ijecs open access www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 2 Issue 10 October, 2013 Page No. 2903-2910

New technique for improving recognize letters E-set

saeed vandaki*, saman zahiri rad**, naser mehrshad***

Department of Electrical Engineering, Islamic Azad University, Gonabad Branch, Khorasan-e-Razavi, Iran,
*Vandakisaeed@yahoo.com, Saman.zahiri.rad@gmail.com, n.*mehrshad*@gmail.com*

Abstract
In any language, Spoken alphabet recognition as one of the subsets of speech recognition and pattern recognition has many applications. The purpose of audio signal processing, they are classified. Speech recognition is one of the issues in computer science and artificial intelligence, which seeks to identify a person based on the person's voice. Alphabet Recognition Speech recognition is below the branches. The different methods of feature extraction and classification, in this paper, a method combining these algorithms are trying to improve the English alphabet recognition. We are also leading to problems such as these problems can be noted that E-set, this collection contains the letters B, C, D, E, G, P, T, V and Z. The problem is similar to the set of waves vocal alphabet E that makes it difficult to recognize in all this is set in this paper by using MFCC feature extraction and SVM classification methods to achieve our desired results. In this paper, a method is said to have achieved 80% accuracy on data-set TI ALPHA.
Keywords:Mel-frequency cepstral coefficients, (MFCC), Support Vector Machines (SVMs).

1. INTRODUCTION

In 1990s has been much research in the field speech classification. Including applications can be classification speech, speech recovery applications in industrial problems, the classification of different classes of speech, video segmentation using speech detection, recognition and speaker verification using speech recognition systems, code Manufacturer of low bit-rate, classification, and music ... Named. A kind of speech classification system for the problem of pattern recognition and classification the most similar pattern recognition system is composed of three stages. Firstly, suitable feature vectors extracted speech file short interval is called the frame. Overall, the speech signal interval properties, is assumed to be static. Secondly, the classification of speech components frames to 1 second, for example based on features extracted from the speech frames or time intervals based on the second. The classification of the problems in pattern recognition is performed by different methods. The third step, is usually to enhance the system's accuracy experimental methods have been used as a final processing.
In one of the most applicable wavelet transformation and mathematical processing, particularly in the areas of signal processing and in wavelet analysis, the short-time Fourier transform is similar. A function of the desired signal (wavelet) multiplied by the fact that the wavelet transform of the window function will be the same. Several studies on classification of speech files using wavelet transform are performed. Binary wavelet transform has been proposed algorithm shows that the proposed method is simple and yet accurate noise is the noise. Frequency signals of wavelet able to have best of configuration.
A binary support vector machine classifier is two classes separated by a linear boundary. The advantages of support vector machine have separate classes according to their distribution. In training stage, the number of feature vectors for a known class, the optimal linear separation boundary between two classes using all the vectors of the training defined and an optimization algorithm, the numbers of educational classes, the boundaries have created. The educational examples of are called support vectors. The educational of support vector is considered as the minimum distance to the boundary separating the two classes and using them to optimize the separation of a boundary line separating the two classes is achieved.
Because of the ability of neural networks in modeling of nonlinear and high-speed response functions in many engineering problems are used. One of the most efficient and intelligent systems is considered most extensive. Perceptron neural network modeling of advanced high strength due to non-linear functions and basic structure of neural networks is one of the most prolific. These papers to classify the feature vectors are extracted by wavelet transform have been used.
In this paper, the classifications of speech, the feature extraction based on speech components are analyzed with 39 features. Then the next step to help support vector machine classification method and a multilayer perceptron neural network, the feature vectors and finally, the test data is the error have been investigated. The results show improvements in fault classification method of support vector machine classification method is multilayer neural network.

The English language consists of 26 characters of which 5 are vowels and 21 are consonant. Although automatic speech recognition performance has improved substantially over the past several years, automatic methods are still far inferior to most human listeners for most tasks. Performance is quite good for limited tasks with clean speech, but High acoustic similarities may cause difficulty in classification while low acoustic similarities causes ease to discriminate among classes for speech recognition systems. An alphabet set which has been detected to be the most confusable for speech recognition is the so called E-set letters. Human listeners are generally able

to better discriminate members of the "E" set (b, c, d, e, g, p, t, v and z) than are the best machine algorithms. The E-set letters are so called because all the nine letters share the same /iy/ (as in 'E') vowel at the back end of its oration [4, 5]. These set of letters may be difficult to detect because the distinguishing sound is short in time and low in energy [5]. In this paper, to overcome this problem, in this letters, given that the original pronunciation of the word, with the excess energy is different, we've removed the excess portion to resolve this issue. In this paper also system performance is studied normalization of feature vectors. We'll

then feature extraction by Mel-frequency cepstral coefficients, (MFCC) [2]; finally, we perform classification by Support Vector Machines (SVMs) [3]. A lot of work done in the field of signal processing, audio processing tasks such as speech recognition and speaker on mental state, things are taken for processing Alphabet, article [1] has to recognize the alphabet, In this paper we see that use the their desired method, detects percent, Between 62.28% to 70.49%, which is set for E-set Here we use the above method, the detection percentage of 80% to obtain.

This paper is organized as follows. MFCC feature extraction, The Support Vector Machine (SVM) is used as classifier, work method, finally results and discussion also CONCLUSION.

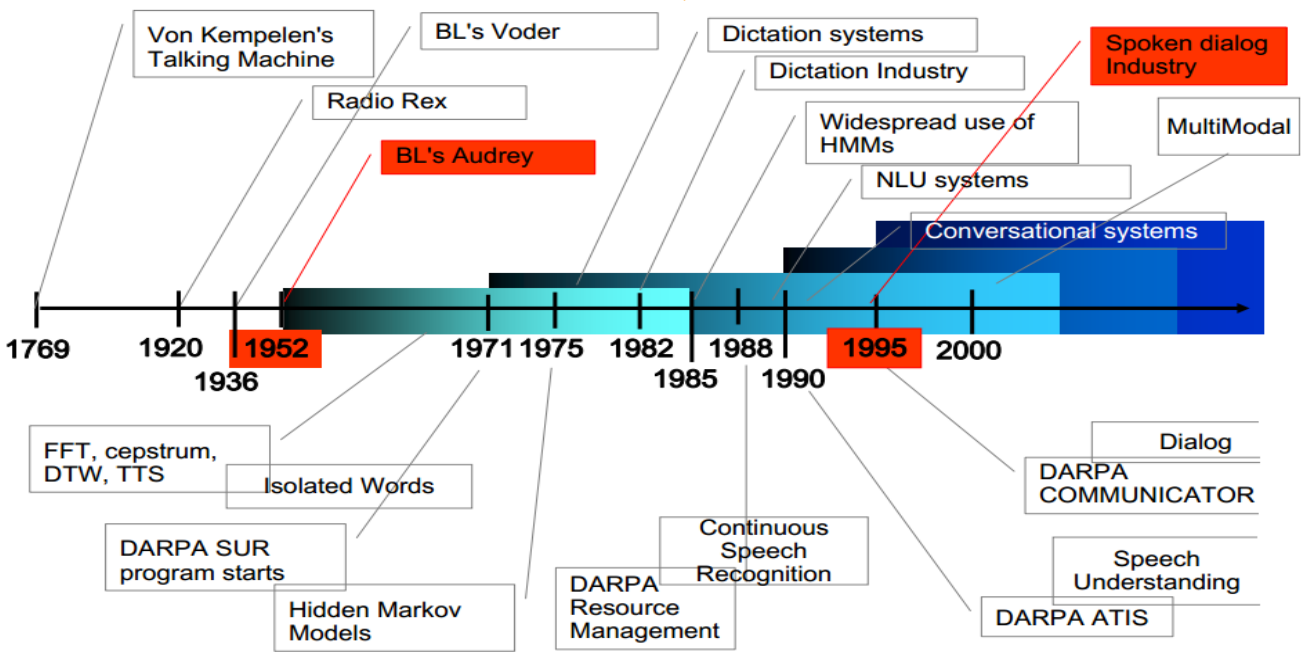


Fig. 1.History of Speech Recognition

2. Feature Extraction (MFCC)

In feature extraction all of the basic speech feature extracted may not be helpful and essential for speech recognition system [20]. If all the extracted features gives as an input to the classifier this would not guarantee the best system performance which shows that there is a need to remove such a useful features from the base features. Therefore there is a need of systematic feature selection to reduce these features. So that for this system we have used only feature that are MFCC. The Mel-frequency cepstral coefficients (MFCCs) introduced by Davis and Mermelstein is perhaps the most popular and common feature for SR systems [8]. A block diagram of the structure of an MFCC processor is shown in the figure 1. The speech input is typically recorded. Frame Blocking is the process of segmenting the speech samples obtained from analog to digital conversion (ADC) into a small frame with the length within 256 samples (16ms). Hamming windowing is employed to window each individual frame 85 samples (5.3ms). The concept here is to minimize the spectral distortion by using the window to taper the signal to zero at the beginning and end of each frame. Fast Fourier Transform converts each frame of N samples from the time domain into the frequency domain. Filtering is out the center frequencies of the 20 triangle band-pass filters corresponding to the Mel frequency scale of individual segments. Estimating is inversely the IDFT to get all 12-order MFCC coefficients. The result of windowing is the signal.

$$y_1(n) = x_1(n)h(n), \quad 0 \le n \le N-1 \qquad (1)$$

Typically the Hamming window is used, which has the form:

$$h(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \le n \le N-1 \qquad (2)$$

$$x_k = \sum_{n=0}^{N-1} x_n e^{-i\,2\pi kn/N}, \quad k = 0,1,2,...,N-1 \qquad (3)$$

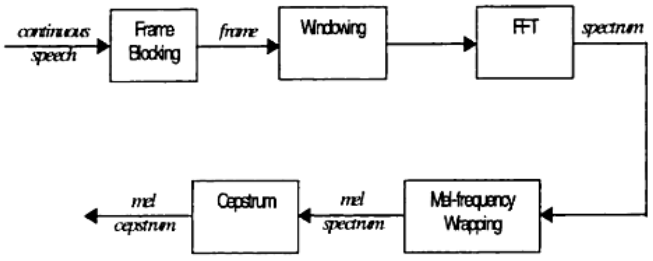
The result after this step is often referred to as spectrum. The Mel-filter bank is a triangular band pass filter which is equally spaced around the Mel-Scale. The mapping between real frequency (Hz) and Mel frequency is given by the following equation as

$$f_{mel} = 2595.\log\left(1+\frac{f}{700}\right) \qquad (4)$$

Finally, we can convert them to the time domain using the Discrete Cosine Transform (DCT).

$$C[n] = \sum_{K=0}^{M-1} \log\left(S[k]\right)\cos\left[n\left(k+0.5\right)\frac{\pi}{M}\right] \qquad (5)$$

Where n=1, 2….K



❖ *In summary, the following configuration is MFCC:*

Pre-emphasis coefficients = -0.95
Frame size = 256 samples (16ms)
Frame overlap = 85 samples (5.3ms)
Number of Triangular band pass filters = 20
Number of MFCCs = 12

3. CLASSIFICATION (SVM)

The Support Vector Machines (SVMs) present one of kernel-based techniques. Support vector machines (SVMs) are receiving increasing attention as a tool for speech recognition applications due to their good popularization properties [9, 10, 11, 12, 13, and 14]. Support Vector Machines are a new approach to classification standards and have recently attracted great concern in the methodical association, specifically in the areas of machine classification, regression and learning. SVM (Support Vector Machine) theory was first presented by Vapnik (1992). The SVM maps the input space to a high dimensional space. By calculating an optimal separating hyper plane in this new space, the SVM learns the border between Fast Fourier Transform converts each frame of N samples from the time domain into the frequency domain.

areas belonging to both classes. The separating hyper plane is chosen to maximize separation interval between the closest training samples. This approach is strongly linked to the theories of statistical learning and constructional risk minimization [Haykin, 2001]. The SVM may be equally used to separate non-linearly separable patterns. The SVM is train according to labeled features. The SVM kernel functions are used in the training process of SVM. Binary classification can be viewed as the task of separating classes in feature space. SVM is a binary classifier, but it can also be used as a multiclass classifier. The confusion matrix represents the percentage of accurate classification and misclassification for the given class. Transforming the original feature set to a high dimensional feature space by using the kernel function is the main thought behind the support vector machine (SVM) classifier, which leads to get optimum classification in this new feature space. A binary SVM classifier estimates a decision surface that jointly maximizes the margin between the two classes and minimizes the misclassification error on the training set. For a given training set $(x1...xp)$ with corresponding class labels $(x1...xp)(y1, ..., yp)$, $yi \in \{+1, -1\}$, an SVM classifies a test point x by computing a score function,

$$h(x) = \sum_{i=1}^{p} \alpha_i y_i k(x, x_i) + b \qquad (6)$$

Where λi are the Lagrange multiplier corresponding to the training sample, xi, b is the classifier bias – these parameters are optimized during training – and K is a kernel function. The class label of x is then predicted as sign (h (x)). While the simplest kernel $K(x, \tilde{}\,x) = hx, \tilde{}\,xi$ produces linear decision boundaries, in most real classification tasks the data is not linearly separable.
Overall, the different aspect of SVM approach is as follows.

❖ the following Binary SVM:

- Find optimal hyper plane
- Minimize cost function:

$$\Phi(w,\xi)=\frac{1}{2}w^T w+C\sum_{i=1}^{N}\xi_i \qquad (7)$$

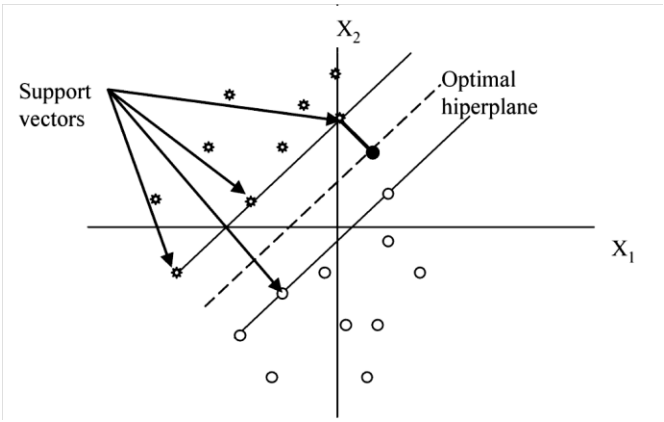


Fig. 2. Binary SVM

- linear separable patterns

$$d_i\left(w^T x_i+b\right)\geq 1 \qquad (8)$$

- non-separable patterns

$$d_i\left(w^T x_i+b\right)\geq 1-\xi \qquad (9)$$

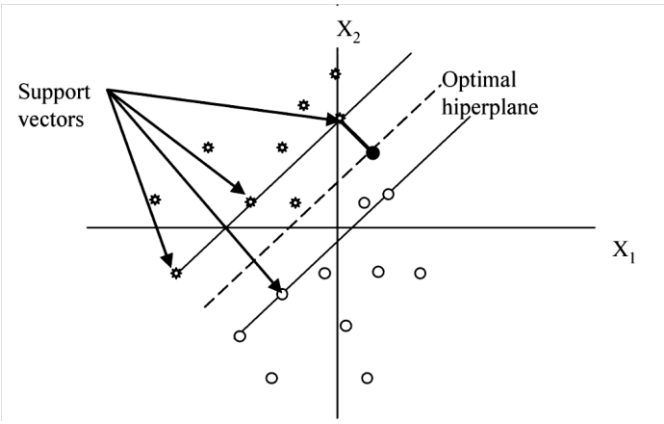


Fig. 3.linear separable patterns

❖ The following Nonlinear mapping: Inner product kernel:

- Polynomial

$$k\left(x,x_i\right)=\varphi^T\left(x_i\right)\varphi\left(x\right)=\sum_{j=0}^{m_1}\varphi_j\left(x\right)\varphi\left(x_i\right) \qquad (10)$$

- radial basis function
- two layer perceptron

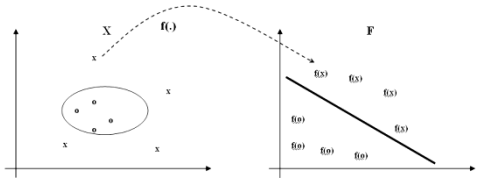


Fig. 4. Nonlinear mapping

❖ The following Multi class SVM:

Algorithms:

- one against all
- one against one (max wins)
- DAGSVM

4. Normalization

There are many ways to normalized. The answer can also be used where it can have a huge impact. In this paper the system performance is studied normalization feature vectors. Thus, for each component of the feature vector of zero mean value and variance are the same. For the training and testing feature vectors, each component of a file, the average low and training components of the vectors are divided by the standard deviation of the component. The normalization method in all experiments thus raises the accuracy of system, feature vectors are normalized in this way. Equation (6) shows normalized to each component.

$$x_i^{ij}=\frac{x_i^{j}-\mu_i}{\sigma_i} \qquad (11)$$

We've used this normalized after feature extraction, which could be the beginning of the input data, use it, but its use in this episode was better. We've used this normalized after feature extraction, which could be the beginning of the input data, use it, but its use in this episode was better.

5. Principle Component Analysis (PCA)

Main components of analysis by multivariate data analysis methods are one of the main objectives are to reduce the dimension of the problem. One of the most important applications of principal component analysis is in the regression. Using principal component analysis can be a lot of explanation variables (independent variables) correlated with a limited number of new explanation variables are uncorrelated the main components have to be replaced. This problem of will be reduced only after some time, but before it does not linear.

Assume that $\underline{X}=(X_1,X_2,...,X_P)^T$ agivennonnegativerandom vectorwithcovariance matrix $\sum$ and $\lambda_1\geq\lambda_2\geq...\geq\lambda_P\geq 0$ Eigenvaluesis $\sum$. So $a_1,a_2,...,a_P$ eigenvalues is respectively correspondingorthogonal $\lambda_1,\lambda_2,...,\lambda_P$. Variables $Y_1,Y_2,...,Y_P$ definedthemain components as following:

$$Y_1=a_{11}X_1+a_{21}X_2+...+a_{P1}X_P$$
$$Y_2=a_{12}X_1+a_{22}X_2+...+a_{P2}X_P$$

$$Y_P=a_{1P}X_1+a_{2P}X_2+...+a_{PP}X_P$$

Y_i The principal component is i and vector $\underline{Y} = (Y_1, Y_2, ..., Y_P)^T$, vector is called PCA.

6. Description E-set

The E-set that include words letters B, C, D, E, G, P, T, V and Z. These words are common in /iy/ (as in 'E') vowel voices, namely voices of addition to main letter in this letters there are additional noise, which makes it difficult to distinguish the words. To better understand is your issue, comparing the two letters A and B. See the following figure. Of course we know letter B part of E-set, but not letter AN E-set.

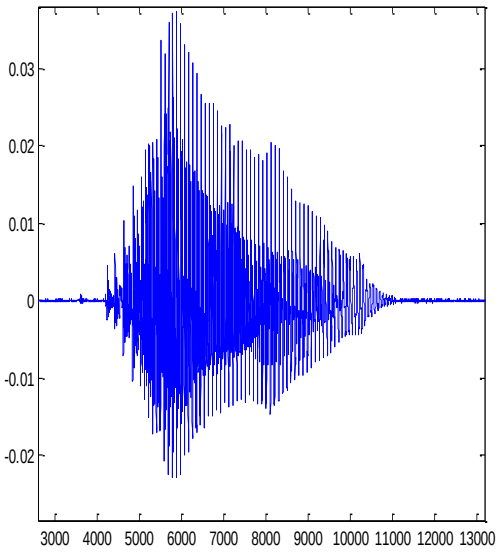


Fig. 5. Display the letter A

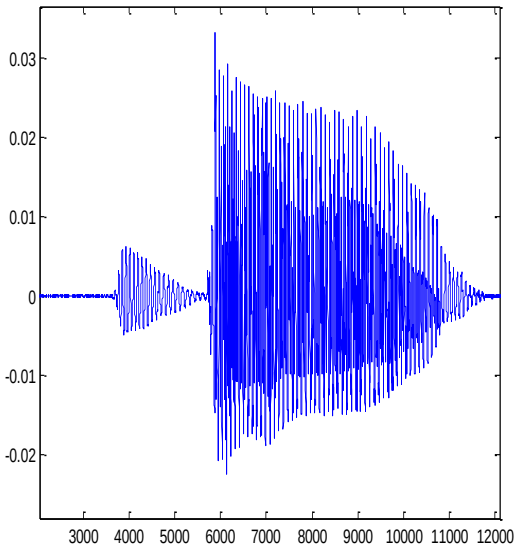


Fig. 6. Display the letter B

If you like the figure the above, you'll notice the difference between the E-set and other letters.

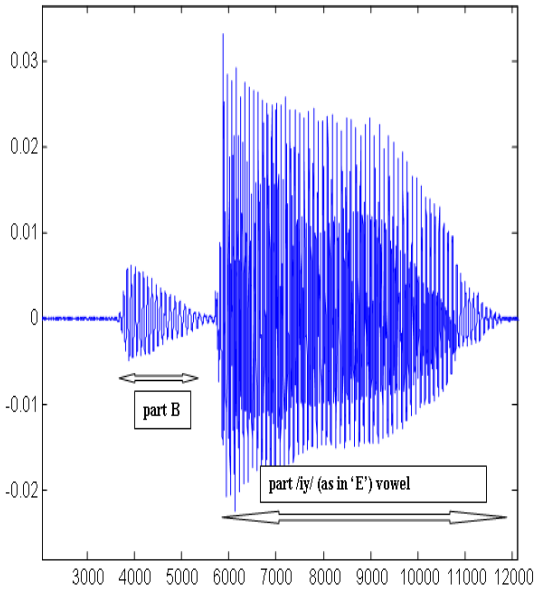


Fig. 7. In the following figure, the letter B parts are shown separately.

We must also remove the silence; remove this part to have a better diagnosis. After removal of silence as well as an additional section at the end of the form is ready for work.

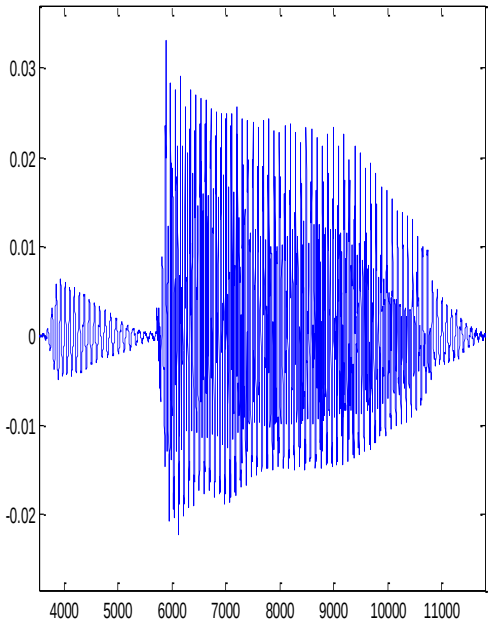


Fig. 8. After removal of the silence

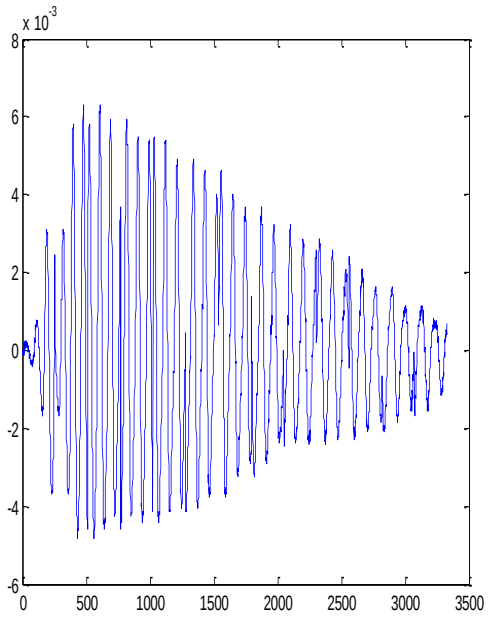


Fig. 9. After removal of the silence and
The /iy/ (as in 'E') vowel

7. METHODOLOGY

In this article, I realized that it must first prepare the data for work. First, we propose to eliminate silent data. In order to eliminate redundant data, the silences of the data are deleted. Then E-set dataset is difficult to resolve the additional part of this collection is that it will remove the same /iy/ (as in 'E') vowel at the back end of its oration. Then turn the feature extraction MFCC is the way it is described in the paper. However, the feature was normalized by the time they are told. In the final stage, which is classified as part of his SVM method for training and testing, we use our data.

In this paper, some methods have been used to improve character recognition. First, for convenience, we eliminate all the letters from the silent feature extraction. After eliminate of silent characters E-set to eliminate the common problem we recognize that this case is incorrect. Currently the data is ready for feature extraction, feature extraction is performed by MFCC method described above has been completed; the MFCC parameters are selected empirically. After the shift to data classification is done by SVM methods, values, SVM has great influence on the results, so we use a genetic algorithm to determine the correct values. Genetic algorithms for the SVM error percentage choose new values. To find the best values, SVM C and γ values are important, and classification is performed. However, before applying for data classification methods to reduce the amount of PCA is used to help lower of high performance computing. In this paper, we experiment various methods for speech recognition, and combining them together, we get the desired result at the end is the best method of setting its parameters. So this place is left open for future researchers interested in using the results of this paper will continue this way.

8. Speech Data

In this paper the speech data (wave files '.wav') are taken from the TI46 database isolated alphabet called TI ALPHA. The TI ALPHA consists of eight male and female speakers. The files were further segregated into training and testing sets. For training, there are 16 patterns for each alphabet A to Z while for testing there are 10 patterns for each alphabet [1].

9. RESULTS

In order to get the optimal solutions, we adjust the parameters in the MFCC and the SVM reached by repeating the following appropriate parameters.

- The following parameters MFCC:

Lowest Frequency = 1.333333e+02
Linear Filters = 13
Linear Spacing = 6.666667e+01
Log Filters = 23
Log Spacing = 1.071170e+00
FFT Size = 256
Cepstral Coefficients = 12
Window Size = 128

- The following parameters SVM:

Parameter C = 2
Parameter value γ = 3.000000e-02

We understand that by changing the parameters to achieve better results, in the following figure you can see answer changes with respect to changes in the parameters C and γ in the SVM method.

Parameter by setting we got good results. The results 80% for E-set letters, the result is very suitable for this letters below you can see the result.

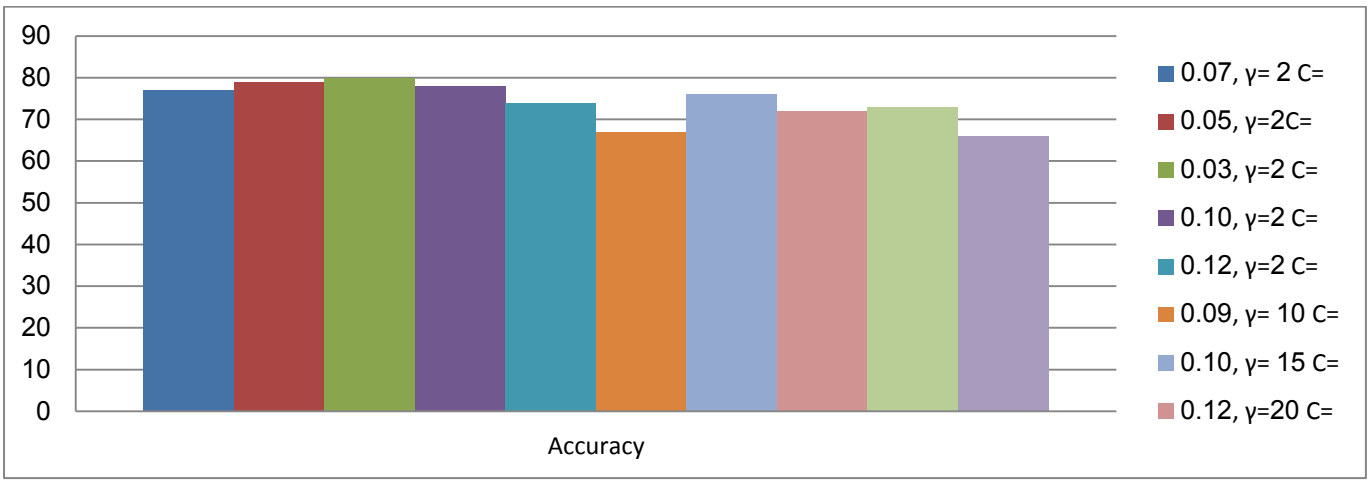


Fig. 10. Changes in response to changes in C and γ

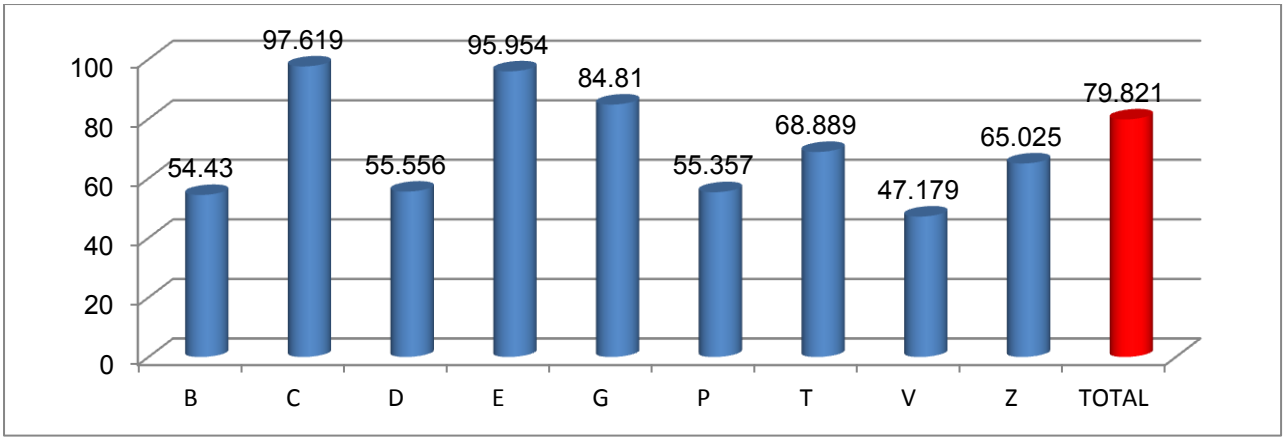


Fig. 11.Chart of results

9. CONCLUSION AND FUTURE WORK

The results show that the parameter setting has a significant impact on improving the solution, the better the response of the normalized, We see combine SVM with MFCC for us to bring good results, We can adjust parameters for future work on the put of intelligent algorithms And so we have changes in the extraction of features to help improve results.In the future, this problem can be a lot of work to improve the methods used, classification methods can be optimized by the optimization algorithm used., Such as genetic algorithms, and fuzzy methods to improve the classification algorithms, as well as of these methods can also be used to select the best features., or a combination of these methods can be used. Hopefully this article useful

REFERENCES

[1] T.B. Adam, Md Salam, "Spoken English Alphabet Recognition with Mel Frequency Cepstral Coefficients and Back Propagation Neural Networks," presented at the International Journal of Computer Applications (0975 – 8887) Volume 42– No.12, March **2012**.

[2] ChadawanIttichaichareon, SiwatSuksri, and ThaweesakYingthawornsuk, "Speech Recognition using MFCC," presented at the International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012) July 28-29, **2012**Pattaya (Thailand).

[3] Aamir Khan, Muhammad Farhan, Asar Ali, "Speech Recognition: Increasing Efficiency of Support Vector Machines," presented at the International Journal of Computer Applications (0975 – 8887) Volume 35– No.7, December **2011**.

[4] M. D. Ibrahim, A. M. Ahmad, D. F. Smaon, and M. S. H. Salam, "Improved E-set Recognition Performance using Time-Expanded Features," presented at the Second National

Conference on Computer Graphics and Multimedia (CoGRAMM), Selangor, Malaysia, **2004**.
[5] K. J. Lang, A. H. Waibel, and G. E. Hinton, "A Time-Delay Neural Network Architecture for Isolated Word Recognition," Neural Networks, vol. 3, pp. 23-43, **1990**.
[6] MontriKarnjanadecha, and Stephen A. Zahorian, "ROBUST FEATURE EXTRACTION FOR ALPHABET RECOGNITION," department of Electrical and Computer Engineering Old Dominion University Norfolk, VA 23529, USA.
[7] Mr.SaurabhPadmawar, Prof. Mrs. P.S. Deshpande, "Classification of Speech Using Mfcc and Power Spectrum," presented at the Research and Applications (IJERA) ISSN: 2248-Vol. 3, Issue 1, January -February **2013**, pp.1451-1454.
[8] D. O'Shaughnessy, "Invited Paper: Automatic Speech Recognition: History, Methods and Challenges," Pattern Recognition, vol. 41, pp. 2965-2979, **2008**.
[9] J. Yousafzai, Z. Cvetkovi´c, P. Sollich, and B. Yu, "Combined Features and Kernel Design forNoise Robust Phoneme Clas-sification Using Support Vector Machines," To appear in the IEEE Trans. ASLP, **2011**.
[10] P. Clarkson and P. J. Moreno, "On the Use of Support Vector Machines for Phonetic Classification," Proc. ICASSP, pp. 585–588, **1999**.
[11] A. Ganapathiraju, J. E. Hamaker, and J. Picone, "Applications of Support Vector Machines to Speech Recognition," IEEE Trans. Signal Proc., vol. 52, no. 8, pp. 2348–2355, **2004**.
[12] S. E. Kr¨uger, M. Schaff¨ner, M. Katz, E. Andelic, and A. Wen-demuth, "Speech Recognition with Support Vector Machines in a Hybrid System," Proc. INTERSPEECH, pp. 993–996, **2005**.
[13] V. N. Vapnik, the Nature of Statistical Learning Theory, Springer-Verlag, New York, **1995**.
[14] J. Louradour, K. Daoudi, and F. Bach, "Feature Space Mahalanobis Sequence Kernels: Application to SVM Speaker Verification," IEEE Trans. ASLP, vol. 15, no. 8, pp. 2465–2475**, 2007**.
[15] J. Yousafzai, Z. Cvetkovi´c, "A High-Dimensional Subband Speech Representation and SVM Framework for Robust Speech Recognition," The authors are with the Division of Engineering and theDepartment of Mathematics at King's College London.
[16] DavoodMahmoodi, HosseinMarvi, Ali Soleimani, FarbodRazzazi, Mehdi Taghizadeh,MarziehMahmoodi," Age Estimation Based on Speech Features and Support Vector Machine," presented at the computer science and electronic engineering conference (CEEC) 3rd **2011**.
[17] SANTOSH GAIKWAD, BHARTI GAWALI AND MEHROTRA S.C, "GENDER IDENTIFICATION USING SVM WITH COMBINATION OF MFCC," presented at the Advances in Computational Research ISSN: 0975-3273 & E-ISSN: 0975-9085, Volume 4, Issue 1, 2012, pp.-69-73.
[18] AiniHafizahMohdSaod, DzatiAthiarRamli, "Preliminary Study on Classification of Apraxia Speech using Support Vector Machine," presented at the Australian Journal of Basic and Applied Sciences, 5(9): 2060-2071, **2011** ISSN 1991-8178
[19] Sujata B. Wankhade, PritishTijare ,YashpalsingChavhan," Speech Emotion Recognition System Using SVM AND LIBSVM," presented at the International Journal Of Computer Science And Applications Vol. 4, No. 2, June July **2011** ISSN: 0974-1003
[20] Paolo Baggia,"Speech Technologies and Standards," DIT Seminars, Povo, Trento June 7th, **2006**.