# MULTICHANNEL CROSS-TALK CANCELLATION IN A CALL-CENTER SCENARIO USING FREQUENCY-DOMAIN ADAPTIVE FILTERING

*Anthony Lombard and Walter Kellermann*

Multimedia Communications and Signal Processing,
University of Erlangen-Nuremberg, Cauerstr. 7, 91058 Erlangen, Germany
Email: {lombard,wk}@LNT.de

**ABSTRACT**

We propose a multichannel residual cross-talk cancellation strategy to operate in environments where each microphone signal is dominated by one of several simultaneously active sound sources, like in a call-center scenario. The method relies on a two-step adaptive filtering scheme derived from an established GSC structure. The filter updates are performed in the frequency domain to allow an effective DFT-bin-wise adaptation control. This new scheme is applicable to an arbitrary number of competing sources and experiments conducted for three and seven speech sources show that it allows to reduce the amount of undesired cross-talk components by about 10dB, without noticeable distortion of the desired signal.

***Index Terms***— Interference cancellation, adaptive filtering

## 1. INTRODUCTION

This paper addresses the problem of recovering the original sources from microphone signals corrupted by some residual (i.e., mostly low-level) cross-talk coming from competing sources. We therefore assume here that each sensor signal is dominated by one of several simultaneously active sound sources. This situation is given, e.g., in a call-center environment, or in any scenarios where multiple sources are captured by a dedicated microphone each. In such situations, even at a relatively low level, the cross-talk may become audible for a remote listener and cause unacceptable annoyance. Interference cancellation techniques can be used in this context to reduce the amount of cross-talk in the sensor signals, while avoiding deterioration of the desired signal.

The rest of the paper is organized as follows. We first introduce the basic principles of adaptive interference cancellation and discuss the problems encountered in our context in Sect. 2 . An appropriate signal enhancement strategy building upon existing concepts is then presented in Sect. 3. Experimental results and conclusions are provided in Sect. 4 and Sect. 5.

## 2. PROBLEM FORMULATION

In this section, we first consider the single-channel adaptive interference cancellation problem where the output of a single filter is subtracted from a target channel.

### 2.1. Signal model

Fig. 1 depicts the functionality of a basic Adaptive Interference Cancellation (AIC) scheme for a single interfering source. The signal to be enhanced (the primary input $p$ in the figure) is composed of a desired signal $p_d$, and an undesired signal $p_i$ uncorrelated with $p_d$ and originating from an interfering source. To remove the undesired signal components from $p$, the output $y$ of an adaptive Finite Impulse Response (FIR) filter $w$ is subtracted from $p$. The filter output signal should therefore be as close as possible to $p_i$.

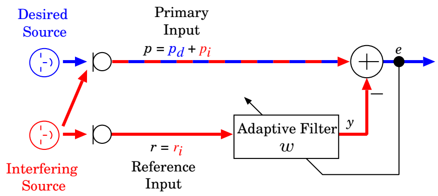


**Fig. 1**. Idealized AIC scenario.

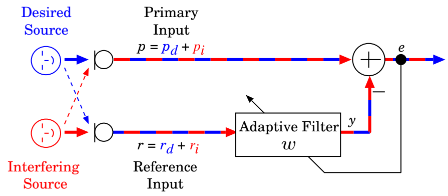


**Fig. 2**. Realistic AIC scenario.

In the idealized model (Fig. 1), the reference input $r$ to the FIR filter is assumed to come exclusively from the interfering source, i.e., $r$ is correlated with $p_i$, but uncorrelated with $p_d$. In most applications, like in the call-center scenario, this assumption does not hold. The more realistic model of Fig. 2 should then be considered, where the reference input also contains some components $r_d$ coming from the desired source. Another important characteristics of the call-center scenario is the limited amount of cross-talk in all microphone channels, i.e., the signals $p_i$ and $r_d$ are weak compared to $p_d$ and $r_i$. This is denoted by dashed lines in Fig. 2.

### 2.2. Performance measures

AIC algorithms are usually assessed in terms of Signal-to-Interference Ratio (SIR). Following the notations of Fig. 2 for the special case of a single interfering source, we can express, in dB, the SIR at the primary input $SIR_{\text{prim}}$ (also called $SIR_{\text{in}}$ in the rest of the paper), the SIR at the system output $SIR_{\text{out}}$ and the resulting gain in SIR:

$$SIR_{\text{in}} = SIR_{\text{prim}} = 10\log_{10}\left(\frac{\hat{E}\left[{p_d}^2\right]}{\hat{E}\left[{p_i}^2\right]}\right), \quad (1)$$

$$SIR_{\text{out}} = 10\log_{10}\left(\frac{\hat{E}\left[(p_d - r_d * w)^2\right]}{\hat{E}\left[(p_i - r_i * w)^2\right]}\right), \quad (2)$$

$$SIR_{\text{gain}} = SIR_{\text{out}} - SIR_{\text{in}}, \quad (3)$$

where "$*$" and $\hat{E}[\cdot]$ denote the convolution and expectation estimate operators, respectively. Similarly, we can also compute the SIR at the reference input as follows:

$$SIR_{\text{ref}} = 10\log_{10}\left(\frac{\hat{E}\left[{r_d}^2\right]}{\hat{E}\left[{r_i}^2\right]}\right). \quad (4)$$

In the idealized AIC model (Fig. 1), we have an infinitely low $SIR_{\text{ref}} = -\infty$ (since $r_d = 0$) and the desired source signal is absent from the filter output $y$. But in a realistic scenario (Fig. 2), the presence of the desired signal in the reference input introduces some propagation $r_d * w$ of the desired source components through the adaptive AIC filter. If this propagation is too strong compared to the desired signal $p_d$ in the primary signal, it will result in distortion/suppression of the desired signal and degradation of the output quality after subtraction of the filter output from the primary signal. This distortion effect can be quantified by computing the Signal-to-Distortion Ratio (SDR) of the system, expressed in dB as follows:

$$SDR = 10\log_{10}\left(\frac{\hat{E}\left[{p_d}^2\right]}{\hat{E}\left[(r_d * w)^2\right]}\right). \quad (5)$$

### 2.3. The achievable interference rejection

Under the conditions depicted in Fig. 1, and assuming an infinitely long cancelling filter, it is theoretically possible to filter the reference input such that the filter output is an exact replica of the undesired signal contained in the primary input. This results in a complete removal of the interfering signal (i.e., $SIR_{\text{out}} = +\infty$) at the system output, regardless of the $SIR_{\text{in}}$ at the primary input [1].

We saw however in Sect. 2.1 that our scenario does not comply with this idealized model since some desired signal components are present also at the reference input. This results in desired signal cancellation at the system output (and thus a low SDR) because the AIC tries to remove all components in the primary input which are correlated with the reference input. Apart from a degradation of the output signal quality, non-ideal reference signals also affect the achievable interference rejection. Assuming an infinitely long AIC filter, it was actually shown in [1] that the achievable SIR at the system output is the exact opposite of the SIR calculated at the reference input:

$$SIR_{\text{out}} = -SIR_{\text{ref}}. \quad (6)$$

Note that, accordingly, the achievable output SIR only depends on the SIR at the reference input, regardless of the input SIR in the primary channel. The quality of the reference signal passed to the AIC filter is therefore crucial.

In this paper, we consider scenarios where each microphone channel is dominated by one of the sound sources. We may think that such scenarios provide good reference signals already. But the difficulty, here, is that the SIR at the primary input is already relatively high. To obtain an attractive SIR improvement, fine tuning of the AIC filter is therefore required, which in turn necessitates a very good reference signal, as shown by (6). However, because of the symmetry of the acoustical setup, we typically have $SIR_{\text{ref}} \approx -SIR_{\text{in}}$. According to (6), it results in a system output with at most $SIR_{\text{out}} \approx -SIR_{\text{ref}} \approx SIR_{\text{in}}$. This means that with the simple AIC scheme depicted in Fig. 2, no gain can be expected in our case. Additional mechanisms are therefore necessary to guarantee the effectiveness of the cancellation scheme.

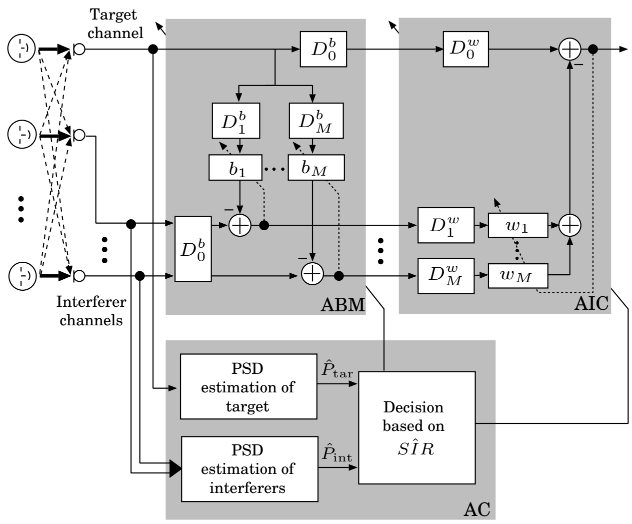


**Fig. 3**. Proposed multichannel residual cross-talk canceler for call-center-like scenarios.

## 3. THE ALGORITHM

### 3.1. Description of the algorithm

Only one interference source was considered in Sect. 2. Fig. 3 illustrates the solution adopted in this paper to address the call-center scenario, for an arbitrary number $M$ of interfering sources. The SIR definitions (1), (2) and (4) defined in Sect. 2.2 for the single-channel (single-filter) case can be easily extended to the multichannel case by replacing the interference terms $p_i$ and $r_i$ by the sum of the $M$ interfering contributions. Similarly, to extend the definition (5) of the SDR, we simply have to replace the distortion signal $r_d * w$ by a sum of $M$ distortion signals (one for each filter, i.e., one for each reference input).

In Fig. 3, one channel among the $M + 1$ sensor signals has been (arbitrarily) chosen as the "target channel" to enhance. The $M$ remaining channels are called "interferer channels". Moreover, we assume here that the number $M + 1$ of sensors is equal to the number of sources and that each source is dominant in one channel. We then have one desired source (dominant in the target channel) and $M$ interfering sources (each dominant in one of the interferer channels). To remove the interference leakage present at the target channel without deteriorating the desired signal coming from the target source, the residual cross-talk canceler combines a multichannel AIC scheme with an Adaptive Blocking Matrix (ABM) (similar in principle to the ABM of [2]). The ABM and AIC filtering units are presented in more detail in Sect. 3.2. Furthermore, an Adaptation Control (AC) stage has been implemented, which ensures correct convergence of the AIC and ABM filters towards the desired solution, as explained in Sect. 3.3.

### 3.2. The two filtering stages

The AIC stage in Fig. 3 generalizes the single-filter structure presented in Sect. 2. $M$ reference inputs built upon $M$ ABM outputs are used as inputs to the AIC filters $w_1, \cdots, w_M$, while the target channel serves as primary input. We saw in Sect. 2 that additional mechanisms were necessary to decrease the SIR at the reference inputs passed to the AIC filters. One way to achieve this is to introduce an ABM stage which minimizes the target signal leakage at the AIC reference inputs, like in the Robust Generalized Sidelobe

Canceler (RGSC) known from the beamforming literature [2]. As can be seen from Fig. 3, the ABM operates as a bank of single-channel AICs, with the target channel as a common reference input and the interferer channels applied as inputs to the adaptive filters $b_1, \cdots, b_M$. Additional delays $D_0^b$ and $D_0^w$ can be applied to avoid acausality of the desired ABM and AIC solutions, respectively, while delays $D_1^b, \cdots, D_M^b$ and $D_1^w, \cdots, D_M^w$ can be used to compensate for possibly large microphone spacings. The presence of $D_0^b$ in the interferer and target signal paths allows to keep all ABM outputs synchronized.

As mentioned above, the algorithm depicted in Fig. 3 resembles the RGSC structure, an extension of the well-known Generalized Sidelobe Canceler (GSC). Developed as a beamforming technique, the (R)GSC exploits some prior knowledge on the position of the target source (typically in the broadside direction of a microphone array) to steer a microphone beamformer towards this source and provide an output with enhanced desired signal. This signal serves both as primary input to the AIC and as reference input for the ABM. In our case, since we assume that the desired source is already dominant in the target channel, the implementation of an additional fixed beamformer is not necessary. But like in the RGSC, diligent control of the ABM and AIC filter updates are needed to avoid misadjustment of the adaptive filters and distortion of the desired signal at the system output. This is done by adapting the AIC and ABM filters in the frequency domain using the MC-BRFDAF algorithm [3] and by controlling the updates in a DFT-bin-wise manner, as shown below.

### 3.3. The Adaptation Control (AC)

The ABM cannot produce an estimate of the desired signal which is completely free of interference. Therefore, the AIC filters should only be adapted when the SIR is low in order to prevent cancellation and distortion of the desired signal (Sect. 2). On the other hand, the interferer channels contain some contributions from the target source. Adaptation of the ABM filters should only occur when the SIR is high to prevent suppression of some interference components by the ABM. The AC unit in Fig. 3 implements therefore an activity detector for "target source only" (adaptation of the ABM), "interference only" (adaptation of the AIC) and "double-talk" (no adaptation) which operates in separate DFT bins.

Applied to the RGSC, the MC-BRFDAF algorithm was controled in [3] by exploiting the directivity of the FBF to form an estimate of the SIR in each DFT bin, and track its minima and maxima. No FBF can be used here but we can compute an SIR estimate $S\hat{I}R(m,k)$ at each DFT bin $k$ and for each processing block $m$, using the sensor signals directly. From the target channel $x_{\text{tar}}$ and the $M$ interferer channels $x_{\text{int}}^i$, $i = 1 \cdots M$, periodograms

$$\hat{I}_{\text{tar}}(m,k) = \left| \sum_{n=0}^{N-1} x_{\text{tar}}(n+mR) e^{-j2\pi kn/N} \right|^2, \quad (7)$$

$$\hat{I}_{\text{int}}^i(m,k) = \left| \sum_{n=0}^{N-1} x_{\text{int}}^i(n+mR) e^{-j2\pi kn/N} \right|^2, \quad (8)$$

$$\hat{I}_{\text{int}}(m,k) = \frac{1}{M} \sum_{i=1}^{M} \hat{I}_{\text{int}}^i(m,k), \quad (9)$$

can be computed from (possibly overlapping) data frames of length $N$ (which is also the FFT length). $R$ is the sample innovation in each frame. Power Spectral Density (PSD) estimates can then be obtained by recursively averaging the periodograms using a forgetting factor

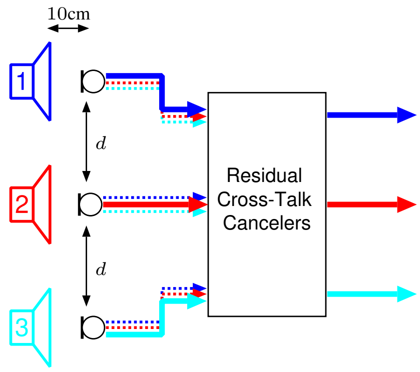


**Fig. 4**. Experimental setup.

$0 < \lambda < 1$:

$$\hat{P}_{\text{tar}}(m,k) = \lambda \hat{P}_{\text{tar}}(m-1,k) + (1-\lambda) \hat{I}_{\text{tar}}(m,k), \quad (10)$$

$$\hat{P}_{\text{int}}(m,k) = \lambda \hat{P}_{\text{int}}(m-1,k) + (1-\lambda) \hat{I}_{\text{int}}(m,k). \quad (11)$$

The SIR estimate, expressed in dB, is finally given as:

$$S\hat{I}R(m,k) = 10 \log_{10} \left( \frac{\hat{P}_{\text{tar}}(m,k)}{\hat{P}_{\text{int}}(m,k)} \right). \quad (12)$$

Using a maximum and minimum tracking in each DFT bin, calculated over a sliding detection window like in [4], the ABM (respectively AIC) filters are adapted whenever $S\hat{I}R(m,k)$ is maximum (respectively minimum) and positive (respectively negative). Such a mechanism allows to exploit the sparseness of the target and interference signals in time and frequency to obtain a frequent adaptation of the filters. Note that the proposed control mechanism does not differentiate between the interfering sources since (9) includes all interferer channels. Therefore, the activity detector enables the adaptation of either all AIC filters, or of all ABM filters.

## 4. EXPERIMENTAL EVALUATIONS

The performance of the presented residual cross-talk canceler (Fig. 3) has been assessed using the experimental setup depicted in Fig. 4, for different microphone spacings $d$. Three loudspeakers broadcasting speech signals at approximately the same level have been placed in front of three linearly-aligned microphones, at a distance of 10cm, in a living-room-like environment ($T_{60} \approx 300$ msec). To allow the calculation of the SIR and SDR values (Sect. 2.2), each source signal was first played separately. The microphone signals were then generated by adding up the source contributions at each sensor.

The AIC and ABM filters were updated every 512 samples at the sampling frequency $f_s = 16$kHz. The FFT length was set to $N = 2L$, where $L = 2048$ is the length of the AIC and ABM filters. The forgetting factor $\lambda = 0.2$ was used in (10) and (11) and the minimum/maximum tracking of the SIR estimate (12) was realized based on a sliding window of length 1.5 seconds (Sect. 3.3). The delay $D_0^b$ and $D_0^w$ were set to zero. Delays $D_1^w, \cdots, D_M^w$ and $D_1^b, \cdots, D_M^b$ were chosen approximately equal to the inter-microphone propagation delay. Note also that the proposed algorithm has been originally developed as a Multiple-Input-Single-Output (MISO) system (Fig. 3). As Fig. 4 indicates however, the algorithm has been extended here to form a Multiple-Input-Multiple-Output (MIMO) filtering system separating all sources simultaneously. This is done

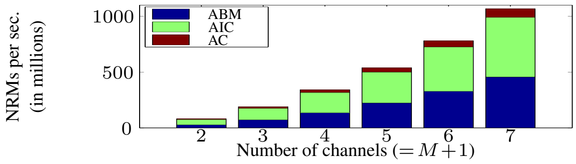

**Fig. 5**. Estimated Number of Real Multiplications per second.

by applying three MISO cancelers, assigning a different target channel to each output. Adaptation control of the ABM and AIC filters (Sect. 3.3) is performed for each output independently. Fig. 5 provides the estimated Number of Real Multiplications (NRMs) involved in each part of the scheme, when applying the set of parameters listed above for different numbers of channels. We can see from the figure that the computational complexity of the (MIMO-extended) scheme increases roughly quadratically with the number of channels.

The experimental results presented in this section have been averaged over all channels. Fig. 6 shows the interference rejection results obtained after convergence of the algorithm, for different spacings $d$ (Fig. 4). The performance of the algorithm including AIC and ABM (Fig. 3) has been compared to the performance achieved without ABM. Adaptation control of the filter updates (Sect. 3.3) was performed in both cases. Moreover, the convergence of the algorithm including AIC and ABM is depicted in Fig. 7. As expected, we see that the input SIR decreased with decreasing $d$. The algorithm performed well for all three microphone spacings, with or with ABM. It could provide an output SIR of at least 20dB on average. The ABM had very little impact on the achieved interference rejection because the double-talk detector involved in the AC unit could ensure a good AIC adaptation in all cases, even in the absence of ABM to pre-process the AIC reference inputs. However, the ABM allowed to improve the quality of the output signals significantly since it increased the SDR of 5 to 10dB, thereby maintaining the SDR at a high level above 35dB. This shows that the mechanisms implemented to provide good reference signals to the AIC were efficient and allowed to remove an important part of the cross-talk without distorting the target signal in each system output.

The scheme depicted in Fig. 3 for an arbitrary $M$, has also been assessed for seven sources (i.e., $M=6$ interfering sources) along a linear axis like in Fig. 4. The distance $d$ between neighboring microphones was 70cm. The input SIR of 4.8dB on average was therefore very low. This partly contradicts the assumption on which the algorithm is relying on, i.e., the presence of a (significantly) dominant source in each sensor signal. But under such adverse conditions, the algorithm could still provide a significant SIR improvement of 9.8dB, while maintaining the SDR at a satisfactory level of 28.5dB.

## 5. CONCLUSION

We proposed a residual cross-talk cancellation strategy to operate in environments where each microphone signal is dominated by one of several simultaneously active sound sources, like in a call-center scenario. It combines a multichannel AIC scheme with an ABM preprocessing stage and performs the update of the adaptive filters in the frequency domain to allow an effective DFT-bin-wise adaptation control. The scheme is applicable to an arbitrary number of competing sources. Experiments conducted for three and seven sources showed that the scheme provides good interference rejection performance without noticeable degradation of the target signal.

Originally developed as an adaptive MISO interference canceler, the algorithm can be easily extended to a MIMO system capable of

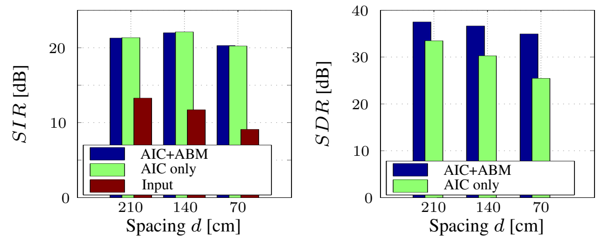

**Fig. 6**. SIR and SDR values calculated at the inputs (averaged over all channels) or after convergence of the residual cross-talk canceler (both with or without ABM), in the three-channel case.

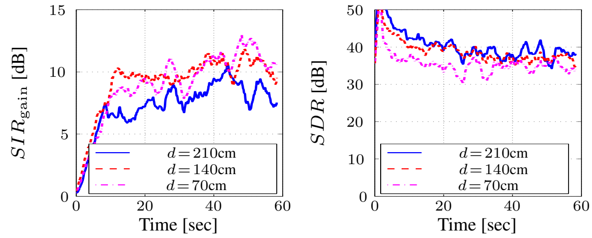

**Fig. 7**. Convergence of the residual cross-talk canceler (AIC+ABM) in the three-channel case.

simultaneously separating all sources. This extension reminds of the Blind Source Separation (BSS) problem aiming at separating independent components from a convolutive signal mixture. But a major difference between both concepts is that BSS does not make use of reference inputs and therefore does not require a separate double-talk detector [5]. BSS constitutes a more general but computationally more expensive solution to the problem of separating several simultaneously active sound sources. The proposed method, on the other hand, makes use of some prior knowledge on the acoustical setup to offer a conceptually simple and robust alternative. Combinations applying the proposed method as a post-processing step to other algorithms like BSS may be of interest, but interaction effects between the involved algorithms would need to be carefully studied.

## 6. REFERENCES


[1] B. Widrow, J. Glover, J. MacCool, J. Kautnitz, C. Williams, R. Hearn, J. Zeidler, E. Dong, and R. Goodlin, "Adaptive noise cancelling: principles and applications," *Proc. IEEE*, vol. 63, pp. 1692–1716, 1975.

[2] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. Signal Processing*, vol. 47, no. 10, pp. 2677–2684, Oct. 1999.

[3] W. Herbordt, H. Buchner, S. Nakamura, and W. Kellermann, "Application of a double-talk resilient DFT-domain adaptive filter for bin-wise stepsize controls to adaptive beamforming," in *Proc. Int. Workshop on Nonlinear Signal and Image Processing (NSIP)*, Sapporo, Japan, May 2005.

[4] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Processing*, vol. 9, no. 5, pp. 504–512, July 2001.

[5] H. Buchner and W. Kellermann, "A fundamental relation between blind and supervised adaptive filtering illustrated for blind source separation and acoustic echo cancellation," in *Proc. Joint Workshop on Hands-Free Communication and Microphone Arrays (HSCMA)*, Trento, Italy, May 2008.