User Interface of an Interactive Evolutionary Computation for Speech Processing

Todoroki, Yukichi
Graduate School, Kyushu Institute of Design

高木, 英行
Department of Acoustic Design, Kyushu Institute of Design

https://hdl.handle.net/2324/4486278

出版情報：6th International Conference on Soft Computing (IIZUKA2000), pp.112-118, 2000-10-01. ファジィシステム研究所
バージョン：
権利関係：

User Interface of an Interactive Evolutionary Computation for Speech Processing

Yukichi Todoroki* and Hideyuki Takagi**
Kyushu Institute of Design, *Graduate School, **Dept. of Art and Information Design
4-9-1, Shiobaru, Minami-ku, Fukuoka 815-8540, Japan
Phone & FAX: +81-92-553-4555, E-mail: takagi@kyushu-id.ac.jp

Keywords: interactive evolutionary computation, user interface, user fatigue, speech processing

Abstract— In this paper, we propose and evaluate three methods to improve interactive evolutionary computation (IEC) by decreasing user fatigue when an IEC displays sequential outputs for tasks such as music creation, speech processing, or sound enhancement. The first method decreases human fatigue by using multiple sound sources. We evaluate two systems based on this method: a system that assigns a different sound source to each individual and a system that assigns a different sound source to each generation. Subjective tests have shown that the first system is not as effective as the second system, which significantly decreases fatigue. The second method displays the previous best individual as a reference to compare subsequent individuals. Subjective tests have shown that this method significantly improves the IEC operability and has a tendency to improve the IEC convergence. The third method sequentially displays individuals that cannot be spatially compared. Subjective tests have shown that this method significantly improves the IEC operability.

1 Introduction

Interactive Evolutionary Computation (IEC) is an optimization method based on human subjective evaluation and has been applied to various tasks [9].

A common problem with these tasks is IEC user fatigue. IEC users must repeatedly observe and evaluate system outputs that match their mental image of a graphic or music. Due to this repetitive operation, they become both mentally and physically tired. When IEC tasks optimize sound output systems, such as music creation or speech processing, IEC individuals cannot be simultaneously displayed, making it difficult to compare sequential outputs. The fatigue problem for these particular tasks becomes more apparent than for other IEC tasks.

A practical use for IEC technology has been requested. To solve the fatigue problem, there have been several research proposals [9]. Some IEC interface improvements have included the implementation of the discrete fitness value input method [5], the use of display methods based on predicting human evaluation [7, 6], the avoidance of predictable non-musical melodies [1], the selection of the best N individuals from many individuals using a prediction function for human evaluation [4], the acceleration of genetic algorithm (GA) convergence for an interactive GA [3], the combination of an IGA with a normal GA, embedding on-line knowledge [2, 10], and Visualized IEC [11].

We propose and evaluate three improved IEC interfaces that can sequentially display speech, sound, or music. We apply these IEC interfaces to the task that adjusts the filter parameters for speech processing and evaluate how they reduce human fatigue with subjective tests. The three improved interfaces discussed in this paper are:

- method changing sound source to relieve the monotony of evaluation,
- method displaying the past elite to make it easy to compare with new individuals, and
- method displaying individuals in series.

The first method uses multiple sound sources to maintain IEC user interest during an evaluation, while the conventional IEC uses the same sound source in favor of a precise comparison rather than usability. The second method displays the best candidate from previous generations to let the IEC user easily compare and evaluate individuals with the displayed reference. The third method displays only one individual per window, while a conventional IEC usually displays all individuals. Although it is difficult for the IEC user to compare the current individual with undisplayed individuals, the number of operations to select an individual can be reduced, which may decrease human fatigue.

We describe the experimental conditions commonly used in subjective tests for three IEC interfaces in section 2, the evaluations of three methods in sections 3, 4 and 5, and the discussion and conclusion in sections 6 and 7.

2 The Conditions of Experimental Evaluation

2.1 IEC speech processing system

The IEC speech processing system [12] used in our experimental evaluation is shown in Figure 1.

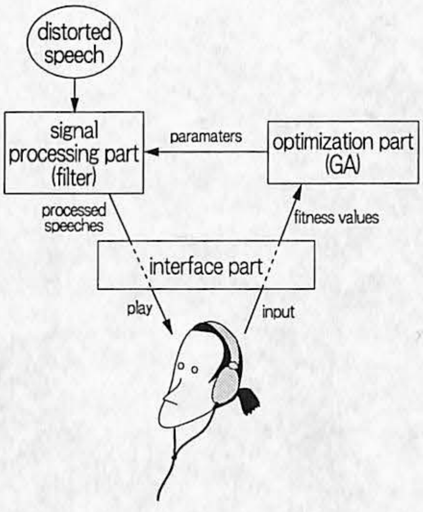

Figure 1: The system diagram of an IEC speech processing system in our experimental evaluations.

In the optimization part, the genetic algorithm (GA) optimizes the filter parameters and transfers them to the signal processing part. The signal processing part determines characteristics of filters based on the transfered parameters and processes distorted speeches using the filters. The interface part displays the processed speeches to the user. The user evaluates the quality of displayed speech, inputs a fitness value for each speech and sends it to the GA part through the interface part. We show the interface in Figure 2.

Figure 3 illustrates the process in the signal processing part. The parameters that the signal processing part receives from GA are amplification levels of frequency bands. The processing part calculates amplification rates from the levels and smoothes the overall frequency range, which provides us with the frequency characteristics of a signal processing filter. Input speeches are processed by the filter window frame by frame, and the processed speech is displayed to an IEC user. All distorted speeches are initially prepared in same way as this processing.

2.2 Subjective tests

Two systems based on different methods are compared and evaluated with subjective tests. The procedure of the subjective tests used in section 3 is shown in Figure 4. Each user operates and evaluates the first system, then operates and evaluates the second system, and finally evaluates a pair of both systems.

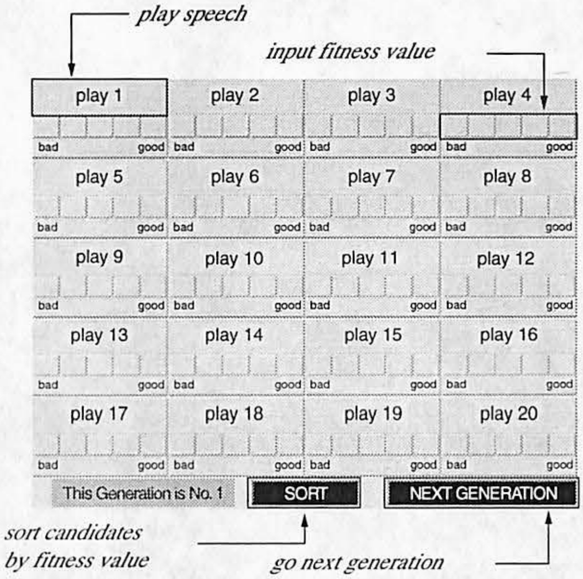

Figure 2: User interface of experimental IEC systems

Users operate the IEC speech processing system mentioned in section 2.1 with three different interfaces. They evaluate each individual and give fitness values in five levels ranging from *hard to hear* to *easy to hear*. Users are instructed to assign the minimum and maximum fitness values to the worst and best individuals, respectively, for each generation except the evaluation in section 5.

Subjects are normal hearing males and females in their early twenties. Speech data used in our experiments are Japanese sentences selected from an audio CD for hearing aid fitting, taking into account the variety of vowels and consonants. The length of speech data ranges from 2 to 2.5 seconds.

The population size, the number of generations, and the number of genes of the GA used in our experiments are 20, 5, and 6, respectively. The fitness values are the subjective values of human evaluation according to the five-grade system. Other experimental parameters are shown in Table 1. The GA searches the amplification levels of six frequency bands of a speech processing filter. The central frequencies of the bands are 125Hz, 250Hz, 500Hz, 1kHz, 2kHz, and 4kHz. The range of each amplification level is $[-32,32]$ dB and is encoded into 8 bits; the length of a chromosome is $8 \times 6 = 48$ bits.

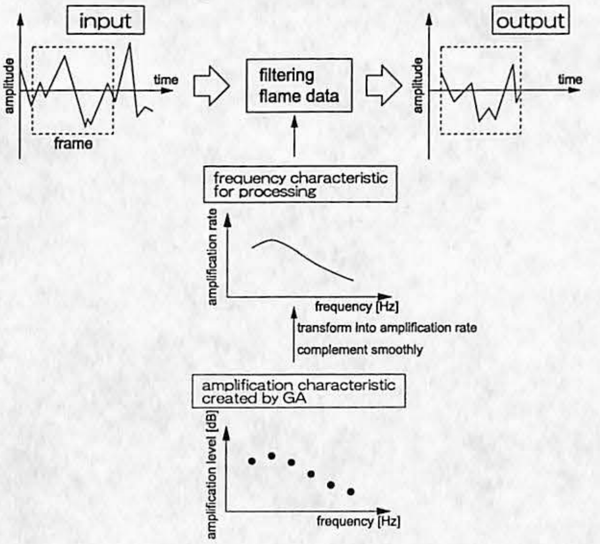

Figure 3: Process of speech filtering in our system.

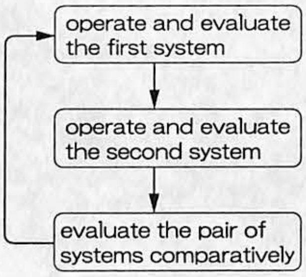

Figure 4: Procedure of experimental evaluation.

3 Evaluation of the IEC displaying multiple speech sources

3.1 The IEC displaying multiple speech sources

Conventional IEC for signal processing develops filters and applies them to prepared images or sounds. The IEC user evaluates the processed images or sounds that are variations of the original source, which often bores and tires human operators. Previous IEC-based speech processing systems [12] also use the same distorted speech. The user listens to only one speech many times until a satisfactory quality of speech is obtained.

The first method uses different kinds of sound sources to decrease the user's fatigue with variations of the source when the IEC is used for image or sound processing.

3.2 Experimental conditions

The IEC with multiple speech sources is evaluated with subjective tests. Three systems with different interfaces compared with the subjective tests are:

Table 1: GA parameters

selection	roulette wheel selection with the elitist strategy
crossover	simplex crossover
crossover rate	0.95 (The19 of 20 individuals are operated.)
mutation rate	0.2% for the half of all individuals and 20% for others except an elite

System #1 assigning only one speech to all individuals in all generations,

System #2 assigning different kind of speech sources to each individual, and

System #3 assigning different kind of speech sources to each generation

We form six pairs from these three systems by considering the order effect. Eighteen subjects are divided into three groups as in Table 2 and operate all these pairs; all subjects do not operate in the same order. Each subject evaluates two pairs a day with a sufficient break between the two pairs, completing all six pairs in three days.

Table 2: Subjects group: the numbers in the table corresponds to system #1, 2, and 3

Group A	1 − 2	3 − 1	2 − 3
	1 − 3	2 − 1	3 − 2
Group B	2 − 3	1 − 2	3 − 1
	2 − 1	3 − 2	1 − 3
Group C	3 − 1	2 − 3	1 − 2
	3 − 2	1 − 3	2 − 1

To evaluate these system pairs under the same conditions, the same random generator seed is used for GA initialization of each pair for the first generation; individuals in the first generation of two systems are same. Except for the seed of GA initialization, randomly selected seeds are used for random values in GA operation and speech source assignment. We use only one speech source in System #1, 20 various kinds of speech sources for 20 individuals in System #2, and 5 various kinds of speech sources for 5 generations in System #3.

The distorted speech sources that are used as the input signal for the experimental systems are made in the same way that the filters are made by IEC process speeches. Table 3 shows the filter coefficients to make the distorted speeches.

Each system and comparison of pairs are evaluated and analyzed with the subjective tests of the method of successive categories and Sheffé's method of paired comparisons, respectively.

Subjects are requested to evaluate three systems on convergence and operability, and their answers to the

Table 3: Filter coefficients to made distorted speech sources.

frequency [Hz]	amplification level [dB]
125	-30.0
250	-30.0
500	-15.0
1000	-15.0
2000	-10.0
4000	-5.0

two pairs are analyzed by the two subjective test methods. Subjects are requested to subjectively evaluate the systems in five levels ranging from -2 to 2, where -2 means *bad* or *worse*, and 2 means *good* or *better*. Questions subjects responded to were:
(1) Evaluation of the convergence; *How is the quality of the best speech in the last generation ?*
(2) Evaluation of the operability; *How easy and how tiring is it to operate this system ?*

3.3 Evaluation results

Subjective evaluation of convergence by the method of successive categories is shown in Figure 5. Systems #1 and 2 belong to different categories, and Systems #1 and 3 belong to the same category in the figure. The Sheffé's method of paired comparisons shows that the psychological difference between Systems #1 and 2 is significant ($p < 0.01$), and the difference between Systems #1 and 3 is not significant ($p > 0.05$).

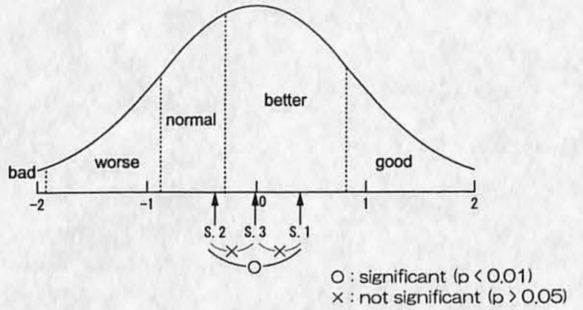

Figure 5: Evaluation on convergence by the method of successive categories. S.1, S.2, and S.3 mean Systems #1, 2, and 3 defined in section 3.2.

These results have shown that IEC systems with only one speech source, a conventional approach, converge faster than IEC systems in which each individual is assigned a different speech source. There is no significance between IEC systems with different speech sources assigned for each generation and System #1 or 2.

Subjective evaluation on operability by the method of successive categories is shown in Figure 6. Systems #1 and 2 belong to the different category to which System#3 belongs in the figure. The Sheffé's method of paired comparisons shows that the psychological difference between System #3 and System #1 or 2 is significant ($p < 0.01$), while there is no significance between Systems #1 and 2 ($p > 0.05$).

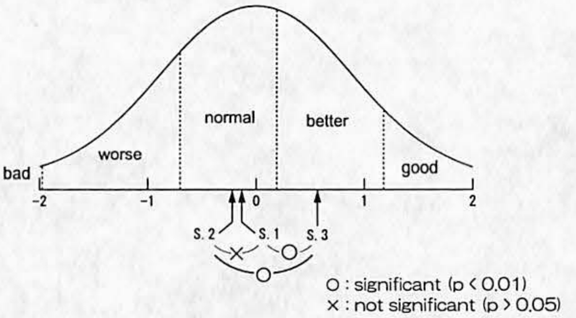

Figure 6: Evaluation on operability by the method of successive categories. S.1, S.2, and S.3 mean Systems #1, 2, and 3 defined in section 3.2.

These results have shown that IEC systems with different speech sources for each generation is easier to operate than IEC systems without changing speech source and with changing speech sources for every individual.

4 Evaluation of the method for displaying the elite

4.1 Method for displaying the elite

IEC users must evaluate similarly displayed stimuli many times. After many evaluations over a number of generations, it becomes difficult for users to judge as to what is good and what is bad. The users lose a basis for their evaluation and begin to evaluate individuals with an imprecise scale, resulting in the system's poor convergence and operability.

To solve this problem, we designed a system to display the best individual from past generations as a standard reference for the users to base their evaluations. The interface system in this section displays the elite individual in the previous generation to the users. The users can then evaluate individuals by comparing them to the elite. We expect that the users can more easily compare the candidates by using the standard.

In this section, using IEC speech processing system, we perform subjective tests on the interface system that displays the elite individual from the previous generation and verify the system's validity.

4.2 Experimental evaluation

We compare the proposed method with one that does not display the elite individual. In the system constructed with the proposed method, we can know which speech is selected as the elite with the interface. The elite has the best fitness value in the previous generation and is left for the next generation. We cannot determine the elite individual in the system that is made by the previous method. When a subject operates the system of proposed method, the subject evaluates the individuals by comparing them with the elite individual after the first generation.

We construct a pair of systems with the proposed method and a previous one and divide the subjects into two groups according to their operating order. The number of subjects is 15.

All experimental conditions follow the ones described for System #1 in section 3.2 except for the system of the proposed method that displays the elite individual. The procedure for the subjective test is the same. However, we analyze the results of the comparative evaluation of the two systems, not by Sheffé's method of paired comparisons, but by the sign test.

4.3 Evaluation results

We show the results of our subjective test in Table 4. With respect to convergence, we omit the answers of 4 of the 15 subjects off the test because their responses were neither good nor bad. Nine of the 11 subjects sensed that the system that displayed the elite individual had converged faster than the system that did not display the elite individual. By applying the sign test, we found that the difference between the two systems was not statistically significant, but it would be significant at 6.7% level.

As to the operability, we omitted the answers of 3 of the 15 subjects because their answers were neither good nor bad. Eleven of the remaining 12 subjects answered that they had sensed the system of the proposed method to be easier to operate and less tiring than the system of past method. By applying the sign test, we found that the difference between two systems to be statistically significant.

Table 4: Results of evaluation of the interface displaying the elite: answers to the question which system subjects have sensed is better.

answers	the number of answers	
"Which is better?"	convergence	operability
proposed method	9	11
past method	2	1

convergence: not significant ($p > 0.05$)
operability: significant ($p < 0.01$)

5 Evaluating the method for displaying the sound source in series

5.1 Method for displaying the sound source in series

To easily compare individuals, the previous interface of an IEC speech processing system displays all the individuals from one generation at once. Users tire from repeatedly moving a cursor to select the next evaluation individual. We then considered that the interface displayed only one candidate on the screen at a time. By doing so, we can minimize the motion required to move the cursor. The user cannot mutually compare individuals, however, the user may roughly evaluate each individual. The operation goes so smoothly that it is expected that the operability increases.

In this section, we perform a subjective test of the interface system of the method for displaying a sound source in series and verify its validity.

5.2 Experimental evaluation

We compare the proposed method with a previous one. The system construct from the proposed method shows one play button and one set of fitness value input buttons for one speech at a time. When a user clicks a fitness value input button, evaluation moves to the next individual and the next speech is automatically played. It is possible to move between the individuals. For example, we can go back to a previous evaluation in the case of making a mistake by inputting an incorrect fitness value. If it is difficult to input value for a candidate, we can defer its evaluation and skip to the next. The user pushes the 'Back' and 'Forward' buttons for moving between individuals. The system of the previous method displays all of the sound sources of a generation at once as shown in Figure 3.

The procedure for this experiment is almost identical to the one described in section 4. However, there is a small difference in the method of evaluation; questions after operating the systems are limited only to operability.

- How easy is the system to operate?
- How tiring is the system?

The number of subjects we used is 15.

In this experiment, we measure the times that the subjects take to operate each system apart from our subjective tests. We analyze the results by statistically testing for significant differences between operating times by the Wilcoxon-Mann-Whitney test [8].

Experimental conditions follow the ones of the experiment described in section 4 except for the system of the proposed method that uses the method for displaying the sound source in series.

5.3 Evaluation results

The results of our subjective test are shown in Table 5. As to the ease of operation, 9 of the 10 subjects whose answers have been used for the test, answered that they sensed the system of proposed method was easier to operate than the system of the conventional method. By applying the sign test, we could find that the difference between two systems was statistically significant.

As to the difficulty to tire, we omitted the answers of 1 of the 10 subjects from the test because their answer was neither good nor bad. All subjects whose answers were used for the test had answered that they sensed the system of the proposed method was less tiring than the system of the conventional method. By applying the sign test, we found the difference between the two systems to be statistically significant.

Table 5: Results of the evaluation of the interface displaying a sound source in series: answers to the question which system subjects sensed was better

answers "Which is better?"	the number of answers	
	easiness to operate	hardness to get tired
proposed method	9	9
conventional method	1	0

easiness to operate: significant ($p < 0.05$)
operability: significant ($p < 0.01$)

Table 6 shows the mean operating times to the 5th generation for each system. By applying the Wilcoxon-Mann-Whitney test, we found that the operating times for proposed method were significantly shorter than for the conventional method.

Table 6: Mean operating times up to 5th generation

	proposal method	past method
minutes	7.08	9.38

significant ($p < 0.05$)

6 Discussion

Observing these experimental results in sections 3, 4, and 5, we may obtain four hints to reduce user fatigue: (1) the easier comparison of individuals, the less user fatigue, (2) refreshment for monotonous evaluations is useful under the condition of the (1), (3) the less number of operations, the less user fatigue, and (4) the interface that causes less obsession causes less user fatigue.

The ease of comparison has an serious effect on user fatigue. the System #2 that assigns different speeches to each individual had poor performance on both convergence and operability than the System #1 that uses only one kind of speech and the System #3 that does not change speech source in same generation. Also, the system that displays the elite individual showed higher operability and possible faster convergence at 6.7% level. Common factor that reduce user fatigue between these two results in section 3 and 4 is how the IEC interface provides easier comparison of individuals.

By decreasing user's weariness, we can improve operability. Refreshment for monotonous evaluations increase operability, but this hint should be under the aforementioned condition of the easier comparison. From the comparison of the Systems #1 and 3, a system changing speech sources in each generation resulted higher operability than a system using same speech source. This result leads us to the hint that refreshment for monotonous evaluation reduce human fatigue. Actually, many subjects reported that they would have got bored anyway. However, this hint must be valid only when the first hint, which is easier comparison is important, is satisfied from the fact that too much change of speech sources showed less operability.

It is better that the number of user operations is small. Fine operability of the method that displays only one individual in a series and has no operation of moving a cursor to select the next individual must have resulted that less operation leads higher operability of systems. In the method for displaying the sound source in a series, a user only must basically click fitness value input buttons. User evaluates each individual without attention to the individuals evaluated once. The result from reducing the operating time shows this fact also.

We can decrease user fatigue by reducing a mental compulsion. In the method for displaying the sound source in a series, it is very difficult to mutually compare individuals. From the users' perspective, all that user has to do is to pay attention to the displayed speech without concern about mutual evaluation. By paying no attention to other individuals, the user may feel less compulsion to take a second look at his or her evaluation by comparing individuals. In addition, subjects have said that they did not feel oppressed by being displayed sound sources at once, and that they could comfortably operate at pleasant pace. However, we must consider the problem that this interface may cause unreliable evaluation because of less adjustment of the first evaluation. The unreliable evaluation may causes poorer convergence.

It is important to comfort the user also. In the experiment in which an interface displays the elite individual, subjects reported that they felt relieved because they could confirm what they had thought was good. We consider the fact that there has been standard for evaluation which led to user's security. As to the system displaying the sound source in series, subjects felt relieved because of existence of 'Back' and 'Forward' buttons. Before conducting a formal experiment, we conducted preliminary tests without these buttons. In the preliminary tests, subjects reported that they became concerned because they could not return to a previous evaluation when they had made a mistake in inputting fitness value. In our formal experiment, however, subjects did not seem to use these button as expected. Some subjects said they felt secure from making a mistake with the presence of these buttons.

7 Conclusion

In this paper, we evaluated three methods to decrease user fatigue for the sound processing system using an interactive evolutionary computation. We verified their validity using a speech processing system with subjective tests.

The first method uses various kinds of speeches. By using the system that changed the kind of speech in every generation, we could improve operability significantly by ensuring that the convergence was not lower than previous methods. As to the system that changed the kinds of speech for each individual, neither convergence nor operability has been better.

The second method displays the elite of previous generation. This improved convergence and operability significantly.

The third method displays the sound source in a series. We have found this system improved operability significantly.

As to the design of IEC interface, it is important to make convergence and operability compatible with each other. If we are to raise the precision of a search, user fatigue will increase. However, if we give a priority to the operability and design of the interface so that the user feels comfortable, the precision of a search will decline. It is important to maintain a high level of precision so that we can put the system to practical use, decreasing user fatigue as much as possible. We have found that two of the three methods suggested in this paper have improved the operability while maintain a high level of precision, except for the method displaying the sound source in a series. We have found the method displaying the sound source in a series has contributed much to the improvement of operability. We must test how much precision convergence this system has and put it to practical use. At present, we are evaluating the convergence of the method displaying the sound source in a series.

References

[1] Biles, J. A., Anderson, P. G., and Loggi, L. W.: "Neural network fitness functions for a musical IGA," Int'l ICSC Symposia on Intelligent Industrial Automation and Soft Computing (IIA'96/SOCO'96), pp.B39–44 (March, 1996).

[2] Caldwell, C. and Johnston, V. S.: "Tracking a criminal suspect through "face-space" with a genetic algorithm," 4th Int'l Conf. on Genetic Algorithm (ICGA'91), San Diego, CA, US, Morgan Kaufmann Publisher, pp. 416–421 (1991).

[3] Ingu, T. and Takagi, H.: "Accelerating a GA convergence by fitting a single-peak function," *IEEE Int'l Conf. on Fuzzy Systems (FUZZ-IEEE'99)*, Seoul, Korea, pp.1415–1420 (Aug., 1999).

[4] Nagao, M., Yamamoto, M., Suzuki, K. and Ohuchi, A.: "Evaluation of the image retrieval system using interactive genetic algorithm," *J. of Japanese Society for Artificial Intelligence*, vol.13, no.5, pp.720–727 (1998) (*in Japanese*).

[5] Ohsaki, M., Takagi, H., and Ohya, K.: "An Input method using discrete fitness values for interactive GA," *J. of Intelligent and Fuzzy Systems*, vol.6, no.1, pp.131–145 (1998).

[6] Ohsaki, M. and Takagi, H.: "Reduction of the burden of human interactive EC operators - Improvement of present interface by prediction of evaluation order -," *J. of Japan Artificial Intelligence Society*, vol.13, no.5, pp.712–719 (1998) (*in Japanese*).

[7] Ohsaki, M. and Takagi, H.: "Improvement of Presenting Interface by Predicting the Evaluation Order to Reduce the Burden of Human Interactive EC Operators," IEEE Int'l Conf. on System, Man, Cybernetics (SMC'98), pp.1284–1289, San Diego, CA, USA (Oct., 1998).

[8] Siegel, S.: *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill Company, 1988

[9] Takagi, H. Unemi, T., and Terano, T.: "Perspective on Interactive Evolutionary Computing," J. of Japan Society for Artificial Intelligence, vol.13, no.5, pp.692–703 (1998) (*in Japanese*).

[10] Takagi, H., and Kishi, K.: "On-line Knowledge Embedding for Interactive EC-based Montage System," 3rd Int'l Conf. on Knowledge-Based Intelligent Information Engineering Systems (KES'99), pp.280–283, Adelaide, Australia (Aug./Sept., 1999).

[11] Takagi, H.: "Active User Intervention in an EC Search," Int'l Conf. on Information Sciences (JCIS2000), pp.995–998, Atlantic City, NJ, USA (Feb./Mar., 2000).

[12] Watanabe, T. and Takagi, H.: "Recovering System of the Distorted Speech using Interactive Genetic Algorithms," IEEE Int'l Conf. on Systems, Man and Cybernetics (SMC'95), vol.1, pp.684–689, Vancouver, Canada (Oct., 1995).