Beamforming based Speech Recognition using Genetic Algorithm for Real-time Systems

Milind U. Nemade, Satish K. Shah

Abstract*— *The speech based applications have been always important in communication for the humans. There are in various essential applications like speech recognition, voice-distance-talk and other forms of personal communications. Most recently, speech based interface has been tried to be employed in almost all the mobile and stationary devices. However, these attempts could not give ultimate response due to variations in surrounding noises, changes in person to person speech and also intra person variation. This scenario leads to further research that will make speech recognition more robust and general and can be applied upcoming electronic devices to be sued for gaming, entertainment, cellular phones. The broad categories of speech enhancement techniques can be listed as speech filtering techniques, beam forming techniques and active noise cancellation methods. In this paper, we have improved the performance of beamforming based speech recognition system using evolutionary computational algorithms (Genetic algorithm, GA). Additionally, the system is made to be working in real-time as time required for classifier has been reduced dramatically. This is particularly achieved by including the zeros at random places and in random amount in initial population chromosomes, which were generated randomly in the range of 0 to 1. This results in the reduction of feature elements in feature descriptor and have feature vector length. The experiments were performed for 20 words including numbers and commands, 10 words of numbers only and 10 words of commands only for different values of filter bank parameters. The results show the effectiveness of the GA optimization in all the subsets of experiments with different parameters of beamforming.

Index Terms*— *Delay and sum beamformer, HMM based classifier, Least Mean Square, MFC, Nearest Neighbor Classifier.

I. INTRODUCTION

Speech is the fundamental and common medium, hence important for us, to communicate and most effective and reliable means for expressing oneself for personal communication. With advancement in hardware technologies, there are so many electronic and mobile personal communication based devices available, today in market and that too in cheaper cost and with easy availability. The applications like speech recognition, mobile and personal communication, public address system are few of the applications from long list of speech based systems. However, undesired noises in environment like sound from heavy machines, vehicles are also present in one or other form everywhere.

These noises cause undesired effects in speech transmission and acquiring systems. Recently, restricted or usable vicinity of applications is moving from one place and close room to more open and multiple locations, leading to several types of undesired signals of mixing with desired speech signal making speech more corrupt with noise. Not only human communications but intelligent machines which trying to automate the things and sometimes also takes decision based on what it receives as a speech, also suffers from the degraded performance. Since last five decades, various approaches for noise reduction and speech enhancements have been investigated and developed. Among, very early and fundamental approach of noise reduction was introduced to use the theory of the optimum Wiener filter. Given a desired signal and an input signal, the Wiener filter produces an estimate of the desired signal that is optimal, i.e. the squared mean error or difference between the signals is minimized. The Wiener filter can also be adaptively estimated used in an environment where the surrounding noise has time-varying characteristics. Adaptive algorithms such as Least Mean Square (LMS) and Recursive Least Squares (RLS) are well known examples and also widely used. Recent advances in CPU and multi-core hardware has provided ample amount of computational power and thus, need for today is to design the complex but yet efficient and realistic approach for noise reduction to achieve speech enhancement. The speech enhancement is not only useful for storage and transmission of speech data but it can play vital role in improving much need system based speech recognition where accurate identification of words and sentences can provide automation in most of the human-machine based interface and also be useful in machine-machine interaction based automation. Robotics is a familiar example where speech recognition systems can become boon for today's advanced society at social level in addition to during natural calamities and on war fields.

It is obvious that speech enhancement can boost up the performance of speech recognition systems by keeping low word error rate (WER). There are various types of advanced speech enhancement algorithms in literature and they can be classified in main three categories, namely; filtering/estimation based noise reduction, beam forming and active noise cancellation (ANC) techniques. In this paper, we have presented the approach of evolutionary computation in form of genetic algorithm to select the features that are responsible for discriminating the different words. In doing so, the amount of feature elements to be used also gets reduced and hence system can be made to recognise the word-speech with real-time performance.

Revised Manuscript Received on 30 May 2013.
* Correspondence Author

Milind U. Nemade, Dept. of Electronics and Telecommunication, University of Mumbai, K.J.Somaiya Institute of Engg. And Information Technology, Sion(E), Mumbai, Maharashtra, India.

Prof. Satish K. Shah, Dept. of Electrical Engineering, M. S. University of Baroda, Faculty of Technology and Engineering, Vadodara, Gujrat, India.

The system is made to be working in real-time as time required for classifier has been reduced dramatically.

This is particularly achieved by including the zeros at random places and in random amount in initial population chromosomes, which were generated randomly in the range of 0 to 1. This results in the reduction of feature elements in feature descriptor and have feature vector length. This is especially an important requirement in the mobile devices where power, memory and processing power are available with large constraints. The in-car infotainment and mobile devices are the potential examples of real-time constraint requirements. The experiments were carried out different filter-bank parameters; filter length and number of subbands. The experiments were performed separately for 20 words including numbers and commands, 10 words of numbers only and 10 words of commands only for different values of filter bank parameters. The results obtained have proved the speech enhancing capability in the beamforming based speech recognition system using genetic algorithm. In beamforming simulated speech, multi-microphone network was generated with noise and echo-interference, which can degrade the original speech signal.

The remaining part of the paper is organized as follows: In next section II, existing work, related to the beamforming and genetic based optimization in speech recognition, have been presented. Section III explains the beamforming structure used in the experiments. The methodology including GA based optimization has been explained in brief in section IV. Section V presents the results obtained and discussion over them. Finally, paper is concluded with summary of the work in section V.

II. RELATED WORK

One of important class of speech enhancement methods is based on the beam-forming, where more than one speech channels (microphones) are used to process the speech. Speech signals are received simultaneously by all microphones and outputs of these sensors are then processed to estimate the clean speech signal. In adaptive beamforming, an array of antennas is exploited to achieve maximum reception in a specified direction by estimating the signal arrival from a desired direction (in the presence of noise) while signals of the same frequency from other directions are rejected. This is achieved by varying the weights of each of the sensors (antennas) used in the array. This kind of speech enhancement techniques can give better performance of the speech applications like automatic speech recognition (ASR) than signal channel processing. Only disadvantage with this class of methods is higher cost of hardware, which can put restriction on using these methods in some speech applications. The basic block diagram of beamformer is shown in figure 1. Frost [1] has suggested constrained minimum power adaptive beamforming, which deals with the problem of a broadband signal received by an array, where pure delay relates each pair of source and sensor. Each sensor signal is processed by a tap delay line filter after applying a proper time delay compensation to form delay-and-sum beamformer. The algorithm is capable of satisfying some desired frequency response in the look direction while minimizing the output noise power by using constrained minimization of the total output power. This minimization is realized by adjusting the taps of the filters under the desired constraint using constrained LMS-type algorithm. Griffiths and Jim [2] reconsidered Frost's algorithm and introduced the generalized sidelobe canceller (GSC) solution. The GSC algorithm is comprised of three building blocks. The first is a fixed beamformer, which satisfies the desired constraint. The second is a blocking matrix, which produces noise-only reference signals by blocking the desired signal (e.g., by subtracting pairs of time-aligned signals). The third is an unconstrained LMS-type algorithm that attempts to cancel the noise in the fixed beamformer output. In [2], it is shown that Frost algorithm can be viewed as a special case of the GSC. The main drawback of the GSC algorithm is its delay-only propagation assumption.

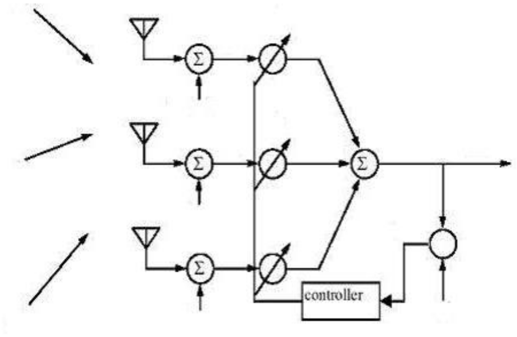

Fig. 1. Beamformer: An Adaptive array system

In another work [3], switching adaptive filters were used to form the beamformer. This beamformer has two sections and interconnected with switch. The first section determines the adaptive look direction and cues in on the desired speech and is adapted only when speech is present. Second section which adapted during silence-only periods is implemented as multichannel adaptive noise canceller. In [4], authors have proposed the solution to GSC algorithm by estimating ratio of transfer functions (TFs), otherwise it is based on TFs which relates source signal and the sensors. The TF ratios are estimated by exploiting the non-stationarity characteristic of the desired signal. This algorithm can be used normally in reverberating room having acoustic environment. One interesting paper [5], describes how optimal finite-impulse response subband beamforming can be used by including coherent multipath propagation into optimality criterion for speech enhancement in multipath environment.

In application point of view, a constrained switched adaptive beamforming (CSA-BF) [6] was used for speech enhancement and recognition in real moving car environment. This algorithm consists of a speech/noise constraint section, a speech adaptive beamformer and noise adaptive beamformer. The performance obtained with this algorithm was compared with classic delay-and-sum beamforming (DASB) using CU-Move corpus and found decrease in word-error-rate (WER) by 31% in speech recognition. The computational complexity of DASB is very low and can be easily implemented for real-time requirement. It is also effective when direction of desired source is known and can be applied in the car as driver's head position is restricted based on seat position. However, as there is possibility of change in drivers head direction, DASB algorithm could be inconsistent and this inconsistency can be solved by employing CSA-BF algorithm which can improve the SNR by up to +5.5 dB on the average.

*Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)*

For the application of hands-free speech recognition, one of the works [7] uses sequence of features to be used for speech recognition itself, to optimize a filter-and-sum beamformer instead of separating the beamformer, to be used for speech enhancement, from speech recognition system. In this work, they used frequency cepstral coefficient (MFCC) and applied to the HMM based classifier for speech recognition.

Optimizing beamformer without knowledge of source or acoustic characteristic of environment is termed as "blind beamforming". One of the papers [8] proposes blind speech enhancement using beamformer which consist of subband soft-constrained adaptive filter using recursive least square (RLS) algorithm, combined with subband weighted time-delay estimator (TDE). Estimation of propagation time difference of arrival of a dominate speech source received by sensor array is based the steered response power with phase transform (SRP-PHAT) algorithm, which was modified to work in subband structure. One recent paper [9] presents phase-based dual-microphone speech enhancement technique based on prior speech model. In this work, it is claimed that around 23% improvement achieved using this algorithm as compared to the delay-and-sum beamformer, where experiments were conducted on the CARVUI database.

In application point of view, the study presented in [10] addresses the problem of distant speech acquisition in multiparty meeting s using multiple cameras and microphones. The camera, used as a multi-person tracker, was used to give the more precise location of each person to the microphone array beamformer. They evaluated the performance of speech recognition using data recorded in a real meeting room for stationary speaker, moving speaker and overlapping speech scenarios. The result obtained with audio-video speech enhancement was better than that with only audio. In one of the recent work [11], adaptive beamformer based on estimation of power spectral density (PSD) and noise statistics update was proposed. An inactive-source detector based on minimum statistics is developed to detect the speech presence and to acquire the noise statistics. The performances of this beamformers were tested in a real hands-free in-car environment. One of the most recent papers [12] uses GSC based speech enhancement using the location of speaker obtained via localization module. This algorithm relies on time delay compensation, DFT computations, fixed channel compensator, adaptive channel compensator.

There are also attempts made by researchers for employing computational intelligence and evolutionary computations in the speech recognition system. Especially, genetic algorithm has been applied in various techniques of speech recognition. In [22], author proposes a genetic algorithm (GA) based beamformer in which beamformer weights are optimized with the help of GA operators, crossover and mutation. It is claimed that this GA based optimization is successful in tackling the non-differentiable and non-linear natures of speech recognition in normal and noisy environment.

Another category of GA application in speech recognition is to select the optimized feature such that it will improve the performance of recognition. The generic sound recognition system that exploits evolutional algorithms for a selection of discriminative acoustic features has been presented in [23]. Similar kinds of objectives have been achieved in [24, 25, 27]. In [30] and other works, feature set itself, in form of codebook dictionary such as in vector quantization, have been optimized.

Apart from optimizing feature set, classifier like multilayer perceptron based neural network was optimized using genetic algorithm in [26, 28]. Another classifier HMM was also made to give optimal performance with the help of GA in [29].

III. BEAMFORMING FILTER STRUCTURE

In order to analyse the performance of the speech recognition in the speech corrupted by noise and echo-interference, the analysis frame work used here is taken from [21]. Adaptive Filtering is an important technique in the field of speech processing including speech enhancement, echo- and interference cancellation and speech coding. Filter banks have been introduced in order to improve the performance of time domain adaptive filters with additional benefits like faster convergence and the reduction of computational complexity with shorter filters in the subbands being processed at reduced sampling rate [13]. Due to inappropriate structure of filter bank in subband processing and improper design of filters, filter bank may yield degraded performance. The subband FIR filter bank scheme [14] to be used for beamforming is shown in figure 2. The design of filters used here is adapted from and given in detail in references [14,15,16,17]. The design includes the prototype analysis and synthesis filter. The filter bank is obtained by using cosine modulation of prototype filter. The analysis-synthesis filter bank structure is shown in figure 3. The multi-microphone network for beamforming based speech recognition was simulated as described in [21].

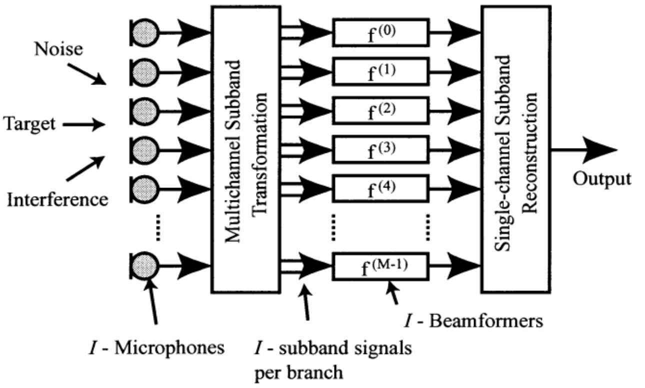

Fig. 2. Subband FIR Beamforming Structure [14]

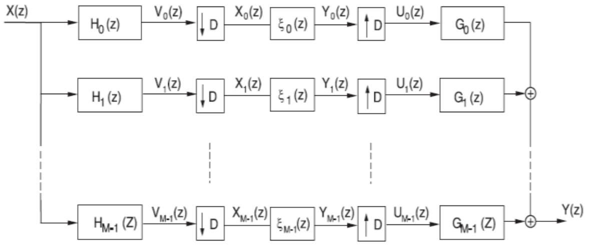

Fig. 3. Analysis and Synthesis Filter Banks with Subband Filtering.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved

IV. METHODOLOGY

A. Beamforming

The signal obtained in each of the microphone is passed through the subband filter bank. The beamformers are formed by using the FIR adaptive filters, whose coefficients are determined by using the LMS algorithm. The beamformer filter is placed between each of analysis subband filter bank and each of microphone branch. This control the gain of each of the subband output from each microphone branch to be passed through the synthesis filter bank for each of the microphone line. The output of entire synthesis filter bank from each of the microphone line is added to form the reconstructed speech output.

B. Recognition

First of all, the features are extracted from the speech of spoken words. The feature Mel frequency cepstral coefficients (MFCC) have been proved to give better performance in case of speech recognition and hence widely used in speech recognition applications [18,19,20]. In speech processing, the Mel-Frequency Cepstrum (MFC) is a representation of the short-term power spectrum of a speech, based on a linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency. The recognition process consists of training the classifier and testing the spoken words with trained classifier. The classier used here is nearest neighbour classifier (NN) based on Euclidian distance metric.

C. Genetic Algorithm

There are numerous attempts made by researchers for evolutionary model in the speech recognition system. Especially, genetic algorithm has been applied in various techniques of speech recognition. The genetic algorithm can be used at two levels. First, the feature elements selection level, where important features are preserved while ignoring remaining, can employ the genetic algorithm. In some cases, the dictionary of features is generated using genetic algorithm as in case of vector quantization codebook [30]. Secondly, the GA can be used at classifier level to determine its optimized parameters. For examples, in neural network the number of hidden layers and number of nodes in each of them can be determined effectively using the evolutionary computations. In case of hidden markov model (HMM), states and state transition parameters can be decides using GA.

The standard MFCC features are very sensitive to additive noise and channel mismatch, therefore the recognition accuracy deteriorates drastically in noisy environments. Thus it is important to remove or suppress the effect of the feature elements which are sensitive to the noise and echo to optimize the recognition performance in high noisy environment. This optimization problem can be handled using genetic algorithm.

The genetic algorithm is applied to the recognition problem of speech words with the objective to find important feature elements that contributes more to classifier for distinguishing one word from others. Additionally, the number of feature elements is also reduced eland into the reduction in the length of feature vector for further step of classification. The chromosome of GA was of length as same as that of feature vector. This chromosome has real value between 0 and 1, randomly generated at each position, in its first form and then it is modified by making 0 to the positions that has lesser value than some randomly generated value between 0 and 1. This modification helped in bringing the wide range variations in usable percentage of total feature elements. This enabled to evaluate the chormosome’s performance of recognition with having even small percentage of elements. This chromosome was multiplied (element wise) with feature vector to be optimized, before using it for recognition.

V. EXPERIMENTAL RESULTS

A. Dataset Preparation

For analysing the performance of speech recognition, we have considered here four speaker’s 20 number of spoken words. These words are listed below in table I and can be categorized on the basis of their use, as numbers and commands. The spoken word from speaker has a length of 2 sec in time. The speech to be used in the experiment is created using multi-microphone mixing environment as described in [21].

TABLE I
LIST OF SPOKEN WORDS

Spoken words (each for 2 sec)	
Numbers	**Commands**
one	yes
two	no
three	hello
four	open
five	close
six	start
seven	stop
eight	dial
nine	on
ten	off

B. Experiments

The prototype filter was designed to construct the filter bank. The frequency spectrums of prototype filter for different length of filters and different numbers of sub-bands are shown in Fig. 4.

For training classifier, 2 speakers’s spoken words used and for testing we used 4 speakers, wherein 2 speakers are unknown and 2 speakers are same as they were in training phase. Each person (speaker) has 20 spoken words, which includes 10 words for numbers and 10 words for commands as listed in table I. The experiments are performed separately with following class of words:

- Numbers and Commands together (20 words)
- Numbers only (10 words)
- Commands only (10words)

The recognition accuracy is calculated as the ratio of correctly recognised words and total words used for recognition test experiment. The above set of experiments was performed twice; firstly without optimization and secondly with GA based optimization. We have used MATLAB® environment for performing all experiments.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

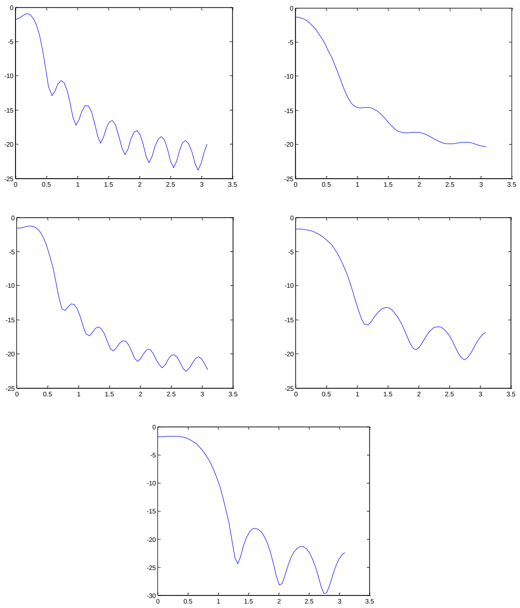

Fig. 4. Frequency Response of Prototype filter with different specifications (no of subbands – filter length): Row-1-Col 1) 16 -16; Row-1-Col-2) 16-8; Row-2-Col-1) 8-16; Row-2-Col-2) 8-8 ; Row-3) 4-8

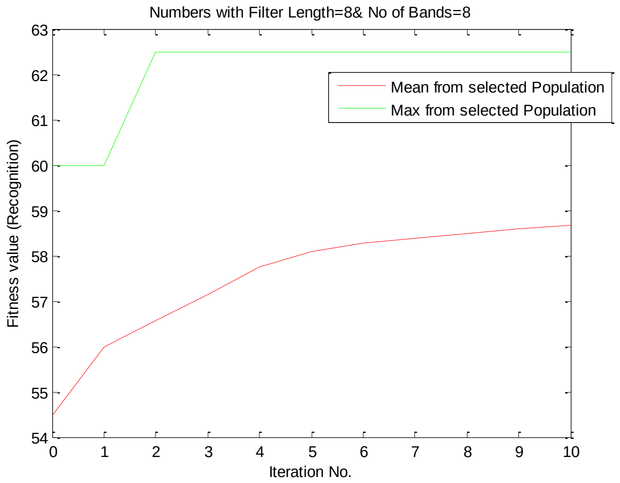

Fig. 5.The graph of fitness value for each of the iteration

TABLE II

RECOGNITION ACCURACY WITH AND WITHOUT BEAMFORMING FOR NUMBERS AND COMMANDS TOGETHER

Filter length	Num of subbands	Multi-mic Beamformed Speech Recognition Accuracy in %			
		Without GA[21]	Best Solution with GA	% improvement with GA	With Least Features
16	16	46.25	51.25	10.8	50.00
8	16	40.00	46.25	15.6	46.25
16	8	42.50	50.00	17.6	48.75
8	8	48.75	56.25	15.4	53.75
8	4	45.00	51.25	13.9	50.00

In first set of experiments, for each class of experiment, the recognition accuracy is calculated in three scenarios. First when pure speech is feed to the recognition experiment without any noise and interference. Secondly, speech was prepared with multi-mic environment with an inclusion of noise and interference (echo). Finally, using beamforming multi-mic speech is enhanced with beamforming-filter bank structure and then fed to the recognition experiment. The last three columns of each of the following tables showing recognition accuracy represents the performance obtained in these three situations.

TABLE III. RECOGNITION ACCURACY WITH AND WITHOUT BEAMFORMING FOR NUMBERS ONLY

Filter length	Num of subbands	Multi-mic Beamformed Speech Recognition Accuracy in %			
		Without GA[21]	Best Solution with GA	% improvement with GA	Least Features GA
16	16	55.00	65.00	18.2	62.50
8	16	47.50	60.00	26.3	57.50
16	8	50.00	62.50	25.0	62.50
8	8	50.00	62.50	25.0	60.00
8	4	60.00	67.50	12.5	67.50

TABLE IV. RECOGNITION ACCURACY WITH AND WITHOUT BEAMFORMING FOR COMMANDS ONLY

Filter length	Num of subbands	Multi-mic Beamformed Speech Recognition Accuracy in %			
		Without GA[21]	Best Solution with GA	% improvement with GA	Least Features GA
16	16	52.50	62.50	19.0	60.00
8	16	50.00	57.50	15.0	57.50
16	8	45.00	55.00	22.2	50.00
8	8	60.00	70.00	16.6	70.00
8	4	52.50	62.50	19.0	60.00

In first set of experiments, for each class of experiment, the recognition accuracy is calculated in three scenarios. First when pure speech is feed to the recognition experiment without any noise and interference. Secondly, speech was prepared with multi-mic environment with an inclusion of noise and interference (echo). Finally, using beamforming multi-mic speech is enhanced with beamforming-filter bank structure and then fed to the recognition experiment. The last three columns of each of the following tables showing recognition accuracy represents the performance obtained in these three situations. In order to analyse the speech recognition performance against the parameters of filter-bank, we selected the two parameters: filter length and number of subbands in filter bank. The experiments were repeated for different values of these two parameters as mentioned in the first two columns of following tables II, III and IV. While performing the experiments with GA based optimization, the feature vector was modified by the each chromosome from the population using inner product operator. Then fitness value was calculated for each of the modified feature vector. In fitness function calculation, recognition ratio is calculated with kNN classifier with two subjects' words in training set and remaining two for testing. The parameters for GA algorithm are:

- Initial Population, 200
- Selected Population in each iteration, 100
- Generated Population using operators, 200
- Elitism, 2 %
- Mutation Rate, 2%
- Uniform crossover rate, 98%

Experiments were conducted separately with and without GA optimization and results are presented in the table II, III and IV.

C. Discussion

It has been observed that GA based optimization was converging in each of the iteration in the sense that mean fitness value of the population in each of the iterations was monotonically increasing. This proves the fact that GA optimization was approaching to find out the optimal solution with the help of GA operations like crossover, mutation in addition to the elitism property of evolution. The graph of fitness value for each of the iteration is shown in Figure 5.

The recognition accuracy obtained in various experiments is shown in tables II, III and IV. The main objective of the speech enhancement is to bring up the speech recognition performance in the presence of noise and echo-interference to the performance obtained with pure speech signals, which is the ideal case. Thus our aim was to boost up the performance of beamforming based speech recognition. Not just that but, performance can be further improved by using GA based optimization. Thus, these tables show the performance with and without GA based optimization.

From the observation of first and second columns of recognition accuracy in each of these tables, it is clear that recognition performance with GA optimized feature vector has been significantly improved. It can be observed that from table II, the speech recognition performance for commands and numbers both can be improved using GA based optimization in the beam forming based speech enhancement. This is also visible in other two experimental results in table III and IV, where only numbers and only commands were used for speech recognition. In addition to this, the best solution with least features elements in last iteration is given in third column of recognition accuracy in each of the tables. The important inference in this point of view is that the recognition performance with least feature elements is not deterrent from the best solution. In other words, we can say that optimal recognition performance with best solution and least amount of feature elements (shortest feature vector) can be easily obtained by the GA optimization, which can be complemented for the real-time computation of classification. This inference further supported by detailed graphs shown in Fig. 6, 7 and 8. These three cases observations in particular are depicted as below:

1. Numbers+Commands: In the case of numbers and commands together in recognition experiment, GA optimization based speech recognition is improved by an average of 15%. The least feature vector length solution also gives similar performance with that of best of solution.
2. Numbers: In case of number recognition, the GA optimization gives very optimal performance and it improves recognition by a factor as high as 26%. This is significant improvement with the additional fact that this improvement can be obtained with shortest feature vector.
3. Commands: In this case, average improvement with GA optimization was around average of 20% with all parameters of filter-bank based beamforming.

The results with percentage of feature elements required to get optimal solution are presented in Table V, VI and VII. The length of feature vector can be reduced as low as 0.5%, saving almost 99% computational power and memory with optimized solution. It is interesting to observe that even with least number of feature elements optimal solution best recognition can be obtained. This fact proves that there so many unnecessary feature elements that are redundant to representation of the word-speech signal. Additionally, there are also feature elements that are sensitive to the noise and echo, removing which performance gets boost up. Another most important shorter feature vector is that in memory required for gallery samples required less and in classification stage computational complexity reduces.

The amount of feature elements that can be used in classification is an important factor, especially in the case of devices with low power(battery operated), low memory and less computational power as in case of mobile hand-held devices. With the smaller size feature vector and yet optimal in recognition performance will take less gallery features. The classifier will take lesser number of computations that that is required with full feature vector. This will also increase the speed of application processing with cheaper hardware, leading to the economical cost of the embedded product.

TABLE V

THE PERFORMANCE IN BOTH CRITERIA IN TERMS OF RECOGNITION ACCURACY AND FEATURE VECTOR LENGTH, BOTH IN PERCENTAGE

Filter Length	Number of Sub-band	Command+Numbers			
		Best Recognition Solution		Least Feature Elements Solution	
		Recognition	FV Size	Recognition	FV Size
16	16	51.25	8.04	50.00	2.24
8	16	46.25	0.78	46.25	0.78
16	8	50.00	5.79	48.75	1.46
8	8	56.25	7.01	53.75	1.85
8	4	51.25	5.55	50.00	1.07

TABLE VI

THE PERFORMANCE IN BOTH CRITERIA IN TERMS OF RECOGNITION ACCURACY AND FEATURE VECTOR LENGTH, BOTH IN PERCENTAGE

Filter Length	Number of Sub-band	Numbers			
		Best Recognition Solution		Least Feature Elements Solution	
		Recognition	FV Size	Recognition	FV Size
16	16	65.00	1.60	62.50	1.07
8	16	60.00	2.87	57.50	0.63
16	8	62.50	1.31	62.50	0.68
8	8	62.50	6.14	60.00	0.68
8	4	67.50	5.70	67.50	2.04

TABLE VI

THE PERFORMANCE IN BOTH CRITERIA IN TERMS OF RECOGNITION ACCURACY AND FEATURE VECTOR LENGTH, BOTH IN PERCENTAGE

Filter Length	Number of Sub-band	Commands			
		Best Recognition Solution		Least Feature Elements Solution	
		Recognition	FV Size	Recognition	FV Size
16	16	62.50	3.43	60.00	1.34
8	16	57.50	4.47	57.50	0.49
16	8	55.00	2.45	50.00	0.49
8	8	70.00	6.37	70.00	1.40
8	4	62.50	9.06	60.00	0.98

VI. CONCLUSION

In this paper, we have presented the approach of evolutionary computation in form of genetic algorithm to select the features that are responsible for discriminating the different words. In doing so, the amount of feature elements to be used also gets reduced and hence system can be made to recognise the word-speech with real-time performance. The system is made to be working in real-time as time required for classifier has been reduced dramatically. This is particularly achieved by including the zeros at random places and in random amount in initial population chromosomes, which were generated randomly in the range of 0 to 1. This results in the reduction of feature elements in feature descriptor and have feature vector length. This is especially an important requirement in the mobile devices where power, memory and processing power are available with large constraints. The in-car infotainment and mobile devices are the potential examples of real-time constraint requirements. The experiments were performed for 20 words including numbers and commands, 10 words of numbers only and 10 words of commands only for different values of filter bank parameters. The results show the effectiveness of the GA optimization in all the subsets of experiments with different parameters of beamforming. The length of feature vector can be reduced as low as 0.5%, saving nearly 99% computational power and memory with optimized solution, leading to a one of the approach to be used in real-time embedded system devices for speech recognition applications.

REFERENCES

1. O. L. Frost, III, “An algorithm for linearly constrained adaptive array processing,” Proc. IEEE, vol. 60, pp. 926–935, Jan. 1972.
2. Griffiths, L.; Jim, C.; , "An alternative approach to linearly constrained adaptive beamforming,", IEEE Transactions on Antennas and Propagation, vol.30, no.1, pp. 27- 34, Jan 1982.
3. Van Compernolle, D , "Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings," ICASSP-90, International Conference on Acoustics, Speech, and Signal Processing, 1990, vol.2, pp.833-836, 3-6 Apr 1990.
4. Gannot, S.; Burshtein, D.; Weinstein, E., "Signal enhancement using beamforming and nonstationarity with applications to speech," IEEE Transactions on Signal Processing, vol.49, no.8, pp.1614-1626, Aug 2001.
5. Grbic, N.; Nordholm, S.; Cantoni, A., "Optimal FIR subband beamforming for speech enhancement in multipath environments," IEEE Signal Processing Letters, vol.10, no.11, pp. 335- 338, Nov. 2003.
6. Xianxian Zhang; Hansen, J.H.L., "CSA-BF: a constrained switched adaptive beamformer for speech enhancement and recognition in real car environments," IEEE Transactions on Speech and Audio Processing, vol.11, no.6, pp. 733- 745, Nov. 2003.
7. Seltzer, M.L.; Raj, B.; Stern, R.M., "Likelihood-maximizing beamforming for robust hands-free speech recognition," IEEE Transactions on Speech and Audio Processing, vol.12, no.5, pp. 489- 498, Sept. 2004.
8. Yermeche, Z.; Grbic, N.; Claesson, I., "Blind Subband Beamforming With Time-Delay Constraints for Moving Source Speech Enhancement," IEEE Transactions on Audio, Speech, and Language Processing, vol.15, no.8, pp.2360-2372, Nov. 2007.
9. Guangji Shi; Parham Aarabi; Hui Jiang; "Phase-Based Dual-Microphone Speech Enhancement Using A Prior Speech Model," IEEE Transactions on Audio, Speech, and Language Processing, vol.15, no.1, pp.109-118, Jan. 2007.
10. Maganti, H.K.; Gatica-Perez, D., McCowan, I, "Speech Enhancement and Recognition in Meetings With an Audio–Visual Sensor Array," IEEE Transactions on Audio, Speech, and Language Processing, vol.15, no.8, pp.2257-2269, Nov. 2007.
11. Hai Huyen Dam; Hai Quang Dam; Nordholm, S., "Noise Statistics Update Adaptive Beamformer With PSD Estimation for Speech Extraction in Noisy Environment," IEEE Transactions on Audio, Speech, and Language Processing, vol.16, no.8, pp.1633-1641, Nov. 2008.
12. Han, S.; Hong, J.; Jeong, S.; Hahn, M., "Robust GSC-based speech enhancement for human machine interface," IEEE Transactions on Consumer Electronics, vol.56, no.2, pp.965-970, May 2010.
13. John J. Shynk, “Frequency-domain and multirate adaptive filtering,” IEEE Signal Processing Magazine, vol. 9, pp. 14–37, 1992.
14. Jan Mark de Haan, Nedelko Grbic, Ingvar Claesson, and Sven Erik Nordholm, “Filter bank design for subband adaptive microphone arrays,” IEEE Trans. Speech Audio Proc., vol. 11, no. 1, pp. 14–23, Jan. 2003.
15. Kumatani, K.; McDonough, J.; Schachl, S.; Klakow, D.; Garner, P.N.; Weifeng Li, "Filter bank design based on minimization of individual aliasing terms for minimum mutual information subband adaptive beamforming," IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2008 , vol., no., pp.1609-1612, March 31 2008-April 4 2008.
16. P. P. Vaidyanathan, “Multirate Systems and Filter Banks”, Prentice Hall, Englewood Cliffs, 1993.
17. Kenichi Kumatani, Tobias Gehrig, Uwe Mayer, Emilian Stoimenov, John McDonough, and MatthiasW¨olfel, “Adaptive beamforming with a minimum mutual information criterion,” IEEE Transactions on Audio, Speech and Language Processing, vol. 15, no. 8, pp. 2527–2541, 2007.
18. L. Rabiner and Biing-Hwang Juang, “Fundamentals of Speech Recognition”, Prentice Hall PTR, 1993.
19. Joseph W. Picone, “Signal Modeling Techniques in Speech Recognition”, Proceedings of the IEEE, vol. 81, No. 9, pages 1215--1247, 1993.
20. Steven B. Davis and Paul Mermelstein, “Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences”, IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-28, vol., No. 4, August 1980.
21. Nemade M. U., Shah S.K., “Improvement in Speech Recognition Performance using Beamforming based Speech Enhancement”, International Journal of Electronics Communication and Computer Engineering (IJECCE), ISSN: 2249-071X (Online, http://ijecce.org) Volume 3 Issue 4, July 2012.
22. Chan, K.Y.; Low, S.Y.; Nordholm, S.; Yiu, K.F.C.; Ling, S.H.; , "Speech Recognition Enhancement Using Beamforming and a Genetic Algorithm," Third International Conference on Network and System Security, 2009. NSS '09. , pp.510-515, 19-21 Oct. 2009.
23. Chmulik, M.; Jarina, R., "Bio-inspired optimization of acoustic features for generic sound recognition," 19th International Conference on Systems, Signals and Image Processing (IWSSIP), 2012, pp. 629-632, 11-13 April 2012.
24. Harrag, A.; Saigaa, D., Boukharouba, K.; Drif, M.; Bouchelaghem, A., "GA-based feature subset selection: Application to Arabic speaker recognition system," 11th International Conference on Hybrid Intelligent Systems (HIS), 2011, pp.383-387, 5-8 Dec. 2011.
25. Gao Wen-xi; Yu Feng-qin, "Feature dimension reduction based on genetic algorithm for mandarin digit recognition," 4th International Congress on Image and Signal Processing (CISP), 2011, vol.5, pp.2737-2740, 15-17 Oct. 2011.
26. Aggarwal, R.K.; Dave, M., "Application of genetically optimized neural networks for hindi speech recognition system," 2011 World Congress on Information and Communication Technologies (WICT), pp.512-517, 11-14 Dec. 2011.
27. Selouani, S., "Evolutionary discriminative speaker adaptation," IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), 2011, pp.164-168, 11-15 Dec. 2011.
28. Shing-Tai Pan; Ching-Fa Chen; Jian-Hong Zeng, "Speech recognition via Hidden Markov Model and neural network trained by genetic algorithm," International Conference on Machine Learning and Cybernetics (ICMLC), 2010, vol.6, pp.2950-2955, 11-14 July 2010.
29. Oudelha, M.; Ainon, R.N., "HMM parameters estimation using hybrid Baum-Welch genetic algorithm," 2010 International Symposium in Information Technology (ITSim), vol.2, pp.542-545, 15-17 June 2010.
30. Yuan Yujin; Zhou Qun; Zhao Peihua , "Vector Quantization Codebook Design Method for Speech Recognition Based on Genetic Algorithm," 2nd International Conference on Information Engineering and Computer Science (ICIECS), pp.1-4, 25-26 Dec. 2010.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

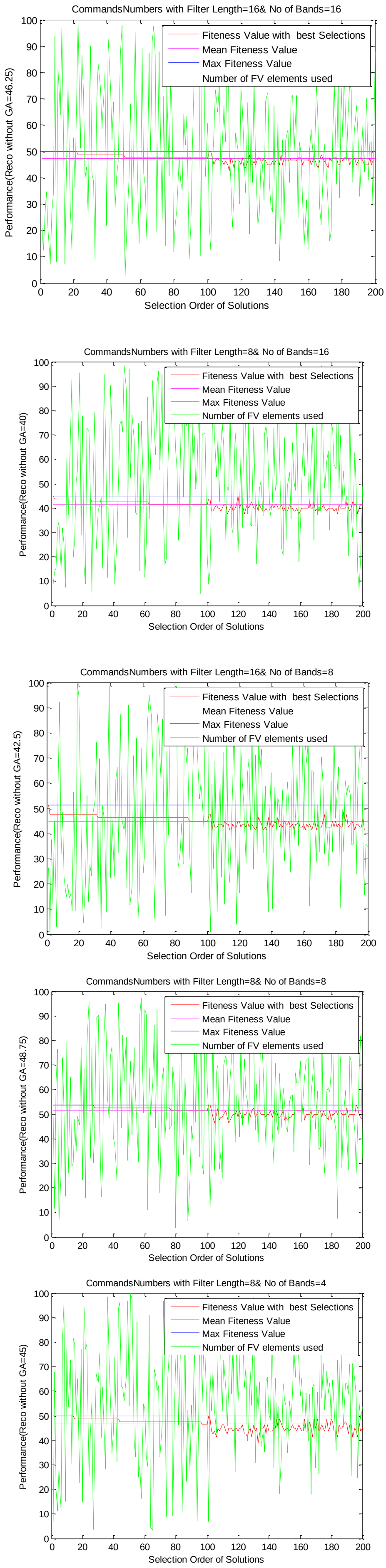

Fig.6. Performance of GA based solutions in last iteration (Commands+Numbers; Rows corresponds to the 5 parameters set of filter-bank)

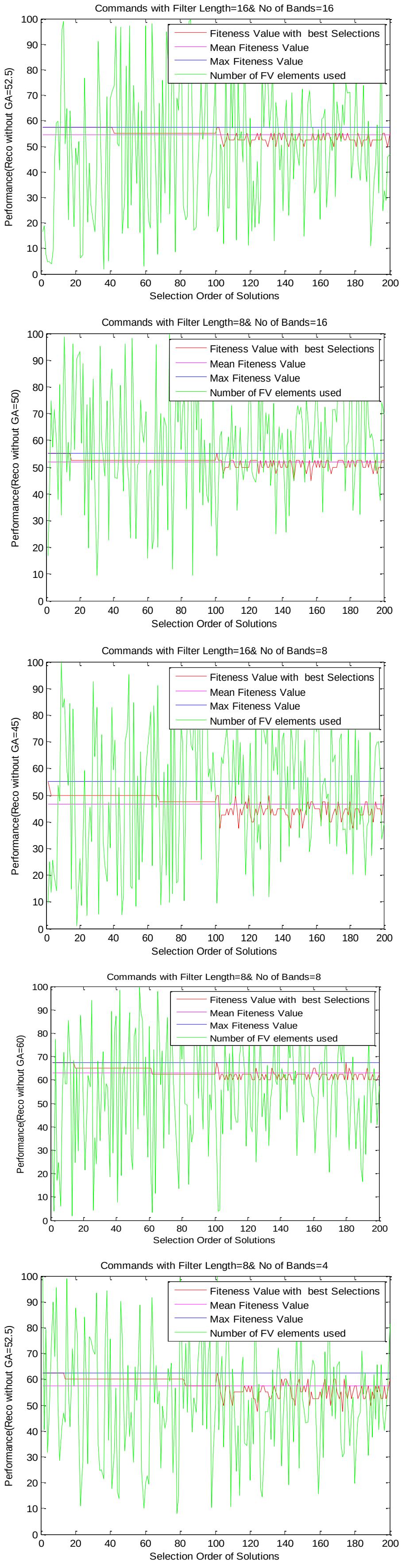

Fig.7. Performance of GA based solutions in last iteration (Commands; Rows corresponds to the 5 parameters set of filter-bank)

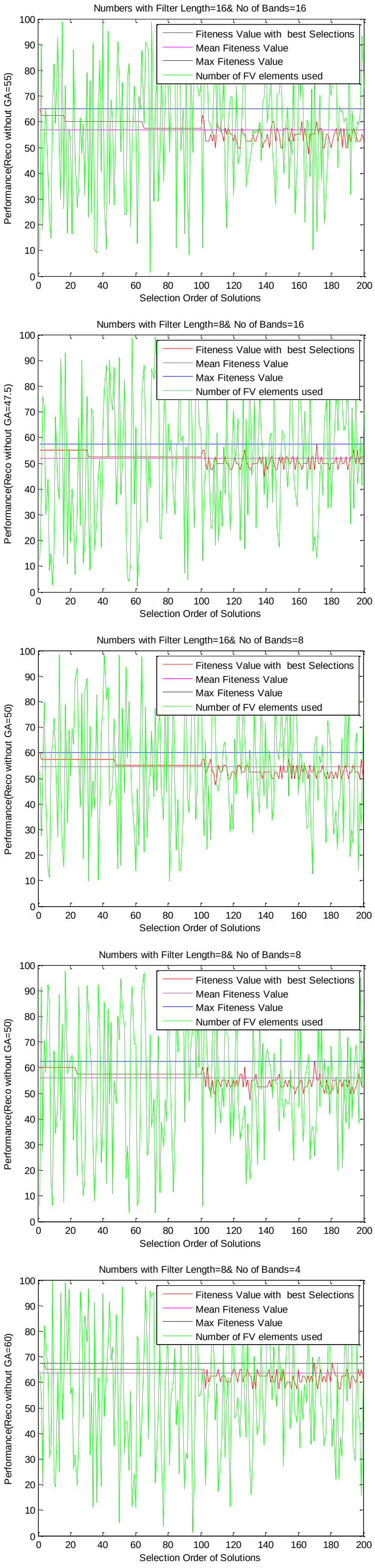

Fig..8 Performance of GA based solutions in last iteration (Numbers; Rows corresponds to the 5 parameters set of filter-bank)

AUTHOR PROFILE

Milind U. Nemade was born in Maharashtra, India 1974. He graduated from the Amaravati University, Maharashtra, India in 1995. He received M.E (Electrical) degree with specialization in Microprocessor Applications from M.S. University of Baroda, Gujrat, India in 1999. Now he is Associate Professor and Head at Department of Electronics and Telecommunication, K.J. Somaiya Institute of Engineering and Information Technology Sion, Mumbai, University of Mumbai, India. He started PhD study at Electrical Department, Faculty of Technology and Engineering, M.S. University of Baroda, Gujrat, India. He presented and published four papers in national conferences three papers in the proceedings of international conferences and two papers in international journals. His research interest includes speech and audio processing.

Prof. Satish K. Shah is a professor in the Electrical Engineering Department at Faculty of Technology, MS University of Baroda for **last Twenty Five** years. He is a fellow of IE(I) for past Fourteen Years and has also served as the member of Committee of Vadodara local center for more than SIX years in past. He has guided more than hundred projects at UG/PG level and completed a research project on DSP based Active Power filter sponsored by AICTE, New Delhi. He has written three books on Embedded System design/ Microprocessors/ Microcontrollers and presented/published more than 25 research papers in national/international conferences/Journals. He has attended and organized several seminars, workshops, and symposiums for UGC, AICTE, IETE, and MSU. He is a fellow of other technical associations such as: IETE, ISA and IEEE (NY) & ISTE. He has served as the member, Hon Secretary and Treasurer of their local executive committees for a span of six-eight years**.**