Pirro *et al. EURASIP Journal on Audio, Speech, and Music Processing* 2012, **2012**:26
http://asmp.eurasipjournals.com/content/2012/1/26

EURASIP Journal on
Audio, Speech, and Music Processing
a SpringerOpen Journal

RESEARCH **Open Access**

Stereophonic hands-free communication system based on microphone array fixed beamforming: real-time implementation and evaluation

Matteo Pirro, Stefano Squartini*, Laura Romoli and Francesco Piazza

Abstract

In this article, the authors propose an optimally designed fixed beamformer (BF) for stereophonic acoustic echo cancelation (SAEC) in real hands-free communication applications. Several contributions related to the combination of beamforming and echo cancelation have appeared in the literature so far, but, up to the authors' knowledge, the idea of using optimal fixed BFs in a real-time SAEC system both for echo reduction and stereophonic audio rendering is first addressed in this contribution. The employment of such designed BFs allows positively addressing both issues, as the several simulated and real tests seem to confirm. In particular, the stereo-recording quality attainable through the proposed approach has been preliminarily evaluated by means of subjective listening tests. Moreover, the overall system robustness against microphone array imperfections and noise presence has been experimentally evaluated. This allowed the authors to implement a real hands-free communication system in which the usage of the proposed beamforming technique has proven its superiority with respect to the usual two-microphone one in terms of echo reduction, and guaranteeing a comparable spatial image. Moreover, the proposed framework requires a low computational cost increment with regard to the baseline approach, since only few extra filtering operations with short filters need to be executed. Nevertheless, according to the performed simulations, the BF-based SAEC configuration seems to not require the signal decorrelation module, resulting in an overall computational saving.

Keywords: Fixed beamforming, Microphone array, Real-time stereophonic acoustic echo cancelation, Hands-free communication, Stereophonic recording

1 Introduction

Nowadays, the flexibility provided by hands-free communication devices has revolutionized the way humans communicate. Nonetheless, when one deals with hands-free systems there are several issues to face and problems to solve. The most relevant one is the echo presence, due to the acoustic coupling between microphones and loudspeakers located in the same room. Thus, an acoustic echo canceler is needed and several algorithmic solutions have been proposed in the literature in the last two decades [1]. More recently, the academic and technology market interest has been attracted by the chance to employ spatial audio techniques to enhance the sound realism in teleconferencing systems. Thus, many solutions have been proposed for multiparty conferencing where more microphones and loudspeakers are involved in each room. As a consequence of this, suitable multichannel AEC algorithms have been developed to deal with the echo problem in presence of multiple audio paths, where the task to be solved is tougher than in the single-channel case study, as rigorously illustrated in [2]. Indeed, the "non-uniqueness problem" occurs in the multichannel scenario, due to the high correlation degree between recorded signals: a very popular involved technique involves the addition of a decorrelation module to allow multichannel adaptive filtering working properly [2-5].

Moreover, microphone arrays have become essential in many applications related to music or voice recording,

*Correspondence: s.squartini@univpm.it
A3LAB - Department of Information Engineering, Università Politecnica delle Marche, Via Brecce bianche 1, 60131 Ancona, Italy

processing, and transmission [6]. More specifically, microphones arrays are particularly useful in order to remove from the desired signal interference due for example to noise, reverberation or acoustic echoes. In fact, spatial filtering, in addition to temporal filtering, is very efficient because the sources of interferences are usually spatially located away from speaker in many environments.

Data-dependent adaptive beamformers (BFs) have proven to be the theoretically optimal choice because they can maximize noise and interference suppression. However, in real environments several reasons (e.g., room reverberation, arrays and microphone imperfections, high temporal variability of the typical interferences) lead to a recorded signal distortion and to a complex integration in advanced systems [7].

Nevertheless, integration of an echo canceler with a microphone array [8], in order to make the most of the positive synergies among them, is suitable to better deal with real-life scenarios: in fact, a crucial issue in AEC is the separation of the local talker from unwanted audio interferences (as background noise and echo itself). With the aim of optimizing the overall AEC performance, several configurations have been thus proposed, even in the multichannel case study [9,10]. Microphones arrays have been also used to enhance stereophonic teleconferencing in [11].

In this contribution a very low computational cost solution based on two optimally designed fixed BFs is proposed to suitably deal with real-life teleconferencing scenarios, where speakers are in two different rooms equipped with a large television display and there are audio recording microphones on the top of the screen and two loudspeakers for stereo reproduction on its sides. The developed architecture allows to improve the ordinary performance of SAEC algorithms: the echo power is significantly reduced with respect to the standard two-microphone based implementation, the distortions often occurring in adaptive beamforming are completely avoided, and the acoustic spatiality effect is guaranteed. The proposed approach also ensures a stereophonic-like audio recording, due to the designed beampattern shapes, resulting in an enhanced quality of communication in terms of spatial localization of teleconference speakers. Up to the authors' knowledge, this represents the first attempt to propose a real-time BF-based hands-free communication system taking both echo reduction performance and stereophonic experience into account. Moreover, it will be experimentally proven that the BF-based SAEC configuration does not practically suffer of the stereo signal correlation problem, allowing to neglect the decorrelation module, thus obtaining a certain computational saving. The issue has been already and preliminarily faced in [12]: this article represents an extended version of the previous contribution, providing more experimental results, in terms of system robustness analysis and behavior in real acoustic environments, together with a more detailed description of employed algorithms.

The article is organized as follows: Section 2 provides an overview of all algorithms used in this study, with a special care to the adopted BF synthesis approach. Section 3 details all the experiments done to evaluate the effectiveness of the idea. First, subjective tests regarding the stereophonic recording capability attained by means of the beamforming approach are discussed. Then, some simulated tests, accomplished to make a first assessment of the system, are described. Finally, system robustness is examined as preliminary and necessary step before going to consider the real system implementation, whose main issues and results are also faced in this study. Section 4 concludes the article and provides some hints for future developments.

2 The overall stereophonic hands-free communication system

The system used in this article for stereophonic acoustic echo cancelation (SAEC) experiments consists of:

- two fixed BFs, one per audio channel in a stereo-based communication;
- a decorrelation block;
- a stereophonic acoustic echo canceler made of four adaptive filters;
- a double-talk detection (DTD) module.

Microphone arrays and beamforming techniques jointly play a relevant role in many applications related to music or voice recording, processing, and transmission [6]. Due to the spatial filtering capability, they allow removing annoying disturbances, like noise, reverberation, and acoustic echoes from the desired signal. Indeed, focusing on the echo cancelation problem, several beamforming based algorithmic architectures have been proposed in the literature so far [9,10], also in the multichannel case study [11]. This study proposes a fixed beamforming algorithm to be included within the SAEC system, typically composed of the decorrelation block, the adaptive filters, and the double-talk detection module.

The use of a two-channel (stereo) system allows one to obtain satisfactory results in terms of spatial information that helps listeners to identify the speaker position. Unfortunately, the problem of SAEC cannot be viewed as a simple generalization of the mono-channel case for different reasons [2]: indeed, more adaptive filters have to be identified and the linear relationship existing between the two channels generated from the same source brings convergence problems related to the ill-conditioned covariance matrix. Therefore, a method to reduce the interchannel coherence must be introduced in order to obtain good echo cancelation performance providing the slightest alteration of the sound perception [1,13].

As regards the adaptive filtering algorithm, the use of a frequency-domain adaptive filtering algorithm is due to its interesting computational cost: as a matter of fact, the adaptation procedure is performed by taking advantage of the fast Fourier transform (FFT) efficiency (Fast-LMS) [14], allowing an improved convergence with low computational requirements [15,16], even in presence of long impulse responses (IRs). The main drawback of this algorithm is the input-output delay, because it is equal to the adaptive filter length: long IRs to be identified require long adaptive filters, that means a high processing latency. In [17], the generalization of frequency-domain adaptive filtering to the partitioning of the adaptive filter is proposed: in this case, the filter length is a positive integer multiple of the block size, which can be opportunely reduced in order to lower the input-output latency.

Finally, the system for SAEC must include a double-talk detection module which "freezes" the adaptation procedure whenever a near-end signal is detected in order to avoid the divergence of the adaptive algorithm [18,19]. Subsequently, a suitable DTD decision variable ξ has to be found. An optimal decision variable for double-talk detection should behave as follows:

- if double-talk is not present, $\xi > T$;
- if double-talk is present, $\xi < T$;

where T is a constant.

The block diagram of the whole system is shown in Figure 1. Details regarding the aforementioned blocks are reported in the following sections.

2.1 Fixed beamforming algorithm

The present section is devoted to the description of the approach used to develop two broadband linear array beamformers characterized by complementary beampatterns. Such a spatial complementarity is aimed at obtaining a stereophonic sound recording capability.

The procedure followed for the BFs design is based on the variably-weighted least squares criterium [20]. Let us consider a linear array consisting of M microphones and that the BFs work under the far-field assumption. Each microphone signal is processed by an Q_{BF}-tap finite impulse response (FIR) filter $\mathbf{w}_m$ (with $m = 0, 1, \ldots, M-1$). According to these assumptions, the array steering vector can be expressed as follows:

$$\mathbf{g}(f,\theta) = \begin{bmatrix} 1 \\ e^{-j2\pi f/f_s} \\ \vdots \\ e^{j2\pi(Q_{\mathrm{BF}}-1)f/f_s} \end{bmatrix} \otimes \begin{bmatrix} e^{-j2\pi f\tau_0} \\ e^{-j2\pi f\tau_1} \\ \vdots \\ e^{-j2\pi f\tau_{M-1}} \end{bmatrix}, \qquad (1)$$

where f_s is the sampling frequency, $\otimes$ is the Kronecker product, θ is the angle of incidence of the plane wave, and $\tau_m = d_m \cos\theta / c (m = 0, 1, \ldots, M-1)$ are the time delays from the mth microphone to the center of the array, with d_m the distance between the mth microphone and the center of the array and $c \approx 340\,\mathrm{m/s}$ the sound speed in air.

The BF response is given by

$$P(f,\theta) = \mathbf{w}^T \mathbf{g}(f,\theta) \qquad (2)$$

where the superscript $()^T$ represents the transpose operator, and

$$\mathbf{w} = [\,\mathbf{w}_0^T, \mathbf{w}_1^T, \ldots, \mathbf{w}_{M-1}^T]^T \qquad (3)$$

is the BF weight vector. In this context, the steering vector is given by

$$\mathbf{g}(f,\theta) = \begin{bmatrix} 1 \\ e^{-j2\pi f/f_s} \\ \vdots \\ e^{j2\pi(Q_{\mathrm{BF}}-1)f/f_s} \end{bmatrix} \otimes \mathrm{diag}\{A(\theta)\} \begin{bmatrix} e^{-j2\pi f\tau_0} \\ e^{-j2\pi f\tau_1} \\ \vdots \\ e^{-j2\pi f\tau_{M-1}} \end{bmatrix}, \qquad (4)$$

where $A(\theta) = \alpha + (1-\alpha)\cos(\theta)$, α describes the microphone polar pattern ($\alpha = 0.25$ for hypercardioid, $\alpha = 1$ for omnidirectional, and $\alpha = 0.5$ for cardioid microphones), and diag{} makes a square matrix whose diagonal has $A(\theta)$ elements.

The original design procedure aims to obtain a BF which satisfies the frequency-angle passband shaping constraints (Ω_p, Θ_p) while holding stopband level (Ω_s, Θ_s) as low as possible. Furthermore, in order to obtain a sound stereophonic recording two different and complementary BFs have been designed. One of them with $\Theta_s = (0° - 20°, 90° - 180°)$ and $\Theta_p = (20° - 90°)$ and the other one with $\Theta_s = (0° - 90°, 160° - 180°)$ and $\Theta_p = (90° - 160°)$. In this way, by using them simultaneously, it is possible to record sound coming from the left and the right side of the room. We have chosen $M = 7$ and $Q_{\mathrm{BF}} = 32$. Figure 2A,B depict the synthesized beampatterns of the left oriented and right oriented BFs with hypercardioid microphones, respectively. Considering that the sampling frequency is 48 KHz, figures above show how the beampatterns have an omogeneous behavior up to 12 KHz, that can be considered adequate to deal with speech signals.

2.2 Decorrelation algorithm

The approach used in this article is described in [5]: it is based on a time-varying phase modulation according to the human ear sensitivity which is high at low frequencies and gradually decreases with increasing frequency. First, signals are decomposed into subbands using the complex

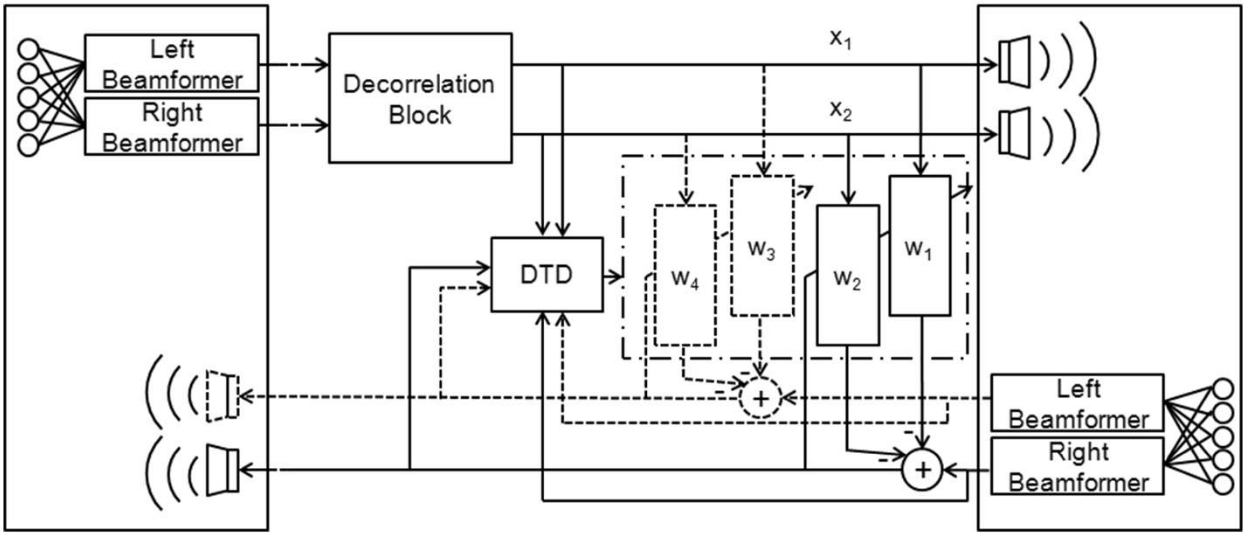


Figure 1 Block diagram of the overall stereophonic hands-free communication system (the echo canceler is shown only for the receiving room).

modulated lapped transform (CMLT) which is a complex value transform that preserves phase information; then, the phase modulation is applied to the transformed signals according to the specific frequency. Taking advantage of the CMLT, phase modulation is applied to the two channels of the stereo signal in a conjugate complex way through a complex multiplication with $e^{\pm j\varphi(t,s)}$, where $\varphi(t,s) = a(s)\sin(2\pi f_m t)$ is the signal phase, depending on the continuous time t and the subband s, with $a(s)$ being the phase modulation term at subband s. It is worth noting that it is possible to develop an efficient CMLT implementation taking advantage of the FFT algorithm, providing a suitable solution for real-time scenarios [21].

2.3 Adaptive filtering algorithm

The main equations describing the adaptive filtering algorithm used in this article are reported [22]. The impulse response of length N is partitioned into K partitions. Assuming a two-channel scenario, where $\mathbf{X}_{k,m}^{(p)}$ and $\mathbf{w}_{k,m-1}^{(p)}$ are the input signal and the filter coefficients for each channel p $(p = 1,2)$ and partition k $(k = 1,2,\ldots,K)$ at each block index m, the estimated echo $\mathbf{y}_m$ is derived as follows:

$$\mathbf{y}_m = \sum_{k=1}^{K}\left(\mathbf{X}_{k,m}^{(1)}\mathbf{w}_{k,m-1}^{(1)}\right) + \sum_{k=1}^{K}\left(\mathbf{X}_{k,m}^{(2)}\mathbf{w}_{k,m-1}^{(2)}\right). \qquad (5)$$

Then, the residual echo signal $\mathbf{e}_m$ is obtained by subtracting $\mathbf{y}_m$ from the microphone signal $\mathbf{d}_m$:

$$\mathbf{e}_m = \mathbf{d}_m - \mathbf{y}_m. \qquad (6)$$

Finally, the filter coefficients are updated as follows

$$\begin{aligned}\mathbf{w}_{k,m}^{(1)} &= \mathbf{w}_{k,m-1}^{(1)} + \mu_m\nabla_{k,m}^{(1)} \text{ for k = 1,2,\ldots,K}\\ \mathbf{w}_{k,m}^{(2)} &= \mathbf{w}_{k,m-1}^{(2)} + \mu_m\nabla_{k,m}^{(2)} \text{ for k = 1,2,\ldots,K,}\end{aligned} \qquad (7)$$

where μ_m is the fixed convergence speed μ normalized by the power of the input signal at each block index m (i.e., block NLMS [16]) and $\nabla_{k,m}^{(p)}$ is the block gradient estimate for each channel p $(p = 1,2)$ and partition k $(k = 1,2,\ldots,K)$ at each block index m. Differently from [22], a fixed step-size has been considered. The FFT efficiency has been exploited for the implementation of the block NLMS [16], e.g., for the filtering operation and for the computation of the block gradient estimate, applying suitable constraints in order to avoid circular convolution.

2.4 Double talk detector

In this article, the algorithm used for the DTD module is based on two control variables, ξ_1 and ξ_2, with the aim of obtaining a more correct detection especially in presence of low-level signals. The former is based on the approach discussed in [18], and employs the cross-correlation between the far-end signal and the microphone signal. The latter is based on the approach described in [19], i.e. on the cross-correlation between the residual echo signal and the microphone signal. Double-talk is detected whenever

$$\xi_1 < T_1 \text{ and } \xi_2 > T_2, \qquad (8)$$

where T_1 and T_2 are properly selected. Experimental results have shown a more correct and robust behavior

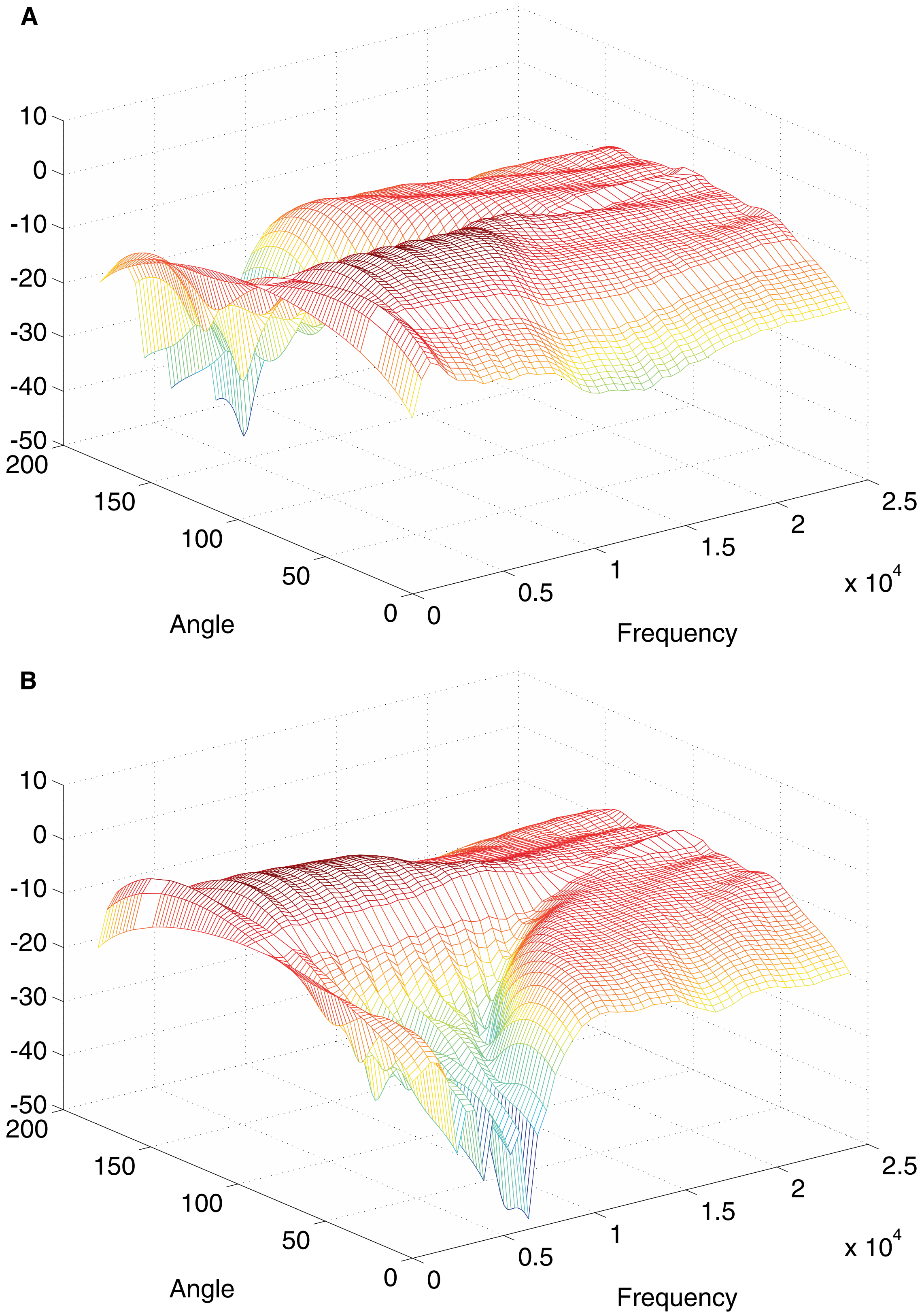


Figure 2 Beampattern of left oriented and right oriented BFs with hypercardioid microphones. A Left oriented beampattern. **B** Right oriented beampattern.

when double-talk occurs, considering $T_1 = 0.8$ and $T_2 = 0.5$.

Algorithm 1 summarizes the main steps to be performed, considering all the aforementioned blocks.

Algorithm 1 Overall stereophonic hands-free communication system description

```
while streaming audio do
    Far-end Input blocks (left and right channel)
    with 50% overlap;
    for each channel $p$ do
        Complex multiplication for phase
        modulation (using CMLT filter
        bank) applied to far-end signals;
    end
    Local reproduction of processed far-end
    signals;
    Near-end Input blocks (microphone signals);
    Optimally designed left BF applied to
    near-end signals, providing the left channel;
    Optimally designed right BF applied to
    near-end signals, providing the right channel;
    Estimated echo signal $y$ computation;
    Residual echo signal $e$ computation;
    DTD control variables $\xi_1$ and $\xi_2$ computation;
    if $\xi_1 < T_1$ and $\xi_2 > T_2$ then
        no adaptation (double-talk)
    else
        for each channel $p$ do
            for each filter section $k$ do
                Filter section $w_k^{(p)}$ adaptation;
            end
        end
    else
    Output blocks (left and right channel);
end
```

2.5 Computational cost

Computational costs of different system blocks have been shown in Table 1 in terms of arithmetic operations per sample. The parameters of interest are described as follows: Q_{SAEC} and K are the framesize and the partitions number used by stereophonic acoustic echo canceler, respectively; K_{DTD} is the partition number used by Double-Talk Detector; Q_{BF} and M are the framesize and the microphone number used by Beamformer; finally, for a generic framesize Q, the term $2Q \log(2Q)$ represents the FFT computational cost.

Let us provide a numeric example. Given the sampling frequency equal to 48 KHz and assuming $Q_{\mathrm{SAEC}} = 512$, $K = 16$, $K_{\mathrm{DTD}} = 8$, $Q_{\mathrm{BF}} = 32$, and $M = 7$, the computational cost in terms of million floating point operation per second (MFLOPS) results $\simeq$ 198.8 for SAEC module, $\simeq$ 125.0 for DTD module, $\simeq$ 2.7 for decorrelation module, and $\simeq$ 2.5 for BF module. It can be observed how the computational burden of this latter module is negligible in comparison with the first two.

Table 1 Computational cost of different system blocks

	Computational cost
Stereophonic acoustic echo canceler	$(2 * Q_{\mathrm{SAEC}} * (8 * (1 + K) * \log(2 * Q_{\mathrm{SAEC}}) + 1)+$ $+(88K + 9) * (Q_{\mathrm{SAEC}} + 1))/Q_{\mathrm{SAEC}}$
Double-talk detector	$((136 * K_{\mathrm{DTD}} + 40) * (Q_{\mathrm{SAEC}} + 1)+$ $+34 * Q_{\mathrm{SAEC}} + 210 + 16 * (1 + K_{\mathrm{DTD}})+$ $+ * Q_{\mathrm{SAEC}} \log(2 * Q_{\mathrm{SAEC}}))/Q_{\mathrm{SAEC}}$
Decorrelation module	$(2 * (2 * (2Q_{\mathrm{SAEC}} * \log(2Q_{\mathrm{SAEC}})) + 2Q_{\mathrm{SAEC}}+$ $+4(Q_{\mathrm{SAEC}} + 1)) + 2 * 2 * Q_{\mathrm{SAEC}} * 4)/Q_{\mathrm{SAEC}}$
Beamformer	$M * (2 * 2Q_{\mathrm{BF}} * \log(2Q_{\mathrm{BF}}) + 6 * 2Q_{\mathrm{BF}})/Q_{\mathrm{BF}}+$ $+((M - 1) * (Q_{\mathrm{BF}} + 1))/Q_{\mathrm{BF}}$

3 Experimental tests

This section is aimed at experimentally proving the effectiveness of the beamforming based technique for stereophonic hands-free communications. The Intel integrated primitives (IPP) library [23], a high performance library for mathematical computation, has been used for the system implementation as NU-Tech Satellites (NUTSs) [24] on the NU-Tech framework [24], a suitable software platform for real time audio processing directly on a PC hardware. Tests have been carried out on two PCs Dell Precision T1500 with Intel i5 Core. Different tests have been performed and they have been arranged in four groups as follows:

- *Subjective evaluation of stereo recordings*: the objective is to evaluate to what extent two BFs designed as discussed in Section 2.1 are able to render the stereophonic effect in recordings;
- *Simulated tests*: some computer simulations are described to show the behavior of the overall system under ideal acoustic conditions;
- *System robustness tests*: with these tests, the authors want to assess the impact of microphone-array perturbations and noise in the overall system performance, always from a simulated scenario perspective;
- *Real tests*: once shown that the system is robust enough and can work well also under real acoustic

conditions, some real tests are performed and related results are discussed to draw final conclusions on the effectiveness of the proposed idea.

3.1 Subjective evaluation of stereo recordings

In this section, the results of the subjective listening tests are discussed. The objective is assessing the stereophonic effect provided by means of the employed beamforming technique.

3.1.1 Subjective tests setup

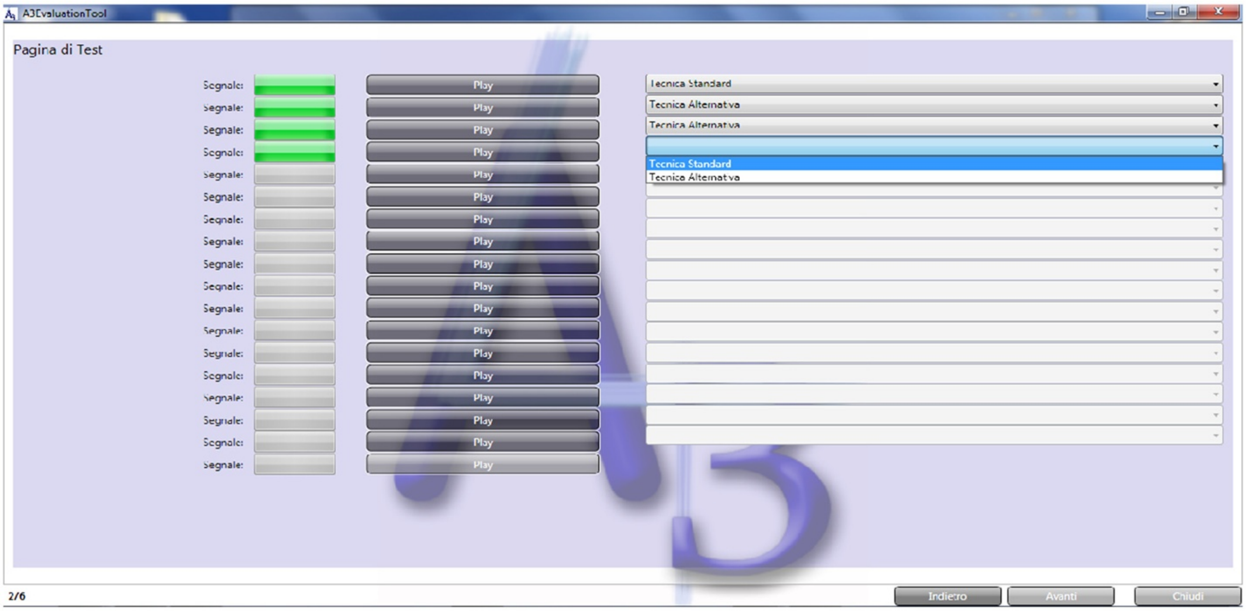
In these tests the opinions of 20 different subjects about 72 audio files have been collected. The goal of the tests consisted in evaluating the subjective difference between standard stereo recording technique and beamforming based approach in terms of stereophonic sound rendering. The tests were organized as follows. First, some audio examples relative to the real stereophonic recording method have been provided to the subjects just before starting their test session. This was made by stressing the focus on the acoustic spatial image, with the aim of providing a useful general reference from this perspective. Afterwards, during the real test, the subjects were asked to listen to a total amount of 72 audio files, recorded with the two techniques and randomly ordered, and then classify them according to their sound experience. Test audio files were synthetically created by using both male and female speakers and by varying different parameters: the position of the speaker, the value of reverberation coefficient, the distance between speaker and microphones, and the used technique.

It must be pointed out that, up to the authors' knowledge, no significant contributions are available as useful terms of comparison in the present application area, in contrast to other fields where many efforts have been made in the literature [25-28]. The subjective methods proposed therein have not been considered in the present experimental study since the objective was not comparing to what extent two stimuli are different or which one is closer to a certain reference, but how much confusable they are in terms of spatial image. Therefore, the listener was asked to classify them in a continuous sequence of listening tests, assessing its ability to distinguish the two techniques used to produce such an acoustic feature. The proposed approach has been inspired by the subjective validation methodology for musical instrument emulation algorithms adopted in [29].

The software platform used for the tests, namely A3LAB Evaluation Tool (Figure 3), is a Microsoft.NET C# application, configured after compile-time by modifying extensible application markup language (XAML) files. It is an interactive tool through which users can listen to the audio files and express their assessment grade or preference according to the evaluation method in usage.

3.1.2 Subjective evaluation of stereo recordings

Results of the performed preliminary subjective tests are reported in this section. By referring to [30,31], confusion



Tecnica Standard
Tecnica Alternativa
Tecnica Alternativa
Tecnica Standard
Tecnica Alternativa

Figure 3 A3LAB evaluation tool for subjective listening tests.

matrix has been created, taking the following definitions into account:

$$\text{Accuracy} = \frac{TM + TB}{TM + FM + TB + FB} \tag{9}$$

where TM, TB are the number of correct answers in identifying the standard ("M" stands for "microphone") and the BF based approach ("B" stands for "beamforming") for stereophonic recording, whereas FM, FB are the number of wrong answers again when "M" and "B" are expected to be recognized. The majority of *Accuracy* values are all included between 0.35 and 0.65 (with an average of 0.48), as depicted in Figure 4A: as pointed out in [30,31], a value equal to 0.5 means random guessing and obtained *Accuracy* values goes along this direction. Similar conclusions were drawn in [32], targeted to the synthesis of musical instrument tones, where the sound indistinguishability threshold was set to 0.75.

Moreover, results in terms of receiver operating characteristic (ROC) [31] are provided in Figure 4B. Each ROC point is relative to a single subject test session. It can be easily observed that all values are close to the square diagonal (and many of them even under it), meaning that a coherent classification of stereophonic recording techniques under test has not been accomplished by the subjects. Such classification would have corresponded to those ROC points located close to the ideal position (0,1). Concluding, reported results seem to reasonably support the similarity between the BF-based approach and the standard microphone configuration for stereophonic recording and spatial image rendering.

3.2 Simulated tests

In this Section, the results of PC simulations under ideal acoustic conditions are shown in terms of echo return loss enhancement (ERLE). More specifically, the following expression of ERLE has been considered:

$$\text{ERLE}(n) = 10 \log_{10} \frac{\varepsilon[e^2(n)]}{\varepsilon[y^2(n)]} \tag{10}$$

where the followings hold:

$$\varepsilon[e^2(n)] = \gamma\varepsilon[e^2(n)] + (1-\gamma)e^2(n) \tag{11}$$

$$\varepsilon[y^2(n)] = \gamma\varepsilon[y^2(n)] + (1-\gamma)y^2(n) \tag{12}$$

The $\varepsilon[.]$ stands for the Expectation operator whereas the terms $e(n)$ and $y(n)$ are the residual echo signal and the acoustic echo signal at sample n, respectively, and $0 < \gamma < 1$ is a forgetting factor. Therefore, a lower ERLE corresponds to a better SAEC performance.

3.2.1 Simulated tests setup

Various techniques for sound recording have been taken into consideration in order to enhance the system performance. Each technique aims at reducing the acoustic coupling between loudspeakers and microphones by focusing on the speaker's voice direction: in this way, two issues are addressed, i.e., the attenuation of the sound from loudspeakers and a partial decorrelation of the stereo signal. The different techniques under test are listed as follows:

- omnidirectional microphones with decorrelation module, taken as standard configuration, and hypercardioid/cardioid microphones with/without decorrelation module;
- hypercardioid microphones in stereo configuration with/without decorrelation module;
- BFs with omnidirectional/hypercardioid/cardioid microphones with/without decorrelation module.

The acronyms HYP, CARDIO, OMNI refer to hypercardioid, cardioid, and omnidirectional polar patterns, respectively. BF and 2MIC refer to beamformer and two-microphone stereo configurations. Finally, DEC indicates the activation of the decorrelation module. The different combinations of these acronyms denote the techniques listed above and tested in our experiments. For instance, the term "OMNI DEC" stands for the hands-free communication system with two omnidirectional microphones (not in stereo configuration) and involving the decorrelation module, whereas the term "BF HYP" stands for the beamforming based system with hypercardioid microphones and without the decorrelation module.

All simulations have been accomplished considering two identical rooms, as far-end and near-end. Room plants with a couple of microphones or with a microphone array are depicted in Figure 5. The talker has been originally positioned on the left side of the microphones. During simulation the source has been moved on the right side after 5 s and again on the left side after 11 s: indeed the SAEC system is particularly sensitive to room changes due to the non-uniqueness problem.

Echo cancelation has been performed using the algorithms described in Sections 2.2, 2.3, and 2.4. The adaptive filters are partitioned into $K = 8$ sections of length 512 taps. A female speaker active over a time range of 18 s and a sampling frequency of $f_s = 44100$ Hz has been considered. The number of samples of the simulated room IRs is equal to 4096 samples.

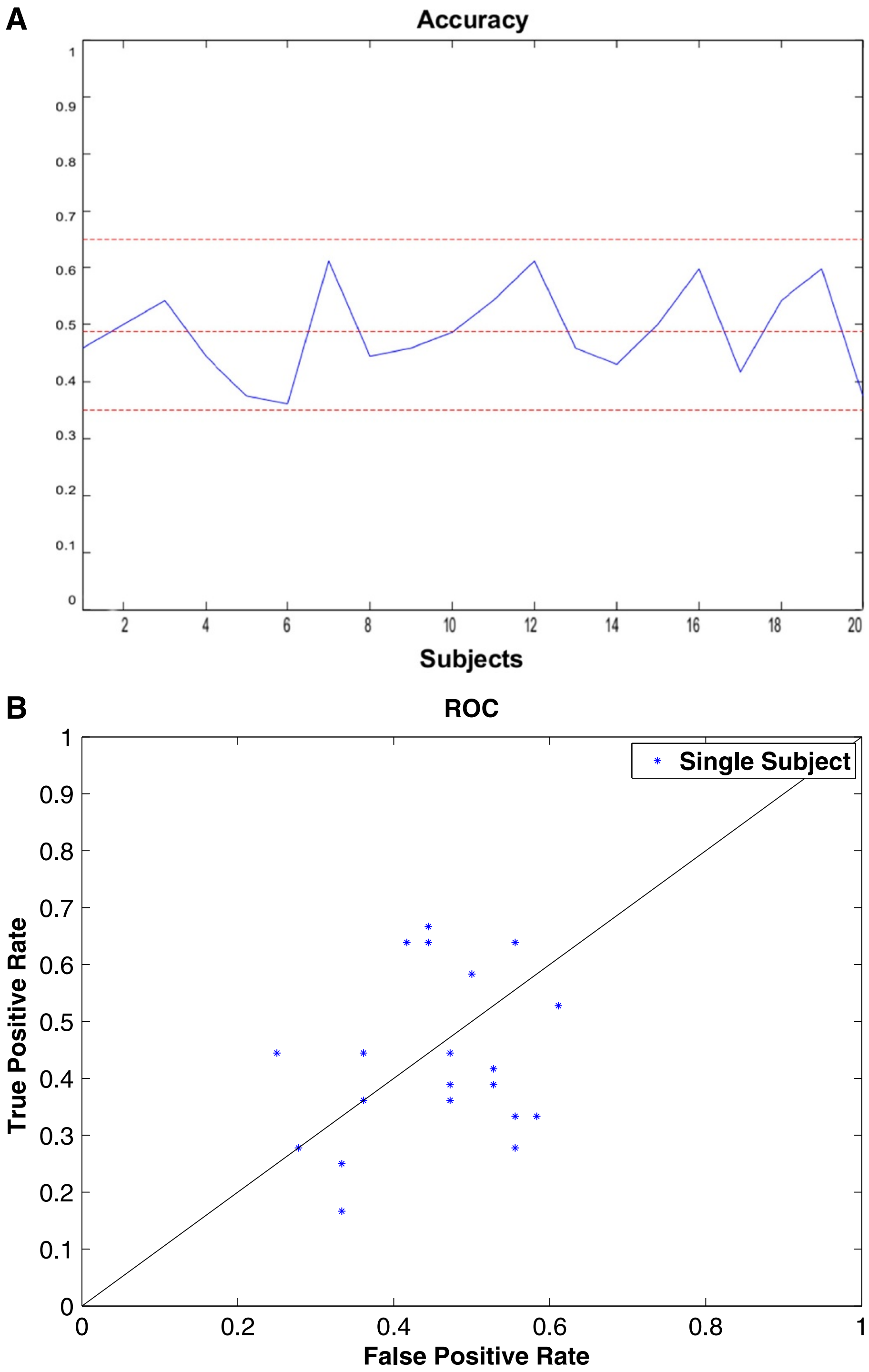


Figure 4 Subjective test results. (A) Accuracy graph. Average value = 0.48. **(B)** ROC graph.

3.2.2 Simulated tests evaluation

The scenario with omnidirectional microphones and the decorrelation module have been taken as the reference situation for the evaluation. ERLE behavior, with (Figure 6B) and without (Figure 6A) decorrelation module, is reported in Figure 6.

First of all, it is evident that techniques using BFs have a positive impact in terms of echo cancelation performance, setting the ERLE 12 dB lower than the reference value: in particular, BFs with hypercardioid microphones seem to have the best conduct, setting the ERLE 17 dB lower than the reference value.

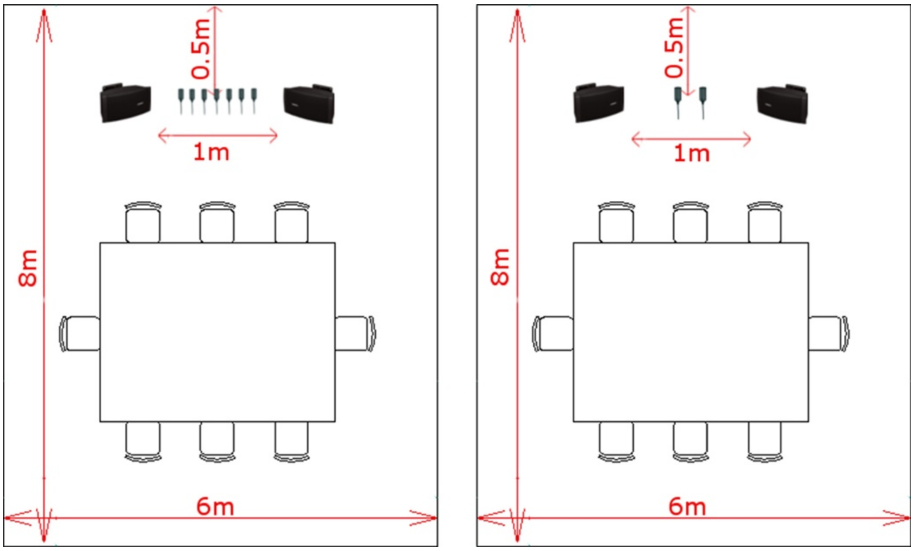


Figure 5 Room plants with microphone array and microphone couple.

Furthermore, it seems that the decorrelation module does not significantly influence the SAEC behavior when stereophonic recording is employed. Indeed, there are no differences between Figure 6A,B after transmission room changes occurring at 5 and 11 s, precisely. This is even more evident if we look at the ERLE trends reported in Figure 7, corresponding to two different stereophonic configurations using hypercardioid microphones and with/without decorrelation module. It is clear that the employment of the decorrelation module seems to not provide a significant boost of echo reduction performance. This is basically due to the inherent decorrelation of audio channels obtained in stereophonic recordings (both in the two-microphone standard configuration and in the BF-based one), caused by the different source contributions, corresponding to diverse acoustic paths in the remote room, included in the left and right channels.

3.3 System robustness tests

Fixed broadband BFs employing small microphone arrays are typically sensitive to array imperfections. Several efforts have been made in the literature to study this problem and to propose efficient solutions to limit its effects. One possible strategy consists in developing BF synthesis methods which take into account the robustness to microphone gain and phase imperfections [33] or to microphone positioning errors [34]. Another interesting approach is the one based on microphone calibration before the usage of beamforming algorithm [35,36].

In this section, the authors analyze the effect of the array imperfections on the characteristics of the designed BF and on the performance of the overall stereophonic echo cancelation system. Also the impact of noise from this perspective will be discussed. Some simulated tests will be presented in the following to evaluate the robustness of the adopted beamforming technique and the overall proposed hands-free communication system.

3.3.1 Beamforming robustness analysis

In this section, the simulated results of the analysis of the influence that the microphone array imperfections have on the BF beampatterns are discussed: such an evaluation is performed both as a function of frequency and incidence angle. The difference between the steering vector attained in these simulated operating conditions and the one related to the ideal case study is also evaluated.

Effect of microphone-array imperfections on beampatterns. For this analysis the ideal beampattern, i.e., the one attainable in absence of microphone array imperfections, is taken as reference and the error between this and the one attained in simulated operating conditions when such imperfections take place is calculated as a function of frequency and incidence angle. Such an evaluation has been accomplished by considering 50 distinct beampatterns, corresponding to the following conditions:

- Each microphone is located around the ideal position according to a Gaussian distribution, with zero-mean and variance equal to 1 mm;
- Each microphone is rotated around the ideal position according to a Gaussian distribution with zero-mean and variance equal to 6°.

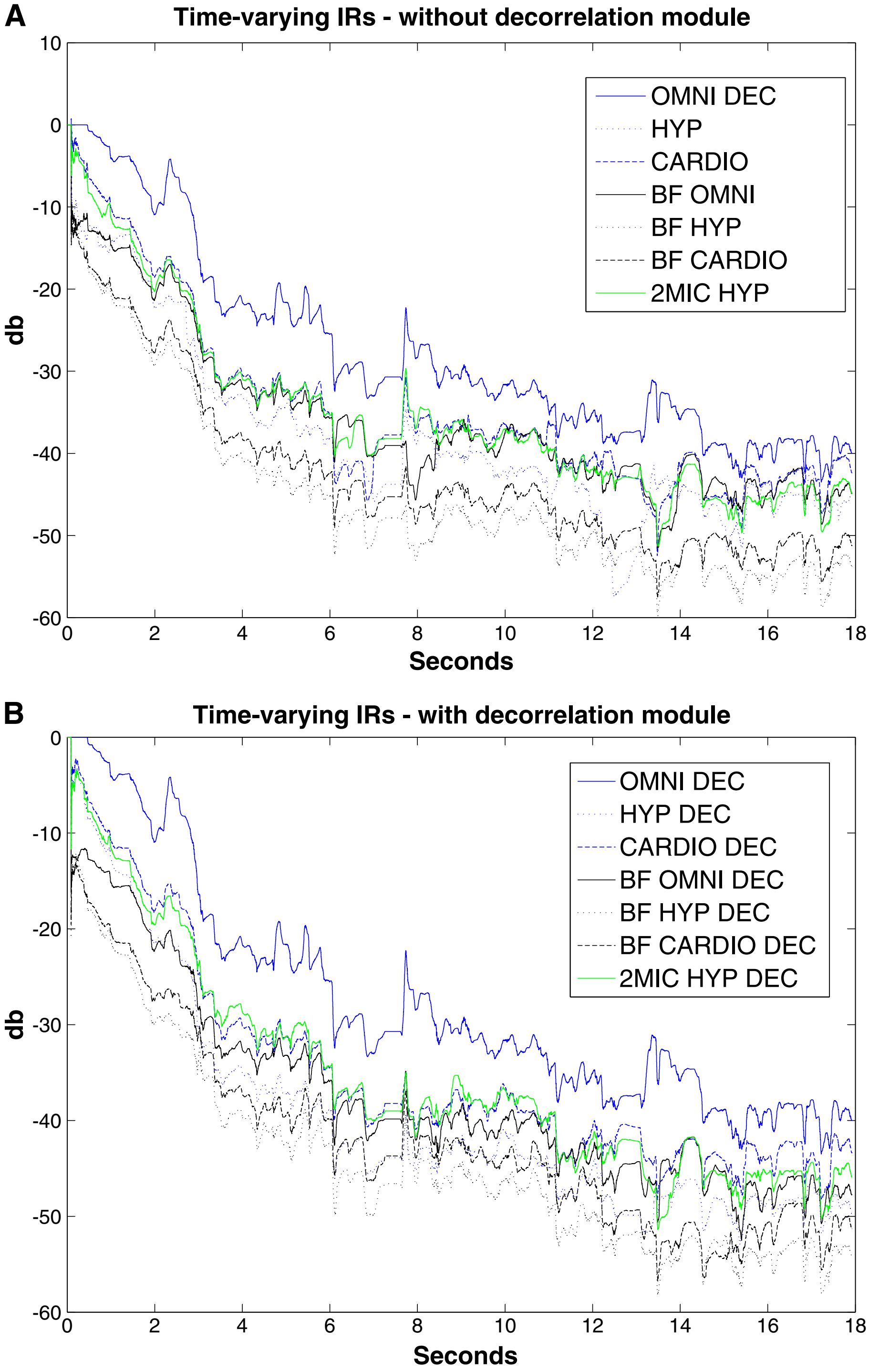


Figure 6 ERLE trends for simulated tests with time-varying IRs. The "OMNI DEC" configuration is used as reference in both plots. **(A)** Without decorrelation module. **(B)** With decorrelation module.

For each point of the grid frequency - incidence angle, the mean square error between the ideal beampattern and the one obtained for each microphone array configuration mentioned above has been calculated; then, all values have been averaged. Figure 8A–C report the errors obtained in three different cases: omnidirectional, cardioid, and hypercardioid microphones. It can be easily seen that the biggest error values are obtained at lowest frequencies, i.e., where the BF is mostly selective.

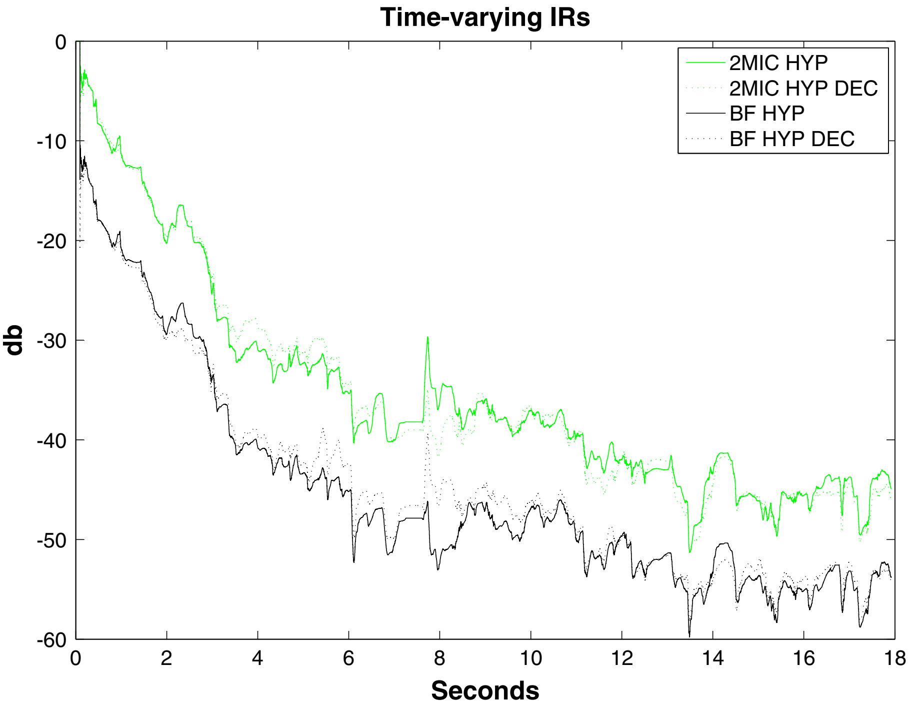


Figure 7 ERLE trends comparison for simulated tests in presence of time-varying IRs, with and without the decorrelation module. Two case studies are here addressed: the two-microphone stereophonic configuration and the BF one, both with hypercardioid microphones.

Effect of microphone array imperfections on steering vector. The following approach has been adopted:

1. A bidimensional grid of 11*8 points has been generated. Each point represents a couple of position—incidence angle values, respectively varying in the ranges $-1; 1$ mm and $-6°; 6°$ (around the ideal microphone array configuration).
2. Each grid point corresponds to certain perturbations in position and orientation for all microphones according to the following rule: x value (both in position and orientation) means that a perturbation of $-x$, x, or 0 occurs for each microphone in the array. 50 different combinations (generated by means of a discrete Gaussian distribution) have been considered for each grid point.
3. The steering vector has been calculated for each grid point.
4. The mean square error between this steering vector and the ideal one has been also evaluated for each grid point and then averaged over all different grid points.

Figure 9A shows the error behavior in the presence of omnidirectional microphones: it is evident that the BF is only sensitive to the position and not to imperfections of the microphone array orientation. Instead, Figure 9B,C report the error surfaces for the cardioid and hypercardioid case studies: the microphone array position imperfections have a reduced impact with respect to the omnidirectional case, but now the orientation perturbations influence the BF characteristics.

3.3.2 Robustness analysis of the whole system

Position, orientation, and gain. Now, it is important to evaluate to what extent the beampattern error due to the array imperfections influences the performance of the overall echo cancelation system. Moving from the same experimental scenario considered in simulations in Section 3.2.1, the ERLE curves have been here calculated under different operating conditions, corresponding to different array imperfections for different microphone polar patterns. Also in this case we have that:

- each microphone is placed around the ideal position according to a Gaussian distribution with zero-mean and variance equal to 1 mm;
- each microphone is rotated around the ideal position according to a Gaussian distribution with zero-mean and variance equal to 6 mm;
- each microphone presents a gain value with respect to the unitary one according to a Gaussian distribution with zero-mean and variance equal to 0.2.

The ERLE curves have been calculated averaging the results of four different configurations (for each operating condition) and then reported in Figures below. Furthermore, in order to better understand the influence that each type of perturbation has on the overall echo

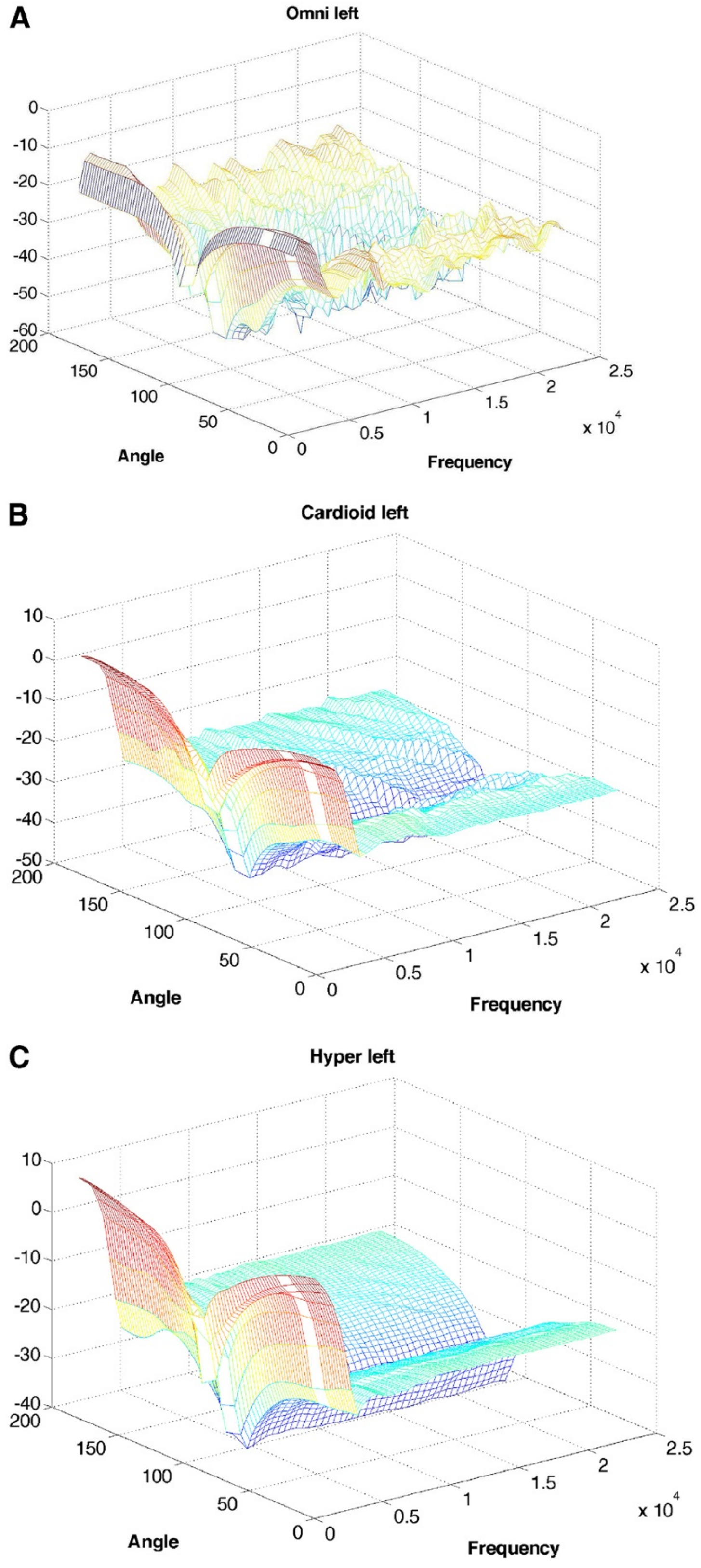


Figure 8 Effect of microphone array imperfections on beampatterns. (A) Omnidirectional microphone. **(B)** Cardioid microphone. **(C)** Hypercardioid microphone.

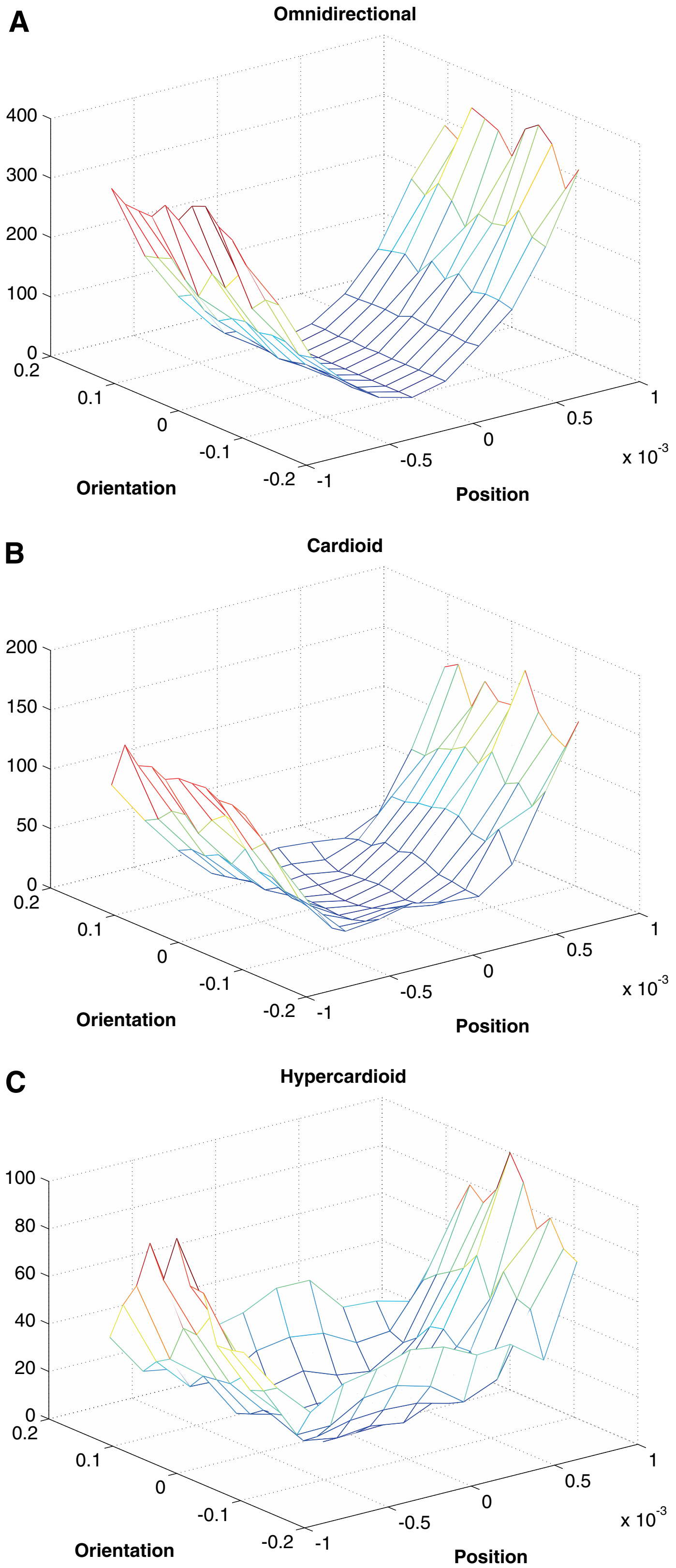


Figure 9 Effect of microphone array imperfections on steering vector. (A) Omnidirectional microphone. **(B)** Cardioid microphone. **(C)** Hypercardioid microphone.

reduction performance, the following tests have been accomplished:

1. real microphone position and ideal orientation and gain;
2. real microphone position and orientation and ideal gain;
3. real microphone position, orientation, and gain.

Microphone position perturbations do not have a negative impact in terms of echo cancelation: indeed the ERLE curve does not differ significantly from the ideal one. On the contrary, imperfections in microphone orientation (Figure 10A) induce a relevant decrement of the overall performance, in particular in the presence of cardioid and hypercardioid microphones. A more evident impact is clearly visible in the third addressed case study, where gain imperfections are also considered: also the BF configuration with omnidirectional microphones is negatively influenced, as depicted in Figure 10B.

Noisy scenario. As final test, the effect of noise presence on echo cancelation performance has been evaluated, without taking the aforementioned microphone array imperfections into account. We can notice that:

- noise has been added both in the transmission and receiving room;
- there is no acoustic coupling in the transmission room since we are dealing with a simulated scenario.

Various tests have been accomplished by varying the signal-to-noise ratio (SNR) in both rooms. In particular the following operating conditions have been tested:

- No noise;
- SNR = 20 dB in the transmission room without noise in the receiving room;
- SNR = 20 dB in the receiving room without noise in the transmission room;
- SNR = 20 dB in the transmission and in the receiving rooms;
- SNR = 20 dB in the receiving room and SNR = 40 dB in the transmission room;
- SNR = 20 dB in the transmission room and SNR = 40 dB in the receiving room;
- SNR = 40 dB in the transmission and in the receiving rooms.

Noise sources are located within the right and left beampatterns at a distance of 3 m from the array, separated by an angle of 50 °. The SNR value has been calculated with respect to the voice signal in the transmission room. The cardioid microphone array has been considered in our simulations.

Figure 11A reports the ERLE curve in absence of noise, confirming what already shown in the previous section. The noise contribution in the transmission room does not have a significant impact on echo reduction performance, as confirmed by Figure 11C. On the contrary, Figure 11B,D–F show how the presence of noise in the receiving room seriously influences the overall system performance; in particular, no advantage in using BF instead of stereo microphones can be registered in this case. It must also be noted that the ERLE value is limited by the amount of noise acquired by the system and therefore lower performance with respect to the ideal case is achieved at convergence.

3.4 Real tests

In this section, the results of the tests performed under real acoustic conditions are discussed in order to prove the effectiveness of the proposed system in real scenarios.

3.4.1 Real tests setup

Tests have been accomplished by using two separate rooms (namely Room A and Room B) at the audio laboratory at the Department of Information Engineering at Università Politecnica delle Marche, as shown in Figure 12A,B. It follows the list of the audio material involved for test setup:

- 2 MOTU Traveler-mk3 audio boards, as depicted in Figure 12A;
- 2 MOTU 8-PRE audio boards, needed to handle 5 microphones inputs for each room;
- 2 PCs Dell Precision T1500 with Intel i5 Core equipped with the software NU-Tech (powered by Leaff Engineering) [24] to manage the audio I/Os at the sampling frequency of 48 kHz;
- 2 Scopia XT1000 modules [37] to allow the TCP/IP based audio streaming between the two rooms, as depicted in Figure 12C;
- 4 GENELEC 6010A loudspeakers;
- 10 AKG C400BL microphones;
- two bases for microphone positioning (Figure 12D).

Two different test sessions have been carried out in order to prove the real-time performance of the proposed system, referred as *test session 1* and *test session 2*, respectively. In *test session 1*, the microphone signal is acquired by using only the two most external microphones of the linear array mounted on the supporting

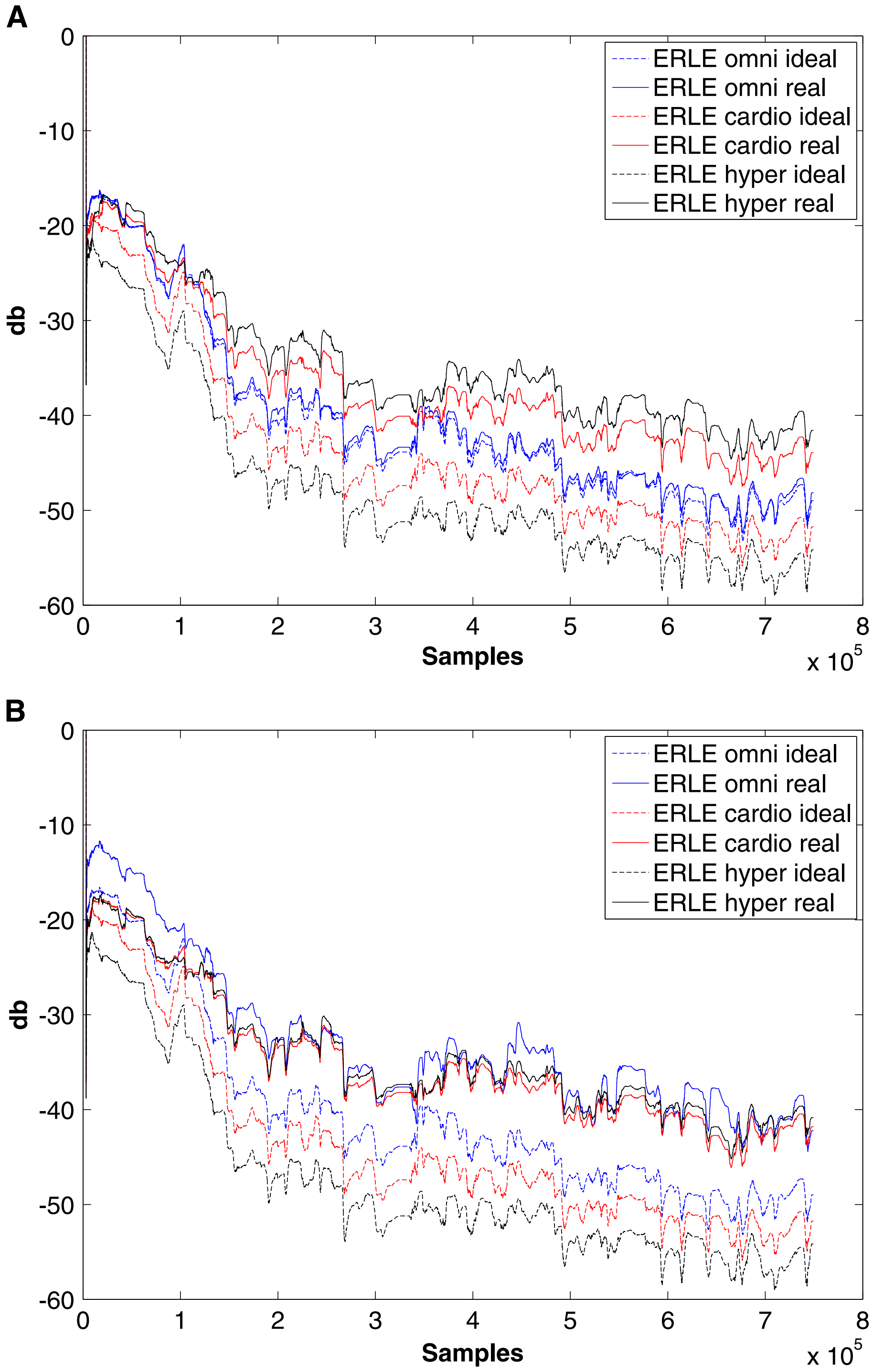


Figure 10 ERLE trends in presence of non-idealities. (A) Non-idealities in position and orientation. **(B)** Non-idealities in position, orientation, and gain.

base; since the microphone distance in the array is equal to 4 cm, the selected microphones in this configuration are 16 cm far away from each other and they are also tilted to form an angle of 70 degrees in order to guarantee the stereophonic effect. In *test session 2*, the near-end signal is acquired through the microphones array; then, the beamforming algorithm is applied, according to the guidelines described in the previous sections. In particular, the beampatterns have been designed to maximize the stereophonic recording capabilities of the array. Loud-

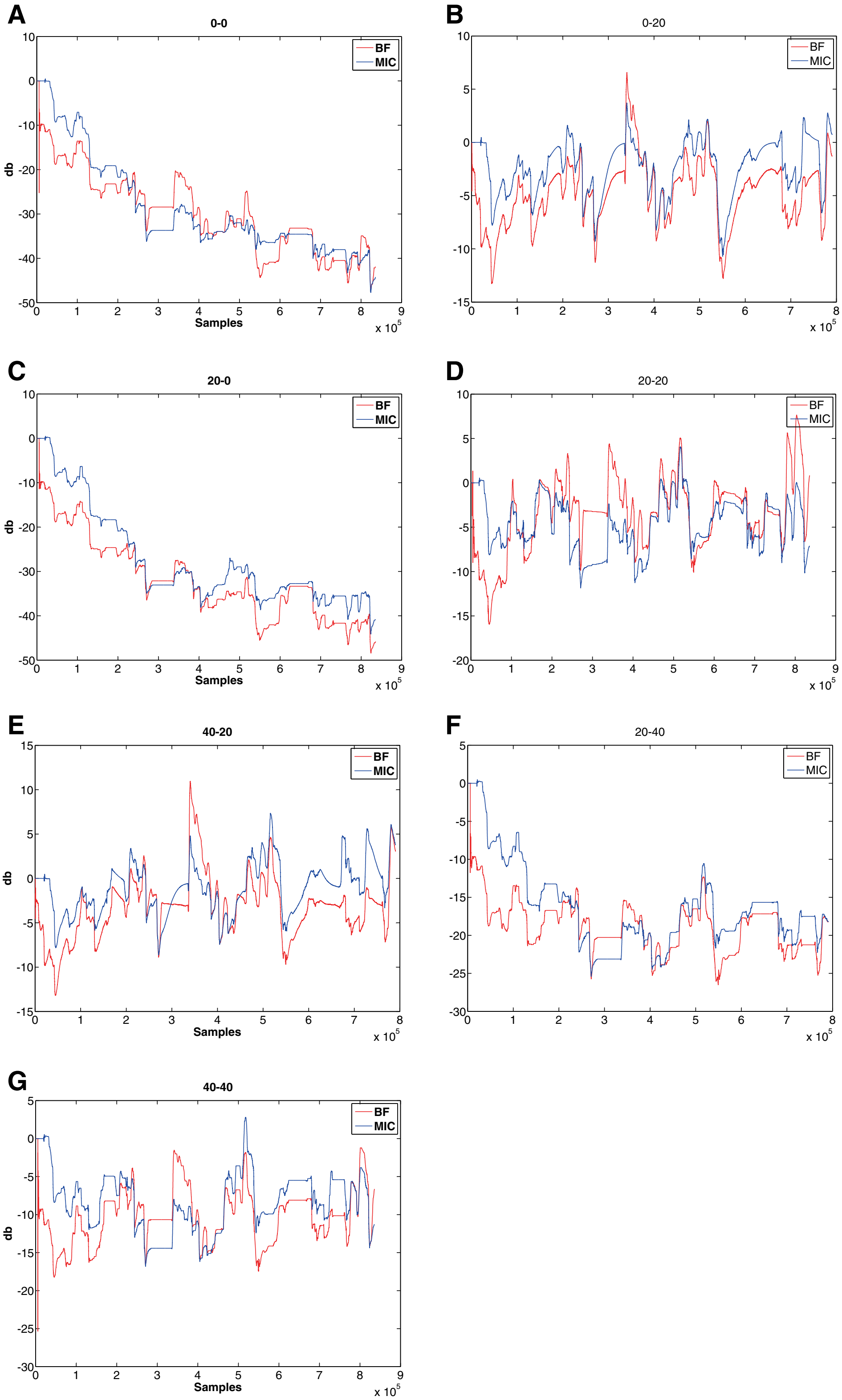


Figure 11 ERLE trends in presence of noise. (A) Absence of noise. **(B)** Absence of noise in transmission room and SNR = 20 dB in receiving room. **(C)** SNR = 20 dB in transmission room and absence of noise in receiving room. **(D)** SNR = 20 dB in transmission room and in receiving room. **(E)** SNR = 40 dB in transmission room and SNR = 20 dB in receiving room. **(F)** SNR = 20 dB in transmission room and SNR = 40 dB in receiving room. **(G)** SNR = 40 dB in transmission room and in receiving room.

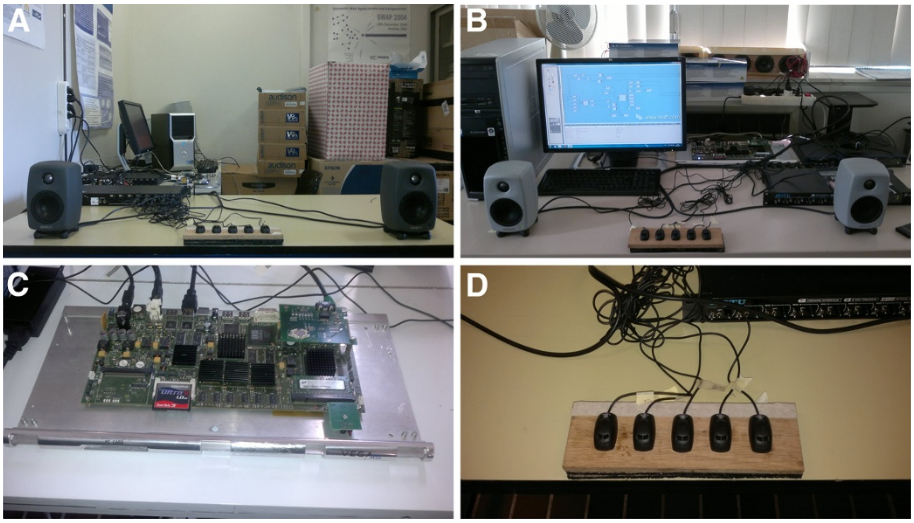


Figure 12 Real tests setup: meaningful details. (A) Room A. **(B)** Room B. **(C)** The SCOPIA XT1000 board (powered by Radvision). **(D)** The microphone array.

speakers have been placed at a distance of 1 m from the center of the array in order to simulate real hands-free communication acoustic conditions. Microphone calibration has been made in order to minimize the negative effect of gain imperfections: microphone gain level has been adjusted so that the energy of all the outputs were the same, having white noise as input. Each simulation test lasts 20 s assuming the following scenario: 0–5 s far-end only, 5–10 s near-end only, 10–15 s double-talk situation, and 15–20 s far-end only.

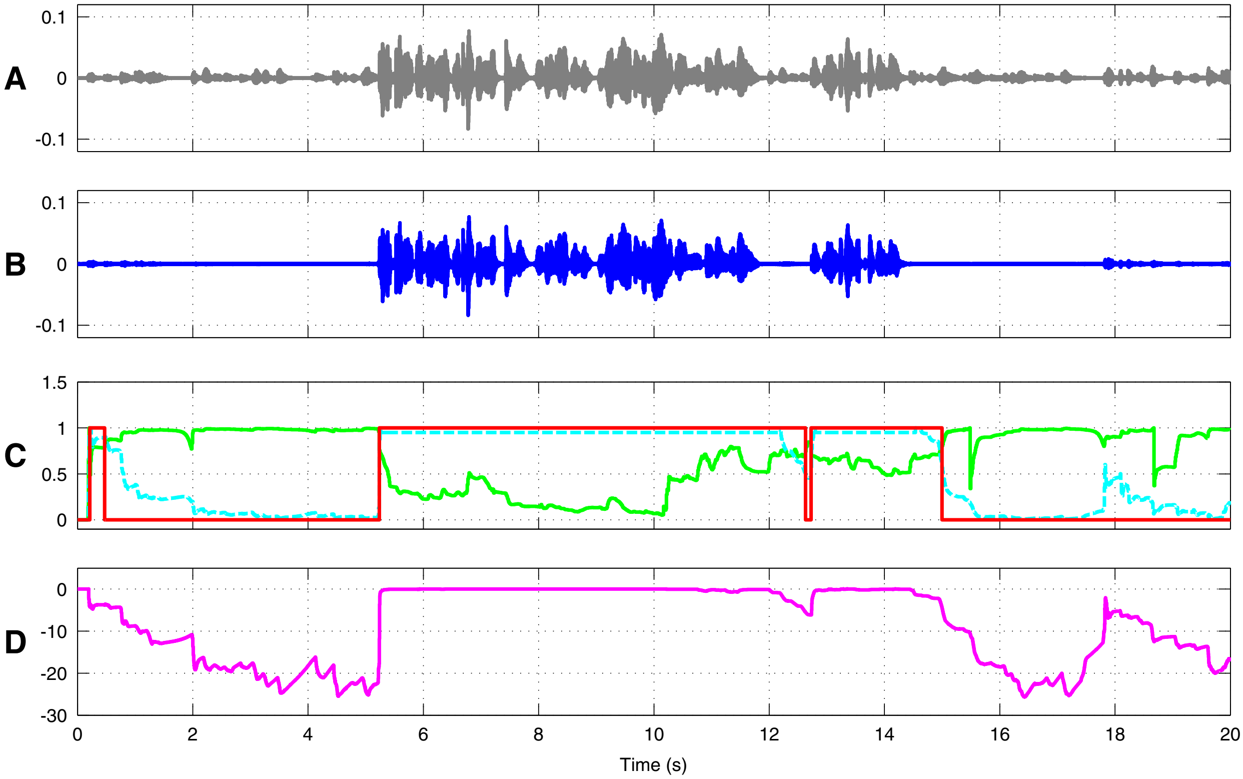


Figure 13 Real tests in the two-microphone configuration: main signals behavior. (A) Microphone signal. **(B)** Signal to be transmitted to the other room. **(C)** DTD signals, i.e., ζ_1 (green line), ζ_2 (cyan line), and DTD (red line). **(D)** ERLE (dB).

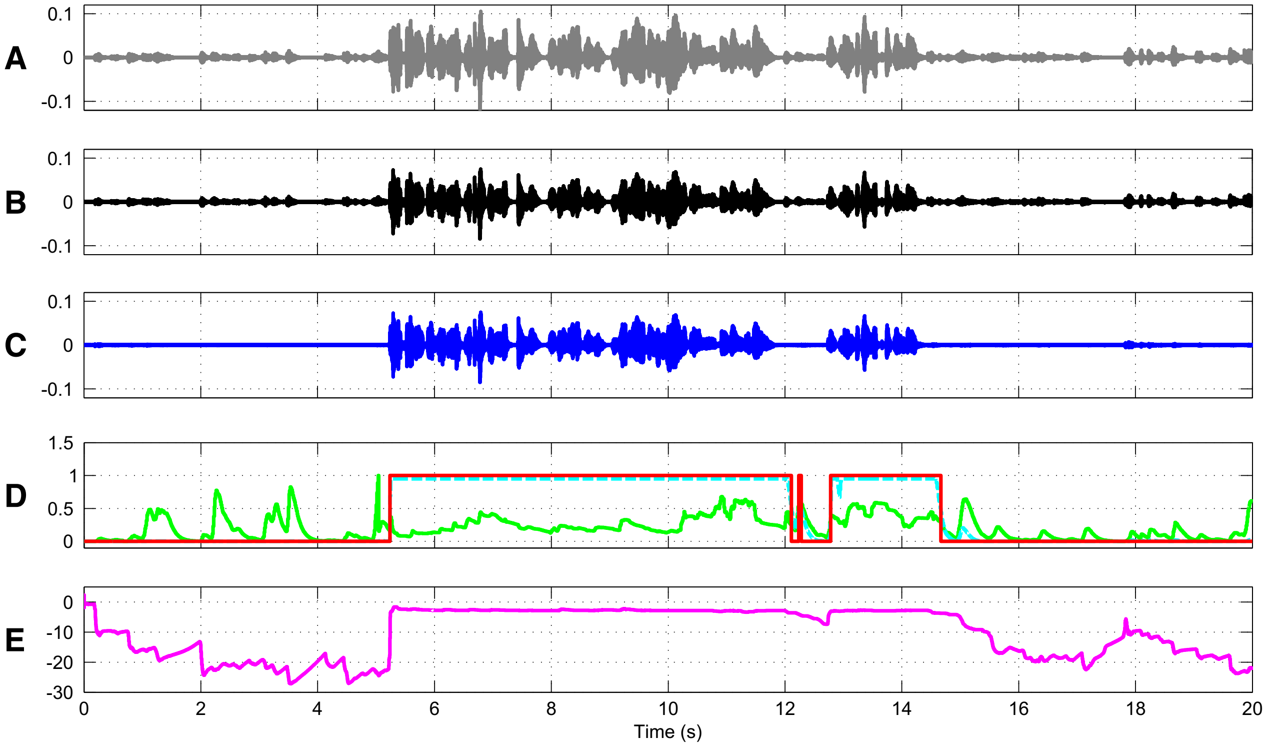


Figure 14 Real Tests in the BF configuration: main signals behavior. (A) Microphone signal. **(B)** Signal processed by the beamformer. **(C)** Signal to be transmitted to the other room. **(D)** DTD signals, i.e., ζ_1 (green line), ζ_2 (cyan line), and DTD (red line). **(E)** ERLE (dB).

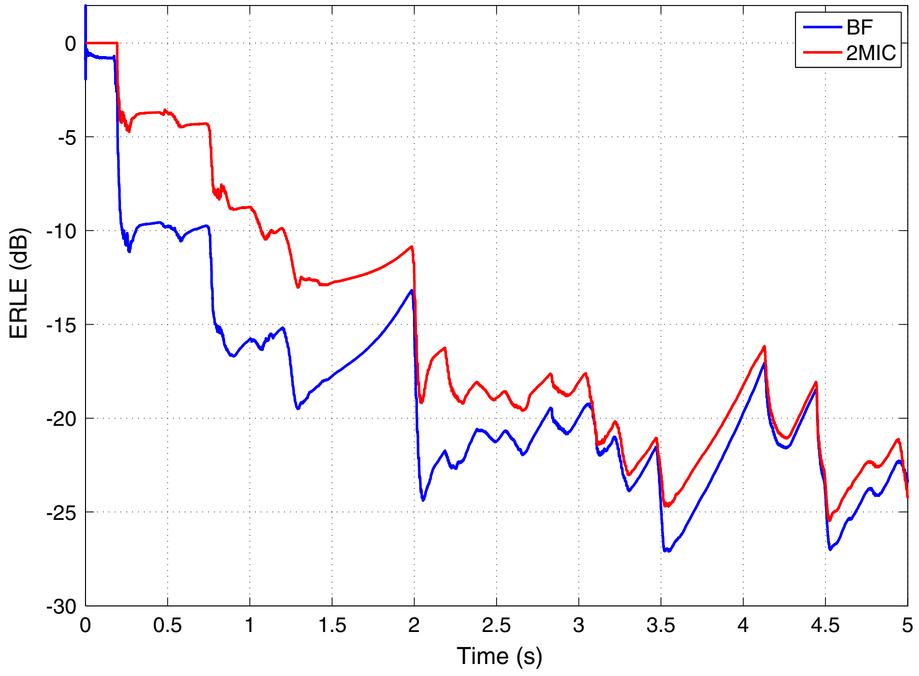


Figure 15 ERLE behavior in the first 5 s for the two-microphone and BF-based approaches.

3.4.2 Real tests evaluation

In these Section, only the results obtained in Room B during *test session 1* and *test session 2* are reported for the sake of brevity. Analogous results have been obtained for Room A. First, it is worth noting that PC simulations (performed with the parametric choices reported in Section 2.5) show that the workload required by the overall system in terms of CPU load is about 41 %, thus guaranteeing real time performance.

Figure 13 shows the behavior of the main signals recorded during *test session 1*, i.e., the microphone signal (a), including the far-end signal and the near-end signal, the residual echo signal (b), the DTD signal (c), and the ERLE behavior (d). It can be easily observed that the DTD algorithm correctly detects the occurrence of double-talk taking advantage of the two control variables ξ_1 and ξ_2. Indeed the DTD signal (c) gets the value "1" and the signal to be transmitted to the other room coincides with the microphone signal in presence of the local speaker. On the contrary, the DTD signal gets the value "0" when no double-talk is present and the signal to be transmitted to the other room is correctly canceled in presence of remote speaker. As previously stated, ξ_2 allows to overcome encountered problems related to very low-level signals acting as a far-end signal detector. Therefore, acoustic echo is correctly canceled as confirmed by the behavior of ERLE (d). Moreover, echo path change scenario has been taken into account considering new echo paths after about $t_1 = 17$ s: Figure 13 proves the robustness of the system showing that no DT situation is revealed.

Regarding *test session 2* and assuming the same scenario as in *test session 1*, analogous considerations can be done taking into consideration Figure 14, where the microphone signal (a), the signal processed by the beamformer (b), the residual echo signal (c), the DTD signal (d), and the ERLE behavior (e) are reported. More specifically, the BF-based solution results more responsive and robust to the echo presence, as confirmed by the plots in Figure 15, depicting the ERLE behavior of the two system configurations in the first 5 s of the experiment. Moreover, also in this case study, the DTD is not subject to the echo-path change problem.

Finally, informal listening tests has also confirmed the quality of the stereophonic rendering in real-time, correctly detecting the movements of the remote speaker just by means of the spatial image effect.

4 Conclusions

In this study, an optimal fixed BF for real-time SAEC systems has been developed: relevant performance in terms of echo cancelation has been reached, providing the possibility of a stereophonic recording of sound and, as a consequence of that, a good decorrelation between channels of stereo signals. Furthermore, no significant increase of computational cost of the whole system occurred, keeping the basic requirement of real-time implementation of algorithms. The robustness of the overall algorithmic framework has been experimentally analyzed with respect to the microphone array imperfections and noise presence, in different configuration case studies. Obtained results allowed concluding that, assuming to operate under low noise conditions, small array perturbations in microphone position/orientation and gain do not significantly affect the behavior of the adopted BF solution within the SAEC system. Taking this aspect into account, the authors developed a real hands-free communication system employing the BF technology discussed in this article (also taking double-talk occurrence into account), which has remarkably confirmed the results already obtained in simulated scenarios.

Future developments are oriented to improve the BF design in order to enhance the sound stereo quality and to use the fixed-BF based SAEC algorithm in those environments where the impact of noise is more influential and a specific algorithm is needed to deal with it. Moreover, the whole system will be implemented in embedded platforms to make the proposed framework more attractive for the hands-free communication market.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The proposed study has been partially supported by the regional project TASCA, publicly funded. The authors wish to thank Radvision Italy for having provided the two SCOPIA XT1000 boards used in experiments. Moreover, a special thanks goes to Emanuele Principi (Università Politecnica delle Marche, Italy) for support in featuring and handling the software platform used for subjective tests.

Received: 23 March 2012 Accepted: 21 September 2012
Published: 22 October 2012

References

1. SL Gay, J Benesty, *Acoustic Signal Processing for Telecomunication* (Kluwer Academic Publishers, Dordrecht, 2000)
2. J Benesty, DR Morgan, MM Sondhi, A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation. IEEE Trans. Speech Audio Process. **6**(2), 156–165 (1998)
3. M Ali, in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing*, vol. 6 Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation. (Seattle, WA, USA, 1998), pp. 3689–3692
4. A Gilloire, V Turbin, in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing*, vol. 6 Using auditory properties to improve the behaviour of stereophonic acoustic echo cancellers. (Seattle, WA, USA, 1998), pp. 3681–3684
5. J Herre, H Buchner, W Kellermann, in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing*, vol. 1 Acoustic echo cancellation for surround sound using perceptually motivated convergence enhancement. (Honolulu, HI, USA, 2007), pp. 17–20
6. M Brandstein, D Ward (eds), *Microphone Arrays: Signal Processing Techniques and Applications* (Springer, New York, 2001)
7. W Herbordt, *Sound Capture for Human/Machine Interfaces. Practical Aspects of Microphone Array Signal Processing* (Springer, New York, 2005)

8. W Kellerman, Intergrating acoustic echo cancellation with adaptive beamforming microphones array. J. Acoust. Soc. Am. **105**(2), 1098–1098 (1999)
9. H Buchner, W Herbordt, W Kellermann, in *Proc. Int. Workshop on Hands-Free Speech Communication,* vol. 1 An efficient combination of multichannel acoustic echo cancellation with a beamforming microphone array. (Kyoto, Japan, 2001), pp. 55–58
10. M Kallinger, J Bitzer, KD Kammeyer, in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing,* vol. 2 Study on combining multi-channel echo cancellers with beamformers. (Istanbul, Turkey, 2000), pp. II797–II800
11. W Chen, Z Zhang, in *Proc. IEEE Int. Workshop on Multimedia Signal Processing,* vol. 2 Enhancing stereophonic teleconferencing with microphone arrays through sound field warping. (445 Hoes Lane Piscataway, N.J. 08854, U.S.A, 2010), pp. 29–33
12. M Pirro, S Squartini, F Piazza, in *Proc. 131st Audio Engineering Society Convention* A fixed beamforming based approach for stereophonic audio-conference systems, (2011)
13. J Benesty, Y Huang, *Adaptive Signal Processing: Applications to Real-World Problems* (Springer, New York, 2003)
14. E Ferrara, Fast implementations of LMS adaptive filters. IEEE Trans. Acoust. Speech Signal Process. **28**(4), 474–475 (1980)
15. S Haykin, *Adaptive Filter Theory* (Englewood Cliffs, NJ, Prentice-Hall, Inc., 1996)
16. JJ Shynk, Frequency-domain and multirate adaptive filtering. IEEE Signal Process. Mag. **9**, 14–37 (1992)
17. J Benesty, P Duhamel, A fast exact least mean square adaptive algorithm. IEEE Signal Process. Lett. **40**(12), 2904–2920 (1992)
18. J Benesty, T Gansler, A multichannel acoustic echo canceller double-talk detector based on a normalized cross-correlation matrix. European Trans. Telecommun. **13**(2), 95–101 (2002)
19. MA Iqbal, SL Grant, JW Stokes, in *Proc. 43rd Asilomar Conference on Signals, Systems and Computers,* vol. 1 A frequency domain doubletalk detector based on cross-correlation and extension to multi-channel case. (445 Hoes Lane Piscataway, N.J. 08854, U.S.A, 2009), pp. 638–641
20. H Chen, Z Ser, in *Proc. IEEE Asia Pacific Conference on Circuits and Systems,* vol. 1 Design of broadband beamformers for microphone arrays using variably-weighted least squares. (Macao, China, 2008), pp. 996–999
21. HS Malvar, Fast algorithm for the modulated complex lapped transform. IEEE Signal Process. Lett. **10**, 8–10 (2003)
22. L Romoli, S Squartini, F Piazza, in *Proc. European Signal Processing Conference,* vol. 1 A variable step-size frequency-domain adaptive filtering algorithm for stereophonic acoustic echo cancellation. (Aalborg, Denmark, 2010), pp. 26–30
23. Intel Corporation: Integrated performance primitives libraries. http://software.intel.com/en-us/intel-ipp/
24. F Bettarelli, E Ciavattini, A Lattanzi, D Zallocco, S Squartini, F Piazza, in *Audio Engineering Society Convention 118th,* vol. 1 NU-Tech: implementing DSP algorithms in a plug-in based software platform for real time audio applications. (Barcelona, Spain, 2005), pp. 1–12. [http://www.aes.org/e-lib/browse.cfm?elib=13105]
25. Audio Engineering Society: AES recommended practice for professional audio - Subjective evaluation of loudspeakers (1996)
26. W Hoeg, L Christensen, R Walker, Subjective assessment of audio quality - the means and the methods within the EBU. *European Broadcasting Union* (1997)
27. International Telecommunications Union: ITU-R BS.1116-1 - Method for the Subjective assessment of small impairments in audio systems including multichannel sound systems (1997)
28. International Telecommunications Union: ITU-R BS.1284-1 - General methods for the subjective assessment of sound quality (1997)
29. L Gabrielli, S Squartini, V Välimäki, in *Audio Engineering Society Convention 131* A subjective validation method for musical instrument emulation. (2011) http://www.aes.org/e-lib/browse.cfm?elib=16087
30. T Fawcett, ROC graphs: Notes and practical considerations for researchers. Mach. Learn. **31**(HPL-2003-4), 1–38 (2004)
31. JA Swets, *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers* (Lawrence Erlbaum Associates, New Jersey, 1995)
32. C Wun, A Horner, Perceptual wavetable matching for synthesis of musical instrument tones. J. Audio Eng. Soc. **49**(4), 250–262 (2001)
33. S Doclo, M Moonen, Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics. IEEE Trans. Signal Process. **51**(10), 2511–2526 (2003)
34. MMS Doclo, in *Proc. Workshop on Acoustic Echo and Noise Control,* vol. 1 Design of broadband beamformers robust against microphone position errors. (Kyoto, Japan, 2003), pp. 267–270
35. P Oak, W Kellermann, in *Proc. Workshop on Acoustic Echo and Noise Control,* vol. 1 A calibration algorithm for robust generalized sidelobe cancelling beamformers. (Eindhoven, The Netherlands, 2005), pp. 97–100
36. N Tashev, in *Proc. Int. Conference Multimedia and Expo,* vol. 2 Gain self-calibration procedure for microphone arrays. (Taipei, Taiwan, 2004), pp. 983–986
37. Radvision, An Avaya Company. Conference Room Systems. http://www.radvision.com/Products/Video-Conference-Systems/Conference-Room-Systems/

doi:10.1186/1687-4722-2012-26
Cite this article as: Pirro *et al.*: **Stereophonic hands-free communication system based on microphone array fixed beamforming: real-time implementation and evaluation.** *EURASIP Journal on Audio, Speech, and Music Processing* 2012 **2012**:26.