

TEXT AND LANGUAGE INDEPENDENT SPEAKER RECOGNITION SYSTEM FOR ROBUST ENVIRONMENT

A Project Report Phase – II

**Submitted in partial fulfilment of the requirements for the award of
Bachelor of Engineering Degree in
Electronics and Communication Engineering**

by

P SEETHA 39130342

P SOUMYA SRI 39130340



**DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
SCHOOL OF ELECTRICAL AND ELECTRONICS**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade “A” by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119.

NAGAR, RAJIV GANDHI SALAI, CHENNAI

APRIL - 2023



SATHABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI -600119

www.sathyabama.ac.in

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **P SEETHA (39130342)**, and, **P SOUMYA SRI (39130340)** who carried out the project entitled "**TEXT AND LANGUAGE INDEPENDENT SPEAKER RECOGNITION SYSTEM FOR ROBUST ENVIRONMENT**" under our supervision from NOVEMBER 2022 to MARCH 2023.

Internal Guide

Dr. S. EMALDA ROSLIN. M.E., Ph.D.,

Head of the Department

Dr. T. RAVI, M.E., Ph.D.,

Submitted for Viva voce Examination held on 24.4.23

Internal Examiner

External Examiner

DECLARATION

We **P SEETHA (39130342)** and, **P SOUMYA SRI (39130340)**, hereby declare that the Project Report entitled “**TEXT AND LANGUAGE INDEPENDENT SPEAKER RECOGNITION SYSTEM FOR ROBUST ENVIROMENT**” is done under the guidance of **Dr. S. EMALDA ROSLIN, M.E., Ph.D.**, is submitted in partial fulfilment of the requirements for the award of Bachelor of Engineering degree in **Electronics and Communication Engineering**.

DATE: 24.4.23

1. 

PLACE: Chennai

2. 

SIGNATURE OF THE CANDIDATES

ACKNOWLEDGEMENT

We are pleased to acknowledge our sincere thanks to Board of Management of **SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their kind encouragement in doing this project and for completing it successfully. We are grateful to them.

We convey our profound gratitude to **Dr. N.M. NANDHITHA, M.E., Ph.D.**, Prof. & Dean, School of Electrical and Electronics and **Dr. T. RAVI, M.E., Ph.D.**, Head of the Department of **Electronics and Communication Engineering**, Sathyabama University for having been a constant source of support and encouragement for the completion of the project.

We consider our self-most fortunate and immensely privileged to have worked under the guidance of **Dr. S. EMALDA ROSLIN, M.E., Ph.D.**, Department of Electronics and Communication Engineering to complete this project successfully.

We express our heartfelt thanks to our family members and all our friends who helped me to complete our project successfully & finally, I thank the exciting Super Natural Power, the Almighty GOD.

ABSTRACT

Forensic speaker recognition is required when an individual voice is presented as evidence, such as in telephone recordings, wiretaps, audio surveillance, or informant recordings. Speech is the primary means of communication in our daily lives, with numerous languages used for this purpose. Speech technologies are widely utilized and have countless applications, allowing machines to respond accurately and dependably to human voices and provide helpful and valuable services. The goal of this project is to create a model that can identify speakers regardless of the language they speak and the volume of their voice, and can even recognize speakers when there is noise present in the speech. In our proposed work, Discrete wavelet transform (DWT) is used for feature extraction and Back Propagation Network (BPN) and Adaptive Resonance Theory (ART) is used for classification. This speaker recognition system can be applied in various fields, particularly in forensics, security, and biometrics where speech can be used as a password or for other purposes.

In this project we will develop a model through which we can identify the speaker who is speaking irrespective of language they are speaking in and the volume in which their voice is in. through this network we can also recognize the speaker even noise is present in speech. This application is used in many areas especially in forensic application. This can also be used in security and biometric applications in which we can have speech as password and many other applications.

TABLE OF CONTENTS

CHAPTER No.	TITLE	PAGE No.
	ABSTRACT	V
	LIST OF FIGURES	viii
	LIST OF TABLES	ix
	LIST OF GRAPHS	X
	LIST OF ABBREVIATIONS	Xi
1	INTRODUCTION	1
	1.1 DESIGN CONSIDERATIONS FOR THE PROPOSED SYSTEM	2
	1.2 BACKGROUND	3
	1.3 MOTIVATION	4
	1.4 OBJECTIVES	6
2	LITERATURE SURVEY	8
	2.1 LITERATURE SURVEY	8
	2.2 RELATED WORK	13
	2.3 EXISTING WORK	14
3	PROPOSED METHODOLOGY	16
	3.1 RESARCH DATA BASE	16
	3.2 SPEAKER IDENTIFICATION	18
	3.3 BLOCK DIAGRAM OF PROPOSED WORK	19

3.4	DESIGN CONSIDERATION FOR THE PROPOSED SYSTEM	22
4	PROPOSED SPEAKER RECOGNITION SYSTEM	24
4.1	FEATURES EXTRACTION USING DWT	24
4.2	SPEAKER RECOGNITION USING BPN CLASSIFIER	27
4.3	SPEAKER RECOGNITION USING ADAPTIVE RESONANCE THEORY CLASSIFIER	27
4.4	BPN CODE ALGORITHM	30
4.5	ART CODE ALGORITHM	31
5	RESULTS AND DISCUSSION	33
5.1	FEATURES FOR CLASSIFICATION	33
5.2	BACK PROPAGATION NETWORK CLASSIFIER FOR SPEAKER RECOGNITION	40
5.3	ADAPTIVE RESONANCE THEORY CLASSIFIER FOR SPEAKER RECOGNITION	44
5.4	SENSITIVITY ANALYSIS OF BPN AND ART	45
6	CONCLUSION AND FUTURE WORK	48
	REFERENCE	49
	PUBLICATION	

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE NO
Fig:3.1	SPEECH SINGLE#1	16
Fig:3.2	SPEECH SINGLE#2	17
Fig:3.3	BLOCK DIAGRAM	19
Fig:3.4	IDENTIFICATION VS VERIFICATION	22
Fig:4.1	MEASUREMENT OF GENERAL FORMS OF KURTOSIS	24
Fig:4.2	NORMAL	25
Fig:4.3	DMEY WAVELET	26
Fig:4.4	HAAR WAVELET	26
Fig:5.1	NEURAL NETWORK	41
Fig:.2	TRAINING	42
Fig:5.3	TRAINING SENSISTIVITY	42
Fig:5.4	NEURAL NETWORK TRAINING	43
Fig:5.5	NEURAL NETWORK TRAINING STATE	43
Fig:5.6	NEURAL NETWORK TRAINING REGRESSION	44

LIST OF TABLES		
TABLE NO.	TITLE	PAGE NO
Table:5.1	APPROXIMATION COEFFICIEENTS FEATURES EXTRACTED USING DWT(Haar)	34
Table:5.2	DETAILED COEFFFFICIENTS FEATURE USING DWT(Haar)	35
Table:5.3	APPROXIMATION COEFFICIEENTS FEATURES EXTRACTED USING DWT(Dmey)	37
Table:5.4	DETAILED COEFFICIENT FEATURE USING DWT (Dmey)	39
Table:5.5	SENSITIVITY DMEY WAVELET	45
Table:5.6	SENSITIVITY HAAR WAVELET	46

LIST OF GRAPHS

GRAPH NO.	TITLE	PAGE NO.
Graph:5.1	DMEY WAVELET	46
Graph:5.2	HAAR WAVELET	47

LIST OF ABBREVIATIONS

CMN	-	Cepstral Mean Normalization
DTW	-	Dynamic Time Warping
EM	-	Expectation-Maximization algorithm
FER	-	Failure to Enroll Rate
FAR	-	False Acceptance Rate
FMR	-	False Match Rate
FNMR	-	False Non Match Rate FRR:
FIR	-	Fundamental frequency
GMM	-	Gaussian Mixture Model GUI:
HMM	-	Hidden Markov Model
IP	-	Internet Protocol
LFCC	-	Linear frequency coefficients
LPC	-	Linear predictive coding
Matlab	-	Matrix Laboratory
MAP	-	Maximum a posteriori
MFCC	-	Mel-frequency cepstral coefficients
PLP	-	perceptual linear prediction
SI	-	speaker identification
Sv	-	speaker verification
SVM	-	support vector machine
UBM	-	Universal background model
VQ	-	vector Quantization
TNMR	-	True Non Match Rate

CHAPTER 1

INTRODUCTION

Speaker recognition consists on automatically recognizing the identity of a speaker based on individual information obtained from his speech waves. This technique is developed in order to make possible the access control by voice in many services such as voice dialing, telephone shopping, banking transactions over a telephone network, information and reservation services, database access services, voice mail, remote access to computers, and security control in many confidential information areas. With the advancement of technology, many methods started to develop in the field of Automatic Speaker Recognition. Speaker recognition can be divided into two applications which are speaker identification and speaker verification. Speaker identification is the process of identifying the identity of a speaker from a group of N registered speakers.

Speaker verification is the process of accepting or rejecting a person claimed identity from his voice. Speaker identification is then the most difficult problem of the two. Similarly, speaker recognition can also be divided into text dependent and text independent methods. In text dependent systems, speaker recognition relies on a specific text being spoken. In text independent systems, a speaker is recognized independent of what is saying. The text-dependent methods generally achieve higher recognition performance than the text-independent method since this method can directly exploit the voice individuality associated with each phoneme or syllable. The structure of textdependent recognition systems is, therefore, rather simple since in text-independent method, the system must be able to recognize the speaker from any text.

Furthermore, speaker identification can also be divided into two modes of operation namely, closed set and open set speaker identification. The closed set problem means to identify a speaker from a group of N known speakers. In the open set problem, a reference

model for an unknown speaker may not exist. So, unknown voices are referred to as unknown speakers.

1.1 DESIGN CONSIDERATIONS FOR THE PROPOSED SYSTEM

Design of a speaker recognition algorithm consists of some evaluation. As a first step, the speech signal is taken from microphone or through a data file. Following the data acquisition algorithm, feature extraction is performed. Here, a representation of input speech signal is obtained, and feature models are extracted to define speaker identity. Feature extraction is to collect the speaker specific information such as articulation modeling, theory of linguistics and phonetics, perceptual cues used by listeners. In training phase, pattern recognition algorithm constructs a model/template for each user for a given set of training vectors. Then, these extracted reference feature models are stored in a reference database and compared with the feature model of input voice in test phase. For a pattern recognition algorithm there are many methods to be used. These are Dynamic Time Warping, Hidden Markov model (HMM), GMM, Artificial Neural Network or Vector Quantization techniques. After recognition process, accept or reject of the user is decided according to a similarity measurement method. Similarity measurement is a comparison between input voiceprint (claimant) and reference voiceprint templates (constructed models during training phase). Decision-making is to make a best possible choice for a given input. It is made according to the similarity measurement result. It is an acceptance/rejection comparison with a threshold to determine whether the two voiceprints are similar enough for the system to accept the identity claim. The input voice is accepted if it is greater than a threshold, otherwise it is rejected. Block diagram of a speaker recognition system training & test

A. Speaker Verification Using GMM-UBM model In GMM method, probability distribution of acoustic feature vectors for a given speaker is modeled by a Gaussian mixture probability density function. Here, the objective is to find a model which has the maximum posteriori probability for a given observation sequence. Compared to HMM systems, the GMM systems use one large model and allow the sharing of training data between different mixtures, disregarding phonetic specific information. This leads to a

better trained mixture parameter. The HMM is a generalized model and HMM state densities can be used as the acoustic class densities in the GMM. For this, each speaker's HMM word models are trained as normal and then a speaker's GMM is formed by pooling the state densities over all of his/her word models. For an input utterance, GMM is a weighted sum of M component Gaussian densities. For a D dimensional x_t feature vector, the mixture density is as in Eq.1. Log likelihood model λ for a sequence of feature vectors $X=\{x_1, x_T\}$ is computed as in Eq.(1) (where w_i : mixture weights, $\pi_i(x)$: Gaussian densities) A background model should be selected to represent the population of expected imposters. A background model act as a normalization to minimize non-speaker variability in decision score.

1.2 BACKGROUND

Speaker recognition was attempted on Radio. In 1939 at IBM Bells lab, speech synthesis Communication with people around the world for different purposes is a daily necessity of our lives. Due to the rich information that can provide, speech is the most powerful, natural and preferred means of communication. The information provided by speech is vast and includes gender, attitude, emotion, health situation and identity of a speaker based on the linguistic feature of the language. These aspects can be extracted using speech processing techniques. Speech processing is a diverse field with many applications and Speaker Recognition (SR) is one of these applications that has a great research attention and rapid advancements.

Historically, the idea of speaker recognition was a research interest since 1920. In 1920 the first machine was developed which can artificially listen and recognize the human speech. In 1950 the speech recognition system investigates spectral resonances during the vowel region of each utterance which can extract from output signal of an analog filter box and logic circuit. In 1960's Japanese laborites entering recognition arena and they built special purpose hardware which can recognize only vowels. In 1970's speaker identification technology was focused on isolated words or discrete utterance identification using linear predictive coding technique. In 1980's this

problem has been solved and robust system which recognize the string of words. In 1990"s many innovations took place in the field of pattern recognition.

Main objective was to achieve least identification error and provide the best fitting of distribution function to given data. In 2000"s the adaptive learning techniques has been develop for artificial learning of speech. In 2005 some improvements were developed in speech identification system as continuous speech identification and performance improvements on large scale speech data base. Identifying is the ability to differentiate or remember someone or something from others based on particular properties. It is common in day to day activities of human beings and some other animals. Identifying an individual during interactions between human beings is the most important for different reasons. For example it is important for making communication, sharing resources, giving and taking services, for other social relationship and etc. Human beings have great ability to recognize a person from others based on different characteristics. They use various physiological and psychological characteristics of an individual to identify the identity of a person. Some of the basic psychological and physiological character are name, face, voice, gait and so on. Identifying individuals with what they have and know are used for the security issue of the various systems. But, biometrics has additional application in addition to security in various areas. For example DNA recognition biometrics is used for identifying an individual's gene

1.3 MOTIVATION

The presentations of speaker identification technology are quite varied and frequently rising. It can be functional for Access Control, Transaction Authentication, Law Enforcement, Speech Data Management, Personalization and many more. To catch all these benefits, we need to conduct research on the area especially for the language of our interest. This is the motivation that involved me to do my thesis on utterer identification.

Selected method for speaker features extraction is Mel Frequency Cepstral Coefficients (MFCCs) because it has low-level features mostly used for speaker identification and

speech identification systems. Mel Frequency Cepstral Coefficient (MFCC) feature is selected and Cepstral Mean Normalization (CMN) is for normalizing the feature. And Gaussian Mixture Methods (GMM) for speaker modelling and the aim of modelling process is to capture the unique voice characteristics of the speaker while maintaining its discriminability with other speakers as well. The aim of speaker identification is to identify/verify the speaker who is claimed with the voice from the speakers registered in the database of the given application/ model. It applies a sequence of steps to perform identification. The speakers should be registered as the first task of the speaker identification and then the identification or verification can be conducted. Both for registering speaker and identifying speaker first speech should be acquired, some of non-important components should be removed and compromised, and then speaker specific characteristics should be collected from the speech and normalized for the better quality of characteristics representation. During registering speaker, for all speaker specific characteristics collected from all speech; a single speaker characteristics representation will be developed and stored in the speaker registration database. During speaker identification, the matching rate of speaker specific characteristics of test speech with the registered speaker characteristics will be computed and the maximum match result will be returned as the result. The model is conduct to represent the above techniques. And the model is experimented and evaluated.

STATEMENT OF THE PROBLEM

There has stayed huge wish to be able to identify a person on the root of his or her voice. Aimed at many centuries, courts, lawyers, detectives, and law implementation agencies wanted to use forensic vocal sound confirmation to examine a doubtful or to confirm a decision of fault or goodness.

To safeguard our communities from illegal acts, we need to have strong forensic analysis system. Speaker identification as a voice biometric can be applied for this purpose because speaker's voice is very problematic to fake for biometrics comparison purposes, since a numerous of qualities are measured ranging from language and speaking style to pitch, spectral magnitudes, and setup frequencies. But we have nobody done based on

the language to be applied for this purpose. So, we need to plan a speaker identification system based on our language. The additional reason for this thesis is the occurrence of audiovisual conferencing in our republic. When video conferencing is announced from unlike seats, identifying the individual verbal is helpful for the audience to know from where he is as well as to easily understand what he is talking about. Speaker identification method can be practical for this determination. The final aim for this study is the use of speaker identification technology for medical remote controlling. Ethiopia is one of the poorest countries in the world. Due to this, we do not have sufficient number of specialists and hospitals. To provide health services for the rural society, medical remote nursing is the best way. But to properly follow and treat patients, we have to identify if the person to be treated is the right one or not. To do this, speaker identification is the top method. Even held speaker identification system is valid in areas stated above, we don't have one designed based on the language. This is the difficult that initiated us to fix this study. In general, this investigation is mainly motivated to develop Text Independent Speaker Identification model which operate in text independent Identification scenario by having good language identification mechanism in open set concept. And a Text prompted concept is incorporated to minimize spoofing and mimicry

1.4 OBJECTIVES

The objective of research is generally the aim of this research is to investigate the possibility of developing a Text Independent Speaker identification Model for speakers.

Specific objectives to achieve the general objective of this research the following specific objectives are sited.

- To selected appropriate feature extraction approaches
- To select techniques for securing the model from mimicry and spoofing during identification
- To design language recognition architecture for the model before speaker identification task were performed
- To evaluate performance of the model

Speaker Identification Method

Speaker identification is one of the tasks of identification phase. It helps to identify the speaker from the entire trained model by comparing test utterance with the entire speaker model stored in the database. The detail discussion, experiment and result of TISIHL model speaker identification is found.

CHAPTER 2

LITERATURE REVIEW

2.1 LITERATURE SURVEY

Text and language independent speaker recognition system for robust environment, Selma Ozaydin, Cankaya University, Ankara, Turkey His paper presents the design of a text independent speaker recognition system based on speech data consist of 134 speakers from YOHO database for different training conditions. The increase of the proposed system performance is observed with the decrease of Equal Error Rate. Experiment results the system gives the best recognition performance for Gaussian mixture model with 64 mixtures [1]

Text and language independent speaker recognition system for robust environment, Zhenyu Wang, Hangzhou Dianzi University, Hangzhou, China Adapting speaker recognition systems to new environments is a widely-used technique to improve a well-performing model learned from large-scale data towards a task-specific small scale data scenarios. However, previous studies focus on single domain adaptation, which neglects a more practical scenario where training data are collected from multiple acoustic domains needed in forensic scenarios. Audio analysis for forensic speaker recognition offers unique challenges in model training with multi-domain training data due to location/scenario uncertainty and diversity mismatch between reference and naturalistic field recordings. It is also difficult to directly employ small-scale domain-specific data to train complex neural network architectures due to domain mismatch and performance loss. Fine-tuning is a commonly-used method for adaptation in order to retrain the model with weights initialized from a well-trained model. Alternatively, in this study, three novel adaptation methods based on domain adversarial training, discrepancy minimization, and moment-matching approaches are proposed to further promote adaptation performance across multiple acoustic domains.

A comprehensive set of experiments are conducted to demonstrate that:

- 1) diverse acoustic environments do impact speaker recognition performance, which could advance research in audio forensics,
- 2) domain adversarial training learns the discriminative features which are also invariant to shifts between domains,
- 3) discrepancy-minimizing adaptation achieves effective performance simultaneously across multiple acoustic domains.
- 4) moment-matching adaptation along with dynamic distribution alignment also significantly promotes speaker recognition performance on each domain, especially for the LENA-field domain with noise compared to all other systems. Advancements shown here in adaptation therefore help ensure more consistent performance for field operational data in audio forensics [2]

Text and language independent speaker recognition system for robust environment
Fei Zhao, Hao Li, Xueliang Zhang Inner Mongolia University, China Recently, deep neural networks (DNNs) have achieved incredible performance in speaker verification. However, most of which remains sensitive to environment noise. In this paper, we propose an end-to-end speaker verification framework to enhance the robustness against background noise. The proposed framework first utilizes convolutional recurrent network (CRN) to address speech separation. Then the output of the middle layer of the CRN is used as the auxiliary feature, and together with the robust Filter banks (Fbanks) feature of noisy speech are fed to the speaker verification system. The speech separation and speaker verification are jointly optimized. Compared with deep speaker and DNN/i-vector, systematic evaluation indicates that the proposed algorithm can obtain a better performance in noisy conditions [3]

Text and language independent speaker recognition system for robust environment
Rania Chakroun and Mondher Frikha Advanced Technologies for Image and Signal

Processing (ATISP) Research Unit, Sfax, Tunisia National School of Electronics and Telecommunications of Sfax, Tunisia National School [4]

An important amount of speech is typically required for speaker identification system development and evaluation. Nowadays, robust speaker identification systems when short utterances are used remains a key consideration for automatic speaker recognition, since a lot of real world applications are able to deal with only limited duration on speech data. This paper presents a new approach based on a low complexity solution based on a new feature vectors to build Gaussian Mixture Models (GMM) for speaker identification systems especially when training and testing utterance lengths are reduced. We compared our proposed system to the state-of-the-art based system in Speaker identification. Experiments on TIMIT database were conducted to demonstrate that this new feature vector can outperform the standard GMM-based system and show that there is no need for extra-data to identify the speakers [5]

Text and language independent speaker recognition system for robust environment
Hema A. Murthy Osmania, University, Hyderabad, India Text-independent speaker identification from samples of speech recorded over the telephone. It focuses on the effects of acoustic mismatches between training and testing data, and concentrates on two approaches:

- extracting features that robust against channel variations and
- transforming the speaker models to compensate for channel effects. [6]

Text and language independent speaker recognition system for robust environment
Francoise Beaufays, Brussels University It focuses on the effects of acoustic mismatches between training and testing data, and concentrates on two approaches:

- extracting features that are robust against channel variations and
- transforming the speaker models to compensate for channel effects. [7]

Text and language independent speaker recognition system for robust environment
Smitha Gangisetty Mineral Resources at West Virginia University This research presents new text-independent speaker recognition system with multivariate tools such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA) embedded into

the recognition system after the feature extraction step. The proposed approach evaluates the performance of such a recognition system when trained and used in clean and noisy environments. Additive white Gaussian noise and convolutive noise are added. Experiments were carried out to investigate the robust ability of PCA and ICA using the designed approach. The application of ICA improved the performance of the speaker recognition model when compared to PCA. Experimental results show that use of ICA enabled extraction of higher order statistics thereby capturing speaker dependent statistical cues in a text-independent recognition system. The results show that ICA has a better de-correlation and dimension reduction property than PCA. To simulate a multi noises and stored in the database. Results also show that ICA outperforms PCA under adverse environments. This is verified by computing recognition accuracy rates obtained when the designed system was tested for different train and test SNR conditions with additive white Gaussian noise and test delay conditions with echo effect [8]

Text and language independent speaker recognition system for robust environment
Amirsina Torfi, Jeremy Dawson, Nasser M. Nasrabadi West Virginia University .In this paper, a novel method using 3D Convolutional Neural Network (3D-CNN) architecture has been proposed for speaker verification in the text-independent setting. One of the main challenges is the creation of the speaker models. Most of the previously-reported approaches create speaker models based on averaging the extracted features from utterances of the speaker, which is known as the d-vector system. In our paper, we propose an adaptive feature learning by utilizing the 3D-CNNs for direct speaker model creation in which, for both development and enrollment phases, an identical number of spoken utterances per speaker is fed to the network for representing the speakers' utterances and creation of the speaker model. This leads to simultaneously capturing the speaker-related information and building a more robust system to cope with within-speaker variation. We demonstrate that the proposed method significantly outperforms the traditional d-vector verification system. Moreover, the proposed system can also be an alternative to the traditional d-vector system which is a one-shot speaker modeling system by utilizing 3D-CNNs.[9].

Text and language independent speaker recognition system for robust environment Tomi Kinnunen,a, Haizhou Li aDepartment of Computer Science and Statistics, Speech and Image Processing Unit University of Joensuu, P.O.Box 111, 80101 Joensuu , FINLAND

This paper gives an overview of automatic speaker recognition technology, with an emphasis on text-independent recognition. Speaker recognition has been studied actively for several decades. We give an overview of both the classical and the state-of-the-art methods. We start with the fundamentals of automatic speaker recognition, concerning feature extraction and speaker modeling. We elaborate advanced computational techniques to address robustness and session variability. The recent progress from vectors towards supervectors opens up a new area of exploration and represents a technology trend. We also provide an overview of this recent development and discuss the evaluation methodology of speaker recognition systems. We conclude the paper with discussion on future directions.[10].

Text and language independent speaker recognition system for robust environment Tanja Schultz, Co-Chair Alex Waibel, Co-Chair Alan W. Black Douglas A. Reynolds, MIT Lincoln Laboratory The automatic speaker recognition technologies have developed into more and more important modern technologies required by many speech-aided applications. The main challenge for automatic speaker recognition is to deal with the variability of the environments and channels from where the speech was obtained. In previous work, good results have been achieved for clean high-quality speech with matched training and test acoustic conditions, such as high accuracy of speaker identification and verification using clean wideband speech and Gaussian Mixture Models (GMM). However, under mismatched conditions and noisy environments, often expected in real-world conditions, the performance of GMM-based systems degrades significantly, far away from the satisfactory level. Therefore, robustness becomes a crucial research issue in speaker recognition field. Text and language independent speaker recognition system for robust environment K. Sreenivasa Rao S.I.T IIT Kharagpur, 721302, India Anil Kumar Vuppala G.S.S.S.T IIT Kharagpur, 721302, India . In this paper we are exploring different models and methods for improving the performance of text independent speaker identification system for mobile devices. (i) presence of varying background environment,

(ii) effect of speech coding introduced by the mobile device, and (iii) impairments due to wireless channel. In this paper, we are proposing multi-SNR multi-environment speaker models and speech enhancement (preprocessing) methods for improving the performance of speaker recognition system in mobile environment. For this study, we have simulated five different background environments (Car, Factory, High frequency, pink noise and white Gaussian noise) using NOISEX data. Speaker recognition studies are carried out on TIMIT, cellular, and microphone speech databases. Autoassociative neural network models are explored for developing these multi-SNR multi-environment speaker models. The results indicate that the proposed multi-SNR multi-environment speaker models and speech enhancement preprocessing methods have enhanced the speaker recognition performance in the presence of different noisy environments.[11].

2.2 RELATED WORKS

For both text dependent and text independent speaker recognition systems, the process of extracting features from the speech signal is a fundamental process. The most widely used method to extract features is MFCC due to its close relationship with perceived frequency of ear. For a similarity measurement between an input speech and a reference, there are four usually used approaches such as template matching, statistical classification, structural matching and neural networks.

The GMM is one of the popular statistical approaches where model parameters are estimated with the maximum similarity measurement of the features and there are many kinds of systems based on MFCC and GMM in the literature. In a text independent speaker recognition system approach based on the combination of the reduced MFCC and the Short Time Zero Crossing Rate features extracted from the speech signal is proposed. TIMIT database is used for training and test. It is stated that the proposed system achieved a reduction of nearly 11 % of EER about EER obtained based on 39-dimensional feature vectors.

A text independent automatic speaker recognition (ASR) system, based on MFCC and GMM is presented in. A database including 450 speakers from the Voxforge.org website was used. For each speaker two different speech files were used: the first one to build the profile database, the second one to test the system performance. It is stated that a recognition accuracy greater than 96% was achieved by the proposed approach. An overview of classical and new approaches for automatic text-independent speaker recognition systems is given in. It gives the fundamentals of automatic speaker recognition such as feature extraction and speaker modeling. It also discusses the evaluation methodologies of speaker recognition systems. It is stated in the paper that even if the recognition accuracy of a speaker recognition system under controlled conditions is high, in practical situations many negative factors can be encountered including mismatched handsets for training and testing, limited training data, unbalanced text, background noise and non-cooperative users.

Therefore, some technical challenges such as text/language dependency, channel effects, speech durations, and cross talk speech should be taken into considerations. In, a database which is designed to evaluate text dependent speaker verification systems under different durations and lexical constraints is proposed. In this scope, it gives an overview of existing databases for text-dependent speaker verification, and it summarizes 23 different databases described in the literature by giving their main characteristics, strengths and weaknesses.

2.3 EXISTING WORK

This paper presents the design of a text independent speaker recognition system based on Mel-Frequency Cepstrum Coefficients and Gaussian Mixture Models. HTK speech recognition toolkit is used in the design of speaker models. The system is aimed to use it as a biometric authentication system. The experiments were performed on speech data consist of 134 speakers from YOHO database for different training conditions. The increase of the proposed system performance is observed with the decrease of Equal Error

Rate. Experiment results show that the system gives the best recognition performance for Gaussian mixture model with 64 mixtures.

In speaker verification, the goal is to determine whether input voice matches to a particular known voice. The system tries to find out from template model if the speaker is exactly same with the person who claims to be. The acceptance or rejection of an identity of the claimed user is defined by a decision algorithm. To make this decision, feature vectors carrying information about speaker's identity are designed and compared with the input speaker feature vector which is extracted from one or more utterances of the speaker. For a speaker identification system, the goal is to determine which input voice best matches to a voice previously recorded from N speakers. There is no identity claim from a speaker. Here, the system is required to make a decision among N speakers in a group. Therefore, instead of identification of a single person, N complete comparison is required and then a best possible user is selected after a minimum error calculation. The system tries to find out who the speaker is, by searching all possible speakers. When user population is increased, the error probability increases. In a speaker recognition system, if the speakers who claim their identity are known to the system, they are called as claimants; the speakers, either known or unknown to the system, who pose as other speakers are called impostors.

CHAPTER- 3

PROPOSED METHODOLOGY

3.1 RESEARCH DATA BASE

In this research data base we have collected 50 speech signals with .wav extension to test the signals from the site <https://www.itu.int/net/itu-t/sigdb/genaudio/AudioForm-g.aspx?val=10000501> . The speech or audio language signals are of different persons and are in different language. The speech or audio language signals are of different persons and are in a different language. Figure 1 shows the sample speech signals taken for our work. The overall block diagram of our proposed speaker recognition system is given in Figure 2. Forward transform is applied over the speech signals collected from the online research database. Over the transformed signal, features were extracted . Half of the extracted features were considered for training the classifier network and remaining were used for testing process. During testing process, the features were compared and decision is made on the selection of correct speaker.

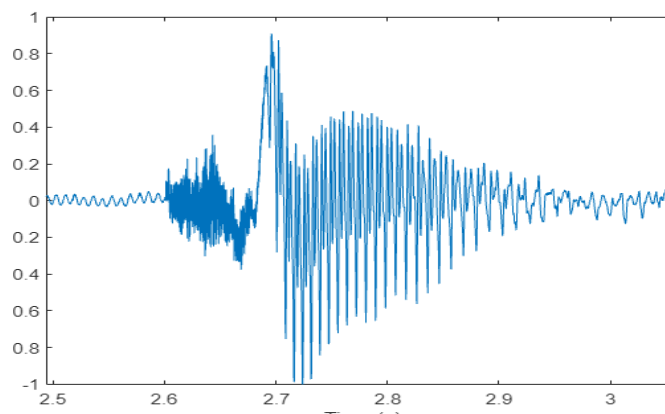


Fig: 3.1 speech signal #1

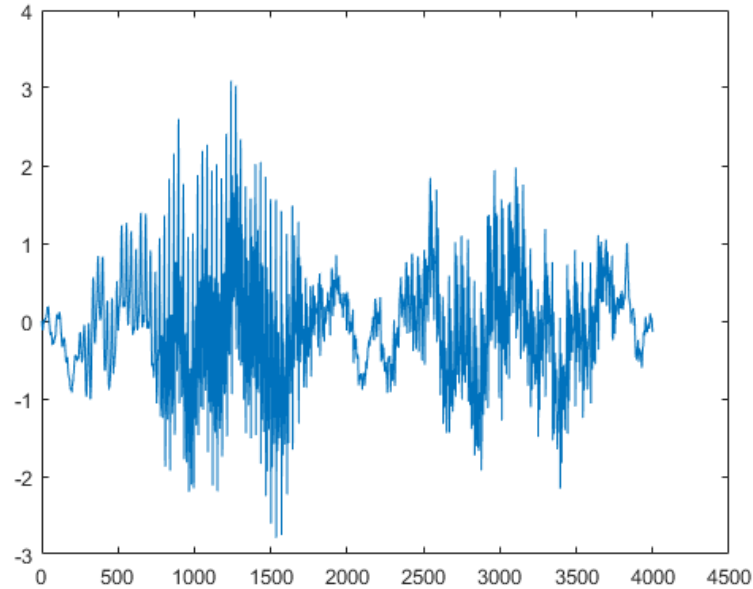


Fig :3.2 speech signal #2

For both text dependent and text independent speaker recognition systems, the process of extracting features from the speech signal is a fundamental process. The most widely used method to extract features is MFCC due to its close relationship with perceived frequency of ear. For a similarity measurement between an input speech and a reference, there are four usually used approaches such as template matching, statistical classification, structural matching and neural networks. The GMM is one of the popular statistical approaches where model parameters are estimated with the maximum similarity measurement of the features and there 2017 International Conference are many kind of systems based on MFCC and GMM in the literature

An overview of classical and new approaches for automatic text-independent speaker recognition systems. It gives the fundamentals of automatic speaker recognition such as feature extraction and speaker modeling. It also discusses the evaluation methodologies of speaker recognition systems. It is stated in the paper that even if the recognition accuracy of a speaker recognition system under controlled conditions is high, in practical situations many negative factors can be encountered including mismatched handsets for training and testing, limited training data, unbalanced text, background noise and non-

cooperative users. Therefore, some technical challenges such as text/language dependency, channel effects, speech durations, and cross-talk speech should be taken into considerations.

3.2 SPEAKER IDENTIFICATION

Design of a speaker recognition algorithm consists of some evaluation steps. As a first step, the speech signal is taken from microphone or through a data file. Following the data acquisition algorithm, feature extraction is performed. Here, a representation of input speech signal is obtained and feature models are extracted to define speaker identity. Feature extraction is to collect the speaker specific information such as articulation modeling, theory of linguistics and phonetics, perceptual cues used by listeners. In training phase, pattern recognition algorithm constructs a model/template for each user for a given set of training vectors. Then, these extracted reference feature models are stored in a reference database and compared with the feature model of input voice in test phase. For a pattern recognition algorithm there are many methods to be used. These are Dynamic Time Warping, Hidden Markov model (HMM), GMM, Artificial Neural Network or Vector Quantization techniques. Discrete Wavelet Transform (DWT) is the forward transform applied over the speech samples. It is splitted as approximation coefficients and detailed coefficients. Out of all the wavelets Discrete meyer (Dmey) and Haar wavelets were considered for our research. Over these coefficients, the features namely mean, standard deviation, kurtosis, skewness, Energy, rate, second order moment and third order moments were extracted. The extracted features were given to a Back Propagation Network (BPN) and Adaptive Resonance Theory (ART) for classification of speaker.

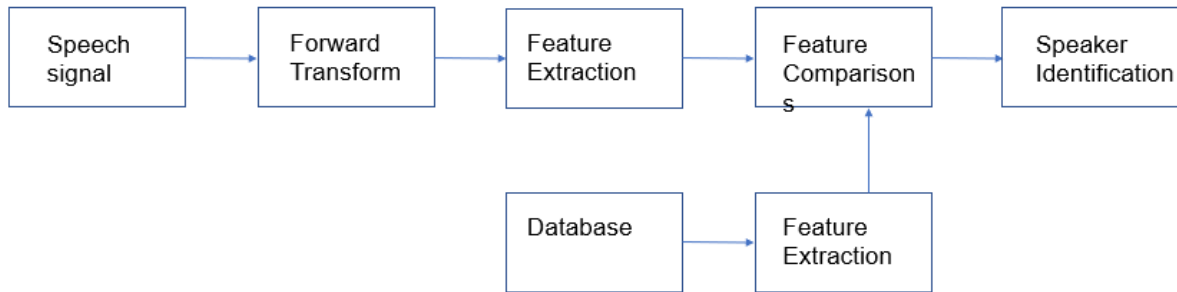


Fig:3.3 Block diagram of proposed speaker recognition system

Speaker identification is the process of automatically recognizing who is speaking. Speaker identification system is often classified into closed-set identification and open-set identification. Just as their names suggest, the closed-set refers to the cases that the unknown voice must come from a set of known speakers; and the open-set means unknown voice may come from unregistered speakers, in which case we could add „ none of the above" option to this identification system. Moreover in practice speaker identification systems could also be divided according to the speech modalities: text-dependent identification, text independent identification. For text dependent SRS, speakers are only allowed to say some specific sentences or words, which are known to the system.

On the contrary, as for the text independent SRS, they could process freely spoken speech, which is either user selected phrase or conversational speech. Compared with text-dependent SRS, text-independent SRS are more flexible, but more complicated. All above speaker identification techniques can be classified into speaker verification and identification. Speaker verification is the process of accepting or rejecting the identity claim of speaker while identification is the process of determining which registered speaker provides a given utterance. Speaker identification system makes it possible to use the speaker's voice to verify their identity and control access of the desired services. Speaker identification system is having three main components: Front –end Processing or Feature Extraction, Speaker Modeling, Pattern Matching or Logical decision. To get the feature

vectors of Incoming voice, front end processing will be performed. Feature vectors are used to create a speaker model. The pattern matching is responsible for comparing the features to speaker models. The decision module analyzes the similarity score (statistical or deterministic) to make a decision. Using a high threshold, system gets high safety and prevents impostors to be accepted, but in the mean while it also takes the risk of rejecting the genuine person, and vice versa.

Basics of speaker identification

Like human beings, computer can recognize individuals based on their physiological and psychological aspects. Speaker identification is the mechanism to recognize an individual depending on the features extracted from the speech signal. Speech signal of two different individuals cannot be the same for the same code being uttered, which is because of difference in internal structure of voice producing components of human being and some behavioral characteristics of an individuals. It follows various procedure, uses various modeling approaches and techniques to perform its task. It operates in two phases such as enrollment phase and identification phase. TSIHL model also follows similar fashion as discussed in the general overview of speaker identification

Phases of Speaker Identification

The whole process performed to recognize a speaker is categorized into two phases: Enrolment phase and identification phase. In each phase the recognizer is designed to follow various steps, procedures and techniques which helps it to identify the individual with minimum error rate.

Enrolment Phase

Enrolment phase of speaker identification is the first phase in any speaker identification system. It is used to register or train the speaker identity to the database/ model. For enrolling speaker to the database, the recognizer accepts speech in various formats. Then the recognizer finds speaker specific information either from physical or from psychological character of the speaker. Finding information from these characters uses various feature extraction technique such as: MFCC, PLP, LPCC and etc. And the recognizer develops a speaker model for each speaker using various speaker modeling approach such as: VQ,

SVM, HMM, GMM, etc. Finally the speaker model will be trained or registered to the speaker database or model using the unique identification key for each speaker.

Identification Phase

Identification phase is the second phase of any speaker identification systems. It is used to recognize (identify) the claimed speaker by comparing the test utterance feature with the model stored in the speaker database. During identification phase the recognizer accepts the speech in various format, performs various activities and releases the identification result.

Speaker Identification

Tasks Identification phase of speaker identification system can be implemented in two ways in various application areas. This way of implementing is called speaker identification task. The basic tasks of speaker identification such as: Speaker identification and Speaker verification tasks. This research is designed to perform the task of speaker identification.

Speaker Identification

Speaker identification is the task of recognizing a test speaker by comparing the test speech features with the speaker model stored in the speaker database. During this task the recognizer compares the test speech feature with all the speaker model which exists in the database.



fig :3.1 Identification Vs Verification

Therefore, Speaker identification is 1: N relation. Speaker identification can be performed in two modes: open set identification mode and closed set identification mode. In closed set identification mode, the recognizer assumes that all the test speaker is from the speaker model registered so far. This is not true in real world scenario, because some speakers who are not registered may claim. Those speakers are called imposters. There is no rejection of test speaker during closed set identification. Open-set identification may be seen as a combination of closed-set identification and speaker verification. For example, a closed-set identification may be conducted and the resulting ID may be used to run a speaker verification session. Open set identification is more flexible and practical than closed set. Speaker identification can be trained and tested based on one of the two modalities either text dependent or text independent modalities. But text independent one is preferable than other. The following figure indicates general structure of speaker identification.

3.4 Design Consideration For The Proposed Methodology

Artificial Neural Networks (ANNs) are trained either using supervised learning or unsupervised learning. Adaptive Resonance Theory (ART) networks use supervised learning where the exemplars with inputs and output are used for training the network. They are capable of self organizing in real time environment and they can able to perform a stable recognition even by getting different input patterns beyond those stored already. As ART 1 is suited for Binary classification/prediction, ART 2 is used in this paper for speaker identification [9][10]. The automated speaker recognition system involves the following steps: Speech signal acquisition, Feature extraction and Classification. A research database has been created with 10 speech signals each of 10 different speakers. Features were extracted using the statistical moments on the approximation and detailed coefficients of the decomposed speech signals using Discrete Wavelet Transform with Discrete Meyer wavelet. In addition to these features statistical moments on the stock well coefficients are also determined. Adaptive Resonance Theory (ART) is chosen as the classifier. Choice of ART is justified due to its plasticity Elasticity feature. It means that ART not only adjusts itself to the new data set but also remembers the old data set. The proposed classifier is trained with three different data sets (DWT features, Stockwell

features, DWT and Stockwell features). Among the exemplars different datasets are used for training and testing the network. Performance of the proposed technique is measured in terms of sensitivity.

CHAPTER- 4

PROPOSED SPEAKER RECOGNITION SYSTEM

4.1 FEATURES EXTRACTION USING DWT

The features namely mean, standard deviation, kurtosis, skewness, Energy, rate, second order moment and third order moments were extracted from the speaker signals.

Mean: The mean (aka the arithmetic mean, different from the geometric mean) of a dataset is the sum of all values divided by the total number of values. It's the most commonly used measure of central tendency and is often referred to as the “average.”

Standard: The standard form of a number is a way of writing the number in a form that follows certain rules. Any number that can be written as a decimal number, between 1.0 and 10.0, multiplied by a power of 10, is said to be in standard form.

Kurtosis: Kurtosis is a measure of the tailedness of a distribution. Tailedness is how often outliers occur. Excess kurtosis is the tailedness of a distribution relative to a normal distribution. Distributions with medium kurtosis (medium tails) are mesokurtic. Distributions with low kurtosis (thin tails) are platykurtic.

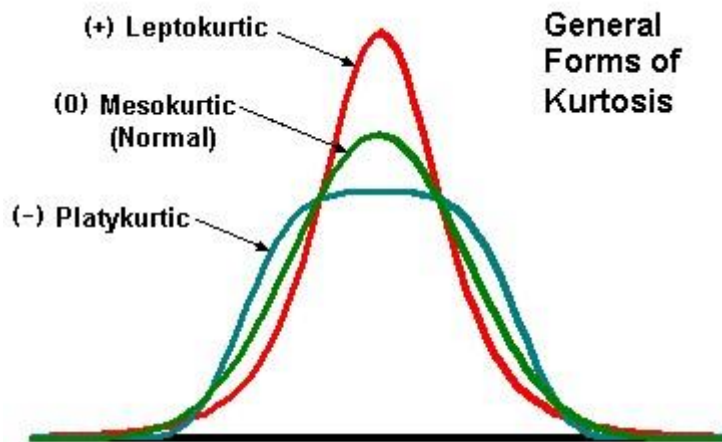


Fig: 4.1 Measurement of general forms of kurtosis

Skewness: Skewness is a measurement of the distortion of symmetrical distribution or asymmetry in a data set. Skewness is demonstrated on a bell curve when data points are not distributed symmetrically to the left and right sides of the median on a bell curve.

Normal: The normal to the curve is the line perpendicular (at right angles) to the tangent to the curve at that point. Remember, if two lines are perpendicular, the product of their gradients is -1.

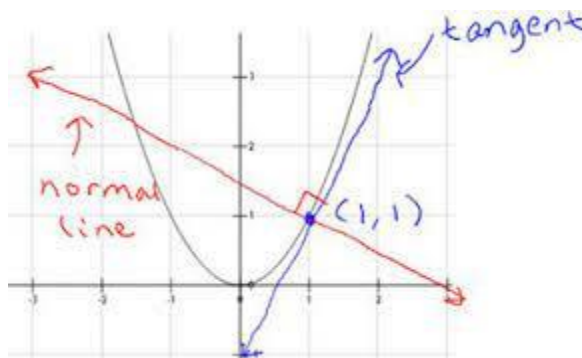


Fig 4.2 Normal

RATE: Rate is usually defined as a ratio of two quantities with different units. Usually, the rate is written as a fraction, with the first quantity as the numerator and the second quantity as the denominator. We can express the rate by reducing them to the lowest form possible. For example, if a person takes 30 steps in 20 seconds, then the rate at which they walk is 30 steps/20 seconds or 3 steps/2 seconds.

Moment: In physics, a moment is a mathematical expression involving the product of a distance and physical quantity. Moments are usually defined with respect to a fixed reference point and refer to physical quantities located some distance from the reference point.

Moment(sm): In mathematics, the second moment method is a technique used in probability theory and analysis to show that a random variable has positive probability of being positive. More generally, the "moment method" consists of bounding the probability that a random variable fluctuates far from its mean, by using its moments.

Moment(tm): The third central moment is the measure of the lopsidedness of the distribution; any symmetric distribution will have a third central moment, if defined, of zero. The normalized third central moment is called the skewness, often γ .

Dmey wavelet: Dmey is a discrete format approximation of the Meyer wavelet, and it can approximate the Meyer wavelet based on a finite impulse response (FIR) filter, as depicted . As a result, the fast wavelet transform can approximate the Meyer wavelet transform, and the discrete wavelet transform (DWT) can be applied.

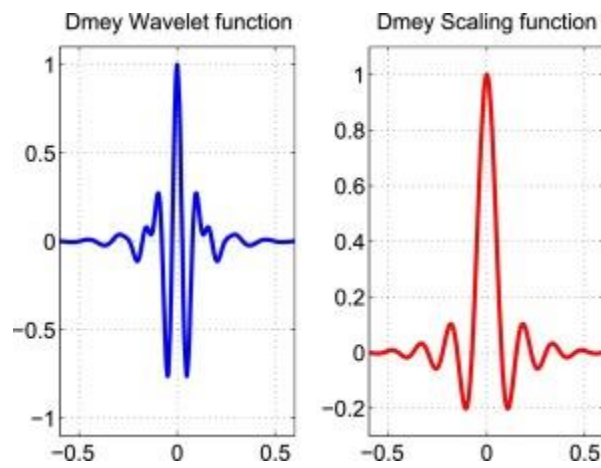


Fig:4.3 Dmey wavelet

Haar wavelet: It is found effective in applications such as signal and image compression in electrical and computer engineering as it provides a simple and computationally efficient approach for analysing the local aspects of a signal. The Haar transform is derived from the Haar matrix.

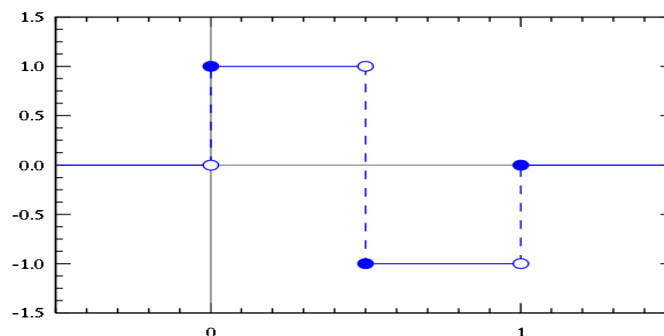


Fig 4.4 Haar wavelet

4.2 SPEAKER RECOGNITION USING BPN CLASSIFIER

A Back-propagation learning feed-forward neural network with tan-sigmoid transfer functions has been used in both the hidden layer and the output layer which is shown in Figure 7. Three-layer Back-propagation learning neural network algorithm has been used to classify the visual speech features.

The major drawbacks of Back-propagation learning neural network algorithm are the training time and local minima. Convergence time of the Back-propagation algorithm is inversely proportional to the error tolerance rate. In learning, effective use of error rate can decrease the convergence time. At first, select the final error rate. Then converge the weights such that all the patterns overcome some of the percentage error of the total system (the error must be higher than the final error rates). Finally, converge the system to the next lower error rate until crossing the final targeted error. For example, if the error rate of the system is 0.001, first the converged error rate for all of the patterns is 0.009, then 0.005, 0.003, and finally 0.001. This process is known as SET-BPL

50 speech utterances are trained in Back-propagation learning neural network and the effects of applying SET-BPL of the proposed system . Sometimes local minima problem occurs in a Back-propagation learning neural network algorithm. As a result, some precautions such as addition of internal nodes and lowering the gain term have been considered to set the learning parameters. The addition of internal nodes and lowering the gain term can increase the convergence time. To overcome these learning difficulties, a momentum term has been used to speed up the convergence process for this proposed speaker identification system.

4.3 SPEAKER RECOGNITION USING ADAPTIVE RESONANCE THEORY CLASSIFIER

Case 1: Feature extraction using S transform features.

Stockwell transform provides the localized time frequency components in contrast to frequency averaged or time averaged amplitude. These features are aggregated using statistical features namely Mean, Standard Deviation, Kurtosis and Skewness for real and

imaginary parts were extracted. These statistical features describe the overall inform of the co-efficients and the distribution of these co-efficients. These features are given to an ART network for classification. From the classifier output the sensitivity is calculated. A sample of one speech signal from all 10 speakers with its corresponding extracted features using S transforms.

Case 2: Feature extraction using DWT features.

Using DWT, the speech signal gets decomposed to approximation coefficients and detailed coefficients. Discrete Meyer (dmey) wavelet is used for decomposition as it has been already found that dmey transform provides features with higher inter-class variance and lesser intra-class variance. In addition to the statistical features namely mean, skewness, kurtosis,skewness,normal,rate,momentsecond order and third order moment were extracted from both approximation and detailed coefficients. Hence a total of 16 features were given to an ART network for speaker classification.

Case 3: Feature extraction using DWT and S transform features.

Eight features extracted from S transform and 16 features extracted from DWT were combined together. As a total, 24 features were given to an ART network for classification i.e. the values from table 1 and 2 were concatenated to form the input data set of an ART network. In all the above cases, performance of the proposed work is evaluated using Sensitivity.

$$\text{Sensitivity (\%)} = \text{TP} / (\text{TP} + \text{FN}) \text{ ----- (1)}$$

ART Network Based Text Independent Speaker Recognition System 477

Where TP is true positive ie. identifying speaker 1as speaker 1 and so on. FN is false negative ie. identifying speaker 1 as some other speaker and so on. Sensitivity based on DWT, S transform and DWT&S transforms.

4.3 Speech recognition based on proposed features

Speech recognition system involves extraction of features from the training and testing data, building BPN codebook models [4] for all digits (0-4) and continuous speeches of speakers and testing each utterance against a certain number of speech models to detect the identity of the speech of that utterance from among the speech models. The speech database used for isolated digit recognition contains isolated digits from TI digits_1 and TI digits_2. Training data for isolated digit recognition system is formed by concatenating the speeches of the set of isolated digits pronounced by 14 speakers. Another set of isolated digits is used for evaluating speaker dependent isolated digit recognition system of digits pronounced by other speakers in the database. Speaker independent continuous speech recognition system is evaluated on training data formed by concatenation of dialect sentences of 14 speakers and test data from 100 speakers in the TIMIT database.

For creating a training model, speech signal is pre-emphasized using a difference operator. Hamming window is applied on differenced speech frames of 16 msecs duration with overlapping of 8 msecs. Then the features such as PLP and MF-PLP [2-4] are extracted. For each speech, BPN codebook model is developed based on K-means clustering procedure [9] for these perceptual features. In this algorithm there is a mapping from L training vectors into M clusters. Each block is normalized to unit magnitude before giving as input to the model. One model is created for each speaker.

For testing, perceptual features are extracted for test speech. Test data can be either isolated digit or continuous speech from the database. Features extracted from each test utterance are fed to the claimant models. Then the minimum distance is found between each test vector and centroid of clusters. Average of minimum distances for each speech model is determined. The test utterance best matches with a speech model which has minimum average value.

The identification system involves extraction of features from the training and testing data, building BPN codebook models [4] for all enrolled speakers and testing each utterance against a certain number of claimant models to detect the identity of the speaker of that utterance from among the claimants. The speech database used for system evaluation contains 14speakers selected randomly from 4dialect regions in 'TIMIT' speech database.

The identification system involves extraction of features from the training data formed by combining TI random contextual variant sentences and MIT phonetically compact sentences and features from the test data formed by combining SRI dialect calibration sentences. It also involves building BPN codebook models for all enrolled speakers and testing each utterance against a certain number of claimant models to detect the identity of the speaker of that utterance from among the claimants. Present study uses the training speech of 15 seconds and test data of 4 seconds duration. Feature vectors of test speech of nearly 7 seconds duration have been considered for evaluating the performance of speaker identification system. Each speaker has been tested on an average of 56 test speech segments. All the speeches taken for analysis have been sampled at 16 kHz.

For creating a training model, speech signal is pre-emphasized using a difference operator. Hamming window is applied on differenced speech frames of 16 msecs duration with overlapping of 8 msecs. Then the features such as PLP And MF-PLP [2-4] are obtained. For each speaker BPN codebook model is developed based on K-means clustering procedure [4] for all the proposed features. In this algorithm there is a mapping from L training vectors into M clusters. Each block is normalized to unit magnitude before giving as input to the model. One model is created for each speaker.

For testing, speech signal is obtained by considering the speeches of SR1 dialect calibration sentences. The features PLP and MF-PLP are extracted for the test speech. To evaluate different test utterance lengths, the sequence of feature vectors was divided into overlapping segments of T feature vectors. The first two segments from a sequence would be. A test segment of length $T=56$ feature vectors. Each segment of T vectors was treated as a separate test utterance. Features extracted from each test utterance are fed to the claimant models. Then the minimum distance is found between each test vector and centroid of clusters. Average of minimum distances for each cluster is determined. The test utterance best matches with a cluster which has minimum average value. The performance evaluation was then computed as a percent of correctly identified T-length segments over all test.

4.4 BPN Code Algorithm

Artificial neural networks use backpropagation as a learning algorithm to compute a gradient descent with respect to weight values for the various inputs. By comparing desired outputs to achieved system outputs, the systems are tuned by adjusting connection weights to narrow the difference between the two as much as possible. The algorithm gets its name because the weights are updated backward, from output to input.

The advantages of using a backpropagation algorithm are as follows:

- It does not have any parameters to tune except for the number of inputs.
- It is highly adaptable and efficient and does not require any prior knowledge about the network.
- It is a standard process that usually works well.
- It is user-friendly, fast and easy to program.
- Users do not need to learn any special functions.
- Approach over a mini-batch approach.
- Data mining is sensitive to noise and irregularities.

The disadvantages of using a backpropagation algorithm are as follows:

- It prefers a matrix-based approach over a mini-batch approach.
- Data mining is sensitive to noise and irregularities.
- Performance is highly dependent on input data.
- Training is time- and resource- intensive

This a BPN code they have a Training and testing. The total single is a 56 single and 14 speaker.

The single odd number is a training and even number a testing single.

Each speaker 4 single.

4.5 ART Code Algorithm

ART models are designed to simulate the processes of perception, learning, and recognition that occur in the human brain. The key idea behind ART is the concept of resonance, which refers to the process of matching incoming sensory data with existing patterns stored in memory. When a new sensory input matches an existing pattern, the system resonates or responds strongly, while also adapting its internal representations to accommodate the new input. If there is no match, the system generates a new pattern in memory to represent the novel input. ART models are characterized by their ability to self-organize, learn incrementally, and adapt to changing environments. They are typically used for tasks that require online, real-time learning and recognition, such as pattern recognition, categorization, and clustering. ART models have been applied to various domains, including image recognition, speech recognition, and cognitive modeling. There are several variants of ART models, including ART1, ART2, ART3, and ARTMAP, each with its own specific architecture and learning rules. ART1, for example, is a simple binary pattern recognition model that uses a competitive learning rule to update its internal representations. ART2, on the other hand, is a more complex model that uses fuzzy set theory and a vigilance parameter to control the learning and recognition processes. ART models have been praised for their ability to adapt to changing environments, handle noisy data, and exhibit robustness in the face of uncertainty. They have been used in various applications where incremental learning, pattern recognition, and online adaptation are critical, such as in robotics, anomaly detection, and adaptive control systems. However, ART models also have limitations. They can be computationally expensive and may require careful tuning of parameters. Additionally, the biological plausibility of ART models and their exact relationship with brain processes are still topics of ongoing research and debate in the field of neuroscience and cognitive science. In summary, Adaptive Resonance Theory (ART) is a cognitive and computational framework that provides neural network models for adaptive and flexible pattern recognition and learning. ART models have been used in various applications and have shown promise in handling dynamic and noisy environments, but they also have limitations and continue to be an active area of research in the field of artificial neural networks.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 FEATURE FOR CLASSIFICATION

Those features consume a lot of time and computing power to get the result. For instance, i needed half an hour to get the result for sentiment analysis with 2algorithms for benchmarking, then something comes to my mind, is it possible to select only some words that importance to some classes.Feature selection has been a research topic for decades, it is used in many fields such as bioinformatics, image recognition, image retrieval, text mining, etc. Theoretically, feature selection methods can be based on statistics, information theory, manifold, and rough set. Feature selection methods can be classified into 4 categories. Filter, Wrapper, Embedded, and Hybrid methods. Filter perform a statistical analysis over the feature space to select a discriminative subset of features. In the other hand approach choose various subset of features are first identified then evaluated using classifiers. While The embedded approach the feature selection process is embedded into training phase of the classification. Hybrid approach takes advantages of both filter and wrapper approaches.In practice, there are some feature selections available on Scikit-learn such as Chi-squared, Variance threshold, and Mutual information. Sklearn also provides Select K Best for how many features you want to proceed. But i recently reading a lot of papers regarding this. I found numerous feature selection techniques especially on filter method, i counted more than 30 filter methods found during my research, and still on going. the features namely mean, standard deviation, kurtosis, skewness, Energy, rate, second order moment and third order moments were extracted from the speaker signals.

Table: 5.1: Approximation Coefficients Features Extracted Using DWT(Haar wavlet)

Speaker	Sample No	Mean	Standard	kurtosis	Skewness	Normal	Rate	Moment (sm)	Moment (tm)
1	1	-0.00033	0.002814	0.348874	-0.44719	0.006426108	0.324752	7.91E-06	-7.95E-08
1	2	-3.61E-07	5.46E-05	0.091069	-0.09185	0.000123915	0.362376	2.98E-09	-1.20E-13
1	3	-1.15E-05	5.98E-05	0.322057	-0.43177	0.000138082	0.366337	3.57E-09	-7.36E-13
1	4	-1.30E-06	4.35E-05	0.126143	-0.17102	9.86092E-05	0.423762	1.89E-09	-1.12E-13
1	5	-2.35E-06	0.000145	0.104145	-0.0272	0.000328428	0.455446	2.09E-08	-6.59E-13
2	1	-0.00016	0.000666	0.62018	-0.74899	0.001552521	0.009901	4.42E-07	-1.76E-09
2	2	0.000133	0.001712	0.101772	0.013562	0.003893382	0.035644	2.92E-06	5.43E-10
2	3	0.00088	0.010321	0.121527	0.143951	0.023490503	0.015842	0.000106	1.26E-06
2	4	-0.001586	0.006992	0.096531	-0.26324	0.016259531	0.009901	4.88E-05	-7.18E-07
2	5	5.77E-07	4.85E-06	1.005579	1.04319	1.10808E-05	0.007921	2.35E-11	9.51E-16
3	1	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
3	2	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
3	3	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
3	4	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
3	5	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
4	1	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
4	2	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
4	3	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
4	4	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
4	5	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
5	1	-0.000963	0.10851	0.04627	-0.04013	0.246073998	0.049505	0.011751	-0.000409
5	2	-0.003952	0.221134	0.051957	-0.00283	0.501539964	0.075248	0.048804	-0.000244
5	3	-0.001428	0.398885	0.033424	-0.03632	0.904546102	0.128713	0.158794	-0.018386
5	4	-0.000526	0.008965	0.037398	-0.00165	0.020365227	0.061386	8.02E-05	-9.50E-09
5	5	-0.013496	0.412357	0.052831	-0.09897	0.935591961	0.112871	0.169702	-0.055348
6	1	0.00575	0.271532	0.032971	0.04421	0.615883404	0.061386	0.073584	0.0070596
6	2	-0.000633	0.030356	0.036575	-0.0005	0.068852721	0.023762	0.00092	-1.12E-07
6	3	0.006813	0.217617	0.041359	0.03058	0.493725732	0.043564	0.047263	0.0025137
6	4	-0.003448	0.44341	0.032282	0.043362	1.005538099	0.033663	0.196223	0.0301525
6	5	-0.001633	0.220779	0.027245	-0.01922	0.500668141	0.029703	0.048647	-0.00165
7	1	0.003853	0.142048	0.048593	0.124516	0.322237934	0.037624	0.020138	0.0028467
7	2	-2.52E-05	0.00808	0.051397	0.088515	0.018323692	0.071287	6.52E-05	3.72E-07
7	3	-1.08E-05	0.028488	0.082602	0.02517	0.064602376	0.063366	0.00081	4.64E-06
7	4	0.00572	0.144684	0.018758	0.023781	0.328353012	0.013861	0.020892	0.0005745
7	5	-9.20E-05	0.009096	0.044702	0.033137	0.020627662	0.069307	8.26E-05	1.99E-07
8	1	-0.000494	0.009657	0.034042	-0.01762	0.021928075	0.39802	9.31E-05	-1.27E-07
8	2	-0.000182	0.006759	0.048812	0.035375	0.015332824	0.471287	4.56E-05	8.71E-08

8	3	0.000773	0.012643	0.054874	0.020026	0.028722901	0.447525	0.00016	3.23E-07
8	4	0.010944	0.097787	0.038793	-0.00012	0.223136072	0.029703	0.009543	-8.66E-07
8	5	-9.41E-05	0.029148	0.044276	-0.0496	0.06609943	0.164356	0.000848	-9.80E-06
9	1	-0.002201	0.361527	0.024861	0.061883	0.819839812	0.025743	0.130443	0.0233235
9	2	1.97E-06	0.006993	0.042786	-0.0168	0.0158575	0.249505	4.88E-05	-4.58E-08
9	3	0.000372	0.024294	0.030315	0.018869	0.055097319	0.041584	0.000589	2.16E-06
9	4	-0.000102	0.008381	0.054138	-0.01178	0.019006386	0.168317	7.01E-05	-5.53E-08
9	5	-0.000858	0.027426	0.023958	-0.02589	0.062223204	0.035644	0.000751	-4.26E-06
10	1	0.003314	0.125768	0.031649	0.056484	0.285299405	0.053465	0.015786	0.0008963
10	2	0.001446	0.131439	0.030997	0.044885	0.298077926	0.029703	0.017242	0.000813
10	3	-0.000549	0.019178	0.030318	-0.01177	0.043507951	0.039604	0.000367	-6.62E-07
10	4	0.005985	0.236657	0.032719	0.022693	0.536831581	0.049505	0.055896	0.0023991
10	5	-0.000334	0.009218	0.037307	-0.01325	0.020917968	0.045545	8.48E-05	-8.28E-08
11	1	0.007567	0.068004	0.05463	0.055943	0.155163256	0.049505	0.004615	0.0001403
11	2	0.002355	0.020266	0.036263	0.050215	0.0462675	0.063366	0.00041	3.33E-06
11	3	0.001697	0.101017	0.042125	0.022778	0.229106198	0.051485	0.010184	0.0001873
11	4	0.000867	0.018415	0.03328	-0.00079	0.041806025	0.055446	0.000338	-3.92E-08
11	5	-0.007395	0.138724	0.026505	0.024015	0.315027135	0.035644	0.019206	0.0005114
12	1	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
12	2	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
12	3	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
12	4	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
12	5	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
13	1	-0.000586	0.105166	0.056425	0.086008	0.238484661	0.071287	0.011038	0.0007979
13	2	-0.004567	0.118048	0.045312	0.072171	0.267894208	0.045545	0.013908	0.000947
13	3	9.90E-05	0.002114	0.076429	-0.04632	0.004799809	0.122772	4.46E-06	-3.49E-09
13	4	-0.00163	0.1906	0.052347	0.080789	0.432234808	0.063366	0.036257	0.0044619
13	5	-0.002344	0.038335	0.051814	-0.07864	0.087094891	0.041584	0.001467	-3.53E-05
14	1	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
14	2	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
14	3	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
14	4	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17
14	5	-1.13E-07	1.24E-05	0.167217	-0.00106	2.81042E-05	0.073267	1.53E-10	-1.62E-17

Table:5.2: Detailed Coefficients Features Extracted Using DWT (Haar wavlet)

Speaker	Sample No	Mean 1	Standard 1	Kurtosis 1	Skewness 1	Normal 1	Rate 1	Moment1 (sm)	Moment 2 (tm)
1	1	-2.38E-07	4.14E-05	0.633383	-0.79692	0.000929	0.49901	1.71E-09	-3.38E-13
1	2	-5.10E-07	8.36E-06	0.016956	0.010547	0.000188	0.619802	6.98E-11	3.69E-17
1	3	3.04E-07	8.56E-06	0.017244	0.009342	0.000192	0.566337	7.31E-11	3.50E-17
1	4	4.80E-08	8.23E-06	0.018637	0.024397	0.000185	0.546535	6.76E-11	8.14E-17

1	5	2.24E-07	1.21E-05	0.023487	0.033471	0.000273	0.530693	1.47E-10	3.58E-16
2	1	9.06E-08	4.96E-06	0.454331	0.533305	0.000111	0.083168	2.46E-11	3.90E-16
2	2	-2.57E-07	3.44E-05	0.069031	-0.02693	0.000773	0.358416	1.18E-09	-6.57E-15
2	3	-1.39E-07	9.26E-05	0.064895	0.06557	0.002079	0.30297	8.56E-09	3.12E-13
2	4	-7.59E-08	6.48E-05	0.123401	0.015202	0.001454	0.281188	4.19E-09	2.47E-14
2	5	2.49E-08	1.04E-06	0.995564	0.570404	2.34E-05	0.015842	1.09E-12	3.87E-18
3	1	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
3	2	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
3	3	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
3	4	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
3	5	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
4	1	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
4	2	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
4	3	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
4	4	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	1.19E-01	1.40E-11	-1.02E-17
4	5	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
5	1	6.85E-08	0.000553	0.030819	-0.02018	0.012408	0.522772	3.05E-07	-2.04E-11
5	2	-1.57E-06	0.001014	0.05001	0.059417	0.022771	0.457426	1.03E-06	3.71E-10
5	3	3.75E-05	0.008834	0.032415	-0.01358	0.198333	0.530693	7.79E-05	-5.60E-08
5	4	-1.07E-06	9.73E-05	0.016074	-0.00422	0.002185	0.50297	9.45E-09	-2.32E-14
5	5	2.11E-05	0.006489	0.081788	-0.01044	0.145682	0.49505	4.20E-05	-1.71E-08
6	1	1.35E-06	0.000504	0.665253	1.006933	0.011311	0.40198	2.53E-07	7.70E-10
6	2	1.22E-07	6.34E-05	1.005156	1.104154	0.001423	0.40198	4.01E-09	1.68E-12
6	3	-7.34E-07	0.000206	0.567235	-0.11317	0.004631	0.437624	4.25E-08	-5.94E-12
6	4	2.70E-06	0.00047	0.204585	0.367387	0.010561	0.50297	2.21E-07	2.29E-10
6	5	8.09E-07	0.000116	0.38023	0.10606	0.002593	0.471287	1.33E-08	9.78E-13
7	1	4.75E-07	7.72E-05	0.378609	-0.09078	0.001734	0.512871	5.95E-09	-2.50E-13
7	2	2.39E-07	3.10E-05	0.019563	0.031621	0.000695	0.560396	9.57E-10	5.62E-15
7	3	2.20E-07	5.26E-05	0.0693	-0.06073	0.00118	0.550495	2.76E-09	-5.27E-14
7	4	1.23E-07	6.21E-05	0.583928	0.708394	0.001393	0.568317	3.84E-09	1.01E-12
7	5	-4.02E-07	2.77E-05	0.02634	0.004712	0.000622	0.467327	7.67E-10	6.01E-16
8	1	1.42E-07	0.001539	0.022279	0.003299	0.034552	0.59604	2.36E-06	7.20E-11
8	2	3.82E-06	0.001076	0.023781	0.011876	0.024158	0.59802	1.16E-06	8.85E-11
8	3	-1.39E-06	0.001625	0.026178	0.030439	0.036489	0.588119	2.64E-06	7.82E-10
8	4	1.47E-06	0.0008	0.028267	-0.04242	0.017953	0.584158	6.38E-07	-1.30E-10
8	5	-2.90E-06	0.001637	0.018713	0.02764	0.03676	0.592079	2.68E-06	7.26E-10
9	1	2.04E-06	0.003244	0.026929	-0.03258	0.072838	0.762376	1.05E-05	-6.66E-09
9	2	2.93E-07	0.00087	0.017416	0.000267	0.019521	0.80198	7.55E-07	1.05E-12
9	3	3.74E-07	0.001208	0.019059	0.002642	0.027112	0.841584	1.46E-06	2.78E-11
9	4	3.51E-06	0.001116	0.032788	0.015973	0.025061	0.752475	1.24E-06	1.33E-10
9	5	-1.97E-07	0.000161	0.048491	-0.05118	0.003605	0.708911	2.57E-08	-1.27E-12
10	1	2.31E-07	0.000523	0.028523	0.055765	0.011733	0.372277	2.73E-07	4.76E-11
10	2	1.43E-06	0.000316	0.058375	-0.05744	0.007096	0.374257	9.97E-08	-1.08E-11

10	3	8.27E-07	4.60E-05	0.38006	-0.42976	0.001033	0.443564	2.11E-09	-2.50E-13
10	4	-1.10E-07	0.000306	0.190899	0.205101	0.006877	0.277228	9.37E-08	3.53E-11
10	5	-2.51E-07	1.31E-05	0.315588	-0.52477	0.000295	0.544554	1.72E-10	-7.11E-15
11	1	-8.17E-08	0.000286	0.028179	0.047043	0.006422	0.635644	8.17E-08	6.59E-12
11	2	-7.88E-09	0.000265	0.026776	0.03195	0.005944	0.655446	7.00E-08	3.55E-12
11	3	-2.42E-07	0.000431	0.019982	0.004489	0.009675	0.574257	1.85E-07	2.15E-12
11	4	-4.35E-07	0.000492	0.021035	0.01042	0.011047	0.653465	2.42E-07	7.43E-12
11	5	-9.86E-07	0.000471	0.023866	0.026843	0.010581	0.629703	2.22E-07	1.68E-11
12	1	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
12	2	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
12	3	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
12	4	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
12	5	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
13	1	-7.59E-07	0.001163	0.023575	-0.03168	0.026117	0.625743	1.35E-06	-2.98E-10
13	2	-3.17E-06	0.000846	0.025738	0.00433	0.018988	0.629703	7.14E-07	1.57E-11
13	3	-2.79E-07	0.000217	0.026875	-0.00934	0.004871	0.663366	4.70E-08	-5.71E-13
13	4	-1.55E-08	0.00076	0.020205	-0.03805	0.017065	0.69505	5.77E-07	-1.00E-10
13	5	1.14E-06	0.000338	0.028842	-0.0576	0.007578	0.625743	1.14E-07	-1.33E-11
14	1	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
14	2	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
14	3	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
14	4	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17
14	5	1.32E-09	3.74E-06	0.097758	-0.03268	8.39E-05	0.118812	1.40E-11	-1.02E-17

TABLE: 5.3 Approximation Coefficients Features Extracted Using DWT(Dmey wavlet)

Speaker	Sample No	Mean	Standard	kurtosis	Skewness	Normal	Rate	Moment (sm)	Moment (tm)
1	1	-0.00073	0.003614	0.268379	-0.43151	0.008639	0.412727	1.30E-05	-1.42E-07
1	2	-3.15E-06	5.58E-05	0.108229	-0.10121	0.000131	0.4	3.10E-09	-1.22E-13
1	3	-1.50E-05	6.43E-05	0.387162	-0.48004	0.000155	0.46	4.12E-09	-8.90E-13

1	4	-3.58E-06	4.64E-05	0.156204	-0.214	0.000109	0.510909	2.15E-09	-1.49E-13
1	5	2.13E-06	0.000151	0.128069	0.003311	0.000353	0.470909	2.27E-08	7.93E-14
2	1	-0.00023	0.000758	0.513367	-0.66103	0.001853	0.027273	5.73E-07	-2.01E-09
2	2	8.83E-05	0.001782	0.126967	-0.02347	0.00418	0.061818	3.17E-06	-9.27E-10
2	3	0.002213	0.011935	0.122339	0.185265	0.028442	0.043636	0.000142	2.20E-06
2	4	-0.00234	0.00781	0.097508	-0.26025	0.019106	0.034545	6.09E-05	-8.66E-07
2	5	7.85E-07	5.59E-06	1.009331	1.029052	1.32E-05	0.063636	3.12E-11	1.26E-15
3	1	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
3	2	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
3	3	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
3	4	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
3	5	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
4	1	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
4	2	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
4	3	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
4	4	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
4	5	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
5	1	-0.0028	0.104341	0.064942	-0.04019	0.244567	0.050909	0.010867	-0.00032
5	2	-0.00704	0.223539	0.065257	0.013953	0.524029	0.072727	0.049879	0.001088
5	3	-0.00324	0.400072	0.044422	-0.04155	0.93743	0.129091	0.159767	-0.01857
5	4	-0.00054	0.00872	0.050931	-0.00342	0.020471	0.065455	7.59E-05	-1.58E-08
5	5	-0.01975	0.426888	0.066373	-0.11431	1.001303	0.112727	0.181902	-0.06208
6	1	0.020145	0.267086	0.04247	0.033808	0.627584	0.063636	0.071205	0.004497
6	2	-0.00162	0.02969	0.049104	0.008219	0.069669	0.023636	0.00088	1.50E-06
6	3	0.007084	0.215334	0.053541	0.02987	0.504817	0.043636	0.046284	0.002082
6	4	-0.00181	0.443868	0.041733	0.053561	1.040025	0.034545	0.196661	0.032698
6	5	0.000904	0.224665	0.034615	-0.01751	0.526412	0.027273	0.050383	-0.00139
7	1	0.013728	0.149507	0.056965	0.12856	0.351782	0.038182	0.022312	0.002999
7	2	6.16E-05	0.007931	0.067337	0.095092	0.018583	0.070909	6.28E-05	3.31E-07
7	3	0.000129	0.02738	0.116787	0.027742	0.064154	0.061818	0.000748	3.98E-06
7	4	0.005237	0.146112	0.024164	0.024886	0.342571	0.012727	0.02131	0.000542
7	5	0.000419	0.009334	0.055685	0.041161	0.021892	0.067273	8.70E-05	2.34E-07
8	1	-0.00073	0.009694	0.044734	-0.0148	0.022778	0.389091	9.38E-05	-9.41E-08
8	2	-0.00025	0.00698	0.065666	0.041961	0.016366	0.478182	4.86E-05	9.96E-08
8	3	0.001045	0.012756	0.074875	0.030643	0.029988	0.449091	0.000162	4.44E-07
8	4	0.01315	0.096272	0.051486	-0.00671	0.22767	0.032727	0.009251	-4.18E-05
8	5	0.000159	0.028126	0.062505	-0.06219	0.065903	0.178182	0.00079	-9.66E-06
9	1	0.001171	0.369765	0.031761	0.065644	0.866393	0.023636	0.136478	0.023168
9	2	-5.28E-05	0.006972	0.056947	-0.01946	0.016335	0.26	4.85E-05	-4.60E-08
9	3	0.000684	0.025338	0.040243	0.020672	0.05939	0.041818	0.000641	2.35E-06
9	4	-0.00018	0.008189	0.073426	-0.01042	0.019193	0.181818	6.69E-05	-4.00E-08
9	5	-0.00211	0.027287	0.031007	-0.01883	0.064126	0.036364	0.000743	-2.67E-06
10	1	0.007828	0.123602	0.040845	0.053831	0.290191	0.052727	0.01525	0.00071
10	2	0.00559	0.134908	0.041728	0.057484	0.316372	0.030909	0.018167	0.000985
10	3	-0.0009	0.019713	0.038324	-0.0141	0.046237	0.038182	0.000388	-7.54E-07

10	5	0.011476	0.239904	0.043477	0.034913	0.562757	0.049091	0.057449	0.003365
11	1	3.82E-05	0.009177	0.05022	-0.01403	0.021502	0.049091	8.41E-05	-7.57E-08
11	2	0.011973	0.070721	0.068271	0.072934	0.168067	0.049091	0.004992	0.00018
11	3	0.001827	0.019597	0.05117	0.067794	0.046116	0.072727	0.000383	3.56E-06
11	4	-0.00017	0.10141	0.055821	0.011906	0.237612	0.050909	0.010265	8.67E-05
11	5	0.001007	0.018575	0.042664	0.005642	0.043588	0.054545	0.000344	2.52E-07
12	1	-0.00776	0.136475	0.035313	0.028116	0.320289	0.036364	0.018592	0.000499
12	2	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
12	3	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
12	4	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
12	5	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
13	1	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
13	2	0.001272	0.106397	0.071447	0.093773	0.249315	0.074545	0.0113	0.000788
13	3	-0.01121	0.116362	0.061758	0.094309	0.27391	0.043636	0.013516	0.001037
13	4	8.25E-05	0.002075	0.100892	-0.04728	0.004866	0.130909	4.30E-06	-2.95E-09
13	5	0.000458	0.18485	0.071105	0.088168	0.433118	0.061818	0.034107	0.003888
14	1	-0.00532	0.040104	0.065938	-0.09634	0.094791	0.04	0.001605	-4.34E-05
14	2	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
14	3	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
14	4	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16
14	5	-3.09E-07	1.31E-05	0.194918	-0.00811	3.06E-05	0.16	1.71E-10	-1.26E-16

TABLE :5.4 Detailed Coefficients Features Extracted Using DWT (Dmey wavlet)

Speaker	Sample No	Mean 1	Standard 1	Kurtosis 1	Skewness 1	Normal 1	Rate 1	Moment1 (sm)	Moment 2 (tm)
1	1	-3.36E-07	2.26E-05	0.754889	0.968165	0.000529	0.607273	5.09E-10	5.57E-14
1	2	-3.72E-07	8.09E-06	0.012126	0.004509	0.00019	0.554545	6.53E-11	1.19E-17
1	3	3.39E-07	7.91E-06	0.013057	-0.00143	0.000185	0.541818	6.24E-11	-3.53E-18
1	4	1.16E-07	7.98E-06	0.012921	0.037046	0.000187	0.592727	6.35E-11	9.38E-17
1	5	2.31E-07	9.13E-06	0.011772	0.01174	0.000214	0.5	8.32E-11	4.46E-17
2	1	-5.82E-08	4.20E-06	0.131508	-0.21665	9.85E-05	0.141818	1.76E-11	-8.02E-17
2	2	-1.01E-07	2.17E-05	0.051641	0.005666	0.000509	0.414545	4.71E-10	2.90E-16
2	3	2.19E-07	6.18E-05	0.066928	-0.0278	0.001448	0.361818	3.81E-09	-3.27E-14
2	4	-1.25E-08	6.10E-05	0.073498	0.015648	0.001428	0.345455	3.71E-09	1.77E-14
2	5	1.09E-12	1.42E-06	0.42366	0.000943	3.32E-05	0.08	2.00E-12	1.34E-20
3	1	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18
3	2	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18
3	3	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18
3	4	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18
3	5	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18
4	1	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18
4	2	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18
4	3	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18

4	4	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18
4	5	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18
5	1	-4.89E-07	6.76E-05	0.468535	0.725311	0.001583	0.558182	4.56E-09	1.12E-12
5	2	6.74E-08	0.000293	0.997097	-1.14353	0.006873	0.563636	8.59E-08	-1.44E-10
5	3	-1.50E-07	0.001915	0.833266	0.859925	0.044865	0.818182	3.66E-06	3.01E-08
5	4	-8.28E-07	1.28E-05	0.157389	0.198968	0.000302	0.529091	1.65E-10	2.10E-15
5	5	8.28E-08	0.000977	0.944932	0.978769	0.022889	0.790909	9.53E-07	4.55E-09
6	1	2.11E-07	0.000258	0.852882	-1.02414	0.006056	0.509091	6.67E-08	-8.82E-11
6	2	1.84E-08	4.15E-05	0.735305	-0.91869	0.000973	0.483636	1.72E-09	-3.28E-13
6	3	-1.74E-07	0.000121	0.508836	-0.17899	0.002826	0.549091	1.45E-08	-1.57E-12
6	4	4.76E-07	0.000196	0.525226	-0.24635	0.004583	0.550909	3.82E-08	-9.19E-12
6	5	3.51E-07	6.05E-05	0.489563	0.021047	0.001417	0.541818	3.65E-09	2.32E-14
7	1	5.54E-07	4.64E-05	0.484771	0.058604	0.001087	0.481818	2.15E-09	2.92E-14
7	2	1.83E-07	9.89E-06	0.030987	-0.04385	0.000232	0.529091	9.76E-11	-2.11E-16
7	3	1.85E-07	1.28E-05	0.346889	0.589216	0.000301	0.501818	1.64E-10	6.20E-15
7	4	2.23E-07	3.49E-05	0.890207	-1.07971	0.000818	0.507273	1.22E-09	-2.29E-13
7	5	-3.61E-07	9.27E-06	0.012752	0.004111	0.000217	0.458182	8.58E-11	1.63E-17
8	1	-2.27E-07	3.90E-05	0.351998	0.442457	0.000913	0.650909	1.52E-09	1.30E-13
8	2	-2.04E-08	0.000134	0.595424	-0.50495	0.003136	0.641818	1.79E-08	-6.03E-12
8	3	-1.48E-07	4.45E-05	0.546581	0.600912	0.001043	0.678182	1.98E-09	2.64E-13
8	4	-4.87E-07	6.89E-05	0.892737	-0.96872	0.001614	0.647273	4.74E-09	-1.58E-12
8	5	-7.63E-07	0.000132	0.790361	-0.6533	0.003083	0.658182	1.73E-08	-7.42E-12
9	1	-5.06E-06	0.002299	0.015246	-0.00241	0.053857	0.876364	5.27E-06	-1.46E-10
9	2	1.93E-06	0.000741	0.01298	-0.00325	0.017372	0.832727	5.49E-07	-6.60E-12
9	3	-2.26E-06	0.001174	0.015943	0.002754	0.027506	0.838182	1.38E-06	2.22E-11
9	4	7.84E-07	0.001035	0.030725	0.021147	0.024246	0.772727	1.07E-06	1.17E-10
9	5	-1.43E-07	0.000136	0.064512	-0.04039	0.003196	0.785455	1.86E-08	-5.11E-13
10	1	-1.13E-07	0.000125	0.514001	-0.80697	0.002919	0.518182	1.55E-08	-7.78E-12
10	2	-5.62E-07	6.76E-05	0.604984	0.732753	0.001583	0.530909	4.56E-09	1.13E-12
10	3	7.63E-07	2.55E-05	0.502615	0.518611	0.000598	0.512727	6.50E-10	4.30E-14
10	5	5.81E-07	0.000125	1.002697	-1.14689	0.002938	0.523636	1.57E-08	-1.13E-11
11	1	-2.46E-07	9.90E-06	0.089024	0.187454	0.000232	0.545455	9.79E-11	9.07E-16
11	2	1.19E-07	5.11E-05	0.801286	-1.01859	0.001197	0.7	2.61E-09	-6.78E-13
11	3	-1.58E-07	3.64E-05	0.613039	0.590324	0.000852	0.656364	1.32E-09	1.42E-13
11	4	-2.85E-07	7.07E-05	0.819026	-0.8698	0.001656	0.734545	4.98E-09	-1.53E-12
11	5	1.12E-07	3.65E-05	0.12702	-0.00013	0.000855	0.776364	1.33E-09	-3.20E-17
12	1	-1.96E-07	9.78E-05	0.886266	-0.94134	0.002292	0.741818	9.55E-09	-4.39E-12
12	2	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18
12	3	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18
12	4	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18
12	5	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18
13	1	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18
13	2	4.88E-07	0.000979	0.017588	0.010363	0.022937	0.605455	9.57E-07	4.85E-11
13	3	-2.38E-06	0.000862	0.02044	0.012647	0.020193	0.654545	7.41E-07	4.04E-11
13	4	-4.45E-07	0.00021	0.016503	0.015562	0.00493	0.696364	4.42E-08	7.23E-13

13	5	-5.49E-07	0.000713	0.0183	-0.05975	0.016703	0.72	5.07E-07	-1.08E-10
14	1	5.96E-07	0.00032	0.021033	-0.03198	0.007494	0.614545	1.02E-07	-5.22E-12
14	2	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18
14	3	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18
14	4	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18
14	5	8.40E-08	3.96E-06	0.055188	0.022913	9.28E-05	0.170909	1.56E-11	7.09E-18

5.2 BACK PROPAGATION NETWORK CLASSIFIER FOR SPEAKER RECOGNITION

All Neurons of a given Layer are generating an Output, but they don't have the same Weight for the next Neurons Layer. This means that if a Neuron on a layer observes a given pattern it might mean less for the overall picture and will be partially or completely muted. This is what we call Weighting a big weight means that the Input is important **and** of course a small weight means that we should ignore it. Every Neural Connection between Neurons will have an associated Weight.

Understanding Image Processing Using Artificial Neural Networks is explained by datamahadev.com. This article examines how ANN works and how it relates to image processing, useful for a beginner who is unfamiliar with this technique. Because of advances in technology, users of today's mobile devices and computers can edit and enhance images. Artificial neural networks are clusters of nodes that represent artificial neural networks. A neural network is a collection of calculations that attempt to acknowledge the fundamental connection between a collection of information by repeating a cycle similar to what the brain does. Image processing using artificial neural networks (ANN) has proven to be profitable in various fields, including Applied science, Mechanics, Geotechnical, and Industrial surveillance. A neural network uses the matrix as an input file.

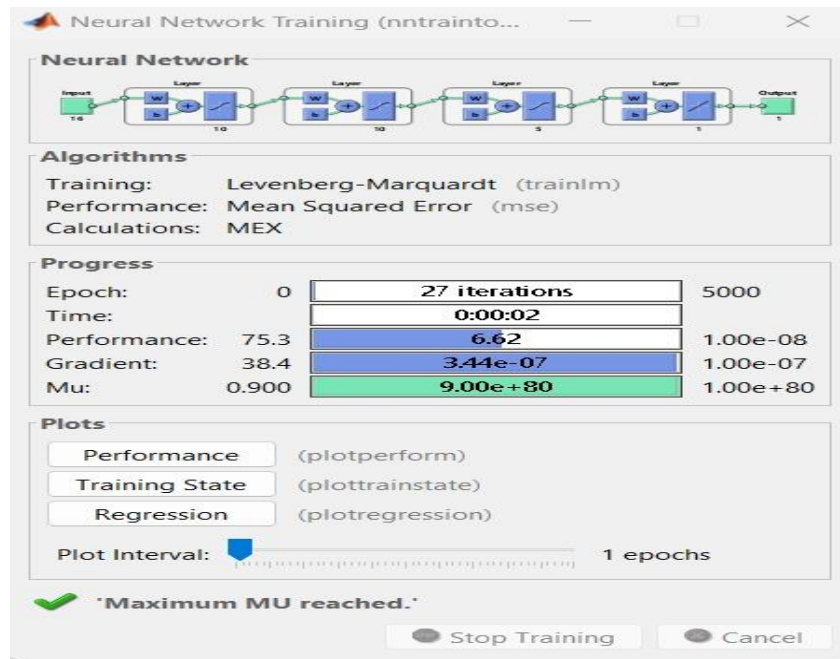


Fig:5.1 Neural network.

TRAINING

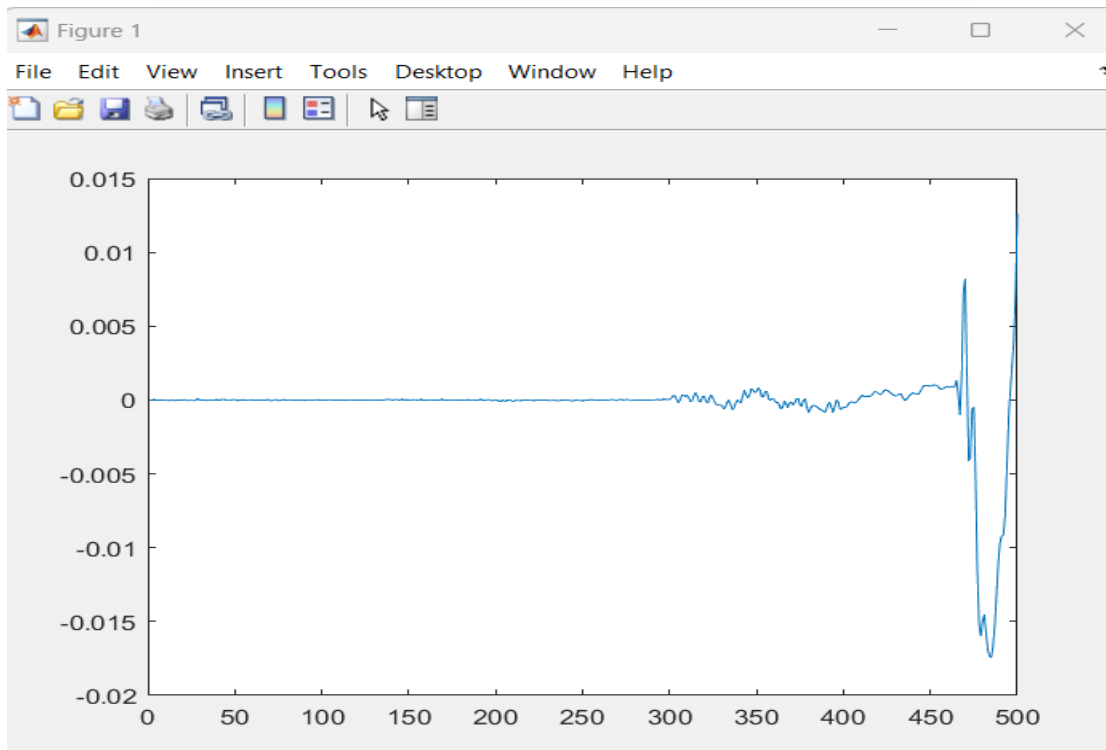


Fig: 5.2 Training

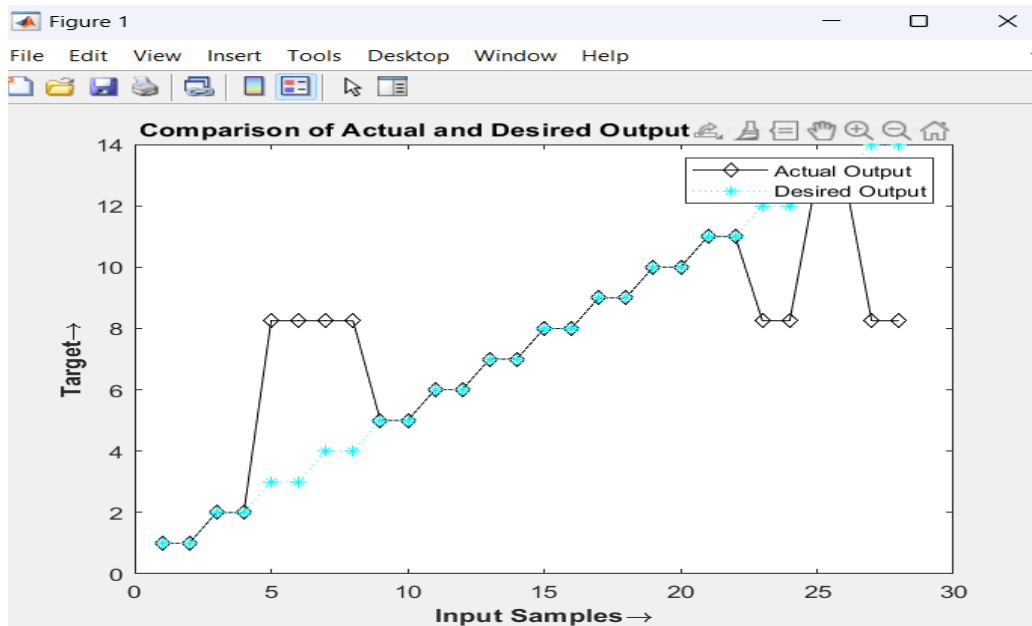


Fig: 5.3 Traing Sensistivity

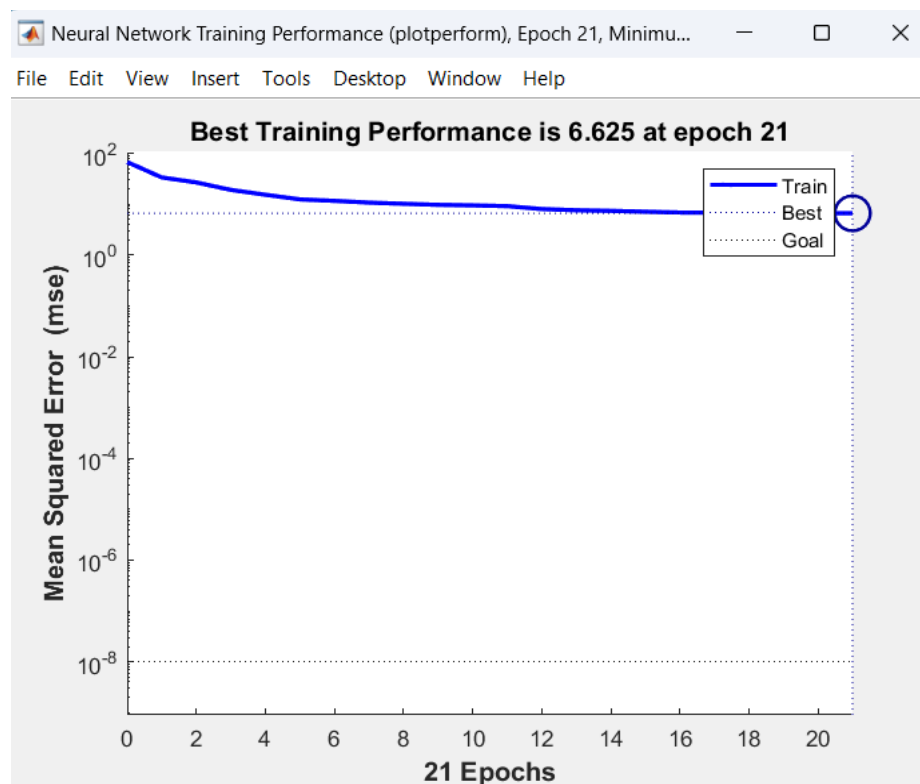


Fig: 5.4 Neural Network training performance

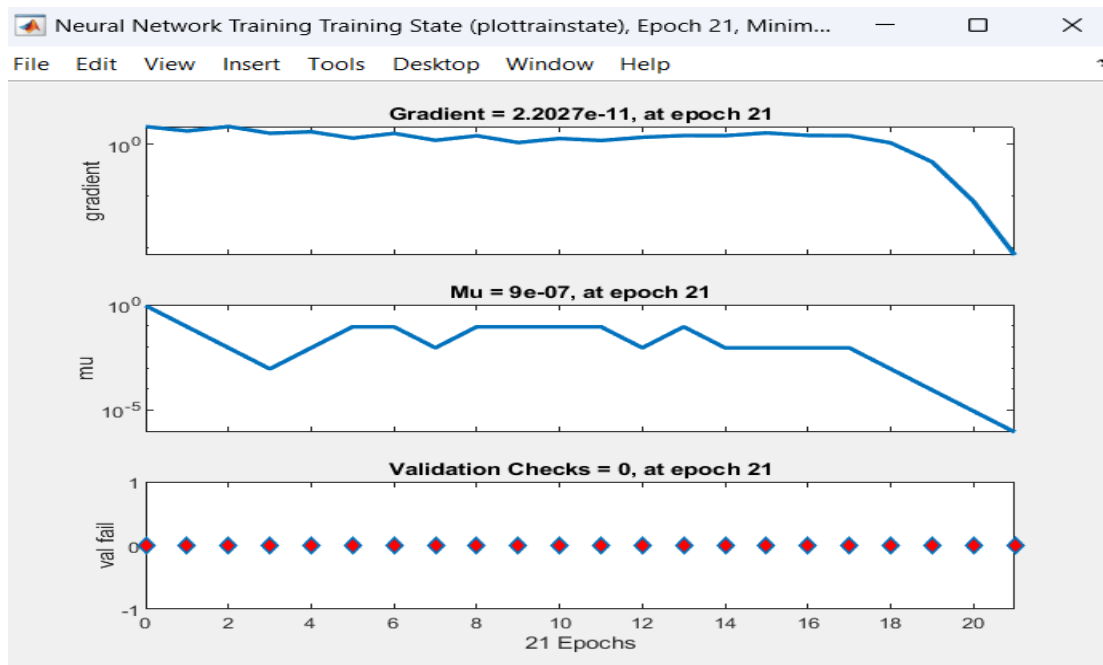


Fig:5.5 Neural Network Training state

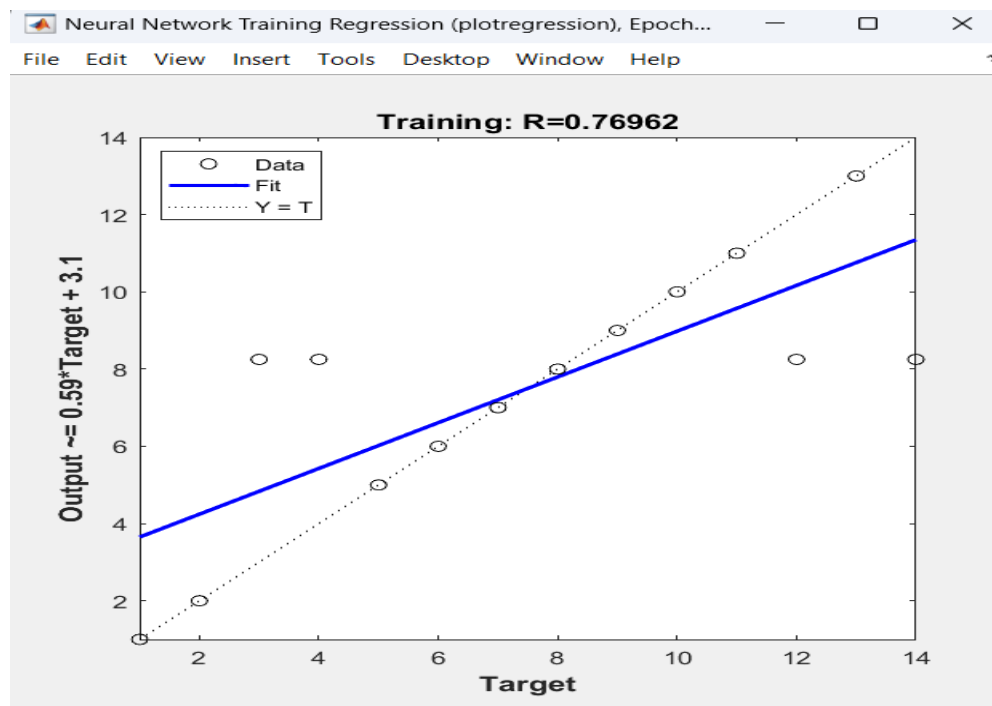


Fig:5.6 Neural Network Training Regression

5.3 ADAPTIVE RESONANCE THEORY CLASSIFICATION FOR SPEAKER FOR SPEAKER RECOGNITION

Stephen Grossberg is widely accepted as being responsible for performing extensive work relating to the development of Adaptive Resonance Theory. Grossberg completed his senior fellowship thesis. At least thirty years of his life have been spent trying to understand and recreate certain thought processes of the human brain. Stephen Grossberg's research has produced at least forty technical papers developing the basic ideas of Adaptive Resonance Theory. Most of his works relate to simulating memory activity through the use of applied mathematics and computer science. Adaptive Resonance Theory has the ability to self-organize and self-stabilize recognition code in response to an arbitrary list of input patterns. It uses a type of adaptive filter to strain out noise from an input pattern. ART has two basic configurations. The first algorithm (ART1) uses unsupervised learning techniques on binary input patterns. Bidirectional connections exist between the input layer and the output layer. Lateral inhibition exists between all of the output nodes. Pure mathematicians like this network because it can be completely described using nonlinear differential equations.

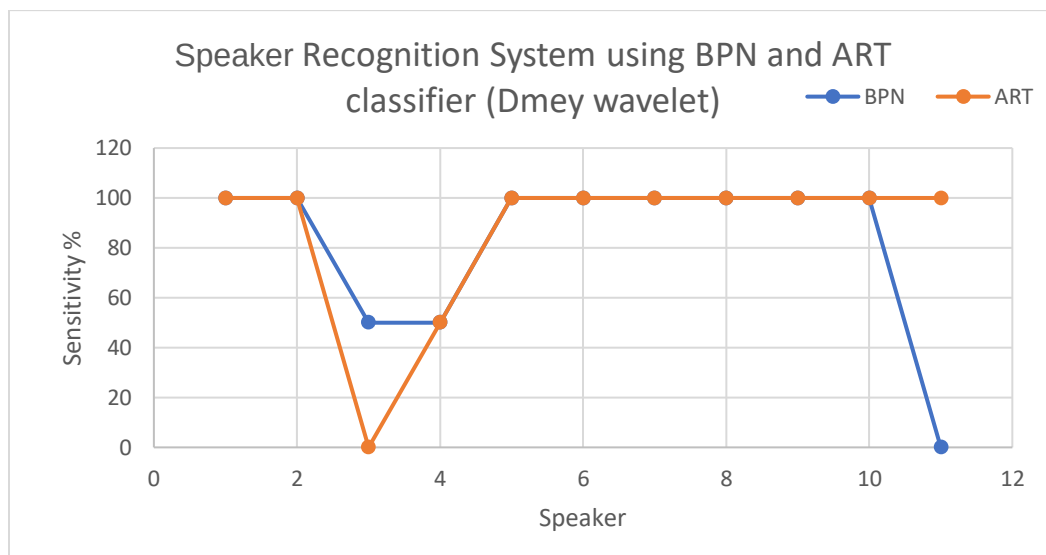
5.4 SENSITIVITY ANALYSIS OF BPN AND ART

The extracted features using Discrete Meyer Wavelet (dmey) and Haar wavelet were given to a Back Propagation Network (BPN) and Adaptive Resonance Theory (ART) for classification of speaker. The sensitivity is calculated for both classifiers and is listed in Table 3 and Table 4. A comparison graph is also plotted for both the classifiers using Dmey and Haar wavelet

Comparison of BPN and ART (Dmey wavelet)

TABLE NO :5.1 Dmey wavelet

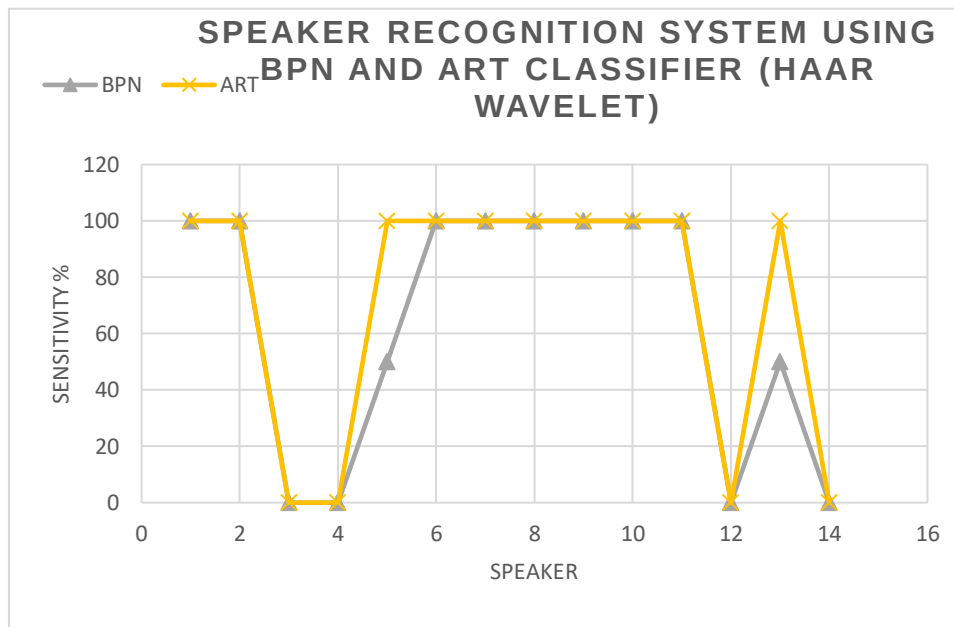
Speaker	Sensitivity using BPN (%)	Sensitivity using ART (%)
1	100	100
2	100	100
3	50	0
4	50	50
5	100	100
6	100	100
7	100	100
8	100	100
9	100	100
10	100	100
11	0	100



Graph no:5.1 Dmey wavelet

Table 5.2: Comparison of BPN and ART (Haar wavelet)

Speaker	Sensitivity using BPN (%)	Sensitivity using ART (%)
1	100	100
2	100	100
3	0	0
4	0	0
5	50	100
6	100	100
7	100	100
8	100	100
9	100	100
10	100	100
11	100	100
12	0	0
13	50	100
14	0	0



Graph: 5.2 HAAR wavelet

the figures 5.4.1 and 5.4.2, it has been proved that Dmey wavelet outperforms Haar wavelet and also ART classifier gives better sensitivity compared to BPN classifier.

CHAPTER 6

CONCLUSION AND FUTUREWORK

6.1 FUTURE WORK

This stands the chief work in talker identification for the Hadiyyisa language. Here are numerous effects to be made to growth the precision of the method. By way of we have seen throughout the investigation, the result we attained here has to be enhanced to be like other languages that are argued in the related works. Another task that we need to perform in the future is preparation of large speech databases that are to be functional in this research area so that those who are interested in the field can use these databases to save the enormous amount of time required to collect those speech utterances. Discovering other speaker dependent features other than those used in this thesis is yet another important future work. The more speaker dependent the speech features are, the better will be the identification rate of the system. In addition to all these, accent and dialect of a person can be used for the speaker identification purpose. So, these and other works not specified here are the works that we need to perform in the future in order to get the highest benefit from this field of study. We are hopeful that this thesis will be used as a start point for researchers who are interested to do their work on the area of speaker identification model designing by considering multilingual purpose.

6.2 SUMMARY

We applied the same techniques used in phonetic speaker recognition to other non-verbal cues recognition tasks including accent identification and language identification. Our classification framework performed equally well in the domains of accent and language identification. We achieved 97.7% discrimination accuracy between native and non-native English speakers. For language identification, we obtained 95.5% classification accuracy for utterances 5 seconds in length and up to 99.89% on longer utterances. The phonetic language identification technique was one component in our Mandarin Broadcast News system for the RT-04f Rich Transcription evaluation. It brought significant gains to the over all system performance.

REFERENCES

- [1] Smitha Gangisetty, Mineral Resources at West Virginia University “Text and language independent speaker recognition system for robust environment”
- [2] Francoise Beaufays, Brussels University “Text and language independent speaker recognition system for robust environment”
- [3] Hema A. Murthy Osmania, University, Hyderabad, India “Text and language independent speaker recognition system for robust environment”
- [4] Rania Chakroun and Mondher Frikha Text and language independent speaker recognition system for robust environment
- [5] G. Heigold, I. Moreno, and S. Bengio, “End-to-end textdependent speaker verification,” 2016 IEEE International Conference on Acoustics, Speech and Signal Processing
- [6] Fei Zhao, Hao Li, Xueliang Zhang “Text and language independent speaker recognition system
- [7] Selma Ozaydin “Text and language independent speaker recognition system for robust environment”
- [8] D. Snyder, D Garcia-Romero, Gregory S., D. Povey, and S. Khudanpur, “x-vectors: Robust dnn embeddings for speaker recognition,” 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
- [9] Amirsina Torfi, Jeremy Dawson, Nasser M. Nasrabadi West Virginia University “Text and language independent speaker recognition system for robust environment”
- [10] Tomi Kinnunen,a, Haizhou Li aDepartment of Computer Science and Statistics, Speech and Image Processing Unit University of Joensuu, P.O.Box 111, 80101 Joensuu , FINLAND “Text and language independent speaker recognition system for robust environment”
- [11] Sreenivasa Rao S.I.T IIT Kharagpur, 721302, India Anil Kumar Vuppala G.S.S.S.T “Text and language independent speaker recognition system for robust environment”
- [12] Y.H. Alotaibi, Y. A. Alsulaiman, S.S Al-Dahri, "A Word-Dependent Automatic Arabic Speaker Identification," 2008.

- [13] R. a. Juang, ""Fundamentals of Speech Recognition," New Jersey," 1993.
- [14] Ajinkya N. Jadhav, "A Speaker Recognition System Using Gaussian,EM Algorithm and KMeans Clustering," I.J. Modern Education and Computer Science, 2018.
- [15] S.B. Dhonde, "Speaker Recognition System using Gaussian Mixture," International Journal of Computer Applications, 2015.
- [16] York Town, USA: D. H. Beigi,"Fundamentals of Speaker Recognition," 2012.
- [17] Addis Ababa, Birhane, "Text Prompted Speaker Verificaion ModelW," 2011.
- [18] M. SRIKANTH, "Speaker Verification and Keyword Spotting Systems for Forensic," IIT Madras, chennai, 2017.
- [19] Aykefam Azene Desta, "TEXT-INDEPENDENT SPEAKER IDENTIFICATION FOR THE AMHARIC LANGUAGE," IEEE, 2015.
- [20] A Comparative Study of Feature Extraction Techniques for Speech Recognition System"