Blind Source Separation in Noisy and Reverberating Environment Using Genetic Algorithm

Masato Katou and Kaoru Arakawa
Department of Computer Science, Meiji University, Kawasaki 214-8571 Kanagawa Japan
E-mail: karakawa@cs.meiji.ac.jp Tel: +81-44-934-7445

Abstract— **A method of blind source separation is proposed using genetic algorithm (GA) for separating mixed voices in noisy and reverberating environment. Generally, the performance of the blind source separation becomes degraded when random noise is added to the mixed voices. Moreover, the reverberating environment must be considered in separating the mixed voices actually obtained. In order to solve the problem of blind source separation in such circumstances, a method to utilize GA is proposed, in which the system parameters are represented as chromosomes and the correlation between the output voices is minimized using GA. Computer simulations show its high performance in separating voices influenced with additive noise and reverberation.**

I. INTRODUCTION

Recently, blind source separation has been actively researched for separating mixed multiple voices obtained from multiple microphones. Various methods have been proposed to realize the separation, but when actual mixed voices are to be separated, additive noise and reverberating environment must be considered. Some methods solve the problem of blind source separation in reverberating environment [1][2], but most of these methods do not consider the additive noise. Generally, the performance of blind source separation gets degraded when noise is added to the mixed voices, because the noise lowers the preciseness in evaluating the independency between different voices.

In this paper, a new method of blind source separation using a genetic algorithm (GA)[3] is proposed in order to effectively separate mixed voices in noisy and reverberating environment.

GA is a search technique to find 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 finds out the optimum solution for the system parameters, on the basis of a certain evaluation function using evolutionary process. The chromosomes also correspond to individuals in a population. This method can realize powerful optimization for the system parameters.

Using GA in the blind source separation, the problem of the affect by additive noise can be solved. In order to consider the reverberation also, GA is introduced into the blind separation system for temporally and spatially mixed voices. The separating system is composed of nonrecursive 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 degree of independence between the output voices be as large as possible.

In this paper, first the system model of the blind source separation in noisy and reverberating environment is shown and then the proposed method introducing GA into the blind source separation is presented. Finally in computer simulations, the performance of the proposed system is shown to be effective for actual voices which are mixed both temporally and spatially, contaminated with additive noise. It is also shown that the conventional method considering only reverberating environment is not effective for noisy input, even if noise reduction is performed before separation.

II. MODEL OF BLIND SOURCE SEPARATION UNDER NOISY AND REVERBERATING ENVIRONMENT

Two microphones located apart are considered here to obtain voices from two sources. Here some noise is supposed to be mixed to each microphone. The sound acquisition system is depicted as shown in Fig.1. In actual environment, some reverberation occurs, also. The signal obtained from the microphones is expressed as the following equation.

$$x_i(n) = \sum_{j=1}^{2}\sum_{k=0}^{K} a_{ij}(k)s_j(n-k) + u_i(n) \tag{1}$$

Here, $x_i(n)$ denotes the mixed signal obtained from the i-th microphone at time n where i is equal to 1 or 2, $s_i(n)$ the j-th

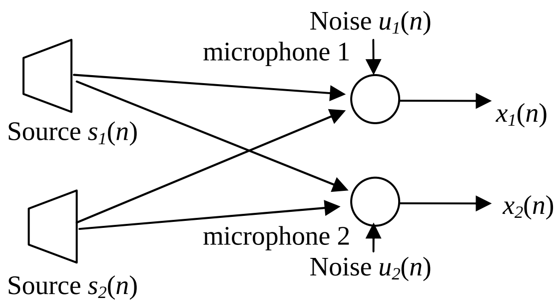


Fig.1 The system model of sound acquisition.

source signal, and $u_i(n)$ the additive random noise to the i-th input. The reverberation is supposed to be expressed as a linear convolutive system express with $a_{ij}(k)$.

The i-th source signal is supposed to be obtained as $y_i(n)$ from $x_i(n-L)$ and the sequence of $x_j(n-k)$ (k=0, 1, …, M) using a linear FIR filter as follows.

$$y_i(n) = x_i(n-L) + \sum_{\substack{j=1 \\ j \neq i}}^{2} \sum_{k=0}^{M} b_{ij}(k) x_j(n-k) \quad (2)$$

$$(i = 1, 2; \quad 0 \leq L < M)$$

Here, L denotes the time lag of this separating system.

The problem of the blind source separation is to determine the values of $b_{ij}(k)$'s, so that the output signals $y_1(n)$ and $y_2(n)$ are independent.

Several methods have been proposed to solve this problem. Typically, they adopt an evaluation function which represents the degree of the independency between $y_1(n)$ and $y_2(n)$, and minimize it with a certain method, such as a gradient method. Kawamoto et al. showed that multiple voices mixed in reverberating environment, which does not include additive noise, can be separated with a gradient method minimizing the following equation[1].

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

Where $y(n)=(y_1(n), y_2(n))^T$, and N is set at 2 for this case. When $y_1(n-L)$ and $y_2(n-L)$ are independent, the value Q is close to zero.

III. Introduction of Genetic Algorithm into Blind Source Separation

A. Principle of Genetic Algorithm

A genetic algorithm (GA) is a technique to obtain the optimal or suboptimal solution of a system by maximizing a certain evaluation function named as a 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 system parameter. Then the fitness function of these individuals is evaluated and those with higher fitness function 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 for all individuals, selection, and reproduction are performed 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. This procedure is show in Fig.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 represents the statistical independency as the fitness function. Here, an individual is generated as a sequence of the filter coefficients $b_{ij}(k)$ in (2) as Fig.3. As to the evaluation function, the following value C is considered.

$$C = \sum_{m=-M'}^{M'} corr(y_1(n), y_2(n-m)) \quad (4)$$

Here, $corr(y_1(n), y_2(n-m))$ denotes the correlation coefficient between $y_1(n)$ and $y_2(n)$ as

$$corr(y_1(n), y_2(n-m)) = \frac{E[(y_1(n)-\bar{y}_1)(y_2(n-m)-\bar{y}_2)]}{\sigma_1\sigma_2} \quad (5)$$

Where E denotes averaging, and $\bar{y}_1$, $\bar{y}_2$ are the means, σ_1, σ_2 are the standard deviations of $y_1(n)$ and $y_2(n)$ respectively. The correlation coefficient with time delay m is considered, because $x_j(n)$ may be added to $x_i(n)$ with some time delay up to M in (2) and the undesired voice may remain in $y_i(n)$ if only the value of $corr(y_1(n), y_2(n))$ is considered in the evaluation function. Unlike the fitness function, the evaluation function (4) is to be minimized for optimization. The new generation is generated so that the value C should be smaller.

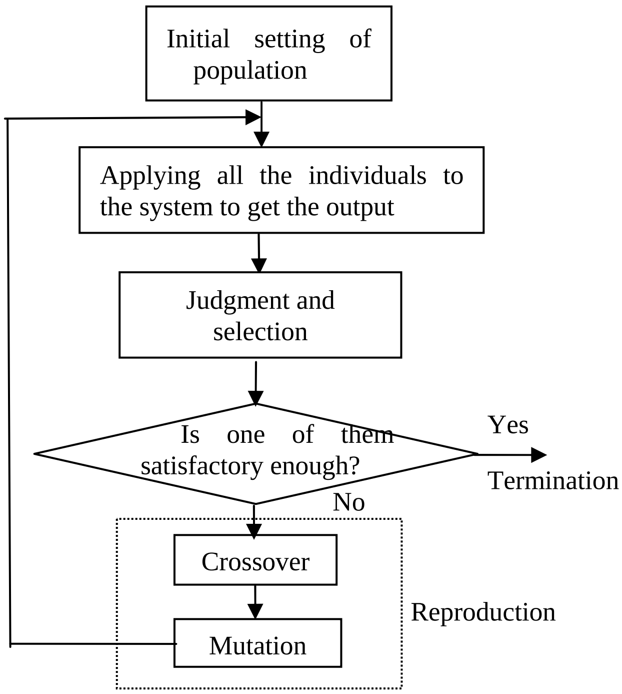


Fig.2 The procedure of the genetic algorithm

$b_{12}(0)$	$b_{12}(1)$	…..	$b_{12}(M)$	$b_{21}(0)$	$b_{21}(1)$	….	$b_{21}(M)$

Fig.3 The structure of an individual.

There are some variants in GA, but here real- valued GA is adopted for correspondence with signal processing. The procedure of this GA is as follows.

(a) Initial setting

P individuals as shown in Fig.3 are generated randomly to make a population.

(b) Selection

Signal processing as (2) is performed for the input mixed voices $x_1(n)$ and $x_2(n)$ using the *P* sets of parameters $b_{12}(k)$ and $b_{21}(k)$. Accordingly *P* output pairs $y_1(n)$ and $y_2(n)$ are obtained. These output pairs are evaluated with the value *C*, and *R* of the individuals which give smaller *C* are selected as survivors to the next generation.

(c) Crossover

S pairs of the survived *R* individuals are randomly selected and crossover is performed for them. Here, uniform crossover as shown in Fig. 4 is adopted in order to mix the characteristics of the parents adequately. In the uniform crossover, each $b_{ij}(k)$ in the individual of children takes the value of the corresponding $b_{ij}(k)$ in either of the parents at probability of 50%. In this stage, 2*S* individuals are newly generated.

(d) Mutation

(*P*-*R*-2*S*) individuals are newly generated by mutation. Totally *P* individuals are prepared for the next generation. There are several methods in mutation, but here in order to avoid the trap of the plateau of convergence, every $b_{ij}(k)$ is to be changed randomly. Here, a random value in a certain range is added to every $b_{ij}(k)$.

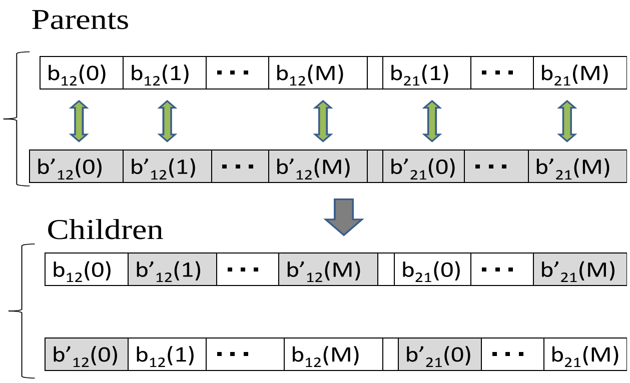


Fig. 4 An example of the uniform crossover.

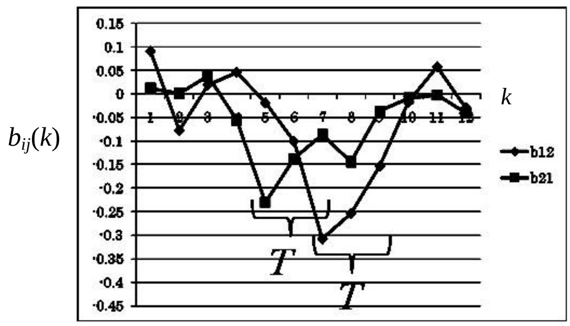


Fig.5 An example of the sequence of $b_{ij}(k)$ obtained during GA process.

(e) Termination

The new population is applied to signal processing (2) and the value *C* is evaluated. If the value *C* of an individual is less than a certain threshold, the individual is obtained as the solution and the procedure is terminated. The signal separation is performed by (2) using the filter coefficients $b_{ij}(k)$ obtained as the solution.

IV. Computer Simulation

A. Parameter Setting in Mixture Model and GA

11kHz sampled 16-bit male voice and female voice are used as source signals $s_i(n)$. They are mixed with a spatiotemporal matrix and noise is added as follows.

$$\boldsymbol{X}(z) = \overline{A}(z)\,\boldsymbol{S}(z) + \boldsymbol{U}(z)$$

$$\overline{A}(z) = \begin{bmatrix} z^{-1} + 0.3z^{-2} & 0.4 + 0.4z^{-1} + 0.3z^{-2} \\ 0.4 + 0.3z^{-1} + 0.2z^{-2} & 1 + 0.4z^{-1} \end{bmatrix} \qquad (6)$$

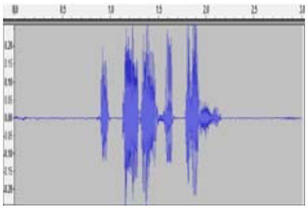
where $\boldsymbol{X}(z)=(X_1(z), X_2(z))^T$, $\boldsymbol{S}(z)=(S_1(z), S_2(z))^T$, $\boldsymbol{U}(z)=(U_1(z), U_2(z))^T$, and $X_i(z)$, $S_i(z)$ and $U_i(z)$ ($i=1,2$) are z-transform of $x_i(n)$, $s_i(n)$, and $u_i(n)$ respectively. $u_i(n)$ is a white gaussian noise with SNR 6.25. In (2) *M* and *L* are set at 11 and 6 respectively and *M'* in (4) is set at 11.

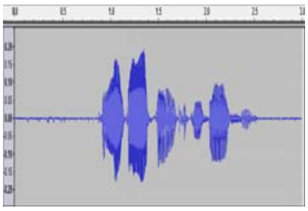
In GA, *P*, *R*, *S* are set at 100, 10, and 20, respectively. The initial values of $b_{ij}(k)$'s in individuals are set close to those obtained by Kawamoto's method[1] minimizing (3) for the input signal (6). Quasi-gaussian white noise ranged from -0.1 to 0.1 is added to $b_{ij}(k)$'s obtained by Kawamoto's method, and 100 individuals are generated to make the initial population.

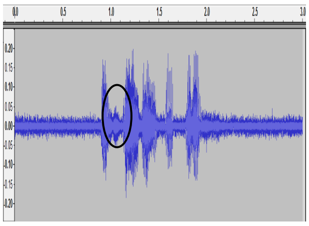
As is shown in Fig.8 in the next section, when the input signal contains additive noise, the values $b_{ij}(k)$'s obtained by Kawamoto's method have less variation, do not have steep gorge, compared with those when the input signal does not contain noise. Thus, in order to realize fast and stable convergence, the change caused by mutation is biased to be negative and the variance of the change is set larger for the middle part of *k*. Moreover, since the reverberating sound enters the microphone later than the direct sound, *k* with the minimum $b_{ij}(k)$ is searched as *k'*, and the amount of the change caused by mutation is set larger at $k=k', k'+1, \ldots k'+K_0$. For example, suppose that the values $b_{ij}(k)$'s are obtained as shown in Fig. 5, quasi-gaussian white noise ranged from -0.3 to 0.1 is added in the area of *T*, while that from -0.1 to 0.1 is added in the other area of *k*. Here, K_0 is set at 2.

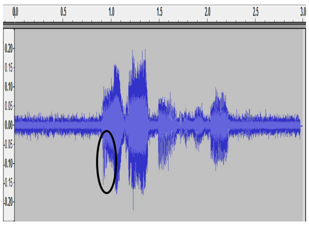
B. Results of Computer Simulations

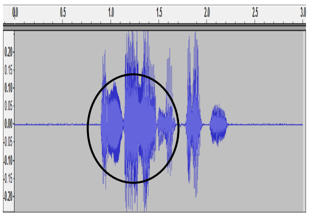
Fig. 6(a)(b) show the source voice signals $s_i(n)$ for i=1, 2 respectively. If noise $u_i(n)$ is not added, the conventional method of blind source separation, such that proposed by Kawamoto, is effective enough and the mixed voices are almost completely separated. However, if noise is added, this

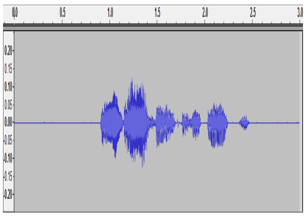

(a) Source voice signal #1: $s_1(n)$.

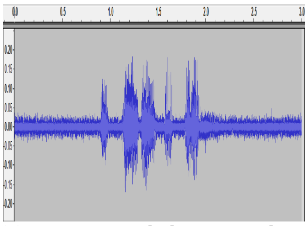

(b) Source voice signal #2: $s_2(n)$.


(c) Output #1: $y_1(n)$ with the conventional method.


(d) Output #2: $y_2(n)$ with the conventional method.


(e)Output #1 with the conventional method after noise reduction is applied.


(f) Output #2 with the conventional method after noise reduction is applied.


(g) Output #1 with the proposed method using GA..

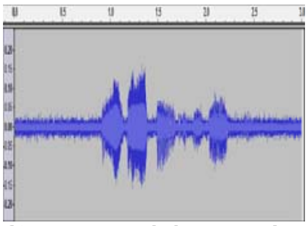
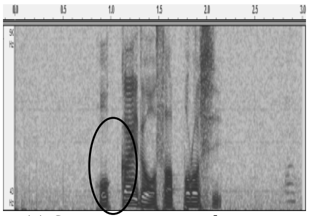
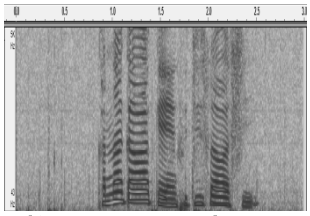
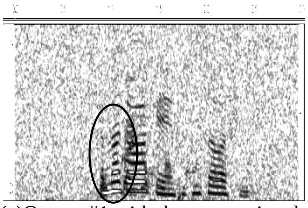
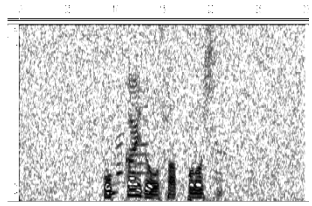

(h) Output #2 with the proposed method using GA.

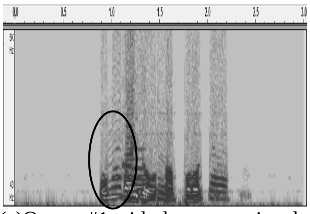
Fig.6 Waveforms of the source and output voice signals.

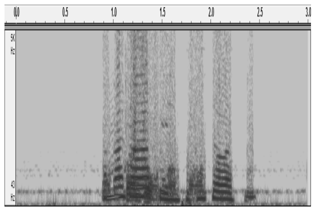

(a) Source voice signal #1

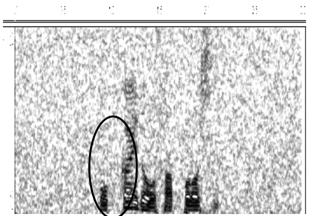

(b) Source voice signal #2


(c)Output #1 with the conventional method.


(d) Output #2 with the conventional method.


(e)Output #1 with the conventional method after noise reduction is applied.


(f) Output #2 with the conventional method after noise reduction is applied.


(g) Output #1 with the proposed method using GA.

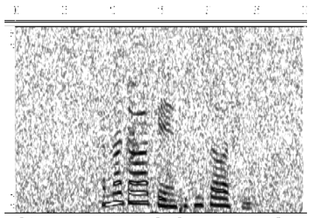

(h)Output #2 with the proposed method using GA.

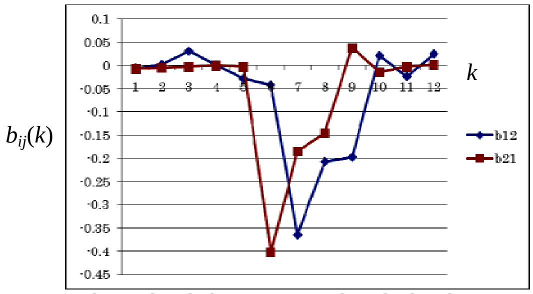
Fig. 7 Spectrograms of the source and the output voice signals.

method is not effective any more. Fig. 6(c)(d) show the separated signals with this method. The circled part indicates the remaining mixed voice. Fig.7(a)(b) show the spectrograms of the original voices and (c)(d) those of the separated ones by this method. We can see that some part of the voice #2 remains in the output #1 in the circled area. Fig. 8 shows the obtained filter coefficients $b_{ij}(k)$'s. (a) shows those obtained by Kawamoto's method when noise is not added, and (b) those when noise is added. We can see that the values of $b_{ij}(k)$'s obtained with this method when noise is added are quite different from those when noise is not added.

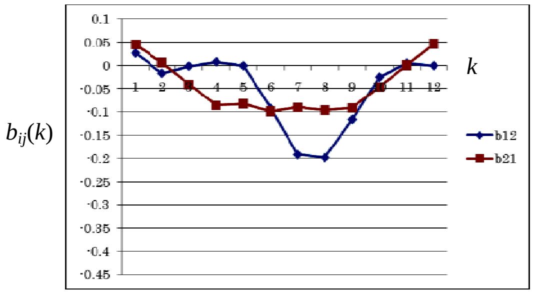
For comparison, Fig. 6(e)(f) show the outputs of blind source separation by Kawamoto's method, when the noisy inputs are preprocessed with a noise reduction filter. Here, a noise reduction using spectral subtraction[4] and a nonlinear smoothing filter[5] is adopted. Fig.7 (e)(f) show the spectrogram of these outputs.

We can see that the noise is clearly removed, but the signal separation is not performed well. Finally, the performance of the proposed method with GA for noisy input is shown in Fig.6(g)(h). The signal component of the other voice does not appear in this case. As is shown in Fig.7(g)(h), the component of the source voice #2 does not appear in the output #1 in the circled area. By auditory test, the source voices are verified to be not mixed in the output #1 and #2. Fig. 8(c) shows the values of $b_{ij}(k)$ obtained by this method. We can see that the shape of this function $b_{ij}(k)$ in (c) is close to that in (a), thus the performance of the blind source separation with GA is close to that when noise is not added.

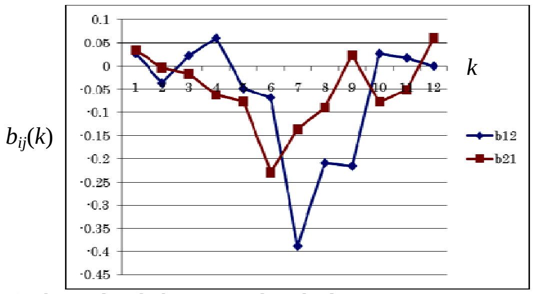
Table 1 shows the final value of C for evaluation in the above four cases. We can see that the proposed method with GA realizes as low evaluation function as that without noise. As to the computational time of the proposed method, it takes



(a) Obtained with the conventional method without noise.



(b) Obtained with the conventional method in noisy environment.



(c) Obtained with the proposed method using GA in noisy environment.

Fig. 8 Filter coefficients $b_{ij}(k)$ obtained by three methods.

TABLE I
THE VALUES OF *C* OBTAINED IN VARIOUS CASES

	Conventional method without noise	Conventional method in noisy environment	Conventional method with noise reduction in noisy environment	Proposed method using GA in noisy environment
The value C	0.134	2.019	1.813	0.140

about 1.5 minutes to get the solution with Intel Core 2 Duo 2.13GHz processor.

V. CONCLUSIONS

A method for blind source separation is proposed using GA in order to separate the mixed voice effectively in noisy and reverberating environment. Generally, when random noise is added to the mixed voice, conventional methods for blind source separation based on a gradient method cannot separate the mixed voice effectively. However, using GA, powerful optimization for the evaluation function is realized and the mixed voices are effectively separated. Here, an evaluation function considering the correlation with time delay is adopted. When the evaluation function is too complicated, conventional gradient method is difficult to be applied, but GA can be applied such complicated evaluation function easily.

The quality of separation is also verified to be high enough in subjective auditory test.

As for further research, the cases with mixture of additive colored noise and mixture of noise from the same noise source are to be considered. The method for fast convergence is also to be developed.

REFERENCES

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

[2] R. Aichner, H. Buchner, S. Araki, and S. Makino, “On-line time –domain blind source separation of nonstationary convolved signals”, Proc. Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA), pp.987-992, April, 2003

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

[4] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” *IEEE Trans. Acoust., Speech Signal Processing*, vol. 27, no. 2, pp. 113–120, April 1979.

[5] K. Arakawa, K. Matsuura, H. Watabe, and Y. Arakawa, “A method of noise reduction for speech signals using component separating ε-filter”, Trans. IEICE, vol. J85-A, no. 10, pp.1059-1069, Oct. 2002.