Blind Source Separation in Reverberant Environments Using Genetic Algorithms

Felipe Augusto Pereira de Figueiredo
Instituto Nacional de Telecomunicações - Inatel
Santa Rita do Sapucaí - MG - Brazil
felipeaugusto@mtel.inatel.br

Carlos Alberto Ynoguti
Instituto Nacional de Telecomunicações - Inatel
P.O. Box 05 - 37540-000
Santa Rita do Sapucaí - MG - Brazil
ynoguti@inatel.br

***Abstract* — This article proposes the use of Genetic Algorithms (GA) to solve a Blind Source Separation (BSS) problem. Although the subject of BSS by means of various techniques, such as ICA, PCA and Neural Networks, has been largely discussed in the literature, to date the possibility of employing genetic algorithms has not been fully explored. The approach presented here makes use of a genetic algorithm to blindly solve the problem of separating speech signals in reverberant environments. The system parameters are represented as chromosomes and the Signal-to-Interference Ratio (SIR) for the output channels is used as the fitness function for the GA. Computer simulations show that the SIR is maximized and when compared to the results produced by a standard time-domain BSS algorithm, the method adopted here, has a little performance gain over that.**

***Index Terms* — BSS, Acoustic Environments, Genetic Algorithms, Cocktail Party Problem.**

I. Introduction

Humans have the ability to focus their attention on a single talker amongst a lot of conversations and background noise, and yet, recognize a specific voice. This ability is known as the "*cocktail party effect*".

To mimic this behavior, the Blind Source Separation problem was proposed. This problem consists of recovering unknown signals or "sources" from several observed mixtures. Typically, these mixtures are acquired by a set of sensors, where each of them receives mixtures of all the sources. The term "blind" is justified by the fact that the only *a-priori* knowledge we have for the signals is their statistical independence. No other information about the signal distortion on the transfer paths from the sources to the sensors is available beforehand.

There are many potential applications for blind source separation techniques. Some of them refer to communication systems [1], biomedical signal analysis such as MEG, ECG, EEG [2], speech enhancement and noise reduction (denoising) [3,4,5,6], and speech recognition [7,8].

The speech recognition technology is still vulnerable when dealing with signals in the presence of acoustic interference. Specifically, one of the most difficult problems encountered is the interfering speech from competing stationary speakers. Robust speech recognition in real environments still remains a challenging task.

Generally, a great number of algorithms for BSS of speech signals have been proposed [9,10]. However, most of them deal with the instantaneous mixture of sources [9,10] and only a few methods examine the case of convolutive mixtures of speech signals [3-8].

In this paper we propose an off-line blind signal separation method in the time domain which uses a genetic algorithm (GA) [13] to separate speakers in a simulated reverberant environment. This environment is simulated through the convolution of the speech signals and the room impulse response generated by the image method [11,12].

GA is a search technique to search for exact or approximate solution to an optimization problem. This method expresses the system parameters as a binary or real-valued array, corresponding to chromosomes, and tries to find the optimum solution for the system parameters, using an evolutionary process. The chromosomes also correspond to individuals in a population. This method can realize powerful optimization for the system parameters.

In order to consider the reverberation, GA is introduced into the blind separation system for temporally and spatially mixed voices. The separating system is composed of non-recursive linear filters. The filter coefficients are concatenated to make a sequence as a chromosome. GA is applied to determine the values of the filter coefficients, so that the SIR for each of the system's output is maximized.

The structure of this paper is as follows: In the next section the system model used to blindly separate speech signals in reverberant environment is shown and then, in section III, we discuss the principle of genetic algorithms and present the basic method for separating convolutively mixed speech signals, based on the maximization of the Signal-to-Interference Ratio (SIR). In section IV, the experiments carried out for the evaluation of the aforementioned algorithm's performance are presented. Finally, in the last section some conclusions are given.

II. Model Of Blind Source Separation in Reverberant Environment

Let us assume that we have Q speech sources, denoted by $s_q(n)$, $q = 1,\ldots,Q$, which are considered to be zero meaned, mutually stochastic independent, in reverberant environment. In addition, let P be the number of sensors (microphones).

These microphones acquire the convolutive mixture of the speech signals denoted by $x_p(n), p = 1,\ldots,P$. Due to the room acoustics, the sensors acquire besides the speakers' speech signal, delayed versions as well as multiple echoes that propagate in the room. In order to solve the BSS problem, we make the assumption that the number of the speech signals that must be separated is known beforehand and that it is equal to the number of sensors, i.e., $Q = P$. Figure 1 shows the MIMO mixing and demixing system generally adopted for BSS in order to model the acoustic environment.

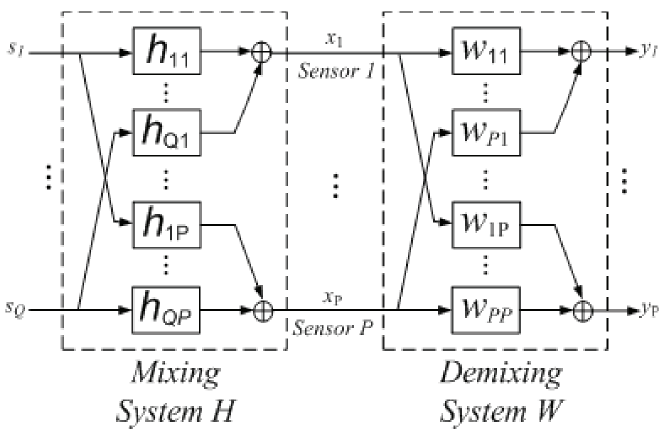

Figure 1: System modeling the acoustic environment.

The reverberation and sound absorption characteristics of a room can be simulated through the convolution of the room's impulse response and the original source signals. An acoustic impulse response can be precisely and efficiently simulated by the Image Method [11][12].

The signal obtained from the microphones is expressed as the following equation.

$$x_p(n) = \sum_{q=1}^{P} \sum_{k=0}^{M-1} h_{qp}(k)\, s_q(n-k) \tag{1}$$

As it can be seen, it is the equation of a M-tap mixing system, where $h_{qp}(k), k=0,\ldots,M-1$ denotes the coefficients of the finite impulse response (FIR) filter model from the q-th source to the p-th sensor.

In BSS, we are interested in finding a corresponding demixing system according to Figure 1, where the output signals $y_q(n),\ q = 1,\ldots,P$ are described by

$$y_q(n) = \sum_{p=1}^{P} \sum_{k=0}^{L-1} w_{pq}(k)\, x_p(n-k) \tag{2}$$

where $w_{pq},\ p = 1,\ldots,P,\ q = 1,\ldots,Q$ are the coefficients of the demixing filters and L is the length of these filters.

It can be shown (see, e.g., [14]) that the MIMO demixing system coefficients $w_{pq}(k)$ can in fact reconstruct [15] the sources up to an unknown permutation and an unknown filtering of the individual signals, where L should be chosen at least equal to M.

The problem of the blind source separation is to determine the values of $w_{pq}(k)$, so that the output signals $y_q(n),\ q = 1,\ldots,P$ are independent.

Several methods have been proposed over the years to solve this problem. Typically, they adopt an evaluation function which represents the degree of the independency between the output signals $y_q(n),\ q = 1,\ldots,P$ and minimize it with a certain method, such as a gradient method. Kawamoto et al. showed that multiple voices mixed in noiseless reverberating environment can be separated with a gradient method minimizing the following equation [16].

$$Q = \frac{1}{2} \sum_{i=1}^{N} \{\log E[y_i(n-L)^2] - \log \det E[y(n-L)\, y(n-L)^T]\} \tag{3}$$

When $y_q(n),\ q = 1,\ldots,P$ are independent, the value Q is close to zero.

In BSS the evaluation function is known as cost function and in GA it is known as fitness function, from now on we adopt the latter term.

III. Introduction of Genetic Algorithms into Blind Source Separation

A. Principle of Genetic Algorithms

A genetic algorithm (GA) is a technique to search for the optimal or suboptimal solution of a system by maximizing a certain evaluation function named as fitness function. GA initially generates a population of individuals, which correspond to chromosomes in genetics. The initial individuals are generated randomly. Each individual represents a whole system configuration (in our case, the weigths of the demixing filter). Then the fitness function value of these individuals is evaluated and those with higher values are selected as survivors to the next generation. Moreover, new individuals are additionally reproduced by crossover and mutation using the survivors, and accordingly a new population is created. Then the evaluation of the fitness function value for all individuals, selection, and reproduction are carried out and this procedure is iterated until the fitness function takes a high enough value or gets saturated adequately. Finally, the solution of the system parameter is obtained from the individual which takes the highest fitness function value. This procedure is shown in Figure 2.

B. Application of Genetic Algorithm to Blind Source Separation

GA can be applied to the problem of blind source separation by adopting an evaluation function which can be maximized. Here, an individual is generated as a sequence of the filter coefficients $w_{pq}(k)$ in (2) as shown in Figure 3.

In this paper we present a brand-new approach for the fitness function. We adopt the Signal-to-Interference Ratio, shown in equation (4), as the fitness function.

$$C = SIR_{y_q} = 10 \log \frac{E\{y^2_{s_r,q}(n)\}}{E\{y^2_{c,q}(n)\}}$$
$$SIR_{y_q} \approx 10 \log \frac{\sum y^2_{s_r,q}(n)}{\sum y^2_{c,q}(n)} \quad (4)$$

where $y_{s_r,q}(n)$ is the component containing the desired source $s_r(n)$ and $y_{c,q}(n)$ is the crosstalk component in the q-th output channel stemming from the remaining point sources that could not be suppressed by the BSS algorithm. In general, the desired source at the q-th output channel can be any of the source signals due to the permutation ambiguity. As shown in (3), the expectation operator E{·} has to be replaced in practice by a time-average.

In order to use (4) as the fitness function, we first have to apply each of the individuals (chromosomes) generated by the GA procedure to equation (2) to obtain the outputs y.

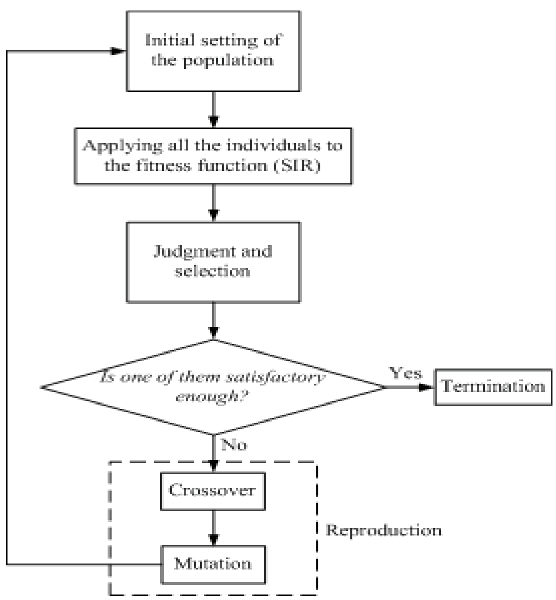

Figure 2: The procedure of the genetic algorithm.

$W_{11}(0)$	·····	$W_{11}(L-1)$	$W_{12}(0)$	·····	$W_{21}(L-1)$	$W_{22}(0)$	·····	$W_{22}(L-1)$

Figure 3: The structure of a chromosome.

In this work we adopt real-valued GA for correspondence with signal processing. In summary, the GA-based BSS algorithm can be implemented as the following iterative procedure.

(a) Initial setting

An initial population of I individuals as shown in Figure 3 is created from a random initial set of parameters. The length of each individual is $4L$, because we have 4 demixing filters of length L.

(b) Selection

The input voice signals $x_p(n), p = 1,\ldots,P$ are processed according to (2), using the I sets of parameters $w_{pq}(k)$, in order to obtain I sets of output signals $y_q(n),\ q = 1,\ldots,P$. Then (4) is used to obtain the fitness evaluation C for each individual. The R individuals with the higher value of C are selected to survive for the next generation. The remaining I-R individuals are discarded

(c) Crossover

Here, we adopt the reproduction technique known as the uniform crossover to adequately mix the characteristics of the parents. This technique is depicted in Figure 4. Then, S pairs of the survived R individuals are randomly selected and crossover is performed with them. In the uniform crossover, each $w_{pq}(k)$ in the descendant chromosomes take the value of the corresponding $w_{pq}(k)$ in either of the parents at a probability of 50%. In this stage, $2S$ new individuals are generated.

(d) Mutation

After reproduction (crossover) takes place, (I-R-$2S$) individuals are newly generated by mutation. After mutation, P individuals are prepared for the next generation. In order to avoid the trap of the plateau of convergence, every $w_{pq}(k)$ constituting a chromosome is randomly changed, i.e., a random value in a certain range is added to every $w_{pq}(k)$.

(e) Termination

Finally, the new population is applied to equation (2) along with the mixed signals $x_p(n), p = 1,\ldots,P$ and the value C for every chromosome in the new population is evaluated. If the value C is greater than a predefined threshold, the individual is chosen as the solution and the procedure is finished. The separation of the mixed speech signals is performed by (2) making use of the filter coefficients $w_{pq}(k)$ obtained as the solution.

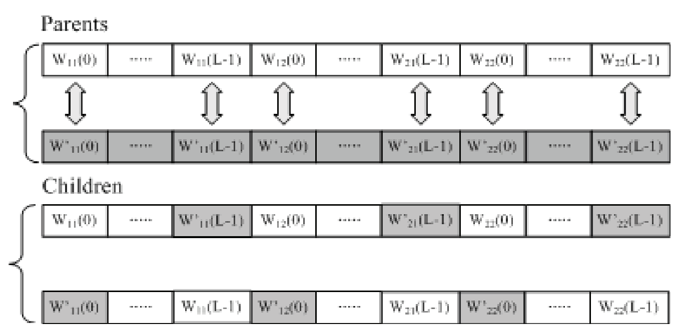

Figure 4: An example of the uniform crossover.

IV. Computer Simulations

A. Simulated Mixture Model and GA

The experiments executed for this work were focused on testing the proposed GA-based BSS method in the case of two simultaneous speakers in a reverberant simulated environment. Filters with 447 taps (M = 447) were generated by the image method with the purpose of simulating the acoustical behavior of a real room. Two audio signals with 5 seconds of speech were convolved with the synthetic impulse response of a room generated by the image method. This speech signals correspond to a male and a female speaker voices. The recordings were taken in a low noise environment with 11025 Hz sampling frequency and 16 bits of resolution.

The length L of the demixing filters is made equal to the length M of the mixing filtes, so we have L=M=447. The demixing filters W_{pq} , i.e., the initial set of chromosomes, are randomly initialized. The parameters I, R and S are made equal to 100, 50 and 25 respectively.

B. Results of Computer Simulations

Figures 5 and 6 present the average SIR for both output channels when the GA-based BSS and the standard BSS methods are respectively adopted. In both experiments, the same set of speech signals and number of filter coefficients L and M were used. The standard BBS method relies on the minimization of a cost function through the use of a gradient descent approach. The technique used to adapt the step size of the standard BSS method is known as fixed step size [17], and it is made equal to 0.002.

According to the results shown in Table 1, when the GA-based BSS method is adopted, the average time for each epoch is approximately 16.9 times smaller than the average time presented for the standard BBS method. It should be also noted that the final SIR for the GA-based BSS method is also greater than the final value achieved by the standard BBS method.

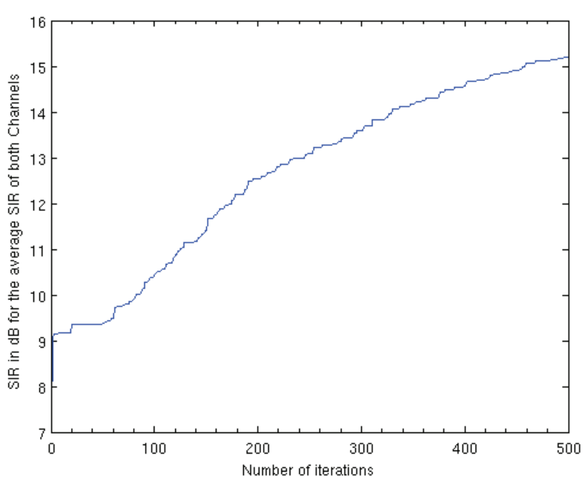

Figure 5: Average SIR for both output channels with GA-based BSS method.

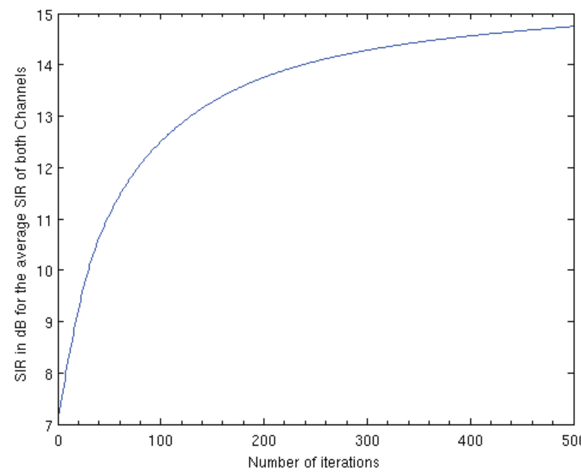

Figure 6: Average SIR for both output channels with standard BSS method.

TABLE I:
COMPARISON BETWEEN BSS METHODS.

Method	Initial SIR [dB]	Final SIR [dB]	Avg. time per epoch [s]
GA-based BSS	7.14	15.23	1.96
Std. BSS		14.75	33.13

V. CONCLUSIONS

A method for blind source separation is proposed using GA in order to separate the mixed speech signals effectively in a reverberating environment. Here, a fitness function considering the maximization of the Signal-to-Interference Ratio (SIR) is adopted. As showed in section 4, the method proposed here outperforms the standard one in both final SIR value and elapsed time per iteration.

During subjective auditory tests, it was noticed that the GA-based BSS algorithm proposed here produced output signals with metallic artifacts. But, although it produced such audible artifacts, additional analysis must be carried out in order to verify its possible use as an front-end tool for Automatic Speech Recognition (ASR) systems once the human auditory system possesses different perception properties of sounds than such systems.

Therefore, as for further research, the use of the GA-based BBS method proposed here as a preprocessing tool for ASR systems is to be studied. The proposed method seems to be able to improve the performance of such systems.

REFERENCES

[1] Feng M. and Kammeyer K., *Application of source separation algorithms for mobile communication environment.*1st Int. Workshop on ICA & Signal Separation, Aussois, France January 1999, pp. 431-436.

[2] Makeig S., Bell A., Jung T., Sejnowski T., *ICA in electroencephalographic data*. Advances in Neural Information Processing Systems, (8), MIT Press, 1996, pp. 145-151.

[3] Lee T., Bell A., Lambert R., *Blind separation of delayed and convolved sources.* Advances in Neural Information Processing Systems, (9), MIT Press, 1997, pp. 758-764.

[4] Girolami M., *Noise reduction and speech enhancement via temporal anti-Hebbian learning*. ICASSP, Vol 2, 1998, pp. 1233-1236.

[5] Choi S., Cichocki A., *Adaptive blind separation of speech signals: Cocktail Party Problem*. International Conference of Speech Processing, Seoul, Korea, 1997, pp. 617-622.

[6] Yen K., Huang J., Zhao Y., *Co-channel speech separation in the presence of correlated and uncorrelated noises.* EuroSpeech, Budapest, Sept. 1999, Vol. 6, pp. 2587-2590.

[7] Koutras A., Dermatas E., Kokkinakis G., *Recognizing simultaneous speakers: A genetic algorithm approach.* EuroSpeech, Budapest, Sept. 1999, Vol. 6, pp. 2551-2554.

[8] Yen K., Zhao Y., *Co-channel speech separation for robust Automatic Speech recognition.* ICASSP, Vol 2, 1997, pp.859-862.

[9] Amari S., Cichocki A., *Adaptive Blind Signal Processing- Neural Network Approaches.* IEEE Proceedings, Vol. 86, No 10, Oct 1998, pp. 2026-2048.

[10] Cardoso J., *Blind Signal Separation: Statistical Principles.* IEEE Proceedings, Vol. 86, No 10, Oct 1998, pp. 2009-2025.

[11] J. H. Rindel, *The use of computer modeling in room acoustics*. Journal of Vobroengineering, vol. 3, no. 4, pp. 41–72, 2000.

[12] L. Savioja, *Modeling techniques for virtual acoustics*. Ph.D. dissertation, Helsinki University of Technology, august 2000.

[13] D.E. Golberg, *Genetic Algorithms in Search, Optimization and Machine Learnning.* Addison-Wesley Pub (Sd) , 1989.

[14] H. Buchner, R. Aichner, W. Kellermann, *A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics.* IEEE Trans. Speech Audio Process. 13 (1) (January 2005) 120-134.

[15] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis.* Wiley & Sons, Inc., New York, 2001.

[16] M. Kawamoto A.K. Barros, A. Mansour, K. K. Matsuoka and N. Ohnishi, *Blind signal separatio for convolved non-stationary signals*. Trans. IEICE, vol. J82-A, no.8, pp.1320-1328, Aug. 1999.

[17] F. A. P. Figueiredo and C. A. Ynoguti, *On the improvement of the learning rate in Blind Source Separation using techniques from Artificial Neural Networks theory*. Proceedings of the International Workshop on Telecommunications – IWT/09.