Indian Journal of Science and Technology, *Vol 9(1), DOI: 10.17485/ijst/2016/v9i1/85763, January 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Fuzzy Logic based NAM Speech Recognition for Tamil Syllables

V. Prabhu[1*] and G. Gunasekaran[2]

[1]Department of Electronics and Communication Engineering, MAHER Deemed University, Chennai – 600 078, Tamil Nadu, India; Prabhu.cvj@gmail.com
[2]Meenakshi Engineering College, Chennai – 600078, Tamil Nadu, India

Abstract

Background/Objectives: In this paper, we propose the speech recognition of tamil words from dump mute persons with NAM sensor. **Methods/Statistical analysis:** The NAM sensor collects the speech signal of the tamil words frequently used in day-to-day life. These frequently used words are classified using fuzzy logic, based on syllables and the words are recognized. **Findings:** Until now, the NAM speech recognition was implemented only for vowels. From the 50 sample of speech sound, we attained 85% of accuracy in recognition. **Application/Improvements:** System helps in recognizing the tamil words from dump mute persons.

Keywords: Fuzzy Logic, NAM Sensor, Syllable, Speech Recognition, Tamil Words

1. Introduction

A quietly uttered speech, received through the body tissue, is the Non- Audible Murmur (NAM). A special acoustic sensor i.e., NAM microphone sensor is attached behind the communicator's head which could sense quiet sounds that are inaudible to other listeners near the talker. The first ever microphone was used by medical doctors to look after patient's internal organ sounds well known to all is a stethoscope. For the recognition of NAM speech automatically stethoscope microphone were used. Here the author describes two algorithms in the form of mixture based Gaussian speech recognition system for reducing complexity in the computation of likelihood. For this the system consider partial distance elimination and nearest-neighbor search method for the computation of likelihood. One algorithm is used to get the best score in state mixture known as BMP (Best Mixture Prediction) that improves PDE technique speed. The other algorithm is a feature component reordering used for final scoring for mean space vector and feature dimensions. As compared to the existing system of this likelihood computation, these two algorithms with PDE technique improve the efficiency to 29.8% by reducing the speed of computation for likelihood and maintains same accuracy without further requirement of memory space[1]. In this paper, the author has conceived an approach of dynamic programming to ease the search problem. Also reviewed about the statistical approach of the search problem and requirements of acoustic and language models of search space results for statistical approach. The idea is further extended from the baseline algorithm to various dimensions. A new method of forming structure by the search space is also dealt with the pronunciation lexicon. To handle large vocabulary, organized lexical prefix tree is used. The combination process of time-synchronous and construction process for recognition has been summarized. In the pruning operation, the language model look-ahead is integrated for the increment of beam search efficiency. Word graph is also included to generate alternative sentences. The model is demonstrated by 64k model with various concepts of search[2]. The paper enumerate, word hidden Markov model for real time speech recognition to achieve high accuracy. Since the computational model uses big data, the processing time is also long.

**Author for correspondence*

2. Proposal Work

A VLSI system with scalable architecture is implemented to achieve high speed operations. The use of parallel computation reduces the overall time consumed in processing the data and recognize vocabulary. The above concept was implemented on a CMOS cell library with speech analysis and noise robustness to check the effectiveness of the system. Results showed that the processing time was 56.9 μs/word at 80 MHz operating frequency[3]. A VLSI chip was designed to process over 5000 word as a standalone speech recognition. The chip is programmed with a HMM based recognition algorithm, comprises of Viterbi beam and emission probability units. A host processor is used to compute vector for speech recognition. The data is moved to external DRAM to improve the memory space in SRAM. To improve the efficiency of read/write consecutive data a custom DRAM controller module is designed. Results showed that the real time factor was about 0.77 for SDRAM and 0.55 for DDR SDRAM[4]. Hardware based real-time large-scale vocabulary speech recognizer demands high memory bandwidth. We have succeeded in development of a FPGA (Field Programmable Gate Array) based 20000-word speech recognizer availing efficient dynamic random access memory (DRAM) access. This system consists of functional blocks for hidden-Markov model based speaker independent continuous speech recognition: Feature extraction, emission probability computation, intraword, and interword Viterbi beam search. Soft-core based CPU software conducts the feature extraction, while the other functional units are implemented using pipelined and parallel hardware blocks. To minimize the number of memory access operations, we use various techniques like bit-width reduction of the Gaussian parameters, multi-frame computation of the emission probability, and two-stage language model pruning. We also use a customized DRAM controller that assists different access-patterns optimized for each functional unit of the speech recognizer. The speech recognition synthesized the Virtex-4 FPGA and it functions at 100 MHz. In Nov 1992, 20000 experimental result, test set demonstrates that the developed system runs 1.39 and 1.52 times faster than real-time with the trigram and bigram language models respectively[5]. We outline our preliminary experiments on building dynamic lexicons for native speaker conversational speech and for foreign-accented conversational speech. Our objective is to develop a lexicon with a group of pronunciation for each word. In this, the probability distribution dynamically computes over pronunciation. The set of pronunciations are derived from hand-written rules (for foreign accent) or clustering (for phonetically transcribed Switchboard data). The dynamic pronunciation probability considers specific characteristics of speaker and the factors like disfluencies, phonetic context, language-model probability and sentence position. This work is still in progress[6]. In this paper, we discussed various types of universal speaker models for NIST speaker recognition and verification performance and benchmarks. A MIT Lincoln Laboratory's describe the major elements in several NIST Speaker Recognition Evaluations (SREs), successfully implemented in Gaussian Mixture Model (GMM)-based speaker verification system. Alternative methods form a Bayesian adaptation to derive speaker models for speaker representation with greater ratio and effective GMMs for likelihood functions from a Universal Background Model (UBM) built a system. Finally, greatly improve verification performance and increase the usage of a handset detector and score normalization on NIST SRE corpora are presented[7]. A computing branch metrics that are combinations of autocorrelation and terms are constructed by calculating and storing autocorrelation and cross correlation components at an instant. Once calculated and stored, predetermined ones of the autocorrelation components and the cross correlation components are selected. The selected components inverse are predetermined and combined to produce a branch metric. Other predetermined combinations of the stored components of autocorrelation and cross correlation terms, or their inverse, are combined to produce other branch metrics at the same symbol instant. All branch metrics associated with the symbol instant can be calculated in this manner. The components of the autocorrelation terms and cross correlation terms associated with the next symbol instant are calculated and stored, and the process of generating the branch metrics are repeated[8]. In this paper, we briefly explained concept of Carlson and Fant and Chistovich and SA integration to demonstrate the analysis of both synthetic and speech that psychoacoustic and in vowel perception, namely the F1, F2' concept and complete speech analysis-synthesis system based on the PLP method. The auditory spectrum from the speech established psychoacoustic concepts based on the linear predictive analysis (PLP), uses to modeling spectrum of the low-order all-pole model and help drive the critical-band filtering, equal-loudness curve

pre-emphasis and Intensity-loudness root compression[9]. In this paper signal processing in speech recognition, techniques that are mostly used are presented. The flow of signal modeling is spectral shaping, analysis, parametric transformation and modeling. The important trends in speech recognition is examined. Namely the time derivative, transformation technique, signal parameter estimation have put together to form speech recognition. The signal processing components of these algorithms are reviewed[10]. Arabic, is the second most spoken language. But not much research has been done on this language in speech recognition system. This paper mainly investigates the vowels in standard Arabic. The vowels are analyzed. The differences and similarities are explored using CVC. A HMM is built to classify and analyze the performance of recognizer to understand the phonetic features. The samples are analyzed in both time and frequency domains and can be used for future speech recognition of Arabic language[11]. Automatic Speech Recognition (ASR) has made great progress with the development of digital signal processing software and hardware. However, despite of all these advances, machines cannot fulfill the performance of their human counter parts in terms of accuracy and speed, especially in case of speaker independent speech recognition. Therefore, today noteworthy portion of speech recognition research concentrates on problems in speaker-independent speech recognition. The reasons are its broad span of applications and restrictions on available techniques of speech recognition. This report briefly examines the signal modeling approach in speech recognition. Then it overviews the basic operation engaged in signal modeling. Further, it discusses the general analyzing techniques like temporal and spectral analysis of feature extraction in detail[12]. By applying 6-dB/octave spectral tilt and by adding white noise for speech degraded and ungraded speech, these were done along with the test compared between speaker independent and dependent isolation-word and connected recognitions test for several acoustic representation. The representation comprises the output of an auditory model, the cepstrum coefficient which is derived from the FFT- based mel-scale filter bank along with various weighting techniques applied to the coefficients, the measure of their rates of change in time for cepstrumcoefficient augmented, the filter bank output are used to derive linear discriminant functions and called IMELDA. Except in noise free connected word test, the cepstrum representatives outperformed, where it had a high insertion rate. Variances derived within the class are the best cepstrum-weighting scheme. The empirical adjustments is explained by its behavior which was found necessary with other schemes[13]. The efficiency of minimum distortion encoding for vector quantization is improved using a very simple method. For a full search vector quantize having a huge number of code words was reduced up to 70 percent in the number of multiplication was seen in the simulation, for a tree search vector it is about 25–40 percent. It can be seen that similar improvement can be seen in other vector quantization systems[14]. The state likelihoods computation is seemed to be an intensive task, in speech recognition system which is based on continuous observation density hidden[16]. The computation of the likelihoods was produced in an efficient way by the author defined by weighted sums of Gaussians. To identify the subset of Gaussian neighbors, this method uses a quantization of the input feature vector. It is shown that instead of computing the likelihoods of all the Gaussian neighbors, one is suppose to compute the likelihoods of only the Gaussian neighbors. By using various data bases the computation reduction can be obtained for significant likelihood, with only small loss of recognition accuracy[15].

3. Methodology

A unit of sequence organization forms the speech sound and each unit is termed as syllable. For example, the word cable comprises of two syllables namely 'ca' and 'ble'. The syllable nucleus is used to make a syllable (mostly vowel) with optional final and initial margins (consonants). Syllables are usually considered as phonological "building blocks" of words. The syllable is capable of influencing the rhythm of the language, its poetic meter, its stress pattern and its prosody. A word that consists of one syllable is called monosyllable. Similarly, two syllable words are called as syllable, three as trisyllable. Capturing the inaudible murmur (NAM speech) is able to achieve using a device called a stethoscope Non-Audible Murmur (NAM) microphone, which is seemed not to be heard by listeners near the talker. The device used here is based on stethoscope, which is used in medical field. By attaching the NAM microphone near the talker's ear, it is capable of collecting the NAM speech from the talker's body, and the speech recognition is done in a conventional way. A possible tool for sound impaired people, privacy and robustness to environmental noise are the advantages of the NAM microphone, when this system is applied in

the speech recognition system. The obtained signal from the NAM microphone is transferred to the computer for analysis process using a loudspeaker microphone holding 3.5mm headset jack holding an Omni-directional microphone directivity, sensitivity of -52 +/- 3dB and 200Hz to 8KHz frequency response. The obtained audio signal is used in Mat lab for analysis purpose where the ANFIS technique is used to distinguish different types of variation in the voice. The GENFIS1, GENFIS2 and GENFIS3 are the different techniques under ANFIS used in the data acquisition process. The GENFIS1 is used at the initial condition of the ANFIS training, GENFIS2 is used to provide initial FIS for ANFIS training when the available input is only one and GENFIS3 generates FIS using Fuzzy C- Means (FCM) is shown in Figure 1.

4. Syllable and Word Recognition

The primary analysis of the paper is on the obtained signals from the dumb people while murmuring. The murmur speech is captured using a special type of sensor called the Non-Audible Murmur (NAM) microphone, which is a stethoscope, and capable of sensing the changes in the vibration created while murmuring as the sensor is kept between the ear and the neck. The sound signals generated are analyzed on the basis of syllable words. The received NAM signal from the NAM microphone is recognized using a technique called ANFIS in Matlab. ANFIS is used to identify the member function parameters of single-output, which is a hybrid learning algorithm using Sugeno type Fuzzy Inference Systems (FIS). To train FIS membership function parameters to model a given set of input/output data it uses a combination of methods such as back propagation gradient descent and least-squares. The

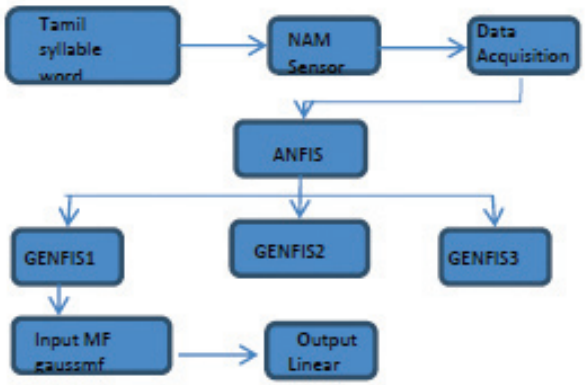

Figure 1. Block diagram.

signal is further analyzed using three more sub techniques of ANFIS namely GENFIS1, GENFIS2 and GENFIS3. The GENFIS1 uses a grid partition technique to generate the initial Sugeno-type FIS for ANFIS training. The input MF used here is gaussmf and the output obtained is of linear type. Using subtractive clustering the GENFIS2 generates sugeno-type FIS. The GENFIS# uses FCM clustering to generate a FIS. Here we use GENFIS for data analysis. The below graph are the different types tamil words and the data in it explains: A graph in general has huge number of data in it and to reduce the errors due to analysis of such data is large and to reduce it a data pre processing technique is used. Sequences of values are taken and those values are replaced by a central value and it is named as bin. The bin represents the complete set of several values. The potential predictive relationship of the obtained value can be identified by using a set of data called as training set. The training set are used to access the predictive relationship that is as the users we know what output will be obtained for the given input and the known function. A set of values has its own individual squares of error, which is averaged to obtain estimator called as Mean Squared Estimator (MSE), which represents the squared error loss. The (RMSD) is a technique used to measure the difference of the estimator to the actual value. The change in the above parameters during different sound can be used to predict its actual meaning.

The different syllable words used here for analysis is

Table 1. Tamil's equivalent English word

S.no	Tanglish	English
1	Ingaeva	Come here
2	Va	Come
3	Poo	Go
4	Naan	Me
5	Irangu	Get down
6	Enakuuthavungal	Help me
7	Enpeyar	My name

The output graph of the above tamil words are

1. Ingaeva

The above word is a two tamil word having a disyllabic and monosyllabic word. The train data and test data details are shown in Figure 2, Figure 3, Figure 4 and Table 1, 2.

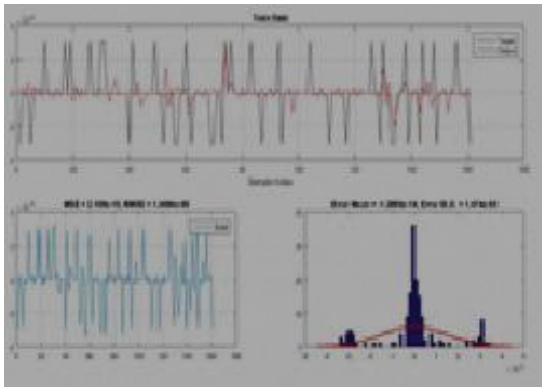

Figure 2. Train data of the word ingaeva.

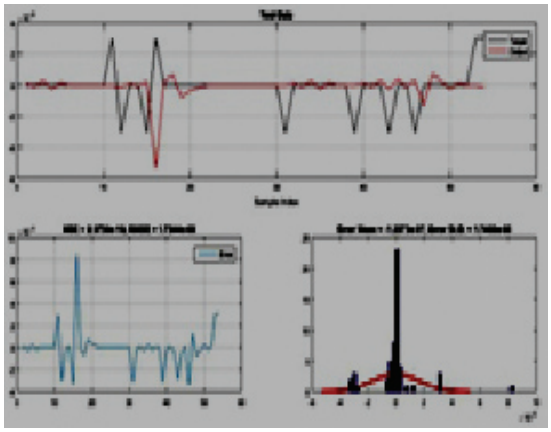

Figure 3. Test data of the word ingaeva.

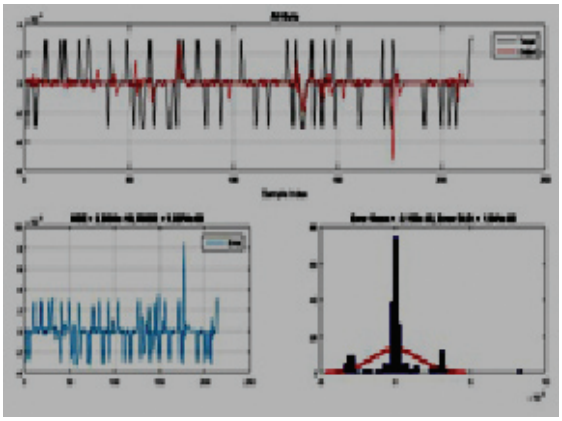

Figure 4. All data of the word ingaeva.

Table 2. Different data of the word ingaeva

Data type	MSE	RMSE	Error mean	Error St.D.
Train	2.154 e-10	1.4694 e-5	–1.2895 e-14	1.474 e-05
Test	2.973 e-10	1.724 e-05	–1.2573 e-07	1.7405 e-05
All	2.3636 e-10	1.5374 e-05	–3.158 e-08	1.541 e-05

1a. Come here

The above word is english word equivalent to the above-explained tamil word which carries two monosyllabic word. The train data and test data details are shown in Figure 5, Figure 6, Figure 7 and Table 3.

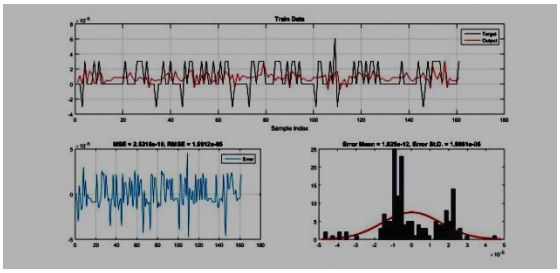

Figure 5. Train data of the word come here.

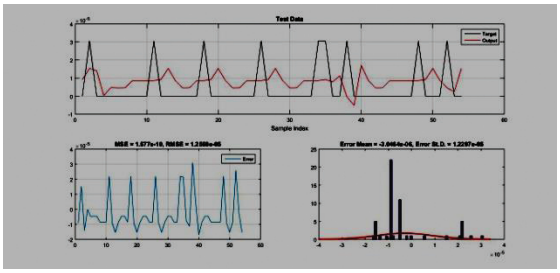

Figure 6. Test data of the word come here.

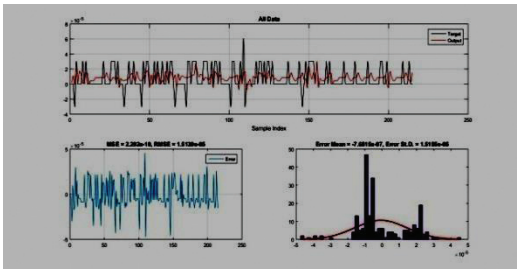

Figure 7. All data of the word come here.

Table 3. Different data of the word come here

Data type	MSE	RMSE	Error mean	Error St.D.
Train	2.5318 e-10	1.5912 e-05	–2.4592 e-13	1.3466 e-05
Test	4.3102 e-10	2.0761 e-05	1.8243 e-06	2.0875 e-05
All	2.432 e-10	1.5595 e-05	4.5819 e-07	1.5625 e-05

2. Va

The above word is a two tamil word having a monosyllabic word. The train data and test data details are shown in Figure 8, Figure 9, Figure 10 and Table 4.

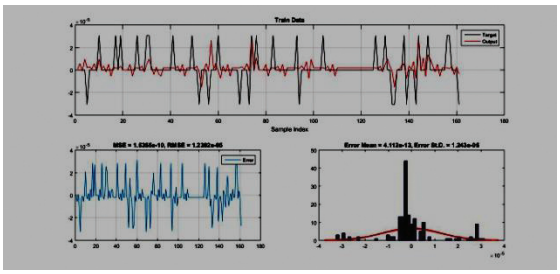

Figure 8. Train dta of the word va.

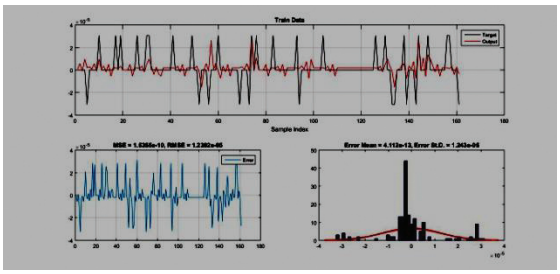

Figure 9. Test data of the word va.

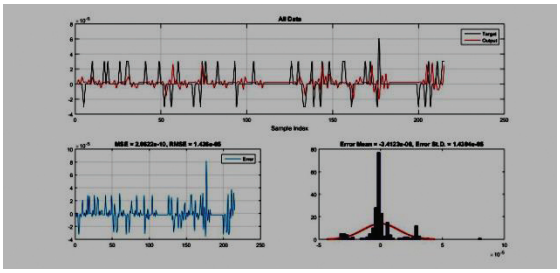

Figure 10. All data of the word va.

Table 4. Different data of the wordva

Data type	MSE	RMSE	Error mean	Error St.D.
Train	1.5355 e-10	1.2392 e-05	4.112 e-13	1.243 e-05
Test	3.6325 e-10	1.9059 e-05	–1.3586 e-07	1.9237 e-05
All	2.0622 e-10	1.436 e-05	–3.4123 e-08	1.4394 e-05

2a. Come

The above word is English word equivalent to the above-explained tamil word which carries one monosyllabic word. The train data and test data details are shown in Figure 11, Figure 12, Figure 13 and Table 5.

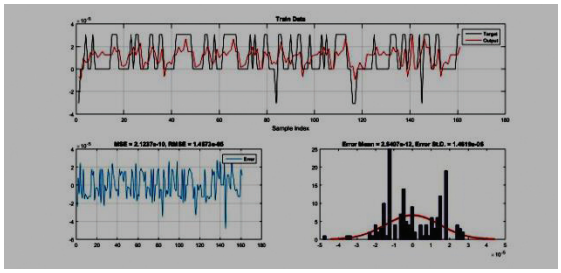

Figure 11. Train data of the word come.

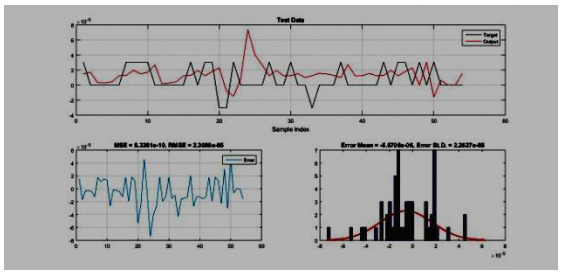

Figure 12. Test data of the word come.

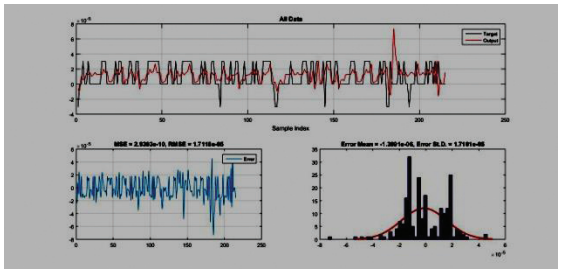

Figure 13. All data of the word come.

Table 5. Different data of the word come

Data type	MSE	RMSE	Error mean	Error St.D.
Train	2.1237 e-10	1.4573 e-05	2.5407 e-14	1.4619 e-05
Test	5.3351 e-10	2.3098 e-05	–5.5075 e-06	2.2627 e-05
All	2.9302 e-10	1.7118 e-05	–1.3991 e-07	1.7101 e-05

3. Poo

The above word is a two tamil word having a monosyllabic word. The train data and test data details are shown in Figure 14, Figure 15, Figure 16 and Table 6.

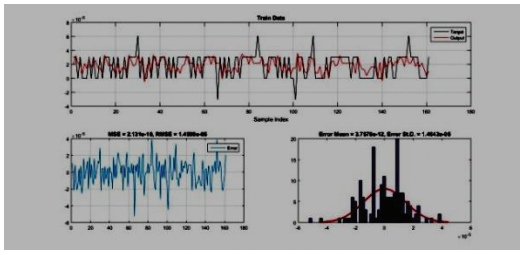

Figure 14. Train data of the word poo.

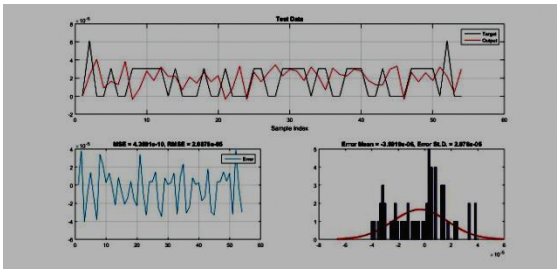

Figure 15. Test data of the word poo.

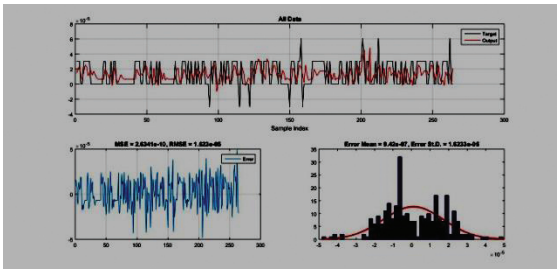

Figure 16. All data of the word poo.

Table 6. Different data of the word poo

Data type	MSE	RMSE	Error mean	Error St.D.
Train	2.131 e-10	1.4598 e-10	3.7578 e-12	1.4643 e-05
Test	2.6341 e-10	1.623 e-05	9.42 e-07	1.6233 e-05
All	4.359 e-10	2.0878 e-05	–3.5919 e-06	2.076 e-05

3a. Go

The above word is english word equivalent to the above-explained tamil word which carries one monosyllabic word. The train data and test data details are shown in Figure 17, Figure 18, Figure 19 and Table 7.

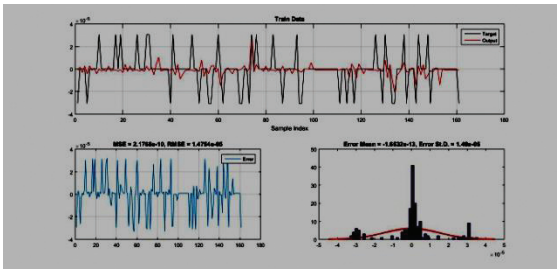

Figure 17. Train data of the word go.

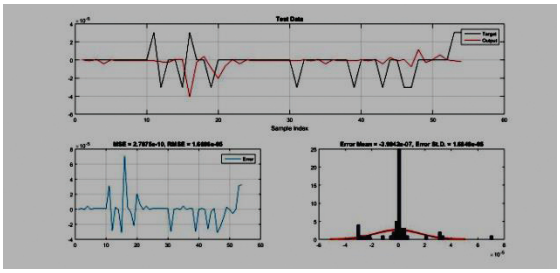

Figure 18. Test data of the word go.

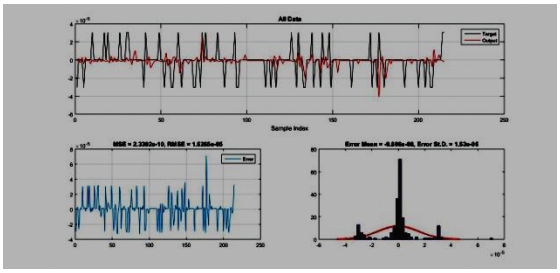

Figure 19. All data of the word go.

Table 7. Different data of the word go.

Data type	MSE	RMSE	Error mean	Error St.D.
Train	2.1765 e-10	1.4754 e-05	1.6532 e-15	1.48 e-05
Test	2.7875 e-10	1.6696 e-05	–3.9042 e-07	1.6848 e-05
All	2.3302 e-10	1.5265 e-05	–9.806 e-07	1.53 e-05

4. Naan

The above word is a two tamil word having a monosyllabic word. The train data and test data details are shown in Figure 20, Figure 21, Figure 22 and Table 8.

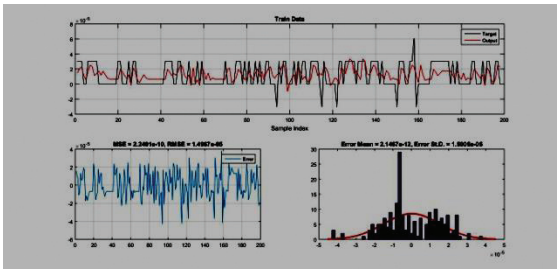

Figure 20. Train data of the word naan.

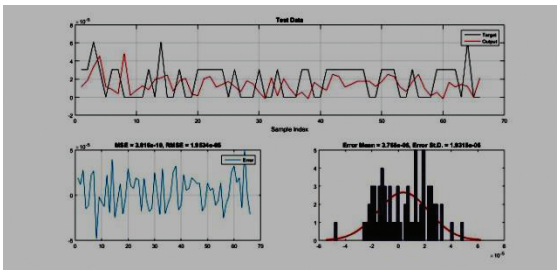

Figure 21. Test data of the word naan.

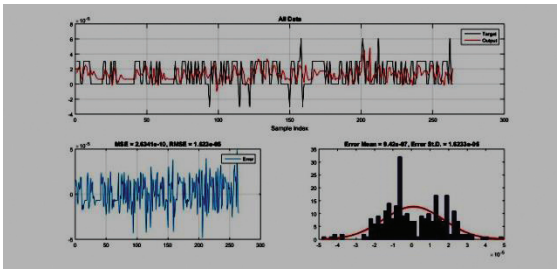

Figure 22. All data of the word naan.

Table 8. Different data of the word naan

Data type	MSE	RMSE	Error mean	Error St.D.
Train	2.2401 e-10	1.4967 e-05	2.1467 e-12	1.5005 e-05
Test	3.816 e-10	1.9534 e-05	3.768 e-06	1.9315 e-05
All	2.6341 e-10	1.6232 e-05	–9.42 e-07	1.6233 e-05

4a. Me

The above word is english word equivalent to the above-explained tamil word which carries one monosyllabic word. The train data and test data details are shown in Figure 23, Figure 24, Figure 25 and Table 9.

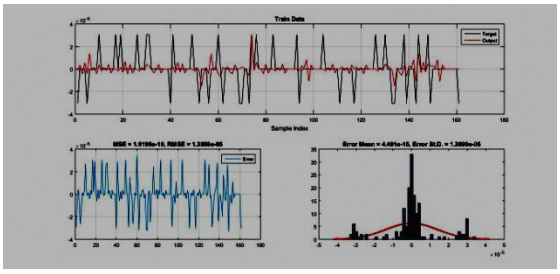

Figure 23. Train data of the word me.

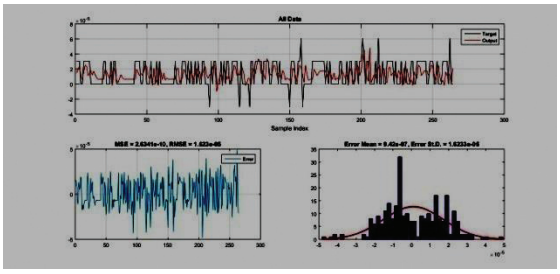

Figure 24. Test data of the word me.

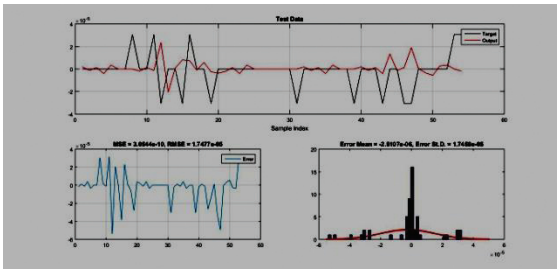

Figure 25. All data of the word me.

Table 9. Different data of the word me

Data type	MSE	RMSE	Error mean	Error St.D.
Train	1.9198 e-10	1.3856 e-05	4.491 e-15	1.3899 e-05
Test	3.0544 e-10	1.7477 e-05	–2.5107 e-06	1.7458 e-05
All	2.2048 e-10	1.4849 e-05	–6.306 e-07	1.487 e-05

5. Irangu

The above word is a two tamil word having a disyllabic word. The train data and test data details are shown in Figure 26, Figure 27, Figure 28 and Table 10.

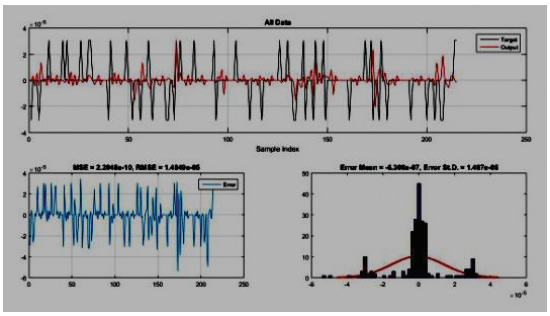

Figure 26. Train data of the word irangu.

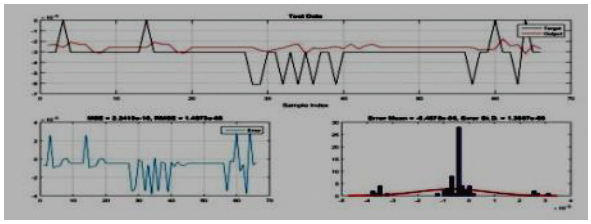

Figure 27. Test data of the word irangu.

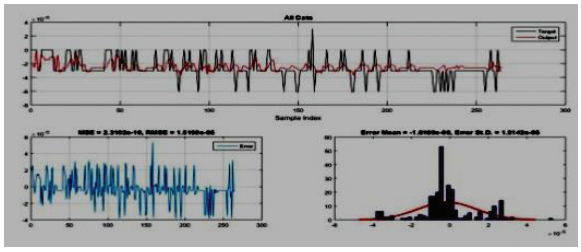

Figure 28. All data of the word irangu.

Table 10. Different data of the word irangu

Data type	MSE	RMSE	Error mean	Error St.D.
Train	2.333 e-10	1.5274 e-05	–3.5599 e-12	1.5313 e-05
Test	2.419 e-10	1.4973 e-05	–6.4675 e-06	1.3607 e-05
All	2.3102 e-10	1.5194 e-05	–1.6169 e-06	1.5142 e-05

5a. Get down

The above word is English word equivalent to the above-explained Tamilword, which carries two monosyllabic word. The train data and test data details are shown in figure 29, figure 30, figure 31 and table 11.

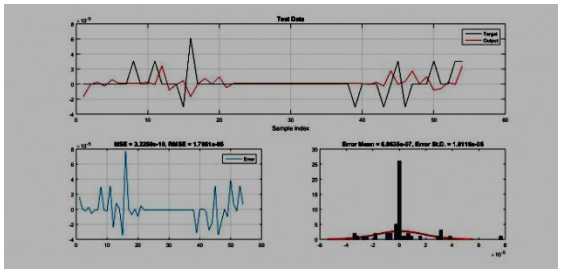

Figure 29. Train data of the word get down.

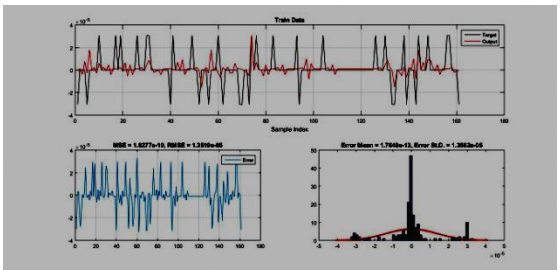

Figure 30. Test data of the word get down.

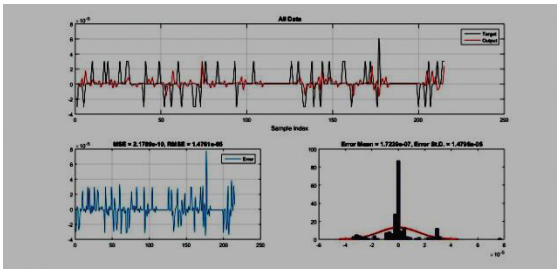

Figure 31. All data of the word get down.

Table 11. Different data of the word get down

Data type	MSE	RMSE	Error mean	Error St.D.
Train	1.8277 e-10	1.3094e-05	3.1771 e-13	1.3135e-05
Test	3.7932 e-10	1.9476 e-05	2.346 e-06	1.9516 e-05
All	2.2367 e-10	1.4955 e-05	5.8924 e-07	1.4979 e-05

6. Enakuuthavungal

The above word is a two tamil word having a disyllabic word and trisyllabic. The train data and test data details are shown in Figure 32, Figure 33, Figure 34 and Table 12.

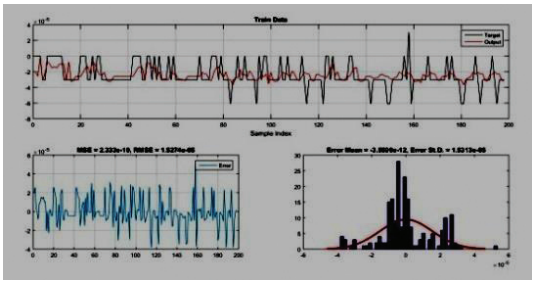

Figure 32. Train data of the word enakuuthavungal.

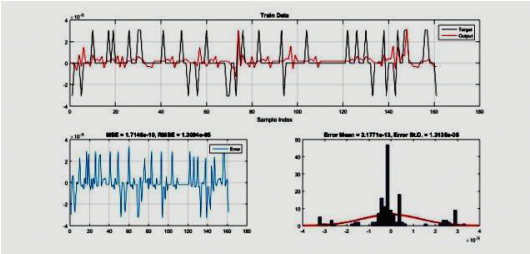

Figure 33. Test data of the word enakuuthavunga.

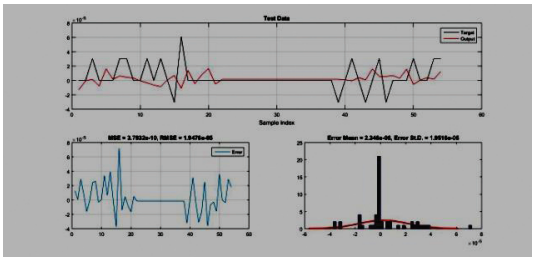

Figure 34. All data of the word enakuuthavungal.

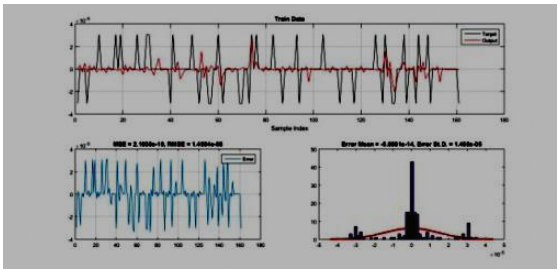

Figure 35. Train data of the word help me.

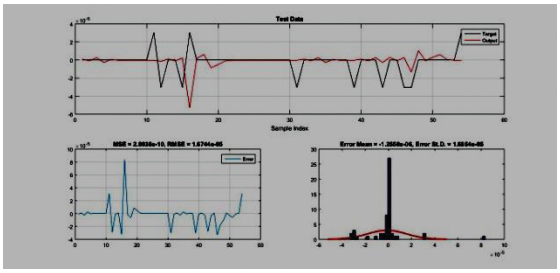

Figure 36. Test data of the word help me.

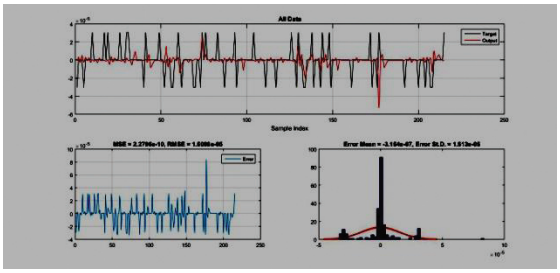

Figure 37. All data of the word help me.

Table 12. Different data of the word enakuuthavungal

Data type	MSE	RMSE	Error mean	Error St.D.
Train	1.7146 e-10	1.3094 e-05	3.1771 e-13	1.3135 e-05
Test	3.7932 e-10	1.9476 e-05	2.346 e-06	1.9156 e-05
All	2.2367 e-10	1.4955 e-05	5.8924 e-07	1.4979 e-05

Table 13. Different data of the word help me

Data type	MSE	RMSE	Error mean	Error St.D.
Train	2.1038 e-10	1.4504 e-05	–5.6051 e-14	1.455 e-05
Test	2.8038 e-10	1.6744 e-05	–1.2558 e-07	1.6854 e-05
All	2.2796 e-10	1.5265 e-05	–9.806 e-07	1.53 e-05

6a. Help me

The above word is english word equivalent to the above-explained tamil word which carries two monosyllabic word. The train data and test data details are shown in Figure 35, Figure 36, Figure 37 and Table 13.

7. Enpeyar

The above word is a two tamil word having a monosyllabic and disyllabic word. The train data and test data details are shown in Figure 38, Figure 39, Figure 40 and Table 14.

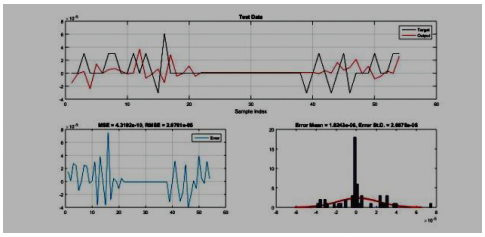

Figure 38. Train data of the word enpeyar.

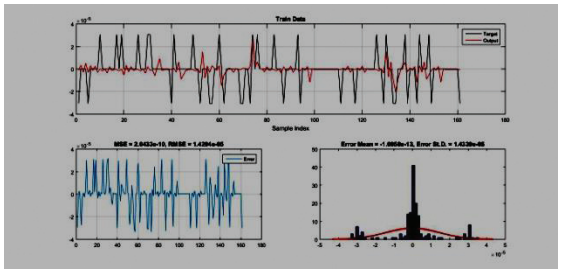

Figure 39. Test data of the word enpeyar.

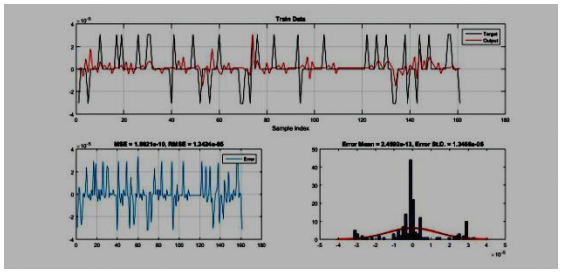

Figure 40. All data of the word enpeyar.

Table 14. Different data of the word enpeyar

Data type	MSE	RMSE	Error mean	Error St.D.
Train	1.8021 e-10	1.3424 e-05	2.4592 e-13	1.3466 e-05
Test	4.3102 e-10	2.0761 e-05	1.8243 e-06	2.0875 e-05
All	2.432 e-10	1.5535 e-05	4.5819 e-07	1.5625 e-05

7a. My name

The above word is english word equivalent to the above-explained tamil word which carries two monosyllabic word. The train data and test data details are shown in Figure 41, Figure 42, Figure 43 and Table 15.

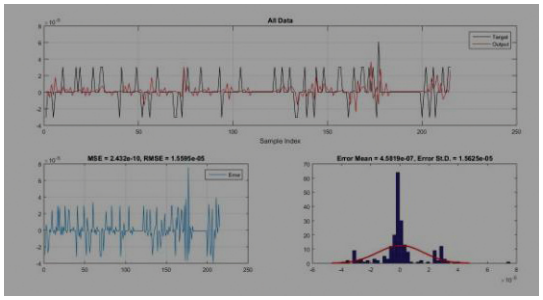

Figure 41. Train data of the word my name.

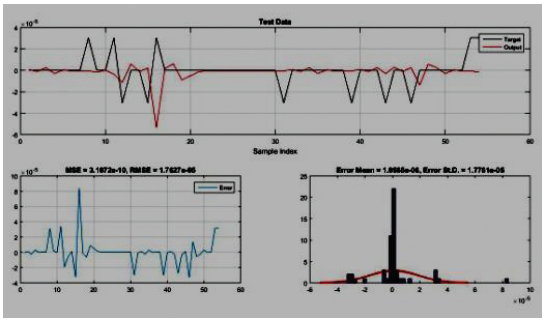

Figure 42. Testdata of the word my name.

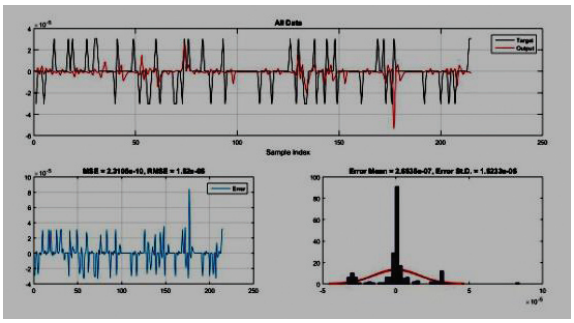

Figure 43. All data of the word my name.

Table 15. Different data of the word my name

Data type	MSE	RMSE	Error mean	Error St.D.
Train	2.0483 e-10	1.4294 e-05	–1.0055 e-13	1.4339 e-05
Test	3.1072 e-10	1.7629 e-05	1.0565 e-06	1.7761 e-05
All	2.3105 e-10	1.52 e-05	2.6525 e-07	1.5233 e-05

The different syllablle found in the above stated tamil words are well recogniable with NAM speech is shown in Table 16.

Table 16. Different syllabic words used for experiment and obtain good results

Tamil word	Syllable
Ingae va	Disyllabic+Monosyllabic
Va	Monosyllabic
Poo	Monosyllabic
Naan	Monosyllabic
Irangu	Disyllabic
Enaku uthavungal	Disyllabic+Trisyllabic
En peyar	Monosyllabic+Disyllabic

5. Conclusion

The above different words explains how much stress need to be given over the particular word while speaking. According to the stress and number of syllable the words shows its variation in the number of bins presented in the all data graph. From the graph it is also understood that the variation of bin amplitude varies for different word. For example lets us take a three different syllabic words namely a monosyllabic, disyllabic and trisyllabic word. The word 'naan'which is a monosyllabic word could be seen from the bin. The word has a larger stress applied on it thus the length of the bin is large in amount. The tamil word 'irangu' being a disyllabic word used here for experiment has somehow same number of bin as compared to the word 'naan'. But the amplitude of the bin varies in contrast to the bin in the word 'naan'. The trisyllabic tamil word used here is 'uthavungal' which varies from the other two words such that the bin of this word has a sharp streap. The equivalent english word for all the above word shows the same characteristics as its equivalent tamil word but the words involved here are all monosyllabic word. Thus the span in terms of bin of those words are short when comparing with the tamil words.

6. References

1. Akram M, Habib S, Javed I. Intuitionistic fuzzy logic control for washing machines. Indian Journal of Science and Technology 2014 May; 7(5):654 –61.
2. Pellom B, Sarikaya R, Hansen J. Fast likelihood computation techniques in nearest-neighbor based search for continuous speech recognition. IEEE Signal Processing Letters. 2001 Aug; 8(8): 221–4
3. Ney H Ortmanns S. Dynamic programming search for continuous speech recognition. IEEE Signal Processing Magzine. 1999 Sep; 16(5):64–83.
4. Optimization of nam speech recognition system using low power signal processor, International Journal of Applied Engineering Research. 2015.
5. Choi Y, You K, Choi J, Sung W. VLSI for 5000-word continuous speech recognition. Proceedings of IEEE ICASSP; Taipei: Taiwan. 2009 Apr. p. 557–60. ISSN: 1520–6149.
6. Choi Y, You K, Choi J, Sung W. A real-time FPGA-based 20,000-word speech recognizer with optimized DRAM access. IEEE Transaction on Circuits Systems I: Regula Papers. 2010 Aug; 57(8): 2119–31.
7. Ward W, Krech H, Yu X, Herold K, Figgs G, Ikeno A, Jurafsky D. Lexicon adaptation for LVCSR: speaker idiosyncrasies, non-native speakers, and pronunciation choice. Center for Spoken Language Research University of Colorado, Boulder. 2002 Sep. p. 83–8.
8. Reynolds DA, Quatieri TF, Dunn RB. Speaker verification using adapted Gaussian mixture models. MIT Lincoln Laboratory. 244 Wood St., Lexington, Massachusetts 02420. Digital Signal Processing. 2000; 10(1–3):19–41. Available from: http://www.idealibrary.com on
9. Cesari RA. Viterbi decoder with reduced metric computation. The application claims priority of provisional application No.60/009337 field. 1995 Dec.
10. Hermansky H, Hanson BA, Wakita H. Perceptually based linear predictive analysis of speech. Proceedings of IEEE International Conference on Acoustic, speech, and Signal Processing; 1985 Aug. p. 509–12
11. Picone JW. Signal modelling technique in speech recognition. Proceedings of the IEEE. 1993 Sep; 81(9): 1215–47.
12. Alotaibi YA, Hussain A. Comparative analysis of arabic vowels using formants and an automatic speech recognition system. International Journal of Signal Processing, Image Processing and Pattern Recognition. 2010; 3(2):11–22.
13. Kesarkar M. Feature extraction for speech recognition. M.Tech. Credit Seminar Report, Electronic Systems Group, EE. Dept, IIT Bombay. 2003 Nov. p. 1–12.
14. Hunt M, Lefebre C. A comparison of several acoustic representations for speech recognition with degraded and undegraded speech. IEEE International Conference on Acoustics, Speech, Signal Processing. Glasgow: U.K. 1989. p. 262–5.
15. Bei CD, Gray RM. An improvement of the minimum distortion encoding algorithm for vector quantization. IEEE Transaction on Communication. 1985; 33(10): 1132–3.
16. Bocchieri E. Vector quantization for efficient computation of continuous density likelihoods. IEEE International Conference Acoustics, Speech, and Signal Processing. Minneapolis: MN. 1993; 2:692–5. ISSN: 1520–6149.