Addis Ababa University
Addis Ababa Institute of Technology
School of Electrical and Computer Engineering

Signal-based Ethiopian Languages Identification using Gaussian Mixture Model

A Thesis Submitted to Addis Ababa Institute of Technology in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Engineering

By

Mikias Wondimu

Advisor: Mr. Menore Tekeba

January, 2017

Addis Ababa University
Addis Ababa Institute of Technology
School of Electrical and Computer Engineering

Signal-based Ethiopian Languages Identification using Gaussian Mixture Model

By: **Mikias Wondimu Mekuria**

APROVAL BY BOARD EXAMINERS

Dr. Yalemzewd Negash	_____
Dean, the School of Electrical	Signature
and computer engineering	
Mr. Menore Tekeba	_____
Advisor	Signature
_____	_____
External Examiner	Signature
_____	_____
Internal Examiner	Signature

DECLARATION

I, the undersigned, declare that this thesis is my original work and has not been presented for a degree in this or any other universities, and that all source of materials used for the thesis work have been duly acknowledged.

Declared by:

Name: Mikias Wondimu

Signature: _____

Date: _____

Place: Addis Ababa institute of Technology, Addis Ababa University, Addis Ababa

This thesis has been submitted for examination with my approval as a university advisor.

Confirmed by:

Advisor's Name: Mr. Menore Tekeba

Signature: _____

Date: _____

ACKNOWLEDGEMENT

First and foremost, I would like to thank God and his mother virgin marry for giving us faith to persevere.

With humbleness and sincerity, I express my deep sense of gratitude to the advisor of my thesis Mr. Menor Tekeba for introducing me with the idea of working on the area of speech recognition. In addition I would like to pass my deepest gratitude for his patience, continuous follow-up, encouragement, insight, valuable guidance, and untiring cooperation during each stage of my research work; right from conception to completion of the thesis.

I would like to thank all instructors who responsibly thought me courses during my graduate study.

My deepest gratitude goes to all 28 volunteer speakers for giving me there valuable time and speech for recording as this work would never have been completed without their cooperation.

Last but not the least, It is my pleasure to express my heartfelt gratitude to my friends and family members for the understanding and moral support.

Table of Contents

ACKNOWLEDGEMENT i

LIST OF TABLES iv

LIST OF FIGURES v

LIST OF ACRONYMS vi

ABSTRACT vii

CHAPTER-ONE 1

1. INTRODUCTION 1

1.1. Statement of the Problem 2

1.2. Scope and Aim 2

1.3. Objective 3

1.3.1. General Objective 3

1.3.2. Specific Objective 3

1.4. Motivation for the Present Work 3

1.5. Major Contribution of the Thesis 4

1.6. Organization of the Thesis 4

CHAPTER-TWO 6

2. LITERATURE REVIEW 6

2.1. Introduction 6

2.2. Approaches to Language Identification 7

2.2.1. LID using Spectral Similarity 7

2.2.2. LID using Prosody 9

2.2.3. LID using Phone-Recognition 10

2.2.4. LID using Word-Recognition 12

2.2.5. LID using Continuous Speech Recognition 12

2.3. Text-Based LID Systems 13

2.4. Signal-Based LID Systems 14

2.5. Summary of related works 14

CHAPTER-THREE 17

3. METHODOLOGY AND DATA ANALYSIS 17

3.1. Preparing the Database Mono Channel 17

3.2. System Model 18

3.3. Pre- Processing 19
3.4. Feature Extraction 20
3.5. Classification 24
CHAPTER-FOUR 28
4. RESULT AND DISCUSSION 28
4.1. System Description 28
4.2. Experimental Setup 30
4.3. LID Performance for Two Languages Task Using GMM 30
4.4. LID Performance for Three Languages Task Using GMM 38
4.5. LID Performance for Four Languages Task Using GMM 42
4.6. Summary of the LID System Accuracy 45
4.7. Speaker Independent LID System 46
CHAPTER-FIVE 47
5. CONCLUSION AND RECOMMENDATION 47
5.1. Conclusion 47
5.2. Recommendation 49
5.3. Future Work 49
REFERENCES 50
APPENDICES 53

LIST OF TABLES

Table 3- 1:- Dataset description of utterance dependent and independent system 18
Table 4- 1:- Test result for utterance dependent/ independent of Amharic and Guragegna language..... 31
Table 4- 2:- Test result for utterance dependent/ independent of Amharic and Oromiffa language 32
Table 4- 3:- Test result for utterance dependent/ independent of Amharic and Tigregna language..... 33
Table 4- 4:- Test result for utterance dependent/ independent of Guragegna and Oromiffa language..... 34
Table 4- 5:- Test result for utterance dependent/ independent of Guragegna and Tigregna language 35
Table 4- 6:- Test result for utterance dependent/ independent of Oromiffa and Tigregna language..... 36
Table 4- 7:- Summary of the performance of the LID system for two languages task..... 37
Table 4- 8:- Test result for utterance dependent/ independent of Amharic/Guragegna/Oromiffa language..... 38
Table 4- 9:- Test result for utterance dependent/ independent of Amharic/Guragegna/Tigregna language. 39
Table 4- 10:- Test result for utterance dependent/ independent of Guragegna /Oromiffa/Tigregna language. 40
Table 4- 11:- Test result for utterance dependent/ independent of Amharic /Oromiffa/Tigregna language..... 41
Table 4- 12:- Summary of the performance of the LID system for three languages task 42
Table 4- 13:- Utterance dependent LID System Accuracy taking four languages at a time..... 43
Table 4- 14:- Utterance independent LID system accuracy taking four languages at a time 43
Table 4- 15:- Summary of the performance of the LID system for three languages task 44
Table 4- 16:- Summary of the performance of the LID system for increasing number of Languages..... 45
Table A- 1:- Sentence taken for testing 54

LIST OF FIGURES

Figure 3- 1:- System Model 19
Figure 3- 2:- MFCC Calculation Steps 22
Figure 3- 3:- Diagram of Gaussian Mixture Model 25
Figure 4- 1:- GMM for Amharic 29
Figure 4- 2:- GMM for Guragegna 29
Figure 4- 3:- Test result for utterance dependent/ independent of Amharic and Guragegna language 31
Figure 4- 4:- Test result for utterance dependent/ independent of Amharic and Oromiffa language 32
Figure 4- 5:- Test result for utterance dependent/ independent of Amharic and Tigregna language 33
Figure 4- 6:- Test result for utterance dependent/ independent of Guragegna and Oromiffa language 34
Figure 4- 7:- Test result for utterance dependent/ independent of Guragegna and Oromiffa language 35
Figure 4- 8:- Test result for utterance dependent/ independent of Oromiffa and Tigregna language 36
Figure 4- 9:- Test result for utterance dependent/ independent of Amharic/Guragegna/Oromiffa language. .. 38
Figure 4- 10:- Test result for utterance dependent/ independent of Amharic/Guragegna/Tigregna language.. 39
Figure 4- 11:- Test result for utterance dependent/ independent of Guragegna /Oromiffa/Tigregna language.40
Figure 4- 12:- Test result for utterance dependent/ independent of Amharic /Oromiffa/Tigregna language. .. 41
Figure 4- 13:- Utterance dependent LID system accuracy taking four languages at a time 43
Figure 4- 14:- Utterance independent LID system accuracy taking four languages at a time 44
Figure 4- 15:- Summary of the performance of the LID system for increasing number of Languages 45
Figure 4- 16:- Speaker Independent LID system accuracy taking four languages task for utterance dependent only 46

LIST OF ACRONYMS

GC	: Gaussian Classifier
GMM	: Gaussian Mixture Model
GMM-LM	: Gaussian Mixture Model based Language Model
GMM-UBM	: Gaussian Mixture Model based Universal Background Model
KNNC	: K - Nearest - Neighbor Classifier
LDA	: Linear Discriminant Analysis
LDC	: Linguistic Data Consortium's
LID	: Language Identification
MFCC	: Mel-frequency Cepstral Coefficient
MLC	: Maximum Likelihood Classifier
NIST	: National Institute of Standards and Technology
LRE09	: Language Recognition Evaluation Plan 2009
OGI	: Oregon Graduate Institute
P-PRLM	: Parallel-Phone Recognition Followed by Language Modeling
SDC	: Shifted Delta Cepstrum
SEAME	: South East Asia Mandarin-English
UBM	: Universal Background Model

ABSTRACT

Language Identification (LID) refers to the task of identifying an unknown language from the test utterances. The core problem in solving the language identification (LID) task is to find a way of reducing the complexity of human language such that an automatic algorithm can determine the language and identify it from a relatively brief audio sample. From the review of the existing approaches for LID, it is observed that very few attempts have been made on Language Identification System for African languages. The importance of Language Identification for African languages is seeing a dramatic increase due to the development of telecommunication infrastructure and, as a result, an increase in volumes of data and speech traffic in public networks. By automatically processing the raw speech data the vital assistance given to people in distress can be speeded up, by referring their calls to a person knowledgeable in that language.

An LID system for four different Ethiopian languages namely Amharic, Guragegna, Oromiffa and Tigregna is done using Gaussian mixture models (GMM). The system developed here is intended to identify which language is spoken by the speaker from these four languages audio utterances of some phrases for some duration. A dataset consisted of recording of 7 different speakers for each languages were prepared and after preprocessing the database mono channel, the features are extracted using Mel frequency cepstral coefficients (MFCC) and classification is done using GMMs.

To test the performance of the LID system experimental scenarios are designed and carried out by taking two, three and four languages at a time. The LID system is tested for both utterance dependent and independent system (i.e. the test is done by taking the same speech for both training and testing (utterance/speech dependent) and also by taking different speech than the training utterance (utterance/speech independent)). It is more challenging to implement and get a better LID system performance with utterance independent system with such a small recorded database. In addition to this the system also tested for the speaker independent system. The utterance dependent LID performance for four language tasks was about 93% accurate and the utterance independent LID performance for four language tasks was about 70% accurate on average. The speaker independent LID system performance for the four language task was about 91%.

Keywords: Language identification, Languages, MFCC, GMM, Accuracy, Utterance.

CHAPTER-ONE

1. INTRODUCTION

Language is a means used for human communication either in the form of speech or test. Speech is primarily intended to convey some message. The speech signal contains not only the intended message, but also the characteristics of the utterance speaker and the language of communication. The language is conveyed through the sequence of sound units. The present work focuses on signal form of speech.

With the growth of global partnership, the demand for communication across the languages is increasing. This has given rise to new challenges for automatic language recognition, followed by speech recognition system before the machine can understand the meaning of the utterance. [1]

Automatic Language Identification (LID) is the task of automatically recognizing a language from a given spoken utterance. With increasing interest in multi-lingual speech systems, such as international telephone-based information access, there has been a great deal of research in LID techniques over the last decades. The importance of having an efficient LID system dealing with large databases of languages is to allow for further processing to be carried out on the hypothesized languages. [2]

LID has various application where one application could be a telephone based front whose main work is to route the call to the corresponding operator who is knowledgeable to that language. Other application of language identification would be in the speech-to-speech translation, shopping, airports and other commercial areas.

The main goal of this research is to develop and test language identification system for the specific case of the selected languages found in Ethiopia. The system development consists of three important steps: it starts by recording and preparing the raw speech data and followed by the training stage and finally to determine the effectiveness of the system the evaluation stage will proceed.

1.1.Statement of the Problem

Due to the advance of telecommunication infrastructure in Ethiopia and, as a result, an increase in volumes of data and speech traffic in public networks. The existing system in the area of telephone based information services is an interactive voice response. Interactive voice response is an automated telephony system that interacts with the callers, gathers information and routes calls to the appropriate recipient. Developing an LID system can automatically process 2-3sec row speech data of the customer and transfer calls to the appropriate operator who is knowledgeable to that language. This will increase the processing time by removing the time it takes to interact with the caller.

In Ethiopia there are more than 18 million mobile phone subscribers [3] and in a day if we assume 100 thousand customers are calling to the operator to get information. By integrating the interactive voice response with the automatic language identification system we can save a significant amount of time and speech traffic in the public network.

Similar works have been carried out on LID systems that concentrate mostly on European and a few Asian languages. In Ethiopia similar researches have been undergone in areas of automatic speech recognition (ASR) and text based language identification. So, this work will add to the researches that have been conducted in the area of natural language processing in the country. The paper presents a signal-based language identification system for Ethiopian languages to address the following research questions:

- How many and which languages to include in the LID systems?
- How can we prepare testing and training dataset?
- Which language models are more appropriate for the proposed LID systems?
- What type of feature selection algorithm is used?

1.2.Scope and Aim

There are more than 80 languages in Ethiopia, addressing all the language at this level is difficult, so based on the number of speakers and popularity in the country. The research is mainly focused on the four languages, namely Amharic, Guragegna, Oromiffa and Tigregna.

The aim of this work is to develop and test language identification systems for the above mentioned four languages. To do so, different literatures had been reviewed in the area of language identification system.

1.3.Objective

1.3.1. General Objective

The main objective of this research work is to develop and test language identification systems based on Gaussian mixture model for Ethiopian languages.

1.3.2. Specific Objective

- To develop language models based on the suitable methods for LID system from the dataset prepared.
- To assess and select best feature selection algorithm suitable for language audio signals.
- To train and test the LID system based on the developed language model.
- To assess the performance of the LID system for an utterance dependent, utterance independent and speaker independent system.
- To enhance computational resources for LID of major local languages so that LID systems can be used for improved services.

1.4.Motivation for the Present Work

Lots of research had been carried out in the area of LID and there has been significant progress in this field. Mostly the researches were done on European and a few Asian languages [2]. One of the most important areas in natural language processing is spoken language identification. In the area of telephone based information services such as customer service, phone banking and call centers, LID system is helpful to transfer the incoming call to the corresponding agent who is knowledgeable to that language.

LID systems can also be used as an input to machine translation. Machine translation is the ability of the machine to translate the speech or a text from one natural language to another. It normally substitutes the words in one natural language for words in another.

We can also use LID systems in defending against terrorism. Government has deployed complicated systems to monitor communications among suspicious subjects. LID technology can detect the sensitive languages used by the terrorists during a telephone conversation.

Depending on the above motives, a research work is proposed to develop the Language identification systems for Ethiopian languages.

1.5.Major Contribution of the Thesis

Significance of the present work has been illustrated below:

- Different language identification approaches have been discussed.
- A research that will add to the researches that has been conducted in the country has been presented.
- The dataset that can also be used in the future similar research works has been prepared.
- The extraction and representation of language specific features based on Mel frequency capestral coefficient for LID system developed.
- The classification of the LID system using GMM was tested in different combination of languages.
- Effectiveness of the GMM for language identification task for an increasing number of language has been demonstrated.
- A GMM based LID system using MFCC has been developed for four Ethiopian languages (i.e. Amharic, Guragegna, Oromiffa and Tigregna).

1.6.Organization of the Thesis

The thesis is organized as five chapters. A brief overview of the reminder chapters and their contents are as follows.

Chapter two gives an overview of research work done in the field of language identification. The chapter begins by giving the background of LID systems. This is extended further with the detailed discussion of existing LID approaches and their related works. At the end of the chapter comparison of similar literatures with the present work is discussed.

Chapter three starts by presenting how the raw speech data and database mono channel prepared. After it shows the general pipe line of the system and the chapter flows by discussing every step of the system model. The chapter concludes how the pre-processed speech data feature extracted and classified.

Chapter four starts by discussing the tools that are used in the experiment and flows through the experimental results of the research work.

Finally, Chapter five concludes the paperwork and gives recommendations. At the end, the chapter shows how this work can be extended.

CHAPTER-TWO

2. LITERATURE REVIEW

2.1.Introduction

In this chapter, various existing research approaches have been discussed for language identification (LID) system. LID applications fall into main categories: pre-processing for machine understanding systems and pre-processing for human listeners. The LID approaches range from simple systems based on acoustic-phonetic information to complex systems based on continuous speech recognition.

There was no research in LID systems up to 1970. Even though it was started in early 1970's, there was no momentum in this area for nearly 20 years. Later, much progress was made taking the advantage of the openly available multilingual corpus of speech [4]. The LID systems implemented till data mainly vary in their methods for modeling languages.

All the LID systems can be broadly classified into two groups, namely, text-based (Explicit) and signal-based (Implicit) LID systems. All the existing LID systems use some amount of language specific information [4]. These two approaches differ only in the extent of information used for the LID task. The performance evaluation of an LID system comprises of accuracy and complexity. It mainly depends on the amount of linguistic information given to the system. [1]

In the training stage signal-based LID systems require only the speech signal and the true identity of the language. In this type of systems language models are created from the speech signals alone, which are given to the system at the time of training. The text-based LID systems may require segmented and labeled speech corpus of all languages at the time of training. Even though the performance of text based language identification systems is better than other systems, inserting a new language into such systems is a difficult task. If the number of languages under consideration is large, it is obvious to make a choice between the performance and simplicity. [1]

2.2.Approaches to Language Identification

Human beings and machines used different perceptual cues (such as phonology, morphology, syntax and prosody) to distinguish one language from the other. Based on this, to solve LID problems, following approaches are used [5]:

Human beings and machines use different perceptual cues (such as phonology, morphology, syntax and prosody) to distinguish one language from the other. Based on this, to solve LID problem, following approaches are used [5]:

I. LID using Spectral Similarity
II. LID using Prosody
III. LID using Phone-Recognition
IV. LID using Word-Recognition
V. LID using Continuous Speech Recognition

It has been observed that human beings often can identify the language of an utterance even when they have no strong linguistic knowledge of that language. This suggests that they are able to learn and recognize language specific patterns directly from the signal [5]. In the absence of higher level of knowledge of a language, a listener presumably relies on lower level constraints such as acoustic-phonetic, syntactic and prosody.

In this section, the different approaches that are used to solve the LID problems and review of some of research efforts have been discussed.

2.2.1. LID using Spectral Similarity

Spectral similarity approach for language identification is used which concentrates on the differences in spectral content among languages [5]. This is about exploiting the fact that speech spoken in different languages contains different phonemes and phones.

Acoustic-phonetic variations for various languages play an important role for language identification in spectral similarity approach. One of the earliest studies of automatic LID is based on spectral similarity approach. Earlier, language identification researchers emphasized on the variations in spectral content as different languages will have different phonemes and phones.

In spectral similarity approach, the prototype of a set of short-term spectra is obtained from training utterance and these are compared with the test speech, i.e. this system performance depends on template matching algorithm.

The main aim of an acoustic feature based LID is to capture the fundamental differences between the languages. These can be captured by modeling the distribution of spectral features which can be done by extracting a language independent set of spectral features from segments of speech. The differences in phoneme inventory, variations in the frequency of occurrence of phonemes and acoustic realization of similar phonemes cause the languages to differ from each other in their short time acoustic features [6].

Calvin Nkadimeng [2]: Demonstrated an LID system for African languages that is based on simple stochastic models has led to the implementation of various approaches. The use of GMMs in various configurations and using various MFCC-based parameterizations were evaluated. It was found that increasing mixtures led to a general improvement, but leveled out above 300 mixtures. For single GMM systems MFCC gave the best performance. However when they train the system using a Universal Background Model (UBM), small further improvements are achieved by also including acceleration coefficients. Using full covariance did not improve with use of diagonal covariance when the number of parameters increased.

Pinki Roy and Pradip K. Das [7]**:** The paper presents the efficiency of a speech dependent LID system for four different Indian languages namely Indian English, Hindi, Assamese and Bengali. The evaluation of languages is done on standard recorded databases where the features are extracted using Mel frequency cepstral coefficients (MFCC) and classification is done using Gaussian mixture models (GMMs). The results show that the accuracy of LID is best for all languages in mixture order 1024. The accuracy of LID is very good lowest up to 93% for Assamese and highest up to 100% for Bengali, Hindi and English.

Shashidhar G. Koolagudi, Deepika Rastogi and K. Sreenivasa Rao [8]: Mel-frequency Cepstral coefficient as features and the language identification model is developed for fifteen Indian languages. The identification of the languages is carried out using Gaussian mixture model. A semi natural read database is used for obtaining the language specific information and they show that the performance of Language identification system is better when trained and tested with twenty

nine features as compared to six, eight, thirteen, nineteen and twenty one MFCC Features. The average language recognition rate over fifteen Indian languages is around 88%.

YonghuaXu, Jian Yang and Jiang Chen [9]: Introduce methods for improving Gaussian mixture model classification which includes GMM-UBM (universal background model) for language identification was proposed. Training with GMM can be very time-consuming with large number of mixture components so GMM in combination with UBM was proposed. LDC consisting of 21 languages divided into training, development and testing set with 8 kHz frequency is used as database. Here MFCC is used as feature extractor and when LDA in combination with GMM-UBM as backend classifier is used, it achieves a higher average accuracy rate 80.6%.

Pedro A. Torres-Carrasquillo, Douglas A. Reynolds and J.R. Deller [10]: Here speech is given as input to the system and after features are extracted using MFCC, GMM tokenizer is used to assign an incoming feature vector which is used to partition the acoustic space. After doing language modeling for 12 languages using probability, GMM is used as a back-end classifier. Results depict that the fusion of scores from Parallel-Phone Recognition Followed by Language Modeling (P-PRLM), GMM tokenizers and GMM acoustic systems produces an error rate of 17% on the NIST 1996 evaluation test set (one of the lowest error rates published on this benchmark test).

2.2.2. LID using Prosody

The prosody offers an enhancement to spectral, phoneme or word-based LID systems which are robust to noise. Languages have characteristic sound patterns which can be analyzed in terms of duration of phonemes, speech rate, intonation (pitch contour), and stress (short term energy) [11].

Prosodic information refers to the duration characteristics of phones, intonation (pitch variation) and stress patterns. Some phones are shared across multiple languages, however their duration characteristics will depend on the phonetic system of the language. Intonation is the variation of tone or pitch used when speaking. Such a variation can convey different interpretations of a sentence in some languages. Variation of pitch can convey different meanings in some languages.

There is a lot of variation in the prosodic properties of languages. In the prosodic structure of the spoken utterance, the main elements like fundamental frequency (F_0), duration and voice intensity

are used. The way in which these are introduced into the prosodic structure of a spoken utterance varies for different languages. The differences across languages can often be observed in the realization of the prosodic features which determine the tones or stress contained throughout an utterance. For example, it has been shown that in American English (a stress language) and Mandarin Chinese (a tone language), the fundamental frequency patterns of continuous speech will display different characteristics [1].

Ann Thyme-Gobbel and Sandra E. Hutchins [11]: Demonstrated a running averages and correlations of prosodic features capturing syllable pitch and amplitude contours, duration and phrase location. Results show that prosody is highly useful in LID if complex perceptual events are broken down into simpler physical events and features are chosen based on task.

Bo Yin, Eliathamby Ambikairajah and Fang Chenv [12]: it combines cepstral features and prosodic features. This combination approach shows a significant improvement on a GMM- UBM based LID system which utilizes modern shifted delta cepstrum (SDC) and feature warping techniques. The proposed system achieves a high accuracy of 87.1% on a 10-language task.

David Martinez, Lukas Burget, Luciana Ferrer and Nicolas Scheffer [13]: an automatic language recognition system that extracts prosody information from speech and makes decisions about the language with a generative classifier based on iVectors is built. The system is tested on the NIST LRE09 dataset. The prosodic system (2048 Gaussians, 400- dimension iVectors) and the fusion of both systems improve performance in all conditions. The relative improvements obtained over the acoustic system are: 10.93% for 3 seconds; 15.24% for 10 seconds; and 9.39% for 30 seconds.

2.2.3. LID using Phone-Recognition

There is a finite set of meaningful sounds which appear in human languages that can be produced physically by humans. Not all these sounds appear in any given language, hence each language has its own finite subset of meaningful sounds. A phoneme is the abstract sound unit of the phonetic system of a language capable of conveying a distinction in meaning. In addition, sounds that are different but accepted as the same phoneme in a language are called allophones (or a phone in terms of the physically produced sound). Phonetic features refer to a sequence of sound unit that can be extracted from speech.

Different languages can have different set of phones representing their phonemes. Phoneme/phone of frequencies of occurrence may also differ, where a phoneme/phone may occur in two languages but it may be more frequent in one language than the other.

Phonotactics, the rules governing the sequences of allowable phonemes, can be different as well. Hence, phonetic information appears to be extremely suitable information for exploiting the characteristics of a language.

L.F. Lamel and J.L. Gauvain [14]: Demonstrated phone-based acoustic likelihoods to the problem of language identification using laboratory quality speech. With 2 sec of speech the LID perfumes around 99% accuracy on average. On spontaneous telephone speech from the Oregon Graduate Institute (OGI) corpus, the language can be identified as French or English with 82% accuracy with 10s of speech. The 10 language identification rate using the OGI corpus is 59.7% with 10s of signal.

J.L. Gauvain, A. Messaoudi, and H. Schwenk [15]: This paper proposes a new phone lattice based method for automatic language recognition from speech data. In this work three phone recognizers were used to produce phone lattices for each training and test segment. On the NIST Eval03 language recognition test set, the lattice based method reduces the equal error rate from 6.8% to 4.0% for the 30s segments, with smaller gains for the shorter segments. When the scores corresponding to the 3 phone recognizers are combined with a neural network, the equal error rate for the 30s segments is further reduced to 2.7%. This makes a very competitive language recognition system running in about 0.5xreal-time (i.e. the lattice-based language identification system runs faster than real-time).

V. Ramasubramanian A. K. V. Sai Jayram T. V. Sreenivas [16]: The paper demonstrated some of the unexplored issues in the parallel phone recognition (PPR) system for automatic language identification (LID). It considers three types of scores for LID, namely, the acoustic score, language model score, and the joint acoustic-language score. Using each of these scores it formulates three types of classifiers for performing LID: maximum likelihood classifier (MLC), Gaussian classifier (GC) and K - nearest - neighbor classifier (KNNC) and compare their performances. Among all the different combinations of scoring methods and classifiers, it is found that MLC with bias-removal performs best for either acoustic or language model score alone; this is closely followed by GC with the joint acoustic language score.

2.2.4. LID using Word-Recognition

LID systems based on words employ sequence modeling at word level. This is an approach to the LID problem where phones are recognized first, followed by words, and eventually language. To model the LID system at word level one should know the morphology and the syntax of the language at word level.

Morphology is the study of word structure. Hence to perform an LID system at word level, examining the characteristics of word forms such as inflection, derivation and compounding is necessary. The word roots, lexicons and vocabulary are usually different from language to languages.

Syntax is the rules that govern how the words in a sentence are connected. The sentence patterns are different between languages. Even when two languages share a word, the word sequence that precedes and follows the word will be different.

Koena R. Mabokela, Madimetja J. D. Manamela and Nalson Gasela [17]**:** This paper presents an approach to the development of the automatic LID system on mixed-language speech. Speech corpus to be used for simulation involves mixed utterances of Northern Sotho (aka Sepedi) and English. Language boundary detection methods are used to identify multiple languages within an utterance. Overall, the research work aims at ultimately enhancing the performance of a general-purpose speech recognizer with automatic LID capabilities.

2.2.5. LID using Continuous Speech Recognition

LID systems based on continuous speech recognition uses more language-specific knowledge to get better performance [1]. It employs one continuous speech recognizer per language. While testing, all these recognizers runs in parallel. The language of the recognizers which yields the highest likelihood is hypothesized as the language of the test utterance.

For the systems which use higher-level knowledge (words and word sequences) rather than lower-level knowledge (phones and phone sequences), the identification performance is better than other simple systems. On the other hand, they require several hours of labeled training data for each language to create separate continuous speech recognizers.

Ngoc Thang, Dau-Cheng Lyu etal [18]: This paper presents first steps toward a large vocabulary continuous speech recognition system for conversational Mandarin- English code-switching speech. The paper applied state-of-the-art techniques such as speaker adaptive and discriminative training to build the first baseline system on the SEAME corpus (South East Asia Mandarin-English). On language model level, it investigated statistical machine translation based text generation approaches for building code-switching language models. The system best 2-pass system achieves a Mixed Error Rate of 36.6% on the SEAME development set.

2.3.Text-Based LID Systems

Text based language identification system is the task of automatically recognizing a language from a given text of document [19]. Text-based LID systems, instead of language models, the phoneme recognizers are used as a front-end. Using phone-recognizers the phoneme sequence is generated. This approach requires large amount of segmented and labeled speech corpus in different languages to train the phone recognizers [1].

In [20] it presents the text based language identification system for Ethiopian Cushitic languages namely, Afaan Oromo, Afar, Sidama and Somali. In this research n-gram frequency rank order and Naïve Bayes were compared as language identifier. To evaluate the models a document size of 15, 100. And 300 characters windows were used. The classification accuracy of 99.78% on average was achieved for text document as short as 15 characters.

In [19] it presents text based language identification system for Indian languages following Devanagiri Script. The paper investigates the performance of statistical measures to determine the text- based language identification system, with an emphasis on five languages used in India. The proposed system was trained using a corpus of languages like Hindi, Sanskrit, Marathi, Bhojpuri and Nepali having size 2MB. The proposed LID uses two level monograms, bigrams and trigrams.

In [21] it investigates the performance of text-based language identification systems on the 11 official languages of South Africa, when n-gram statistics are used as features for classification. In particular, it compare support vector machines, likelihood and frequency difference-based classifiers on different amounts of input text and for various values of n of the n-gram (i.e. an n-gram is a contiguous sequence of n items of letters or words from sequence of text). the it is found that acceptable language-identification accuracy can be obtained with as few as 15 words of input

text (in fact, even with 2 words somewhat useful results are obtained). During the tests, the support vector machine performs with better accuracy than the likelihood and frequency difference- based classifiers when employed under the same circumstances.

2.4.Signal-Based LID Systems

In signal-based LID systems, the features are extracted directly from speech signal and using these features, the language models are created. Signal based LID systems neither depend on the phone recognizers nor do they require segmented and labeled speech corpus of the target language. In other words, these LID systems need only the raw speech utterance along with the true identity of the language being spoken. Here, the language models or language specific information is derived only from the speech data.

The features selected for differentiating languages may be different in existing signal based LID systems, i.e. they differ mainly at the feature extraction stage. The classification used in LID systems may be also different depending on the researcher's selection.

2.5.Summary of related works

From the review of the existing LID systems in this chapter, it is noticed that the choice between the signal-based and text- based LID systems is a compromise between performance and complexity. The text-based LID systems that make use of phonotactics and word level knowledge perform better than the signal-based LID systems which rely on acoustic-phonetics and prosody features extracted directly from the speech signal. But the higher performance of text-based LID systems is achieved at the cost of additional complexity of using phone recognizer at a front-end. Creating such a phone recognizer is a difficult task, because it requires segmented and labeled speech corpus of target languages [1]. Such a segmented and labeled speech corpus may not be available in many languages. This is especially true for Ethiopian languages.

Hence, for the Ethiopian languages using a text-based LID systems is not desirable due to the following reasons:

- Getting a segmented and labeled speech corpus especially for Guragegna language is difficult.
- Similarities in letters between the languages used in this research.
- Inserting a new language into text-based LID system is a difficult task.

Therefore, from the reviewed different approaches that are used to solve LID problems the researches focuses on the signal-based LID systems.

From the review of the signal-based approaches (acoustic-phonetics and prosody approaches) the research focuses on the acoustic approaches of an LID system. Acoustic-phonetic approaches is advantageous over other signal-based LID systems is that it doesn't require language specific knowledge. Because of this, the development and insertion of new language into the system is not a difficult task.

In [7] the paper presents a signal based utterance dependent language identification system for four Indian languages. The paper uses MFCC as a feature extractor and GMM as a classifier. In preparation of the dataset 15 speakers is used for each language. Each speaker instructed to utter the same sentence 20 times and out of which 15 were used for training and 5 were used for testing the model. They have tested the LID system with increase in GMM mixture order (2, 4, 8, 16… 1024). The performance of the utterance dependent language identification is best for GMM mixture order of 1024.The accuracy of LID system is very good lowest up to 93% for Assamese and highest up to 100% for Bengali, Hindi and English.

In [8] the paper presents a signal based speech independent language identification system for fifteen Indian languages. They use MFCC as feature extractor and GMM as a backend classifier. For preparation of the dataset 5 mints of speech data were collected for each language, out of which 3.5 minutes are used to train the model and 1.5 mints of data has been used for testing the model. The average language recognition rate over fifteen Indian languages is around 88%.

In comparison with others, this paperwork demonstrated a signal based language identification for both speech dependent and independent system for four Ethiopian languages. In addition to this, the paper also presents a speaker independent language identification system. This paper also uses the most common feature extraction technique called MFCC and a backend classifier called GMM. Some works in the field of LID system use a higher GMM mixture order for a better accuracy of

the system, but higher mixture order needs a higher hardware resource. Due to this, this work is tested for only 16 GMM mixture order depending on the hardware resource used for the experiment.

CHAPTER-THREE

3. METHODOLOGY AND DATA ANALYSIS

3.1.Preparing the Database Mono Channel

For preparation of the database mono channel recording was done for 28 speakers in the Amharic, Guragegna, Oromiffa and Tigregna language in relatively closed and quiet noise-free room. For digitization, 16 kHz of sampling frequency and 16-bit quantization were used. All the speakers were male speakers in the age group of 20-30. Speakers are taken in such a way so that they are native speakers of their respective language. Appendix 'A' shows the paragraph and the sentence that was taken for training and testing. A recording of 7 different speakers for each language was done.

The LID system is tested for both utterance dependent and independent system (i.e. the test is done by taking the same speech for both training and testing (utterance/speech dependent) and also by taking different speech than the training utterance (utterance/speech independent)). The last experiment is done to test the performance of the speaker independent LID system. The speaker independent system is tested by creating biometrical disjoint sets in the training and testing dataset (i.e. out of 7 speakers of each language, 4 speakers utterances is used to train the system and the other 3 speakers utterances is used to test the accuracy of the system).

To train and test the utterance dependent and independent system properly, each speaker is instructed to utter the same paragraph (∼1min long) for 15 times and the same sentence ((3-5) sec long) for 5 times. Out of 15 recorded ∼1min long database, 10 were used for training the system and the remaining 5 utterances were used for testing the accuracy of the model for the utterance dependent LID system. All (3-5) sec long speech database was used to test the accuracy of the model for the utterance independent system. Thus, for training we have a total of 280 samples and the detail on the database mono channel is shown in Table 3-1 below.

Table 3- 1:- Dataset description of utterance dependent and independent system

Language	No. of Speaker (N)	No. of times each speaker utters	Total training sample (N*10)	Total testing(N*5)		Sample Frequency
				Utterance dependent sample(1 min long each) (N*5)	Utterance Independent sample(3-5sec long each) (N*5)	
Amharic	7	15	70	35	35	16KHz
Guragegna	7	15	70	35	35	
Tigrigna	7	15	70	35	35	
Oromiffa	7	15	70	35	35	
Total	**28**	**60**	**280**	**140**	**140**	

3.2.System Model

The LID system comprises of pre-processing, feature extraction and classification. First the raw speech of known language is given to pre-processing. Once the speech is pre-processed the next step is extracting important features from the speech signal. The important spectral features are next given to GMM for training and creating a model for each language. Once the system creates the unique model for each language and it will be compared to the speech of the unknown language. After choosing the best and the most likelihood languages from all models the system gives its decision. Figure 3-1 shows the whole processes that have been carried on for language identification task.

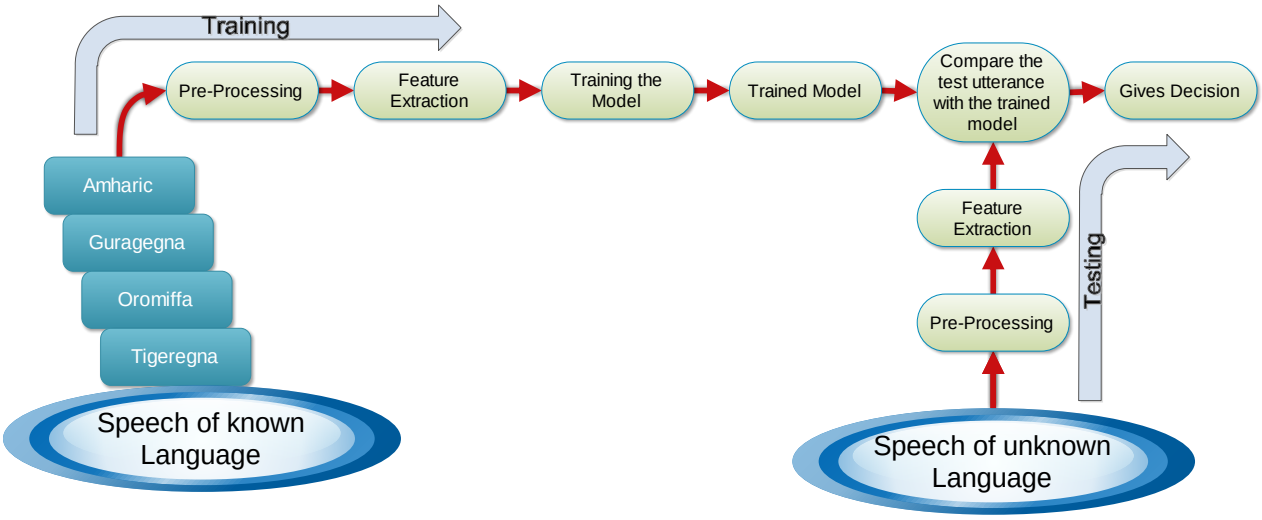

Figure 3- 1:- System Model

3.3.Pre- Processing

The first step in LID system is processing of the row speech data to be compatible with the feature extraction used for the study. Pre-processing speech is converting of the analog representations (first air pressure, and then analog electrical signals in a microphone) into a digital signal. The process of analog-to-digital conversation of speech signal has two steps: sampling and quantization.

A signal is sampled by measuring its amplitude at a particular time and to have a better sampling it is necessary to have at least two samples in each cycle (i.e. one measuring the positive part of the wave and one measuring the negative part of the wave). More than two samples per cycle increases the amplitude accuracy, but less than two samples will cause the frequency of the wave to be completely missed. According to the Nyquist-Shannon sampling theorem a time-continuous signal $x(t)$ that is bandlimited to a certain finite frequency fmax needs to be sampled with a sampling frequency of at least 2fmax [22].

Most information in human speech is in frequencies below 10 KHz; thus taking Nyquist theorem into account a 20 KHz (2fmax) sampling rate would be necessary for complete accuracy. A 16 KHz sampling rate (sometimes called wideband) is often used for microphone speech. Since the speech is recorded using microphone the appropriate sampling for the present work is 16 KHz sampling. 16 KHz sampling rate requires 16000 amplitude measurement for each second of speech, and so it is important to store the amplitude measurement efficiently. They are usually stored as integers, either 8 bit or 16 bit. This process of representing real-valued numbers as integers is called

quantization.

The process of analog to digital conversion of the wave form is done using software called wavesurfer 1.8.8p4. Now that we have a digitized, quantized representation of the waveform, the next step is extracting spectral feature vector from the speech signal.

3.4.Feature Extraction

The speech signal cannot directly given to the LID system (i.e. weaker signal has to be amplified, longer silences have to be removed and speech with background noise is to be extracted for further processing).

A feature vector should emphasize the important information regarding the specific task and suppress all other information. As the goal of automatic speech recognition is to transcribe the linguistic message the information about this message needs to be emphasized. The speaker dependent characteristics, the characteristics of the environment and recording equipment should be suppressed because these characteristics do not contain any information about the linguistic message. Including this non-linguistic information introduces an additional variability, which could have a negative impact reparability of the phone classes. Furthermore, the feature extraction should reduce the dimensionality of the data to reduce the computation time and the number of training samples [23].

The feature analysis component of the LID system plays a crucial role in the overall performance of the system. Many feature extraction techniques are available, these include

- Linear Predictive Coding (LPC)
- Perceptual Linear Predictive Coefficients (PLP)
- Dynamic Time Warping (DTW)
- Relative spectral filtering of log domain coefficients (RASTA)
- Mel-frequency cepstral coefficients (MFCC)

Linear Predictive Coding (LPC):

The idea behind the Linear predictive coding analysis is that a speech sample can be approximated as a linear combination of past speech samples. LPC is a frame based analysis of the speech signal which is performed to provide observation vectors of speech [24]. To compute LPC features, initially the digitized speech signal is put through a low order digital system. The output of the pre-empesizer network is blocked into frames of N samples. After frame blocking, the next step is to window each individual frame so as to minimize the signal discontinuities at the beginning and end of each frame. The next step is to auto correlate each frame of windowed signal. Finally, the LPC analysis that converts each frame of autocorrelations into LPC parameter set by Durbin's method. [25]

Perceptual Linear Predictive Coefficients (PLP):

The perceptual linear prediction model developed by Hermansky. The PLP speech analysis technique is based on the short-term spectrum of speech and it provides the human speech based on the concept of the psychophysics of hearing. PLP discards irrelevant information of the speech and thus improves speech recognition rate. It is identical to LPC except that its spectral characteristics have been transformed to match the characteristics of the human auditory system. [26]

Dynamic Time Warping (DTW):

DTW is a time series alignment algorithm developed originally for speech recognition. It aims at aligning two sequences of feature vectors by warping the time axis iteratively until an optimal match between the two sequences is found.

Dynamic time warping (DTW) is an algorithm for measuring similarity between two temporal sequences which may vary in time or speed. This technique is also used to find the optimal alignment between two time series. If one time series may be "warped" non-linearly by stretching or shrinking it along its time axis. This warping between two time series can then be used to find the corresponding regions between the two time series or to determine the similarity between the two time series. [24]

Relative Spectral Filtering log domain coefficients (RASTA):

The term RASTA comes from the word RelAtive SpecTrA. RASTA processing is studied in a spectral domain which is linear-like for small spectral value and logarithmic-like for a large spectral value. The rate of change of non-linguistic components in speech often lies outside the typical rate of change of vocal tract shape. RASTA filtering is often coupled with PLP for robust speech recognition. It is a separate technique that applies a band-pass filter to the energy in each frequency sub band in order to smooth over short-term noise variations and to remove any constant offset resulting from static spectral coloration in the speech channel [27].

Mel-frequency cepstral coefficients (MFCC):

Till now, for conventional LID systems, features are extracted using Mel-frequency cepstral coefficients (MFCC) [1]. This paper focuses on this most popular, prevalent, widely used and efficient technique for feature extraction [2] [1] [27]. Humans have the ability of distinguishing languages without having a much knowledge of that language. The idea behind language identification is the representation of human capability into machine understanding. The speech generated by humans is filtered by the shape of the vocal tract and representing this shape accurately is the main job of feature extraction techniques. The advantage of MFCC is that its ability to represent this shape in a more appropriate and accurate way.

After getting a digitized, quantized representation of the raw speech data, the next step is extracting the spectral feature vector from the speech signal. MFCC can be a tool to represent the signal with its important spectral feature vector.

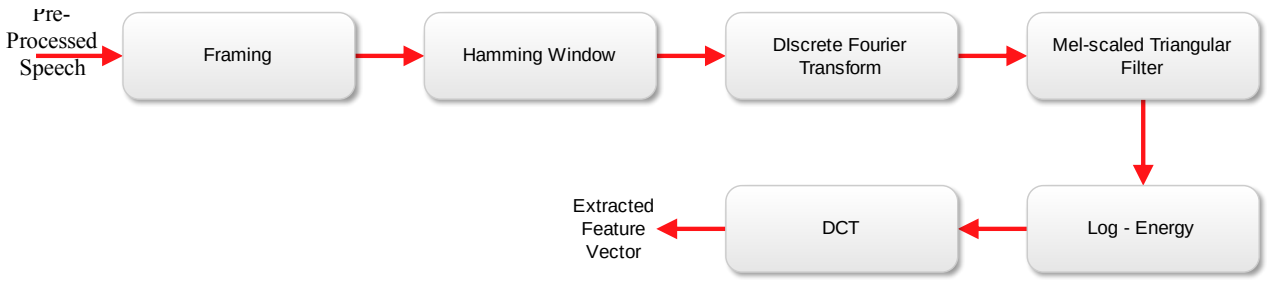

Figure 3- 2:- MFCC Calculation Steps

As stated in the above Fig 3-2 here are the steps in calculating MFCC feature vectors

- Frame the signal into short frames.
- For each frame calculate the periodogram estimate of the power spectrum.
- Apply the mel filter bank to the power spectra, sum the energy in each filter.
- Take the logarithm of all filter bank energies.
- Take the DCT of the log filter bank energies.

The main goal of feature extraction is to provide a spectral feature that can help us to build phone or sub-phone classifier. Hence, extracting spectral feature from the entire utterance is impossible, since the spectrum changes very quickly. Speech is a non-stationary signal, meaning that its statistical properties are not constant across time. Instead, we want to extract spectral features from a small window of speech that characterizes a particular sub-phone and for which we can assume that the signal is stationary. The speech extracted from each window is called a frame.

The more common window used in MFCC extraction is the Hamming window, which shrinks the values of the signal near to zero at the window boundaries, avoiding discontinuities. The hamming window is described by: [28]

$$w[n] = 0.54 - 0.46\cos(2\pi n/L) \qquad 0 \leq n \leq L-1 \quad \text{..... 3-1}$$

Where,

$w[n]$ is the value of the window at time n.

L is the speech extracted from each window (frame)

The next step is to extract spectral information for our windowed signal, we need to know how much energy the signal contains at different frequency bands. The extraction of spectral information for discrete frequency bands for a discrete time signal is the Discrete Fourier Transform (DFT).

The result of DFT will be the information about the amount of energy at each frequency band. Human hearing, however, is not equally sensitive at all frequency band. The advantage of applying the mel-scale is that it approximate the nonlinear frequency resolution of the human ear. [8]

The formula for converting from frequency to Mel scale is:

$$M(f) = 1125 \ln(1 + f\backslash 700) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3\text{-}2$$

Where;

$M(f)$ − Denotes the mel scale in frequency domain

f −Denotes frequency

Once we have the mel fiter bank energies, we take the log of each of the mel spectrum values. In general the human response to signal level is logarithmic; humans are less sensitive to slight differences in amplitude at high amplitudes than at low amplitudes. In addition, taking logarithms allows us to use cepstral mean subtraction, which is a channel normalization technique. [28]

Finally, a discrete cosine transform is applied to the log of the filter bank outputs results in the raw MFCC vector. The highest cepstral coefficients are omitted to smooth the cepstral and minimize the influence of the pitch which are irrelevant to the language identification process.

3.5.Classification

The most popular classification techniques in the area of language identification are Deep Neural Network (DNN) and Gaussian Mixture Model (GMM).

A deep neural network (DNN) is a feed-forward, artificial neural network that has more than one layer of hidden units between its inputs and its outputs. DNNs are more accurate classifier. DNN training is extremely time consuming, even with the aid a graphical processing unit (GPU) or lots of CPU. The DNN training algorithm is very complicated because it's not guaranteed to converge to an optimal point. [29]

Compared to DNNs, GMM is faster to compute, easier to learn. GMM training is reliable, if we have a clean data, it guaranteed to train a good system. The GMM approach to classification has been widely used in a variety of language processing applications. The system structure of a GMM based LID system is very simple. Accordingly, the computational requirements for processing are low. This simplicity advantage also extends to the development phase for such a system. The GMM LID system has significant potential advantage over LID systems since they do not require orthographically or phonetically transcribed speech and are far more computationally efficient. [10]

GMM Based Language Identification System

Basically, a GMM based LID system is a classifier with each class (a language in this task) modelled by a GMM. Language classification is performed according to the likelihood score calculated by the language-GMMs against a given feature vector. To determine the language in LID testing, multiple feature vectors are used. That is, likelihood scores are accumulated for each language and the decision making is delayed until all the feature vectors are processed. [4]

In a GMM model, the probability distribution of the observed data takes the form given by the following equation, [1]

$$\rho(\bar{x}|\lambda) = \sum_{i=1}^{M} \rho_i b_i(\bar{x}) \quad \text{.....3-3}$$

Where, M is the number of component densities, $\bar{x}$ is a D dimensional observed data, $b_i(\bar{x})$ is the component density and ρ_i is the mixture weight for $i = 1, \dots, M$ as shown in Fig. 3-3.

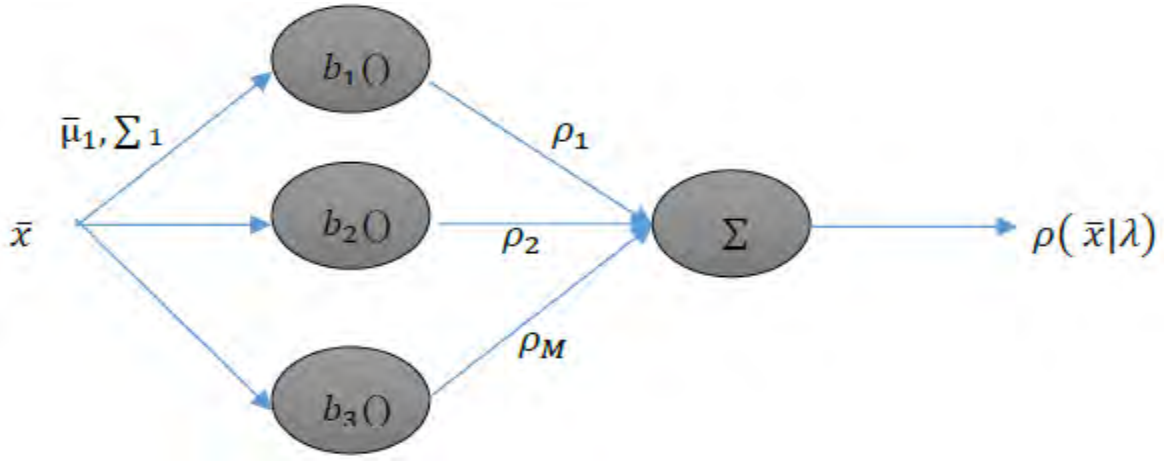

Figure 3- 3:- Diagram of Gaussian Mixture Model

$$b_i(\bar{x}) = \frac{1}{(2\pi)^{D/2}|\sum_i|^{1/2}} \exp\{-\frac{1}{2}(\bar{x} - \bar{\mu}_i)' \sum_i^{-1} (\bar{x} - \bar{\mu}_i)\} \quad \text{.....3-4}$$

Each component density $b_i(\bar{x})$ denotes a D-dimensional normal distribution with mean vector $\bar{\mu}_i$ and covariance matrix $\sum_i$. The mixture weights satisfy the condition $\sum_{i=1}^{M} \rho_i = 1$ and therefore represent positive scalar values. These parameters can be collectively represented $\lambda = \{ \rho_i, \bar{\mu}_i, \sum_i \}$ for $i = 1, \dots, M$. Each language in a language identification system can be represented by one distinct GMM and is referred by the language models λ_i, for $i = 1, 2, 3 \dots N$, where N is the number of languages.

3.5.1.1. Training the Model

In the training phase, a multivariate GMM for the spectral or cepstral feature vectors is created for each language. In the recognition phase, the likelihood of the test utterance feature vectors is computed given each of the training models. The language of the model having the maximum likelihood is anticipated as the language of the utterance.

The EM Algorithm is an iterative optimization of the means, variances and mixture weights of the M basis distributions of a Gaussian mixture model. The aim is to optimize the likelihood that the given data points are generated by the mixture of Gaussians. The EM algorithm alternates between performing an expectation (E) step, and a maximization (M) step [2].

- E - Computes an expectation of the likelihood by including the latent variables as if they were observed variables.
- M - Estimates the parameters by maximizing the expected likelihood found in the E step.

This technique is commonly referred to as the Expectation Maximization (EM) algorithm. The main idea of EM is to estimate the densities by taking an expectation of the logarithm of the joint density between the known and the unknown components, and then maximize this function by updating the parameters that are used in the probability density function. In order to find the updated parameters (i.e., means, variances and mixture weights) that give a good representation of the true distribution, the parameters must be updated iteratively using the EM algorithm until the expected likelihood converges to a stable value, indicating that an optimum has been reached [2].

An iterative approach is followed for computing the GMM model parameters using Expectation-Maximization (EM) algorithm [1]. The aim of training is to obtain the mean, variance, and weighting of each Gaussian distribution (λ).

Steps for training:

1. Begin with an initial model λ then calculate the new mean, variance weighting for the new model $\bar{\lambda}$.
2. Check if the newly calculated parameters are more suitable to model the language by using the following formula.

$$\rho(i|\bar{x}_t, \lambda) = \frac{\rho_i b_i(\bar{x}_t)}{\sum_{k=1}^{M} \rho_k b_k(\bar{x}_i)} \quad \text{.....3-5}$$

3. If the $\rho(X|\bar{\lambda})$ is larger than the $\rho(X|\lambda)$, then the new model $\bar{\lambda}$ is used to do the training again. i.e.

$$\rho\left(X\middle|\bar{\lambda}\right) \geq \rho(X|\lambda) \text{……...….3-6}$$

4. Continue to do the training by repeating step (2) and step (3).

Where;

λ_i is model for $i = 1, 2, 3 \ldots N$ and N is the number of languages.

$\bar{x}$ is a D dimensional observed data,

$b_i(\bar{x})$ is the component density ,

ρ_i is the mixture weight for $i = 1, \ldots, M$ and M is the number of component densities,

$\rho(x|\lambda)$ is the conditional probability and vector $X = \{x_1, x_2, \ldots, x_t\}$

When procedure is repeated to train the new model $\bar{\lambda}$, the new parameters are more close to the actual parameter for modeling the language. The error between the actual parameter for the model and λ become smaller and smaller through training. This procedure is repeated until the error is reached to certain threshold [1].

CHAPTER-FOUR

4. RESULT AND DISCUSSION

4.1.System Description

To train and test the LID model an auditory tool box were used. Appendix B shows Matlab code that is used for the LID system in addition to the auditory tool box [30]. The mfcctrain.m is the feature extraction module and has built in training GMM module. The mfcctest.m is the test module and has feature extraction module inside. Once the test data are feature extracted the classification is done using newgmmtest.m module.

In [31] the iteration of the EM- algorithm was tested in 500, 250, 100, 50 and 10. The research shows that the EM algorithm converges quickly to a good state due to its local search nature even in 10 iterations. Due to this in this research an iteration of 100 was chosen for the EM algorithm. At every iteration the estimated parameters provide an increase in the likelihood function until a local maxima is achieved, at which point the likelihood function cannot increase (but will not decrease) and this is the advantage of the EM algorithm even if the iteration exceeds the maximum iteration since the likelihood function stay on the last stage without any change the algorithm only shows warning rather than creating errors.

The GMM mixture order is dependent with the processing power of the device used for the training and researches at [7] shows with the increase number of GMM mixture order the classification accuracy increases. In this research with respected to the device used for the training 16 GMM mixture order is used. The decision length to train the system is about 56 min for the four languages.

The decision time of the LID systems for all conditions and tests were so fast. For any test either for 1min or (3-5) tests it gives decision with almost less than a second.

After all process the GMM creates a unique model for every language. This uniqueness can be seen clearly from the sample Figure 4-1 and Figure 4-2 of the GMM of Amharic and Guragegna respectively.

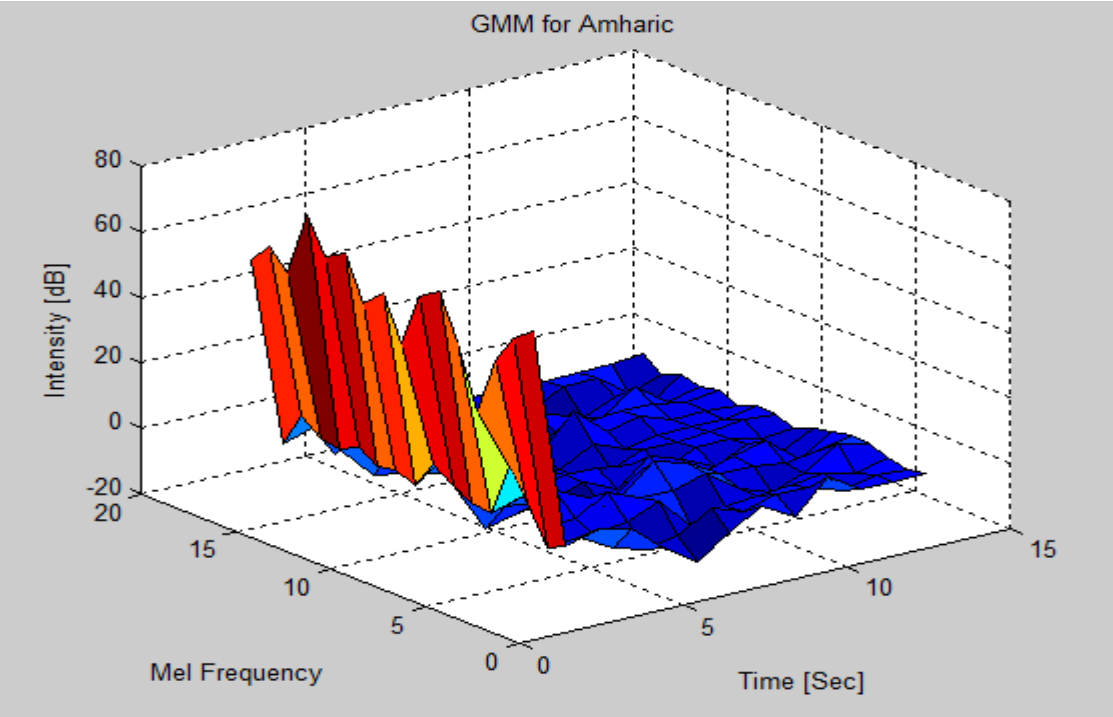

Figure 4- 1:- GMM for Amharic

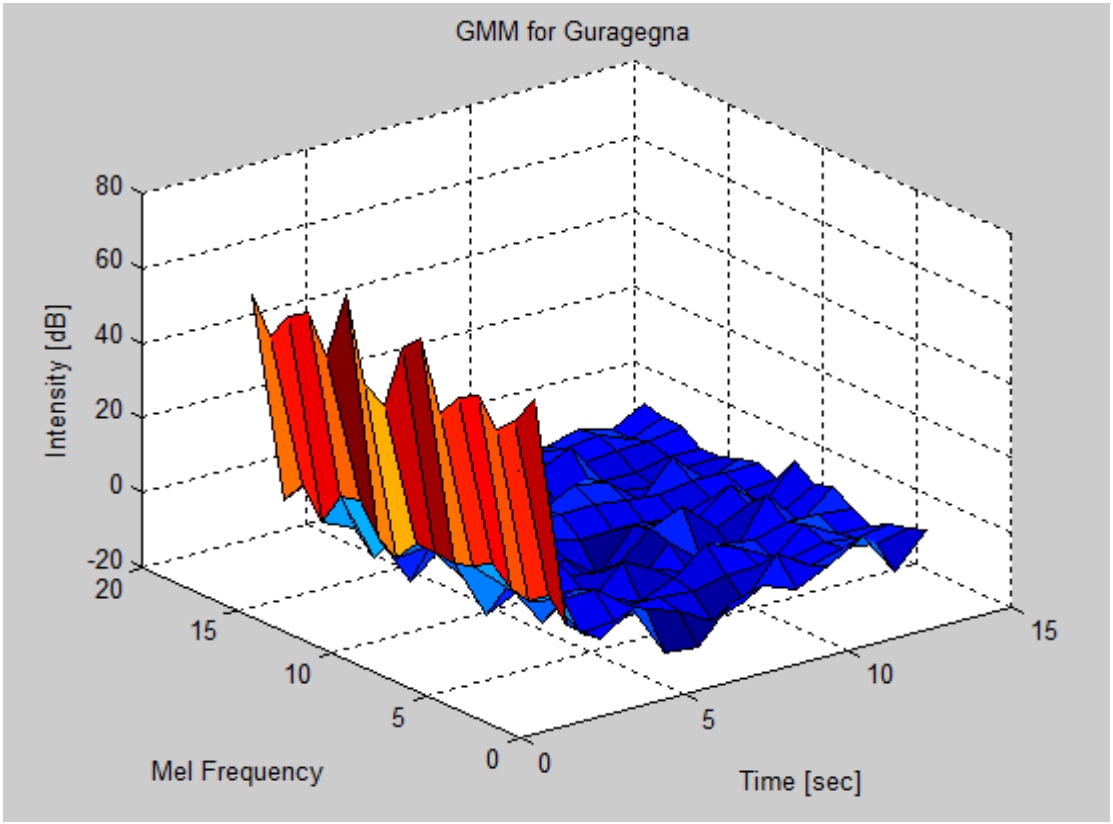

Figure 4- 2:- GMM for Guragegna

4.2.Experimental Setup

The system has been implemented in Matlab 7.12.0.635 (R2011a) on windows 7 ultimate Intel® Xeon® platform. The database mono channel was prepared with the help of Adobe Audition 3.0 and preprocessing of the utterance were done using wavesurfer 1.8.8p4. To train each model, 280 minutes of language training speech is used. From each seven speakers of the language 5 utterances were used for testing of the performance of the model.

The LID system accuracy is tested for an utterance dependent, an utterance independent and speaker independent systems.

Here, System Accuracy (%) and Accuracy (%) is defined:

$$Accuracy(\%) = \left(\frac{correct}{total}\right) * 100 \text{.....4-1}$$

$$System\ Accuracy\ (\%) = \frac{Accuracy(\%)\ L1 + Accuracy(\%)\ L2 + \cdots + Accuracy(\%) Ln}{n} \text{...4-2}$$

Where, $correct$ =Number of samples correctly classified,

$total$ =Total number of samples given for testing,

$Accuracy(\%) L1, \ldots, Ln =$ $Accuracy$ of individual language and

$n =$ Number of languages

It is clear that it is difficult to implement and get a better LID system performance with an utterance independent system with such a small recorded database. But even under such condition, the test has shown an excellent result for utterance dependent LID and a promising result for the utterance independent LID system. The experimental results of each test for the LID performance for two, three and four languages task are discussed in the following sections.

4.3.LID Performance for Two Languages Task Using GMM

The experiments are carried out for varying the combination of languages. The experimental results for Amharic and Guragegna language, Amharic and Oromiffa language, Amharic and Tigregna language, Guragegna and Oromiffa language, Guragegna and Tigregna language and Oromiffa and Tigregna language task are shown in Table 4-1, Table 4-2. Table 4-3, Table 4-4, Table 4-5, and Table 4-6 respectively, and also presented in Fig 4-3, Fig 4-4. Fig 4-5, Fig 4-6, Fig 4-7, and Fig 4.8.

Table 4- 1:- Test result for utterance dependent/ independent of Amharic and Guragegna language

	Utterance dependent		Utterance independent	
Test Utterance	Amharic (1min Long)	Guragegna (1min Long)	Amharic ((3-5)sec Long)	Guragegna ((3-5)sec Long)
Accuracy (%)	91.43	94.29	91.43	65.71

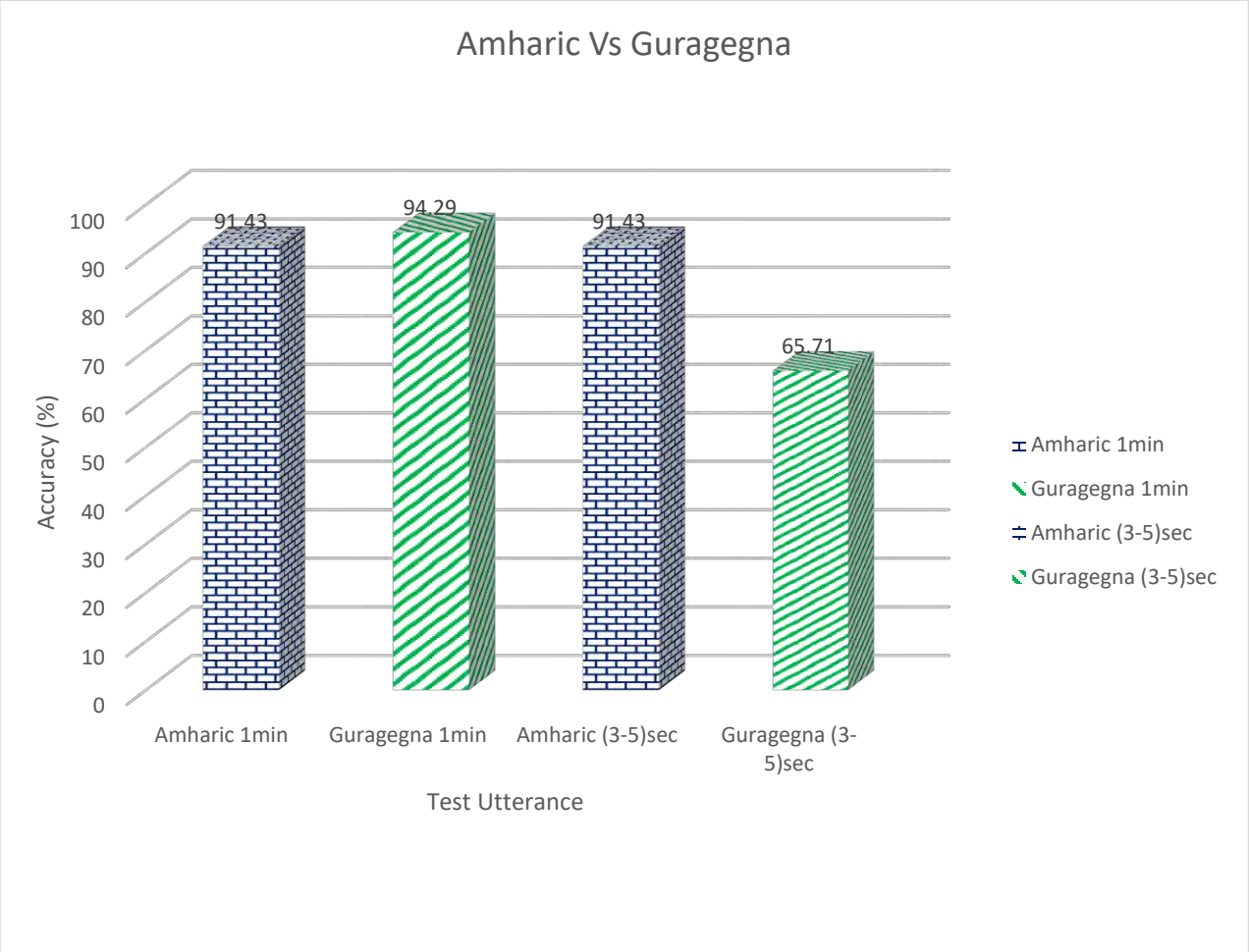

Figure 4- 3:- Test result for utterance dependent/ independent of Amharic and Guragegna language

Table 4- 2:- Test result for utterance dependent/ independent of Amharic and Oromiffa language

	Utterance dependent		Utterance independent	
Test Utterance	Amharic (1min Long)	Oromiffa (1min Long)	Amharic ((3-5)sec Long)	Oromiffa ((3-5)sec Long)
Accuracy (%)	100	100	91.43	77.14

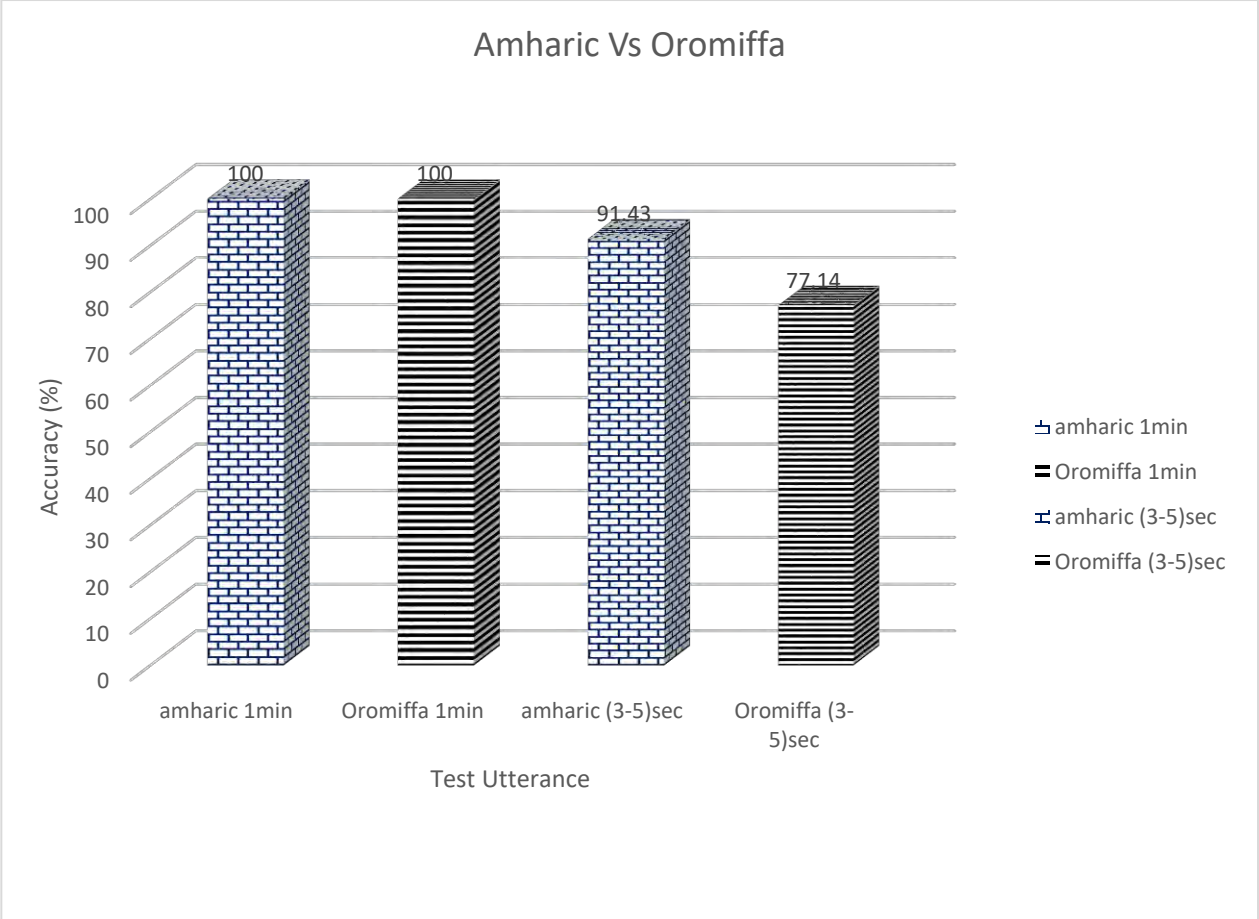

Figure 4- 4:- Test result for utterance dependent/ independent of Amharic and Oromiffa language

Table 4- 3:- Test result for utterance dependent/ independent of Amharic and Tigregna language

	Utterance dependent		Utterance independent	
Test Utterance	Amharic (1min Long)	Tigregna (1min Long)	Amharic ((3-5)sec Long)	Tigregna ((3-5)sec Long)
Accuracy (%)	100	100	97.14	80

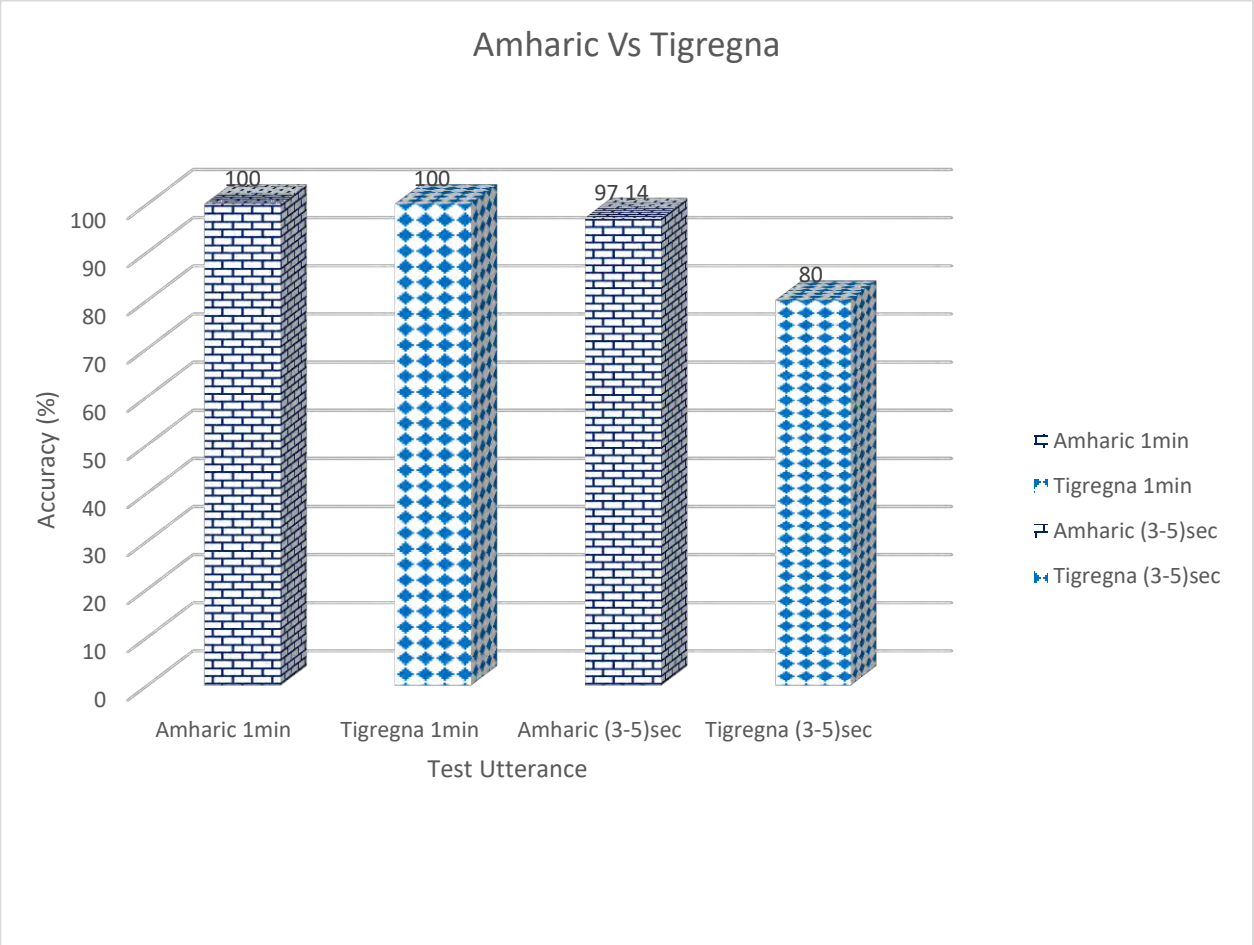

Figure 4- 5:- Test result for utterance dependent/ independent of Amharic and Tigregna language

Table 4- 4:- Test result for utterance dependent/ independent of Guragegna and Oromiffa language

	Utterance dependent		Utterance independent	
Test Utterance	Guragegna (1min Long)	Oromiffa (1min Long)	Guragegna ((3-5)sec Long)	Oromiffa ((3-5)sec Long)
Accuracy (%)	100	100	100	80

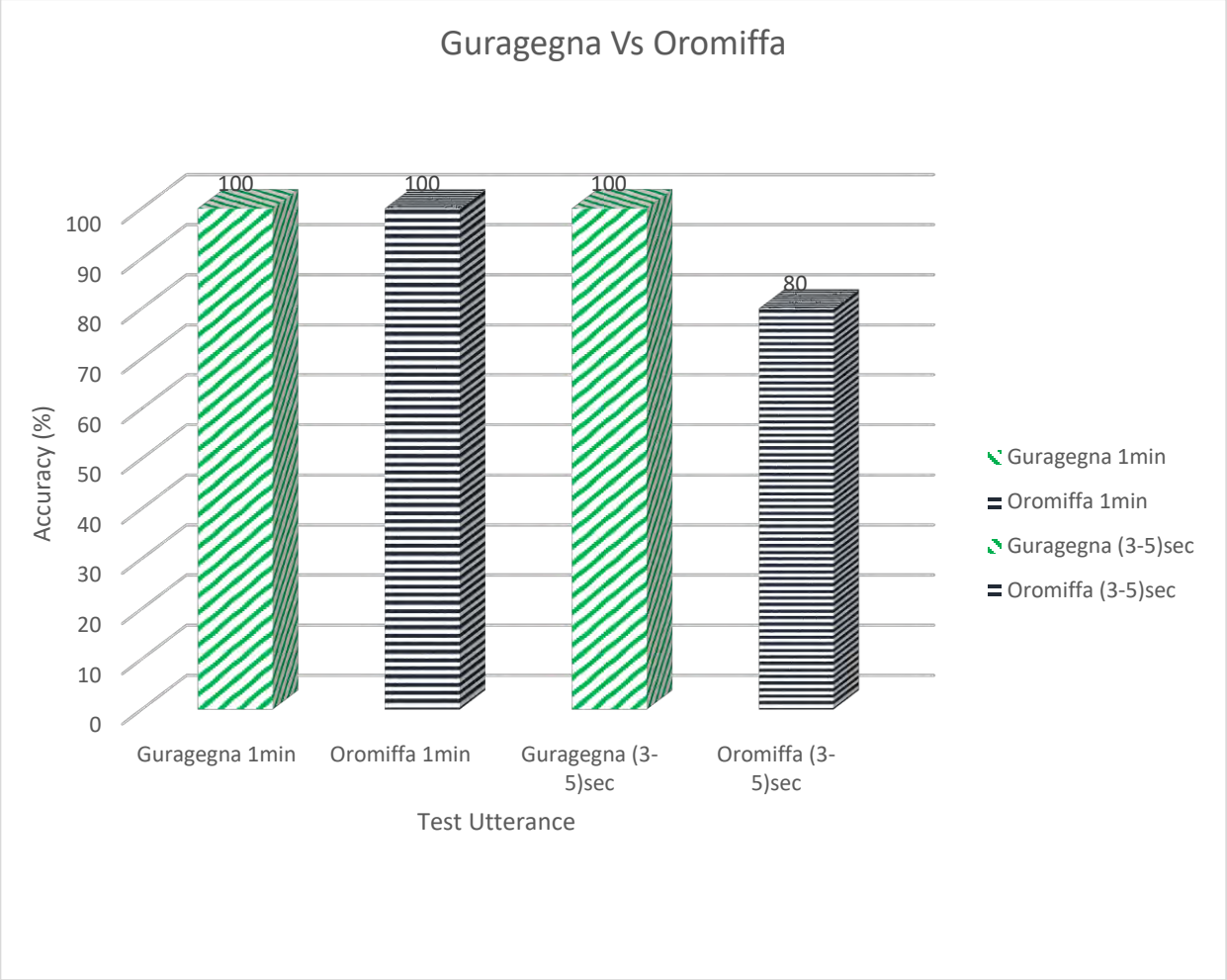

Figure 4- 6:- Test result for utterance dependent/ independent of Guragegna and Oromiffa language

Table 4- 5:- Test result for utterance dependent/ independent of Guragegna and Tigregna language

	Utterance dependent		Utterance independent	
Test Utterance	Guragegna (1min Long)	Tigregna (1min Long)	Guragegna ((3-5)sec Long)	Tigregna ((3-5)sec Long)
Accuracy (%)	100	100	100	88.57

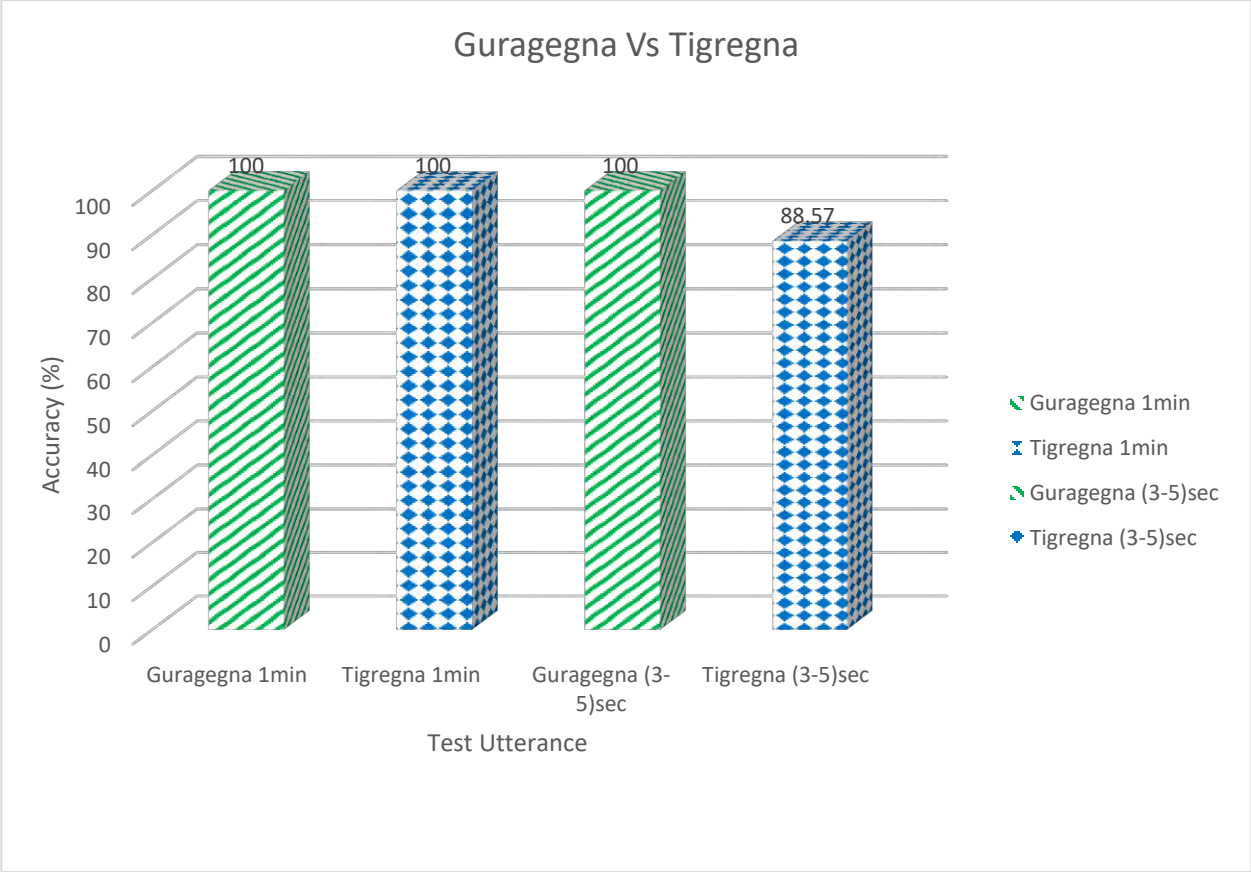

Figure 4- 7:- Test result for utterance dependent/ independent of Guragegna and Oromiffa language

Table 4- 6:- Test result for utterance dependent/ independent of Oromiffa and Tigregna language

	Utterance dependent		Utterance independent	
Test Utterance	Oromiffa (1min Long)	Tigregna (1min Long)	Oromiffa ((3-5)sec Long)	Tigregna ((3-5)sec Long)
Accuracy (%)	94.29	97.14	62.85	88.57

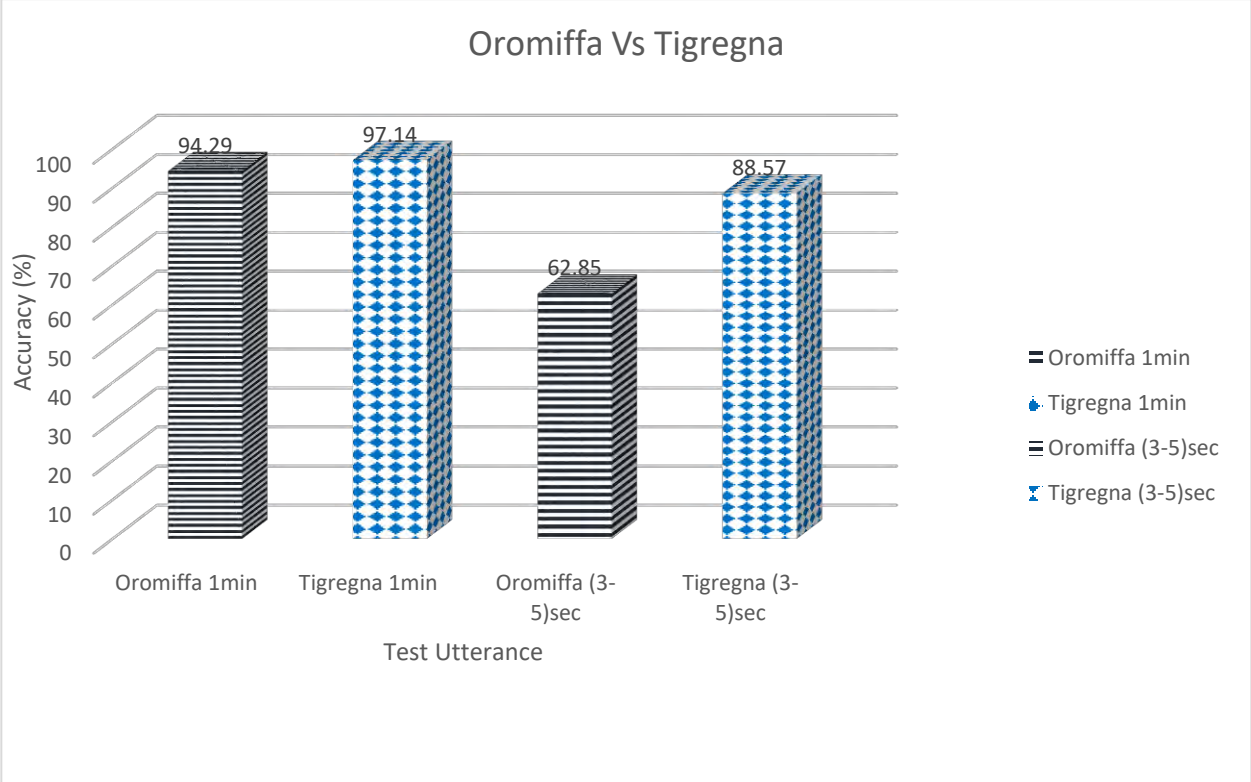

Figure 4- 8:- Test result for utterance dependent/ independent of Oromiffa and Tigregna language

Result analysis and summary of LID system for two languages task

Table 4- 7:- Summary of the performance of the LID system for two languages task

Test Languages	System Accuracy (%)	
	Utterance Dependent	Utterance Independent
Amharic/ Guragegna	92.86	78.57
Amharic/ Oromiffa	100	84.28
Amharic/ Tigregna	100	88.57
Guragegna/Oromiffa	100	90
Guragegna/Tigregna	100	94.28
Oromiffa/Tigregna	95.71	75.71
LID Accuracy for two languages task	**98.095**	**85.235**

The above Table 4-7 summarizes the LID system accuracy for two languages task of both an utterance dependent and independent system. The recognition performance of the LID system is mainly dependent on the feature extraction performance. To acquire relevant features of the speech database the raw data should be as clean as possible (i.e. free of any background noise). For the utterance dependent system, we can see that the accuracy for taking Amharic/Guragegna and Oromiffa/Tigregna is lower compared to the other similar tests. The reason for this decrement is that the unwanted background noises of the recorded dataset used for training and testing; i.e. if the recording is done in an isolated room/free of background noise/ the accuracy will be higher.

As we can see from the Table 4-7 the accuracy of an utterance independent LID system goes lowest up to 75% for Oromiffa/Tigregna and highest up to 94% for Guragegna/Tigregna. Here also the quality or cleanness of the dataset that is used for training and testing plays an important role in the performance of the system. The other main reason for the decrement in performance of the utterance independent system is the unfitness of the training dataset in representing accurately the language. For the training data to be rich the database should be bigger. Since the dataset used for training is comparatively small and getting this performance at this level is acceptable.

4.4.LID Performance for Three Languages Task Using GMM

The experiments are carried out for varying the combination of languages. The experimental results for Amharic/Guragegna/Oromiffa language, Amharic/Guragegna/Tigregna language, Guragegna / Oromiffa /Tigregna language and Amharic/ Oromiffa /Tigregna task are shown in Table 4-8, Table 4-9. Table 4-10 and Table 4-11 respectively and also presented in Fig 4-9, Fig 4-10. Fig 4-11 and Fig 4-12.

Table 4- 8:- Test result for utterance dependent/ independent of Amharic/Guragegna/Oromiffa language.

	Utterance dependent			Utterance independent		
Test Utterance	Amharic (1min Long)	Guragegna (1min Long)	Oromiffa (1min Long)	Amharic ((3-5)sec Long)	Guragegna ((3-5)sec Long)	Oromiffa ((3-5)sec Long)
Accuracy (%)	94.29	94.29	100.00	88.57	62.86	62.86

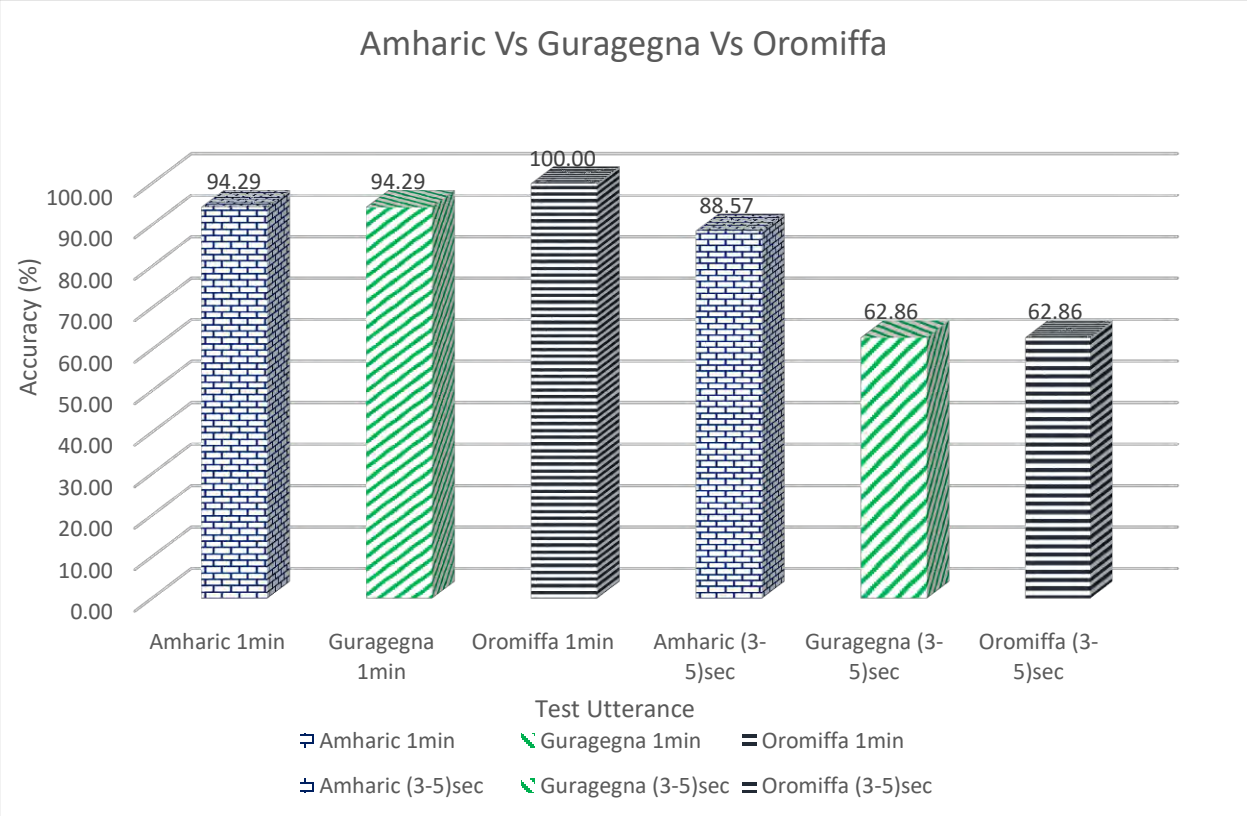

Figure 4- 9:- Test result for utterance dependent/ independent of Amharic/Guragegna/Oromiffa language.

Table 4- 9:- Test result for utterance dependent/ independent of Amharic/Guragegna/Tigregna language.

	Utterance dependent			Utterance independent		
Test Utterance	Amharic (1min Long)	Guragegna (1min Long)	Tigregna (1min Long)	Amharic ((3-5)sec Long)	Guragegna ((3-5)sec Long)	Tigregna ((3-5)sec Long)
Accuracy (%)	100.00	100.00	100.00	97.14	62.86	74.29

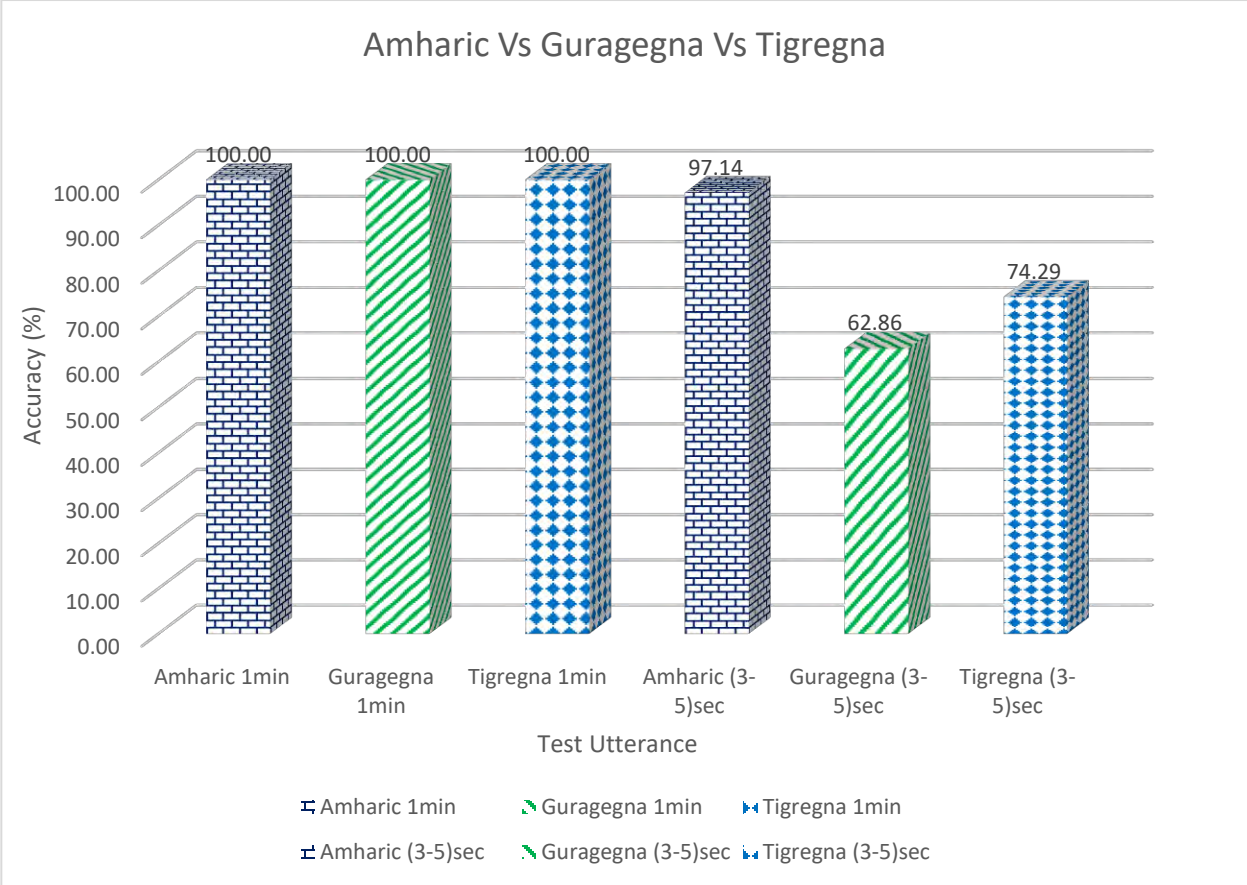

Figure 4- 10:- Test result for utterance dependent/ independent of Amharic/Guragegna/Tigregna language.

Table 4- 10:- Test result for utterance dependent/ independent of Guragegna /Oromiffa/Tigregna language.

	Utterance dependent			Utterance independent		
Test Utterance	Guragegna (1min Long)	Oromiffa (1min Long)	Tigregna (1min Long)	Guragegna ((3-5)sec Long)	Oromiffa ((3-5)sec Long)	Tigregna ((3-5)sec Long)
Accuracy (%)	100.00	97.14	80.00	100.00	62.86	80.00

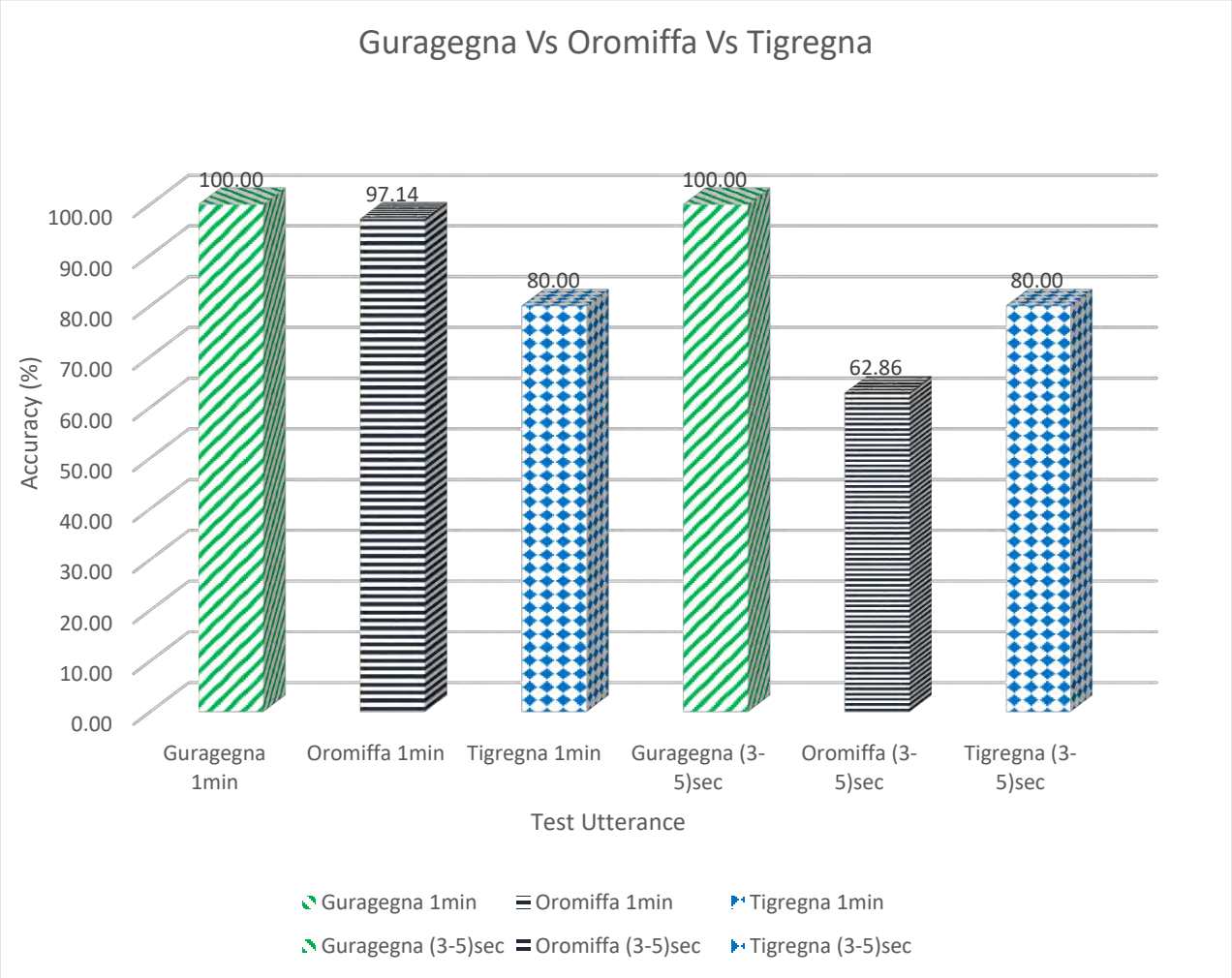

Figure 4- 11:- Test result for utterance dependent/ independent of Guragegna /Oromiffa/Tigregna language.

Table 4- 11:- Test result for utterance dependent/ independent of Amharic /Oromiffa/Tigregna language.

	Utterance dependent			Utterance independent		
Test Utterance	Amharic (1min Long)	Oromiffa (1min Long)	Tigregna (1min Long)	Amharic ((3-5)sec Long)	Oromiffa ((3-5)sec Long)	Tigregna ((3-5)sec Long)
Accuracy (%)	100.00	94.29	97.14	97.14	60.00	68.57

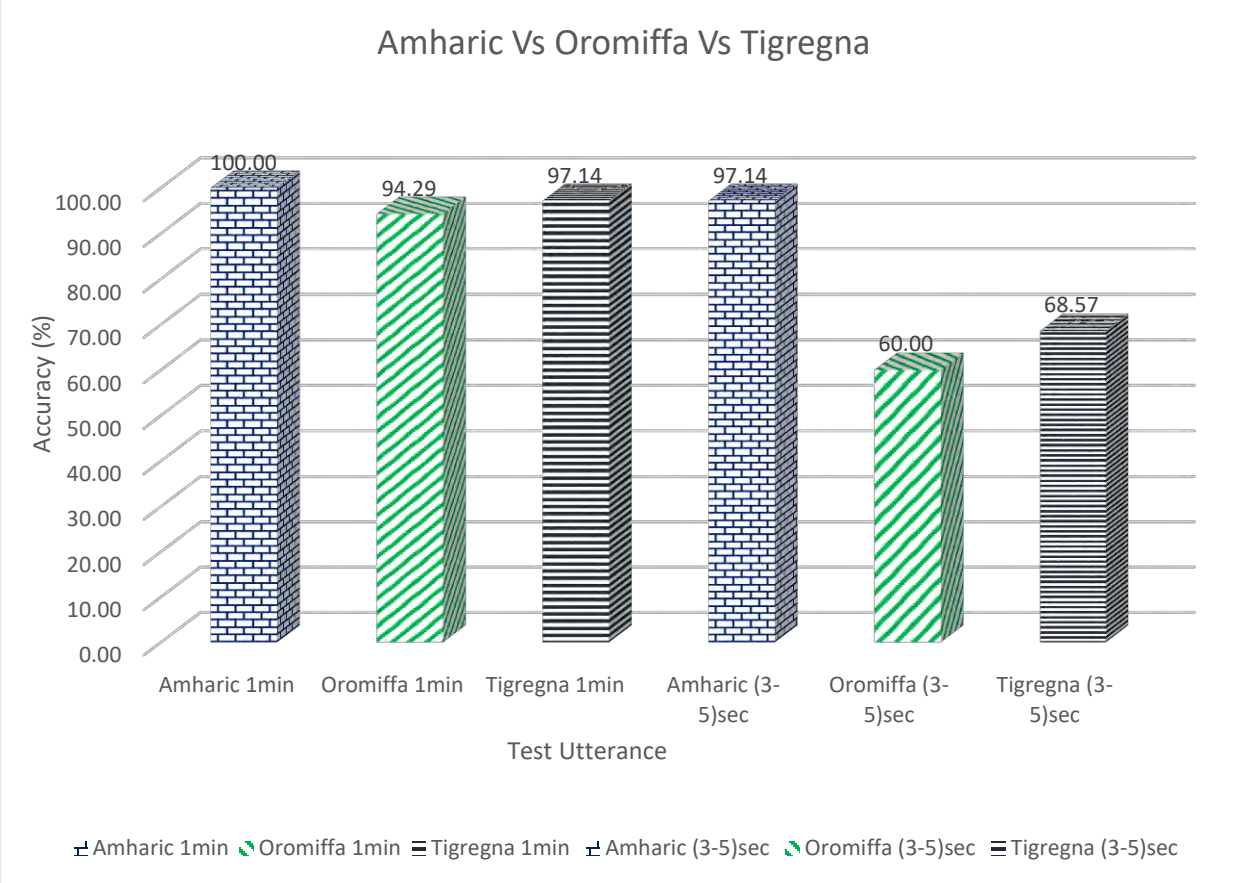

Figure 4- 12:- Test result for utterance dependent/ independent of Amharic /Oromiffa/Tigregna language.

Result analysis and summary of LID system for two languages task

Table 4- 12:- Summary of the performance of the LID system for three languages task

Test Languages	System Accuracy (%)	
	Utterance Dependent	Utterance Independent
Amharic/Guragegna/Oromiffa	96.2	71.43
Amharic/ Guragegna/Tigregna	100	78.09
Guragegna/Oromiffa/Tigregna	92.38	80.95
Amharic/ Oromiffa/Tigregna	97.14	75.4
LID Accuracy for three languages task	**96.43**	**76.47**

The above Table 4-12 summarizes the LID system accuracy for three languages task of both utterance dependent and independent system. Here also the accuracy of both systems is affected by the cleanness and the size of the dataset used for training and testing. It can be observed that for both tests the accuracy of the three languages task is lower compared to the system accuracy of the two languages task. This is due to for two languages task in the training phase the system tries to create two distinct models of each language and in the testing phase the classifier chooses the most likelihood by calculating the probability density from the two models. But increasing the language to three will increase the models by one and this has a direct impact on the calculation of probability density function. It is clear that calculating probability for two sample space is higher than calculating probability of three sample spaces.

4.5.LID Performance for Four Languages Task Using GMM

The next experiment is the LID system accuracy by taking all the four languages i.e. Amharic, Guragegna, Oromiffa and Tigregna. The testing is done for both utterance dependent and utterance independent. The result is shown in Table 4-13 and Table 4-14 respectively and also presented in Fig 4-13 and Fig 4-14.

Table 4- 13:- Utterance dependent LID System Accuracy taking four languages at a time

	Utterance dependent			
Test Utterance	Amharic (1min Long)	Guragegna (1min Long)	Oromiffa (1min Long)	Tigregna (1min Long)
Accuracy (%)	91.43	97.14	97.14	85.71

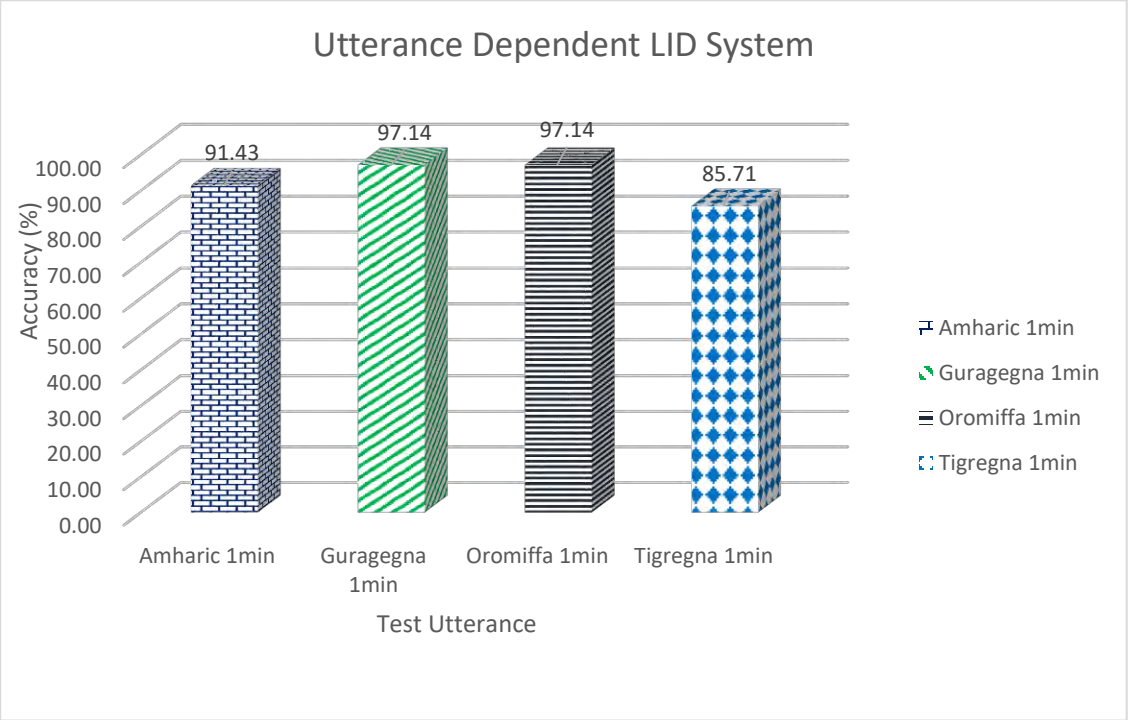

Figure 4- 13:- Utterance dependent LID system accuracy taking four languages at a time

The above Figure 4-13 shows the LID system accuracy for the four languages task of an utterance dependent system. Here also the accuracy of the system is mainly affected by the cleanness of the dataset used for training and testing. The system accuracy will be better if the recording is done in an isolated room/free of background noise/.

Table 4- 14:- Utterance independent LID system accuracy taking four languages at a time

	Utterance independent			
Test Utterance	Amharic ((3-5)sec Long)	Guragegna ((3-5)sec Long)	Oromiffa ((3-5)sec Long)	Tigregna ((3-5)sec Long)
Accuracy (%)	91.43	62.86	60.00	65.71

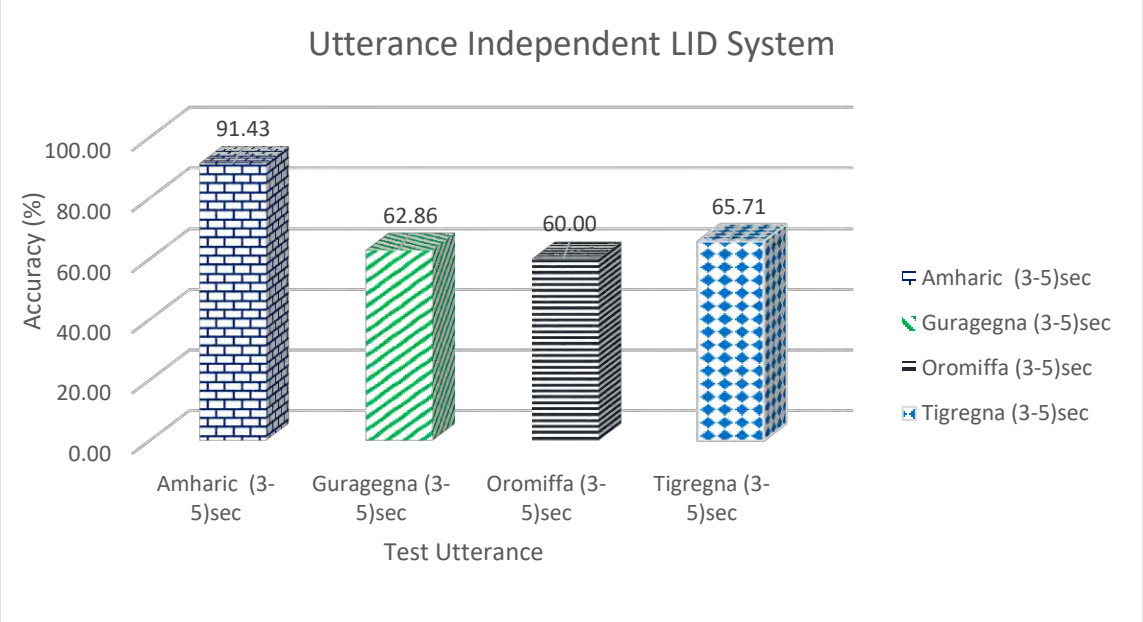

Figure 4- 14:- Utterance independent LID system accuracy taking four languages at a time

The above Figure 4-14 shows the LID system accuracy of the four languages task of an utterance independent system. It is observed that the identification accuracy for the Amharic is higher than the other tests. The possible reason for this is that the spectral pattern match of the training and the testing data is higher for Amharic language then others. To have a better accuracy for the whole system the training dataset should be rich in spectral distribution in representing the respective language accurately. This is done by increasing the size of the dataset used for training the system.

Table 4- 15:- Summary of the performance of the LID system for three languages task

Test Languages	Accuracy (%)	
	Utterance Dependent	Utterance Independent
Amharic/Guragegna/Oromiffa/Tigregna	92.85	70

4.6.Summary of the LID System Accuracy

As we can see from the result below the LID system for both utterance dependent and utterance Independent tasks the accuracy decreases when the number of languages increases. This is due to for a multivariate GMM system, increasing the number of language has a direct impact on the classification performance of the system (i.e. in the testing phase the classifier chooses the most likelihood by calculating the probability density from the models). It is clear that calculating probability for two sample space is higher than calculating the probability of three sample spaces and the same is true for calculating the probability for three and four sample spaces.

Table 4- 16:- Summary of the performance of the LID system for increasing number of Languages

Test Languages	Accuracy (%)	
	Utterance Dependent	Utterance Independent
LID by taking Two languages task	98.10	85.24
LID by taking Three languages task	96.43	76.47
LID by taking Four languages task	92.85	70.00

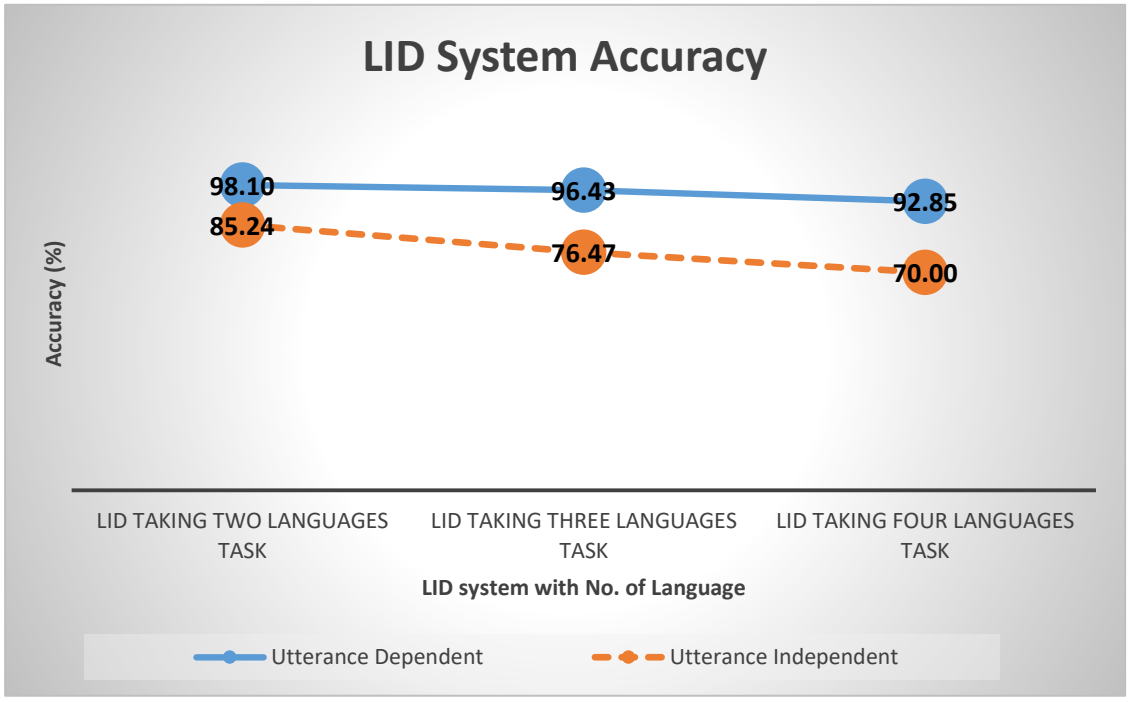

Figure 4- 15:- Summary of the performance of the LID system for increasing number of Languages

4.7.Speaker Independent LID System

The speaker independent system is a system where the speech patterns are constructed (or adapted) to a multiple speakers. The system is tested by creating biometrical disjoint sets in the training and testing dataset (i.e. out of 7 speakers of each language, 4 speakers utterances is used to train the system and the other 3 speakers utterances is used to test the accuracy of the system).

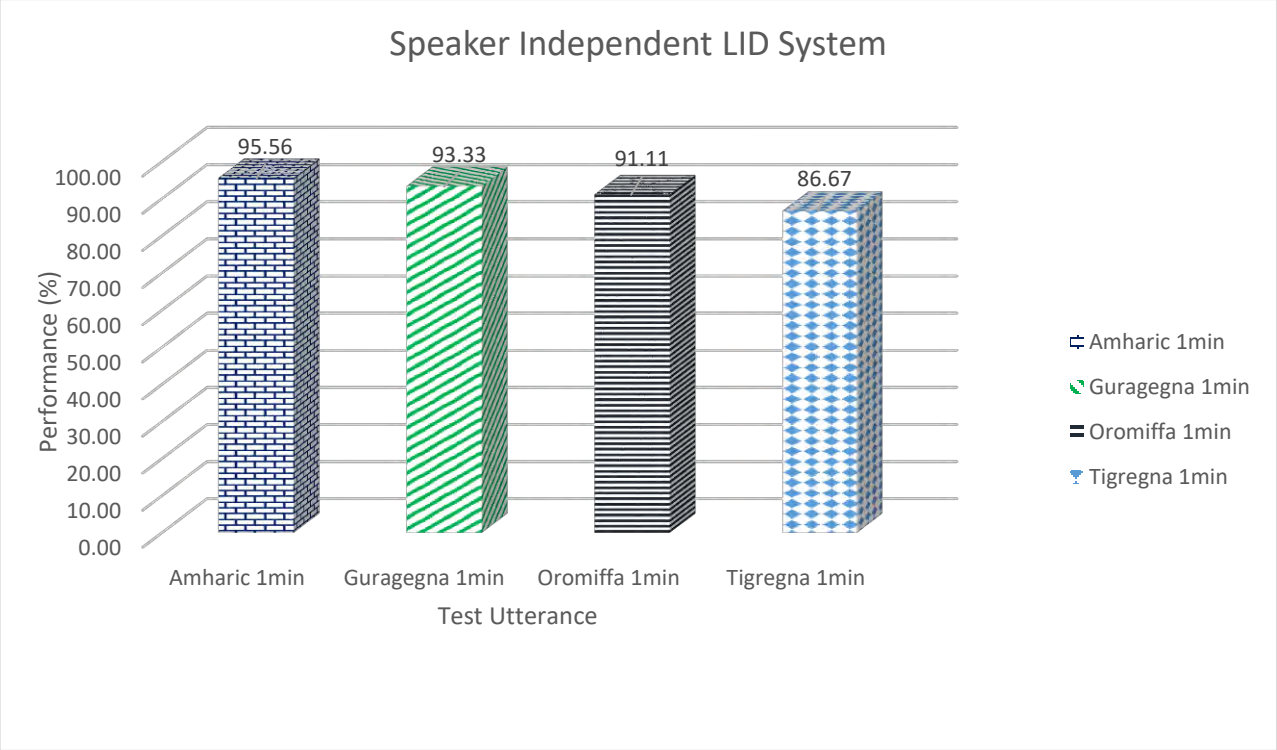

Figure 4- 16:- Speaker Independent LID system accuracy taking four languages task for utterance dependent only

As we can see from the Figure 4-16 the speaker independent system performs lower form 86.67% up to higher 95%. The average accuracy of the speaker independent LID system is 91.67%. Here also the accuracy of the systems is affected by the cleanness of the dataset used for training and testing. The number of subjects/speakers used for training the model has also a role in the decrement of the accuracy. In this system the speaker used for training and testing is different. This makes the system to have a lower accuracy than an utterance speaker dependent system.

CHAPTER-FIVE

5. CONCLUSION AND RECOMMENDATION

5.1.Conclusion

In this thesis, the aim of this work is to develop and test language identification systems for four Ethiopian languages namely Amharic, Guragegna, Oromiffa and Tigregna. To develop the LID system properly first a dataset was prepared by recoding 7 speakers of each language. After preparation of the database pre-processing is done using software called waveserfer.

Once the recording is pre-processed the extraction and representation of language specific features of the LID task was done. To do this, MFCC is used to extract feature vectors from the speech signal. Every MFCC feature vector has been transformed into a feature vector using GMMs. Since training and processing of languages with any model need a higher computer processing power a smaller GMM mixture order (i.e.16 mixture order) is used for classification.

To train and create the model for each language it almost takes a total of 56min and this processing time depends on the hardware resource that is used for the experiment. If the processing power of the device is higher, then the time it takes to train the system will be lower.

The paper encloses three ways of experimentation for testing of the LID system accuracy, namely utterance/speech dependent, utterance/speech independent and speaker independent system. It is more challenging to implement and get a better LID system performance with an utterance independent system with such a small recorded database. But even under such condition, the test has shown an excellent result from the utterance dependent and speaker independent LID systems and a promising result of the utterance independent system. The decision time of the LID systems for all conditions and tests were so fast. For any test either for 1min or (3-5) tests it gives decision with almost less than a second.

To test the performance of the LID system, experimental scenarios are designed and carried out by taking two, three and four languages task at a time.

The LID system accuracy by taking two languages task for the utterance dependent LID system was excellent and it was about 98% accurate on average. For the utterance independent system

even though the performance was decreasing compared to the utterance dependent system, but it shows a good performance about 85% accurate on average.

The result for taking three languages task for the utterance dependent LID system was also about 96% accurate on average, whereas the LID system accuracy for the utterance independent system was about 76% accurate on average.

The next experiment was done by taking four languages at a time for both the utterance dependent and independent system. The utterance dependent system performs with about 92% accuracy and on the other side the utterance independent system shows a performance of 70% accuracy.

The last experiment was done by taking four languages at a time for speaker independent LID system and the system show performs of 91% accuracy on average.

From all employed scenarios, as it is expected to see the decrement of LID system accuracy with an increase in the number of languages tested in the system. The rate of decrement is acceptable, but we could have decreased the rate of decrement with a higher speech database and better hardware resource.

Finally, we can conclude that this research will add to the researches that have been conducted in the country.

5.2.Recommendation

The following recommendations are forwarded:

- The investigation revealed that the cleanness and size of the dataset used for training has a direct effect on the performance of the LID system. To have an improved LID system performance, it is therefore recommended that the researcher should develop large enough and studio quality speech corpus.
- In Ethiopia there is no common, public domain corpus of speech in different languages, which may be used as a benchmark in comparatively evaluating system performance. So it is advised to develop a common speech corpus for local languages that will contribute to the researches that has been conducted in the area of natural language processing.

5.3.Future Work

The following future works can be performed for further analysis and study:

- As the GMM LID system provides an efficient means to identify spoken languages automatically, it is worth the effort to develop techniques to further improve the accuracy of the system since most LID applications require faster and accurate language identification system.
- The system can be tested with a higher hardware resource with different GMM mixture order to implement the accuracy with increase in number of mixture order.
- The research can be expendable by increasing the number of local languages in addition to the languages used for the research.
- The dataset can be used with other feature extraction and classification techniques to compare the result with the implemented research.
- The LID system can also be tested using other classification algorithms and its performance can be compared with this research work.

REFERENCES

[1] A. Nagesh, "Automatic Text Independent Language Identification," *International Journal of Emerging Technology & Research,* April 2013.

[2] Calvin Nkadimeng, "Language Identification Using Gaussian Mixture Models," *Stellenbosch: University of Stellenbosch,* March 2010.

[3] Wikipedia, "List of countries by number of mobile phones in use," [Online]. Available: https://en.wikipedia.org/wiki/List_of_countries_by_number_of_mobile_phones_in_use. [Accessed 30- 01- 2017].

[4] Kim-Yung, "Automatic Spoken Language Identification Utilizing Acoustic and Phonetic Speech Information," 2004.

[5] Bhanu Prasad and S.R. Mahadeva Prasanna (Eds.), Speech, Audio, Image and Biomedical Signal Processing using Neural Networks, vol. 83, 2008.

[6] Manas A. Pathak and Bhiksha Raj, "Privacy-Preserving Speaker Verification and Identification Using Gaussian Mixture Models," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 21(2), February 2013.

[7] Pinki Roy, Pradip K. Das, "Language Identification of Indian Languages Based on Gaussian," *International Journal of Wisdom based Computing,* vol. 1(3), pp. 54-59, December 2011.

[8] Shashidhar G. Koolagudi, Deepika Rastogi and K. Sreenivasa Rao, "Identification of Language using Mel-Freqency Coefficients," *Elsevier Ltd,* 2012.

[9] YonghuaXu, Jian Yang, Jiang Chen, "Methods to improve Gaussian Mixture Model for Language Identification," *Proceedings of the 2010 International Conference on Measuring Technology and Mechatronics Automation,IEEE Computer Society,* 2010.

[10] Pedro A. Torres-Carrasquillo , Douglas A. Reynolds and J.R. Deller, "Language Identification using Gaussian Mixture Model Tokenization," *IEEE Conference on ICASSP,* May 2002.

[11] Ann Thyme-Gobbel and Sandra E. Hutchins, "Prosodic Features in Automatic Language Identification Reflect Language Typology," *International Congress of Phonetic Sciences,* August 1999.

[12] Bo Yin, Eliathamby Ambikairajah and Fang Chen, "Combining Cepstral and Prosodic Features in Language Identification," *The 18th International Conference on Pattern Recognition,* 2006.

[13] David Martinez, Lukas Burget, Luciana Ferrer and Nicolas Scheffer, "Ivector-Based Prosodic System for Language Identification," *IEEE Conference on ICASSP,* 2012.

[14] L.F. Lamel and J.L. Gauvain, "Language Identification Using Phone-based Acoustic Likelihoods," *ICASSP-94,* 1994.

[15] J.L. Gauvain, A. Messaoudi, and H. Schwenk, "Language Recognition Using Phone Lattices," *Proc. International Conferences on Spoken Language Processing,* pp. 1283-1286, 2004.

[16] V. Ramasubramanian, A. K. V. Sai Jayram and T. V. Sreenivas, "Language Identification Using Parallel Phone Recognition," *International speech communication association,* 2004.

[17] Koena R. Mabokela, Madimetja J. D. Manamela and Nalson Gasela, "Automatic Language Identification Using Word Segments on Mixed-Language Speech," *Telkom Centre of Excellence for Speech Technology,* 2011.

[18] Ngoc Thang Vu, Dau-Cheng Lyu etal, "A First Speech Recognition System For Mandarin-English Code-Switch".

[19] Indhuja K, Indu M, Sreejith C and P. C. Reghu Raj, "Text Based Language Identification System for Indian Languages Following Devanagiri Script," *International Journal of Engineering Research & Technology,* vol. 3, no. 4, April 2014.

[20] Legesse Wedajo, "Modeling Text Language Identification for Ethiopian," July 2014.

[21] Gerrit Botha, Victor Zimu and Etienne Barnard, "Text-Based Language Identification for South African," *South African Institute of Electrical Engineers,* vol. 98(4), December 2007.

[22] Julian Heuser, "Speech Recognition Wiki," 28- 12- 2014. [Online]. Available: http://recognize-speech.com/preprocessing. [Accessed 15- 04- 2015].

[23] H. Bourlard, H. Hermansky, N. Morgan, "Towards increasing speech recognition error rates," *Speech Communications,* vol. 18, pp. 205-231, 1995.

[24] Pratik K. Kurzekar, Ratnadeep R. Deshmukh, Vishal B. Waghmare, Pukhraj P. Shrishrimal, "A Comparative Study of Feature Extraction," *International Journal of Innovative Research in Science,* vol. 3, no. 12, 2014.

[25] Eslam Mansour mohammed and Mohammed Sharaf Sayed,, "LPC and MFCC Performance Evaluation with Artificial Neural Network for Spoken Language Identification," *International Journal of Signal Processing, Image Processing and Pattern Recognition,* vol. 6, no. 3, 2013.

[26] Poonam Sharma and Anjali Garg, "Feature Extraction and Recognition of Hindi Spoken Words using

Neural Networks," *International Journal of Computer Applications,* vol. 142, no. 7, 2016.

[27] Varsha Singh, Vinay Kumar Jain and Dr. Neeta Tripathi, "A Comparative Study on Feature Extraction Techniques for Language," *International Journal of Engineering Research and General Science,* vol. 2, no. 3, 2014.

[28] James Lyons, "Practical Cryptography," 2012. [Online]. Available: http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/. [Accessed 05- 12- 2014].

[29] Leo Liu, "Acoustic Models for Speech Recognition Using Deep Neural," *Massachusetts Institute of Technology,* 2015.

[30] "https://github.com/lucyd/independent_study/tree/master/GMM," [Online]. [Accessed 26- 11- 2014].

[31] Tommy Stromhaug, "Discriminating Music,Speech and other," *Norwegian University of Science and Technology,* August 2008.

[32] Daniel Kibret's views, "http://www.danielkibret.com/," [Online]. [Accessed 15- 02- 2014].

APPENDICES

Appendix A

This encloses the paragraph and the sentence that were taken for training and testing. The translation of each paragraph and sentence has the same meaning with each individual language.

Amharic paragraph taken for training and testing

“ወገኖቼ እኛ አሁን ጨዋ ማነው? የሚለውን ከመምረጣችን በፊት ጨዋነት ምንድን ነው? በሚለው ላይ መስማማት ነው ያለብን፡ ፡ ጨዋነት መከራከር፣ የተሻለ ሃሳብ ማቅረብ፣መሞከር፣ መፍጠር፣ መንቃት፣መጠየቅ ስራን አክብሮ መስራት፣ የተሻለ የስራ አካባቢ መፍጠር፣ ለምን ብሎ መጠየቅ? ይመለከተኛል ብሎ ማሰብ፣ እኔ የለሁበትም ሳሆን እኔም አለሁበት ብሎ ማመን፣ ሌሎች ሰዎችንም ከተጉበት መቀስቀስ ማለት መሆን አለበት፡፡ ትርጉሙን ስላበላሸነው ልጆቻችን አካባቢያቸውን ከማወቅ ይልቅ አርፈው እንዲቀመጡ፣ ከመሞከር ይልቅ እጃቸውን አጣጥፈው እንዲፈሩ ፣ ከመፍጠን ይልቅ ዘገምተኛ እንዲሆኑ ከመጠየቅ ይልቅ ዝም እንዲሉ፣ ከመከራከር ይልቅ ዝም እንዲሉ ከመከራከር ይልቅ ሁሉን አሜን ብለው እንዲቀበሉ የራሳቸውን ሃሳብ ከማቅረብ ይልቅ የሌላውን ብቻ እንዲጋቱ፣ ሃሳባቸውን ከመግለጥ ይልቅ ሃሳብ አልባ ሆነው በዝምታ እነዲቀመጡ፣ ከመናገር ይልቅ እንዲያልጎመጉሙ፣ አደረግናቸው፡፡ ጨዋነት ግን ይኼ አይደለም፡፡ ይኼ ጥሬ ጨዋነት ነው፡፡” [32]

Guragegna paragraph taken for training and testing

“ሰቦ አኋ ወሄቃሩ ሚኑ ንብኔ ተበሮተንዳ ይፍቴ ወሄቃር ምቃሩ ወበር ነረብንደ? ወሔ ወባጇ ፣ ይፈዝቃር አትባዶት፣ ሰክቶት፣ ቤጥነት፣ ተሳሮት፣ ሜና ወሄቃር ቶቶት፣ የሜና መደር ወሄቃር ሳቦት፣ የምንቃር ባረንታ ተሳሮት፣ ዝህ ዘንጋ እያም እቾትኔ ያንቢ ቃሩ በሮት፣ እያ ኤያገቤ ተባሮት ያገቤ ቃሩ በሮቱ ፣ ያንሰማ ሰብ ሠማምታ ይቾት ኸማ ኣምሮቱ፣ የቃረንዳ ፍቾታ ዲንጋንዳ ቄያሁና ተህሮተሁና ይርቅ ቁሳባሮም፣ ይችኖህማ አመነዉዮም፣ ሸር ተዘንጎት ይርቅ እንምቃር ቄስባሮም ወሔቃሩ ባሮም ይብሮኸም አመነውዮም፣ ታንህሮተሁና የቾነ የገሁና ይዘንጋ ታቦምታ ያስብቃር ዮዶኸማ ቾቾዮም::የመሰረኖቃር ተዘንጎት ይርቅ አትቃር ኤሕሮቃርኸማ ይረብሮ ፣ ተዘንግቶ ይርቅ እመጠጠቦማ ቁሰ ይብሮኸማ አመነቦዬም፡፡ ወሔቃርበሮት ዝኸታ አንኸረ፡፡ ዝህታ ይራ ወሔነቱ፡፡”

Oromiffa paragraph taken for training and testing

”Firoottan koo eenyu mee kan saalfataan? Kana filachuun dura saalfatummaan maal jechuudha? Kan jedhu irrati walii galuutu nurra jira. Saalfatummaan mormuu, yaada fooyya'aa dhiyyeessuu, yaaluu, uumuu, sochaa'uu ykn dammaquun gaaffachuu, hojii kabajanii hojjachuu, bakka hojjii fooyya'aa uumuu, maaliif jedhanii gaaffachuu, na ilaalata jedhanii yaaduu, na hin ilaalatu jedhanii yaadurra na ilaalata jedhanii amanuun fudhachuu. namoota biroos bakka hafanii baanansuu ta'uu qaba jechuu. Hiikaa isaa waan dogongorsiineef ijaoleen keenya naannoo isaanii beekuu irraa callisaanii akka ta'aan, yaaluu irraa harkaa isaanii sassaabanii akka soodaatan, wantoota uumuu irraa dadhaboo akka

ta'an, gaafachuu irraa cal akka jedhan, mormuurra wantaota hundaa amananii akka fudhatan, yaada isaanii dhiyyessuurra kan namoota biroon qofa akka wal ga'aman, yaada isaanii ibsuurra yaadamaleessa ta'aan callisaanii ta'uu, dubbachuurra akka gungummaan isaan taasiseera. Salffatummaan garuu kana miti, kun salfatummaa hin bilchaaneedha."

Tigregna paragraph taken for training and testing

"ወገነይ ሕጅ ጨዋ መን እዩ? ዘብል ቅድሚ ምምራፅና ጭውነት እንታይ እዩ? አብ ዝብል ክንስማእማእ አለና፡፡ጭውነት ምክርካር፣ ዝሓሸ ሓሳብ ምምፃእ፣ ምሙካር፣ ምፍጣር፣ ምንቃሕ፣ ምጥያቅ፣ ስራሕ ላክቢርካ ምስራሕ፣ዝሐሸ ናይስራሕ አካባቢ ምፍጣር፣ ንምንታይ ኢልካ ምጥያቅ? ይምለልከተኒ ኢልካ ምሕሳብ፣ የለኩልን ዘይኽነስ አለኩሉ አልካ ምእማን፣ ከልኦት ካብ ዝደቀሱሉ ምቅስቃስ ማለት ክኽውን አለዎ፡፡ ትርጉሙ ስለ ዘባለሸናዬ ደቅና አካባቢእም ካብ ምፍላጥ ዓሪፍም ክቅመጡ ፣ካብ ምሙካር ኢዶሞ ዓጺፎም ክቅመጡ፣ ካብ ምፍጣን ዝሑላት ክኮኑ፣ ካብ ምጥይቅ ሱቅ ክብሉ፣ ካብ ምክርካር ኹሉ አሜን ኢሎም ክቐበሉ፣ ናይባዕሎም ሓሳብ ካብ ምቐራብ ናይኽሊእ ሱቕ ኢሎም ክቐበሉ ፣ ሓሳቦም ካብ ምግላጽ ሱቅ ምባል፣ ካብ ምንጋር ከዕነምጉሙ ጌርናዩም፡፡ ጨዋነት ግን እዚ ኣይኮነን፡፡ እዚ ጥረ ጨዋነት እዩ፡፡"

Table A- 1:- Sentence taken for testing

Language	Sentence taken
Amharic	"ጤና ይስጥልኝ የተንቀሳቃሽ ስልክ መሰመሬ ስለተዘጋብኝ ልትረዱኝ ትችላላችሁ?"
Guragegna	"አረመምር ነርሁ ስልከና ተዘጋቢማ ትከፍቶኒሸዌ?"
Oromiffa	"Harka fuune sararrii bilbila sochaa'a koo waan na jalaa cufameef na gargaruu dandeessuu?"
Tigregna	"ጥዕና ይሃበላይ ናይ ተንቀሳቓሲ ስልክይ መስመር ስለ ዝተዓፀወኒ ክሕግዙኒ ይኽእሉ ዶ?"

Appendix B

Matlab code that are used for the LID system in addition to the auditory tool box:

```
%%mfcctrain.m
%%Feature extraction module and has built in training GMM module
clear all;
clc;
tic
for m=1:4
 dname='amharic';
 if(m==2)
  dname='guragegna';
 end
 if(m==3)
  dname='oromiffa';
 end
 if(m==4)
  dname='tigrigna';
 end
 x1=0;
 dir2 = ['C:\Users\hp1\Desktop\Desktop
Folders\research\TOOLBOX\Four Language\trainwavdata1\' dname '\'];
 %Read all names in dir
 a=dir(dir2);
 %Open file for writing
 fid = fopen([dname '.mat'],'w','a');
 for j=3:length(a);
  sprintf('%s %2.0f','Processing trainfile :',j-2);
  %filename of training files
  fname=[dir2 a(j).name];
  %read wave
  [y fs]=wavread(fname);
  clear fname;
  sig=y.*y;
  E=mean(sig);
  Threshold=0.05*E;
  k=1;
  for b=1:100:(length(sig)-100)
   if((sum(sig(b:b+100)))/100 > Threshold)
   dest(k:k+100)=y(b:b+100);
   k=k+100;
   end;
  end;
  clear FS Threshold E sig y ;
```

```
  dest=dest';
  if j==1
   x1=dest;
  else
   x1=vertcat(x1,dest);
  end;
  clear dest;
 end;
 y1=mfcc_rasta_delta_pkm_v1(x1,16000,13,26,20,10,0,0,1);
 save(fullfile('mfcc_train1',dname),'y1');
 %clear y1 x1;
 b=dir('mfcc_train1');
 for i=3:length(b)
  dim=13;
  centres=16;
  MIX=gmm(dim,centres,'diag');
  load(fullfile('mfcc_train1',dname));
  options(14)=100;
  MIX=gmminit(MIX,y1,options);
  %MIX.priors
   OPTIONS(1)=-1;
   OPTIONS(14)=100;
  [MIX,OPTIONS,ERRLOG]=gmmem(MIX,y1,OPTIONS);
  if (m==1)
   save(fullfile('allcleanmodels_16(1)',dname),'MIX');
  end
  if (m==2)
   save(fullfile('allcleanmodels_16(2)',dname),'MIX');
  end
  if (m==3)
   save(fullfile('allcleanmodels_16(3)',dname),'MIX');
  end
  if (m==4)
   save(fullfile('allcleanmodels_16(4)',dname),'MIX');
  end
 %clear y1 MIX;
 end;
end
toc
```

```
%%mfcctest.m
%%Feature extraction of the test data
clear all;
clc;
tic
a=dir('testwavdata1/*');
for i=3:length(a)
 allwav=dir(fullfile('testwavdata1',a(i).name,'*.wav'));
 for j=1:length(allwav)
  fname=fullfile('testwavdata1',a(i).name,allwav(j).name);
  [y,FS,FFX]=wavread(fname);
  sig=y.*y;
  E=mean(sig);
  Threshold=0.05*E;
  k=1;
  for b=1:100:(length(sig)-100)
   if((sum(sig(b:b+100)))/100 > Threshold)
    dest(k:k+100)=y(b:b+100);
    k=k+100;
   end;
  end;
  dest=dest';
  clear FS FFX Threshold E sig y ;
  % clear fname dest;
  y1=mfcc_rasta_delta_pkm_v1(dest,16000,13,26,20,10,0,0,1);
  %%mkdir(fullfile('mfcc_test1',a(i).name));
  save(fullfile('mfcc_test1',a(i).name, regexprep(allwav(j).name,
'.wav', '')),'y1');
  %clear y1 dest fname ;
 end;
end;
toc
```

```
%%newgmmtest.m
%%GMM Classification Module
clc;
tic
totalError=0;
Error1=0;
Error2=0;
total1=0;
total2=0;
errorInstances2=0;
errorInstances1=0;
language1='amharic';
language2='guragegna';
language3='oromiffa';
language4='tigregna';
%%Load models
for m=1:4
 dname='amharic';
 load(fullfile('allcleanmodels_16(1)',dname));
 model1 = MIX;
 temp1 = -log10(gmmprob(model1,y1));
 if(m==2)
  dname='guragegna';
  load(fullfile('allcleanmodels_16(2)',dname));
  model2 = MIX;
  temp2 = -log10(gmmprob (model2,y1));
 end
 if(m==3)
  dname='oromiffa';
  load(fullfile('allcleanmodels_16(3)',dname));
  model3 = MIX;
  temp3 = -log10(gmmprob (model3,y1));
 end
 if(m==4)
  dname='tigregna';
  load(fullfile('allcleanmodels_16(4)',dname));
  model4 = MIX;
  temp4 = -log10(gmmprob (model4,y1));
 end
end
global1=0;
global2=0;
global3=0;
global4=0;
```

```
%Dir for test files
for j=3:length(a)
 % %Read all names in the test dir
    a=dir(fullfile('mfcc_test1',a(j).name,'*.mat'));
end
lang4=0;
lang3=0;
lang2=0;
lang1=0;
 %Probability dinsities
lang1=lang1+(temp1);
lang2=lang2+(temp2);
lang3=lang3+(temp3);
lang4=lang4+(temp4);

candidate = min(lang1,lang2);
candidate2 =min(lang3,lang4);
candidate3 =min(candidate,candidate2);

a=nnz(candidate3);
b=nnz(candidate3==lang1);
c=nnz(candidate3==lang2);
d=nnz(candidate3==lang3);
e=nnz(candidate3==lang4);
percentage1=(b*100)/a;
percentage2=(c*100)/a;
percentage3=(d*100)/a;
percentage4=(e*100)/a;
%% print the candidate
if ((percentage1 > percentage2)&&(percentage1 >
percentage3))&&(percentage1 > percentage4)
 sprintf('%s %2.2f',' The language is amharic')
elseif((percentage2 > percentage1)&&(percentage2 >
percentage3))&&(percentage2 > percentage4)
 sprintf('%s %2.2f',' The language is guragegna')
elseif((percentage3 > percentage1)&&(percentage3 >
percentage2))&&(percentage3 > percentage4)
 sprintf('%s %2.2f',' The language is oromiffa')
else
 sprintf('%s %2.2f',' The language is tigregna')
end

if(strcmp(dname, language1))==1
 Error1 = ((global2)/(global1+global2))*100;
 errorInstances2 =errorInstances2+global2;
 total1=global2+global1;
 l11=total1-(global2);
```

```
 l12=global2;
end
if(strcmp(dname, language2))==1
 Error2 = ((global1)/(global1+global2))*100;
 errorInstances1=errorInstances1+global1;
 total2=global2+global1;
 l22=total2-(global1);
 l21=global1;
end
toc
```