11

COMBINED ARCHITECTURES FOR ADAPTIVE BEAMFORMING

Contents

ADAPTIVE combination of filters is a very effective and flexible approach to balance the compromises inherent to the settings of adaptive filters. In this chapter we exploits the capabilities of adaptive combination of filters in order to introduces novel adaptive beamforming methods for speech enhancement applications, designed to be robust against adverse environment conditions. The proposed architectures derive from the *generalized sidelobe canceller* (GSC); the novelty relies on the use of hybrid adaptive sidelobe cancelling structures which allow the system to achieve robustness in nonstationary environments. The novel structures are based on the convex combination of two *multiple-input single-output* (MISO) adaptive systems with complementary capabilities. The whole beamformer benefits from the combination and results to be able to preserve the best properties of each system. Experiments show that the proposed beamforming systems are capable of enhancing the desired speech signal even in adverse environment conditions[1].

11.1 INTRODUCTION

In immersive speech communications, taking place in multisource environments, the presence of interfering signals and reverberation may cause the loss of spatial information, thus resulting in compromising the speech intelligibility. In order to tackle this problem, speech enhancement systems are widely employed in distant talking applications. Microphone array beamforming represents a class of such speech enhancement techniques which are highly effective in acquiring a desired source signal while reducing the interfering components, thus resulting in recovering the binaural perception. Beamforming systems exploit the properties of microphone interfaces which facilitate binaural hearing.

The *generalized sidelobe canceller* (GSC) [54] is one of the most popular beam-

[1]The work in this chapter has been performed while the author was a Ph.D. student collaborating with the Fondazione Ugo Bordoni.

forming techniques for speech enhancement. The potency of a GSC system strictly relies on the adaptive algorithm chosen to perform the sidelobe canceller in the adaptive path. Generally the adaptation of filters in time-domain may be performed by *gradient*-based adaptive algorithms (see Section 4.4), such as the LMS-type algorithms. Although this family of algorithms is computationally quite cheaper, when the filter length is quite large a rather slow convergence occurs [120], thus the adaptation of the filter weights becomes unpractical in hands-free applications. Another time-domain standard approach is *Hessian*-based adaptive filtering, which is typical of algorithms such as the RLS. The latter approach displays a faster convergence rate compared with gradient-based algorithms [120]. However, RLS adaptive filtering entails a high computational complexity; therefore, adaptation may become prohibitively expensive, thus compromising real-time implementations. Moreover, the RLS may perform worse than LMS algorithm in nonstationary environment, depending on the statistics of acquired source signals [41]. A good compromise between performance and computational load may be obtained by using the family of APA [98], which is quite used in adaptive beamforming [163, 30], since it shows better convergence rates and manageable computational complexity compared with other time-domain algorithms. Moreover, APA is the best suitable algorithm to process speech signals compared with other classic time-domain adaptive algorithms. However, despite its good capabilities, APA suffers adverse environment conditions, especially in presence of multiple nonstationary sources which make the adaptation process unstable and reduce speech enhancement performance.

In order to address this problem we propose robust microphone beamforming architectures based on the adaptive combination of MISO systems, that are nothing but filter banks. Combined adaptive schemes are usually adopted with filters of the same family and complementary properties, e.g. using different step sizes, different filter lengths; however, they are used even with filters of different families using different updating rules or different cost functions [82, 160, 17, 126, 74]. A combined architecture is capable of adaptively switching between filters according to the best performing filter, thus always providing the best possible filtering (see Chapter 10).

In this chapter we propose two different beamforming architectures based on the combination of MISO systems using different updating approaches. In particular, we propose a *system-by-system* combined architecture, in which the overall output of the first MISO system is convexly combined with the overall output of the second MISO system, and a *filter-by-filter* combined architecture, in which each adaptive filter of the first MISO system is convexly combined with the correspondent filter of the second MISO system. Moreover, in order to use the best parameter setting for each filter and further improve the tracking performance we use both the combination of filters with different updating approaches and the combination of filters with different step size values in a *multi-stage* combined architecture in which the filtering process is carried out in two steps [73].

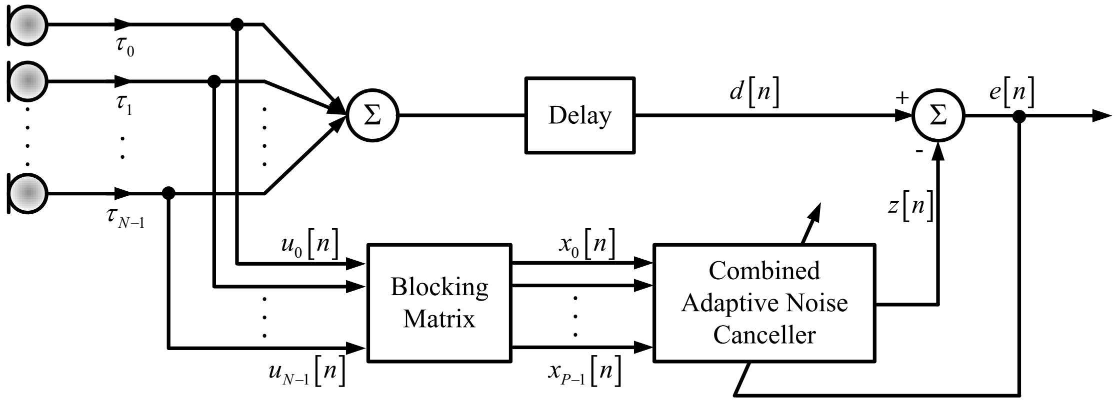


Fig. 11.1: *Microphone array beamforming architecture.*

11.2 COMBINED MICROPHONE ARRAY BEAMFORMING ARCHITECTURE

The beamforming architecture adopted in this paper is a typical GSC configuration [54] composed of a microphone array interface, a fixed *delay-and-sum beamformer* (DSB), and an *adaptive noise cancelling* (ANC) path, as depicted in Fig. 11.1. Let us consider a microphone array interface composed of N sensors. The signal $u_i[n]$ acquired by the i-th microphone, with $i = 0, \ldots, N-1$, is a delayed replica of the target signal $s[n]$ convolved with the (AIR) $\mathbf{a}_i$ between the i-th microphone and the desired source with the addition of background noise $v_i[n]$. The DSB spatially aligns the microphone signals with reference to the desired source direction, yielding the speech reference signal $d[n]$:

$$\begin{aligned} d[n] &= \sum_{i=0}^{N-1} u_i[n] \\ &= \sum_{i=0}^{N-1} \sum_{m=0}^{M-1} a_i[m]\, s[n-m-\tau_i] + v_i[n] \end{aligned} \tag{11.1}$$

where we suppose that each AIR between the desired source and the i-th microphone has the same length denoted with M. τ_i represents the delay relative to the i-th microphone.

In the adaptive path of the beamformer, the *blocking matrix* (BM) generates the noise references $x_p[n]$, with $p = 0, \ldots, P-1$, being $P = N-1$. The BM is implemented by pairwise differences between microphone signals [20], i.e. the sum of the elements of each column, except the first one, is null.

The noise reference signals are then processed by means of the *combined adaptive noise canceller* (CANC), whose structure will be detailed in the next section. The goal of the CANC is to remove any residual noise components in the speech reference signal, minimizing the output power and yielding the beamformer output signal $e[n]$.

11.3 ADAPTIVE COMBINATION OF MISO SYSTEMS

11.3.1 Convex combination of adaptive filters using APA

The trademark of the proposed beamforming approach is represented by the structure of the CANC. Generally, a conventional ANC is composed of an adaptive filter bank forming an MISO system. However, the adopted architecture results from combinations of adaptive filters. In particular, the structure is composed of two or more different MISO systems, each bringing different filtering capabilities to the whole beamformer. Each MISO system receives the same input signals, which are the noise reference signals resulting from the BM. Taking into account a number J of MISO systems, the p-th filter of the j-th MISO system, with $j = 0, \ldots, J-1$, receives as input a noise reference matrix $\mathbf{X}_{n,p}^{(j)}$, defined similarly to (5.2), but using a projection order K_j relative to all the filters of the j-th MISO system. We denote the coefficient vector of the p-th filter belonging to the j-th MISO system at n-th time instant as $\mathbf{w}_{n,p}^{(j)} \in \mathbb{R}^M$, which contains the same number of coefficients, M, and is adapted according to the *affine projection algorithm* (APA) [98], whose updating rule is derived similarly to (4.42):

$$\mathbf{w}_{n,p}^{(j)} = \mathbf{w}_{n-1,p}^{(j)} + \mu_j \mathbf{X}_{n,p}^{(j),T} \left(\delta_j \mathbf{I} + \mathbf{X}_{n,p}^{(j)} \mathbf{X}_{n,p}^{(j),T} \right)^{-1} \mathbf{e}_n^{(j)} \tag{11.2}$$

where $\mathbf{e}_n^{(j)} \in \mathbb{R}^{K_j}$ is the error vector of the j-th MISO system containing the last K_j samples of the j-th error signal, which results from:

$$\mathbf{e}_n^{(j)} = \mathbf{d}_n^{(j)} - \sum_{p=0}^{P-1} \mathbf{y}_{n,p}^{(j)} \tag{11.3}$$

where $\mathbf{d}_n^{(j)} \in \mathbb{R}^{K_j}$ is the vector containing the last K_j samples of the desired signal and $\mathbf{y}_{n,p}^{(j)} \in \mathbb{R}^{K_j} = \mathbf{X}_{n,p}^{(j)} \mathbf{w}_{n-1,p}^{(j)}$ is the vector containing the K_j projections of the output signal relative to the p-th filter of the j-th MISO system. Moreover, in equation (11.2), the parameters μ_j and δ_j are respectively the *step size* and the *regularization factor* common for all the filters of the j-th MISO system.

Using the updating rule described by (11.2) it is possible to differentiate the considered MISO systems simply changing the values of the step sizes or of the projection orders. However, aside from the chosen distinguishing parameters, there are two ways to combine the MISO system. The first way is to convexly combine the outputs of the two MISO systems and the second is to combine each filter of the first MISO system with the correspondent filter of the second MISO system under a convex constraint. We denote the former way as *system-by-system combined architecture* and the latter as *filter-by-filter combined architecture*, which are both described in the following two subsections.

11.3.2 System-by-system combined architecture

The first proposed scheme is the *system-by-system* CANC, depicted in Fig. 11.2 (a). The output of each MISO system, that we denote as $y^{(j)}[n] = \sum_{p=0}^{P-1} y_p^{(j)}[n]$, yields two system outputs that are then convexly combined generating the overall CANC output:

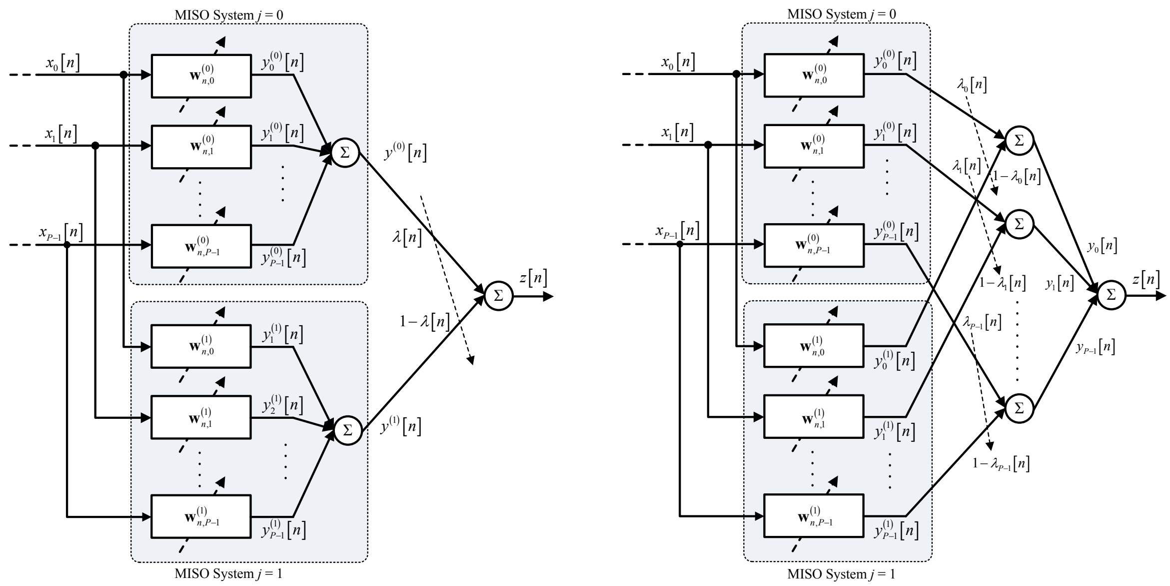
$$z[n] = \lambda[n]\, y^{(0)}[n] + (1 - \lambda[n])\, y^{(1)}[n] \tag{11.4}$$

where $\lambda[n]$ is the *mixing parameter* (see Chapter 10). Therefore, the beamformer output signal $e[n]$, using the system-by-system combination, is achieved as $e[n] = d[n] - z[n]$.

The mixing parameter in (11.4) is usually updated using a gradient descent rule through the adaptation of an auxiliary parameter, $a[n]$, related to $\lambda[n]$ by a sigmoidal activation function, similarly to (10.12).

11.3.3 Filter-by-filter combined architecture

The second proposed scheme is the *filter-by-filter* CANC in which the output signal $z[n]$ is built in a different way. As it is possible to see in Fig.



(a) (b)

Fig. 11.2: *Combined adaptive noise canceller architectures: (a) system-by-system and (b) filter-by-filter combination schemes.*

11.2 (b), the p-th filter output of the first MISO system is convexly combined with the correspondent p-th filter output of the second MISO system, thus generating $P-1$ outputs, each relative to a noise reference:

$$y_p[n] = \lambda_p[n]\, y_p^{(0)}[n] + (1 - \lambda_p[n])\, y_p^{(1)}[n] \tag{11.5}$$

where $\lambda_p[n]$ is the p-th *mixing parameter*, adapted using the p-th auxiliary parameter, $a_p[n]$, similarly to (10.12). Once computing the convex combinations, it is possible to achieve the CANC output signal $z[n]$ by summing the individual output contributions deriving from the combinations, as it is possible to see in Fig. 11.2 (b):

$$z[n] = \sum_{p=0}^{P-1} y_p[n] \tag{11.6}$$

from which we derive the overall beamformer output signal $e[n] = d[n] - z[n]$, relative to the filter-by-filter combination scheme.

Both the combined architectures presented above improve the tracking capabilities of CANC giving robustness to the overall beamforming system in nonstationary environments.

11.4 MULTI-STAGE MICROPHONE ARRAY BEAMFORMING

The microphone beamforming schemes described in Section 11.3 are effective in presence of multiple nonstationary sources both choosing different step size values (μ_0 small and μ_1 large) and different projection orders ($K_0 = 1$ and $K_1 > 1$). However, further improvements may be achieved if we consider the joined capabilities deriving from choosing both different step size values and projection orders. To this end we propose a *multi-stage combined architecture* in which the filtering process may involve more convex combinations of MISO systems.

In particular, in order to yield an adaptive beamforming architecture robust against adverse conditions, we may consider a CANC composed of a number $J = 4$ of MISO systems, as depicted in Fig. 11.3, each bringing different capabilities to the whole architecture. We differentiate by twos the four systems according to the step size values and the projection orders. In particular, we choose a small step size $\mu_j = \mu_{\mathrm{A}}$ for $j = 0, 2$ and a large step size value $\mu_j = \mu_{\mathrm{B}}$ for $j = 1, 3$. Moreover, we update the first two MISO systems using a gradient-based algorithm and the second two systems with a Hessian-based algorithm. This is obtained by setting a unitary projection order $K_j = 1$ for $j = 0, 1$ and a superior projection order $K_j > 1$ for $j = 2, 3$.

The choice of different step size values affects the convex combinations on the first stage, in which the first MISO system is combined with the second and the third with the four. In this stage the convex combination may follow the system-by-system scheme or the filter-by-filter scheme. In Fig. 11.3 a multi-

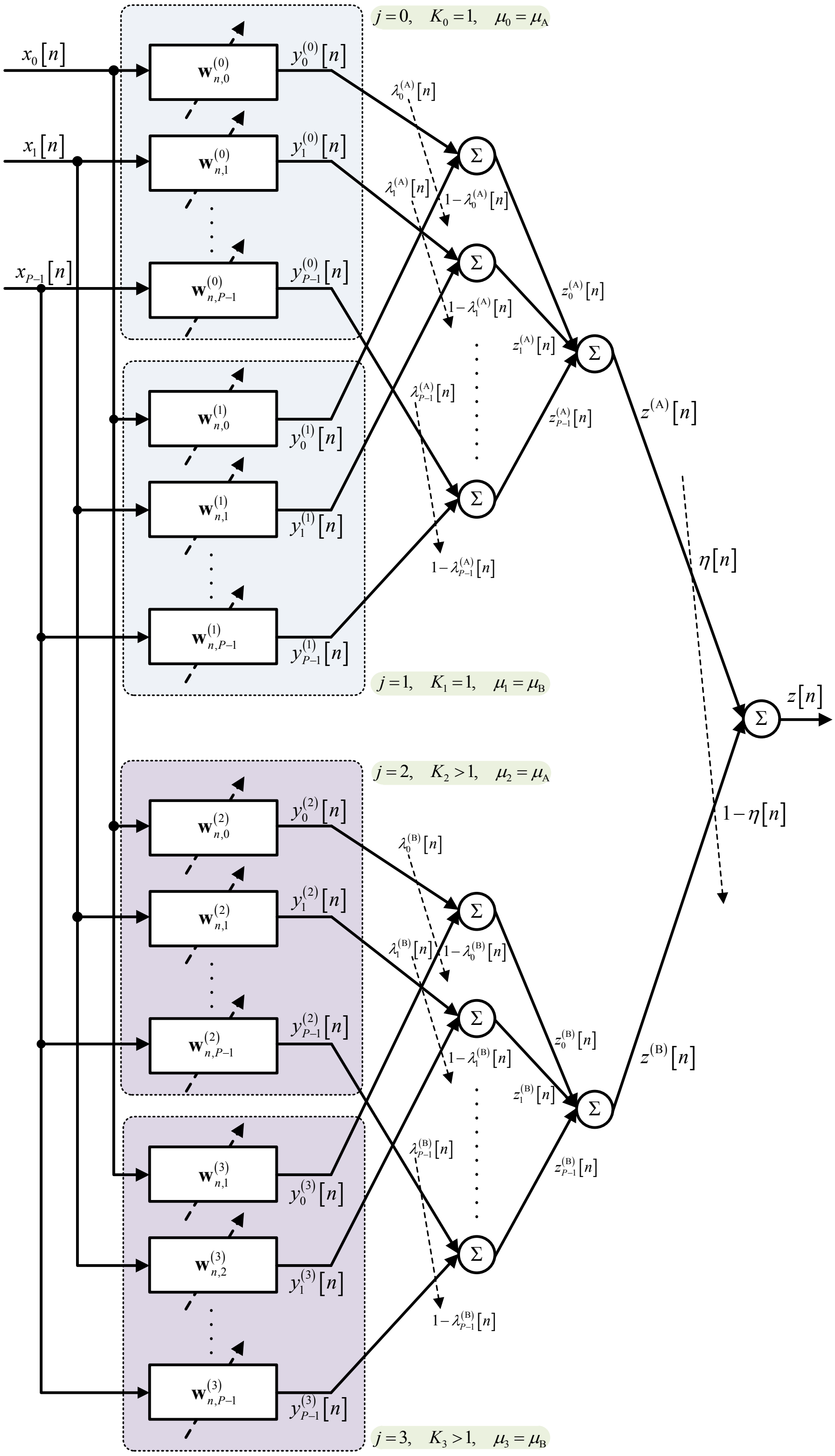


Fig. 11.3: *Multi-stage combined adaptive noise canceller.*

stage beamformer with a filter-by-filter scheme on the first stage is depicted. On the other hand, the choice of different projection order affects the convex combination on the second stage, in which the output signal resulting from the combination of the first and the second MISO systems is in turn combined with the output signal resulting from the combination of the third and the four MISO systems. The convex combination on the second stage follows the system-by-system combination scheme.

Output signals of the convex combinations on the first stage, denoted as $z^{(\mathrm{A})}[n]$ and $z^{(\mathrm{B})}[n]$, may be achieved similarly to (11.4), according to a system-by-system combination scheme, or similarly to (11.6), according to a filter-by-filter combination scheme as depicted in Fig. 11.3. In turn, the convex combination on the second stage may be achieved according to a system-by-system scheme, thus resulting the following output signal from the multi-stage CANC:

$$z[n] = \eta[n]\, z^{(\mathrm{A})}[n] + (1 - \eta[n])\, z^{(\mathrm{B})}[n] \tag{11.7}$$

where $\eta[n]$ is the mixing parameter of the second stage, even adapted using an auxiliary parameter.

Once computing the second stage convex combination, it is possible to derive the overall multi-stage beamformer output signal $e[n] = d[n] - z[n]$, as done for the single-stage combination schemes in Section 11.3.

The multi-stage beamforming architecture introduced above exploits the capabilities of each MISO system, thus improving speech enhancement performance compared to both conventional beamformers (using a single MISO ANC) and single-stage combined beamformers in presence of nonstationary interfering signals.

11.5 EXPERIMENTAL RESULTS

In the this section we carry out two different sets of experiments: the first set, in Subsection 11.5.1, aims at assessing the effectiveness of the described combined filtering schemes adopted in the proposed beamforming method; the second set of experiments, detailed in Subsection 11.5.2, is performed to evaluate the proposed combined beamforming architectures for speech enhancement application in a multisource scenario.

11.5.1 Convergence performance of combined architectures

In the first set of experiments we prove the filtering effectiveness of proposed CANC schemes through a tracking analysis which describes the convergence performance. To this end we use conventional ANC MISO systems and the proposed combined architectures to identify an unknown nonstationary system and to compare their performance.

The initial optimal solution is formed with $M = 7$ independent random values between -1 and 1. In the following examples the initial system is: $\mathbf{w}_1^{\mathrm{opt}} = \begin{bmatrix} 0.4125 & 0.7632 & -0.5484 & -0.6099 & -0.4622 & -0.4826 & -0.5296 \end{bmatrix}^T$. 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 i.i.d. Gaussian random process. The length of the input signal is of $L = 10000$ samples. However, in order to study the ability of combined schemes to react to nonstationary environments, at time instant $n = L/2$ the system changes into $\mathbf{w}_2^{\mathrm{opt}} = \begin{bmatrix} -0.4223 & 0.0848 & -0.1228 & 0.3876 & 0.9950 & 0.9806 & -0.2700 \end{bmatrix}^T$. Furthermore, an additive i.i.d. noise signal $e_0[n]$ with variance $\sigma_0^2 = 0.01$ is added to form the desired signal.

In order to identify the unknown solutions $\mathbf{w}_1^{\mathrm{opt}}$ and $\mathbf{w}_2^{\mathrm{opt}}$ we use both conventional MISO systems and the adaptive combined filtering schemes described in Sections 11.3 and 11.4 and we compare their performance in terms of *excess mean square error* (EMSE), defined as $\mathrm{EMSE}[n] = \mathrm{E}\left\{\left(e[n] - e_0[n]\right)^2\right\}$,

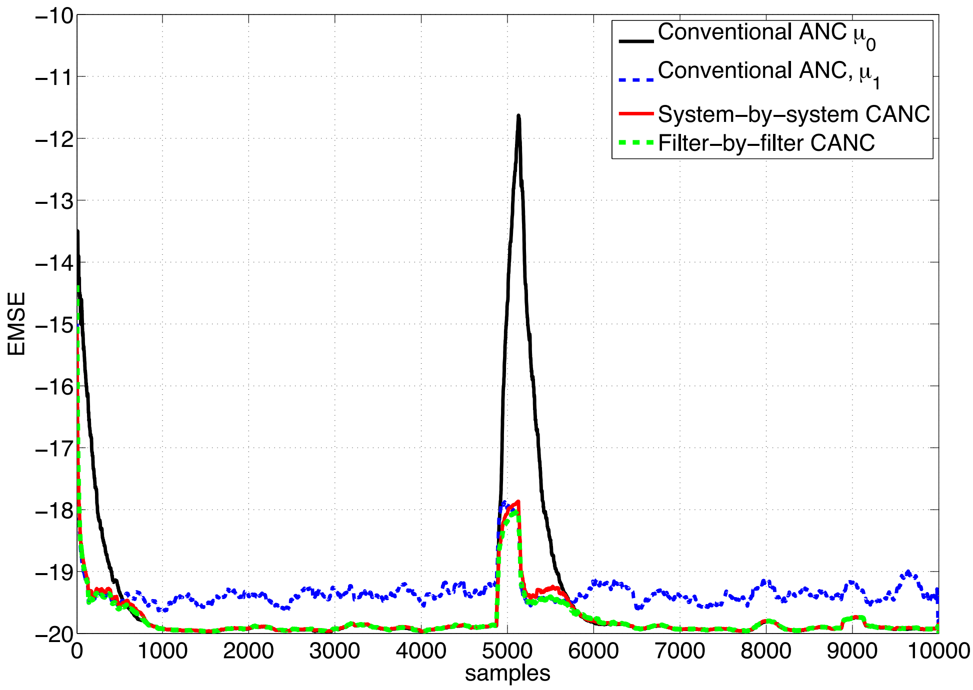


Fig. 11.4: *EMSE comparison between single-stage combined filtering architectures and conventional ones using the same projection order and different step size values.*

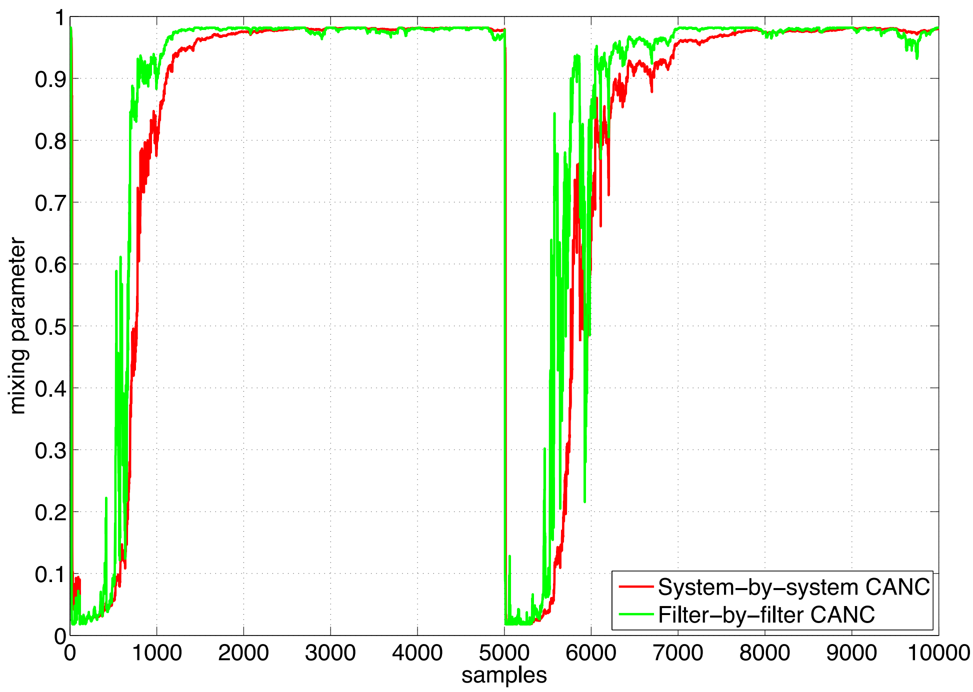


Fig. 11.5: *Behaviour comparison between the mixing parameter of the system-by-system CANC and the mixing parameter relative to the first channel of the filter-by-filter CANC.*

where $e[n]$ is the error signal of the filtering architecture, $e_0[n]$ is the additive noise signal (which is the same for all the filtering architectures) and the operator $\mathrm{E}\{\cdot\}$ is the mathematical expectation. The EMSE of each filtering structure is evaluated over 1000 independent runs. Moreover, in order to facilitate the visualization, the EMSE curves are filtered by a moving-average filter. All the filtering architectures, included the conventional ones, use MISO systems with $P = 4$ channels.

In a first experiment, we compare a conventional MISO architecture and both single-stage combined architectures described in Section 11.3, i.e. the system-by-system CANC and the filter-by-filter CANC. Both the system-by-system and the filter-by-filter schemes are composed of two MISO systems, as depicted in Fig. 11.2. All the MISO systems use APA filters. For the adaptation of the mixing parameter of the system-by-system filtering architecture we use a step size value of $\mu_{\mathrm{s}} = 10^2$, while a step size value of $\mu_{\mathrm{f}} = 10^3$ is adopted for the adaptation of all the mixing parameters of the filter-by-filter scheme. Both the step size values provide the best performance in each case. We evaluate the filtering architectures choosing the same projection order $K = 2$ for all the MISO systems and different step size values for the MISO systems of the combined schemes: a slower one $\mu_0 = 0.01$ and a faster one $\mu_1 = 0.1$. In Fig. 11.4 we have compared the performance of combined architecture with those of conventional ANC using both μ_0 and μ_1. As it is possible to see, both system-by-system and filter-by-filter schemes take advantage from using the combined filtering with respect to conventional filtering. In fact, combined schemes always show the behaviour of the best performing system and in transient state they behave even better than the best conventional filtering. Both the combined schemes provide good convergence performance, however the filter-by-filter scheme is slightly better than the system-by-system one due to the fact that the adaptation of the mixing parameters in the filter-by-filter scheme is faster than the system-by-system one, as can be seen in Fig. 11.5. This results in a quality improvement of the processed signal that can be

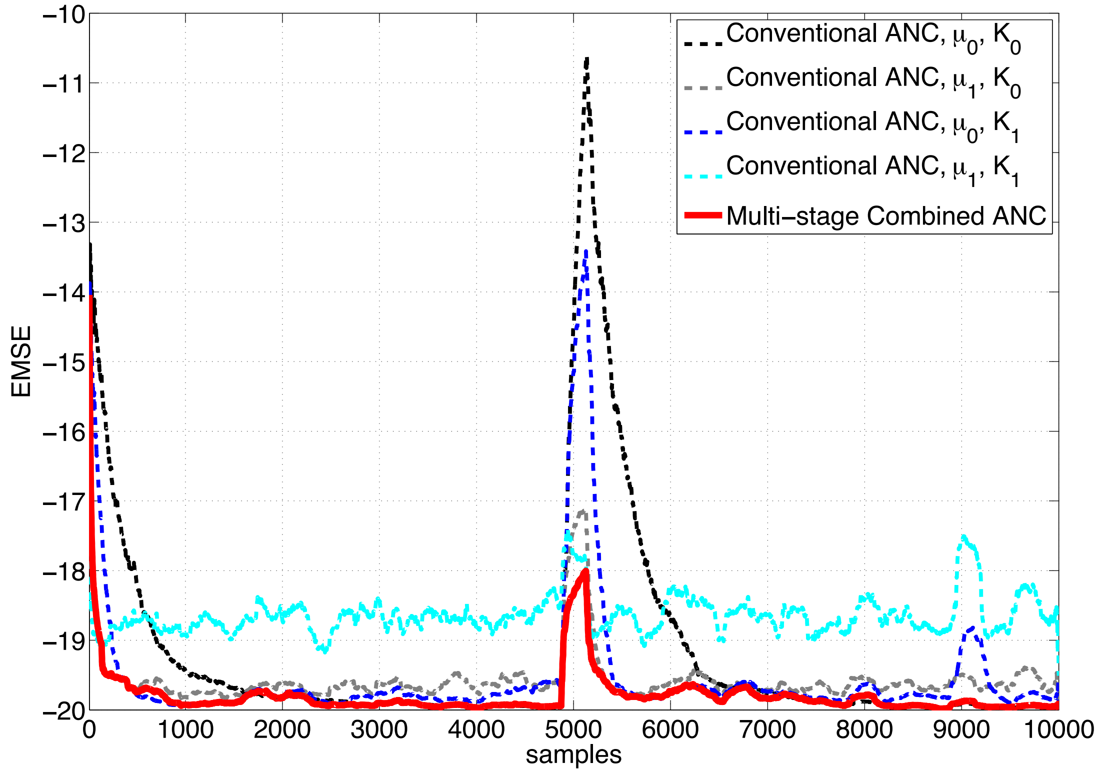


Fig. 11.6: *EMSE comparison between multi-stage combined filtering architectures and conventional ones.*

decisive in speech applications. A similar result was achieved choosing the same step size value and different projection orders.

In a second experiment, keeping the same scenario, we study now the convergence performance of the multi-stage combined architecture. As stated in Section 11.4, in a multi-stage combined scheme the combinations on the first stage may be performed in both system-by-system or filter-by-filter way. However, in light of previous result we take into account the performance of a multi-stage scheme whose combinations on the first stage are performed according to a filter-by-filter scheme, as depicted in Fig. 11.3. Therefore, we consider a two-stage combined scheme composed of four different MISO systems and we choose two different step size values, $\mu_0 = 0.01$ and $\mu_1 = 0.1$, and two different projection orders $K_0 = 1$ and $K_1 = 4$. In Fig. 11.6 the comparison between the multi-stage combined filtering architecture and the

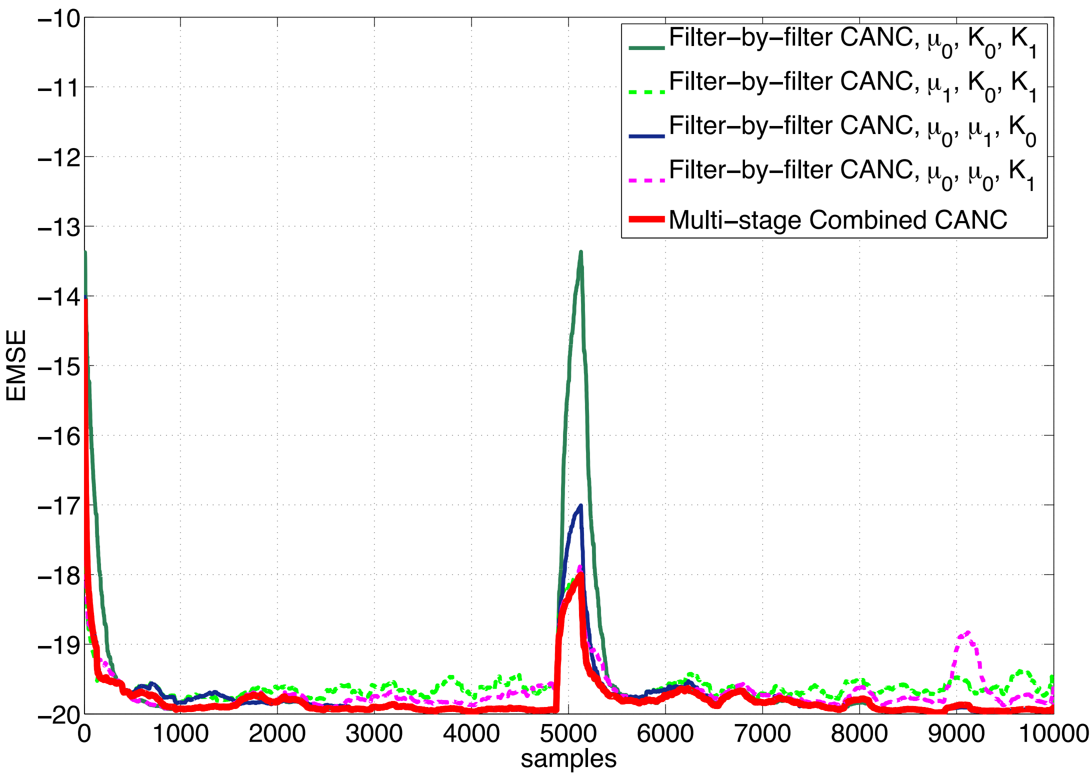


Fig. 11.7: *EMSE comparison between multi-stage combined filtering architectures and single-stage ones. Multi-stage combined architecture always provide the best overall performance.*

individual conventional filterings shows that the multi-stage filtering results the best performing architecture. Moreover, the performance improvement of the multi-stage architecture results even from the comparison with the single-stage filter-by-filter architectures, as depicted in Fig. 11.7.

Results achieved in this subsection the filtering ability of proposed combined schemes compared to conventional filtering. Moreover, a slightly preference is given to the filter-by-filter schemes which show a better reaction to abrupt changes in the environment due to the fact that the adaptive combination is performed for each channel. Furthermore, filter-by-filter schemes may exploit spatial diversity and thus different step size values for the adaptation of the mixing parameters may be chosen according to the scenario requirements. Finally, it has been shown that the multi-stage combined filtering always achieves the best convergence performance.

11.5.2 Speech enhancement evaluation of combined beamformers

In the second set of experiments we assess the effectiveness of the proposed combined beamforming architectures in terms of speech enhancement in multisource nonstationary environments. Experiments take place in a $6 \times 5 \times 3,3$ m room with a reverberation time of $T_{60} \approx 120$ ms. The source of interest is a female speaker located 70 cm from the center of the microphone array, as depicted in Fig. 11.8. Two interfering sources are initially located respectively $1,9$ m and $2,8$ m about from the center of the acoustic interface: the first source is a female speaker located on the left of the array, while on the right is located the second source which is a male speaker. White Gaussian noise is added at

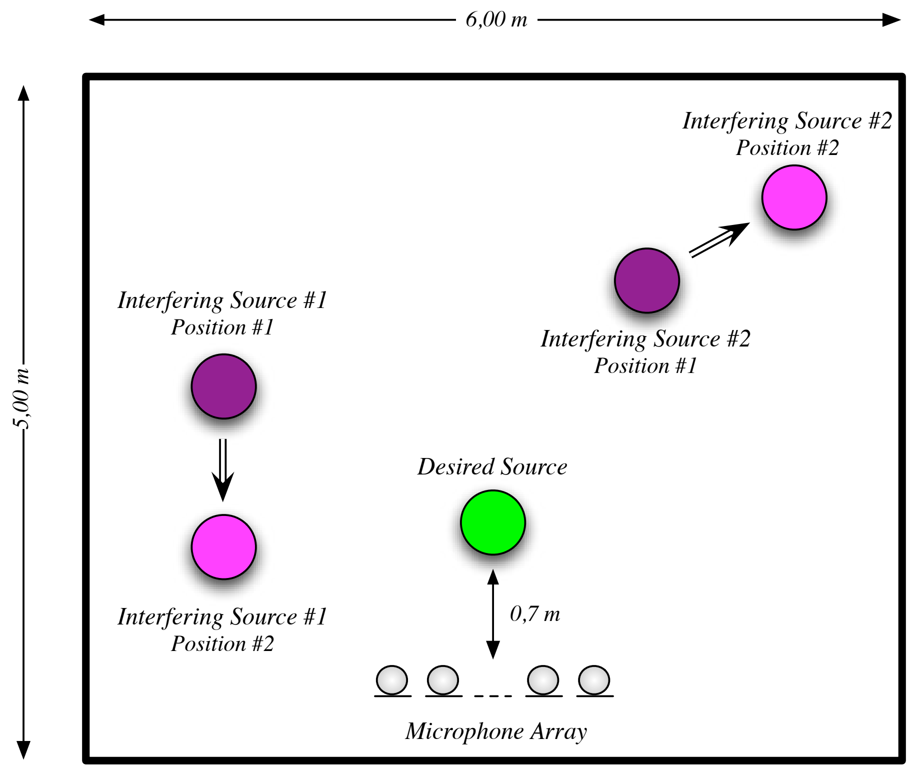


Fig. 11.8: *Speech enhancement nonstationary scenario. The source of interest is a female speaker located in front of the microphone array and two interfering speakers are located respectively on the left and on the right of the desired source. After 5 seconds from the start of the experiment the first interfering source moves to position 2 and at second 10 also the second interfering source changes its position.*

microphone signals as diffuse background noise, thus providing 20 dB of SNR (*signal-to-noise ratio*) with respect to the desired source. The overall input SNR level, measured for each microphone signal, is of about 3 dB. After 5 seconds from the start of the experiment the first source changes its position and at second 10 also the second source changes its position. The overall length of the experiment is of 15 seconds.

The AIRs between sources and microphones are simulated by means of *Roomsim*, which is a Matlab tool [24]. Each AIR is measured by using an 8 kHz sampling rate and it is truncated after $M = 340$ samples, which is also the length of each filter. The microphone interface is a classic *uniform linear array* (ULA) composed of 8 omnidirectional sensors equally spaced with a distance of 5 cm, thus having a good spatial resolution even at mid-low frequencies.

The enhancement of the speech, provided by the beamformer, and the resulting noise reduction, are usually associated with an SNR improvement, defined as [20]:

$$\mathrm{SNR} = 10\log\left[\frac{\mathrm{E}\left\{s_{in}^2\left[n\right]\right\}}{\mathrm{E}\left\{s_{in}^2\left[n\right]\right\} - \mathrm{E}\left\{s_{out}^2\left[n\right]\right\}}\right] \tag{11.8}$$

where $s_{in}\left[n\right]$ is the generic input clean signal and $s_{out}\left[n\right]$ is the processed signal. We compute the SNR level over the total length of the experiment ($0 - 15$ seconds) and also in 3 different sub-intervals of time: the initial state, from $0-5$ seconds, when the two interfering sources are located in their initial position; the first change, from $5-10$ seconds, which includes the position change of the first interfering source and the consequent readaptation of the filtering system; the second change, from $10 - 15$ seconds, when also the second interfering source changes its position. We compare GSC beamformers having different ANCs: conventional ANCs with with different parameter settings, the single-stage filter-by-filter combined ANC with different parameter settings, and the two-stage combined ANC. Filter parameters μ_0, μ_1, K_0, K_1, μ_s and μ_f are the same used in the first set of experiments. Results are collected in

GSC	0-5 s	5-10 s	10-15 s	0-15 s
Conventional ANC, μ_0, K_0	17.2	14.2	14.9	15.6
Conventional ANC, μ_0, K_1	17.8	16.7	16.8	16.9
Conventional ANC, μ_1, K_0	18.1	16.3	16.5	16.8
Conventional ANC, μ_1, K_1	13.4	13.2	13.4	13.4
FF CANC, μ_0, K_0, K_1	18.4	17.0	18.1	18.0
FF CANC, μ_1, K_0, K_1	18.2	17.5	18.0	17.9
MFF CANC, μ_0, K_0, K_1	18.8	18.1	19.1	18.7

Table 11.1: *SNR comparison in dB. We evaluate the beamformers over three sub-intervals of time, 0-5, 5-10 and 10-15 seconds, and over the total length of the experiment, 0-15 seconds. Multi-stage combined beamformer always performs the best reduction of interfering signals.*

Table 11.1, in which it is possible to notice the behaviour of the different beamformers taken into account and their contribution to the noise reduction in terms of SNR improvement. We could have shown performance of both system-by-system and filter-by-filter combination schemes and both varying the step size values and the projection order. However, for a better ease of reading results, we only show the performance relative to filter-by-filter combination schemes, which achieve the more relevant results, and we only vary the step size value for the single-stage combined ANCs. From Table 11.1 can infer that all the conventional ANCs show difficulties when a source change its position, thus decreasing speech enhancement performance. The more stable conventional ANC is the one having a large step size value and a large projection order, however, its performance is the poorest in terms of SNR. A significant enhancement is achieved by means of the filter-by-filter combined ANC (FF CANC) and a further improvement is provided by the multi-stage filter-by-filter combined ANC (MFFC ANC) which achieves the best performance in each time interval in terms of SNR.

SNR values obtained from this experiment are not definitely the best achievable values, since better results may be obtained using more sophisticated GSC beamformers, i.e. involving any *voice activity detectors* (VADs) and post-filters. However, the obtained results are sufficient to show the effectiveness of the proposed combined beamformers compared to conventional methods. Further results can be found in [28].

11.6 CONCLUSIONS

In this chapter we have introduced novel beamforming methods whose goal is to improve the performance, in terms of speech enhancement, in presence of a multisource nonstationary environment . The trademark of proposed methods relies on the use of combined filtering schemes in the ANC block. These combined schemes are based on the adaptive combination of MISO systems with different parameter settings thus involving complementary capabilities. The whole beamforming system benefits from the different capabilities of each MISO systems, yielding improved performance. We introduced two different way of combining the MISO systems which are the system-by-system scheme and the filter-by-filter one. Both the combined architectures provides better performance compared to conventional beamformers, however filter-by-filter schemes are slightly preferable due to the fact that the adaptive combination is performed for each channel. This allows filter-by-filter beamformers to better react to abrupt changes in the environment and to exploit spatial diversity by choosing different step size values for the adaptation of the mixing parameter of each channel. Finally, a multi-stage combined beamformer has been introduced in which the adaptive combination of MISO systems can be performed in subsequent stages. In particular, we have taken into account a two-stage combined beamformer which outperforms the single-stage schemes, thus always providing the best performance when nonstationary sources interfere with the enhancement of a desired speech signal.