International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970)
Volume-4 Number-1 Issue-14 March-2014

Voice Matching Using Genetic Algorithm

Abhishek Bal[1], Nilima Paul[2], Suvasree Chakraborty[3], Sonali Sen[4]

Abstract

In this paper, the use of Genetic Algorithm (GA) for voice recognition is described. The practical application of Genetic Algorithm (GA) to the solution of engineering problem is a rapidly emerging approach in the field of control engineering and signal processing. Genetic algorithms are useful for searching a space in multi-directional way from large spaces and poorly defined space. Voice is a signal of infinite information. Digital processing of voice signal is very important for automatic voice recognition technology. Nowadays, voice processing is very much important in security mechanism due to mimicry characteristic. So studying the voice feature extraction in voice processing is very necessary in military, hospital, telephone system, investigation bureau and etc. In order to extract valuable information from the voice signal, make decisions on the process, and obtain results, the data needs to be manipulated and analyzed. In this paper, if the instant voice is not matched with same person's reference voices in the database, then Genetic Algorithm (GA) is applied between two randomly chosen reference voices. Again the instant voice is compared with the result of Genetic Algorithm (GA) which is used, including its three main steps: selection, crossover and mutation. We illustrate our approach with different sample of voices from human in our institution.

Keywords

Genetic Algorithm, Voice Recognition, Euclidean Distance, Threshold Value, Fitness Function.

This work was supported in part by the Department of Computer Science, St. Xavier's College (Autonomous), affiliated by Calcutta University, Kolkata, India.

Manuscript received March 24, 2014.

Abhishek Bal, Department of Computer Science, St. Xavier's College(Autonomous),Kolkata, India.

Nilima Paul, Department of Computer Science, St. Xavier's College(Autonomous),Kolkata, India.

Suvasree Chakraborty, Department of Computer Science, St. Xavier'sCollege(Autonomous),Kolkata, India.

Sonali Sen, Department of Computer Science, St. Xavier's College(Autonomous),Kolkata, India.

1. Introduction

Voice recognition technique can be classified into identification and verification. This is the process of automatically recognizing the speaker on the basis of individual information included in speech waves. This technique makes it possible to use the speaker's voice to verify their identity and control access to services such as biometry, voice dialling, database access services, voice mail, and security control for confidential information areas, and remote access to computers [14].

This report defines a set of evaluation criteria and test methods for voice recognition [1, 11, 12] system used in real time system. The sound based applications have been always important in communication for the humans. Some applications are speech recognition, voice-distance-talk and other forms of personal communications. So the speech based interface [1,12] has been employed in mobile and stationary devices. But it could not reach its expectations due to variations in surrounding noises, changes in person to person speech and intra person variation [1]. This scenario leads to further research that will make voice recognition more robust and general and can be applied upcoming electronic devices. We have improved the performance of voice recognition system using evolutionary computational algorithms (Genetic algorithm, GA) [2, 3, 4, 13].

Genetic algorithms [9] have been traditionally considered as robust techniques [6], easily applicable to almost any domain. Genetic Algorithms (GAs) are adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics. Genetic algorithm used to solve optimization problems and also helpful in random search. Optimization [10, 13] is a process of finding a best or optimal [10] solution for a problem. It is better than conventional AI (Artificial Intelligence); It is more robust [6, 13]. Unlike older AI systems, the GA's do not break easily even if the inputs changed slightly or in the presence of reasonable noise.

In Figure 1, the voice recognition technique is shown using Genetic Algorithm.

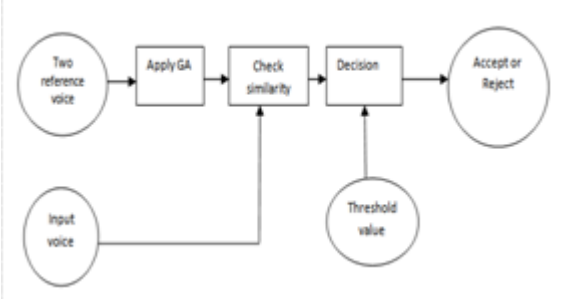

Figure 1: Basic Structure of Human voice verification

2. Literature Review

In 1990s the key technologies developed during this period were the methods for stochastic language understanding, statistical learning of acoustic and language models, and the methods for implementation of large vocabulary speech understanding systems.

After the five decades of research, the speech recognition technology has finally entered marketplace, benefiting the users in variety of ways. Designing of a machine is still big challenge because it cannot function truly like an intelligent human.

In 2006, Patricia Melin, Jerica Urias, Daniel Solano, Miguel Soto, Miguel Lopez, and Oscar Castillo [1] proposed that speaker recognition by analyzing the sound signal with the help of intelligent technique, such as the neural network and fuzzy system. In this paper, neural network for analyzing the sound signal of unknown speaker and also used the set of type-2 fuzzy rules. Here, a fuzzy rule used due to the uncertainty of the decision processing and also the genetic algorithm is used to optimize the architecture of the neural network.

In 2013, Preeti Saini and Parneet Kaur [12] represented a study of basic approaches to speech recognition and their results shows better accuracy. This paper also presents what research has been done around for dealing with the problem of ASR (Automatic Speech Recognition).

In 1996, K.F.Man, K.S.Tang, and S.Kwong [3] represent a comprehensive coverage of the techniques involved describing the characteristics, advantage and constraints of Genetic algorithm as well as discussing genetic operations such as crossover, mutation, and reinsertion in terms of parallelism, multi-objective and multi-model etc.

3. Proposed Work

This section describes the method of capturing the acoustics of a human and details of the experimentation procedure.

3.1. Recoding of Sound

We use microphone as a voice recorder to record voice in a closed room and save the voice using audio file format (.wav). Then we extract the amplitude values of the voice file in excel sheet. Here, we get two columns (one for left channel and the other for right channel) of amplitude values because we capture voice using stereo type recorder. We choose any one of the columns as amplitude values to process.

3.2. Voice Recognition using Genetic Algorithm

As genetic algorithm is a well-known procedure for handling noisy functions well, hence, the methodology adopted here uses for recognizing the voice features and matching the voice from the available training voice database. The voice features are recognized using genetic algorithm following a systematic technique which includes repeated iterations consisting of genetic operators that are selection, crossover, mutation and reproduction until we get the optimal solution.

3.2.1. Selection of Two Sound Files

The parameter (population size) [5, 6] is very important things in GA [2, 3, 4, 13]. So population size is taken as the minimum size of two voices. We want to match the instant voice with the reference voices that store in database. For this checking we calculate the difference between i^{th} amplitude values of two voices. We select a threshold value 0.09 after a lot of tests. If the difference is below the threshold value, then counter is increased, otherwise loop. If the counter is greater than from the half of the file size; then we can say that the two voices are same. If two files are not same then selection [6] is needed for choosing two files from database.

The first operator which is usually applied on population is reproduction. We apply Roulette Wheel Selection [13] mechanism for choosing two sound files depending on fitness [7, 8] function $f(x)=x^2$ [13]. That's why Roulette Wheel Selection [13]

mechanism also known as Fitness Proportionate. Here some modification is done in Roulette Wheel Selection mechanism. We calculate the fitness value of each individual rather than choosing the best individual among all individuals depending on fitness value. This is the main concept of Roulette Wheel Selection mechanism [3, 13].

3.2.2. Crossover between Two Sound Files

We consider the fitness function $f(x)=x^2$ [13] and cross point $\log_2(n+1)$ where n is the number of bits in each individual. After choosing the cross point, we apply one point crossover mechanism to crossover [7,13] between amplitude values of two selected voice files and copy everything before this point from the first voice and then everything after the crossover point copy from the second voice. Now, we choose best offspring after crossover as a new offspring based on fitness value [13].

Now we show the following figure to understand the working principle of one point crossover.

Consider the two parents selected for crossover.

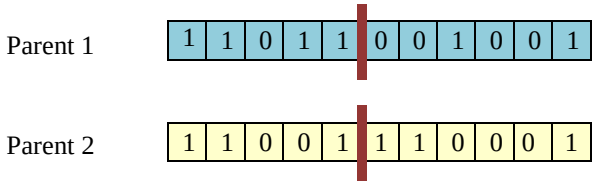

Figure 2: Bit representation of two parents

Interchanging the parent chromosomes after the crossover points, the offspring produced are:

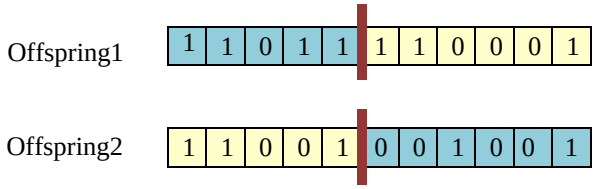

Figure 3: Crossover of two parents

3.2.3. Mutation for New Offspring

After crossover is performed, mutation [7, 8, 13] takes place. Mutation is the genetic operator that is used for keeping the genetic diversity from one generation of a population of chromosomes to the next. With the new offspring value the genetic algorithm may be able to arrive at better solution from previously possible solution. So, the new generated offspring is mutated [13].

We consider the mutation point $\log_2(n+1)$ where n is the number of bits in each individual. After selecting mutation point, we use flip bit mutation to produce mutated offspring. Then calculate the fitness value for this new mutated offspring [4]. We do the crossover and mutation procedure several times [8, 10, 13] until better offspring is generated rather than its parents.

Now we show the following figure to understand the working principle of flip bit mutation operator that simply inverts the value of the chosen gene. i.e. 0 is replaced by 1 and 1 is replaced by 0.

Consider the original offspring1 selected for mutation.

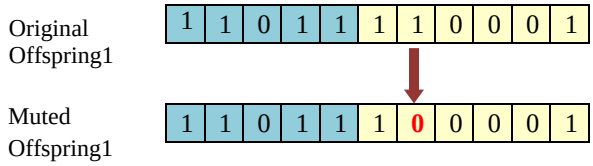

Figure 4: Mutation of offspring1 at 7th position

Consider the original offspring2 selected for mutation.

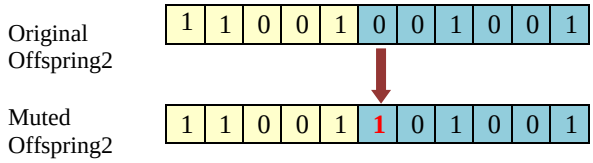

Figure 5: Mutation of offspring2 at 6th position

3.3. Comparison between Two Sound Files

In our approach, we compare the genetic resultant sound with other sound in following two ways -

a) Compare the genetic resultant sound with instant sound using Threshold Value.

b) Compare the genetic resultant sound with instant sound file using Euclidean Distance.

a) Compare the Genetic Resultant Sound with Instant Sound using Threshold Value

After getting a better offspring, we compare this offspring (sound file) with instant voice file. We compare these two sound files using a threshold value. In environment the undesired noise like sound from heavy machines, vehicles are present in one or other form in everywhere. So, these undesired noise effects in speech transmission and acquiring systems. So the same sounds are varying due to this noise. In this case comparing amplitude values of sound would not give better solution. So we filtering the noise and then applying GA so that the good features are come and then we compare

the sound with reference sound for getting the optimal solution.

b) Compare the Genetic Resultant Sound with Instant Sound using Euclidean distance

The Euclidean distance or Euclidean metric is the "ordinary" distance between two points that one would measure with a ruler, and is given by Pythagorean formula. The formula for this distance from the point $X(x_1,x_2,x_3,...x_i)$ to $Y(y_1,y_2,y_3,...y_i)$ is given by :-

$$d=\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

After getting a better offspring, we calculate the Euclidean distance between this offspring (sound) and input sound. Here, Euclidean distance is inversely proportional to the probability of voice matching.

4. Algorithm for Voice Matching using Genetic Algorithm

```
Algorithm VOICE_MATCHING
Begin
input1=sound1; (* initialize a sound *)
refference1=sound2; (*sound from database*)
input1_score=SCORE (input1);
refference1_score=SCORE (refference1);
if (input1= refference1)
PRINT "two voices are same";
Exit;
end if
/*if the two sounds are not same*/
refference2=sound3;(*another sound from database*)
repeat
(* crossover section *)
Offspring1=CROSSOVER(refference1,refference2);
Offspring1_score=SCORE(Offspring1);
until (better offspring is not found)
repeat
(* mutation section *)
Offspring2=MUTATION(Offspring1);
Offspring2_score=SCORE(Offspring1);
until (better offspring is not found)
if(Offspring2=input1)
PRINT "same voice";
else
PRINT "not same voice";
end if
End.
```

In our algorithm, the SCORE() method is basically selection process based on the fitness value which is calculated by function $f(x)=x^2$. The reference voices are chosen by Roulette Wheel Selection mechanism.

5. Experimental Result and Discussion

In this section we describe the results of some test cases using our approach with the use of Genetic Algorithm. We now show some example to illustrate our experimental result. In this scenario, two sound file's amplitudes are taken. Then, absolute differences are calculated. A counter is initialized to count how many number of times the absolute difference is crossed the threshold value. Each of the difference is compared with the threshold value which is set to 0.09 after a lot of test experiments. If the difference is below the threshold value, then counter is increased. If the counter value is greater than the half of the file size, then we can say that the two voices are same, otherwise not same.

A. Experiment 1:

Now we illustrate our first experiment where two voices are same. For doing this we take an input voice from the human and any reference voice from the other reference voices that are stored in database. First voice is of the word "hello" in English as input which is shown in Figure 6.

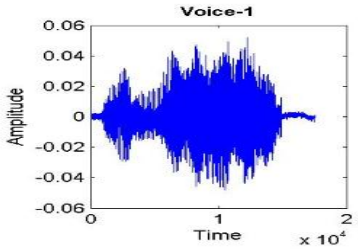

Figure 6: Input voice signal of the word "hello"

Second voice is of the word "hello" in English, store in database is shown in Figure 7.

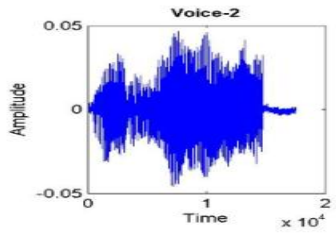

Figure 7: Reference voice signal of the word "hello"

We give an example of data set in the following Table 1.

Table 1: Comparison between Two Voices

Input voice amplitude	Reference voice-1 Amplitude	Absolute Difference	Abs diff > Threshold value	Result
0.006439	-0.002014	0.008453	No	SAME
0.006378	0.001038	0.007416	No	
0.005096	0.000702	0.004395	No	
0.003113	0.002991	0.000122	No	
0.001038	0.005493	0.004456	No	
-0.000488	0.007874	0.008362	No	
-0.001129	0.009766	0.010895	No	
-0.000854	0.010925	0.011780	No	
0.000153	0.011230	0.011078	No	
0.001404	0.010620	0.009216	No	
0.002441	0.009216	0.006775	No	

Here, the Figure 6 and Figure 7 consists more or less one lakh amplitude values but it is not possible to show all the values. So, we highlight some of the portion of the experimental result in Table 1. From Table 1, we conclude that the absolute differences of amplitude values of two voices do not cross the threshold value most of the time. So, we can say that these two voices are same.

B. Experiment 2:

We illustrate another experimental part when the input voice is not matched with the reference voice. We take an input signal which is shown in Figure 8 as input and then compare it with any reference signal shown in Figure 7 which is stored in database. But, these two files are not same.

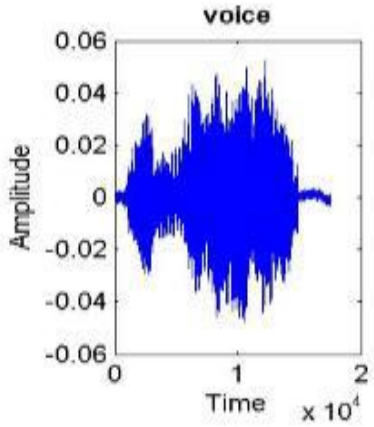

Figure 8: Input voice signal of the word "hello"

We give the experimental table in Table 2 where voices are not same.

Here, in Table 2 we calculate the absolute difference between two voice files (one is input voice and another is reference voice signal), then compare the difference value with the threshold value and see that

Table 2: Comparison between Two Voices

Input voice amplitude	Reference voice-1 amplitude	Absolute Difference	Abs diff> Threshold Value	Result
0.008148	0.000335	0.007813	No	NOT SAME
0.089996	0.008789	0.081207	No	
0.167450	0.009674	0.157776	Yes	
0.219696	0.003326	0.216370	Yes	
0.229889	-0.007233	0.237122	Yes	
0.191071	-0.017639	0.208710	Yes	
0.109344	-0.024109	0.133453	Yes	
0.003204	-0.024414	0.027618	No	
-0.101135	0.018707	0.082428	No	
-0.177246	0.008972	0.168274	Yes	
-0.206177	-0.001617	0.207794	Yes	

the half or more differences of two voices are greater than the threshold value. So, we can say that these two voices are not same.

As the two voices are not same, we apply Genetic Algorithm that generate a new voice with some new characteristic that makes the closest match to input voice. So, now we choose two reference voices of the same person from database using selection process and then apply Genetic Algorithm on these. After applying Genetic Algorithm, we get this new signal with some new characteristics is shown in Figure 9.

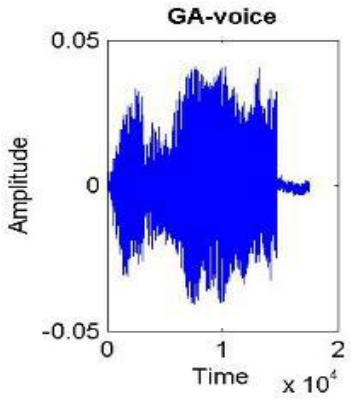

Figure 9: Genetic result of two reference voices

Now we compare the Genetic resultant voice with the input voice which is shown in Figure 8 in two ways so that we can get accurate match. Firstly, compare the genetic result with input sound file using Threshold value. Secondly, compare the genetic result using Euclidean distance.

i) Using Threshold Value

In this case, we compare these two files using a threshold value. The threshold value is set to 0.09 after lot of sample test. We calculate the difference between the amplitude values of Genetic Result shown in Figure 9 and input voice signal which is shown in Figure 8 and then compare the differences with the threshold value. Here, most of the times absolute difference does not cross threshold value. So we can say that these voices are same.

Now, we give the experimental table which is shown in Table 3 where we compare the Genetic resultant voice with input voice using threshold value.

Table 3: Comparison between GA result and input voice

Input voice amplitude	GA result amplitude	Absolute Difference	Abs diff> Threshold Value	Result
0.008148	0.000477	0.007671	No	SAME
0.089996	0.070562	0.019434	No	
0.167450	0.099123	0.068327	No	
0.219696	0.004526	0.21517	Yes	
0.229889	-0.000256	0.230145	Yes	
0.191071	0.125679	0.065392	No	
0.109344	0.094389	0.014955	No	
0.003204	-0.024872	0.028076	No	
-.101135	-0.113260	0.012125	No	
-.177246	-0.087712	0.089534	No	
-.206177	-0.258941	0.053664	No	

Here, in Table 3 we calculate the absolute difference between two voice files (one is input voice and another is Genetic result), then compare the difference value with the threshold value and see that the half or more differences of two voices are not greater than the threshold value. So, we can say that these two voices are same.

ii) Using Euclidean Distance

Secondly, we compare the input voices with the GA voice using Euclidean Distance. We give an experimental result of measuring Euclidean Distance from one voice signal to different voice signals. Here we take 12 voices of same word of same person and then calculate the Euclidean distance to see that how much the GA result voice is close to the other reference voices. In this section, we calculate the Euclidean distance between two sound files using the Equation (1) where x_i represents the amplitude values of one sound file and y_i represents the amplitudes value of another sound file. Now, we give the experimental table which is shown in Table 4 where we calculate Euclidean distance and the percentage of similarity using threshold value.

Table 4: Experimental value using Euclidean Distance

Genetic Sound	Compare with	Using f(x)=x*x	
		% of Same (using threshold)	Euclidean Distance
GAvoice_1_2	Voice_1	55.8358	33.8098
	Voice_2	95.6582	0.4806
	Voice_3	56.7977	35.4254
	Voice_4	55.6276	35.5585
	Voice_5	57.1877	32.2513
	Voice_6	55.2815	33.5545
	Voice_7	54.3255	34.3444
	Voice_8	57.3724	33.6477
	Voice_9	52.9707	35.5144
	Voice_10	58.0645	33.0312
	Voice_11	54.7713	33.9864
	Voice_12	55.6276	35.5585

From above experiments, we observe that the comparison between input voice and genetic voice using Euclidean distance method is better than using threshold value method. Because choosing the threshold value requires lots of computation and if choosing the threshold value is not good enough then the total experimental process is discarded due to the wrong result. But Euclidean distance method is reportedly faster than most other methods. It also compares the relationship between actual sound files and gives the average result. But the Euclidean distance method suffers in such cases where there is a high noise-to-signal. Overall, we can say that the comparison using Euclidean method is more accurate than using threshold value.

6. Conclusion

The goal of the experiment was purely practical to find reduce set of features for voice matching. We have described in this paper an intelligent approach for voice recognition using Genetic algorithm. It is found genetic algorithm for optimization gives better result for voice recognition. We have performed tests on the samples of different type of human voice signal with and without noise and very good result is achieved during recognition.

Recognition accuracy has been improved with a larger set of reference templates. To found good solution it requires many trials and trials. Genetic Algorithm has been used for difficult problems (such as-hard problems), for machine learning and also for evolving simple programs. It is also useful for evolving pictures and music. Advantage of Genetic Algorithm (GAs) is in their parallelism. The main advantage of evolutionary algorithm is that they do not have much mathematical requirements about the optimization problems.

7. Future Scope

In fact the proposed method gives better results than the algorithms proposed earlier but there is always a window of improvement. Better results may be achieved by putting genetic algorithm with a combination of neural network and fuzzy logics. Better feature extraction techniques can be implemented to achieve higher degree of accuracy.

Acknowledgment

The authors are very much grateful to the Department of Computer Science for giving the opportunity to work on voice matching using Genetic Algorithm (GA). One of the authors Sonali sen sincerely expresses her gratitude to Fr. Dr. Felix Raj, Principal of St. Xavier's College (Autonomous) for giving constant encouragement in doing research in the field of sound processing.

References

[1] Patricia Melin, JericaUrias, Daniel Solano, Miguel Soto, Miguel Lopez, and Oscar Castillo,"Voice Recognition with Neural Networks,Type-2 Fuzzy Logic and Genetic Algorithms", Engineering Letters, 13:2, EL_13_2_9,4 August 2006.

[2] Georges R. Harik, Fernando G. Lobo, and David E. Goldberg ,"The Compact Genetic Algorithm", University of Illinois at Urbana-Champaign Urbana,IL 61801,IlliGAL Report No. 97006,August 1997.

[3] K. F. Man, Member, IEEE, K. S. Tang, and S. Kwong, "Genetic Algorithms: Concepts and Applications", Member, IEEE Transaction On Industrial Electronics,Vol-43,No-5,October 1996.

[4] Kumara Sastry(2), David Goldberg(2), Graham Kendall(3), "GENETIC ALGORITHMS", Springer,2. University of Illinois, USA, 3.University of Nottingham, UK, 2005.

[5] Davis, L. D. (ed)," Genetic Algorithms and Simulated Annealing, Pitman publishing", London, 1987.

[6] Sastry, K., "Evaluation-relaxation schemes for genetic and evolutionary algorithms", Master's Thesis, General Engineering Department, University of Illinois at Urbana-Champaign, Urbana, IL,2001.

[7] Sastry, K. and Goldberg, D. E., 2, Let's get ready to rumble: Crossover versus mutation head to head, in: Proc. 2004 Genetic and Evolutionary Computation Conf. II, Lecture Notes in Computer Science, Vol. 3103, Springer, Berlin, pp. 126–137,2004.

[8] Srivastava, R. and Goldberg, D. E.,Verification of the theory of genetic and evolutionary continuation, in: Proc. Genetic and Evolutionary Computation Conf., pp. 551–558, 2001.

[9] David E. Goldberg, Addison-Wesley, "Genetic Algorithms in Search Optimization and Machine Learning", chapter 1-8, page 1-432, 1989.

[10] Randy L Haupt, "Practical genetic Algorithm", John Wiley and Sons Inc, chapter 1-7, page 1-251, 2004.

[11] David J. White, Andrew P. King, And Shan D. Duncan,"Voice Recognition Technology, As A Tool For Behavioral Research",34 (1), 1-5,Springer ,Indiana University, Bloomington, Indiana,February 2002.

[12] Preeti Saini, Parneet Kaur, "Automatic Speech Recognition: A Review", International journal of Engineering and Technology, 2CSE Department, Kurukshetra University, ACE, Haryana, India,Vol 4 Issue2-2013.

[13] Artificial Intelligence: Course Content, Lecture hours – 42, RC Chakraborty, June 01, 2010. Available: http://www.myreaders.info/html/artificial_intelligence.html.

[14] T. Matsui, and S. Furui, "Concatenated phoneme models for text-variable speaker recognition", Proceedings of ICASSP'93, pp. 391-394, 1993.

Abhishek Bal received his B.Sc (Hons.) degree from West Bengal State University, Kolkata, India in 2012. He is currently pursuing his M.Sc in Computer Science at St. Xavier's College (Autonomous), Kolkata, India. He was born in Kolkata on 02.08.1991. He is presently involved in research work in Sound Processing.

Nilima Paul received her B.Sc(Hons.) degree from West Bengal State University, Kolkata, India in 2012. She is currently pursuing her M.Sc in Computer Science at St. Xavier's College (Autonomous), Kolkata, India. She was born in Kolkata on 30.08.1991. She is presently involved in research work in Sound Processing.

Suvasree Chakraborty received her B.Sc(Hons.) degree from Calcutta University, Kolkata, India in 2012. She is currently pursuing her M.Sc in Computer Science at St. Xavier's College (Autonomous), Kolkata, India. She was born in Kolkata on 15.04.1991. She is presently involved in research work in Sound Processing.

Sonali Sen is the Associate Professor in Department of Computer Science at St. Xavier's College (Autonomous), Kolkata, India. Her area of interest covers Advance Operating System, Microprocessor, Micro-controller, VLSI design, Object Oriented Programming etc. Her research field is concerned with Bio-informatics, Sound Analysis.