2020-03-24

DESIGNING AUTOMATIC PRONUNCIATION DETECTION FOR þÿ G • ¾ • Z L A N G U A G E

Abeje, Misganaw

BAHIR DAR UNIVERSITY

BAHIR DAR INSTITUTE OF TECHNOLOGY

SCHOOL OF RESEARCH AND GRADUATE STUDIES

FACULTY OF COMPUTING

DESIGNING AUTOMATIC PRONUNCIATION DETECTION FOR GƏʾƏZ LANGUAGE

MSc. Thesis

By

Misganaw Abeje

January, 2018

Bahir Dar, Ethiopia

DESIGNING AUTOMATIC PRONUNCIATION DETECTION FOR GƏʾƏZ LANGUAGE

By: MISGANAW ABEJE DEBASU

A Thesis Submitted to the Faculty of Computing, Bahir Dar University in Partial Fulfilment of the Requirements for the Degree of Master of Science in Computer Science.

Main Advisor: Dr. Gebeyahu Belay

Co-Advisor: Zelalem Belay

Bahir Dar, Ethiopia

10 January 2018

DECLARATION

I, the undersigned, declare that the thesis comprises my own work. In compliance with internationally accepted practices, I have acknowledged and refereed all materials used in this work. I understand that non-adherence to the principles of academic honesty and integrity, misrepresentation/ fabrication of any idea/data/fact/source will constitute sufficient ground for disciplinary action by the University and can also evoke penal action from the sources which have not been properly cited or acknowledged.

Name of the student: Misganaw Abeje__ Signature: _____

Date of submission: 10/01/2018__

Place: Bahir Dar, Bahir Dar University, Ethiopia

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

Advisor Name: _____Dr. Gebeyahu Belay__

Advisor's Signature: _____

Bahir Dar University

Bahir Dar Institute of Technology-

School of Research and Graduate Studies

Faculty of Computing

THESIS APPROVAL SHEET

Student:

Misganaw Abeje — Signature — 22/02/2018

Name — Date

The following graduate faculty members certify that this student has successfully presented the necessary written final thesis and oral presentation for partial fulfilment of the thesis requirements for the Degree of Master of Science in Computer Science. Name and signature of approval members of the examining board are:

Approved By:

Advisor:

Dr. Gebeyahu Belay — Signature — 22/02/2018

Name — Date

External Examiner:

Million Meshesha — Signature — February 15, 2018

Name — Date

Internal Examiner:

Dr. Tesfu Tegegne — Signature — 22/02/2018

Name — Date

Chair Holder:

Asegahegn Endalew — Signature — 22-02-2018

Name — Date

Faculty Dean:

Dawed N. — Signature — 22-02-2018

Name — Date

ii

ACKNOWLEDGEMENT

First and foremost, I would like to praise the GOD, who favors me from beginning to end of this study, Without GOD noting happen forever.

I would like to express my deepest gratitude to my advisor, Dr. Gebeyehu Belay, for his continuous support and guidance throughout the various stages of this thesis. Without his contribution and interference at critical stages, it would have been very challenging to complete the thesis. He provided critical and useful feedback and suggestions on how to address the research problems systematically and tactfully. He showed me how research will produce, how to contact, and follow-up his students. His encouragement has always inspired me and help to see in all direction, to overcome good result and the completion of work in time, thank you so much.

And also, Memihir Zelalem Belay, who is my co-advisor and his support in my life to live meaningful life. His wished to see my success and very happy due to my research area focussed on geez language in modern technology, his full-time volunteer to help me to complete this work from the beginner (formulation of the problem) to final stage, am saying thank you so much. And his friend also (memhir Temework, memhir Friew, memhir Sibhat) am really thank you, your supporters are very useful for completion of my works by answering my question and who have participated for data collection.

I would like to thank you Bahir Dar Institute of technology school of research and postgraduate studies and faculty of computing for their financially supported for this thesis and MSc class.

I would also like to give special thanks to my family (Abeje Debasu, Zewuditu Kebede, all my brothers and sisters, and my special family, who is my lovely wife Maza Abineh and my little daughter Yeabsira Misganaw). For their help and support from the early stage of school life to MSC class level until the completion of this work.

Finally, I would like to thank everybody who was important to the successful realization of a thesis, I would like to express my apology that I could not mention personally.

ABSTRACT

Language is one of the essential aspects of human behavior, which has a major role in our daily life activities. It helps in transferring information and knowledge for generations in the form of written and speech format. Natural Language Processing (NLP) is used to perform useful tasks involving human language using an electronic device, tasks like enabling and improving human-machine, human-human communication, or simply doing useful processing of text or speech. Pronunciation detection is a technique used in speech processing for identifying the pronunciation style from speech signals.

Nowadays computer-assisted language learning (CALL) system can provide many possible benefits to both the language learner and teacher by processing text and speech data. They allow the continuous reaction to the student without requiring the individual helpfulness of the teacher or scholars, however in the area of pronunciation training, which often requires the full attention of the teacher for only single learners, therefore the aim of this study to design pronunciation detection. Whereas Gəʾəz language used the religious language in Ethiopia, specifically in Ethiopian Orthodox Church at this time. And also it came to a research area in different fields. In Ethiopia, different kinds of literature for several ancient manuscripts, arts (qɗne), scriptures, and heritages, historical, ethical and religious chronicles are written by Gəʾəz language. Those literature, books, and any manuscripts contain different thought and attitudes on the philosophy, creativity, knowledge, civilization, ethical and cultural principle of Ethiopia. The usability of those books like the bible and any pray of books are a day to day activates of the people. The main problem to understand that literature and books are the difficulties of Gəʾəz language pronunciation. Mispronunciation is lead to distortion of the meaning and alter of POS. So, this thesis is focused to design a model for solving such a type of problems and challenges.

The acoustic features are used for pronunciation detection using MATLAB with ANN learning algorithm. The chosen acoustic features of speech signals could be easily represented into the system and training for the target pronunciation. After suitable times of training process, extra test signals are load into the system for pronunciation detection, which gives an accuracy of 76.9% classification rate.The result shows that the selected acoustic features (such as speech rate or duration, energy, pitch, formant and MFCC) are proven to be good representations of Gəʾəz language pronunciation for the speech signal. Finally, this thesis has comprehended the design automatic pronunciation detection for Gəʾəz language.

.

TABLE OF CONTENT

List of figures

List of Tables

Abbreviations

ACF	Autocorrelation Function
AMDF	Average Magnitude Difference Function
ANN	Artificial Neural Network
CALL	Computer-Assisted Language Learning
CAPT	Computer Assisted Pronunciation Training
EPD	End-Point Detection
FFT	Fast Fourier Transforms
EOTC	Ethiopia Orthodox Tewahido Church
FF	Fundamental Frequency
FS	Sample Frequency
HMM	Hidden Markova Model
KNN	K- Nearest Neighbor
MATLAB	Matrix Laboratory
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
POS	Part of Speech
QBSH	Query by Singing/Humming
SIF	Simple Inverse Filter
SVM	Support Vector Machine
TTS	Text-To-Speech
VAD	Voice Activity Detection
ZCR	Zero-Crossing Rate

CHAPTER ONE:

INTRODUCTION

Language is one of an essential aspect of human life, which has a major role in our day to day activities. In its written form, it used to keep information and knowledge transferring for generations. In its spoken form it supports to cooperate in the human activities with each other (Allen, 1995). Because of the environmental and sociological attachment, languages can be characterized as natural language (Jurafsky and Martin., 2006), such as Amharic, Afaan Oromo, Gəʾəz, Tigrinya, English, French, and Arabic and so on.

Natural Language can be explored in various disciplines each of these disciplines has their own set of problems. For example, linguists are concerned with the structure of the language, how words are arranged to form sentences, how words are pronounced and why certain word arrangement cannot produce the correct sentence. On the other hand, computational linguists develop a computational theory of the natural language and model based on the concept of algorithm and data structure of computer science (Allen, 1995). Computational linguistics (also known as Natural Language Processing (NLP)) is a field of study that processes natural language (Jurafsky and Martin., 2006). This field is used to perform useful tasks using Computers that involve human language, tasks like enabling and improving human-to-machine, and human-human communication, by processing of text or speech (Jurafsky and Martin., 2006). It also aims at designing and building applications to understand, imitates a language where human can use to the extent their possible communication with the machine via computer (Allen, 1995). Understanding of natural language involves knowledge what concepts a word or phrase stands for and how to link those concepts together in a meaningful way. Natural language is understandable for the humans to learn and use, but it is complex for a computer system (Bird, Klein and Loper, 2009).

NLP usually comprises one or more levels of linguistic analysis such as word, phrase, sentence, semantic levels, etc. It is thought that humans normally utilize all of these levels since each level conveys different types of information. According to Allen (1995), NLP can be classified into many subfields shown as:

- **Phonetic and Phonology**: studies how words are related to the sounds. Speech synthesis, speech recognition etc.
- **Morphology**: concerned with how words are constructed from more basic meaning units called morphemes. Like sentence grammar, we have also word grammar or rule that govern how words can be formed from other words or changed by a bit modification to other words.
- **Syntactic study**: deals with how words can be put together to form right sentences and determine what structural role each word plays in the sentence and what phrases are subclasses of other phrases. Different languages may have different word categories, or they might associate different properties to the same word (Allen, 1995). Such behaviours of natural language are studied under this subfield. Taggers, word category disambiguation is grouped under the syntactic study of natural language.
- **Semantic study**: concerned with what the word means and how the meaning come together in sentences to form a meaningful sentence. It is the study of context meaning independently, the meaning of a sentence has regardless of the context in which it is used.
- **Discourse**: studies how the instantaneously preceding sentences affect the interpretation of the next sentence. This information is especially important for the interpretation of pronunciation and to interpret the temporal aspects of the information conveyed.

Nowadays computer-assisted language learning (CALL) system can provide many possible benefits to both the language learner and teacher by processing text and speech data. They allow a continuous reaction to the student without requiring the individual helpfulness of the teacher or scholars, it facilitates self-study and inspires the interactive use of the language in preference to rote learning (Witt and Young, 1998). Computer-based language learning systems potentially can provide some advantages over traditional methods, especially in the area of pronunciation training, which often requires the full attention of the teacher for only single learners. If the computer could provide the type of feedback that a pronunciation teacher provides, it would be a much good alternative, accessible at any time and at any place, and certainly tireless (Franco, Neumeyer, Ramos and Bratt, 1999). In such kind of system, the computer provides the response of the kind that an instructor would produce, such as an assessment of the quality of pronunciation or pointing to specific mispronunciation. Finally,

they can be used to evaluate the learners, to determine whether correctly pronounced or not to provide their status of the language or language competence. Speech recognition technology is crucial to the automatic evaluation of pronunciation quality. However, standard speech algorithms were not designed with the goal of evaluation of pronunciation quality. Therefore, new methods and algorithms must be planned to match the perceptual capabilities of human listeners to grade speech quality (Neumeyer, Franco, Digalakis and Weintraub, 2000). The aim of this study is to design a model that identify correctly pronounced the Gəʾəz language at the word level based on the acoustic feature and machine learning approach, which is pronunciation detection is used to add a judgment of the fluently of speech.

Gəʾəz language is one of the natural languages under Semitic language family. Pronunciation in Gəʾəz language has a great role for study and learn the language resource. To pronounce Gəʾəz language we must know the way of pronunciation for each pronunciation style. According to different Gəʾəz scholars and books, Gəʾəz language pronunciation categorized into two broad categories. Those categories are major pronunciation and minor pronunciation of the language (አፈወርቅ፣ 2005; ያሬድ፣ 1997; ደሴ ፤ 2007). Major pronunciation styles are four types. The minor pronunciation styles of Gəʾəz language are more than eight, which does not well defend the exact types.

There is no previous work for Gəʾəz language pronunciation identification. Therefore, this study is proposed to design a model automatic pronunciation detection of speech signal.

1.1 Statement of the Problem

For a given country, a language is a fundamental tool which has a great role to study and analyses the culture and historical information. It used to, keep information and knowledge transferring for generations. Language is also essential to protect country's' philosophy, tradition, history, knowledge, and sovereignty, which reflects, their identity (Desta, 2010).

Gəʾəz is an ancient and historic language of Ethiopia, which discovers in 4000 years ago (Leslau, 1987; Fikire, 2008E.C). Gəʾəz language is the base of Ethiopia because of many of Ethiopian literature and manuscripts are written in this language. Several ancient manuscripts, arts (qȅne), scriptures, heritages, historical, ethical and religious chronicles are a primary reference from Gəʾəz manuscripts. Some of the literature and manuscripts are bibles, ገድላት /gädəlatə መልክታት/mäləkətatə፣ ድርሳናት/dərəsanatə፣ ፍትኀ ነገስት/fətəhä nägäsətə፣ hymn books, ʾአቡሻህር/äbušahərə, and different historical books are written in geez language. Those literature,

books, and any manuscripts contain different thought and attitudes on the philosophy, creativity, knowledge, civilization, ethical and cultural principle of Ethiopia. The usability of those books like the bible and any pray of books are a day to day activates of the people. The main problem to understand that literature and books is difficult to pronounce Gəʾəz language. Mispronunciation is led to a distortion of the meaning and alter of POS.

Currently, Gəʾəz language used as the religious language in Ethiopia, specifically in Ethiopian Orthodox Church. And also it came to a research area in different fields and opening the programs in universities. According to Desta, (2010) anyone who is proposed to survey or conduct a research on issues related to the history, tradition, custom, politics of the Ethiopians, and to explore the works handed down from the ancient to the present generation. It needs to go to traditional or church schools, the learner can gain the skills as its standard and methodology. As per the review of these multi-functional scriptures and literature, anyone who has the skill can have multi-dimensional benefits for the ancient Gəʾəz scripts. It has many problems, including, the literature has to be translated into spoken languages manually, which led to taking long time, and meaning distortions due to mispronunciation.

Additionally, there is a limited research that has been done in the area of computational linguistics in relation to Semitic languages like Gəʾəz. In some extent designed the model for morphological analyzer for Gəʾəz verbs, by using rule-based with consonant-vowel based for analysis of morph syntactic features including affixes together with their syntactical functions was done (Desta, 2010). In his study, the subjects and objects along with their person, gender-number features, tense-mood of the verbs were identified. And another research conducted on verb classification of Gəʾəz verbs into heads and troops (Muluken, 2007). According to muluken's the main criteria used to classify verbs are forms of letters. In Gəʾəz language pronunciation identification is challenged for reading Gəʾəz words, because mispronunciation is led to the incorrect meaning of the given words. When we read Gəʾəz texts with mispronunciation, it is difficult to understand the speeches, which requires the scholars. Even speech that is intelligible, but enforced by Amharic and other language reading style, can elicit negative reactions in the scholars. To teach the pronunciation of geez language is, it requires more practice time and teacher feedback and it's different their teaching-learning methodology of geez language from the modern education system. For this reason. As to the researcher knowledge, there is no research that conducted on pronunciation detection for Gəʾəz language, which is becoming a barrier for researchers such as speech to text translation and speech

recognition. Therefore, we propose to develop a model for automatic pronunciation detection for Gəʾəz language word at level.

1.2. Research Question and Motivation

Ethiopia is one of the ancient countries in the world. It has a well-defined history of more than three thousand years (Lulie 1986), an ancient and well-developed educational system, philosophy and literature, which are unique attributes. Additionally, the country has its alphabetic language numerical system, manuscripts, arts, calendar, and hymns, which make it unique country (Bender, 1976.; Dillmann, 1899). Most of such kinds of literature that are used as identities of the country are found being written in Gəʾəz language (Desta, 2010).

Anyone who is proposed to survey or conduct a research on issues related to the history, tradition, custom, politics of the Ethiopians and to explore the works handed down from the previous generations, first review these multi-functional Scripts and kinds of literature. Unlike another language speaking and reading skill Gəʾəz language is difficult because of it follows different pronunciation style. Therefore, this is the motivation to conduct this research. The following research questions are formulated and answered:

- How to identify one pronunciation style of the word to others? Are there any unique features that can be identified Gəʾəz language pronunciation style?
- How to develop an automatic pronunciation detection in speech data?

1.3. Objective of the study

1.3.1. General Objective

The main objective of this thesis is design automatic pronunciation detection for Gəʾəz language pronunciation.

1.3.2. Specific Objectives

In order to achieve the general objective stated above, the study attempts the following specific objectives:

- To identify pronunciation style features of Gəʾəz language;
- To study the pronunciation ways of Gəʾəz words for identifying important features' pronunciation.
- To study the acoustic features of Gəʾəz word's speech waveform for used to pronunciation detection.
- To develop a prototype of pronunciation detection for Gəʾəz words;
- To evaluate the performance of the model

1.4. Methodology of the study

1.4.1. Research Design

The research design is the conceptual structure within which research is conducted; it constitutes the blueprint for the collection, measurement, and analysis of data. There are different types of research design. However, in this thesis, we are followed an experimental research design approach to achieve the objective of the study. The purpose of the study is, to identify the pronunciation style of geez words based on the acoustic features of the speech signal. To evaluate the acoustic features factors on the detection of pronunciation style, we conduct an experiment through MATLAB and machine learning approach.

1.4.2. Feature Selection and Corpus Preparation

To identify the relevant features of Gəʾəz language pronunciation, different kinds of literature are thoroughly reviewed. The primary sources of data and information for this study are scholars of Gəʾəz in the Ethiopia Orthodox Tewahido Church (EOTC) that has studied Gəʾəz both traditionally and modern scholars on qɔ̂ne (ቅኔ). In this thesis requires many contacts and discussion with scholars of Gəʾəz language to have a better understanding and to identify the feature of the pronunciation of the language. Furthermore, kinds of literature that are reviewed including books, research reports, journal articles, manuals, and other published and unpublished documents. Therefore, by the help of Gəʾəz scholars, we prepared the speech corpus. The scholars, who have certified in both school of qene and Gəʾəz books.

1.4.3. Feature Extraction of the Pronunciation

So far, there are no ready-made acoustic features of Gəʾəz language pronunciation or dataset that help to identify the style of Gəʾəz language pronunciation, which is the main process of this thesis. In this study, features are referring to ways or accent of the pronunciation of the language. Those features are extracted based on the acoustic features of speech signal or waveform representation of the speech.

1.4.4. Designing Model

In this thesis, to design a model we are followed machine learning approach. There are different algorithm in machine learning approach such as an artificial neural network (ANN), Hidden Markova Model (HMM), K-Nearest Neighbor (KNN), Support Vector Machine (SVM). From those, we are used artificial neural network (ANN) to detect the style of pronunciation of the words from a speech signal, which is most popular algorithms for speech recognition problems, pattern prediction by summarizing and generalizing solution for the problems (Tebelskis, 1995; Mohammed A. , 2012) .

1.4.5. Tools and Implementation

To develop the proposed model of automatic pronunciation detection, we used MATLAB and acoustic feature based on the representation of speech signals, and also praat and audacity used for the purpose of editing speech signals. MATLAB is more user-friendly to machine learning algorithms and popular tool. When we Link with MATLAB a Neural Network Toolbox that provides many of the functions needed to implement any type of neural network. For the speech processing, MATLAB is powerful tools and open source. MATLAB is a numerical computing environment which allows matrix manipulation, implementation of algorithms and plotting of functional data (Liu etal, 2010).

1.4.6. Evaluation procedure

Pronunciation detection is used ANN algorithm to learn the relevant features from the input layers. The input layers are referred to the features of the speech signal of Gəʾəz words. ANN is required more training times and adjusting the value of hidden layers to get the best performance of result. After getting the good performance of the model we evaluated by testing data set, which has 10% of the training data. The test data set are separate data from the training data. Finally, the pronunciation detection output is cross-checked with detecting by the experts,

and the model performance measured based on the number of correctly detected the pronunciation styles.

1.5. Scope and Limitation the study

In Gəʾəz language there are different pronunciation styles, However in this thesis, we are focused on the major pronunciation style of the language, those styles are ተነሽ ንባብ/tänäšə nəbabə (rising up reading), ወዳቂ ንባብ/Wädaqi nəbabə (folding reading), ሰያፍ ንባብ/säyafə nəbabə (slant reading) and ተጣይ ንባብ/tätʾayə nəbabə (throwing reading). Studying the major pronunciation styles before the minor pronunciation is very important as many of Gəʾəz words, phrase sentence, books, are pronounced in this pronunciation type. Whereas the minor pronunciation is obtained from major pronunciation and Gəʾəz words may not be pronounced by minor pronunciation. Additionally, types of minor pronunciation style were not well defined. Therefore, the minor pronunciation style of the language does not cover. In this study, the model design is based on the acoustic feature of the speech signal at the word level.

Further, the research is limited to the major pronunciation styles of Gəʾəz language because they have a dominant and representative nature among the other styles. As to the survey conducted by the researcher on the words that found in Gəʾəz language literatures and books are pronounced by this pronunciation styles. And hence, studying this pronunciation will imply studying roughly all other styles with some modifications on the algorithms developed for this pronunciation.

1.5. Significance of the Study

Automatic pronunciation detection is a type CALL system can provide many possible benefits to both the language learner and teacher and used for higher form NLP, which is a components speech processing. Therefore, this model is contribute to speech recognition, speech to text translation by detection the correct the way of pronunciation. In addition to this:

- It can help Gəʾəz learners for reading and speaking skill Gəʾəz language pronunciation
- Assigning word class of a given word, it can be used for statistical work such as counting the distribution of different word classes in corpora
- To help easily understand any documents written in Gəʾəz language.

1.6. Organization of the Thesis

The whole thesis is organized into six chapters. The first chapter describes the introductory part. The second chapter is for literature review, which approach design the model related work are discus in this part. Chapter three focused on methodology of the study, which is the features of pronunciation style of a language, acoustic feature identification and feature extraction. Chapter four deals with architecture and design of the pronunciation detection. The fifth chapter deals with corpus preparation and experimental results and discussion of the model. Finally, the last chapter is on conclusion and recommendation of the thesis.

CHAPTER TWO

Literature Review

2.1. Overview

As described in the introductory part of the thesis, the main objective of this study is to design an automatic Gəʾəz language pronunciation detection, which takes speech signal at the word level. In this chapter, we review a literatures about Gəʾəz language automatic pronunciation identification, model design approaches and related works from the existing literatures that related to automatic pronunciation detection.

2.2. Phenomena of Gəʾəz Language

So many researchers are conducted about Gəʾəz language phenomena as we reviewed from different researches, Gəʾəz is the classical language of Ethiopia within the Semitic language family (Leslau, 1987). And also, according to (ፍቅሬ, 2008e.c; በላይ, 2007e.c) Gəʾəz language or alphabet is discovered in Ethiopia by Ethiopian before 4000 years, and each alphabet/or letter have its own meaning and equivalent number.

A study of the Ge'ez writing systems is essential to understanding the history of Ethiopia and the evolution and also modern usage of the Roman alphabet. This is not to say, by any means, that Ge'ez is merely a "bridging" system that serves to connect only ancient pictograms to the modern western alphabet, though that relationship may be unjustly implied in a western study concerning roman letter forms. In comparisons with the ancient Ethiopic script since their origins are essentially the same, to say that Ge'ez is an ancient language whose evolution stopped yet where roman letter forms began is a very easy trap to into, especially in a distinctly Eurocentric society. In the case of Christianity in both Roman and Ge'ez systems, the philosophical and religious sacred connections of a writing system took precedence over a commonly spoken language. (Scelta, 2001).

2.3. Automatic Pronunciation

Pronunciation learning is one of the most important parts of second language acquirement. Which, is Computer Aided Language Learning (CALL) has received a considerable attention in recent years. Many research efforts have been done for development of such systems especially in the field of second language teaching. The two desirable features of speech enabled computer-based language learning applications are the ability to recognize accented or mispronounced speech produced by language learners, and the ability to deliver meaningful feedback on pronunciation quality (Srikanth and Salsman, 2012). Also learning a second language takes time and dedication, not only from learners, but also from teachers, hence both face-to-face, and personally online language learning are very expensive and not accessible at all. A large and still growing number of CALL in the market has shown a clear trend. (Ai, 2015) Suggest that language learning is going to be web-based, interactive, multimedia and personalized, so that learners are flexible as to times and places for learning.

Computer-based solutions for pronunciation training are becoming gradually an ordinary for foreign language learning purposes (Dalby, Kewley and Sillings, 1998; Menzel , Herron, Bonaventura and Morton, 2000). Nevertheless, currently the available solutions offer considerable area for improvement, mainly with respect to the feedback generated by the system. In many cases a simple playback facility, the visual demonstration of the signal form, a scoring mechanism, or even the identification of the mispronounced word or words might not be sufficient to give students helpful hints on how to improve their pronunciation. Only by finding out the precise nature of the student's mistake is a system in a position to provide error explanations, problem-specific speech stimuli, or individualized suggestions for improvement. Furthermore, such a low-level diagnosis makes it possible to guide the student into specifically tailored opportunities for practice.

Automatic pronunciation error detection is developed in different language, which used to compare existing measures to a metric that takes account of the error patterns observed to capture relevant acoustic differences studies do indeed show that error patterns bear information that can be usefully employed in weighted automatic measures of pronunciation quality at phoneme level (Doremalen, Cucchiarini and Strik, 2013). However, language has its own systematic rules governing pronunciation, word formation, and grammatical pronunciation and son on. Therefore, in this study, we develop pronunciation detection which act as scholar or teacher by showing the pronunciation capability of the speakers the language.

2.4. Acoustic Features

In this section, we discuss some concept of sound or speech and its feature used for detection of Gəʾəz language pronunciation according to the speech signal. The sound is a continuous wave that travels through a medium. Whereas sound wave is an energy that causes the disturbance in a medium that made of pressure differences by propagated through a medium. Acoustic is the study of sound that is, its generation, transmission, and reception of sound. To detect the pronunciation style of Gəʾəz language first, we use an audio signal as input file and extract the basic sound feature or acoustic features. Audio signals can be defined as any physical quantity that varies with time, space or any other independent variables. In this section, we describe speech pre-processing, (starting from recording the data (words), sampling, normalization, and segmentation), acoustic feature selection, and feature extraction from the speech signals.

2.4.1. Speech Pre-Processing

Pre-processing is an essential signal processing, that applied before extracting features from a speech signal, for the purpose of enhancing the performance of feature extraction methods. When we analyze audio signals to detect the pronunciation of Gəʾəz language among the four types or styles of pronunciation. We usually used the method of short-term analysis since most audio signals are more stable within a short period of time. In this study, we design a model for Gəʾəz language pronunciation detection in single word speech signal. So before selecting and extracting the features of speech, we performed some pre-processing steps of speech signals. Those are mainly including first we record the Gəʾəz word with the native speaker, sampling, normalization, and segmentation.

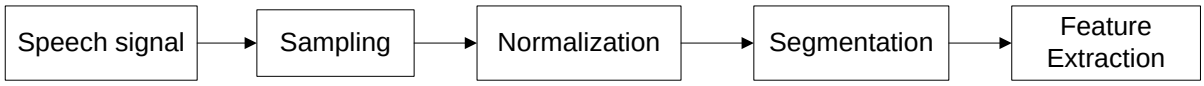

Figure 1: Speech signal Pre-Processing (Rui, 2014)

Speech voice is an Analog signal, and it needs to be converted into a digital signal to process in digital device or computers. When we see sampling theory, it provides a way to transform the Analog signals x (t) into a discrete time signal x (n) and remains the characteristics of the original Analog signal. In another word, we can say that sampling is measuring the amplitude of the signal at time t. According to sampling theorem (Proakis and Manolakis, 1998). When the sampling frequency is larger or equal to two times of the maximum of Analog signal frequency,

then the discrete time signal is able to reconstruct the original Analog signal. Sampling is performed by collecting points from Analog signals in certain time rate that perform is performed by collecting points from Analog signal in a certain rate. Generally, the common sampling frequencies for speech signals are 8000HZ, 16000HZ and 44100HZ (Rui, 2014; Proakis and Manolakis, 1998). In this thesis, we have used 16000HZ sample frequency rate 32-bit rate. In mat lab sampling frequency is applied automatically by calling audio read function after accepted the signals.

After the sampling pre-processing techniques, we perform the normalization process. When we say normalization, it makes the audio files or the data inconsistency, which is also an important aspect that needs to be considered for a robust automatic pronunciation detection by reducing the source of variability, preserving the differences in recording environment or conditions. The quality and features of the speech highly depend on the sensor, and any device and the technology used to capture and transmit the speech (e.g., mobile versus landline systems, close-talking versus far-field environment microphones). An unmatched in the recording between the training and the testing speech sets can affect the features extracted from the signals. Generally recorded signals often have been varying energy levels due to speaker volume and microphone distance. Amplitude Normalization can cancel the inconsistent energy level between signals, thus can enhance the performance of energy-related features. There are several methods to normalize a signal's amplitude. One of them is achieved by a point-by-point division of the signal by maximum absolute value so that the dynamic range of the signal is constrained between 0.0 and +1.0.

For this purpose, we use a normalization process which is, all signals of the input audio divided by their RMS (root mean square) value of each signal is the same. Values of a root-mean-square (RMS) is used as an estimated of signal loudness. The formula of normalization is given below in Equation 1 (Rui, 2014) .

$$S`(n) = \frac{S(n)}{Smax}\ , n = 1,2,3 \dots N \quad \text{..... (1)}$$

Where s (n) is the original sampled signal, S` (n) the signal after normalization, s_{max} is the absolute maximum value of the signal sequence, and N is the length of the sequence.

Speech is random signal and also its feature is changing with time, but this change is not instant. Generally, in this thesis we assume that speech is in a short duration or a single Gəˀəz word

speech duration so that the signal is stable. In the segmentation process, it divides the signal sequence into many frames with overlap. Overlapping is used to avoid loss of data due to aliasing (Proakis and Manolakis, 1998).

2.5. Machine Learning Approach for Designing the Model

Machine learning is the study of algorithms that allow computer programs to automatically improve through experience (Mitchell, 1997). The major focus of machine learning is to extract information from data automatically, by using computational and statistical methods. Its techniques are being used for solving various tasks of Natural Language Processing. This includes speech recognition, morphological analysis, document categorization, document segmentation, part-of-speech tagging, and word-sense disambiguation, named entity recognition, parsing, machine translation and transliteration (Kumar, 2013). There are two main tasks involved in machine learning; learning/training and prediction. The system is given with a set of examples called training data. The primary goal is to automatically acquire effective and accurate model from the training data. The second phase of machine learning is the prediction, where a set of inputs is mapped into the corresponding target values. The main challenge of machine learning is to develop a model; with good prediction performance on the test dataset, i.e., the model with good generalization on unknown data (Kumar, 2013).

Machine learning is a branch of Artificial Intelligence (AI) concerned with the design of algorithms that learn from examples. Machine learning algorithms can be supervised or unsupervised. The input and corresponding output data are used in supervised learning. In unsupervised learning, only input samples are used. The goal of machine learning approach is to use the given examples and find out generalization and classification based on the features of the pronunciation automatically. The feature refers to the acoustic features of the pronunciation of the language. After identifying the relevant acoustic feature, the machine learning algorithms learn from training data by considering the feature to detect the pronunciation styles of words. From severe algorithms of machine learning approach, the most popular widely used in speech processing are ANN, KNN, HMM and SVM (Mitchell, 1997).

Support Vector Machine (SVM): It is a commonly used "eager learning" method. It finds the best separating hyper-plane between two sets of classes in such a way that the distance between the two classes is maximized. Using different kinds of kernel functions, the separating hyper-plane can be found in a space of higher dimensionality than the data itself. It performs

especially well with sparse data (Samer, 2011). Advantages of SVMs High accuracy, nice theoretical guarantees regarding overfitting, and with an appropriate kernel they can work well even if you're data isn't linearly separable in the base feature space. Especially popular in text classification problems where very high-dimensional spaces are the norm (Dumais,Platt and Heckerman, 2010).

Hidden Markov Model (HMM): It is a popular statistical tool for modelling a widespread of time series data. It is a dominant statistical tool for modelling generative sequences that can be characterized by an essential process to generating an observable sequence (Blunso, 2004). HMMs have initiate application in many areas interested in signal processing, and in particular speech processing, but also have been applied with achievement of low level NLP tasks such as phrase chunking, POS tagging, and extracting target information from documents. A key benefit of the statistical approach to speech recognition is that the required models are trained automatically on data (Gales and Young, 2008).

The k-Nearest-Neighbours (k-NN): is a method for categorizing objects based on closest training instances in the feature space. K-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification (Nikhath, Subrahmanyam and Vasavi, 2016). It is a case-based learning method, which keeps all the training data for classification. Being a lazy learning method prohibits it in many applications such as dynamic web mining for a large repository and for text classification. For a data record t to be classified, its k-nearest neighbours are retrieved, and this forms a neighbourhood of t. Majority voting among the data records in the neighbourhood is usually used to decide the classification for t with or without consideration of distance-based weighting. However, to apply k-NN we need to choose an appropriate value for k, and the success of classification is very much dependent on this value. In a sense, the k-NN method is biased by k. There are many ways of choosing the k value such as by calculating Euclidian distances, but a simple one is to run the algorithm many times with different k-values and choose the one with the best performance (Nikhath, Subrahmanyam and Vasavi, 2016).

Artificial neuron network (ANN): is a computational model that inspired on the structure and functions of biological neural systems. Information that flows through the network affects the structure of the ANN because a neural network changes - or learns, in a sense - based on that input and output (Alpaydın, 2010). ANNs are considered non-linear statistical data modelling

tools where the complex interactions between inputs and outputs are patterns are found. An ANN has several advantages but one of the most recognized of these is the fact that it can actually learn from observing data sets. In this way, ANN is used as a random function to estimate general solution. These types of tools help estimate the most cost-effective and ideal methods for arriving at solutions while defining computing functions or distributions. ANN takes data samples of the entire data sets to arrive at solutions, which saves both time and money. ANNs are considered fairly simple mathematical models to enhance existing data analysis technologies.

Artificial neural network (ANN) is provide a general, real method for learning vector-valued real-valued, and discrete-valued functions from data instances. Algorithms such as backpropagation use gradient to tune network parameters to best fit a training set of input-output pairs. ANN learning is robust to errors in the training data and has been successfully applied to problems such as interpreting visual scenes, speech recognition, and learning robot control strategies (Mitchell, 1997). Neural networks are composed of simple computational elements operating in parallel. The network function is determined largely by the connections between elements, the element is refers to the layer of the neural. We can train a neural network so that a particular input led to a specific target output (Wouter, Georgi and Valeri, 2010).

ANNs have three layers that are interconnected. The first layer consists of input neurons. Those neurons send data on to the second layer, which in turn sends the output neurons to the third layer that is, the inputs are summed and sent through an activation function. An output neuron is the connection to the outside world. The second layers is hidden layer of the network, which is not directly reachable to the outside world. It has been shown that with enough neurons in the hidden layer any continuous function may be learned (Lippamann, 1987). In ANN activation function are applied, the most and widely usable activation function is sigmoid function. The simplest transfer activation function is linear. This is activation function used in the original perceptron's. A linear threshold function is simple non-linear activation function, in which the output of the neuron is in between+1 and-1. The advantage of using such functions is that very small and very large summations of layers with weights can still be processed and the neuron can operate over a wide range of input levels. The third layers is an output layer Training an artificial neural network involves choosing from allowed models for which there are several associated algorithms.

The backpropagation learning algorithm was made widely popular (Rumelhart, Hinton and Williams , 1986). Learning with backpropagation involves by defining the proper set of connection weights to estimate a given training set. A training set consists of expected outputs for specific inputs. The learning process involves solving the network for a set of inputs, comparing the outputs to the expected values, and then using the errors to estimate a correction to each weight value in the network.

The training process in ANN is repeated iteratively until the network has highest matched its outputs with the training set, which is refers the performance of the classifier. A trained network will have the property of generalization. This property is evaluated by testing the network with a data set which is similar procedure on the training data, but non-intersecting with the training set. If the results for the test set are the approximately the same to training set, then the network may be said to have generalized. If the network has converged, but has not generalized, then the network may said to have memorized the training set. If the network generalizes, then it should be able to handle any problem that is similar to the training set. The back propagation learning rule is simply a gradient descent algorithm, which minimizes the squares of the differences between the actual and desired outputs, summed over the output neurons for all training examples. The initial state in the network has assigned by random set of connection weights. Because off, when a system starts with all connection weights being identical, the network begins at a kind of local optimum, and will not converge. The rule that used to modifying the connection weights for a single neuron is called the delta rule. The weights on each input should change by an amount delta value, which is proportional to the error signal and the input signal of the neuron.

ANN had the ability to learn how to do task based on the data given for training, learning and initial experience. It can be used in many different applications such as pattern recognition, which is a powerful technique for harnessing the data and generalizing about it. And also, ANN used for speech recognition, speech translation, voice detection (Kudirp, Said and Nayan, 2012; Tebelskis, 1995; Murphy, 2014; Zegers, 1998). The neural networks are composed of simple computational elements that processed in parallel. The network function is determined mostly by the connection between the elements. We can train a neural network so that a particular input leads to a specific target output (Haykin, 1999; Wouter, Georgi and Valeri, 2010).

To detect the pronunciation style of Gəʾəz word the model accepts speech signals of a word as an input and then the model identifies the given word pronunciation styles ተነሽ/tänäšə/ rising-up, ወዳቂ/Wädaqi nəbabə/folding reading, ሰያፍ/säyafə/slant reading, and ተጣይ/täṭʾayə/ throwing reading). As (Mohammed A. , 2012) suggested to design the model for this kind of purpose, we can follow different methods or algorithm. The first one is being rule-based that are describe in section **Error! Reference source not found.**. The second is statistical, and the third one is NN. His works on machine learning approach voice detection, which is the classification of a speech segment as voiced/unvoiced/silence with voicing detection system. According to Mohammed's research, supervised method of voiced/unvoiced/silence speech segment detection and both text and speech corpuses are used. As it showed in the result, or the evaluation of the experiments shows that best performance from the selected classifier model ANN more accurate than others.

Acoustic features are a sound or speech properties that used for analysis of pronunciation features, which is the sound units of a language based on speech features that extracted from the audio signal (Solomon, Martha and Wolfgang, 2009). In which Gəʾəz language pronunciation have different features /properties that are obtaining from speech signals. Sound or speech properties are perceptual properties such as pitch, loudness, timbre and physical properties such as frequency, amplitude, envelope and spectra. Pitch is one of perceptual property of speech that allows their ordering on a frequency-related scale, or more commonly, pitch is the quality that makes it possible to determine sounds as "higher" and "lower. Pitch used to determine in sounds frequency that checks clarity and stability, to differentiate from unwanted signal or noisy. Pitch is a major auditory attribute of musical tones, along with duration, loudness, and timbre (Lee, Mower, Busso, Narayanan,, 2011). According to (Lee, Mower, Busso, Narayanan,, 2011). Prosodic features have also been used to classify emotions in the speech signal. Pitch-max, pitch-min, pitch-mean and pitch-median are the famous features that have been used to classify such emotions. Pitch or the intensity of speech signals vary from person to person, and even from time to time for same person. It also depends on recording through different devices and the input length of the signal (Kudirp, Said and Nayan, 2012) .

2.6. Related Works

CMUSPhnix was used for automatically evaluating the pronunciation and detecting specific phone level segments that have been mispronounced by a non-native student of a foreign language (Srikanth and Salsman, 2012). The researcher develops for American English at the phone-level information allows a language instruction system to provide the student with feedback about specific pronunciation mistakes by means of HMM.

A system called Fluency is developed to correct and detect foreign speaker's pronunciation errors in English (Seymore, 1996). The researcher also used automatic speech recognition to detect pronunciation errors and to provide suitable correct information. The other researcher (Peabody, 2011) focused on the problem of identifying mispronunciations made by non-native speakers using a CALL system. From the acoustic features, Mel-frequency cepstral coefficients (MFCCs) method is used for transforming into a feature space that represents four key positions of English vowel production for robust pronunciation evaluation are used.

Digalakis and Moustroufas (Digalakis and Moustroufas, 2007) presented several techniques to evaluate the pronunciation of students of a foreign language, again without using any knowledge of the uttered text. The researchers are used the native speech corpora for training purpose to evaluate the pronunciation.

Doremalen, Cucchiarini and Strik (Doremalen, Cucchiarini and Strik, 2009) are attempted to compare and combine the confidence measures with MFCCs and phonetic features. Due to the frequent pronunciation errors made by second language learners of Dutch is often concern vowel substitutions. To detect such pronunciation errors, ASR-based confidence measures (CMs) are used. The researchers suggest that the best results are obtained by using MFCCs.

(Zhao etal , 2011) Was proposed an English lexical stress detection approach using acoustic features. The approach done through by classifies the vowels of English words into two patterns: primary stress and unstressed, the research is done at the phone level by help of SVM learning algorithm, which used stressed ,duration the loudness of the speech.

The other work for designing of the model of emotional speech detection, which pre-process speech signal and detect or identify the speaker filling by considering the acoustic features (Demircan and Kahramanli, 2014; Rui, 2014). And also Speech recognition and voice/

unvoiced/silence detection from speech signals are related work for the model (Mohammed A. , 2012; Felber, 2001; Murphy, 2014).

There is a limited research have done in the area of computational linguistics approach in the relation of semantic language like Gəʾəz. In fact, a few theses are conducted in this area, (Muluken, 2007; Desta, 2010; Yitayal, 2014). According to Desta's thesis, a morphological analyser for geez verbs contributed to the developments of full-fledged NLP application. For Gəʾəz language Desta's thesis is used to analysis morph-syntactic features including affixes together with their syntactical functions. Those features subjects and object along with their person, gender-number features, and tense-mood and stem type of the verb. The author achieved the research objective using rule-based approaches specifically CV-based and Two-Level Morphology (TLM) are adopted to design the model.

Yitayal (Yitayal, 2014) have done on develop of a model for morphological analyser of Gəʾəz verbs using memory-based learning approach which is the analysis phase performs instance making by extracting features from the given text to have similar structure of features during comparison. Then the extracted features are accepted to the morpheme identification process to be compared with individual instances in memory, and stems are extracted with their morpheme functions. After that, the roots are extracted from the stems. The author's said that, the study for role of derivational morphology allows existing words to be used as the base for forming new words with different meanings and different functionality, and for the role of inflectional morphology deals with syntactic feature of the language such as person, number, gender, tense, case, and degree. These thesis shows that, components of NLP can be applied in Gəʾəz language for development of a full-fledged NLP application.

The other researcher that have done on Gəʾəz Verb Classification into heads and troops or bases and their derivation (Muluken, 2007). Muluken's research are conducted on base of traditional Ethiopian scholars of three centres of Qane, and foreign and modern scholars of linguistic. His research said that main criteria of classification where the form of letters in verbs. If two verbs are similar in their pattern of letters in the conjugation, they are classified under one head. According to Muluken's Traditional Ethiopian scholar's verbs classify on the basis of criteria such, as germination non-germination, number and position of radicals as well as the positions of gutturals and semi-vowels in written verbs and the conjugation pattern of the verbs. Classification by foreign scholars is based the stem vowel, and the root consonants and classification of modern linguistic are based on root consonants and patterns of vowel in

the formation. The study showed that Gəˀəz verbs or words can classify or identify based on the properties of letters or word by itself. In this study, we identifying Gəˀəz words pronunciation style based on the features of word or Gəˀəz alphabets. And to develop a pronunciation detection we used the most five common acoustic features of the speech signal those are energy, duration, pitch, formants and MFCC at word level. As we seen in the literature most literature ate used at phoneme level and some additional acoustic features, which does not use the acoustic features that used in this study. Additionally, language has its own systematic ways that govern the pronunciation of language, word formation, and grammatical correction and son on. Therefore, designing an approach for pronunciation detection for Gəˀəz language is a relevant components of NLP.

CHAPTER THREE

Gəʾəz Language Pronunciation Features

3.1. Overview

In this chapter, we discuss about an aspect of Gəʾəz language and its features of pronunciation styles. To detect the pronunciation styles of Gəʾəz words, we use several techniques. First, we are identifying the pronunciation style, then we select relevant features of the pronunciation style and in its. After that, selection of acoustic features and extraction of the acoustic feature are the main method to detection the pronunciation style. Therefore, this chapter including writing system of Gəʾəz language, alphabets of Gəʾəz language, Formation of Gəʾəz Words, types of Gəʾəz language pronunciation and Features of Gəʾəz language pronunciation, and acoustic features extraction for speech signals the pronunciation.

3.2. Gəʾəz Writing System

Gəʾəz (ግዕዝ) also known as Ethiopic is a script used as an abugida (syllable alphabet) for several languages of Ethiopia and Eritrea. It originated as an abjad (ኣበገደ) consonant-only alphabet and is used in Ge'ez writing system, still now, Ge'ez language is used in Ethiopian Orthodox Tewahedo Church and the Eritrean Orthodox Tewahedo Church. In Amharic and Tigrinya, the script is often called Fidel (ፊደል), meaning "script" or "alphabet".

The evolution of Gəʾəz language writing system can be seen most clearly in evidence from inscriptions. By the first centuries, what is called "Old Ethiopic" or the "Old Ge'ez alphabet" arose, an abgd writing system is write right-to-left with letters basically identical to the first-order forms of the modern vocalized alphabet (e.g. for all category of letter ‘ሠ/mäʾ it taken only the first order ‘ሠ/mäʾ). Vocalization of Ge'ez occurred in the 4th century, and though the first completely vocalized texts known are inscriptions by Ezana, vocalized letters pre-exist him by some years, as an individual vocalized letter exists in a coin of his predecessor Wazeba.

According to the beliefs of the Ethiopian and Eritrean Orthodox Tewahedo Church scholars, the original consonantal form of the Ge'ez fidel is divinely revealed to Henos "as an instrument for codifying the laws", and the present system of vocalisation is attributed to a team of Aksumite scholars led by Frumentius (Abba Selama), He became the first Abune a title given to the head of the Ethiopian Church. the same missionary said to have converted the king Ezana to Christianity in the 4th century AD. It has been argued that the vowel marking pattern of the

script reflects a South Asian system, such as would have been known by Frumentius. The first modification performed by Abba Fremɨnatus, which is by changing writing system of Gəʾəz language, from right–left to left-right, from ABGD (አበገደ) to HLHM (ሀለሐመ) order. The Gəʾəz alphabet only the first order that is called Gəʾəz that has 26 consonantal letters. But it increases the number of alphabet to 182 by adding of other six vowels letters or alphabet with 26 consonantal letters (በላይ, 2007 EC; Kidanewold, 1934 EC).

A study of the Gəʾəz writing systems is essential to understanding the history of Ethiopia and the evolution and modern usage of the language. Among from the language of ethiopia,Gəʾəz language has its own alphabets. However other languages like Amharic and Tigrigna adopted their alphabets fully from Gəʾəz (Bender, 1976.; Dillmann, 1899). An alphabet or Fidel of a language represents its sound. They are also called hohyat /ሆህያት in Amharic/. Ge'ez sounds can be studied by dividing them into simple-sounds and complex-sounds (zeradawit, 1996). Simple-sounds are represented with 182 alphabets. From these, seven of them represent vowel sounds: አ/ə/, ኡ/u/, ኢ/i/, ኣ/a/, ኤ/e/, እ/ɨ / and ኦ/o/. Whereas, others represent consonant sounds. As we seen the evolution of Gəʾəz writing system, After the modification, every consonant is combined with other six vowel sound alphabets to produce six more derived alphabets (zeradawit, 1996). For example, the combination of the original alphabet ሀ/ha/ with the six vowel sound alphabets ኡ/u/, ኢ/i/, ኣ/a/, ኤ/e/, እ/i/ and ኦ/o/ yields six more derived alphabets as: ሁ/*hu*/, ሂ/*hi*/, ሃ/*ha*/, ሄ/*he*/, ህ/*h*/ and ሆ/*ho*/ respectively.

Generally, Ge'ez has essentially 26 main alphabets, all consonants, and each with six more derivations, while the rest are essentially those with additional strokes and modifications added on to the main forms to indicate a vowel sound associated with it or to make aural adjustments in the basic consonant sound (Scelta, 2001). Hence, it has a total of 182 alphabets, with twenty-six alphabets and seven vowels (26×7 = 182 alphabets). Every twenty-six alphabets, a matrix of 26×7 size is produced. Each of the columns are labelled as ግእዝ /Ge'ez/ (first order), ካእብ /*kaib*/ (second-order), ሳልስ /*salɨs*/ (third-order), ራብእ /*rabɨ*/ (fourth-order), ሓምስ /*hamɨs*/ (fifth-order), ሳድስ /*sadɨs*/ (sixth-order), and ሳብእ /*sabɨ*/ (seventh-order) of alphabets. The orders represent the sound of each of the vowels. For example, if we take the verb "ሖረ/horä/" (he gone), the alphabet ሖ/*ho*/ is read as ሖኦ/ho/, where ኦ/o/ is a vowel with seven-order sound. Hence, it is sorted under the seventh column /seventh -order/ category. The complex-sounds are represented with in twenty-six letters those alphabets are four in number. These are: ኰ /*kwə*/, ጐ /*gwə*/, ቈ /*ḱwə*/, and ኾ /*xwə*/ each complex letter have five orders, which is only missing

the second and seventh order. Alphabets representing complex-sounds have only four derived alphabets, which are produced after the combination of the simple sound alphabets ከ /*kə*/, ገ /*gə*/, ቀ /*ḳə*/ and ኀ /*hə*/ with the vowels አ /*ə*/ and ኧ /*ɨ*/ and the semi-vowels ወ /*wə*/ and የ /*yə*/ in different patterns. For instance, the alphabets ጐ/gwo, ጒ /gwi, ጓ/gwa, ጔ /gwe, ጕ/gwu are derived due to the combination of the vowel and the semi-vowel sounds with the consonant ገ /*gə*/ E.g: ገ /*ge/* +አ /ə/ ዎ /*wo*/ = ጐ /*gwo*/. For more detail Gəʾəz alphabetsa describ below and alphabets adopted and used in this study and detailed discussion on Ge'ez writing system is found in (Dillmann, 1899); ደሴ፣ 2002e.c; Kidanewold, 1934; Zeradawt A, 1996e.c).

Table 1፡ The Gəʾəz Alphabet /Fidel/ (ባላይ፡ 2007e.c)

First order/Gəʾəz	Second-order/kaib	Third-order/ salis	Forth-order /rabe	Fivth-order/ hamus	Sixth-order/ sadis	Seventh-order/ sabe
ሀ (he)	ሁ (hu)	ሂ (hi)	ሃ (ha)	ሄ (hie)	ህ (h)	ሆ (ho)
ለ (le)	ሉ (lu)	ሊ (li)	ላ (la)	ሌ (lie)	ል (l)	ሎ (lo)
ሐ (h.e)	ሑ (h.u)	ሒ (h.i)	ሓ (h.a)	ሔ (h.ie)	ሕ (h.)	ሖ (h.o)
መ (me)	ሙ (mu)	ሚ (mi)	ማ (ma)	ሜ (mie)	ም (m)	ሞ (mo)
ሠ (sse)	ሡ (ssu)	ሢ (ssi)	ሣ (ssa)	ሤ (ssie	ሥ (ss)	ሦ (sso)
ረ (re)	ሩ (ru)	ሪ (ri)	ራ (ra)	ሬ (rie)	ር (r)	ሮ (ro)
ሰ (se)	ሱ (su)	ሲ (si)	ሳ (sa)	ሴ (sie)	ስ (s)	ሶ (so)
ቀ (qe)	ቁ (qu)	ቂ (qi)	ቃ (qa)	ቄ (qie)	ቅ (q)	ቆ (qo)
በ (be)	ቡ (bu)	ቢ (bi)	ባ (ba)	ቤ (bie)	ብ (b)	ቦ (bo)
ተ (te)	ቱ (tu)	ቲ (ti)	ታ (ta)	ቴ (tie)	ት (t)	ቶ (to)
ኀ (hhe)	ኁ (hhu)	ኂ (hhi)	ኃ (hha)	ኄ (hhie)	ኅ (hh)	ኆ (hho)
ነ (ne)	ኑ (nu)	ኒ (ni)	ና (na)	ኔ (nie)	ን (n)	ኖ (no)
አ (a)	ኡ (u)	ኢ (i)	ኣ (ea)	ኤ (eie)	እ (e)	ኦ (o)
ከ (ke)	ኩ (ku)	ኪ (ki)	ካ (ka)	ኬ (kie)	ክ (k)	ኮ (ko)
ወ (we)	ዉ (wu)	ዊ (wi)	ዋ (wa)	ዌ (wie)	ው (w)	ዎ (wo)
ዐ (eee)	ዑ (eeu)	ዒ (eei)	ዓ (eea)	ዔ (eeie)	ዕ (ee)	ዖ(eeo)
ዘ (ze)	ዙ (zu)	ዚ (zi)	ዛ (za)	ዜ (zie)	ዝ (z)	ዞ (zo)
የ (ye)	ዩ (yu)	ዪ (yi)	ያ (ya)	ዬ (yie)	ይ (y)	ዮ (yo)
ደ (de)	ዱ (du)	ዲ (di)	ዳ (da)	ዴ (die)	ድ (d)	ዶ (do)
ገ (ge)	ጉ (gu)	ጊ (gi)	ጋ (ga)	ጌ (gie)	ግ (g)	ጎ (go)
ጠ (t.e)	ጡ (t.u)	ጢ (t.i)	ጣ (t.a)	ጤ (t.ie)	ጥ (t.)	ጦ (t.o)
ጰ (p.e)	ጱ (p.u)	ጲ (p.i)	ጳ (p.a)	ጴ (p.ie)	ጵ (p.)	ጶ (p.o)
ጸ (tse)	ጹ (tsu)	ጺ (tsi)	ጻ (tsa)	ጼ (tsie)	ጽ (ts)	ጾ (tso)
ፀ (tze)	ፁ (tzu)	ፂ (tzi)	ፃ (tza)	ፄ (tzie)	ፅ (tz)	ፆ (tzo)
ፈ (fe)	ፉ (fu)	ፊ (fi)	ፋ (fa)	ፌ (fie)	ፍ (f)	ፎ (fo)
ፐ (pe)	ፑ (pu)	ፒ (pi)	ፓ (pa)	ፔ (pie)	ፕ (p)	ፖ (po)

First order /Gəʾəz	Third order /salis	Forth order/rabe	Fith order/ hamis	Sixth order /sadis
ቈ (que)	ቊ (qui)	ቋ (qua)	ቌ (quie)	ቍ (qu')
ኈ (hhue)	ኊ (hhui)	ኋ (hhua)	ኌ (hhuie)	ኍ (hhu')
ኰ (kue)	ኲ (kui)	ኳ (kua)	ኴ (kuie)	ኵ (ku')
ጐ(gue)	ጒ (gui)	ጓ(gua)	ጔ (guie)	ጕ (gu')

3.3. Formation of Gəʾəz Words

As we describe in section 3.2, Gəʾəz writing system has its own writing system, which is a syllabic system. There are 26 consonant alphabets or fidels with seven variations, symbol, that are according to the vowel that is coupled with the consonant. Each letter or symbol usually represents a whole syllable. Combination of two and more than two consonants alphabets can formulate Gəʾəz words. In Gəʾəz language there are seven POS, these are noun, verb, adjective, preposition, conjunction, adverb, and pronouns (በላይ, 2007 e.c). In this thesis, to detect pronunciation styles of the word, we use all POS of the language but the word is a single word and in verb category, only the past or perfective tense /ቀዳማይ/ታላፊ from ዐቢይ አንቀጽ/ ʾäbiyə ʾänəqäs'ə of the main verbs are used. Additionally, numerical words (that describes numerals) are also included..

3.4. Types of Pronunciation in Gəʾəz Language

Pronunciation is a skill of Gəʾəz language, which is one of the most basic components of a language to understand the correct meaning of words, phrases, and sentence. Pronunciation of the language is different from one language to another language. As we describe in chapter one in the methodology section, the styles or types of pronunciation are detected by the model. Generally, Gəʾəz language follows its own pronunciation styles. Each style has its own different pronunciation ways or features to pronounce a word. Therefore, designing automatic pronunciation detection are depend on this features. According to different scholar and Gəʾəz grammar books, the pronunciation style of Gəʾəz language are grouped into two broad categories, those are major pronunciation and minor pronunciation (ያሬድ፣1997፤በላይ, 2007 e.c፤ አፈወርቅ፣2005፤ደሴ፣2007). To identify the features of Gəʾəz language pronunciation styles, we reviewed different Gəʾəz books. Which, written by Ethiopian scholars and foreign linguistic professional to identify the properties or features of Gəʾəz word, and characteristics of the pronunciation style. It is describing Gəʾəz verb formulation, classification, writing system, Gəʾəz language pronunciation and any properties of part of speech of Gəʾəz language. Those are the book of the verb and grammar /Matsaffe giss wasewasew by (ያሬድ፣ 1997e.c; አፈወርቅ ፣ 2005; በላይ፣ 2007; Kidanawald 1948; ደሴ፣ 2007; ፍቅሬ 2008; (Dillmann, 1899; Leslau, 1987).

The major pronunciation style are (ሲሳይ, 2005)**:**

I. ተነሽ ንባብ/tänäšə nəbabə (rising up reading): - is a style of pronunciation in Gəʾəz language, which read in the form of aggressive speeches that are exerts more force or speech with the highest sound.
II. ወዳቂ ንባብ/Wädaqi nəbabə (folding reading): - is a style of pronunciation in Gəʾəz language, which read or pronounce in the form of politely and slowly which are exerts less force and seized the last letter.
III. ሰያፍ ንባብ/säyafə nəbabə (Slant reading): - is a style of pronunciation in Gəʾəz language, which is the read or pronounce in the form of aggressive speeches that are exerts more force on preceding of last letters. Which is similar to ተነሽ/tänäšə nəbabə.
IV. ተጣይ ንባብ/tät'ayə nəbabə (Throwing reading): - is a style of pronunciation in Gəʾəz language, which read or pronounce in the form of slowly and politely speech that are exerts less force and seized predecessor of the last letter.

The minor pronunciation style: Minor pronunciation styles are obtained from the major pronunciation style of the language, which shows the way of pronunciation of words in a sentence. Minor pronunciation styles are more than eight. However, most scholar are agreed upon the following pronunciation styles (ሲሳይ, 2005).

I. ዐራፊ/ቀዋሚ/ʾärafi/qäwami፡ this pronunciation type is a way of saying words without engaging words, which reads separately with gab in between words.
II. ተናባቢ/tänababi፡ this type of pronunciation is read two and more than two words at the same time. However, the second or the third words depend on the last letter of the previous word.
III. ቆጣሪ/qot'ari፡ this style of pronunciation, which pronounces by counting each letter of a word.
IV. ጠቅላይ/t'äqəlayə: this type of pronunciation in the form of voiceless, which is some voice of letters in the word are voiceless.
V. ጠባቂ/t'äbaqi/Geminated፡ in most case this type of pronunciation is read by stressing or hold ing the middle letters of a word.
VI. ልህሉህ/ləhəluhə: this type of pronunciation is the opposite of ጠባቂ/t'äbaqi/Geminated, which reads the word by unstressed middle letters
VII. ጎራጅ/gorağə፡ this type of pronunciation is read by left the letters from the beginning or the middle or the last in the word.

VIII. ጥያቄያዊ/t'əyaqeyawi: this type of pronunciation is the form of questioner pronunciation which uses special letters to read the words.

IX. ትርአስ ንባት/tərəʾäsə nəbabatə፡ this type of pronunciation is used to change the pronunciation style of words from ተነሽ ንባብ/tänäšə to ወዳቂ ንባብ/Wädaqi and vice versa by adding special letters.

In this thesis, we studied only the four-major pronunciation style of the language. The minor pronunciation is obtained from the major pronunciation styles, which is under major pronunciation style. Therefore, the minor pronunciation styles are left as a future work. These major types of pronunciation styles have its own features or attributes. In this section, we describe for each feature of Gəʾəz language pronunciation styles features

3.4.1. Features of Rising up Reading/ተነሽ ንባብ/tänäšə nəbabə

ተነሽ ንባብ/tänäšə nəbabə or rising up reading is one of Gəʾəz language pronunciation style. As we study ተነሽ ንባብ/tänäšə nəbabə has the following feature:

I. Which reads in the form of agressive with high loundeness.
II. More energy is required to pronounce ተነሽ ንባብ/tänäšə nəbabə words
III. Short duration of speech signals
IV. Low fundamental frequency or pitch when compare to others pronunciation style

Table 2: sample Gəʾəz words under ተነሽ/tänäša pronunciation style(ሰላይ, 2007)

Gəʾəz word	Amharic	English	words	Amharic	English
ሀለ	አወደሰ	Praised	ረቀቀ	ረቀቀ/ቀጠነ	grew thin
ሐለመ	አለመ	Dreamed	ረበ	ዘረጋ/ወጠረ	stretched
ሀለለ	በራ/ተቃጠለ	Burn	ረከበ	አገኘ/ደረሰ	found
ሐለቀ	ክብ አደረገ	Make it circle	ሰለበ	ቆረጠ	cutting
ሀለወ	ኖረ/ተገኘ	Lived	ሰለተ	ነፋ/ለየ	separated/blew
ሐለበ	አለበ	Milked	ሰለጠ	ጨረሰ	finished
ለሀየ	ተጫወተ	Play	ከሰተ	ገለጠ/አብራራ	explained
ሐለተ	መለመለ	Selected	ለሊነ	እኛ	We
ሀበበ	ጨመቀ	Compress	ለልየ	እኔ	Me
ለሐመ	ለሰለሰ	Became soft	ለሊከ	አንተ	You
ሀበወ	ዘነበ	Rained	ሐልዘዘ	አፈራ	Give frut
ለሐቀ	አጣበቀ	Glued	ሐልየነ	ጉቦ ሰጠ	Bribe
ሐመመ	ተመቀኘ/ነፈገ	miserly/stingy	ሀብለየ	በዘበዘ	Plundered
ሀከከ	አሸበረ	terrorized	ሀውለየ	ነቀፈ/አፌዘ	Criticized
ለመደ	ለመደ	became accustomed	ልሕየ	ወዛ	Transpired
መሀረ	አስተማረ	taught	ሎሀ	ጻፈ	Wrote
መለጠ	ላጠ	peeled	ሄደ	ቀማ	Grabbed
መሐለ	ተረገመ/ማለ	oath	ሆበ	ዞረ	Rotated
መሠጠ	ነጠቀ/ቀማ	grabbed	ለብሰ	ተሸፈነ	Covered
መነዘረ	ተለዋወጠ	changed	መርሐ	መራ/አሳየ	Led
መአተ	በዛ/ፈደፈደ	multiplied	መጽአ	መጣ	Came
መከረ	ፈተነ	examined	መድቅሐ	ሠራ/ገነባ	Built
መየነ	ከዳ/ተጠበበ	preoccupied	ሠብሐ	ወፈረ	Became think
መደበ	መደበ	allotted	ሜመ	ሾመ	Gave rank
መደደ	መደመደ	made level	ሥዕየ	በተነ	Scattered
ሠበጠ	በላ/ጐረሰ	ate	ርሕሰ	ራሰ	Got wet
ሠተረ	ቀደደ	tore/cut out	ርእየ	አየ	Saw
ሠተነ	ተወለደ	engendered	ርዕደ	ፈራ	Was afraid
ሠነቀ	ስንቅ ያዘ	provisions	ሮሐ	ረጨ	Sprinkled
ሠነየ	አማረ/ተዋበ	beautified	ሮጸ	ሮጠ	Ran
ሠአነ	ተጫማ/ለበሰ	put shoes	ሰምዐ	ሰማ	Heard
ሠከረ	ተከራየ	rented	ነፅረረ	ከበደ	Was heavy
ሠዐለ	ሣለ	painted	ተኬሰ	ተሸከመ	Carried
ሠገረ	ፖሊስ	police	ከብረወ	መታ/ጐሰመ	Struck/hit
ሠገወ	በቀለ/አደገ	germinated	ዜነወ	ነገረ/አወራ	Narrated
ሠጠቀ	ሰነጠቀ/ከፈለ	split	ዳሕመመ	ቆፈረ	Dug
ሠጠየ	ተሳበ	pulled	ጼጸየ	አለፋ	Had tailored
ሠጠጠ	ቀደደ	tore			

3.4.2. Features of Folding down Reading/ወዳቂ ንባብ/wädaqi nəbabə

It is one of the style of pronunciation of Gəʾəz language, as we study from the experiment and different literature ወዳቂ/wädaqi nəbabə has the following feature: Which is read in the form of slowly and politely

I. Less energy is required to pronounce ወዳቂ/wädaqi nəbabə words which is low sounds
II. It keeping up or stressed on the last speech signals
III. Long duration speech rate of speech signals
IV. High formant of acoustic or speech signal
V. high fundamental frequency or pitch when compare than others pronunciation style

Table 3፡ sample Gəʾəz words under ወዳቂ/wädaqi pronunciation style (በላይ, 2007))

Words	Amharic	English	Words	Amharic	English
ሐለስትዮ	ጦጣ	Monkey	ሰናዲ	አጠናቃሪ	Complete
ህልውና	መኖር	to live	ሰናፔ	ሰናፍጭ	mustard
ሀሎ	ኖረ/አለ	He lived	ሰንቃዊ	የሚመታ	who Hit
ሀርበዲ	ቅምጥል	Pleasant	ቀናዪ	አሸናፊ	winner
ሐላዩ	ዐሳቢ	Considerable	ቅንጻዌ	መዝለል	jump
ሕሊና	ሀሳብ/ምኞት	Wish	ቀንጦሳጢ	ሰልፈኛ	Warrior
ሀባቢ	ጭማቂ	Juice	ቅያሜ	መቀየም	Rancour
ህቦ	ወጨፎ/ውሽንፍር	Rusty	ቀዳማዊ	ቀዳሚ	Previous
ሀቢ	ዋስ/መያዣ	Warranty	ቅዳሴ	መቀደስ	Sanctification
ሐማሚ	የሚመቀኝ	I'm scared	ቀዳዊ	የሚያምር	Beautiful
መናኒ	መናኝ,----------	Bread	ባሲልያ	ንጉሥነት	Kingship
መና	ዳቦ/እንጀራ	Duality	ባቡቴ	ድፎ ዳቦ	Dough bread
መንታዌ	መንታነት/ሁለትነት	Big bang	ባውላ	ቡድን	group
መንኩቢያ	አውራ እጣት	Problem	ባዜቃ	መብራት	Lights
ምንዳቤ	ችግር	Drowning	ባዝራ	እንስት ፈረስ	Female horse
መንደልቶ	ጭራሮ		ቤዛ	ካሳ	compensation
መንደቢ	አስጨናቂ	Disturbing	አርማሚ	ሚዛን	Balance
መንድያ	ቀሚስ	Dress	አርዌ	አራዊት	Beasts
ሰብዑ	ሰብአቶች	Multiple seven	ኪታቢ	ጻፊ	secretary
ሱባዔ	ሰባት ቀን	Seven days	ከናሲ	ሰብሳቢ	Collector
ሳብዓዊ	ሰባተኛው	Seventh	ኰናኔ	መቀጣት	Punishment
ሰቦ	እበት	Dung	ውሳጤ	ውስጥ	interior
ሰታሪ	የሚዘርር	Spreading	ወቃሪ	ቀራጭ	tax collector
ሰታዩ	ጠጭ	Drinker	ውቅሮ	ዋሻ	Cave
ስቴ	መጠጥ/አልኮል	Beverage	ዐመፃ	በደል/ግፍ	Abuse /violence
ሴፈላ	ሸለቆ	Valley	ዝማሬ	መዘመር	Sing Calm
			ጽማዌ	እርጋታ	

3.4.3. Features of Slant Reading /ሰያፍ ንባብ /säyafə nəbabə

ሰያፍ/säyafə nəbabə or slant reading is one of Gəʾəz language pronunciation style it is more similar to ተነሽ ንባብ/tänäšə, as we studied from the experiment and different literature ሰያፍ/säyafə nəbabə has the following feature, which is read in the form of slowly and politely

I. Which reads in the form of agressive with high loundeness.
II. More energy is required to pronounce ሰያፍ/säyafə nəbabə words next to ተነሽ ንባብ/tänäšə
III. Short duration of speech signals which is greater than from ተነሽ ንባብ/tänäšə, however it affected by the number of letters in a word
IV. high fundamental frequency or pitch next to ወዳቂ/wädaqi when compare to others pronunciation style

Note: mostly in Gəʾəz language proper nouns or nouns are ሰያፍ/säyafə /slant reading pronunciation style. This proper nouns are obtained from different language. Therefore pronounce such type of words may not be perfect bay the speakers or native. As we tagged Gəʾəz words manually most of the words that are tagged as ሰያፍ/säyafə/slant reading style are proper nouns. Those words have some effects on the model.

Table 4፡ sample Gəʾəz words under ሰያፍ/säyafə pronunciation style (በላይ, 2007)

Words ሰያፍ					
ሄሮድያኖስ	ሽመላ	Stork	ሮዳስ	ሮዳስ	Reubeለ
ለንዴዎን	መወልወያ	Whack	ቂርቆስ	ቂርቆስ	Rods
ሌጌዎን	ሠራዊት	Force /armiy	ቄድሮስ	ቄድሮስ	Kirkos
ሔሴሜት	ድመት	Cat	ቆጵሮስ	ፍግ	Kidros
ሕዝቅኤል	የሰወ ስም	Ezekiel	ቆባር	ትብያ	Love
ማርቆስ	ሰልፈኛ	Markos	ቆጵን	ኩፍ ጫማ	Story
ማቴዎስ	የሰወ ስም	Matthew	ታቦር	ተላቅ ተራራ	Kum shoes
ሰሎሞን	ሰላማዊ	Solomon	ትያጥሮን	የጫፍራ ቤት	A tall mountain
ሲኖዶስ	ጉባኤ	Conference	ቶማስ	ቶማስ	Dockyard
ሲኦል	ሲኦል(ነባ)	Hell	ቶሬን	የመርከብ ምሰሶ	Thomas (S.)
ሳሙኤል	ሳሙኤል	Samuel		እንቁ	Shipwreck
ሜሎስ	የሚያንፀባርቅ	Reflective	ሰንፔር	አቤል	Pearl
ገብርኤል	የመልአክ ስም	Gabriel (Angel name)	አቤሜሌክ	አቤሜሌክ	Abel
ራምኖን	ዶግ	Dog	ጽዮን	የዳዊት ከተማ	Ebed-melech
ራጉኤል	የመልአክ ስም	Raguel)	ፊልሞን	ስውር ቦታ	melech David's City
ሮቤል	ሮቤል				Hidden place

3.4.4. Features of Throwing Reading (ተጣይ ንባብ/tät'ayə nəbabə)

It is one of the style of pronunciation of Gəʾəz language, as we study from the experiment and different literature ተጣይ ንባብ/tät'ayə nəbabə has the following feature, which is read in the form of slowly and politely.

I. Less energy is required to pronounce ተጣይ ንባብ/tät'ayə nəbabə words, which is low sounds
II. It keeping up or stressed on the last speech signals
III. Long duration speech rate of speech signals
IV. High formant of acoustic or speech signal next to ወዳቂ/wädaqi nəbabə
V. high fundamental frequency or pitch next to ወዳቂ/wädaqi nəbabə when compare than others pronunciation style

However, the pronunciation style of ተጣይ ንባብ/tät'ayə nəbabə and ሰያፍ/säyafə nəbabə are both end with 6th order of the letters, which is similar sound the end of speech signal, therefore the acoustic features are might be interchangeably.

Table 5: sample Gəʾəz words under ተጣይ/tät'ayə pronunciation style (ቦላይ, 2007)

Gəʾəz words	Amharic meaning	English	Gəʾəz words	Amharic meaning	English
ለሊክን	እናንተ	You	አንስት	ሴት	Woman
ማኅፈር	ወዳጅ	Friend	አንቀጽ	በር/ደጃፍ	Door / Door
ልሁም	መጭመቂያ	Press	አንቃለል	ጭቃ	Mud
ሐልክ	የደቀቀ	Crushed	ከርካዕ	ቀርካሐ	Too bad
ሕልያን	ጉቦ	Bribery	ከርደል	ድንክ	Thumbnail
ህብለት	የአንገት ጌጥ	Necklace	ወርቅ	ወርቅ	Gold
ሕሙም	በሽተኛ	Patient	ወርኀ	ጊዜ	Time
መታግር	እግር(የእቃ)	Foot	ዐራት	አልጋ	Bed
ምቱሕ	ስግብግብ	Greedy	ዕርቃን	ባዶነት	Empty
ምትን	ጅማት	Nails	ዐረቅ	ላበት	Play for free
ርቱዕ	ቀጥ ያለ	Straight	ዕርበት	ምሽት	evening
ርካብ	መውጫ	Exit	እግዚአብሔር	ጌታ	Lord
ሰት	ቄጠማ	Recluse	ግዝአት	መግዛት	to buy
ቀንዲል	ማብርያ	Mirror	ጥቅዕ	አጠገብ	Near
ቀውስ	ቀሰት	Come down	ጥብ	ጡት	Breastfeeding
ብእሲት	ሚሰት	MISSION It's	ጥነፍ	ዙርያ	Round
ብውሕ	ይገባል	okay	ፈውስ	መዳን	Salvation
ባዕል	ሀታም	Hey	ፍውስ	የዳነ	He got saved
ነባር	ቋሚ	Permanent	ፈያት	ቀማኛ	Tired
ነባቢት	ተናጋሪ	Speaker			
ንባብ	ንግግር	Speech			

3.5. Feature Selection for Pronunciation Detection

As we describe in section3.4, Gəˀəz language pronunciation has its own features in each style. That means the ways of the pronunciation of each style is different from one pronunciation to the other pronunciation style. When we say pronunciation means it is the way you say a word or the way in which a language is usually spoken. To select the relevant features of each pronunciation style, we have reviewed related works and discuss the Gəˀəz scholars. As we describe in the literature review section, the features for emotional speech recognition are more related to in this work, because of when we pronounce Gəˀəz language it has different energy, speech rate, and the loudness of the audio signal. Based on this, we have selected some relevant features of sound signals or acoustic features. There are five main features that are used in this thesis to feed into the classifier of the neural network. Each feature is extracted from the audio file signal, and the variance and mean for each feature over these blocks is found for single labeled Gəˀəz word. The acoustic features that are used in this thesis are pitch, speech rate, energy, formants and MFCC (Mel frequency cepstral coefficient). From these features, we have used some statistical methods those are mean, maximum, minimum, and variance of pitch, mean, maximum, and variance of formant and the mean and variance of MFCC. So, we describe those features extraction for each pronunciation style.

3.6. Feature Extraction

Feature extraction is the process by which the measurements of the given audio input can be taken to differentiate among the pronunciation style. This process is a fundamental requirement of any speech recognition system. It is the mathematical representation of the speech file (Milwaukee, 2007; Vijayalakshmi and Leema, 201`4). The primary goal of feature extraction is to simplify recognition and representation of the speech data by summarizing and obtaining the acoustic properties. In Gəˀəz language pronunciation detection, the goal is to classify the speech signals of Gəˀəz words into their pronunciation style using a relevant representation of acoustic features. In this thesis, when design the model of pronunciation detection, different speech features are used. The integrated features are grouped into the following five categories: Energy of the signal, pitch frequency detection, other time domain features or speech rate, formant detection, and spectral feature features that is MFCC. The graphical representation of some of the acoustic features that are used in this thesis are shown below diagram:

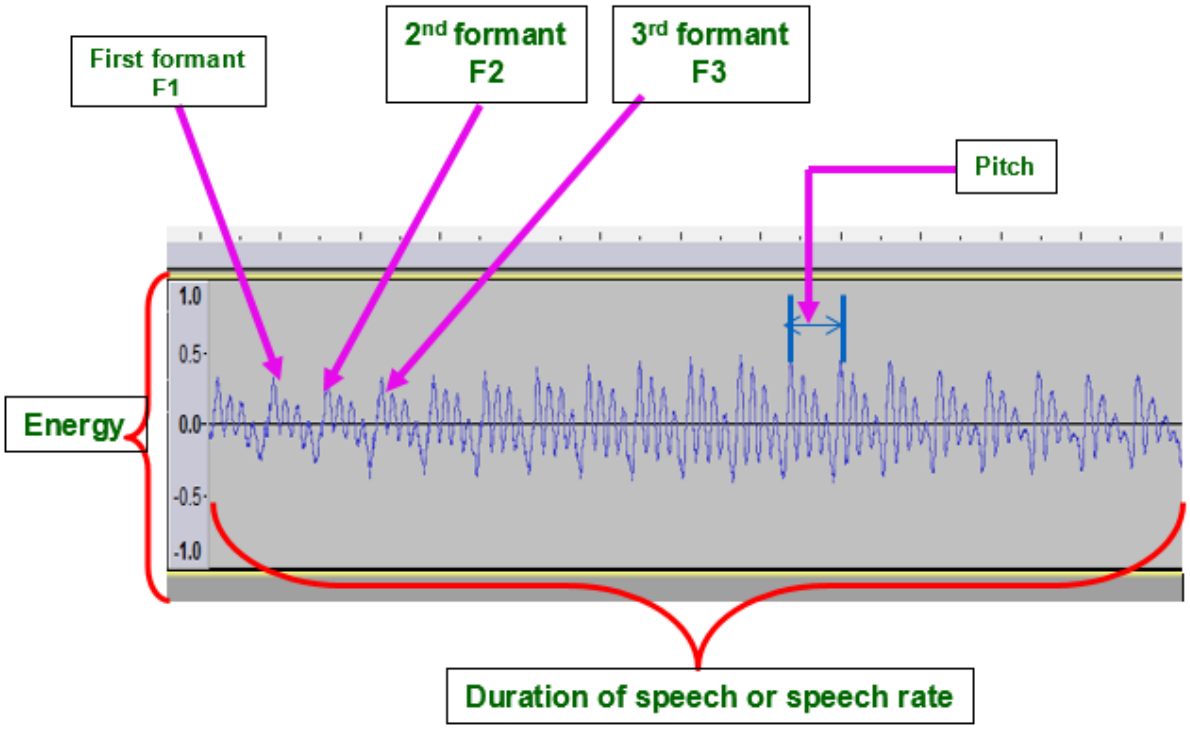

Figure 2: Graphical representation of Acoustic features for Gəʾəz word ሀለሜ/häläwä.

3.6.1. Energy

Energy is a basic feature in speech signal process, and it plays an important role in Gəʾəz language pronunciation detection, because of in Gəʾəz language pronunciation had used different energy in each style. Let us see an example that required more energy from Gəʾəz language pronunciation style, speech signals of ተነሽ/tänäšə /rising up reading and ሰያፍ/säyafə/slant reading styles are required more energy but speech signals of ወዳቂ/wädaqi /folding reading and ተጣይ/tät'ayə styles are required are less energy are used when compare to the first two pronunciation style. The signal energy, which is highly reduced the effect of noise that energy of speech signal could simply is calculated using Eq. (2) after normalization (Rui, 2014)

$$E = \sum_{i=1}^{N} x(n)2 \quad \text{..... (2)}$$

Where E stands for energy and x (n) is the signal sequence after sampling, N is sequence length. Each sampled point will be multiplied by itself and added up for the overall signal.

3.6.2. Pitch

Pitch refers the quality of a sound governed by the rate of vibrations producing it; which is the degree of highness or lowness of tone. This feature represents the vibration rate of audio signals, which can be represented by the fundamental frequency, or equivalently, the reciprocal of the fundamental period of voiced audio signals. It correlates to the vibration frequency of your vocal cord. A high pitch corresponds to a high vibration frequency. It is the most frequently used feature in speech analysis (Rui, 2014).

Pitch tracking **is** used to identify the pitch from the stream of audio signal, then we need to have some reliable methods for such pitch tracking. In general, pitch tracking is a fundamental step toward other important tasks for audio signal processing. Related research on pitch tracking has been going on for decades, and it still remains a hot topic in the literature (Rui, 2014). Therefore, we need to identify the basic concept of pitch tracking steps for other advanced audio processing techniques. Pitch tracking follows the general pre-processing of short-term analysis for audio signals. Pitch tends to low for ተነሽ/tänäšə and ሰያፍ/säyafə, too high for ተጣይ/tät'ayə and ወዳቂ/wädaqi.

There are many ways to estimate pitch value from a speech signal. In this thesis, time-domain autocorrelation method is used for it is a commonly used method and is easy to practice. The method uses short-term analysis technique to maintain characteristic for each frame, which means a pre-process should be fully applied before extract pitch. Since autocorrelation can decide the period of a periodic signal, for each frame apply the autocorrelation by using Eq. (3) (Rui, 2014)

$$A(k) = \sum_{i=1}^{FL} x(m)x(m+k) \quad \text{..... (3)}$$

Where FL is frame length, x (m) is the signal frame and k is shift parameter and A (k) represents t5he result of autocorrelation.

And also, the pitch is extracted from one of the voice records by using the windowing technique. The mean, minimum, maximum and variance of the pitch value are used for single word and these features are taken as the input of the classifier to detect the pronunciation style.

3.6.3. Speech rate

Speech rate is representing of how fast the pronunciation of Gəʾəz language words. It has strong connection with to detect the pronunciation style of Gəʾəz words like ተነሽ/tänäšə, ሰያፍ/säyafə, ተጣይ/täṭʾayə, ወዳቂ/wädaqi, feature is extracted from Eq. (4) (Rui, 2014)

$$Sr = TD/nw \quad \text{..... (4)}$$

Where Sr is speech rate, TD represented total duration of voice part and 'nw' is the number of words used but in this thesis, we focus identifying pronunciation style of at the single word level. Therefore, the number of word is one. By studying waveform of the recorded speech signal in a time domain (after normalization) as shown in Figure 2. Duration of the voice part could be calculate using MATLAB function to find the points that are bigger than zero and then divide by the sampling frequency.

3.6.4. Formant

The feature formant is referring or correlates to accents, which mean a distinct mode of pronunciation of a language, especially one associated with a particular nation, locality, social class, gender, sex. Speech formant can be used for emotional recognition, sex discrimination or diagnosing different neurological diseases (Belean, 2013)**.** The formants are defined as the spectral peaks of the sound spectrum of the voice. A formant is also used to mean an acoustic resonance and, in speech science and phonetics, a resonance of the human vocal tract (Lindblom etal, 2010). It is often measured as an amplitude peak in the frequency spectrum of the sound, using a spectrogram or a spectrum analyser, the frequency of the resonance may lie between the widely spread harmonics and hence no peak is visible.

In this thesis, formant is taken as one of the feature of Gəʾəz word pronunciation, because of Gəʾəz languages had different accents to pronounce a word in each style of the pronunciation. So, such kind of feature can identify by using the acoustic feature of formants. To extract this feature, we used formant detection methods. There are different methods for format detection. In this thesis, we have used a complex analysis of LPCC (linear predictive coding coefficient) formant detection. Linear predictive coding (LPC) is a tool used mostly in audio signal processing and speech processing for representing the spectral envelope of a digital signal of speech in a compressed form, using the information of a linear predictive model (Belean, 2013). The LPC as its name indicate that it predicts the current samples linear compensation of its past samples (Snell and Milinazzo, 1993). It is one of the most powerful speech analysis

techniques, and useful methods for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters (Belean, 2013)

LPCC method is used linear predictive coding (LPC) to detect formant. The two common pre-processing steps that can be applied to speech wave form before LPC, those are windowing and pre-emphasis (high-pass) filtering. The window in the speech segment using a Hamming window (Snell and Milinazzo, 1993). To obtain the LPCC, we specify the model order, by using the general rule, that is the order are two times the expected number of formants plus 2. Therefore, set the model order equal to 8 for three formants. Find the roots of the prediction polynomial returned by lpc command. Because the LPC coefficients are real-valued, the roots occur in complex conjugate pairs. Hold only the roots with one sign for the imaginary part and control the angles corresponding to the roots. Convert the angular frequencies in rad/sample represented by the angles to Hz and calculate the bandwidths of the formants. The bandwidths of the formants are represented by the distance of the prediction polynomial zeros from the unit circle. The frequencies greater than 90 Hz with bandwidths less than 400 Hz represent formants (Loizou, 1999).

3.6.5. MFCC

In speech processing, the Mel-frequency cepstral (MFC) is a representation of the short-term power spectrum of a speech, based on a linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency. Mel-scale is used for estimating the human's perception scale. The Mel-scale is a logarithmic mapping from physical frequency to perceived frequency (Milwaukee, 2007).

Mel-frequency cepstral coefficients (MFCCs) are coefficients that are commonly made up an MFC. It is the resulting from the cepstral representation of the audio clip (a nonlinear spectrum). The difference between the cepstral and the MFC is, In MFC the frequency bands are equally spaced on the Mel scale, which approximates the human auditory system's response more closely than the linearly spaced frequency bands used in the normal cepstral. This frequency warping can allow for better representation of sound. In feature extraction techniques, MFCC is one of the most commonly functioning techniques. Since a speech signal varies over the time. It is more appropriate to analyse the signal in short time intervals where the signals more stationary. Because of these MFCC method is used as feature extraction technique (Karakas, 2010)**.** MFCCs are commonly derived as follows (Karakas, 2010)

- Take the Fourier transform of a signal.
- Map the power of the spectrum obtained above onto Mel scale, using triangular overlapping windows.
- Take the logs of the powers at each of the Mel frequencies.
- Use the discrete cosine transform of the list of Mel log powers, as if it is a signal.
- The MFCCs are the amplitude of the resulting spectrum.

So, to compute MFCC feature extraction of the speech signal we explain step by step in this section the formulas are obtained from (Demircan and Kahramanli, 2014; Karpagavalli and Chandra , 2016; Milwaukee, 2007)

Pre-emphasis: The first stage in MFCC feature extraction is to boost the amount of energy in the high frequencies. This pre-emphasis is done by using a filter, which is the speech signal s (n) is sent to a high-pass filter: A common form of the pre-emphasis filter is:

$s2(n) = s(n) - a * s(n-1)$ ..... (5)

Where s_2 (n) is the output signal and the value of 'a' is usually between 0.9 and 1.0. The z-transform of the filter is

$H(z) = 1 - a * z^{-1}$ ..... (6)

The goal of pre-emphasis is to compensate the high-frequency part that are suppressed during the sound production mechanism of humans or to increase the performance by removing unwanted voice like noise.

Frame blocking: The input of speech signal is segmented into a block of frames, which might have not more than 30Ms with optional overlap of 1/3~1/2 of the frame size. Usually, the frame size (in terms of sample points) is equal to power of two in order to facilitate the use of FFT. If we have the sample rate is 16 kHz and the frame size is 200 sample points, then the frame duration is 200/16000 = 0.0125 sec = 12.5 ms. Additional, if the overlap is 100 points, then the frame rate is 16000/ (200-100) = 160 frames per second.

Hamming windowing: To keep the continuity of the first and the last points in the frame, each frame has to be multiplied with a hamming window (to be detailed in the next step). If the signal in a frame is denoted by s (n), n = 0…N-1, then the signal after Hamming windowing is s (n)*w (n), where w (n) is the Hamming window defined by:

$w(n, a) = (1 - a) - a\cos(2pn/(N-1)), \quad 0 \leqq n \leqq N-1$ ..... (6)

Fast Fourier Transform (**FFT)**: Spectral analysis shows that different timbres in speech signals corresponds to different energy distribution over frequencies. Therefore, we usually perform FFT to obtain the magnitude frequency response of each frame. Which is convert each frame of N samples from the time domain into the frequency domain. The FFT is a fast algorithm to implement the Discrete Fourier Transform (DFT) (Karakas, 2010).

When we perform FFT on a frame, we consider that the signal within a frame is periodic, and continuous when wrapping around. If not, we can still perform FFT but the in continuity at the frame's first and last points is likely to introduce undesirable effects in the frequency response. To deal with this problem, we have two strategies:

a. Multiply each frame by a Hamming window to increase its continuity at the first and last points.
b. Take a frame of a variable size such that it always contains an integer multiple numbers of the fundamental periods of the speech signal.

The second strategy encounters difficulty in practice, since the identification of the fundamental period is not a trivial problem. Moreover, unvoiced sounds do not have a fundamental period at all. Consequently, we usually adopt the first strategy to multiply the frame by a Hamming window before performing FFT.

Without the use of a Hamming window, the discontinuity at the frame's first and last points will make the peak in the frequency response wider and less understandable. With the use of a Hamming, the peak is sharper and more distinct in the frequency response. But with the use of a Hamming window, the harmonics in the frequency response are much sharper.

In other words, the harmonic of the frequency response is generally caused by the repeating fundamental periods in the frame. However, we are more concerned in the envelope of the frequency response instead of the frequency response itself. To extract an envelope-like this feature, we use the triangular bandpass filters, as explained in the next step.

Triangular Bandpass Filters: multiplying the magnitude frequency response by a set of 20 triangular bandpass filters to get the log energy of each triangular bandpass filter. The positions of these filters are equally spaced along the Mel frequency, which is related to the common linear frequency f by the following equation:

$$\text{mel(f)} = 1125 * \ln(1 + \text{f}/700) \quad \text{..... (7)}$$

Mel-frequency is proportional to the logarithm of the linear frequency, reflecting similar effects in the human's subjective auditory perception.

Discrete cosine transforms (**DCT)**: In this step, we apply DCT on the 20-log energy E_k obtained from the triangular bandpass filters to have L Mel-scale cepstral coefficients. The formula for DCT is shown next.

$$\text{Cm} = \text{Sk} = 1^N \cos\left[m * (k - 0.5) * \frac{p}{N}\right] * Ek, m = 1,2,3 \ldots L \quad \text{..... (8)}$$

Where N is the number of triangular bandpass filters, L is the number of Mel-scale cepstral coefficients. Usually, we set N=20 and L=12. Since we have performed FFT, DCT transforms the frequency domain into a time-like domain called frequency domain. The obtained features are similar to cepstral; thus, it is referred to as the Mel-scale cepstral coefficients, or MFCC. MFCC alone can be used as the feature for pronunciation detection. To improve the performance, we can add the log energy and perform delta operation, as explained in the next two steps.

Log energy: The energy within a frame is also an important feature that can be easily found. Hence, we usually add the log energy as the 13th feature to MFCC. And also, we can add some other features at this step, including pitch, zero cross rate, high-order spectrum momentum.

Delta cepstral: It is also advantageous to have the time derivatives of (energy+MFCC) as new features, which shows the velocity and acceleration of (energy+MFCC). The equations to compute these features are:

$$\Delta\ \text{Cm(t)} = \frac{[\text{St}=-M^M \text{Cm(t+t)t}]}{[\text{St}=-M^M\ t^2]} \quad \text{..... (9)}$$

The value of M is usually set to 2. If we add the velocity, the feature dimension is 26. If we add both the velocity and the acceleration, the feature dimension is 39. Most of the speech recognition systems on PC use these 39-dimensional features for recognition. In this study, we have used 39 dimensional features of MFCC.

All of the above MFCC feature extraction steps are represented by the following MFCC feature extraction flow diagram depicted in figure3.

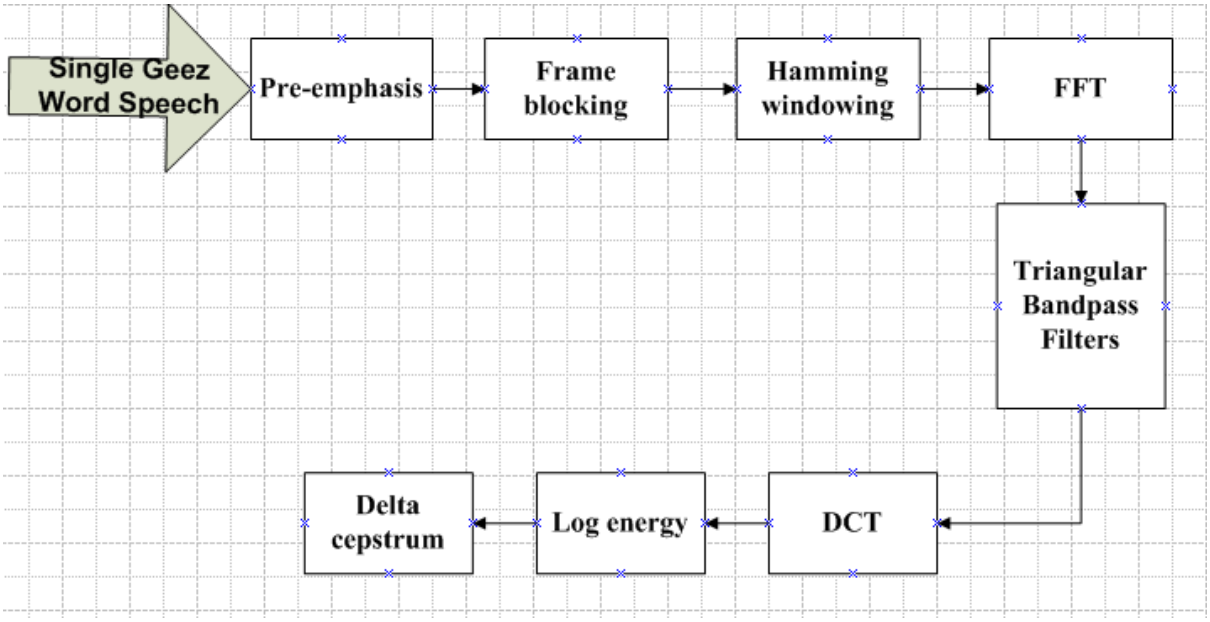

Figure 3: MFCC feature extraction process diagram (Karakas, 2010; Milwaukee, 2007)

CHAPTER FOUR

Model Design for Pronunciation Detection

4.1. Overview

Designing process is the main components of development of the model, which is based on the features pronunciation styles of the language that are discussed in chapter three and the approaches using related work in chapter two. Therefore, in this chapter, we discuss in detail about the process feature extraction and architecture of the pronunciation detection. We have two diagrams for the model, the first one is describing the process of extracting or selection of the feature of the pronunciation. The second is describing the process of detecting the pronunciation style of a given word.

4.2. Steps in Feature Extraction

Feature extraction is a process, which is used for measurements of the given audio input waveform, by producing the statistical data with their corresponding representations of the audio. This data used as to differentiate among the pronunciation style. This process is a fundamental requirement of any speech recognition system; it is the mathematical representation of the speech file. The primary goal of feature extraction is to simplify recognition by summarizing the vast amount of speech data and obtaining the acoustic properties that define speech individuality. Before describing the overall architectures of the model, first, we showed the process of feature extractor diagram figure4. The output of the feature extraction methods are used for the input of the classifier or detection model, as shown figure 5:

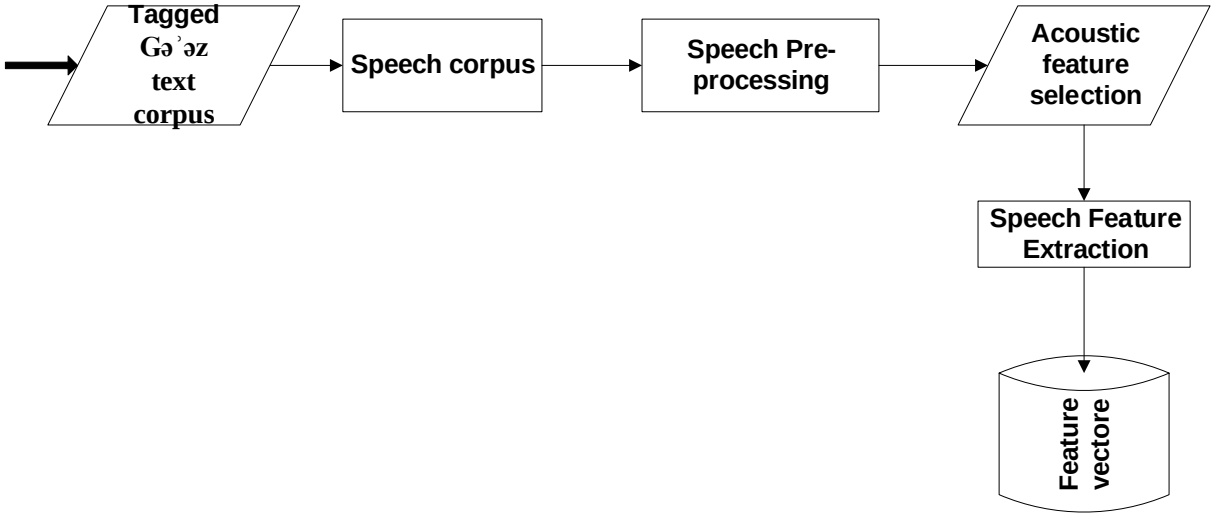

Figure 4: Speech Feature Extraction Process Diagram

As it is shown in figure 6, to extract the features of speech first prepare gəˀəz text data. After that, we record those data with gəˀəz scholars, who is a native speaker of the language. We segment the sequence of word speeches in a single word speech level using audacity tool. In speech analysis or processing, pre-processing techniques is performed before the extraction of the speech signal features. As we described in section2.4.1, there are many pre-processing techniques in speech processing. Some of these techniques are sampling and quantization, however, this process performs by recorder or instruments, when we record the text we converted into wav file format, and normalization of the speech. Speech framing is one of the processes to extract the speech features because of speech is time dependent. It may a varying or make a difference when the time is long. So, framing also used, which indicates a block of the speech into short time interval, like as 30ms. After selected the relevant acoustic features, we extract the features and use its statistical computations like mean max, min and variance of the features are computed. Finally, '15' total feature vectors are used for input layer of the classifier ANN for single Gəˀəz word, which are obtained from the acoustic features of speech signal see section 3.6.

4.3. Architecture of the Pronunciation Detection

The overall architectures of Gəˀəz language pronunciation detection model, which accept speech signals of a single word or speech corpus as shown figure 5:

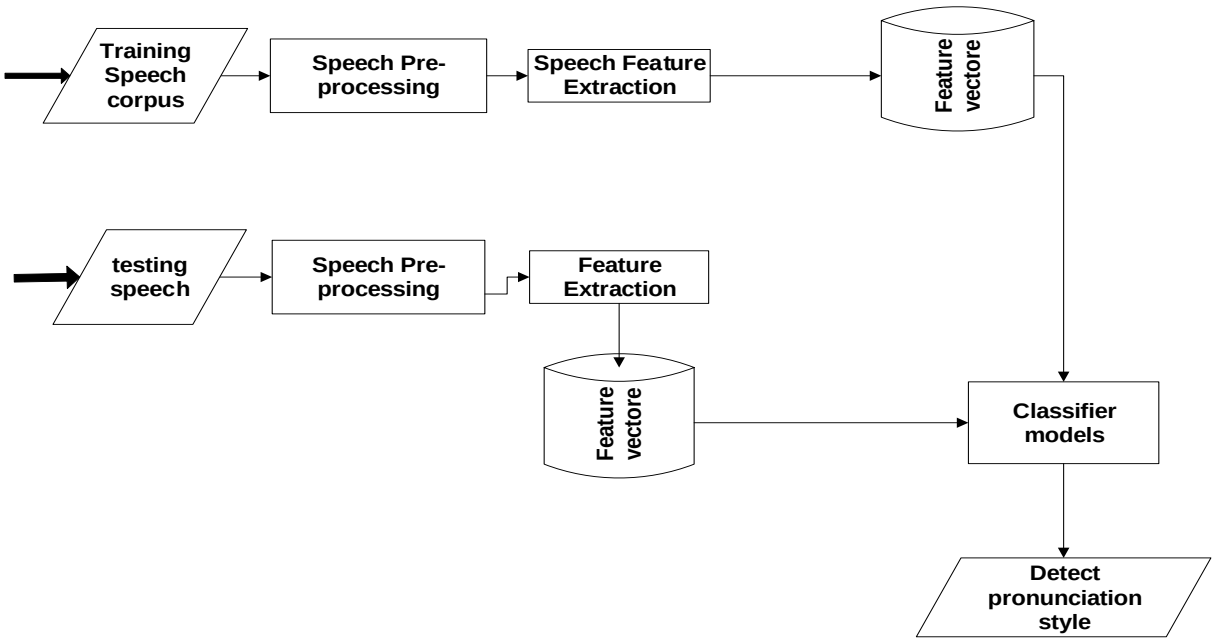

Figure 5: Architecture of Gəʾəz language pronunciation detection model

As it is shown in the figure, the architectures of the Gəʾəz language pronunciation detection process accept single Gəʾəz word speech signal, which means, the first process of the model to be pre-processing after the recording and segmenting the speech signals, the speech records from native speakers of the language scholars.

The second process of the model is speech signal pre-processing, as we describe in the section 2.4.1. The pre-processing stage in speech recognition systems is used in order to increase the efficiency of subsequent feature extraction and classification stages. Therefore, to improve the overall recognition performance. Commonly the pre-processing method that includes the sampling step, a normalization step. At the end of the pre-processing, the compressed and filtered speech frames are forwarded to the feature extraction stage. When we see the sampling in order to that a computer can process the speech signal, it first has to be digitized. Therefore, the time-continuous speech signal is sampled and quantized that performed by the recorder device. The result is a time- and value-discrete signal. Since human speech has a relatively low bandwidth (mostly between 100Hz and 8 KHz). For the purpose of a speech recognition tasks sampling frequency of 16 KHz is sufficient. So, in this thesis, 16 kHz sampling frequency is used. For having a value of the discrete signal, the sampled values are quantized. This led to a significant reduction of data. Usually, speech recognition systems encode the samples with 8 or 16 bits per sample depending on the available processing power. 8 bits per sample would mean 28 = 256 quantization levels, 16 bits per sample provide 216 =

65536 quantization levels. Concluding, if you have enough processing power, a higher bit resolution for the sampled values is preferable.

In addition to sampling, normalization is performed in the pre-processing stage, which supports for making data consistency by reducing the variety of signals. It created by different recording condition, which is the methods of processing the spectrum and segmenting the speech signals for stable speech recognition in the presence of frequency distortions are proposed.

Speech feature extraction: After the pre-processing step, feature extraction is the next component of a pronunciation detection system. This component should derive descriptive features from the windowed, and enhanced speech signal to enable a classification of sounds. This stage the pronunciation detection of the language process makes too easy by summarizing the vast amount of speech data and obtaining the acoustic features of single geʿez word. The feature extraction is needed because the speech signal contains information besides the linguistic message and has a high dimensionality. Both characteristics of the speech signal would be unfeasible for the classification of sounds and result in a high word error rate. Therefore, the feature extraction algorithm derives a characteristic feature vector with a lower dimensionality, which is used for representing the speech signal easily and classification of sounds. The wave data is represented is crucial to speech pattern recognition.

Feature vectors: It is the way numerical data representation of speech features in any dimensional vector, and it can be used for speech pattern recognition and in machine learning. In machine learning, feature vectors are used to represent the numerical property of an aspect of the sound or speech of words that are called features, which is the purpose of the ways of easily analyzable. A feature set, when used as input for classification or clustering methods, is referred to as the feature vector. An individual scalar value in this feature vector is also referred to as an attribute or feature of the speech. They are important for many different areas of machine learning and pattern processing. Machine learning algorithms typically require a numerical representation of speech of a word in order for the algorithms to do processing and statistical analysis. A vector is a series of numbers, like a matrix with one column but multiple rows or vice versa, that can often be represented spatially. A feature vector is a vector containing multiple elements of Gəʾəz word speech signal. Putting feature vectors for speech together can make up a feature space. So, in this Stage, Feature vectors are the equivalent of vectors of explanatory variables that are used in statistical techniques, which is for detecting or

classifying of the pronunciation styles. The feature vectors contain 15 acoustic features of a speech signal. Those are speech energy, duration, mean of the pitch, maximum of the pitch, minimum of the pitch, the variance of the pitch, mean of formats of speech, the variance of formants of speech, maximum formants of speech, mean and variance of three deltas of MFCC.

Detection of pronunciation style: In this process, the pronunciation detection is used to show the output of the testing data of Gəʾəz language pronunciation styles. Which is performed after the building of the model based on the training feature vectors, the model showed the pronunciation style of a given Gəʾəz word. Gəʾəz language pronunciation styles that are detected by the model, which are ተነሽ/tänäšə /rising up reading, ወዳቂ/wädaqi /folding down reading, ሰያፍ/säyafə /slant reading and ተጣይ/tät'ayə /Throwing reading. The pronunciation styles are the output of the model, has been a different format. It might be in the form of a confusion matrix by computing correctly detected and incorrectly detected, and in the form of graphical representation.

4.4. Building Classifier Techniques

In this process, we describe the techniques or a machine learning approach for purpose of building the classifier of the patterns of data or vector feature of the speech signal data. The features vectors are numerical data, which is used as training and testing input data for the classifier. At this stage, feature vectors are taken as training and testing data but the training data is labeled to their pronunciation style, which is said to be patterned. Pattern recognition is the process of classifying input data the pronunciation style of the Gəʾəz language based on the features. There are two classification methods in speech pattern recognition: supervised and unsupervised classification. We use a supervised classification method, which is the input data in the pattern recognition method uses supervised learning algorithms, which create classifiers based on training data with labeled different word pronunciation styles or classes. The classifier then accepts input data and assigns the appropriate pronunciation style or class label of the word speech. Pattern recognition has applications in computer vision, radar processing, speech recognition, and text classification. There are many machine learning approaches for classifiers of the speech pattern recognition, from those approaches we described ANN.

ANN is one of the most common classifier learning algorithm, which used for speech and speech emotion recognition. It is also well known for their ability to learn complex functions, generalize effectively, and can tolerate noise (Tebelskis, 1995). ANN is famous for its ability

to find the nonlinear boundaries for separating the patterns of the data (Mohanta and Sharma, 2015). We used Feed-forward neural network learning rule, which is based on supervised learning for pronunciation detection. In feed forward method it takes once input as a training sample and other one as a target sample. The advantages of feed-forward neural network are fixed computational time, high speed, learns general solution and tolerant the noisy.

Neural networks are composed of simple computational elements operating in parallel. The network function is determined largely by the connections between elements, the element s referred to the layer of the neural. We can train a neural network so that a particular input led to a specific target output (Wouter, Georgi and Valeri, 2010). ANNs have three layers that are interconnected. The first layer refers to of input layers that send a data to the second layer, which in turn sends the output neurons to the third layer that is, the inputs are summed and sent through an activation function. In this thesis, the input layer is the acoustic features of word-level speech signals, which stored in the database, which obtained during feature extraction process. In pronunciation detection, multilayer perception layer neural networks are also very well-known for its well-defined training algorithm and easy implementation. Hence, feature vectors are taken as input layer for ANN which consists 15/fifteen acoustic features of a word-level speech signal. An output layer of the neuron is the connection to the outside world, which is the pronunciation style of the language. Therefore, in this study, we have four the output layer of ANN. The second layer is a hidden layer of the network, which are not directly accessible to the outside world are called hidden neurons. It has been shown that with enough neurons in the hidden layer any continuous function may be learned (Lippamann, 1987). Hidden layers are used to make the network faster and efficient by identifying the important information from the input layers. In ANN activation function are applied, the most and widely usable activation function is the sigmoid function. The purpose of activation function is a help to convert the input data into a more useful output, it captures the non-linear relationship between the layers and it limits the value of the output layers. However, the numbers of layers and nodes affect the time complexity of the use of a neural network. The number hidden layer is more nodes it high time complexity and it can be led to a good performance for the classifier.

. After the classifier model develop using interconnected layers of ANN with number of iteration for training dataset and it assigns detecting each input of the data into their pronunciation style, which means one of them from four pronunciation style is more matched pattern considered as an output.

CHAPTER FIVE

Implementation of the Model for GƏʾƏZ Language Pronunciation

5.1. Overview

In the previous chapter, we discussed the design of the model of gəʾəz language pronunciation. The model of pronunciation detection is taken speech data. In this chapter, we describe the prototype implementation of the models that are designed in the previous chapters, and corpus or data preparation of the model in both text and speech data, the environment of the experiment, experimental result analysis, performance analysis, and discussion of the results.

5.2. Corpus Preparation

Corpus can be defined as a systematic collection of texts in the form of both written and spoken as the nature of the language. When we say systematic the text of structure and contents of the corpus that follows certain linguistic principles ("sampling principles", i.e. principles on the basis of, which the text included is chosen). Corpus (plural corpora) it can be used as starting points of linguistic description or a means or tools of verifying hypotheses about the language (Mohammed A. , 2012). A corpus can be a flat text, which is a text with not additional linguistic information or a text, which have each word in the text that labelled with linguistic information. The additional linguistic information in the corpus can be referred to as annotated or tagged corpus. Such linguistic information in the annotated corpus can be part of speech information/ word class category, sound category, sentiment information that specified the words/ sentiment category, pronunciation category of the words. In this thesis, we prepared the corpus in the form of both in written and spoken text of the language used for detection or labelling the pronunciation style of word. So, the annotated corpus that are tagged by experts and used for evaluation the model output.

The corpus for this thesis represents all word class categories or part of speech of Gəʾəz language. The gəʾəz word's class category, that we have used nouns, adjectives, prepositions, pronouns, Adverbs, conjunctions, verbs. In the category of verbs past or perfective tense /ቀዳማይ/ታላፊ/ qädamayə/halafi from main verbs /ዐቢይ አንቀጽ/ ʾäbiyə ʾänəqäs'ə and their

morphology of the main root verbs by አእማድ/ ʾäʾəmadə[1] are used. አእማድ/ ʾäʾəmadə in Gəʾəz language system has the grate role to shows the time, the completion of activities, and condition of activity completion. There are five most common አእማድ/ ʾäʾəmadə/pillars in Gəʾəz language ተደራጊ/tädäragi፣ አድራጊ/ädəragi/፤ አስደራጊ/äsədäragi/፣ ተደራራጊ/tädäraragi/፣ አዳራጊ/ädaragi. The most corpus of the thesis is obtained from the book of (በላይ፣2007) ፣ which is Amharic-Gəʾəz dictionary, and some Gəʾəz words are obtained from the part of bible which is Psalms book and book of (ያሬድ፣1997). The reason that have been, the corpus of the thesis obtained from different books, In Gəʾəz language, it requires to care letter missing or spelling error in the word, because of alphabets or letters in Gəʾəz language have grate factors, in other words all linguistic activities based on the alphabets or letters of Gəʾəz language. If it be change the word may come a meaning change, pronunciation change, type (part of speech) change. These corpus is important for any natural language processing application especially for pronunciation style labelling and detection. Preparing well organized and relevant corpus for some model has a great contribution for model performance improvement. Therefore, we used language experts for preparing the corpus for both written and spoken format. In this corpus to measure the performance of the model, it requires manually tagged labelled or classified with their corresponding pronunciation style. Speech corpus also, it need to record those text by native speaker of the language.

When we prepared the corpus for both written and spoken data, we used four Gəʾəz scholars. For preparing the data four scholars of the language are used for tagging and recording purpose. The words are tagged into their corresponding pronunciation style and reviewed the whole Gəʾəz words by scholars. For speech corpus, we record two scholar whose native speakers of Gəʾəz language using a tablet sound recorder. The total words in speech corpora are 3400 speech signals. However, to equalize for all pronunciation styles we have used 1600 words, 800 of words from one scholars and 800 words from the second scholars. The following recorder and voice properties or specifications are used and after getting the proper setting of speech, the texts are recorded by the native speaker properly within a quiet environment:

[1] አእማድ/ ʾäʾəmadə in Gəʾəz language system has the grate role to shows the time, the completion of activities, and condition of activity completion. There are five most common አእማድ/ ʾäʾəmadə/pillars in Gəʾəz language ተደራጊ/tädäragi፣ አድራጊ/ädəragi/፤ አስደራጊ/äsədäragi/፣ ተደራራጊ/tädäraragi/፣ አዳራጊ/ädaragi.

- Set the microphone volume to 3.0.
- Set the default Sampling Rate to 16000 KHz.
- Set the default Sample Format to 32-bit float.
- The Channels to 1 (Mono).
- After recording 3gp format we are changed to WAV format and edit using Audacity speech editor.

Those data when we extract the acoustic feature into separate data for training and testing data purpose, testing dataset are 10% training dataset. More detail corpora, look the following table 7, and the text that where records re attached on appendix.

Table 6: Spoken Gəʾəz word speech corpus

No	Pronunciation style	Number of words	Total (%)
1	ወዳቂ/wädaqi/	400	25
2	ተጣይ/ tät'ayə/	400	25
3	ሰያፍ/ säyafə/	400	25
4	ተነሽ/ tänäšə/	400	25
	total	1600	100

5.3. Experimentation Environment

In this thesis, as we listed the research questions in chapter one section1:2, such as "How to identify the pronunciation style of words for Gəʾəz language". This question is answered, by the model, which follows an experimental research design approach, by implementing the acoustic feature and the help of machine learning classifier. As we described in the previous chapters to do the experiment, the features of the data are determined by the selection of tools and experiment type. In this section, we discuss the procedures and tools for implementation of the experiment of the model.

To perform Gəʾəz language pronunciation detection model, we have been used different tools such as recorder device/tablet, microphone, Audacity, Praat, MATLAB, and ANN as a classifier algorithm. The tools, audacity, and Praat are used for audio file editors. For design and implement the model, we use ANN learning algorithm through the program MATLAB

tools. When we perform the experiment for detection of the pronunciation style. The first process is preparing the corpus and then read or import the speech signal/file or data into the model using MATLAB functions. From those speech files, we extract the acoustic features, which representation of a given word. As well as the programming behind the neural network is used for classifier model that learns from the training dataset. MATLAB is more user-friendly to machine learning algorithms. When we Link with MATLAB a Neural Network Toolbox that provides many of the functions needed to implement any type of neural network. For the speech processing, MATLAB is powerful tools, which accept a wav file from the file folder after recording the data. However, we can record an audio through the computer soundcard using wave record MATLAB functions. Once the wav files imported and read to the program. We used a function feature, which is user-defined functions that are called all acoustic features and the results stored as feature vectors. The acoustic features that are called by features function are pitch, Formant, MFCC, energy and speech rate or duration of the speech. For details of acoustic features, we can refer feature extraction section 3.6. And for MATLAB source code of each acoustic feature see appendix section.

The total speech words, which used the model are 1600. However, after extraction the acoustic features, we separate the training and the testing dataset. The mode used 1440 speech data for training, which is 90% total data and 10% of the total data used as a testing dataset. After extracting the acoustic features from each input audio or speech file, statistical functions are used for computation of maximum, minimum, variance and mean of acoustic features. The total feature of the speech signal for single Gəʾəz word is fifteen (‘15’). Those are speech rate/duration, Energy, maximum of the pitch, minimum of the pitch, mean of the pitch, variance of the pitch, maximum of formant, mean of formant, the variance of formants, variance and mean of the first three MFCC. Thus, fifteen features are taken as input of neuron network, and we are used fifteen (15) ANN hidden layer. The output layers of ANN are styles of Gəʾəz language pronunciation that is referred to 4/four classes. The structure of ANN as it is shown the on figure10 and which support sigmoid activation function.

The total feature vectors for used to build pronunciation detection are 15X1440 training corpus and for each of the pronunciation style taken as the 15X360 feature vectors. The reason for minimizing the data, the word that is tagged pronunciation style as ስያፍ/ säyafə/ is small compared to other pronunciation styles in the corpus. From those total data, we divided the dataset randomly choose the percentage of input data into 3 categories to build the model

namely training, validation and testing, where training set is used to fit the parameters of the classifier i.e. to find the optimal weight for each feature. Whereas validation set is used to track the parameters of a classifier that is to determine a stopping point for the training set. And finally, the test set is used to test or measure the final performance of the model and estimate error rate. Those data which have 80%, 10%, 10% of the total training data. The training datasets are labeled with their corresponding pronunciation style or target class before ANN trained, that is the training data is 15X1440 and the target class is 4X1440, which indicates the pronunciation style for these 1440 Gəˀəz words. It is followed the supervised learning method. In another word, the once a network has been structured for a particular application, the neural network is ready to be trained. To start this process the initial weights are chosen randomly. Then, the training, or learning, begins. There are two approaches to training - supervised and unsupervised. Supervised training involves a mechanism of providing the network with the desired output either by manually "grading" the network's performance or by providing the desired outputs with the inputs. Unsupervised training is where the network has to make sense of the inputs without outside help.

The massive majority of networks apply supervised training. Unsupervised training is used to perform some initial characterization on inputs. However, in the full-blown sense of being truly self-learning, it is still just a shining promise that is not fully understood, does not completely work, and thus is relegated to the lab. In supervised training, both the inputs and the outputs are provided. The network then processes the inputs and compares its resulting outputs against the desired outputs. Errors are then propagated back through the system, causing the system to adjust the weights which control the network. This process occurs over and over as the weights are continually tweaked. The set of data, which enables the training is called the "training set." During the training of a network, the same set of data is processed many times as the connection weights are ever refined. When finally, the system has been correctly trained, and no further learning is needed, that stops the iteration of the network. In this thesis for the classifier, we use a function of MATLAB code (see appendix) a for ANN classifier. By adjusting a number of hidden layers of the neural we have a good performance. So, in this thesis, we use 15 input layers, 20 hidden layers and 4 output layers of ANN classifiers. The general structure of ANN for pronunciation detection as shown figure8.

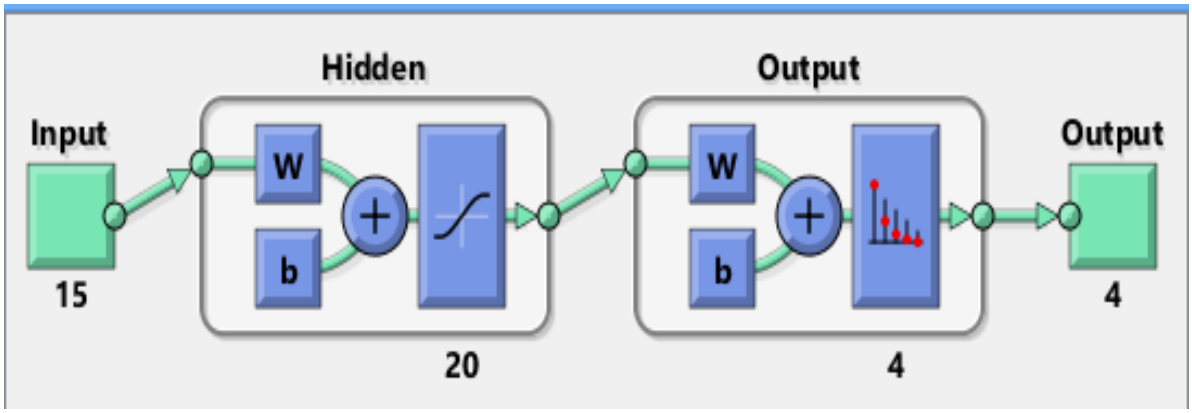

Figure 6: Structure of Neuron Network for Pronunciation Detection

5.4. Experimented Result for Pronunciation Detection

ANN algorithm is used for speech patterns classification problem. Patterns are the numerical data which the representation of acoustic features of pronunciation style for Gəʾəz words. These acoustic features include speech rate, energy, pitch, formants, and MFCC. After extraction of these features it generates numerical data as a representation of each acoustic feature. Those numerical data are a feature vector used as input of the classifier or ANN. In this thesis the model is trained five times with different hidden layers as shown table7. Generally, after properly trained the model ANN has fifteen inputs layers, twenty hidden layers, four outputs, it uses sigmoid function, and the total training corpus are divided into 80% for training, 10% for validation and 10% for testing to build a model. After building the model we use different speech corpus for testing the model. The result of the model is generating in the form of confusion matric and MSE with training time as shown table7 and 8 respectively.

Table 7: Experiment Result Based on Training times and hidden layers that shown on confusion matrix

No hidden layer	No of training from total	Testing dataset result %	No of training times				
			1st training model performance %	2nd training model performance %	3rd training model performance %	4th training model performance %	5th training model performance %
10	80%	76.9	77.2	73.7	77.0	73.7	78.5
15	80%	72.5	75.8	76.0	75.1	76.3	74.8
20	80%	76.9	78.4	80.3	77.6	74.6	78.1

Table 8：Error rate performance validation Based on Training times and hidden layers

No hidden layer	No of training data from total	No of training times				
		1st training model error rate	2nd training model error rate	3rd training model error rate	4th training model error rate	5th training model error rate
10	80%	0.29	0.33	0.30	0.31	0.29
15	80%	0.289	0.31	0.307	0.327	0.31
20	80%	0.258	0.252	0.32	0.277	0.36

As shown in table 7and 8, the best performance of the model is indicates highest accuracy and minimum error rate value. When we use 20 number of hidden layers and after number of iteration training time occurred, the second training time is best performance of the model.

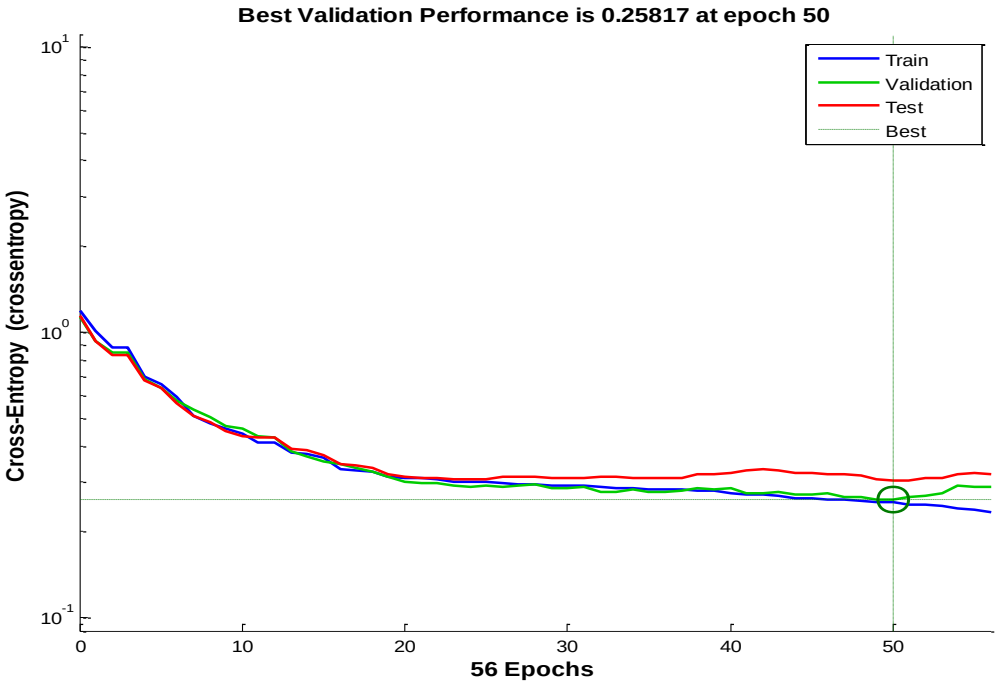

Figure 7: Mean square first times training

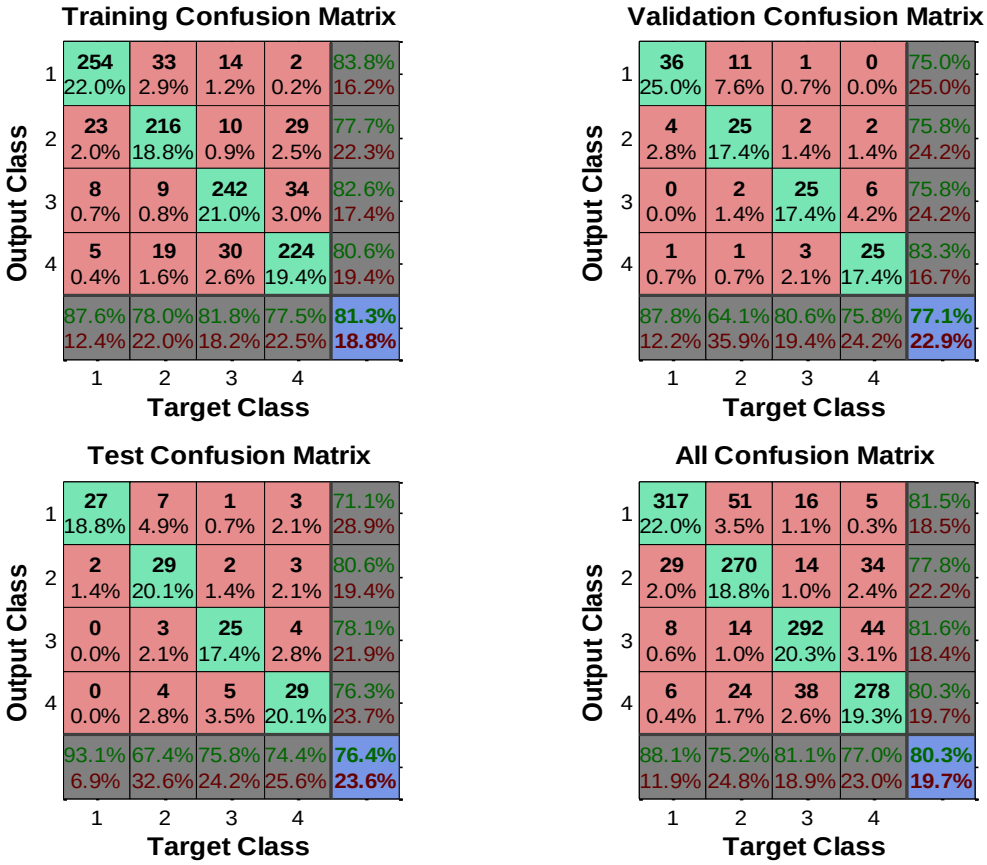

Figure 8: Model result first time training

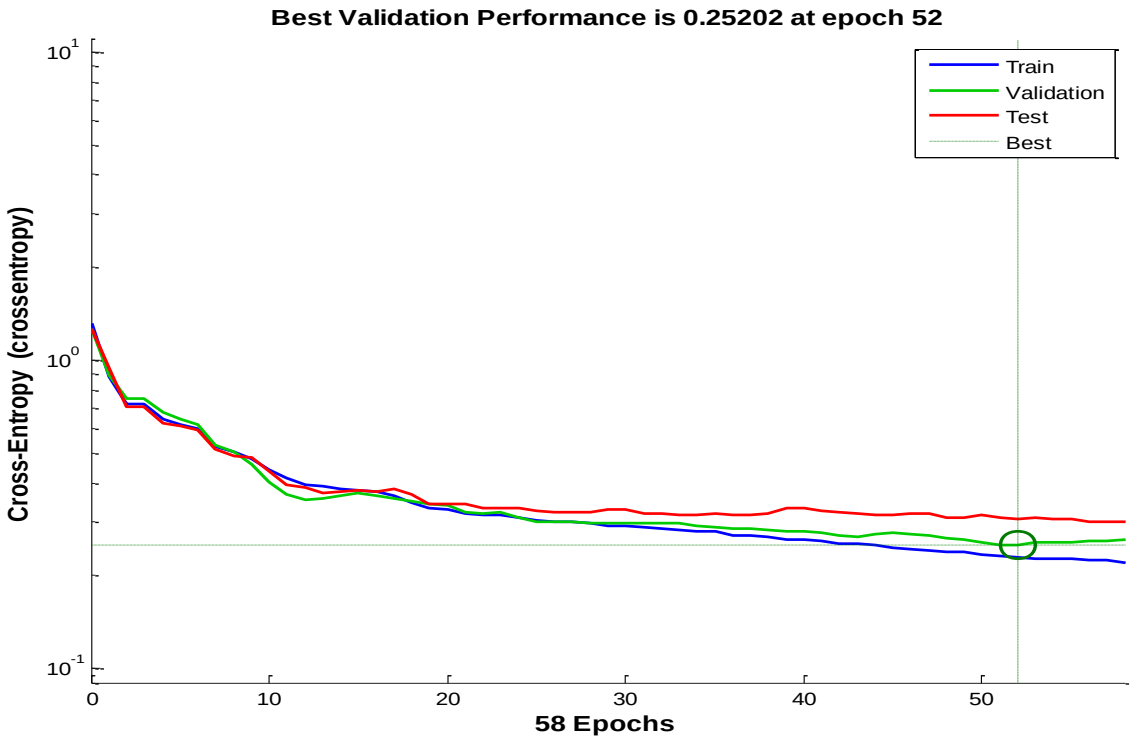

Figure 9：Mean square after second times training

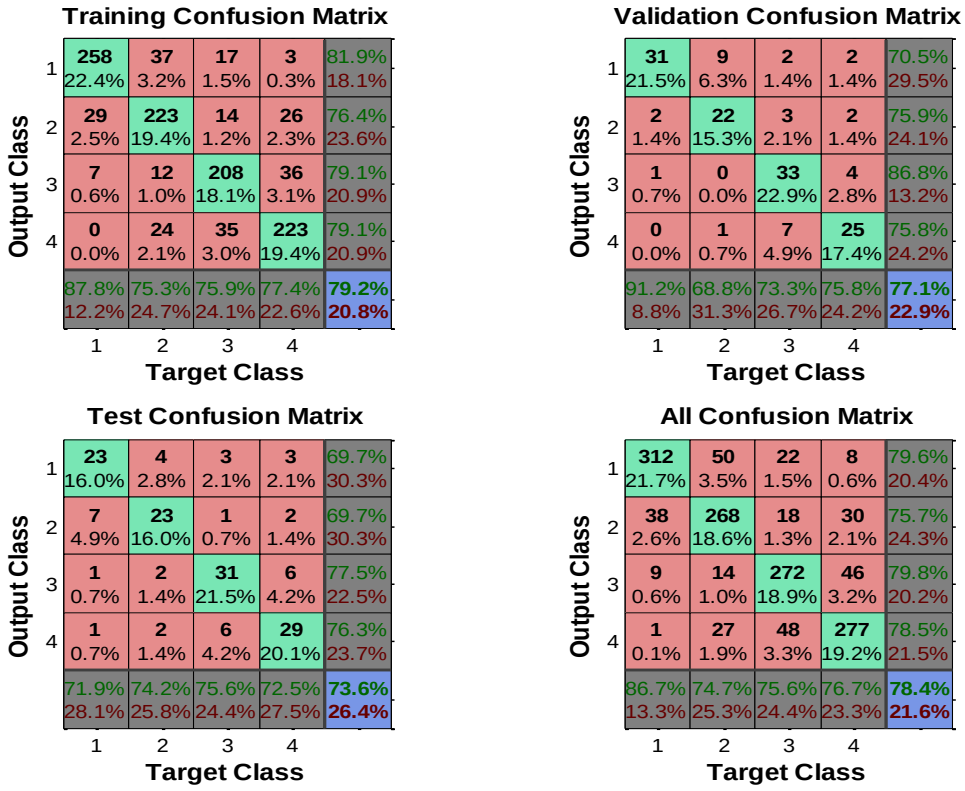

Figure 10：Model result after second time training

5.5. Model Performance Evaluation

For the performance evaluation for pronunciation detection, to show the result we use a confusion matrix and mean square of errors (MSE). The confusion matrix is a tabular format that is used to describe the performance of classification of the model or classifier on the given set of test data for which the true values are well-known. Whereas, the mean squared error (MSE) or mean squared deviation (MSD) of an estimator (of a procedure for estimating an unobserved quantity) measures. The average of the squares of the errors or deviations is the difference between the estimator and what estimated or actual value.

Furthermore, after a suitable train process i.e. with a low error rate performance the neural network is ready to classify extra speech signals. For example, we have trained five times from those training times the second training is best classifier performance at epoch 52 as shown table7 and 8, and figure 10 and 11. An epoch means a number of times for all training feature vectors used once to update the weights to the features. Therefore, the system is now assigned to classify a completely new speech corps. In this thesis, we use 160 different words for testing dataset. ‘40’ words for each pronunciation style. The classification result in terms of error rate is shown in Fig 10. Therefore, Figure 12 shows that the testing result for pronunciation detection. In the confusion matrix, each row and each column represents for actual or target class and output/desired class, which shows the performance of the model. The number in cell ‘1’ is the intersect point between the 1st row and the 1st column, which represents for how many ወዳቂ/wädaqi speech words have been classified into the ወዳቂ/wädaqi output. In cell ‘2’ is the intersect point between the 1st row and the 2nd column, shows that how many ተጣይ/tät'ayə/ speeches have been misclassified into ወዳቂ/wädaqi pronunciation style or class. Classification results of all four sets of data and overall result are shows on figure12, this result gives a clear idea of how the classification model is good for unknown data. Let see an example in figure12, from 40 of ወዳቂ/wädaqi words 33 have been classified in the correct output but 7 of ወዳቂ/wädaqi words speeches are misclassified. From 7 speech 4 speech is misclassified into ተጣይ/tät'ayə/ pronunciation style, 3 of speeches are misclassified into ሰያፍ/ säyafə/ pronunciation style. For more detail classification result of the model is shown in figure12. The total classification rate is 76.9% for the new speech samples.

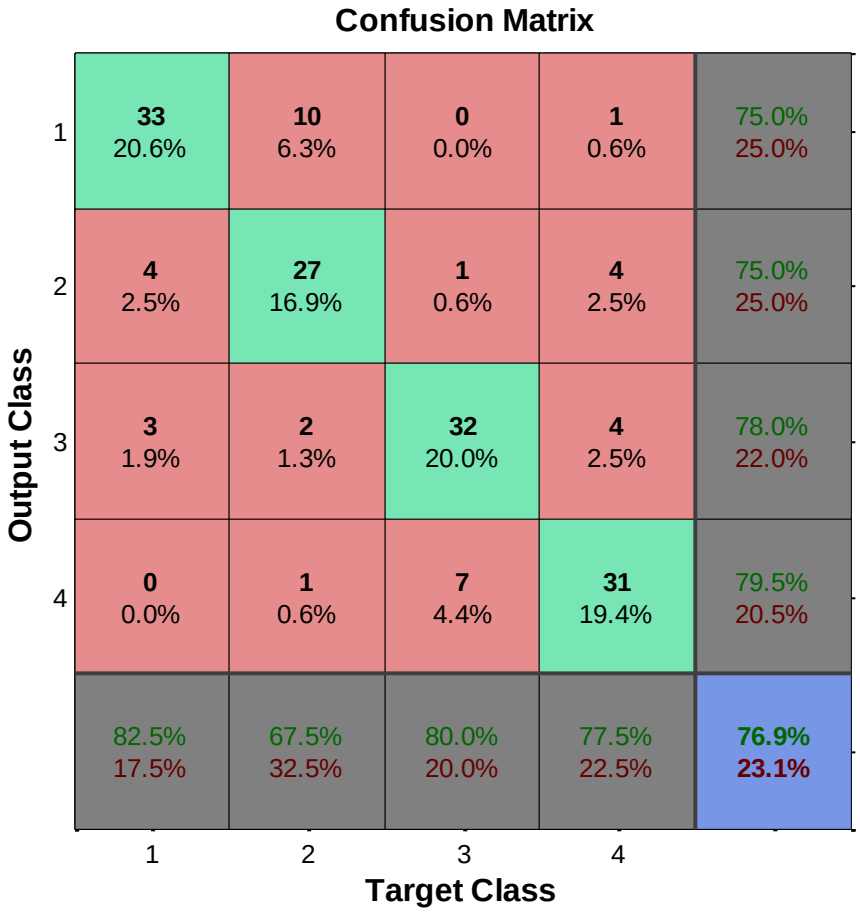

Figure 11: Classification result of testing dataset

5.5. Discussion of the Results and Finding

First, for the recording process, author recorded all the speech corps. Besides, each pronunciation style is performed in a native speaker and in similar conditions. Therefore, pronunciation style ተነሽ/ tänäšə/ and ስያፍ/säyafə/ speeches are high sound, short duration and powerful, less speech duration, less pitch when compare wädaqi and ተጣይ/tät'ayə/ pronunciation style. Those make it comparably easier for the classifier to decide the pronunciation style of each speech. The further work could be for a more accurate system to classify speeches use with additional different features of pronunciation and native speaker.

However, there is some cause of errors for the model development. Those are during time of recording the words that not be perfect for the condition of the environment. Mode or status of the experts who is a record his voice or speech.

In the feature extraction process, the accuracy of features is an important factor. For speech rate, the number of letters in a word is challenged in this study. Therefore, it could be examined to compare the simulate result and counting of the letters. From all pronunciation style ሰያፍ/ säyafə/ have more alphabets or letters in Single words and ተጣይ/tät'ayə has less letters in a word. For pitch and formant, it is not likely to examine the accuracy for each frame, so the reasonable ranges for pitches and formants are defined to filter other error values out. This is a quick and simple implementation as well, but it might also cause the edge effect. One way to improve this is by applying a start and end point detection algorithm to indicate the start and end sample point of real voice part.

For the classification process also, the error rate is highly depended on the training times compare the confusion results in figure 9 and figure 11. The overall error rate varies from 21.6% to 19.7 in different training times. However, the suitable number of training times is neither fixed nor predictable in neural network system. For the random process of choosing training, validation and test data, the weights assign for each feature changes the choice. Thus, for the best of the model in its desirable result, it is necessary to test the training process in several times. In the pronunciation detection is challenged by native speakers the corpus of the model is depend on the speakers, if we got more native speaker, the model can more accurate and used as a guidance for non-native speakers.

Based on the result of the constructed model of pronunciation style detection in Geez language, the findings of the study are summarized as follows:

- Gəʾəz language had its own pronunciation acoustic feature for each pronunciation style.
- To identify the pronunciation style of the words, we need to know the pitch, speech rate, energy formants and MFCC. Each pronunciation style has own Acoustic feature, which might differentiate each other's.
- Based on the acoustic feature of Gəʾəz speech words and using machine learning approach, can detect the pronunciation style from speech signals automatically.

CHAPTER SIX

Conclusion and Recommendation

6.1. Conclusion

The purpose of this thesis is to design pronunciation detection for Gəʾəz language, which used for learners or non-native speakers' to improve their pronunciation skill. The acoustic features of Gəʾəz language pronunciation style are useful for identified the pronunciation style of the language automatically. Then the various techniques to pronunciation detection are reviewed. MATLAB neural network is a powerful tool for pattern recognition and classification. And MATLAB has the ability to learn from the experience of the training feature, the chosen acoustic features of speech signals could easily be represented into the system and training for the target pronunciation. After suitable times of training process, extra test signals are load into the system for pronunciation detection. Which gives with the desired result of 76.9% classification rate those selected acoustic features are (speech rate or duration, energy, pitch, formant and MFCC) are proven to be good representations of Gəʾəz language pronunciation for the speech signal.

The model development based on pronunciation features of the speech signal. In this thesis, we have focused only the four-major pronunciation style of Gəʾəz language. Gəʾəz language has its own structure, a formation of words, character or alphabets and also numbers the rule that governs the pronunciation style. Therefore, in this thesis designed a model which evaluate the pronunciation skill of speakers or by detecting correctly pronounced and incorrectly pronounced in percentile. The main challenged and focusing process is preparing corpus by native speakers of Gəʾəz language pronunciation. Language experts are used for preparing the training data for the building of the model and additional testing data evaluated the performance of the model.

Furthermore, pronunciation accepts in the form of the speech signal at the word level, which follows some procedures. Primarily preparing speech corpus, the words that are recorded are obtained Amharic-Gəʾəz dictionary books. And then identifying acoustic features that used to represent pronunciation style of the language. Acoustic features are the most important approaches for detecting or recognizing speech signals. The process of feature extraction, which is extracting acoustic features with the help of MATLAB tools and compute the statistical methods of each feature. After extracting the acoustic feature, we stored those

features as the feature vector used as input of classifier. The last process builds the classifier using ANN classifier with proper training times.

Generally, Gəʾəz language pronunciation style can be automatically identified by the help of machine learning and acoustic-based features. Identifying the pronunciation style of the language is used for higher level of NLP application such as TTS translation, speech recognition and guidance for learns. In this study, the challenges we faced are the facts that the language does not have a native speaker who can give its basic linguistic facts and noisy is a crucial challenge for preparation of speech corpus, and that there is no study, as far as we know, done on the language so far from the computational perspective.

6.2. Recommendation

Currently, there are many research areas in NLP that can be done for different language in locally as well as internationally. In Gəʾəz language a few researchers conducted some areas of the NLP components. However, when we see Gəʾəz language is the founder or base of other semantic languages. Gəʾəz language has more complex and rich morphological process, grammar and formulation of words and sentence, it is also rich and complex than another semantic language specifically Amharic, tgriea, Tigre. Most sematic languages structure and properties are under Gəʾəz language structure and properties. We can say that almost all other semantic languages are proper set of Gəʾəz language.

Therefore, in this thesis, we recommended the following research areas as a way forward:

1. Design and develop pronunciation identifier model for minor pronunciation style of Gəʾəz language which are not covered by this work (those are፡ ዐራፊ/ቀዋሚ፣ተናባቢ፣ ቆጣሪ፣ ፣ ጠቅላይ፣ ጠባቂ፣ ልህሉህ፣ ጎራጅ፣ ጥያቄያዊ፣ ትርአስ ንባባት).
2. Phoneme based pronunciation detection or recognition for Gəʾəz language pronunciation.
3. Design an automatic pronunciation detection using other machine learning approach for Gəʾəz language.

References

Kudiri, Said and Nayan. (2012). Emotion Detection Using Relative Amplitude-Based Features through Speech. *IEEE*, 522-525.

Neumeyer, Franco, Digalakis and Weintraub. (2000). Automatic scoring of pronunciation quality. *Elsevier Science B.V*, 11.

Ai, R. (2015). Automatic Pronunciation Error Detection and Feedback Generation for CALL Applications. *DFKI GmbH, Language Technology Lab Alt-Moabit 91c, 10559, Berlin, Germany*, 12.

Allen, J. (1995). *Natural Language Understanding. .* New York / Ontario / Wokingham: U.K.The Benjamin/Cummings Publishing Company, Inc.

Alpaydın, E. (2010). *Introduction to machine learning.* USA: The MIT Press Cambridge, Massachusetts Institute of Technology.

Belean, B. (2013). Comparison of Formant Detection Methods Used in Speech Processing Applications. *Research Gate*, 5.

Bender, M. L. (1976.). *Language in Ethiopia.* London: Oxford University Press,.

Bird, Klein and Loper. (2009). *Natural Language Processing with Python.* United States of America: O'Reilly Media, Inc.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* New York, NY 10013, USA: Springer Science+Business Media, LLC.

Blunso, P. (2004). *Hidden Markov Models.*

Dalby, Kewley and Sillings. (1998). *language specific pronunciation training using the HearSay system.* Marholmen, Sweden: Proc. Speech Technology in Language learning.

Demircan and Kahramanli. (2014). Feature Extraction from Speech Data for Emotion Recognition. *Journal of Advances in Computer Networks, Vol. 2, No. 1*, 3.

Desta, B. (2010). *Design and Implementation of Automatic Morphological Analyzer for Ge 'Ez Verbs.* Addis Ababa, Ethiopia: Thesis In Addis Ababa university.

Digalakis and Moustroufas. (2007). Automatic pronunciation evaluation of foreign speakers using unknown text. *speakers using unknown text. In Comput. Speech Language*, page 219-230.

Dillmann, A. (1899). *Ethiopic Grammar.* (C. Bezold, Ed.) London: Amsterdam Hilo Press.

Doremalen, Cucchiarini and Strik. (2013). Automatic pronunciation error detection in non-native speech: The case of vowel errors in Dutch. *Centre for Language and Speech Technology, Radboud University Nijmegen, Erasmusplein 1, 6525HT,*, 1336–1347.

Doremalen, Cucchiarini and Strik. (2009). Automatic Detection of Vowel Pronunciation Errors Using Multiple Information Sources. *Department of Linguistics, Radboud University Nijmegen The Netherlands*, 6.

Dumais,Platt and Heckerman. (2010). Inductive Learning Algorithms and Representations for Text Categorization. *Mehran Sahami Computer Science Department Stanford University,Stanford, CA 94305-9010*, 8.

Felber, P. (2001). *SPEECH RECOGNITION Report of an Isolated Word experiment.* Illinois Institute of Technology: ECE 566 Statistical Pattern Recognition.

Franco, Neumeyer, Ramos and Bratt. (1999). AUTOMATIC DETECTION OF PHONE-LEVEL MISPRONUNCIATION FOR LANGUAGE LEARNING. *6th European Conference on Speech Communication and Technology* (p. 4). USA: ISCA Archiev.

Gales and Young. (2008). The Application of Hidden Markov Models in Speech Recognition. *Foundations and Trends in Signal Processing*, 195–304.

Grace Ngai, F. R. (2001). *Transformation-Based Learning in the Fast Lane. , .* MD 21218, USA, Weniwen Te Hnologies Hong Kong.: Johns Hopkins University Baltimore.

Haykin, S. (1999). *Neural Networks: a comprehensive foundation.* 2nd Edition, Prentice Hall.

Jurafsky and Martin. (2006). *Word classes and part-of-speech tagging. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech recognition2,.*

Karakas, M. (2010). *COMPUTER BASED SYSTEM CONTROL USING VOICE INPUT.* İZMİR: Dokuz Eylül University.

Karpagavalli and Chandra . (2016). A Review on Automatic Speech Recognition Architecture and Approaches. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 12.

Kudirp, Said and Nayan. (2012). Emotion Detection Using Relative Amplitude-Based Features through Speech. *Universitiy Teknologi PETRONAS. in Universiti Teknologi PETRONAS.*, 5.

Kumar, A. (2013). *Morphology Based Prototype Statistical Machine Translation System for English to Tamil Language, .* Tamilnadu, India: A Thesis Submitted for the PHD in the School of Engineering, Amrita School Of Engineering Amrita Vishwa.

Lee, Mower, Busso, Narayanan,. (2011). Emotion recognition using a hierarchical binary decision tree approach. Speech Communication,. *Elsevier B.V. All, in University of Texas at Dallas, Dallas, TX 75080, USA*, 1162-1171.

Lengyel, L. (2015). *Validating Rule-based Algorithms, 12(4), 59–75.*

Leslau, W. (1987). *Comparative dictionary of Ge'ez (Classical Ethiopic) : Gə'əz English/English-Gə'əz with an index of the Semitic roots".* Germany: otto Wiesbaden: Harrassowitz.

Lindblom etal. (2010). The Gunnar Fant Legacy in the study of Vocal Acousics. *10'eme congres Francais Acoustique*, 6.

Lippamann, R. (1987). An inroduction to computing with neural nets, ,. *IEEE Acoustics*, Speech and Signal Processing ,4(2:4-22).

Liu etal. (2010). *Algorithms in Signal Processors Audio and Video Applications.* Sweden: Dept. of Electrical and Information Technology, Lund University.

Loizou. (1999). *COLEA: A MATLAB software tool for speech analysis.* Dallas: University of Texas.

Menzel , Herron, Bonaventura and Morton. (2000). *Automatic detection and correction of non-native English pronunciations.* University of Hamburg, under its Language Engineering project ISLE LE4–8353.

Milwaukee, W. (2007). *SPEech Feature Toolbox (SPEFT) Design and Emotional Speech Feature Extraction.* Marquette University.

Mitchell, T. M. (1997). *Machine Learning.* McGraw-Hill Science/Engineering/Math.

Mohammed, A. (2012). *Machine Learning Approach for Voicing Detection.* ADDIS ABABA, ETHIOPIA: Thesis in ADDIS ABABA UNIVERSITY.

Mohanta and Sharma. (2015). *Human Emotion Recognition through Speech.* Assam Don Bosco University: Krishi Sanskriti Publications.

Muluken, A. (2007). *Ge'ez Verb Classification in the Three Traditional Schools of Q. .* Addis ababa, Ethiopia : thesis in AddisAbaba University.

Murphy, A. (2014). *Implementing Speech Recognition with Artificial Neural Networks.* Sault Ste. Marie, Ontario: Algoma University.

Nikhath, Subrahmanyam and Vasavi. (2016). Building a K-Nearest Neighbor Classifier for Text Categorization. *(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (1)*, 254-256.

NLTK. (2016, October 17). Retrieved from Natural Language Toolkit: http://nltk.org/book_1ed

Peabody, M. A. (2011). *Methods for Pronunciation Assessment in Computer Aided Language Learning. PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.* USA.: PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Proakis and Manolakis. (1998). *Digital signal Processing.* Northeastern Uniersity boston,Massachusettes: 4th edition,Pearson Education inc. Upper Saddle River.

Python . (2016, October 17). Retrieved from Python programming: www.python.org

Rui, M. (2014). *Parametric Speech Emotion Recognition Using Neural Network.* UNIVERSITY of GAVLE.

Rumelhart, Hinton and Williams . (1986). *learning internal representations by error propagation, .* parallel Distributed Processing: Explorations in the Microstructure ofc Cognition, volume 1, pages 318-362. : MIT Press.

Samer, A. a. (2011). *Automatic Prominence Classification in Swedish Centre for Speech Technology, .* Stockholm: Royal Institute of Technology, Lindstedtsvägen 24, SE-10044.

Scelta, G. (2001). *The Comparative Origin and Usage of the Ge ' e z writing system of Ethiopia.*

Seymore, a. R. (1996). Language and pronunciation modelling in the cmu 1996 hub-4 evaluation. *In Proc. of DARPA Speech Recognition workshop, chantilly,Virginia, Morgam kaufmann* .

Shaalan, K. (2010). Rule-based Approach in Arabic Natural. *International Journal on Information and Communication Technologies, Vol. 3*, 8.

Snell and Milinazzo. (1993). Formant Location From LPC Analysis Data. *IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 1, NO.2*, 6.

Solomon, Martha and Wolfgang. (2009). Amharic Speech Recognition: Past, Present and Future. *In: Proc. 16th International Conference of Ethiopian Studies* (p. 12). Hamburg, Germany: University of Hamburg.

Srikanth and Salsman. (2012). Automatic Pronunciation Evaluation And Mispronunciation Detection Using CMUSphinx. *Proceedings of the Workshop on Speech and Language Processing Tools in Education,*, 61–68.

Tanzim , Azharul , Mahtab. (2015). A Rule Based Approach for NLP Based Query. *Proceedings of International Conference on Electrical Information and Communication Technology (EICT 2015)* (p. 14). Khulna , Bangladesh: Khulna University of Engineering & Technology (KUET).

Tebelskis, J. (1995). *Speech Recognition using Neural Networks.* Pittsburgh, Pennsylvania : Carnegie Mellon University.

Vijayalakshmi and Leema. (201`4). Real-time Speech Emotion Recognition Using Support Vector Machine. *International Journal of System and Software Engineering, Volume 2*(Issue 1). Retrieved from http://www.publishingindia.com

Witt and Young. (1998). Phone-level pronunciation scoring and assessment for interactive language learning. *Elsevier Science*, 14.

Wouter, Georgi and Valeri. (2010). Neural Networks used for Speech Recognition. *AUTOMATIC CONTROL, UNIVERSITY OF BELGRADE,*, 7.

Yitayal, A. (2014). *Morphological Analysis of Ge'ez Verbs Using Memory Based Learning.* Addis Ababa, Ethiopia: Thesis in Addis Ababa University.

Zegers, P. (1998). SPEECH RECOGNITION USING NEURAL NETWORKS. *ARIZONA*, 98.

Zelalem, M. (2013). *Design and Development of Part-of-speech Tagger for Kafi-noonoo Language.* Addis Ababa, Ethiopia: Thesis in Addis Ababa University.

Zhao etal . (2011). Automatic Lexical Stress Detection Using Acoustic Features for Computer-Assisted Language Learning. *APSIPA ASC ,State Key Laboratory on Transducing Technology, Institute of Electronics,*, 5.

ሊቀ ጎሩያን በላይ መኰንን፤ 2007 ዓ.ም ፤ሕያው ልሳን ግዕዝ-አማርኛ መዝገበ ቃላት ፤ አዲስ አበባ።
ዶ/ር ሲሳይ ደመቀ ፤ 2005 እ.ኢ.አ፤ ሰዋስው ወልሳነ ግእዝ፤ ጎንደር ፤ ያልታተመ የመጻፍ ክፍል

አለቃ ኪዳነ ወልድ ክፍሌ፤ መጽሐፈ ሰዋስው ወግስ ወመዝገበ ቃላት ሐዲስ፣ አዲስ አበባ፣ አርቲስቲክ ማተምያ ቤት።

መምህር አፈወርቅ ተክሌ የቅኔ ና የመጽሐፍ መምህር፤2005ዓ.ም፤ መጽሐፈ ታሪክ ወግስ፤ አዲስ አበባ።

ሊቀ ሊቃውንት ያሬድ ሽፈራው፤ የቅኔ ና የሐዲሳት ትርጓሜ መምህር፤ 1997 ዓ.ም፤መጽሐፈ ግስ ወሰዋሰው

መርዖ መጻሕፍት፤ ባህርዳር በቅዱስ ጊዮርጊስ ማተሚያ ቤት።

ዘርአዳዊት አድሐና ፣ 1996፣ መርዖ ሰዋስው ዘልሣነ ግእዝ፣ አዲስ አበባ፣ ብርሃን ና ሰላም ማተምያ ድርጅት።

መምህር ደሴ ቀለብ፤2007 ዓ.ም፤ ትንሳኤ ግዕዝ ፤ አዲስ አበባ፤ በኢትዮጵያ ኦርቶዶክስ ተዋህዶ

ቤተክርስቲያን ማኅበረ ቅዱሳን።

ፕሮፌሰር ፍቅሬ ቶሎሳ ፤ 2008 ዓ.ም ፤ የኦሮሞ እና የአማራ እውነተኛው የዘር ምንጭ ፤ አዲስ አበባ ፤ NEBADSN m.c PLC

Appendix

Appendix A: sample corpus

ሀለ ሀለለ ሀለወ ሀላዌ ሀመድ ሀምር ሀርበደ ሀሰለ ሀሴን ሀበበ ሀበባሊ ሀበወ ሀቡ ሀቢ ሀባቢ ሀብለት ሀብለየ ሀብተ ሀብት ሀንጻጳ ሀንጹጵ ሀከከ ሀከክ ሀከየ ሀካይ ሀኬት ሀዋኪ ሀውለየ ሀውል ሀውክ ሀየል ሀየየ ሀየደ ሀያዲ ሀያዲት ሀያጢ ሀይመነ ሀይማኒ ሀይከል ሀይድ ሀይዶ ሀደመ ሀደየ ሀድአ ሀገረ ሀገራዊ ሀገር ሀግረተ ሀግረት ሀነለ ሀነል ሀዣል ሀፈወ ሀፍ ሁከት ሂጣን ሂጳጳ ሃራራ ሃይማኖት ሄላ ሄርማ ሄሮድያኖስ ሄደ ሄጠ ሄጳ ሄጹጤን ሄጵዲያቆን ህላዌ ህልያን ህሳሊ ህሳሊት ህሳል ህብሉይ ህብልያ ህቦ ህንደኬ ህንጻጼ ህኩክ ህየ ህየንተ ህዩድ ህያይ ህዱእ ህድአት ሆሄ ሆበ ሆባይ ሆከ ለሀየ ለህሀ ለሊሁ ለሊሃ ለሊሆሙ ለሊሆን ለሊነ ለሊከ ለሊኪ ለሊኪን ለሊክሙ ለልየ ለሐመ ለሐሰ ለሐቀ ለሐኬ ለሐፀ ለሐፈ ለሓመ ለሓኩ ለሓኩት ለሓጺ ለሓጺት ለሓፂ ለሓፂት ለሕሐ ለመደ ለመጸ ለመጽ ለምለመ ለምሐ ለምንት ለምዐ ለምድ ለምጳኔ ለምፓ ለሮን ለሰነ ለሰደ ለስሐ ለስደ ለቀ ለቀሰ ለቀወ ለቀየ ለቀደ ለቃሒ ለቃሒት ለቅለቀ ለቅሕ ለቅዳ ለቆ ለበበ ለበነ ለበወ ለበየ ለበጠ ለባሲ ለባሲት ለባጢ ለባጢት ለቤሳት ለብሐ ለብሐዊ ለብሐዊት ለብሓ ለብሕ ለብሰ ለብን ለተመ ለተረ ለተተ ለታሪ ለታሪት ለትሐ ለጐሰ ለነ ለንዴዎን ለንጸሰ ለንጻሰ ለንጽ ለኞሳስ ለአትከለ ለአከ ለአፈ ለኣኪ ለኣኪት ለከ ለከወ ለከፈ ለኪ ለካኢ ለካኢት ለካፌ ለካፌት ለኬንያቂ ለክሙ ለክአ ለኰመ ለወለ ለወረ ለወሰ ለወወ ለወየ ለወገ ለውለወ ለውሕ ለውር ለውዝ ለውግ ለዐለ ለዐተ ለዘ ለዘረ ለዝለዘ ለዝሊዝ ለይ ለይቀለ ለይቅል ለገበ ለገነ ለገገ ሊጋዒ ለግለገ ግሊግ ለግነተ ለግዐ ለጐመ ለጐነ ለጠረ ለጠነ ለጥለጠ ለጥሊጥ ለጥሐ ለጥሕ ለጸቀ ለጽለጸ ለጽሊጽ ለጽላጺ ለፈቀ ለፈየ ለፈደ ለፈጸ ለፈጽ ለፈፈ ለፌ ለፓ ሉላዊ ሉል ሉቃስ ሉዓላዊ ሉዓሌ ሉያ ሊሉይ ሊሊት ሊል ሊሎ ሊሳ ሊቃኖስ ሊቅ ሊባ ሊባኖሳዊ ሊባኖስ ሊተ ሊትር ሊቶስጥራ ሊጣ ሊጠስ ሊጠን ላሀየ ላህም ላህብ ላህይ ላሐወ ላሓዊ ላሕ ላሕለሐ ላሕሎ ላሕም ሕይ ላሜዳ ላባ ላቲ ላእከ ላእክት ላከወ ላኰየ ላኳ ላዕላዊ ላዕላዕ ላዕላይ ላዕል ላዕልዐ ላይዳ ላጋዒት ላጲስ ላጸየ ላጻዩ ላጻዩት ላጽቂት ሌለየ ሌሊት ሌሐ ሌቀ ሌወወ ሌዊ ሌዋዊ ሌውቄን ሌጌዎን ልሁም ልሁብ ልሂቅ ልህመ ልህሰ ልህቀ ልህበ ልህድ ልላሒት ልሑም ልሑይ ልሑፅ ልሑፍ ልሔም ልሕሉሕ ልሕመ ልሕኩት ልሕየ ልሕደ ልሕጸ ልሕጽ ልሕፅ ልሙዕ ልሙድ ልማዓት ልማድ ልምሉም ልምላሜ ልሱሕ ልሱን ልሳነ ልሳናዊ ልሳናዊት ልሳን ልስሐት ልስሕ ልቀተ ልቁሕ ልቃሕ ልቅሶ ልቡና ልቡይ ልቡጥ ልባብ ልብ ልብሐት ልብሕ ልብሰት ልብስ ልብኔ ልብን ልብው ልብጠት ልቱት ልኡክ ልኡፍ ልኩእ ልኩፍ ልክው ልውር ልውስ ልውግ ልዑል ልዕል ልዕልና ልዩ ልደት ልገት ልጋግ ልግን ልጐት ልጓም ልጕን ልጡሕ ልጥር ልጹቅ ልጹይ ልፋፈ ልፋፋ ልፋፍ ሎሀ ሎለወ ሎሎ ሎሐ ሎሕ ሎሙ ሎሚን ሎሰ ሎቀ ሎቅሳ ሎተሪ ሎታፌ ሎዘ ሎዛ

Appendix B: MATLAB code for pronunciation detection:

Feature extraction sample

```
function sound_features=features(traindir, n)
 for i=1:n
    file = sprintf('%s%d.wav', traindir, i);
    disp(file);
    [nw, sr]=speechRate(file);%Speech rate
    nSig = file / max(abs(file)) %normalization
    eg = sum(nSig.^2);%energy
    mfps=pitchs(file);%pitch
    meanPich=mean(mfps);
    variancePitch=var(mfps);
    maxPitch=max(mfps);
    minPitch=min(mfps);
    mFormant=formant(file);%formant
    meanFormant=mean(mFormant);
    varianceFormat=var(mFormant);
    maxFormant=max(mFormant);
    cc=mfcc(file);%MFccs
    mf1=cc(:,1);
    mf2=cc(:,2);
    mf3=cc(:,3);
    mfcc1mean=mean(mf1);
    mfcc2mean=mean(mf2);
    mfcc3mean=mean(mf3);
    mfcc1variance=var(mf1);
    mfcc2variance=var(mf2);
    mfcc3variance=var(mf3);
    features(i,:)=[sr eg meanPich variancePitch maxPitch minPitch meanFormant varianceFormat
    maxFormant mfcc1mean mfcc1variance mfcc2mean mfcc2variance mfcc3mean mfcc3variance];
    sound_features= features;
     display(sound_features);
     save features.mat sound_features
 end
end
```

```
%% pitch detect use autocrosscorrelation
function mfps=pitchs(Sig)
[sig,fs]=audioread(Sig);
length(sig);
FrameDuration =length(sig)/fs;
%info=audioinfo(Sig);
%FrameLen = info.TotalSamples;
FrameLen =200;
Num_frame= length(sig);
FrameInc = 20;
win=hamming(FrameLen,'periodic');
length(win);
fSig=enframe(sig,win,FrameInc);%hamming(FrameLen,'periodic')
length(fSig);
for i=1:200 % frames in speech signal
x=fSig(i,:);%calculate autocorrelation for each frame
rxx=xcorr(x);
y=rxx(255:end);%define distance to first peak
[k,I]=max(y);%find the first peak, value and position
if I<30 % define the unvoice part
fps(i)=0;
elseif I>160
fps(i)=0;
else
fps(i)=fs/(I+14);%add the cut points back to
%give the right position to
%calculate the pitches
end
end
stem(fps)
ind=find(fps);%find those nonzero values in fps for calculation
mfps=fps(ind);
maxPitch=max(mfps);
```

```
% Geez Language pronunciation detection classifier using ANN
%import trinning data and target data
Training_Data ='ThesisTrainningData.csv';
Training_class ='ThesisTraningClass.csv';
inputs = importdata(Training_Data);
targets = importdata(Training_class);
% Create a Pattern Recognition Network
hiddenLayerSize = 15;
net = patternnet(hiddenLayerSize);
% Set up Division of Data for Training, Validation, Testing
net.divideParam.trainRatio = 80/100;
net.divideParam.valRatio = 10/100;
net.divideParam.testRatio = 10/100;
% Train the Network
[net,tr] = train(net,inputs,targets);
% Test the Network
outputs = net(inputs);
errors = gsubtract(targets,outputs);
performance = perform(net,targets,outputs)
% View the Network
view(net)
% Plots various plots.
% figure, plotperform(tr)
% figure,plottrainstate(tr)
% figure confussion
plotconfusion(targets,outputs)
```