adopted according to objective measurements, such as the mean-square error minimization (see for example [5]). Such error-driven strategies do not take into account any subjective parameter but are only based on the processed signal. This is a challenging problem since even a wrong parameter setting may change the results of an enhancement technique in terms of perceived quality. As a consequence, the qualitative constraints fixed by user cannot be assured.

In order to address this problem, we propose a new framework which includes a user feedback, related to the subjective perceived quality, in the setting optimization of audio enhancement techniques. The proposed method is based on the inclusion into a classical optimization strategy of *interactive evolutionary computation* (IEC) methods.

Albeit speech enhancement has not been a prominent line of research for IEC, all applications developed up to now using this methodology show interesting capabilities in adapting to user preference and generalizing unknown situations. In the following we detail three representative examples that we consider quite significant and that strongly motivated this work. For a general review of IEC applications up to 2001 we refer the interested reader to [6]. According to it, we define all error-driven approaches as *analytical*, and to user-driven approaches as *interactive*.

Watanabe et al. [7] applied IEC to the design of an FIR filter for recovering distorted speech. The interactive system showed statistical advantage over the analytical approach in all the performed tests. IEC was also applied to optimize a system in the context of hearing aid [8, 9]. This resulted in reduced expert intervention, and the possibility of using everyday audio signals for the optimization process, including musical samples. More in general, IEC was also applied for the automatic generation of simple melodic samples [10, 11]. Users were then able to generate short melodies that received positive feedback.

The aim of this paper is to provide a common framework for interactive audio enhancement using IEC to foster future research and applications. As a preliminary result, we take into account an applicative case study dealing with the *acoustic echo cancellation* (AEC). One of the main problem in AEC is the parameter setting of the adaptive algorithm, which may change according to devices and scenarios. This leads even well-performing cancellers to yield unsatisfying results from a point of view regarding the signal quality perceived by the user. Simulation results prove the effectiveness of the proposed interactive optimization setting in providing the best quality from a subjective perspective.

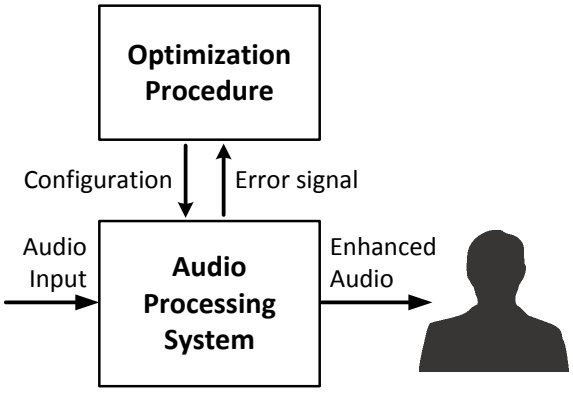


**Fig. 1:** A classic configuration procedure for audio signal processors.

The paper is organized as follows: in Section 2 the problem formulation of audio quality enhancement is described. In Section 3 the proposed method based on the interactive quality enhancement is introduced. Section 4 details a set of possible applications for interactive audio signal processing, while Section 5 focuses on an applicative case study dealing with AEC, on which experiments are conducted. Results and statistical significance are discussed in Section 6, and finally our conclusions and future lines of research are drawn in Section 7.

## 2. QUALITY ENHANCEMENT FOR PROCESSED AUDIO

The analytical approach to audio quality enhancement is represented in Fig. 1. It is composed of a main audio signal processing system and an optimization procedure that provides the system with the optimal parametric configuration to use. Therefore, the enhanced audio signal returned to the user depends on the accuracy of both the audio processor and the optimization procedure.

Audio signal processing systems may be simple al-

gorithms, adaptive filters or more complex model based on psychoacoustic or cognitive techniques. Their development may be based on different kinds of quality optimization. In particular, the goodness of an audio system may be evaluated according to objective methods or measurements, or also using subjective tests in an attempt to provide an enhanced audio signal satisfying the quality requirements demanded by customers. However, a perfect development of the audio processor is not as much sufficient to guarantee a certain audio quality since the optimization procedure affects the configuration of the system.

Unlike the audio processor, the techniques used to evaluate the goodness of the optimization procedures are often restricted to objective measurements. In particular, also depending on the audio processor to optimize, these procedures may be based on the optimization of certain cost functions, such as the minimization of the mean-square error, or on some objective indices. However, objective procedures may be not 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 audio processor, whose potential quality is higher than the perceived one. This is the main reason why we propose a new method to improve the optimization procedure by including a user feedback.

## 3. INTERACTIVE QUALITY ENHANCEMENT USING IEC

### 3.1. Interactive Evolutionary Computation

*Evolutionary algorithms* (EA) [12] are 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. *Interactive EA* (IEA) can be defined as any EA in which an interaction between the algorithm and the end user occurs [6]. A simpler definition, also useful here, is considering IEA as any EA in which the function to be optimized, denoted in 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 IEA is able to efficiently search the psychological space of user preferences.

The main drawback arising in the use of an IEA is *user fatigue*. This effect is of tantamount importance in audio applications where a single fitness evaluation can require several seconds, up to minutes. For practical purposes, any IEA has only a short time to successfully converge to an optimum. This requirement in turn imposes strong constraints on the required convergence speed of the global algorithm. A second related problem is that of *user discrimination*. In fact, only intensively trained ears can discriminate between very similar sounds. Thus, a medium user only enforces a partial ordering over possible solutions to the optimization problem.

Both problems can be partially solved by allowing the fitness to take only a small set of values (typically between 5 and 10). This eases the burden on the users, thus evaluating a single sound becomes easier and reduces the optimization problem to that of finding a macroregion in the search space where the fitness is sufficiently small. Any representative sound from that region is then an acceptable solution to the original problem.

Another solution that is widely used in IEC research is inserting an adaptive layer between the user and the IEC, such as a *neural network* (NN) [13]. The NN learns the user preferences in the beginning and is then used to predict future evaluations with a given confidence that increases with time. The number of fitness evaluations required to the user is then gradually reduced as the capabilities of the NN improve.

### 3.2. Proposed Framework for Interactive Audio Enhancement

The proposed framework for interactive audio quality enhancement is summarized in Fig. 2. As can be seen, IEC is now used as an active part of the optimization procedure. The analytical error of the classical approach is thus substituted with a real-time feedback from the user. The subjective evaluation is then used to iteratively search for an optimal configuration for the audio enhancer.

The main advantage of this architecture, which has already been extensively discussed, 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. The optimization process can

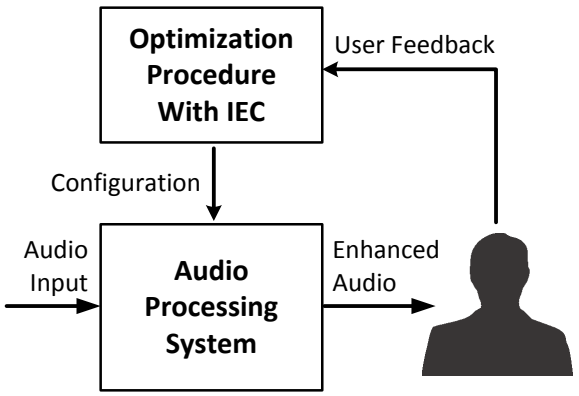


**Fig. 2:** Interactive configuration procedure for audio signal processors.

now be run in an initialization phase, spaced at regular intervals, or directly enforced by the user when performance degrades.

In addition, there is nothing preventing a successful mix of analytical and interactive approaches. In fact, a carefully designed optimization strategy could combine an analytical error coming from the audio enhancer with a subjective evaluation coming from the user. We expect this to be a natural line of research for the future.

## 4. INTERACTIVE APPLICATIONS FOR AUDIO SIGNAL PROCESSING

In this Section we detail some possible applications of our framework. For a review of already existing works on IEC for audio processing, we refer the reader to Section 1.

### Immersive Speech Communications

Modern speech communications are evolving towards a new direction which involves users in a more perceptual way. That is the *immersive experience*, which may be considered as the "last-mile" problem of communications. One of the main features of immersive communications is the *distant-talking*, i.e. the hands-free (in the broad sense) speech communications without body-worn or tethered microphones that takes place in a multisource environment where interfering signals may degrade the communication quality and the intelligibility of the desired speech source [14]. In order to preserve the speech quality, *intelligent acoustic interfaces* may be used to capture and enhance speech signals [15]. Immersive communications involve several problems, from interfering noise reduction to acoustic echo cancellation, thus assuring a high quality of the processed speech signals is absolutely challenging. In this context, an interactive control of the perceived quality through IEC can guarantee to users a high intelligibility of the acquired speech and, consequently, a high quality of the speech communications.

### Microphone Array Calibration

When an acoustic interface is composed of a microphone array, the disposition of microphones with respect to the involved sources plays a fundamental role in providing a high quality enhancement of acquired signals, especially in immersive audio applications or in speech recognition. Moreover, when sources move within a certain environment, the optimal geometry of the microphone array might vary in time, thus it could be useful to use only a selection of available microphones. In this scenario, an interactive optimization of the array configuration may facilitate a run-time calibration, thus preserving the quality of the processed acquired signals.

### Acoustic Spatial Sound Reproduction

An intelligent acoustic interface may be composed not only of a microphone array, but also of a loudspeaker array [14]. In this regard, one of the main tasks of an intelligent acoustic interface is to reproduce the desired acoustic information taking into account that the listener would hear the sound exactly as in the original sound field. This feature indeed is known as *spatial sound reproduction*. However, the spatial perception of sound may vary according to user relish, therefore a user feedback can improve his satisfaction. To this end, IEC may be used to reproduce a user-liking spatial sound.

### Audio Equalization

Audio signal equalization is another application in which the concept of quality strictly depends on the subjective perception of the user. From music signals to speech, a user may express his own pleasure in listening, so that an equalization can be considered optimal with respect to each individual user. This problem may be addressed by using an interactive feedback (through IEC) from the user to enhance the perceived quality of an audio signal.

In this process a fundamental role is played by the adaptive filter which estimates the acoustic impulse response. However, adaptive filters are very sensitive with respect to the setting of their parameters, such as filter length, step size, regularization factor and others (see for example [16]). In fact, even a small change of such parameter values may imply a disease of the canceller, notwithstanding the appropriate choice of the adaptive filter. 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 order to optimize the parameter setting procedure IEC can be used. As it is possible to see in Fig. 3, the IEC block, containing the user feedback, supplies the adaptive filter with the parameter setting that yields the best perceived quality to user. In the *interactive acoustic echo cancellation*, 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.

## 6. EXPERIMENTAL RESULTS

### 6.1. Test Setup

We evaluated the proposed architecture on five target signals, all of which are interfered by the addition of a female continuous voice. The target signal is then recovered by using an acoustic echo canceller based on an *affine projection algorithm* (APA) (see for example [5]) which implies 4 free parameters to set (i.e. filter length, step size value, regularization factor and projection order). By denoting a configuration of the filter as $\lambda$, our problem is then finding an optimal configuration $\lambda^*$. The meaning of "optimal" must be defined in an analytic way according to the classical approach, whereas it is defined in terms of "user optimality" in the interactive method. In our experiment, such configuration is chosen first using an *interactive genetic algorithm* (IGA), and then by a classical GA minimizing the *normalized misalignment* defined as:

$$\mathcal{M} = 20 \log \frac{\|\mathbf{w}_0 - \mathbf{w}_n\|}{\|\mathbf{w}_0\|} \qquad (1)$$

| Number | Original signal |
|---|---|
| Test 1 | White Gaussian noise |
| Test 2 | Continuous male voice |
| Test 3 | Continuous male voice and additive white Gaussian noise |
| Test 4 | Background music |
| Test 5 | Background music and continuous male voice |

**Table 1:** Original audio signals

where $\mathbf{w}_0$ is the unknown impulse response to estimate and $\mathbf{w}_n$ is the filter estimate. The choice of the normalized misalignment as measure to optimize is due to the fact that it is based on the least perturbation property [5], thus reflecting the perceived quality of the processed speech signal. In fact, unlike error minimization-based measures, when the processed signal introduces some artifacts, such as the musical noise, the normalized misalignment shows a jumpy behaviour [14].

The processings of the five signals are then subjectively evaluated by asking users an assessment regarding a randomly extracted couple of processed signals resulting both from GA and IGA implementations. Assessment results are averaged to draw the overall conclusions. It is important to note that equation (1) cannot be minimized in a realistic situation since it is defined in terms of the unknown impulse response $\mathbf{w}_0$. Also, it is a typical choice for comparing analytic algorithms. Thus, an estimate $\mathbf{w}_n$ can be expected to approximate the theoretical optimum that could be reached by an expert fine-tuning of APA parameters.

More in detail, we have evaluated the proposed IGA-based AEC for all the tests described in Table 1. As can be seen, in order to obtain an uniform testing, different combinations of voice, music and white noise were tested. The parameters for the IGA and GA are detailed in Table 2. A set of experts was asked to serve as evaluators for the IGA. Then, we asked a set of 30 people, equally subdivided by sex, to evaluate processed audio signals for two of the five original signals randomly chosen. Each person

| Parameter | GA | IGA |
|---|---|---|
| Population | 50 | 5 |
| Iterations | 20 | 10 |
| Cost function | Misalignment $\mathcal{M}$ | Subjective evaluation in the scale 0-5 |
| Elitism | Yes (5 individual) | Yes (1 individuals) |
| Crossover | Uniform crossover (both) at 0.8 rate | |
| Mutation | Random mutation (both) with 0.05 probability | |

**Table 2:** Configuration of the IGA and GA algorithms

listened to the chosen processed signals coming both from the IGA and the GA. Finally, he was asked if he preferred the first, the second, or none of the two. In this manner we obtained a set of 60 listening experiments. Next, we summarize and analyze the results.

### 6.2. Results Analysis

Results of the test detailed in the previous subsection are presented in Figure 4. Slightly more than half of the tests (55%) resulted in a clear preference for the configuration obtained by the IGA, with only 28% that preferred the classical GA and 17% left undecided. The 27% difference between the two algorithms is more striking if we consider that the result obtained by the GA is proximal to a theoretical optimum for analytical approaches, while there is still room for many improvements for the IGA.

Next, we proceed to a brief statistical analysis to show that our result is significant. This analysis is necessary to show that the aforementioned percentages were not due to random choices of the users. For further details on the methodology used, we refer the reader to [17]. We define significant as having less than 5% probability of happening by sheer chance. If configurations obtained by IGA and GA were equivalent, we could expect the subjective probability of choosing the result obtained by IGA, denoted as $\mu_0$, to be approximately 0.5. This is our *null hypothesis*:

$$H_0 : \mu_0 = 0.5 \tag{2}$$

In our case, 33 users have chosen IGA, and 17 users have chosen GA. Thus, the estimated value for $\mu_0$

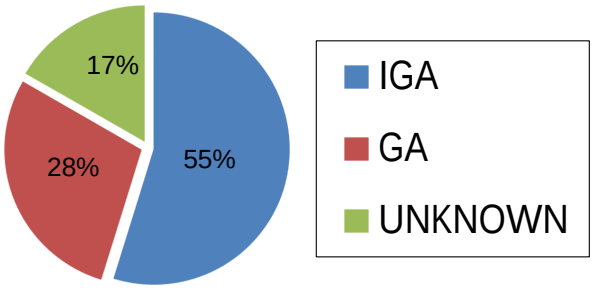


**Fig. 4:** Results of subjective test

is $\hat{\mu_0} = 33/50 = 0.66$. If the null hypothesis is true, since the number of observations is large, we can assume that the probability of observing $\hat{\mu_0}$ follows a *t-student distribution* with $df = n-1 = 49$ degrees of freedom, where $n = 33+17$ is the total number of samples. The *standardized t-value* in our experiment is then given by:

$$t = \frac{\hat{\mu_0} - 0.5}{s/\sqrt{n}} = 2.3634 \tag{3}$$

where $s = 0.4785$ is the sample standard deviation computed from our data. The corresponding one-tailed *p-value* is $p(2.3634) = 0.0110595$. Thus, our result has only 1.1% probability of being due to chance. Since this is lower that our allowed significance level, we can reject the null hypothesis and assert that the better performance of IGA is significant from a statistical point of view.

## 7. CONCLUSIONS AND FUTURE WORK

In this paper we have introduced a general framework for user-driven audio enhancement, using a

mixture of classical and interactive optimization strategies. Our results in the case study of echo cancellation have shown that a system designed in this way is able to efficiently adapt to user preferences and generalize to future users in a significant way. A natural line of research is now extending our methodology to the problems considered in Section 4. Research is also needed to find more sophisticated techniques for reducing user fatigue and improving convergence time. In this respect, a particular importance is given by the remarks discussed in Section 3.2 on the possibility of mixing analytical errors and user-driven evaluations.

## 8. REFERENCES


[1] ITU-R Recommendation BS.1387. Method for objective measurements of perceived audio quality, 1998.

[2] P. Loizou. *Speech Enhancement: Theory and Practice*. CRC, Boca Raton, FL, 2007.

[3] D. Campbell, E. Jones, and M. Glavin. Audio quality assessment techniques – a review, and recent developments. *Signal Processing*, 89(8):1489–1500, August 2009.

[4] J. G. A. Barbedo and A. Lopes. A new cognitive model for objective assessment of audio quality. *Journal of the Audio Engineering Society*, 53(1/2):22–31, February 2005.

[5] A. H. Sayed. *Fundamentals of Adaptive Filters*. John Wiley & Sons, Inc., Hoboken, NJ, 2003.

[6] H. Takagi. Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation. *Proceedings of the IEEE*, 89(9):1275–1296, 2001.

[7] Tatsumi Watanabe and H Takagi. Recovering system of the distorted speech using interactive genetic algorithms. *IEEE International Conference on Systems, Man and Cybernetics, 1995*, pages 684–689, 1995.

[8] Hideyuki Takagi and Miho Ohsaki. Interactive evolutionary computation-based hearing aid fitting. *IEEE Transactions on Evolutionary Computation*, 11(3):414–427, 2007.

[9] A. Schelsinger and M. M. Boone. Evolutionary optimization for hearing aids of computational auditory scene analysis. In *126th AES Convention*, Munich, Germany, May 2009.

[10] Brad Johanson and R Poli. GP-music: An interactive genetic programming system for music generation with automated fitness raters. *Genetic Programming 1998: Proceedings of the Third Annual Conference*, 1998.

[11] J Biles. GenJam: A genetic algorithm for generating jazz solos. *Proceedings of the International Computer Music Conference (ICMA)*, 1994.

[12] Sean Luke. *Essentials of metaheuristics*. 2009.

[13] 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'96)*, 1996.

[14] D. Comminiello. *Adaptive Algorithms for Intelligent Acoustic Interfaces*. PhD thesis, 'Sapienza' University of Rome, December 2011.

[15] D. Comminiello, M. Scarpiniti, R. Parisi, and A. Uncini. Intelligent acoustic interfaces for immersive audio. In *134th AES Convention*, Rome, Italy, May 2013.

[16] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay. *Advances in Network and Acoustic Echo Cancellation*. Springer-Verlag, Berlin, Heidelberg, New York, 2001.

[17] R Peck, C Olsen, and JL Devore. *Introduction to statistics and data analysis*. 2011.