An Interactive Optimization Procedure for Stereophonic Acoustic Echo Cancellation Systems

Laura Romoli, Stefania Cecchi, Francesco Piazza,
Danilo Comminiello, Michele Scarpiniti, and Aurelio Uncini

***Abstract*—Acoustic echo cancellers are used in teleconferencing systems in order to reduce undesired echoes due to coupling between microphones and loudspeakers. Stereophonic systems provide more realistic experience than single-channel systems, since listeners have spatial information that helps to identify the speaker position. Assuming this scenario, a suitable choice for the system parameters becomes essential to improve the audio reproduction quality. Error-driven optimization strategies are usually used to obtain an optimal system configuration but there is no relationship with the quality desired by the user. In this paper, an interactive evolutionary algorithm is adopted for a stereophonic acoustic echo cancellation system in order to meet subjective specifications in the optimization stage. In this way, the optimal system configuration is derived according to a user-driven approach in order to satisfy the quality requirements demanded by users availing such stereophonic systems. Experimental results prove the effectiveness of the proposed interactive architecture according to both objective and subjective measures.**

I. Introduction

Multiple microphones and loudspeakers are used in sound reproduction systems for capturing and reproducing the desired signals while offering an immersive experience to the listener [1], [2]. Adaptive signal processing systems are often introduced to overcome quality degradation due to interfering sources and reverberation, thus resulting in intelligent acoustic interfaces (IAIs) [3]–[5]. Acoustic echo caused by the multiple coupling between microphones and loudspeakers is one of the main problems of such immersive scenarios. Therefore, acoustic echo cancellers are used in teleconferencing systems in order to reduce undesired echoes [6]. In the two-channel scenario, stereophonic acoustic echo cancellation systems (SAECs) are exploited to provide spatial information to listeners helping the identification of the speaker position.

Differently from the single-channel scenario, the linear relationship existing among the loudspeaker signals worsens the performance and a method to reduce interchannel coherence must be introduced [7]. Several efforts have been done in the field of signal decorrelation as summarized in [2], [7], [8], mainly focusing on the stereophonic scenario. These approaches can be divided into two main categories: methods based on the direct alteration of the stereo signal (e.g., half-wave rectifier [6], time-varying all-pass filters [9], [10], phase modulation [11], frequency shifting [12], "missing-fundamental" theory [7]) and other techniques based on the introduction of an external signal to both channels, exploiting psychoacoustic phenomena (e.g., introduction of noise masked according to the human auditory system properties [13], [14]).

Then, the effectiveness of an echo cancellation system strictly relies on the design of a multiple-input multiple-output (MIMO) filtering system, whose main task is to estimate several acoustic impulse responses (AIRs), depending on the number of microphones and loudspeakers. A large number of coefficients has to be adapted, therefore an appropriate choice of the adaptive algorithm becomes essential [1], [15]. Several adaptive algorithms have been studied in the literature. The time-domain first-order adaptive algorithms, such as the least mean squares (LMS), are very attractive due to their simplicity and low computational cost. Hessian-based algorithms, such as the recursive least squares (RLS), improve convergence abilities but entails a larger computational cost and, moreover, they may perform worse than first-order algorithms in adverse environments [16]. The affine projection algorithm (APA) could be used to adapt MIMO filters, since it can be seen as a generalization of the normalized LMS (NLMS) algorithm involving cross-correlations of the input signals [17].

One of the main problems of such systems lies in the overall number of parameters, which is rather high and may produce a hard and flawed tuning of the system. This problem may be overcome by applying any structured stochastic optimization procedure that produces an optimal tuning of the parameters. This technique was extensively studied in the literature and proposed for adaptive systems [18]–[20]. However, the most crucial point for an optimization procedure in IAIs is represented by the quality constraints that must be satisfied. The relevance of audio signal quality using IAIs is due to the necessity to locally reproduce the immersive sound perception that user would have [3]. Therefore, audio quality is strictly related to the perception of users since the higher the audio quality, the more realistic the sound field perception of a user. In that regard, structured optimization techniques very often do not meet such quality requirements since they are "error-driven" strategies and do not take into account any subjective quality index, thus the qualitative constraints fixed by user cannot be assured.

In order to address this problem, an optimization procedure for a stereophonic system is proposed in this paper. The solution includes both the solution for overcoming the well-known non-uniqueness problem and the adaptive algorithm. Decorrelation is performed according to psychoa-

L. Romoli, S. Cecchi and F. Piazza are with the Department of Information Engineering, Università Politecnica delle Marche, 60131 Ancona, Italy (email: l.romoli@univpm.it)

D. Comminiello, M. Scarpiniti and A. Uncini are with the Department of Information Engineering, Electronics and Telecommunications (DIET), "Sapienza" University of Rome, 00184 Rome, Italy (email: danilo.comminiello@uniroma1.it.

The work of Danilo Comminiello was partly funded by bdSound.

coustic criteria as discussed in [21], and echo cancellation is addressed using a stereophonic affine projection algorithm (SAPA). The optimization procedure involves an *interactive evolutionary computation* (IEC) method [22] based on user feedback [23], [24]. The proposed strategy aims at providing echo cancellation performance that satisfies the final users from a perceptual point of view. Experimental results prove the effectiveness of the proposed architecture in optimal parameters setting both from objective and subjective points of view.

The paper is organized as follows. The interactive quality enhancement procedure is presented in Section II and the proposed stereophonic interactive system is fully described in Section III. Then, its effectiveness is proved in Section IV through objective (Section IV-B) and subjective (Section IV-C) evaluations. Finally, conclusion and future work are drawn in Section V.

II. Interactive Quality Enhancement Procedure

One of the main problems of IAIs involving a decorrelation system and an echo canceller is represented by the overall setting of parameters, which is difficult to initialize and may need to be tuned also during the learning. A classic solution is represented by structured (or *analytical*) stochastic optimization procedures, which are "error-driven" approaches providing the system with the optimal parametric configuration to be used [22]. However, apart from the chosen optimization procedure, they strictly depend on the accuracy of the system that produces the error signal [23], [24]. Moreover, objective procedures may not be optimal from a perceptual point of view, thus not reflecting the quality desired by the user. Therefore, a wrong parameter setting may debase even a well-performing IAI, whose potential quality is higher than the perceived one. This is the reason why, instead of an analytic procedure, an *interactive* optimization procedure for IAIs has been preferred in this paper, since it is based on a "user-driven" approach.

Evolutionary computation includes a family of stochastic optimization procedures that iteratively improve a set of candidate solutions by selective application of recombination and mutation operators, inspired to the process of natural selection [25]. IEC can be defined as any evolutionary algorithm in which an interaction between the algorithm and the end user occurs [22]. In other words, the IEC function to be optimized, denoted in the literature as *fitness*, is replaced by user evaluations. In this way, a user can be seen now as a black box guiding the search process. As a result, an interactive evolutionary algorithm (IEA) is able to efficiently search the psychological space of user preferences. A human listener presents two main differences from a "standard" fitness function: the *user fatigue*, i.e., the performance degradation after a series of evaluations, and the *hypothesis discrimination*, i.e., the impossibility for a non-expert user of distinguishing very similar sounds. Both problems can be partially solved by allowing the fitness to take only a small set of values, or even by inserting an adaptive layer between the user and the IEC that learns the user preferences and reduces the number of fitness evaluations [26].

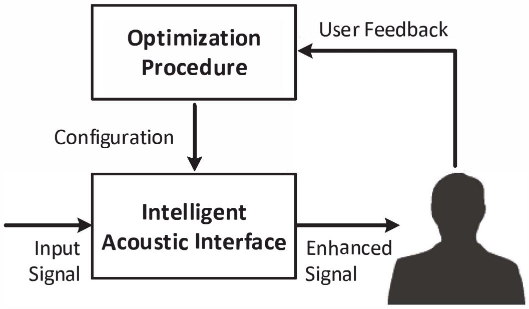

Fig. 1. Interactive optimization procedure for intelligent acoustic interfaces.

The interactive quality enhancement procedure is summarized in Fig. 1. As can be seen, IEC is now used as an active part of the optimization procedure, therefore the subjective evaluation is used to iteratively search for an optimal configuration for the IAI. The main advantage of this architecture is that the system is now able to "close the loop" on the user, optimizing directly its own preferences rather than an analytical approximation. A second important advantage is flexibility, since the process can be run at an initial phase, or during the processing at regular intervals, or even directly enforced by the user when performance degrades.

III. Proposed Stereophonic Interactive System

The proposed interactive stereophonic IAI is depicted in Fig. 2. At the n-th time instant, the far-end speech signals, denoted as $u_p\left[n\right]$, with $p = 1, 2$, arrive at the near-end side where the listener is located. Here the signals are processed by the stereophonic decorrelator. The decorrelated signals $x_p\left[n\right]$, with $p = 1, 2$, are reproduced by $P = 2$ loudspeakers and then acquired by $Q = 2$ microphones. The acoustic coupling between microphones and loudspeakers is characterized by four AIRs, which also contain information about environment reverberations. The desired signals $d_q\left[n\right]$ ($q = 1, 2$) acquired by the microphones represent the echo signals, which may be possibly superimposed on any near-end contribution, containing the near-end speech signal $s\left[n\right]$ with the addition of background noise $v\left[n\right]$. At the same time, the decorrelated signals $x_p\left[n\right]$ are processed by the SAEC in order to estimate the AIRs between microphones and loudspeakers. The near-end user provides a quality feedback to the IAI. The feedback is received by the the IEC block, which supplies the adaptive filters with the parameter setting that yields the best perceived quality to the user. In the following, the main parts of the whole system are described.

A. Stereophonic decorrelation

The first processing of the input signals is performed by a stereophonic decorrelator. The two-channel decorrelator has been recently introduced for stereophonic reproduction systems [21]. It is based on the combination of the missing-fundamental theory with frequency shifting. More specifically, the solution includes an adaptive notch filter $H\left(z\right)$ acting in the low-frequency range and estimating an adaptive parameter related to the fundamental frequency that is also used for controlling the desired frequency shift value applied

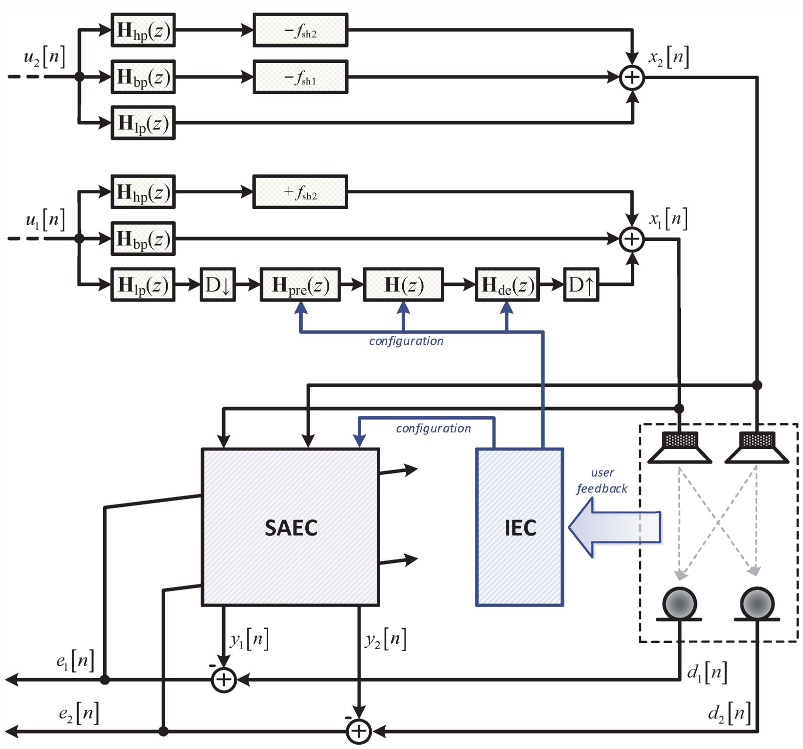
$u_2[n]$
$\mathbf{H}_{\text{hp}}(z)$
$-f_{\text{sh2}}$
$\mathbf{H}_{\text{bp}}(z)$
$-f_{\text{sh1}}$
$\mathbf{H}_{\text{lp}}(z)$
$x_2[n]$
$u_1[n]$
$\mathbf{H}_{\text{hp}}(z)$
$+f_{\text{sh2}}$
$\mathbf{H}_{\text{bp}}(z)$
$x_1[n]$
$\mathbf{H}_{\text{lp}}(z)$
D↓
$\mathbf{H}_{\text{pre}}(z)$
$\mathbf{H}(z)$
$\mathbf{H}_{\text{de}}(z)$
D↑
configuration
configuration
user feedback
SAEC
IEC
$y_1[n]$
$y_2[n]$
$e_1[n]$
$e_2[n]$
$d_1[n]$
$d_2[n]$

Fig. 2. The proposed interactive stereophonic IAI.

in the medium and high frequency range. As it is possible to see in Fig. 2, each far-end signal $u_p[n]$, with $p = 1, 2$, is divided into three subbands by using high-pass $H_{\text{hp}}(z)$, band-pass $H_{\text{bp}}(z)$, and low-pass $H_{\text{lp}}(z)$ filters.

In the low-frequency band, the "missing-fundamental" phenomenon is exploited for reducing the interchannel coherence taking advantage of the human brain capability of processing the information present in the highest harmonics to calculate the "missing-fundamental" [27]. More specifically, the fundamental frequency $f_{\text{est}}(n)$ is estimated and removed from one channel of the stereo signal by using an adaptive second-order lattice form notch filter whose transfer function is defined by the following equation:

$$H(z) = \frac{1 + 2k_0(n)z^{-1} + z^{-2}}{1 + k_0(n)(1+\alpha)z^{-1} + \alpha z^{-2}}. \tag{1}$$

This function is described by the pole-zero contraction factor α, which controls the filter's bandwidth, and the adaptive coefficient $k_0(n)$ [28], related to the tracked frequency $f_{\text{est}}(n)$ as follows:

$$f_{\text{est}}(n) = \frac{f_{\text{s}}}{D} \cdot \frac{1}{2\pi} \cos^{-1}[-k_0(n)], \tag{2}$$

where f_{s} is the sampling frequency and D is the downsampling factor. A sigmoid function is used as constraint function in order to prevent the filter from diverging [29]. Moreover, pre-emphasis $H_{\text{pre}}(z)$ and de-emphasis $H_{\text{de}}(z)$ filters [7] are introduced, as depicted in Fig. 2, for improving the estimation and tracking of the fundamental frequency, especially when the adaptive notch filter range includes the fundamental frequency and some harmonics. A more detailed scheme can be found in [21]. It is worth noting that fundamental estimation accuracy is an important aspect because, if the fundamental is not correctly detected, it is possible to have signal distortion due to the removal of a wrong part of the spectrum [27].

Regarding the medium and high frequency range, frequency shifting is applied according to the human sensitivity at different frequencies. The frequency shifter is implemented using cosine and sine as modulation functions at frequency f_{sh} together with a Hilbert filter [12] of length $2N_{\text{h}}+1$ whose impulse response can be computed as follows

$$h_{\text{h}}(k) = \begin{cases} 0, & k \text{ even} \\ \frac{2}{k\pi}, & \text{otherwise} \end{cases} \tag{3}$$

for $k = -N_{\text{h}}, \ldots, N_{\text{h}}$. The desired frequency shift is varied at any time instant according to the estimated adaptive coefficient $k_0(n)$ bounded in the range $(-1; 1)$, as shown in Fig. 2. This is performed in order to have a slow and continuous time variation of the parameter [30], enhancing the decorrelation performance as in the case of time-varying all pass filtering [9]. Shift values are chosen considering the human sensitivity to phase differences [11], [31] and limiting the change in the time of arrival of each frequency within the just noticeable interaural delay [9], in order to obtain the best compromise between decorrelation and audio quality preservation. Therefore, as it is possible to see in Fig. 2, $f_{\text{sh1}} = k_0(n)/10$ and $f_{\text{sh2}} = 10 \cdot k_0(n)$ are applied in the medium-frequency range and in the high-frequency range, respectively. In this way, the variation results bounded in the range $(-0.1; 0.1)$ Hz and in the range $(-10; 10)$ Hz, respectively. Moreover, still according to human perception, the frequency shifter is applied only on one channel in the

Algorithm 1: Pseudocode of the interactive genetic algorithm.

```
Data: Maximum number of generations $N$,
      Population size $S$, Selection rate $r$, Mutation
      rate $m$, Test sound T
Result: Best individual overall
$\mathbf{C}_0 \leftarrow$ InitializePopulation()
for $i \leftarrow 1$ to $N$ do
    ComputeInteractiveFitness($\mathbf{C}_i$, T)
    $\mathbf{C}_{i+1} \leftarrow$ Select($\mathbf{C}_i$, $r$)
    $\mathbf{C}_{i+1} \leftarrow \mathbf{C}_{i+1} \cup$ Reproduce($\mathbf{C}_{i+1}$)
    Mutate($\mathbf{C}_{i+1}$, $m$)
end
```

Algorithm 2: Pseudocode of the interactive fitness computation.

```
Data: Population $\mathbf{C}_i$, Test sound $T$
Result: Interactive fitness vector $\mathbf{g}_i$
for $s \leftarrow 0$ to $S-1$ do
    Filter $\leftarrow$ TrainAPAFilter($\mathbf{c}_{i,s}$, $T$)
    $g_i[s] \leftarrow$ GetUserEvaluation(Filter, $T$)
end
```

medium band in order to avoid audible artifacts [32]. Differently, in the high band, a frequency shift with opposite sign is applied on the two channels thanks to the lower sensitivity at these frequencies [32]. Besides the decorrelation provided by the frequency shifter, the time variation of the shift value improves the interchannel coherence reduction [9], [11] still preserving audio quality. This is due to the behavior of the coefficient $k_0(n)$ that is characterized by small and limited changes, thanks also to the sigmoid function used for the fundamental frequency tracking.

B. Stereophonic acoustic echo cancellation

The decorrelated input signals $x_p[n]$, with $p = 1, 2$, are collected in an input data matrix $\mathbf{X}_n \in \mathbb{R}^{K \times MP}$, where M is the length of the adaptive filters and K denotes the number of previous entries to keep in memory, i.e., the projection order. The input matrix is processed by the SAEC that is represented by a coefficient matrix $\mathbf{W}_n \in \mathbb{R}^{MP \times Q}$, which contains all the individual filters $\mathbf{w}_{n,pq} \in \mathbb{R}^{M \times 1}$. The filter output $\mathbf{Y}_n \in \mathbb{R}^{K \times Q} = \mathbf{X}_n \mathbf{W}_{n-1}$ can be seen as a concatenation of $Q = 2$ individual output vectors, i.e., $\mathbf{Y}_n = \begin{bmatrix} \mathbf{y}_{n,1} & \mathbf{y}_{n,2} \end{bmatrix}^T$. The error matrix is achieved as:

$$\mathbf{E}_n = \mathbf{D}_n - \mathbf{Y}_n. \tag{4}$$

where $\mathbf{D}_n \in \mathbb{R}^{K \times Q}$ is the matrix containing the Q desired signals. Also the error matrices can be seen as a concatenation of Q individual error vectors, i.e., $\mathbf{E}_n = \begin{bmatrix} \mathbf{e}_{n,1} & \mathbf{e}_{n,2} \end{bmatrix}^T$. The MIMO filters are individually updated according to the stereophonic affine projection algorithm [17]:

$$\mathbf{W}_n = \mathbf{W}_{n-1} + \mu \mathbf{X}_n^T \left(\delta + \mathbf{X}_n \mathbf{X}_n^T\right)^{-1} \mathbf{E}_n \tag{5}$$

where μ and δ are the step-size parameter and the regularization factor, respectively.

C. Interactive parameters configuration

It is quite clear from the previous subsections that the number of parameters to be configured is rather high. The choice of the parameters of both the decorrelator and the canceller may significantly affect the performance of the whole system. Very often the parameter setting is made *a priori*, according to preliminary tests, or it can be based on some analytical procedure. However, the resulting parameter configuration does not always provide customers with a satisfying quality of the processed signal. In fact, the wrong choice of even one of these parameters may result in a serious disease of the cancellation performance and, consequently, of the speech quality perceived by users.

In order to optimize the parameter setting, IEC can be used. As depicted in Fig. 2, the IEC block, containing the user feedback, supplies both the decorrelator and the SAEC with the parameter setting that yields the best perceived quality to user. In particular, an *interactive genetic algorithm* (IGA) is used and its pseudocode is briefly described in Algorithm 1. As can be seen, the algorithm starts by initializing a population $\mathbf{C}_0$ of S individuals. Considering a maximum number of generations N, the i-th population, with $i = 1, \ldots, N$, can be represented as a data matrix $\mathbf{C}_i = \begin{bmatrix} \mathbf{c}_{i,0} & \mathbf{c}_{i,1} & \ldots & \mathbf{c}_{i,S-1} \end{bmatrix}$. Each individual is a column vector $\mathbf{c}_{i,s}$, with $s = 0, \ldots, S-1$, containing the parameters to be set, i.e., the pre/de-emphasis factor ν, the contraction factor α, the step-size value μ, the regularization factor δ, the projection order K, and the adaptive filter length M. Their values are chosen within the following ranges: $\nu \in [0,1]$, $\alpha \in [0,1]$, $\mu \in [0.001, 2]$, $\delta \in [0.0001, 1]$, $K \in [1, 10]$, and $M \in [200, 1000]$. At each generation, the fitness function is computed for each individual, then the following operations are performed:

1) A fraction $r|S|$, with $r \in [0,1]$ *a priori* chosen, is selected for the next iteration. The probability of being selected is directly proportional to the inverse of its fitness (*fitness proportionate selection*).
2) The remaining $(1-r)|S|$ individuals are constructed from the previous individuals by *uniform crossing-over* [25].
3) A fraction $m|S|$, with $m \in [0,1]$ chosen *a priori*, is randomly mutated bit-wise.

The interactive fitness computation, which actively involves the user, is detailed in Algorithm 2. Note that in the interactive process, the user feedback can be received at each iteration, or once in a while, or simply when the user feels the need to enhance the perceived quality of the listened audio signal.

IV. Experimental results

Tests have been carried out to prove the effectiveness of the proposed interactive architecture in estimating an optimal parameter setting for SAEC systems. In the following, the adopted simulation setup and the obtained results are described, considering both objective and subjective

TABLE I. AUDIO SIGNALS UNDER TEST.

Test number	Audio signals
1	Continuous female voice and interfering male voice
2	Continuous female voice and background gaussian noise with standard deviation 0.0001 and interfering male voice
3	Background music and interfering male voice
4	Continuous female voice and background music and interfering male voice

TABLE II. CONFIGURATION FOR THE GA AND IGA ALGORITHMS.

Parameter	GA	IGA
Population	5	5
Maximum number of iterations	10	10
Cost function	Misalignment $\mathcal{M}$	Subjective evaluation in the scale 0-5
Elitism	Yes (3 individuals)	Yes (3 individuals)
Crossover	0.8	0.8
Mutation	$m = 0.05$	$m = 0.05$

evaluations. MATLAB code for the IGA is based on the `iga-audio` package freely available on BitBucket[1].

A. Simulation setup

The proposed interactive architecture has been applied to a SAEC application as shown in Fig. 1 but considering only one microphone in the receiving room. The input signals sampled at 8 kHz are listed in Table I and the echo paths have been modeled with IRs of length 256 samples simulated using the image method [33]. The target signal has then been recovered by using an APA-based SAEC, thus implying the setting of the aforementioned six free parameters, i.e., pre/de-emphasis factor ν, contraction factor α, step-size value μ, regularization factor δ, projection order K, and adaptive filter length M.

The goal of the proposed architecture is to find an optimal system configuration according to user preference, i.e., by using the IGA-based approach described in Section. III-C. In particular, user should be focused on three main aspects:

- the quality perceived after the decorrelation procedure,
- the effectiveness of echo cancellation, and
- the clearness of the residual echo to be transmitted to the remote room.

For this reason, the signal to be judged includes a local speaker interfering with the remote speaker. In particular, a male voice processed with the decorrelation algorithm has been added directly to the the residual echo signal in order to allow user to correctly rate the effect of decorrelation without involving the introduction of the double-talk detection procedure that is out of the scope of this paper. The results obtained with the IGA-based approach are compared with those obtained using a GA-based approach, i.e., a procedure based on the minimization of the normalized misalignment defined as follows:

$$\mathcal{M} = 20log_{10}\frac{\|\mathbf{W}_0 - \mathbf{W}_n\|}{\|\mathbf{W}_0\|}, \qquad (6)$$

being $\mathbf{W}_0$ and $\mathbf{W}_n$ the coefficients matrices including the unknown impulse responses and the filter estimates at time instant n, respectively. The parameters for GA-based and IGA-based approaches are detailed in Table II.

B. Objective evaluation

A first evaluation of the proposed architecture has been carried out from an objective point of view, i.e., providing at each iteration the misalignment computation as reported in Eq. (6). More specifically, the misalignment achieved with the GA-based approach and the IGA-based approach are provided for each iteration considering the best system configuration for that iteration.

Regarding the interactive procedure, the optimal configuration is obtained according to user preference. In particular, the user is asked to judge the clearness of the signal to be transmitted to the remote room in the following discrete scale: 0 (Perfect), 1 (Very Good), 2 (Good), 3 (Average), 4 (Bad), and (5) Very Bad. Regarding the GA-based approach, the optimal configuration is obtained according to the best misalignment value reported in Eq. (6) as previously stated.

The misalignment values are reported in Table III. The GA-based approach seems to perform better (improvements are lower than 2 dB) in terms of objective evaluation of echo cancellation since it outperforms the IGA-based approach in Test 1, Test 2, and Test 3. Anyway, a better misalignment value does not necessarily implies a better perceived audio quality. In fact, the trade-off between effectiveness of echo cancellation and clearness of the residual echo signal results an important issue for the assessment of the validity of such systems as it will be highlighted in the following section.

C. Subjective evaluation

The results of the previously described tests were subjectively evaluated. A listening panel was asked to judge four randomly extracted couple of processed signals resulting from the GA-based and IGA-based procedures. More specifically, subjects were asked to express a preference for each couple of processed signals. The subjects involved in the experiment were 10 expert listeners (8 males and 2 females), with ages ranging from 25 to 40. They were asked to evaluate

[1] https://bitbucket.org/ispamm/iga-audio/

TABLE III. MISALIGNMENT COMPUTED AT EACH ITERATION GIVEN THE BEST SYSTEM CONFIGURATION CONSIDERING THE GA-BASED APPROACH AND THE IGA-BASED APPROACH.

	Test 1		Test 2		Test 3		Test 4	
Iteration	GA	IGA	GA	IGA	GA	IGA	GA	IGA
1	-33.23	-37.32	-33.71	-36.92	-35.22	-37.73	-36.53	-36.53
2	-33,71	-34,71	-39,61		-37,73	-37,73	-31,13	
3	-33,71	-36,16	-39,87		-39,50	-37,73	-30,41	
4	-32,01	-36,16	-38,07		-37,73		-30,30	
5	-36,21	-34,71	-38,07		-37,73		-29,89	
6	-38,71		-37,17		-37,73		-30,40	
7	-37,32		-37,17		-39,11		-30,40	
8	-38,55		-38,50		-39,76		-30,40	
9	-37,69		-38,50		-39,82		-30,41	
10	-36,03		-38,50		-38,30		-30,33	

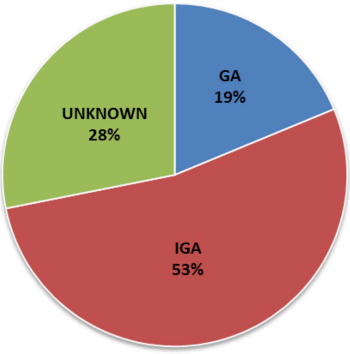

Fig. 3. Results of subjective listening tests in terms of percentage of users preference.

the overall quality of the audio signals to be transmitted to the remote room, merely making their preference between the two processed signals without knowing which procedure was used for each tests listed in Table I. Figure 3 shows the obtained results as percentage of users preference. Listening test results proved that the most of subjects preferred the human-driven genetic algorithm although from an objective point of view a better echo cancellation is obtained in terms of misalignment. In particular, the 53% of involved subjects rated as better the clearness of the signals processed with the configuration obtained through the IGA-based approach whereas only the 19% preferred the configuration obtained with the GA-based approach, and the 28% left undecided. No significant trends were found for the different sexes and ages. The results of the subjective listening tests have underlined the significance of the trade-off between effectiveness of echo cancellation and clearness of the residual echo signal. In fact, the listening panel rated as better the audio signals processed through the interactive procedure also in that cases where a better misalignment value was reached by the GA-based procedure, confirming issues previously underlined in Section II.

V. CONCLUSIONS

An interactive evolutionary algorithm has been applied to meet subjective requirements in the optimization stage of a stereophonic acoustic echo cancellation system. Starting from a previous approach focused on the optimization of the adaptive filter configuration in a single-channel scenario, the solution has been extended and generalized to a two-channel scenario, thus including both the decorrelation procedure required to overcome the well-known non-uniqueness problem and the adaptive algorithm introduced to estimate the echo paths. A human-driven architecture has been developed to obtain the optimal system configuration according to the quality requirements demanded by users availing such stereophonic systems. Different tests have been carried out to prove the effectiveness of the proposed approach according to both objective and subjective measures, also making comparisons with a misalignment-driven procedure. Objective results have proved the effectiveness of both approaches in correctly identifying the echo paths. As expected, subjective results confirmed that objective procedures may be not optimal from a perceptual point of view, since the subjects involved in the listening tests preferred the audio signals processed through the IGA-based procedure over the audio signals processed through the GA-based procedure. Future work will be oriented to the generalization of such interactive architecture to a multichannel scenario, thus involving a higher number of parameters.

REFERENCES

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

[2] L. Romoli, S. Cecchi, D. Comminiello, F. Piazza, and A. Uncini, "Novel decorrelation approach for an advanced multichannel acoustic echo cancellation system," in *22nd European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, Sep. 2014, pp. 651–655.

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

[4] D. Comminiello, S. Cecchi, M. Gasparini, M. Scarpiniti, A. Uncini, and F. Piazza, "Advanced intelligent acoustic interfaces for multichannel audio reproduction," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, Beijing, China, Jul. 2014, pp. 3577–3584.

[5] D. Comminiello, S. Cecchi, M. Scarpiniti, M. Gasparini, L. Romoli, F. Piazza, and A. Uncini, “Intelligent acoustic interfaces with multisensor acquisition for immersive reproduction,” *IEEE Transactions on Multimedia*, vol. 17, no. 8, pp. 1262–1272, Aug. 2015.

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

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

[8] L. Romoli, S. Cecchi, and F. Piazza, “A novel decorrelation approach for multichannel system identification,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014.

[9] M. Ali, “Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 6, Seattle, WA, USA, May 1998, pp. 3689–3692.

[10] A. Hirano, K. Nakayama, and K. Takebe, “Stereophonic acoustic echo canceler with pre-processing - second-order pre-processing filter and its convergence,” in *Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Kyoto, Japan, Sep. 2003, pp. 63–66.

[11] J. Herre, H. Buchner, and W. Kellermann, “Acoustic echo cancellation for surround sound using perceptually motivated convergence enhancement,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 1, Honolulu, HI, Apr. 2007, pp. 17–20.

[12] B. C. Bispo and D. D. S. Freitas, “Hybrid pre-processor based on frequency shifting for stereophonic acoustic echo cancellation,” in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, Bucharest, Romania, Aug. 2012, pp. 2447–2451.

[13] M. M. Sondhi, D. R. Morgan, and J. L. Hall, “Stereophonic acoustic echo cancellation - An overview of the fundamental problem,” *IEEE Signal Processing Letters*, vol. 2, pp. 148–151, Aug. 1995.

[14] A. Gilloire and V. Turbin, “Using auditory properties to improve the behaviour of stereophonic acoustic echo cancellers,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 6, Seattle, WA, May 1998, pp. 3681–3684.

[15] A. Uncini, *Fundamentals of Adaptive Signal Processing*, ser. Signal and Communication Technology. Cham, Switzerland: Springer International Publishing AG, 2015, ISBN 978-3-319-02806-4.

[16] E. Eweda, “Comparison of RLS, LMS and sign algorithms for tracking randomly time-varying channels,” *IEEE Transactions on Signal Processing*, vol. 42, no. 11, pp. 2937–2944, Mar. 1994.

[17] J. Benesty, P. Duhamel, and Y. Grenier, “A multichannel affine projection algorithm with applications to multichannel acoustic echo cancellation,” *IEEE Signal Processing Letters*, vol. 3, no. 2, pp. 35–37, Feb. 1996.

[18] K. Tang, K. Man, S. Kwong, and Q. He, “Genetic algorithms and their applications,” *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 22–37, Nov. 1996.

[19] D. Krusienski and W. Jenkins, “Design and performance of adaptive systems based on structured stochastic optimization strategies,” *IEEE Circuits and Systems Magazine*, vol. 5, no. 1, pp. 8–20, Mar. 2005.

[20] N. Amiri and S. Fakhraie, “Digital network echo cancellation using genetic algorithm and combined ga-lms method,” in *Proceedings of the IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, Singapore, Dec. 2006, pp. 1822–1825.

[21] L. Romoli, S. Cecchi, and F. Piazza, “A combined approach for channel decorrelation in stereo acoustic echo cancellation exploiting time-varying frequency shifting,” *IEEE Signal Processing Letters*, vol. 20, no. 7, pp. 717–720, Jul. 2013.

[22] H. Takagi, “Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation,” *Proceedings of the IEEE*, vol. 89, no. 9, pp. 1275–1296, Sep. 2001.

[23] D. Comminiello, S. Scardapane, M. Scarpiniti, and A. Uncini, “User-driven quality enhancement for audio signal processing,” in *134th Audio Engineering Society Convention*, Rome, Italy, May 2013.

[24] ——, “Interactive quality enhancement in acoustic echo cancellation,” in *36th International Conference on Telecommunications and Signal Processing (TSP)*, Rome, Italy, Jul. 2013, pp. 488–492.

[25] S. Luke, *Essentials of Metaheuristics*. Lulu, 2009. [Online]. Available: http://cs.gmu.edu/~sean/book/metaheuristics/

[26] J. Biles, P. Anderson, and L. Loggi, “Neural network fitness functions for a musical IGA,” in *Proceedings of the International ICSC Symposium on Intelligent Industrial Automation (IIA)*, Reading, UK, Mar. 1996. [Online]. Available: https://ritdml.rit.edu/handle/1850/3065

[27] E. Larsen and R. M. Aarts, *Audio Bandwidth Extension*. J. Wiley & Sons, 2004.

[28] J. Lee, E. Song, Y. Park, and D. Youn, “Effective bass enhancement using second-order adaptive notch filter,” *IEEE Transactions on Consumer Electronics*, vol. 54, pp. 663–668, May 2008.

[29] L. Romoli, S. Cecchi, P. Peretti, and F. Piazza, “A mixed decorrelation approach for stereo acoustic echo cancellation based on the estimation of the fundamental frequency,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 2, pp. 690–698, Feb. 2012.

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

[31] A. J. Blauert, *Spatial Hearing*. MIT press Cambridge Massachusetts, 1997.

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

[33] J. B. Allen and D. A. Berkeley, “Image method for efficiently simulating small - room acoustics,” *Journal of the Acoustic Society of America*, vol. 65, pp. 943–950, 1979.