016 AES Int. Conf. on Sound Field Control*, Jul 2016.

[9] I.F. Akyildiz, T. Melodia, and K.R. Chowdhury, “A survey on wireless multimedia sensor networks,” *Computer Networks*, vol. 51, no. 4, pp. 921 – 960, 2007.

[10] A. Bertrand, “Applications and trends in wireless acoustic sensor networks: A signal processing perspective,” in *18th IEEE Symp. Communications and Vehicular Technology in the Benelux (SCVT)*, 2011, pp. 1–6.

[11] H. Kim, S. Lee, J.-W. Choi, H. Bae, J. Lee, J. Song, and I. Shin, “Mobile maestro: Enabling immersive multi-speaker audio applications on commodity mobile devices,” in *Proc. ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing (UbiComp ’14)*, 2014, pp. 277–288.

[12] A. Bertrand, S. Doclo, S. Gannot, N. Ono, and T. van Waterschoot, “Special issue on wireless acoustic sensor networks and ad hoc microphone arrays,” *Signal Processing*, vol. 107, pp. 1–3, 2015.

[13] O. Kirkeby, P.A. Nelson, H. Hamada, and F. Orduna-Bustamante, “Fast deconvolution of multichannel systems using regularization,” *IEEE Trans. Speech and Audio Processing*, vol. 6, no. 2, pp. 189–194, 1998.

[14] G. Piñero, J. Estreder, F. Martinez-Zaldivar, M. Ferrer, and M. De Diego, “Sound-Field Reproduction System Over a Two-Node Acoustic Network of Mobile Devices,” in *IEEE 2nd World Forum on Internet of Things (WF-IoT)*, 2015, pp. 652–657.

[15] Q.M. Chaudhari, “A simple and robust clock synchronization scheme,” *IEEE Trans. on Communications*, vol. 60, no. 2, pp. 328–332, 2012.

[16] G.H. Golub and C.F. Van Loan, *Matrix Computations (3rd Ed.)*, Johns Hopkins University Press, 1996.

[17] K. Ozeki and T. Umeda, “An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties,” *Electronics and Communications in Japan (Part I: Communications)*, vol. 67, no. 5, pp. 19–27, 1984.

[18] S.L. Gay and S. Tavathia, “The fast affine projection algorithm,” in *1995 Int. Conf. on Acoustics, Speech, and Signal Processing*, 1995, pp. 3023–3026.

[19] J. Benesty and S.L. Gay, “An improved PNLMS algorithm,” in *IEEE Int. Conf. on Acoustics Speech and Signal Processing*, 2002, vol. 2, pp. 1881–1884.

[20] C. Paleologu, S. Ciochina, and J. Benesty, “An efficient proportionate affine projection algorithm for echo cancellation,” *IEEE Signal Processing Letters*, , no. 2, pp. 165–168, 2010.

[21] J. Benesty, Y.A. Huang, J. Chen, and P.A. Naylor, “Adaptive algorithms for the identification of sparse impulse responses,” in *Topics in Acoustic Echo and Noise Control*, pp. 125–153. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[22] Audio and Communications Signal Processing Group (GTAC), “CCS prototype over a WASN of commercial devices,” in *http://gpinyero.webs.upv.es/ccs-in-lab/*.