2020-03-17

DESIGNING AUTOMATIC SPEECH

þÿR E C O G N I T I O N F O R G E E Z L A

ABEBE, TSEGAYE

Wisdom at the source of the Blue Nile

BAHIR DAR UNIVERSITY

BAHIR DAR INSTITUTE OF TECHNOLOGY

SCHOOL OF RESEARCH AND POSTGRADUATE STUDIES

FACULTY OF COMPUTING

DESIGNING AUTOMATIC SPEECH RECOGNITION FOR GE’EZ LANGUAGE

BY TSEGAYE ABEBE ALTAYE

BAHIR DAR, ETHIOPIA

July 19, 2018

DESIGNING AUTOMATIC SPEECH RECOGNITION FOR GE'EZ LANGUAGE

BY TSEGAYE ABEBE ALTAYE

A thesis submitted to the school of Research and Graduate Studies of Bahir Dar Institute of Technology, BDU in partial fulfillment of the requirements for the degree of
Master of Science in the Computer Science in Computing Faculty.

Advisor Name: Tesfa Tegegne Asfaw (PhD)

Bahir Dar, Ethiopia
July 19, 2018

DECLARATION

I, the undersigned, announce that this proposition includes my own particular work. In consistence with globally acknowledged practices, I have recognized and refereed all materials utilized as a part of this work.

Name of the student: Tsegaye Abebe Altaye Signature ______________

Date of submission: July 19, 2018

Place: Bahir Dar

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

Advisor's Name: Tesfa Tegegne Asfaw (PhD)

Advisor's Signature: ______________________________

Bahir Dar University

Bahir Dar Institute of Technology-

School of Research and Graduate Studies

Faculty of Computing

THESIS APPROVAL SHEET

Student:

Tsegaye Abebe ______________________________ 5/7/2018

Name Signature Date

The following graduate faculty members certify that this student has successfully presented the necessary written final thesis and oral presentation for partial fulfillment of the thesis requirements for the Degree of Master of Science in Computer Science

Approved By:

Advisor:

Dr. Tesfa Tegegne ______________________________ 5/7/2018

Name Signature Date

External Examiner:

Dr. Million Meshesha ______________________________ 4/7/2018

Name Signature Date

Internal Examiner:

Dr. Gebeyehu Belay ______________________________ 9/7/2018

Name Signature Date

Chair Holder:

Shankar ______________________________ 9/7/2018

Name Signature Date

Faculty Dean:

Dawed Nesru ______________________________ 9/7/2018

Name Signature Date

ACKNOWLEDGEMENTS

Above all else, I should want to express the profound gratitude to almighty God; He is the owner of the wisdom and helped me from the beginning up to the ending and thus again for His mother virgin Mary, who always came for the day of distress.

Next to my advisor Tesfa Tegegne (PhD) who followed up, supported, and guided me continuously throughout all processes and steps of the study. Following to this, I want to express my thanks and appreciation to Muluken Andualem (PhD) who is a scholar of Ge’ez and ‘Qəne’ supported me about the Ge’ez language with great motivation. Currently, he is a director of Abbay Research Center for Culture and Development at Bahir Dar University.

I also thank again to all Ge’ez readers who have shown me their interest and donated their participation during the recording of Ge’ez speech to develop the corpus. Consequently, I would like to describe my heartfelt gratitude to Bahir Dar University which is my sponsor and organizer of materials like computer, internet access, and office; as well as to faculty of computing which helped me in different issues for example by reducing the loads of courses, giving cooperation letter for collaborating of data collecting and so.

In addition to this, I wished to be delivered my excellence thanks for all my friends and colleagues who helped me by giving and sharing idea directly and indirectly. Finally, I would like to put my great appreciative and respective for my parents who have been patient on everything and who gave me support and moral on each and every day.

ABSTRACT

Language is a typical spiritual communication tool used by human beings in order to share or exchange their knowledge, skills, opinions wishes, commands, thanks, wisdoms and cultures to other people by speaking or using different ways. Currently, human beings are communicating with electronic devices using their speech with the help of automatic speech recognition system. Automatic speech recognition is the process of converting acoustic speech signals into its equivalent text form.

Researches on automatic speech recognition have done on foreign or local languages Amharic. Amharic and Ge’ez languages have redundant letters with the same sound. Researches in Amharic speech recognition are conducted by normalization of the repeated letters. Though, those redundant letters have different usage and meaning in Ge’ez language. Ge’ez language is a classical language of our country which is looking to speech recognition research. Ge’ez language has its own letters and numbering system. However, some letters have lost their sound and they are making confusion during the formation of words in its writing system. The aim of this study is to investigate the possibility of developing automatic speech recognition for Ge’ez language. In this study hidden Markov modeling technique is applied using sphinx 4 trainer. Since there is no recorded or prepared Ge’ez corpus, the investigators developed both text and speech corpora and among the developed corpus 4818 sentences for training and 433 sentences for testing used by selecting seven speakers’ Ge’ez audio randomly. Two experiments were performed using two different language models and two testing methods (online and offline) were performed for evaluation of the system. Both experiment 1 and experiment 2 have shown 90.88% and 68.48%-word accuracy rate respectively by testing with sphinx tool and the average word accuracy is 79.70%. As well as for testing the system using the developed interface with the same testing data the word accuracy is 67.79%. Homophones and hetero-phones in Ge’ez are challenges for speech recognition. In order to increase the accuracy of recognizer, maximizing the size of corpora is the future direction. Phone based, syllable based and gemination were the other future works.

Keywords: Automatic speech recognition, Ge’ez language, hidden Markov model, language model, acoustic model, offline testing, online testing

TABLE OF CONTENTS

LIST OF ABBREVATIONS

AD	Anno Domini
ANN	Artificial Neural Network
ASR	Automatic Speech Recognition
BC	Before Christ
BCE	Before the Common/Christian Era
CMN	Cepstral Mean Normalization
CMU	Carnegie Mellon University
CV	Consonant Vowel
DFT	Discrete Fourier Transform
EOC	Eritrean Orthodox Church
EOTC	Ethiopian Orthodox Tewahido Church
HTK	Hidden Markov Modeling Toolkit
HMM	Hidden Markov Model
LPCC	Linear Predictive Cepstral Coefficient
MFCC	Mel-Frequency Cepstral Coefficient
NLP	Natural Language processing
OOV	Out of Vocabulary
PLP	Perceptual Linear Prediction

SER	Sentence Error Rate
SLM	Statistical Language Modeling
SR	Speech Recognition
STT	Speech to Text
WAR	Word Accuracy Rate
WER	Word Error Rate

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

1. INTRODUCTION

This chapter introduced the general impression of the study. The chapter incudes sections starting from the background of the area of the study, statement of the problem, research questions, general and specific objective of the study, scope and limitation of the study, development tools and methods and significance of the study.

1.1. Background of the study

Language is a typical communication tool for the people to interact with one another, for transferring their history, knowledge, skills, beliefs, opinions, wishes, threats, commands, thanks, promises, and for reflecting their wisdom and culture for other worlds (Patnaik, 2016). Human beings can communicate in different ways for example by speaking, tuning in, making motions, utilizing specific hand signals (e.g. traffic laws), communications through signing for the hard of hearing individuals, or utilizing the types of script. By script implies that words that are composed or imprinted on a level surfaces called papers, cards as well as street signs or that are shown on a screen or electronic gadget with a specific end goal to be perused and understood by the people (Kibble, 2013). Because computers are notably influencing the way of human beings live and their usage is growing at an alarming rate. The interaction with a computer is a huge significance for anyone; currently due to the fact, input units for instance keyboard and mouse have boundaries for entering an input via oral communication.

Nowadays, human beings are communicating with computers or machines (human-computer interaction) by means of natural language i.e. speech[1] to access and retrieve information. Speech technology allows users a hand free communication with their devices. Therefore, developing a speech recognition in any language as a paramount importance. It means that, in the recent years, speech technology started to change the way of our life and work style and became one of the main ways for human beings to interact

[1] Speech is a sound pressure wave that must be converted to an electrical signal, and then a digital signal, to be processed (Picone Joseph, 2002).

with electronic devices (Deng, 2015). For this reason, a lot of researches have been conducted over the area of natural language processing (NLP) for analyzing and understanding the humans speech by computers to develop automatic speech recognition system (Bengio & Keshet, 2009).

Natural Language Processing (NLP) is a branch of computer science, artificial intelligence and linguistics which is involved with interaction among human languages and computers (Reshamwala, Mishra, & Pawar, 2013, Ann Copestake, 2003). NLP is an approach for processing and analyzing the natural language speech or text with the help of computers (Liddy, 2001).

NLP has many applications (sub-fields) such as speech recognition, spelling and grammar checking, optical character recognition (OCR), display screen readers for blind and partly sighted users, machine aided translation (i.e. systems that assist a human translator, for example by means of placing away interpretations of expressions and supplying on the web lexicons coordinated with word processors), information retrieval, file classification and clustering, data extraction, question answering, summarization, text segmentation, report generation, machine translation, natural language interfaces to databases, email understanding and dialogue systems (Ann Copestake, 2003), (Liddy, 2001). However, the consideration of this study is focused on the speech recognition part only.

Speech Recognition is a computer science field that deals about the design of computer systems that can recognize voiced words. It is also recognized as Automatic Speech Recognition (ASR) or only Speech To Text (STT) (Gupta, Pathak, & Saraf, 2014). Additionally, speech recognition can be described as the way toward changing over an acoustic or sound signal, taken by a receiver or portable device, to an arrangement of words (manuscript) (Cole et al., 1997).

Rabiner Lawrence & Juang, (1993) stated that, “Speech recognition is the multidisciplinary sub-field of computational linguistics which incorporates knowledge and research in the linguistics.”. By it mean multidisciplinary, it includes many disciplines during the developing time; which means that, effective speech recognition systems involve knowledge and expertise from a wide range of disciplines such as: signal processing which is the process of extracting relevant information from the speech signal, and linguistics which is the relationships between sounds (phonology), words in a language (syntax), meaning of spoken words (semantics), and sense derived from meaning (pragmatics).

Speech recognition has been done for many languages that are spoken across the world. Although, some countries have their own different domestic languages that are under-resourced languages or not. Among those, Ethiopia which is owner of Nations, Nationalities and Peoples, has many local languages such as Ge'ez that is looking to the study of automatic speech recognition. And since, the ultimate objective of speech recognition systems is that converting spoken words those taken via a microphone to their equivalent representation of written format (Himanshu & Kaur, 2014), the intention of this research is to develop speech to text conversion for Ge'ez language using Hidden Markov Model (HMM).

How automatic speech recognition works?

Automatic speech recognition, encompasses capturing the input utterances, digitizing those utterances, converting them to fundamental language units called phones or phonemes, building words from phonemes, and contextually analyzing the words to certify the correct spelling for those new words that sound comparable. The following figure 1 has shown the general process of automatic speech recognition.

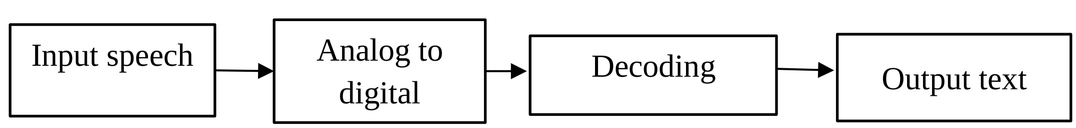


Figure 1: General process of automatic speech recognition

1.2. Challenges in automatic speech recognition

Because of the human language complexness and other reasons, developing automatic speech recognition with the help of a computer is a challenged task. There are various reasons that the development of ASR to be difficult. In the first place, there is repetitive data in the acoustic signal that is not valuable to segregate between classes of a picked utterance unit like words or phonemes. the second reason is that, the presence of interferences in acoustic signal raised from different sources for instance, echo, environmental noise, type of microphones and mutilated acoustics. The other issue is that, there is fleeting and recurrence changeability which incorporate intra-speaker fluctuation in pronunciation and between speaker inconstancy like local dialects. Another important issue is disfluencies in speech, for example usual speech is filled with repetitions, hesitations, subject changing in the middle of an utterance, slips of the tongue.

Why automatic speech recognition is needed?

As it is known, the goal of automatic speech recognin is to provide efficient way of interacting humans with their electronic devices. Hence, automatic speech recognition is an alternate way that has a capability to substitute the traditional methods for interacting with a computer, such as text input through a keyboard (Gupta et al., 2014). An effective speech recognition system can replace the use of mouse and keyboards input. This also leads to assist the People with less experience and skills about keyboard. Because speech recognition is three to four times faster than typewriters and eight to ten times faster than handwriting to insert information to computers (Sadaoki Furui, 2001). As well as, it can support physical disabilities people that worried to enter data to a computer and illiterate People who are unable to write and read (e.g. use of Telephone or mobile system).

1.3. Problem Statement

Speech recognition technology is a useful technology to interact with electronic devices like computers using natural languages instead of input devices in order to enter or access any information to/from those electronic devices. A number of studies are conducted on speech technology over the world as well as in Ethiopia for different languages like Amharic. For example, Teferra & Menzel, (2007) and Gebremedhin et al., (2013) developed syllable-based speech recognition using HMM. Amharic and Ge'ez languages have redundant letters with the same sound. The above two and other researchers conducted Amharic speech recognition by normalization of the repeated letters; which means that they avoided those paired letters and used only one of them. However, those redundant letters have different usage and meaning in Ge’ez language. As a result, the variety of natural languages is the main problem for the development of automatic speech recognition system; because the developed speech recognition for a given language cannot work to other language; it works only for that targeted language. This guides to design and develop a speech recognition system to specific language; for example, Ge’ez language. Ethiopia, which is multilingual country, has many script resources those were written in Ge’ez language like the Bible, liturgical literature, theological scripts, magical texts, stories of martyrs and saints, religious poetry, hymns in honor of Christ, the Virgin, the martyrs, the saints, and angels, as well as secular writings (histories and romances, books of law, chronicles, mathematical, and medical texts) (Leslau, 1991) and those sources are located in different places (e.g. in libraries, monasteries, sanctuaries, personal hands). Those scripts are precious heritages and references for current generation to know the background of the country. It is impossible to use those scripts without knowing the Ge'ez language. Hence, understanding, knowing and investigating of Ge’ez language is a mandatory with the help of speech technology.

The study of Ge’ez language in our country is still in a traditional way even so which is offered mostly at the North Ethiopia except a few schools (primary and secondary), and colleges those are under EOTC. So, the Ge’ez language cannot grow and expand as long as it needed and it is not merged or joined to the information technology’s domain. Because the information communication technology is a real instrument to study and develop any language in this period. Generally, Ge’ez language is one of the Semitic languages that are looking for technological considerations of researchers in the area of ASR. Hence, designing and developing speech recognition in the case of Ge’ez language is possible in order to address the gap and the problems mentioned above.

1.4. Research Questions

This study has tried to answer the following questions.

- What are main features of Ge’ez language?
- What are the challenges in the designing of ASR for Ge’ez language?
- How the accuracy of ASR model for Ge'ez language?

1.5. Objective of the study

General objective: The general objective of this study is to develop an automatic speech recognition for Ge’ez language.

Specific objectives: To accomplish the general objective of this study, the following specific objectives of the study are included to:

- understand and formulate the Ge’ez language with its characteristics.
- identify the challenges in the development of Ge’ez ASR
- build a Ge’ez language speech and text corpus for training and testing purpose
- build language model which holds n-gram words with probabilities of their occurrence in a sentence
- build pronunciation dictionary that contains lists of words that can be recognized by the ASR system and their pronunciations (phoneme sequences they consisted).
- build acoustic model that contains the properties and knowledge about acoustics, phonetics, environmental variation, different pronunciations and differences in dialect of speakers.
- develop a prototype for ASR of Ge’ez language.
- test the developed system in order to evaluate its accuracy and performance.

1.6. Scope and limitation of the study

The main focus of the researcher is to explore the possibility of developing a prototype of speaker independent speech recognizer for Ge’ez language. The recognizer is designed to recognize Ge’ez words and Ge’ez numerals. The approach we used to conduct the research is stochastic or statistical method using HMM model. The size of the developed Ge’ez text corpus is 5251 sentences and the length of speech corpus is 13.31 hours long only. The limitation of this study is that the unbalanced gender ratio in the participation of speech corpus preparation. It means that the ratio of females and male speakers is 14% to 86% respectively. However, we covered different age group and environments like nose.

1.7. Research Methodology

Rajasekar & Philominathan, (2013) stated that "research methods are the various procedures, schemes and algorithms used in a research to be conducted". They help the investigators to gather samples, data and find a suitable solution to a problem.

On the other word, research methodology is the way to solve a given problem systematically. It is a discipline of studying in what way research is to be carried out scientifically (Rajasekar & Philominathan, 2013). It means that the research methodology is the backbone for any research in order to accomplish the tasks those are mandatory for any study. In order to satisfy the objectives of this research, the study followed the experimental research approach. Experimental or empirical research is a procedure to contribute to the previously existed knowledge (Singh, 2006). It is a types of research that depends on gathered information thinking of conclusions which are equipped for being checked by making experiment on the data (Kothari, 2004). Hereafter, the following methods and tools were included during the process of this study and below in Figure 2 all those research methods and procedures have been shown for the clarity and ease of understanding.

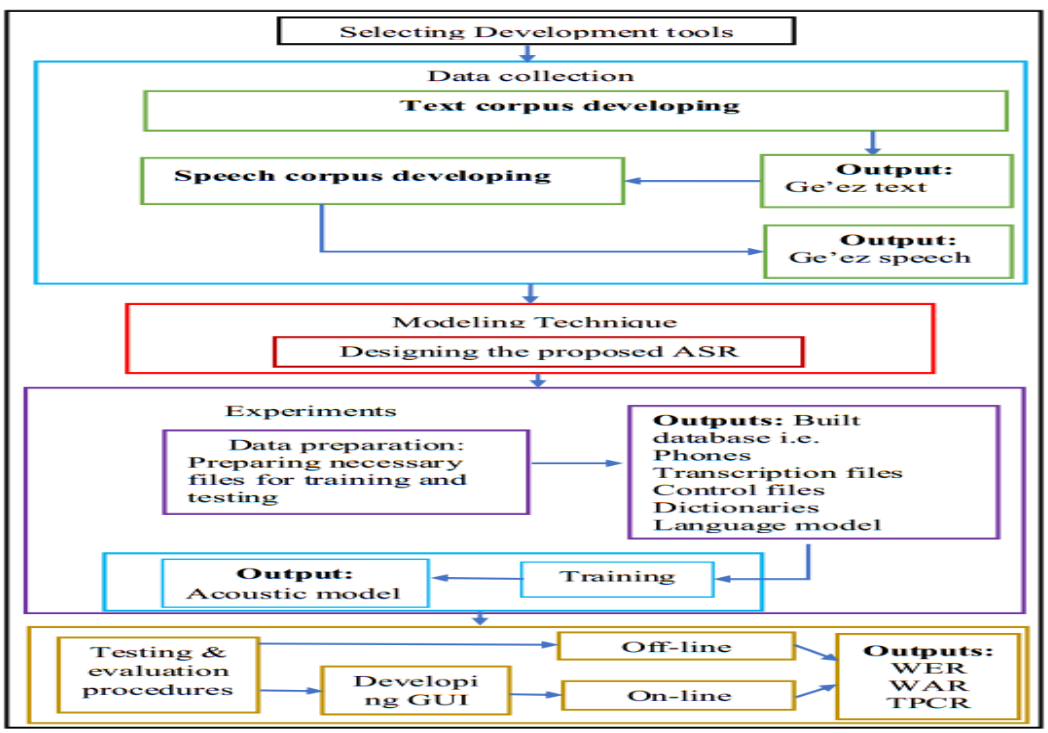


Figure 2: An over view of research methodology and procedures

1.7.1. Literature reviewing

The literature review is a mechanism to investigate, understand and assemble knowledge about the area of study, and to know the past related works. So, to understand techniques or approaches, models and experiences in automatic speech recognition we used books written by domestic and foreign scholars, magazines. As well as we discussed with Ge'ez national scholars and previous researchers.

1.7.2. Development Tools

As a development tool, the CMU-Sphinx[2] is used for training the acoustic model because CMU-Sphinx supports Unicode/UTF-8 format (Solomon, Yifiru, & Abebe, 2015) and (Ghai & Singh, 2012). Since there is a latest version of Sphinx series of speech recognizer tools, the researchers used Sphinx-4 version which is one of the most widely used open source speech recognition toolkit developed at Carnegie Mellon University (CMU). Since Sphinx-4 is developed with java programming language, it is platform independent. The 'CMU-Cam_Toolkit_v2' language modeling tool used for the developing the N-gram language model. The Audacity software which is free, open source software available with latest version of 2.1 used for speech corpus processing (i.e. recording or/and editing sounds), Audio Silence Trimmer Pro for trimming silences from the beginning and ending of each record of sentence. Furthermore, java programming language for implementation, Notepad++ is applied for text corpus preprocessing at different level, Microsoft office word-2016 for document processing, Mendeley-Desktop-1.17.8-win32 for the citation of the referenced materials.

1.7.3. Data collection

Speech and text data are needed for both training and testing processes. In this study, all the speech data is recorded from scratch since there is no speech corpus for Ge'ez language previously as well as the text corpus is collected using different Ge'ez documents. See the next section for more detail information.

[2] Sphinx stands for Site-oriented Processor for HTML INformation eXtraction is one of speech recognition engine that developed at Carnegie Mellon University (CMU) and the Sphinx works based on Hidden Markov Model (HMM) algorithm (Dewi, Firdausillah, & Supriyanto, 2013).

1.7.4. Modeling Technique

The modeling technique used in our study was Hidden Markov Model (HMM) for speech recognition development process that widely used for modeling the temporal speech signal in the acoustic model. The motivation behind why HMM prevalent is on account of it can be prepared naturally and is basic and computationally plausible to utilize (Trivedi, 2013), (Rabiner Lawrence & Juang, 1993). It means that it provides natural and highly consistent way of recognizing speech and the required amount of computation in the HMM method is minimum to others. An HMM has the flexibility of the outcoming recognition system wherever one can simply change the type, size, or model architecture to fit specific words or any sub-word unit (Teferra & Menzel, 2007).

1.7.5. Testing and Evaluation Procedure

To measure the performance of an automatic speech recognition system, accuracy and speed are the basic standards (Ghai & Singh, 2012). The accuracy is measured using WER method which is computed comparing the test data set (reference) to the new output (hypothesis) and then counting the number of substitutions (S), deletions(D), and insertions (I), dividing by the total number of words in the test set and multiplying with 100. The speed of ASR is measured with the aspect of real time factor. It means that real time factor is defined as the ratio between the time it taken by the process of the input and the duration time of the input (Ivanov et al., 2016). However, for this study only accuracy (using word error rate metric) was applied as a measurement of ASR accuracy.

Actually, there are two types of testing mechanisms for automatic speech recognition system; namely off-line and on-line (Chan et al., 2007). On-line evaluation is performed by the speakers directly using speech recognition interface and it needed to develop a prototype interface for calling the recognizer system. The off-line evaluation is done through CMU decoder using pre-recorded audio and text for testing purpose. And this testing mathematically expressed as:

$$WER = \frac{S + I + D}{N} * 100 \qquad (1)$$

And the system accuracy is computed as:

$WAR = \frac{(N - (S+I+D))}{N} * 100$ *, then by substituting equation (1) and by simplifying we get:*

$$WAR = (1 - \left(\frac{WER}{100}\right)) * 100 \qquad (2)$$

Finally, the total percent rate of correctly recognized words is calculated as:

$$TPCR = \frac{(N - (S+D))}{N} * 100$$

Where: WER = word error rate,

WAR = word accuracy rate,

TPCR =total percent correct rate,

N = total number of words occurred in testing data (the reference),

S = number of errors caused by substitution of words,

I = number of errors caused by insertion of words, and

D = number of errors caused by deletion of words.

1.8. Significance of the study

Generally, the tangible significances of this research are the following:

The study is a pioneer for the other NLP researches for example development of the dictation system, speech translation for Ge'ez language. Similarly, the developed text and speech corpora are very useful to any researcher who may wish to endeavor for developing automatic speech recognition of Ge'ez language and other NLP researches.

Since Speech technology is the technology of today and tomorrow, the results of this research can help many scholars of Ge'ez language, traditional students and other Ge'ez speakers to take advantage of many benefits of ICT's ideology. Hence the desired system would help the teaching learning process of Ge'ez language for both modern and tradition students using the electronic devices like computers and projectors to view texts at real time. For example, in EOTC teaching and learning of Ge'ez grammar, all processes are done orally; specially during the production of verbs and creating of poetry. During this process, ASR is used to convert and save those spoken words and poetries using microphone and computer.

Besides, in the Ethiopian Tewahdo churches, at least two poems are presented or spoken orally in Ge'ez language on every Sunday sermon. These poems, however, contain many religious, historical, social, economic, political views and favorite linguistic arts, but is not saved and transmitted or communicated for the society. So, the study is used to solve this problem by capturing and saving those oral poetries in to electronic devices. Furthermore, the study is used for retrieving information from Ge'ez text that is electronically stored and for browsing webpages those contained data in Ge'ez language.

1.9. Organization of the Thesis

The remainder of this thesis is presented as follows. Chapter 2 describes the literature review about speech recognition, components of SR, types of speech recognition, approaches for ASR and related works. Chapter 3 discuses about Ge'ez language writing system, syllable structure, phonetics, phonology, morphology and syntax. Chapter 4 discuses on designing and model Ge'ez speech recognition system, research methodology, reviewing related works, development tools, data collection, developing text and speech corpora. It includes modeling techniques, ASR model for Ge'ez language, data preparation, feature extraction, training, testing and developing an interface for Ge'ez language. Chapter 5 includes results and discussions. And the final chapter contains conclusion and recommendations.

CHAPTER TWO

2. LITERATURE REVIEW

This chapter intends to discuss about general existing knowledge of ASR such as basic components of ASR, approaches, Hidden Markov Models and types of field of automatic speech recognition. And review the previous studies or related works of automatic speech recognition in order to show the strengths, gaps, limitations of the conducted works over the ASR.

2.1. Overview of Automatic Speech Recognition

The concept of speech recognition started in the early 1900s. Hence, the machine called Radio Rex that can recognized the speech was created in1922 by Elmwood Button Company (Jurafsky & Martin, 2007), (Gold, Morgan, & Ellis, 2011) which was the first success story in the field of speech recognition. The Radio Rex was brown bulldog that came out of its dog-house when it heard its name. Homer Dudley invented the vocoder (stands for 'voice coder') at Bell Labs in New Jersey in 1928, which was the first machine that could generate human speech electronically when a person entered the words into a special keyboard (Gold, et al., 2011), (Study.com, 2016). Although much of the work in voice-coding and related speech analysis in the 1930s and 1940s was relevant to speech recognition, the first speech recognition system which recognizes the spoken digits was built in 1952 at the bell labs, by Davis, Biddulph, and Balashe that was about only recognition of digits (Rabiner Lawrence & Juang, 1993), (Davis, Biddulph, & Balashe, 1952). This effort for automatic recognition of speech was essentially focused on the working up of an electronic circuit for recognizing ten digits of telephone quality (Ghai & Singh, 2012). The approaches to speech recognition, evolved thereafter, had a major stress on searching speech sounds and providing suitable labels for those speech sounds. Various approaches and types of speech recognition schemes came into existence in last five decades (i.e. during that time) gradually. This evolution has led to a remarkable impact on the development of speech recognition systems for various languages across the world. However, to cover the whole speech recognition history is outside of scope of this study. Automatic speech recognition has been viewed as successive transformations of audio micro structure of speech signal into its implicit phonetic macro structure. As a general, it means that speech recognition system is a speech to script transformation where in the yield of the framework shows the content (text) comparing to the perceived speech.

2.1.1. Components of Automatic Speech Recognition

Speech recognition is the strategy of consequently extricating and discovering linguistic data conveyed via a speech-wave [3] by applying electronic tools called computers. Linguistic information additionally referred to as phonetic information which is the most substantial data in a speech wave (Sadaoki Furui, 2001). As a general explanation, speech recognition encompasses catching and digitizing the sound waves as input, changing them to the primary language units or phonemes, developing words from phonemes, and relevantly examining the words to ensure the correct spelling for words for phrases that sound alike. Largely, ASR is the procedure of automatically recognizing what the speaker is said and presenting the recognized speech in the form of text. In order to understand speech technology, it is necessary to understand the following ASR sources:

Voice: This is an input datum to a system with the help of audio microphone; the sound card of computers produces the corresponding digital representation for the established audio.

Digitization: is the process of converting the analog audio signal into the form of digital representation. It involves both processes of quantization and sampling. Quantization is the process of approximately representing continuous range of waveform values. Sampling is the process of converting continuous speech signal into discrete or distinct audio signal.

Acoustic Model: is one of the fundamental segments of ASR and used to interface the viewed features of utterance signals with the normal hypothesis sentence phonetics.

Pronunciation dictionary: is a plotting table that maps words into their sequences of set of phonemes (their pronunciation in the grapheme form) and required for both training and

[3] "The speech wave conveys several kinds of information, which consists principally of linguistic information that indicates the meaning the speaker wishes to impart, individual information representing who is speaking, and emotional information depicting the emotion of the speaker." (Sadaoki Furui, 2001)

decoding phases. It is designed to contain only unique set of words (without any repetition vocabulary) that existed in the designed corpus and delivers pronunciation of these vocabularies for the transcription file via the phoneme set. Basically, vocabularies are categorized in to two types namely: closed and open vocabularies. The closed vocabulary contains all unique words in a text corpus, and so the test set contains words from this vocabulary only (i.e. it does not contain unknown words or OOV). Whereas, open vocabulary contains words from corpus and/or without corpus. Thus, the test contains the unknown words and used pseudo word <UNK> to represent the unknown words (Jurafsky & Martin, 2007).

Pronunciation dictionaries also divided in to: canonical pronunciation dictionary which provides a single pronunciation (using standard phoneme to represent) for each word without considering variation of the pronunciation and realized or alternative pronunciation dictionary that contains all alternative pronunciations (using actual phoneme for representation) those are supposed to be pronounced for each word by considering pronunciation variations of different speakers, dialects or coarticulations (Fukada, Yoshimura, & Sagisaka, 1999). Pronunciation dictionary assists as a transitional among the acoustic and language model in the decoding phase.

Language Model: It is the component of ASR that contains a representation of probability of occurrence of words in a sentence. It is used in many NLP applications such as speech recognition and machine translation and it attempts to capture the language properties and to predict the following word in speech sequence.

Speech Engine/Decoder: is a module of ASR used to change an input audio data into text and to complete this, it utilizes all data, algorithms and statistic model. As discussed earlier, the first operation of the decoder is digitization for converting the input to a digital format for additional processing. Then it searches best match by considering words it knows, after the signal is recognized, it yields its matching text.

2.1.2. Types of Automatic Speech Recognition

Speech recognition is classified into different classes based on speaker model, vocabulary, channel and utterance type in their capability to recognize (Agrawal & Raikwar, 2016).

2.1.2.1. Types of ASR based on Speech Utterance

An utterance is a bit of spoken language or the word vocalization that denote a single meaning to the machine. “Utterances might be a single word, a few words, a sentence, or even multiple sentences” (Vimala, 2012).

a) Isolated/ discrete Words

Isolated word recognizers typically need a silent between each and every utterance on each aspects of sample window. Meaning that, within a time, it needs single utterance only. It is popular for recognizing digits, commands and one-word response. It is easy to implement because the boundaries of words are isolated and pronounced clearly (Vimala, 2012).

b) Connected Words

Connected word speech recognition is a type of ASR system that the words are separated by using pauses. Connected word speech recognition is a class of fluent speech strings where the set of strings is derived from small-to-medium size vocabulary for example digit strings, spelled letter sequences, combination of alphanumeric (Arora & Singh, 2012).

c) Continuous/ Fluent Speech

Continuous speech recognition offers with the speech in which words or phrases are linked together rather of being separated via pauses. Consequently, unidentified boundary facts about co-articulation, words, production of adjacent phonemes and step of speech impact the continuous speech performance awareness systems. Recognizers with non-stop speech abilities are difficult to create due to the fact that they make use of different methods to decide utterance boundaries (Arora & Singh, 2012). Continuous word schemes cannot represent all feasible inputs, however have to collect patterns for minor speech actions (e.g. words) into higher sequences (e.g. sentences).

d) Spontaneous Speech

A spontaneous speech is a speech which is natural sounding and no longer learned. Spontaneous speech recognition structures allow the possibility of pause and false starts in the utterance, the utilization of words not found in the lexicon, etc. So, the speech like this is natural and difficult to learn.

e) Read Speech

A read-speech deals with the speech where speakers read sentences from constructed or prepared text corpus but which now included a component that involved speaker-independent recognition (Jurafsky & Martin, 2007).

2.1.2.2. Types of ASR based on Speaker Model

All speakers have their specific voices, regarding to their unique personality and physical body. Hence, speech recognition system can roughly categorize into speaker dependent, independent and adaptation based on model of speaker.

i) Speaker dependent models

Speaker dependent systems are designed to accept or listen a particular speaker. They have more accurate for a specific speaker and less accurate to different speakers; as well as they are simpler to develop, inexpensive and better with accuracy; but the developed system be dependable system because it adapted only one person.

ii) Speaker independent models

Speaker independent systems are types of ASR systems that designed for different speakers. It can recognize the speech come from a different group of people. The system is difficult and expensive to develop and the accuracy is less than speaker dependent;

because different speakers have different speech characteristics and speech parameters representation is dependable on the speakers (Lee, Reddy, & Allen, 1989).

iii) Speaker Adaptive Models

Speaker adaptive speech recognition system practices on the data of speaker dependent. It has the ability to adapt to the best suitable speaker for recognizing the speech and their error rate decreases by adaption the operation based on characteristics of speakers.

2.1.2.3. Types of ASR based on size of Vocabulary

The size of the vocabulary for a speech recognition system can affect the processing requirements, complexity and speech recognition accuracy. Different applications may need different size of word vocabularies: small, medium, large or very large. According to (Saksamudre, Shrishrimal, & Deshmukh, 2015), Small Vocabulary is limited with size of 1 to 100 words, Medium Vocabulary is limited with size of 101 to 1000 words, Large Vocabulary and Very-large vocabulary is limited with size of 1001 to 10,000 words and more than 10,000 words respectively. However, according to (Jurafsky & Martin, 2007), the number of large-vocabulary in the systems is roughly 20,000 to 60,000 words.

2.1.3. Approaches for ASR

According to (Rabiner Lawrence & Juang, 1993), approaches to automatic speech recognition are categorized into three approaches, namely: Acoustic-phonetic, Pattern recognition, and Artificial intelligence approaches.

2.1.3.1. The acoustic-phonetic approach

This approach works based on acoustic phonetics theory that guesses the finite and typical phonetic units in a given language and phonetic units are categorized by group of properties that are obvious in speech signal or its spectrum. In this approach, segmentation at phonetic level is the first step and then attempting to identify real word from a sequence of phonetics segmented previously with the help of dictionary.

2.1.3.2. The Pattern recognition approach

Basically, the pattern-recognition method to speech recognition is one in which speech patterns are used at once except specific feature determination (i.e. in the acoustic phonetic sense) and segmentation. In pattern recognition approach, first the system is trained with utterances which act as reference patterns and then the unknown utterances are compared to these references to know their identity. Generally, this approach includes two essential steps namely, speech patterns training, and those patterns recognition through the comparison of patterns (Rabiner Lawrence & Juang, 1993).

The feature for this approach is that it utilizes mathematical framework and founds constant representations of speech pattern, for comparison of pattern, from set of training samples through training algorithm. The representation of speech pattern can be in the form of speech template otherwise statistical model (e.g. HMM) and can be implemented to sound (may smaller than word), word, or phrase. During the pattern comparison approach stage, the comparison done between unrecognized speeches with every all likelihoods pattern that is learned during training stage in order to decide the identification of the unknown test and class reference pattern based totally on the sample match. There are two methods in this approach called template method and stochastic model method (Arora & Singh, 2012).

A) Template Method

In this method, the unknown speech is compared with set of templates or patterns in order to search best match among all. A collection of typical speech patterns is kept as reference patterns by representing a candidate words dictionary. At this point recognition is performed with the aid of matching an unrecognized utterance with every one of the reference formats and choosing the best category of coordinating pattern. This technique provides good recognition performance for different practical applications. Be that as it may, its drawback that varieties in speech can be demonstrated by utilizing numerous templates per word, which at long last ends up impracticable.

B) Stochastic method

Stochastic techniques delineate the utilization of probabilistic models to manage inadequate data. In speech recognition, inadequacy/uncertainty emerges from different sources for example, speaker fluctuation, confusable sounds, homophones or hetero-phones words and logical impact/contextual effects; and this inadequacy information deals with HMM (Ghai & Singh, 2012). Nowadays the most common stochastic or probabilistic approach is Hidden Markov Modelling (HMM).

2.1.3.3. The Artificial Intelligence Approach

This approach is a combination of two approaches called acoustic phonetic and pattern recognition approaches (Rabiner Lawrence & Juang, 1993). The data respects to spectrogram, linguistic as well as phonetic is used by knowledge-based (artificial intelligence) approach. The fundamental concept is to arrange and coordinate information from different collections of knowledge sources (acoustic knowledge[4], lexical knowledge[5], phonemic knowledge) and to take it to put up on the problem.

2.2. Related works

A lot of works have been done on the development of Speech Recognition systems for various languages through the world. The following is a brief review of the work done on Automatic Speech Recognition Systems for the selected languages and approaches.

2.2.1. Automatic speech recognition for Arabic

Arabic language is family of Semitic language that has 34 basic phonemes, of which six are vowels, and 28 are consonants and the phonemes have two classes namely pharyngeal and emphatic phonemes which are found in Semitic languages only. (Mostafa, Tolba, Mahdy, & Fashal, 2008) developed the syllable-based automatic Arabic speech recognition for Egyptian Arabic speech using syllables. One important factor that supports the use of syllables as the acoustic unit for recognition is the relative insulation of syllable from

[4] "Acoustic knowledge- is evidence of which sounds (predefined phonetic units) are spoken on the basis of spectral measurements and presence or absence of features." (Rabiner Lawrence & Juang, 1993)
[5] "Lexical knowledge- is the combination of acoustic evidence so as to postulate words as specified by a lexicon that maps sounds into words (or equivalently decomposes words into sounds)." (Rabiner Lawrence & Juang, 1993)

pronunciation variations arising from addition and deletion of phonemes as well as co-articulation. Speaker-independent Hidden Markov Models (HMMs)-based speech recognition system was designed using Hidden Markov model toolkit (HTK). The database used for both training and testing consists from 44 Egyptian speakers (22 for training and 22 for testing). Experiments show that the recognition rate using syllables over the rate obtained using mono-phones, triphones and words by 2.68%, 1.19% and 1.79% respectively. A syllable unit spans a longer time frame, typically three phones, thereby offering a more parsimonious framework for modeling pronunciation variation in spontaneous speech. Moreover, syllable-based recognition has relatively smaller number of used units and runs faster than word-based recognition.

Generally, according to the researchers' report, the researchers concluded that the selected mono-phone-based, triphone based, word-based and syllable-based recognition rate is 90.75%, 92.24%, 91.64% and 93.43% respectively using 5-states of HMM-based but at 13-states of HMM-based, the recognition rate is 97.01%. The syllable-based system is the highest recognition rate using 5-states of HMM-based. Although word-based recognition rate in 13-states is higher than syllable-based recognition rate in 5-states, but syllable-based recognition is preferred because it has relatively smaller number of used units (syllables) and runs faster than word-based recognition.

2.2.2. Automatic speech recognition for Afaan Oromo

Gelana, (2010) tried the possibility of developing Afaan Oromo continuous, speaker independent speech recognizer using HMM and sphinx system. The research work for 70 selected Afaan Oromo long words, phrase and simple sentence constituting of 2100 utterances that uttered by 30 selected peoples those are in different age group (<15, 16-30, 31-45) and sex. The collected data was divided in the ratio of 2/3 by 1/3 for training and testing purpose respectively. He constructed two types models namely, Context dependent model that takes the entire words, phrases and sentences for the dictionary and other requirements and context independent model which is directly related to phoneme distance measures and triphone based. As well as the classification and other preparation of the training and test data sets were performed manually. Whereas the training received from the sphinx train was taken to the decoder and the task of measuring the performance is done automatically using the appropriate scripts for the selected words, phrases and sentence to check the applicability of the recognizer by evaluating its performance.

The performance evaluation is performed on two models using test data sets and the recognizer performance is 68.514% with sentence accuracy of 28% for context dependent model (continuous Afaan Oromo speech) and 89.459% with sentence accuracy of 42% for context independent (a phoneme-based trigram). So, the result obtained from the context independent which is a phoneme based has shown a good result for both recognizer performance and sentence accuracy (89.45%, 42%) and indicated the context independent phoneme level by far better than the context dependent.

2.2.3. Automatic speech recognition for Amharic

As it mentioned above in the overview of ASR part, the research's idea of Automatic Speech Recognition emerged in the early1900s across the world. However, in our country, Amharic speech recognition was started in 2001 by Solomon Berhanu which developed an isolated Consonant-Vowel syllable recognition system utilizing the HTK (Hidden-Markov Modeling Toolkit). As per Abate et al., (2009), he designated 41 CV syllables of Amharic language out of 234. Speech data of the selected CV syllables has been recorded from eight speakers (with equal gender ratio) with the age scope of 20 to 33 years. The achieved average recognition accuracy was for speaker dependent 87.68% and 72.75% for speaker independent systems. As his conclusion, the result was low compared to other speech recognition studies for other languages. This might be due to the recording environment problems and lacking of training data.

Teferra & Menzel, (2007) developed syllable-based speech recognition for Amharic via HMM as a model and CV syllables as recognition units. They used bigram language model using HTK development toolkit. They included both types of pronunciation dictionaries called canonical or alternative dictionary with 50,000 and 25,000 words respectively. The training is performed using Baum-Welch re-estimation procedure and the re-estimation is achieved in bootstrapping and flat start approaches. However, since, the bootstrapping approach didn't overcome error of labeler which affects the performance of model, they followed the flat start initialization method. In order to solve the problem of HMM when working with insufficient speech corpus, they applied variety of sharing mechanisms and used diagonal covariance matrices to perform a good re-estimation of model components from limited training data. The HMM topology is left to right with and without jumps and skips (to solve irregular occurrence of six order vowel and glottal stop consonant) including different number of emitting states and Gaussian mixtures (to assign them for different CV syllables based on their frequency number).

According to their report, the model with 5 emitting states, 12 Gaussian mixtures, with no skips and jumps shows that best word recognition accuracy (89.80%) and its memory requirement is less than other recognizers. The other model with 11 emitting states, including skip and hybrid Gaussian mixtures shows that 89.21% for word recognition accuracy and requires high memory space. Using the same corpus, they also developed word internal triphone based recognizer and the model which has 3 emitting states, 12 Gaussian mixtures including skips. Its word recognition accuracy (91.31%) is best from others triphone based recognizers previously developed by the researchers and syllable based (with 90.43% to word recognition accuracy). However, in terms of memory usage the triphone based recognizer asks more than syllable based. At a reverse, processing speed of syllable based was 37% faster than triphone based. As a result, they summarized that their used pronunciation dictionary does not solve problem of gemination of consonants, irregular occurrence of sixth order vowel and glottal stop consonant.

Gebremedhin et al., (2013) forwarded “A new approach to develop a syllable based, continuous Amharic speech recognizer” to prove that a smaller number of acoustic models are sufficient to build a syllable based, speaker independent, continuous Amharic ASR and they use the UASR (Unified Approach to Speech Synthesis and Recognition) Tool kit. The grammar was performed with finite state transducers instead of language model. Their speech corpora were new speech corpus and previous researchers’ corpus. In the new speech corpus 50 speakers (29 males and 21 females) were participated. However, the other corpora were obtained from previous researcher Solomon and Radio stations (Deutsche-Welle and Voice of America radio) included 91speakers (48 males and 43 females). The speech corpora were recorded in three different atmospheres to make the recognizer less sensitive for recording environment and microphone changes and the time duration is more than 35 hours.

The experiments revealed that a left-right structure with 5 states (for longer syllables) and 3 states (for shorter syllables) per HMM and acoustic models for only 93 syllables were trained. And the speech recognizer is tested with a data set that has 4,000 words collected from 10 speakers (4 males and 6 females). According to the report, their new approach reduces size of the acoustic models by training a common acoustic model for similarly pronounced syllables (e.g. syllables pronounced with vowels [a, A, and e], [u and o], [i and E]/ for example [ሀ፡ ሃ፡ ሀ], [ሁ፡ ሆ], [ሂ፡ ሄ]) were combined respectively.

Commonly, the recognition accuracy is 93.26% using smaller number of acoustic models. And it is possible to perform a recognition task on different applications without retraining the acoustic models on a new database; the database has all the syllables in a fairly similar proportion.

2.2.4. Automatic speech recognition for Ge’ez

The study of ASR is not conducted on Ge’ez language without “designing a stemmer for Ge’ez text” by Belay, (2010), “design and implementation of automatic morphological analyzer for Ge’ez verbs” by Weldegiorgis, (2010) and “Morphological Analysis of Ge’ez Verbs Using Memory Based Learning” by Abate, (2014). So, this circumstance invites the researcher to study an automatic speech recognition depending on the Ge’ez language.

To conclude that, as we have understood from above reviews, researches conducted by Mostafa et al., (2008), Gelana, (2010), and Solomon Berhanu were with small amount of corpus. Solomon Berhanu and Gebremedhin et al., (2013) used only 41 and 93 CV syllables respectively. However, our study involved all Ge’ez phones and all numerals. Teferra & Menzel, (2007) conducted with large amount of data related to our corpora but they developed only bi-gram language model. All those researchers directed their ASR study for Arabic, Afaan Oromo and Amharic languages. This phenomenon has shown that there is a gap for conducting automatic speech recognition study for Ge’ez language. Hereafter, the researchers have tried to fill this gap related to Ge’ez language.

CHAPTER THREE

3. THE GE’EZ LANGUAGE

Ge’ez language is the wealth of our country that reflects the religious, social, political, historical aspects of our country to the world and it seeks many studies. “ማእረሩሰ ለግዕዝ ብዙኅ ወፍድፉድ ወገባሩ ኅዳጥ ወውኁድ ሰአልዎ እንከ ኦ አርዳእ ለበዓለ ማእረር እግዚእ ከመ ይወስክ ካዕበ እምድሩ ገባረ ለማእረሩ” (Kidanewold, 1948) “The harvest (knowledge) of Ge'ez is so numerous, but the laborers/staffs of Ge’ez are very few and too small. Therefore, Disciples, please! ask the owner and lord of Ge’ez harvest to add more laborers or staffs for the development of Ge'ez harvest.”

This chapter discuses about the general overview of Ge’ez language. It describes the background of the Ge’ez language its letters, numbers (writing system), as well as its linguistics properties such as syllable structures, phonetics, phonology, morphology and syntax.

3.1. Ge’ez language (ልሳነ ግዕዝ)

The Ge’ez is a classical language of Ethiopia which belongs to the family of Semitic language (Leslau, 1991). From the viewpoint of its origin and essence, (Dillmann, 1899) stated that Ge’ez (Ethiopic) is a pure Semitic language, transplanted by people who migrated from Yemen to Abyssinia, and (Kidanewold, 1948) also stated that Ge’ez had come to the landscape of Ham from Yemen territory of Asia in 3600 BC by the Shem’s clan. In addition to that (Weninger et al., 2011), suggested that “Semitic-speaking peoples left their homeland on the Arabian Peninsula at the end of the 1st millennium B.C. by crossing the Red Sea, and migrated into today’s Ethiopia and Eritrea.”. And (Hetzron, 1997) said that “It is presumably derived from one or more forms of South Semitic brought from Yemen, probably in the first half of the first millennium BCE.”

The name Ge’ez is an original name for the classical language of Ethiopia. However, according to Western discourse, the language is frequently mentioned to as either ‘Old Ethiopic/Classical Ethiopic’, or simply as ‘Ethiopic’ (Weninger et al., 2011). The words

Ethiopic and Ge’ez are used interchangeably in different books and articles. But the researchers of this study used the word Ge’ez instead of Ethiopic.

Let's go back to the ancient stuff (Ge’ez history), as written attestations indicated that, before 5th century B.C the writing direction of ancient Ge’ez was from right to left like Arabic, Syriac and Hebrew and the letters were only the Ge’ez (first order) without 2nd, 3rd, 4th, 5th, 6th and 7th (ካዕብ፣ ሣልስ፣ ራብዕ፣ ኃምስ፣ ሳድስ ፣ ሳብዕ) orders. Below some Ge’ez words show the comparison of right to left with left to right direction of the writing.

Right to left (እም የማን ኀበ ፀጋም)	Left to right (እም ፀጋም ኀበ የማን)	Meaning
መሐጸበወ	ወበጺሖሙ	When they reached
መረሐበ	ብሔሮሙ	Their country
መሀሰሰነአ	እንስሳሆሙ	Their animal
መሰለፈአ	አፍለሶሙ	He migrated them

As well as Pre- and post of Christ (from 5th century B.C up to 4th century A.D), the inscription of Ge’ez (“ዘ ሐወለተ ዘአገበረ አገዘ ለአበወሀ ወሰሐበ…”) founded in the 5th century BC at Metera (መጠራ) in Eritrea shown that the direction of writing became right to left (Diakon Kibret, 2006), (Kidanewold, 1948) but the Ge’ez (first order) letters not changed.

During 4th century AD letters of Ge’ez language were modified by Frumentius (Abba Selama) (Kidanewold, 1948), (Weninger et al., 2011). He added 6 orders from 2nd up to 7th (ካዕብ፣ ሣልስ፣ ራብዕ፣ ኃምስ፣ ሳድስ ፣ ሳብዕ) orders; for example, the letter ‘ሀ’ (hä) before modification was only the first order ‘ሀ’ then after modification it became ሀ፣ ሁ፣ ሂ፣ ሃ፣ ሄ፣ ህ፣ ሆ (hä, hu, hi, ha, he, hə, ho) and those modified letters are used in modern Ge’ez language today.

The Ge’ez language was a spoken language up to Zagwe dynasty (13th century) until the Amharic language replaced the place of Geez language (Kidanewold, 1948), (Weninger et al., 2011). However, (Hetzron, 1997) stated that Ge'ez ceased as a spoken language before the tenth century but it continues today as the liturgical language of the EOTC, and was the only official written language of Ethiopia practically up to the end of the 19th century. In contrast, many people say that Ge'ez language is a dead language. But it is not a dead language since it is still learned and used by church scholars in Ethiopia and Eritrea as a classical language (Weninger, 2010). The investigators of this study also shared Weninger's idea because, at present, it is being offered in a few primary and secondary schools, universities, as well as being a topic of studies in higher educational institutions beyond traditional schools.

3.1.1. Ge'ez writing system (አቡጊዳ/ 'abugida')

Ge'ez has its own writing system called an abugida or alpha-syllabic, in which each character or symbol consisting of one consonant followed by a or zero vowel (Ullendorff, 1951). Meaning that each character represented in Ge'ez script with CV syllable pattern without 6th order. Currently 26 letters/ 'ፊደላት' (each has 7 orders) present in Ge'ez; among those twenty-four symbols are adapted from the South Arabian script called Saba as shown in Appendix 3, Figure 22 and two symbols (ጰ and ፐ) are added later from Greek letters (Meyer, 2016) and (Kidanewold, 1948). But, (ለይኩን, 2006) did not agree on this idea. He said that Ge'ez letters are mother of other letters.

As a general, Ge'ez writing system has 202 symbols (i.e. 182 CV syllables = 26*7 + 20 CWV labiovelars = 4*5), 20 numerals and eight punctuation marks (Kidanewold, 1948), (Dillmann, 1899) excluding mathematical operations and rhyme song of St. Yared. In appendix 3.1, at Table 13, 14, and 15 have shown the all Ge'ez letters, numerals and punctuations respectively.

Until 4th century AD, the order of Ge'ez letters was 'abegede' (አበገደ) vertical line or 'abudida' (አቡጊዳ) horizontal line (see Appendix 3.1, Figure 23) and all letters were called alphabet/ 'አሌፋት' together. However currently the order is 'heleheme' (ሀለሐመ) after it revised and letters are called hohyat/ 'ሆህያት' at one.

According to (Coe, 2012) writing systems of world's languages follow one of the following categories namely: a) alphabet which consists of a set of characters that represent the phonemes of a language in writing. it contains separate letters for both consonants and vowels (Goudi & Nocera, 2014). For instance, the abjad is a subdivision of alphabets. It is a consonantal alphabet in which each character regularly standalone for a consonant (i.e. it has independent letters for consonants but vowels are not present or seen), b) syllabary which is a combination of consonant with vowel (CV) for a syllable. An abugida called alpha-syllabary (it lies between alphabet and syllabary), in which each grapheme represents one syllable. This means that consonant symbols are inherently associated with the following vowel and c) logographic (ideographic) writing system in which a character represents a word or a morpheme. As a result, the Ge'ez script fall in an abugida (alpha-syllabary) writing system.

As described in the Section 3.1 above, each basic (first order) grapheme of Ge'ez has six successor grapheme orders that are produced by the fixed sequence of vowels. The sequences of vowels are shown using diacritics (circles, horizontal or vertical strokes) by attaching them to the basic syllable. For example, the orders of grapheme 'ለ' (lä) are 'ሉ' (lu) represented by adding a horizontal stroke to middle of right side, 'ሊ' (li) written by adding horizontal stroke at the bottom of right side, 'ላ' (la) shortening the left side leg (adding vertical stroke at the right side), 'ሌ' (le) represented by adding a circle at the bottom of right side, 'ል' (lə) written by adding a circle at the internal side of left bottom, and 'ሎ' (lo) by adding a circle at the middle of right side.

The Ge'ez number graphemes were adapted from the Greek letters as shown in appendix 3, Figure 21; because Greeks were used their alphabet as numerical system by assigning numerical value for each letter (Meyer, 2016), (Dillmann, 1899); and (Kidanewold, 1948) also agree without number '፸' among those. Each digit of Ge'ez numbers is multi-syllabic words. In other words, one Ge'ez character is represented by many syllables; for example, ፫ (3) is denoted with three syllables like 'ሠለስቱ'. The Ge'ez numbering system includes ciphered additive and multiplicative of 20 numerals. Which means that the numeral value for digits one up to ten (፩-፲), there is no need additive and multiplicative; their value is itself like ፭ = 5, ፲= 10. The values for numbers from 11- 199 (፲፩ - ፻፺፱), is determined by a linear combination of the numbers from highest to lowest values and each digit added to other. For example, the number 178 is represented as ፻ ፸ ፰ in Ge'ez numbering system (100+70+8). Though, the numbers from and above 200 (፪፻) can be encoded by the mixed pattern (additive and multiplicative). For instance, the numbers 495 and 2010 are encoded like ፬ ፻ ፺ ፭ (4*100+90+5), ፳ ፻ ፲ (20*100+10) respectively. As the same thing, the Ge'ez numbers can be written as ፬ ፻ ፺ወ፭ (495), ፳ ፻ወ፲ (2010), ፲ወ፬ ፼ (140,000) by inserting the letter 'ወ' that is called number wrapper (አኃዝ ጠቅላይ) which is used as adder its predecessor and successor digits. In addition to this Ge'ez numbers used to create words by followed letters like ፳ኤል (ዕሥራ + ኤል), ፪ሆሙ (ክልኤ + ሆሙ), ፪ሆን (ክልኤ + ሆን). However, the advantage is here may only to minimize or abbreviate the length of letters during writing. In Ge'ez numerals, there is no symbols for zero (አልቦ/ዘሮ) and also, they represent integer numbers as well as it can express fraction numbers (e.g. ፬ ትዕሥርት (4/10), ፫ ትኅምስት (3/5)) (Dillmann, 1899). Some Ge'ez numbers (mostly from ፩-፲) have one and more name; meaning that they can named with different their own names for masculine, feminine and days (show Table 1). However, most of the time writing style of number adjective for feminine is in letter form (form of text) instead of numerals for example instead of ፩ ብእሲት. It be written as አሐቲ/አሐት ብእሲት (a woman) and for a day of a week or a date of a month, they can be written as አመ ዐሡሩ ወሰኑዩ ለወርኀ ኅዳር (November 12) in its place አመ ፲ወ፪ ለወርኀ ኅዳር (Kidanewold, 1948) even if, different Ge'ez scripts used the same as 'አመ ፲ወ፪ ለወርኀ ኅዳር' (e.g. Synaxarium/ስንክሳር). The other issue related to Ge'ez numerals (i.e. numbers: ፩-፲, ፻ and sometimes ፼) is that, they contain a letter in front of them in order to show the active action (ገቢር). For instance, number '፩' followed by 'ደ' which is written as '፩ደ' by combining a number with letter together (፩+ደ) and others all written as by followed letter ተ (e.g. ፪ተ, ፫ተ, … ፻ተ) where as ፼ followed ፊ (i.e. ፼ፊ). The analysis of this is that, since Ge'ez numbers are multi-syllables (as defined above), their grapheme is short representation of writing system using letters. For example, number '፭' (5) = 'ኀምስቱ' ('hemmstu') and 'ቱ'/ 'tu' will convert to 'ተ'/ 'te' during active action, then it became 'ኀምስተ' ('hemmste') according to Ge'ez numbers rule. From this fact, '፭' ('hemmstu') is changed to ፭ተ for active action. But in the researcher's consideration, still its internal character representation is 'ኀምስቱ' + 'ተ' since '፭ተ' = 'ኀምስቱ' + 'ተ' from the perspective of grapheme representation. It might be good if first change ፭ to ኀምስቱ then convert 'ቱ' to 'ተ' which became 'ኀምስተ' for the processes of speech recognition system and others NLP researches. Furthermore, in some Ge'ez texts, there exists for example ፩ዱ (፩/አሐዱ+ዱ), ፪ቱ (፪/ክልኤቱ+ቱ) instead of '፩', '፪' respectively. There is no reason to add 'ዱ' in front of '፩' and others like since '፩' equivalents to 'አሐዱ' and '፪' also to 'ክልኤቱ' [i.e. ፩=አሐዱ, ፩ዱ=አሐዱ+ዱ=አሐዱዱ].

Table 1: Ge'ez numerals and their names

Numbers	Numbers with double/triple name for:			Numbers	Numbers with single name for both feminine and masculine
	masculine	feminine	days		
፩	አሐድ/አሐዱ	አሐድ[6]/አሐቲ/አሐት[7]	አሚሩ/ር	፳	ዕሥራ
፪	ክልኤት/ክልኤ/ኤቱ	ክልኤ/ክልኤቲ	ሰኑዩ/ዩ	፴	ሠላሳ
፫	ሠለስት/ሠለስቱ	ሠላስ	ሠሉሱ/ስ	፵	አርብዓ
፬	አርባዕት/አርባዕቱ	አርባዕ	ረቡዑ/ዕ	፶	ኀምሳ
፭	ኀምስት/ኀምስቱ	ኀምስ[8]	ኀሙሱ/ስ	፸	ሰብዓ
፮	ሰድስት/ሰድስቱ *	ስሱ	ሰዱሱ/ስ	፹	ሰማንያ
፯	ሰብዓት/ሰብዓቱ	ሰብዑ	ሰቡዑ/ዕ	፺	ተስዓ
፰	ሰምንት/ሰምንቱ *	ሰማኑ/ሰማኒቱ	ሰሙኑ/ን	፻	ምእት
፱	ተስዓት/ተስዓቱ	ተስዑ	ተሱዑ/ዕ	፼	እልፍ
፲	ዐሥርት/ዐሥርቱ *[9]	ዐሥሩ	ዐሡሩ/ር		
፷	ስድሳ/ስሳ				

Adopted from (Kidanewold, 1948)

3.1.2. Syllable structure of Ge'ez

A syllable is a vowel-like sound together with a portion of the encompassing consonants that are most intently connected with it. It is the combination of consonants and vowels (Jurafsky & Martin, 2007). (Carol & Adelman, 2014) defined as syllable is a linguistic gathering that comprises of a single peak, which might be bordered on one or two sides by consonants. A vowel in the core of the syllable is called nucleus or syllable peak. The syllable peak includes a vowel which is the most essential type of sound (the loudest part of syllable). Where consonants that come first from the nucleus in a syllable are called syllable onset whereas consonants that appear after the nucleus are termed as the syllable coda. And the combination of syllable nucleus and syllable coda is called rime or rhyme (Jurafsky & Martin, 2007).

[6] አሐድ read as 'ahadd' since it is used for feminine.
[7] አሐት read as 'ahatt' since it is used for feminine.
[8] This 'ኀምስ' differs from 'ኃምስ' and this 'ኃምስ' used for masculine to represent the 5th.
[9] Remember that we cannot say 'ሰደስቱ', 'ሰመንቱ' and 'ዐሠርቱ' instead of 'ሰድስቱ', 'ሰምንቱ' and 'ዐሥርቱ' for cardinal number respectively. Because, 'ሰደስቱ', 'ሰመንቱ' and 'ዐሠርቱ' are used for associations (for both masculine and feminine). (in Ge'ez they are called መድበል)

As any language, Ge’ez also has its own syllable structure. (Dillmann, 1899) stated that every syllable in Ge’ez script must begin with a consonant, no syllable begins initially with a double consonant and the syllable may terminate either in a vowel or a consonant. As well as a syllable may end even in two consonants, but only in the termination of a word. Each syllable has one vowel only and no syllable can have above one unless it be two vowels which merge in a single vowel-sound or diphthong[10] (vowel + semivowel in the closed syllable).

And to consolidate this, (Meyer, 2016) stated that words usually begin with a consonant (with the exception of C+r sequences like the word for Christ ክርስቶስ/krəstos which starts with two consonants), but almost never end in the vowel ə. In order to conclude the above reflected concept, the general pattern of Ge’ez syllable is CV(C)(C), where C stands for Consonants and V stands for Vowels. On the other hand, the sub syllables are CV, CVC, CVCC. As it can see from this syllable pattern, Ge’ez supports both open syllable (ends in a vowel e.g. ዝ) and closed syllable (ends in a consonant e.g. ኪያሆን) and in both cases, a vowel of the syllable might be short or long.

In Ge’ez language, syllables can be mono-syllables like ዛ (that), ቃል (word), ለ (to), ላ (to her), ቦ (present), bi-syllables like ንጉሥ (king), ቃልከ (your word), tri-syllables like ቀደስዎ (they blessed him), የሐውር (he goes), poly-syllables like ቀበያውበጠ (go from house to house without invitation), እግዚአብሔር (God), አስተሰናአወ (he made agreement between things).

3.1.3. Ge’ez phonetics

The term Phonetics refers to a study of linguistic sounds which are produced by the human vocal system regardless of their associated languages (Beigi, 2011). In other hands, phonetics is the learning of speech sounds at which how those sounds are produced, perceived, and what are their physical characteristics. The phonetics includes the three areas: Articulatory phonetics (deals to answer the question on how the vocal organs produce speech sound), Acoustic phonetics (studies about the physical features of speech, like frequency, duration, and sounds intensity) and Auditory phonetics (observes the

[10] “A vowel in which the tongue position changes markedly during the production of the vowel is a diphthong.” (Jurafsky & Martin, 2007)

speech perception by the humans' auditory system). They are interrelated, since changing the articulatory setup of the vocal tract brings about acoustic changes which impact the view of a sound (Jurafsky & Martin, 2007), (Carol & Adelman, 2014).

During the process of ASR system, words need to be pronounced in terms of individual speech units (basic pronunciation units) called phones which are the key elements of speech recognition process (Jurafsky & Martin, 2007). Hereafter a sentence is created from sequence of words, in which each word is collected of syllables, wherein each syllable is also made up of phonemes[11] and those are classified in terms of vowels or consonants. Most of the time, phonemes are not exceeded more than 50 in a language (Sadaoki Furui, 2001). There are 37 phonemes (30 consonantal phonemes + 7 vowel phonemes) in Ge'ez language (Weninger et al., 2011). The consonants are 30 in number and categorized as stops, fricatives, approximants, ejectives, nasals and trills according to their articulatory manner. Table 2 illustrates all Ge'ez consonants with their manner and place of articulation. The phonetic transcription of Ge'ez consonants those have same sounds like English consonants is indicated by English letters within brackets ([]) and some consonants those have not same sound like English are represented with other English capital or small letters (e.g. [x] for ኀ, [P] for ጰ, [H] for ሐ) and additional symbols for instance [s'] for ጸ, ['] for ኧ. This indicated that, there are some phonemes presented in Ge'ez language but not in English language and those are ቅ, ቊ, ጥ, ጸ/ፀ and ጵ represented by [q], [qʷ], [T], [s']/ [ḍ] and [P] respectively. Based on place of articulation, Ge'ez consonants also classified as labial, labio-dental, alveolar, palatal, velar, labio-velar, pharyngeal and glottal. In contrast, (Kaye & Daniels, 1997) shown that some difference about place of articulation for Ge'ez consonants. They put the consonants [t], [T], [d], [S], [s'], [n], [r], [l] and [y] under dental category as well as [f] under labial. In addition to that (Kidanewold, 1948) classified as [h, H, x, ', '] (ሀ, ሐ, ኀ, ኧ, ዐ) are spoken with glottal, [g, y, k, q] (ግ, ይ, ከ, ቅ) with palate, [d, T, n, l, t] (ድ, ጥ, ን, ል, ት) with tongue, [b, w, m, f, P, p] (ብ, ው, ም, ፍ, ጵ, ፐ) with labial, and [z, s, s', ḍ, r, S] (ዝ, ስ, ጽ, ፅ, ር, ሥ) are spoken with dental based on their place of articulation.

[11] phoneme is the smallest element of speech units that makes a difference in the meaning of a word. It is a set of sounds or phones.

Each phoneme is represented with one grapheme. The phonemes ጰ and ፐ are used only for borrowed words like ጴጥሮስ, ጰራቅሊጦስ, ፒሉፓዴር, ሰናፔ. In Ge’ez language, phonemes ሀ, ሐ and ኀ merged to h, ሰ and ሥ to s, አ and ዐ to ’, ጸ and ፀ merged to s’/ḍ (Weninger et al., 2011), (Kidanewold, 1948) and (Hetzron, 1997). Although, the graphemes of those merged phonemes have great impact in Ge’ez writing system and semantic. For example, the influence of grapheme ሠ in word ሠረቀ (sereqe) ወጣ/ተወለደ/ተገኘ/ባተ [rise (sun or moon)/born/fund /began (for month, year)] and ሰ in ሰረቀ (Sereqe) ሰረቀ/ሌባ ሆነ [steal/ became a thief]. Those two words (ሠረቀ & ሰረቀ) have same pronunciation but different writing orthography and meaning. For more example see in section 3.1.4 the Ge’ez phonology sub topic.

Table 2: Ge’ez Consonants

Manners of articulation		Places of articulation							
		labial	Labio dental	alveolar	palatal	velar	Labio-velar	pharyn - geal	glottal
Stops	voiceless	ፐ [p]		ት [t]		ክ [k]	ኵ [k^w]		እ [’]
	voiced	ብ [b]		ድ [d]		ግ [g]	ጕ [g^w]		
	ejective	ጵ [P]		ጥ [T]		ቅ [q]	ቍ [q^w]		
Fricatives	voiceless		ፍ [f]	ስ [S]		ኅ [x]	ኍ [x^w]	ሕ [H]	ህ [h]
	voiced			ዝ [z]				ዕ [‘]	
	lateral			ሥ [s]					
Approxi mants	voiced	ው [w]			ይ [y]				
	lateral			ል [l]					
Ejectives	affricate			ጽ [s’]					
	lateral			ፅ [ḍ]					
Nasals		ም [m]		ን [n]					
Trills / Sonorants				ር [r]					

adapted from (Weninger et al., 2011), (Weninger, 2010)

The number of vowels in Ge’ez is 7 and they are categorized as long and short vowels. The short vowels are ä and ə whereas the long vowels are u, i, a, e and o. Furthermore, (Dillmann, 1899) differentiated e and o as mixed sounds (diphthong i.e. ‘e’ as ‘ay’ and ‘o’ as ‘ow’). The fundamental vowel ä contains the most predominance place in Ge'ez and has a great role in the formation of word as a short and long vowel (Meyer, 2016), (Dillmann, 1899). Table 3 depicts all Ge’ez vowel with their position and manner of articulation.

Table 3: Ge’ez vowels

Articulation manner	Articulation place		
	Front	Center	Back
High	ኢ[i]	እ[ə]	ኡ[u]
Middle	ኤ[e]	አ[ä]	ኦ[o]
Low		ኣ[a]	

(adapted from (Kaye & Daniels, 1997))

3.1.4. The Ge’ez phonology

Phonology is the investigation of how sounds systematically act on their distribution in words, their interaction with another cluster of sounds, and their statuses (for example phonemes distinguish the meanings of words) in a given language. Generally, it deals about the significance of phonetics and their internal relation for a particular or different language(s) (Carol & Adelman, 2014), (Beigi, 2011).

There are 202 phones[12] in Ge’ez language derived from 37 phonemes. In Ge’ez language, phones and phonemes can change the meanings of words during the following situations:

a) when consonants or vowels replaced with another consonants or vowels in a given word at the beginning, middle and ending of word within the same environment.

From the beginning of words (their pronunciation is also different):

ቆመ = ‘stand up’, ጾመ = ‘fasted’, and ኖመ = ‘slept’

ዛቲ = ‘this for female’, ላቲ = ‘for her’ and ባቲ = ‘by her’

ቦአ = ‘enter’, and ሞአ = ‘win’

At the middle of words:

ሰሰለ = ‘withdraw/be put off’, ሰቀለ = ‘suspend/crucify’, ሰአለ = ‘ask/demand’, and ሰከለ = ‘become fruitful’, ሰብለ = ‘grow/become grain’

ዋነየ = ‘played’, ዋከየ = ‘shined/lighted’

ስብሐት = ‘glory/glorification’ and ስብከት ‘preaching/proclamation’

From the ending of words:

ወረደ = ‘go down’ and ወረሰ = ‘inherit’

---

[12] Typically, a Phone is an individual and basic sound of speech that occurs in a given language.

ዕፅ = ‘tree’ and ዕድ = ‘man/male’

ሰበረ = ‘break’, ሰበከ = ‘preach’

ምዕራግ = ‘ascent/ladder’, ምዕራፍ = ‘chapter’ and = ‘’ ምዕራብ = ‘West’

የዐርግ = ‘he will raise’ and የዐርብ = ‘it will set (for sun)’

b) when the pronunciation of consonants exchanged in a given word at middle of words within the same environment (making hetero-phones)[13]. In other words, the phonemes in Ge’ez make the gemination by doubling the consonants. For example, the following Ge’ez words illustrate this concept.

ከልአ [käl’ä] (hindered, forbidden/excluded) and ከልአ [källə’ä] (make two/make other)

ጸብሐ [s’äbhä] (become morning, grow light) and ጸብሐ [s’äbbəhä] (pay duty, collect taxes), መሰለ [mäsälä] (be like/look like), and መሰለ [mässälä] (compare/ speak in probable or in proverb),

ጸለለ [s’älälä] (floated, hover), and ጸለለ [s’ällälä] (make darken, shade, cover, slaughtered)

እምነ [’mənnä] (from) and እምነ [‘mmənä] (our mother)

c) when the grapheme representation of consonants exchanged by at least one grapheme at any place in a given word using the *same pronunciation* at the same environment (homo-phones) [14]. Actually, the reason for the sameness of pronunciation is that the missing of the original sounds of some letters as explained in Section 3.1.3. Some Ge’ez words are listed below to show this reflected idea.

አዘዘ [’äzzäzä] (ordered, command) and ዐዘዘ [‘äzzäzä] (be strong, be vigorous),

ሠዐለ [Sä‘älä] (painted, portrayed, shaped), ሰዐለ [sä‘älä] (coughed) and ሰአለ [sä’älä] (asked),

ሠምረ [Sämrä] (liked, agreed, consent) and ሰምረ [sämrä] (flourished, be fruitful),

መሀረ [mähärä] (teach, instruct) and መሐረ [mäHärä] (give mercy),

ርሕበ [rHəbä] (be wide) and ርኀበ [rxəbä] (hungry)

d) when the grapheme representation of consonants exchanged by at least one same sound grapheme at any place in a given word using the different pronunciation at the same environment. (words with homonym/similar phones [e.g. ሠ and ሰ sound as ‘se’] but different syllables/letters)

---

[13] “Hetero-phones are words that have the same orthographic representation but different pronunciations.” (Alkhairy & Jafri, 2016)

[14] Homophones are words that have the same pronunciations but different orthographic and meaning. (Coe, 2012)

ሡብሐ [SäbHä] (be grow/fat) and ሰብሐ [säbbəHä] (glorify, praise),

ሥን [Sən] (beauty) and ስን [sənnə] (teeth),

መልሐ [mällHä] (make tasty) and መልኀ [mäləxä] (tear out, pluck out),

ሠርሐ [SärrəHä] (bring success, make prosper) and ሰርሐ [särHä] (be tired, toil, labor),

አበየ [’äbäyä] (refused, revolt) and ዐበየ [‘äbye] (be great)

e) when the preposition እም (without following ነ) preceded the words that started with መ፣ ሙ፣ሚ፣ማ፣ሜ፣ም and ሞ, it merges the double ም to one and avoids the other.

Examples with መ: እም + መንግሥት = እመንግሥት (from government), by ሙ: እም + ሙዳይ = እሙዳይ (from box), with ሚ: እም + ሚናስ=እሚናስ (from Minas/name of person), by ማ: እም + ማርያም = እማርያም (from Mary), with ሜ: እም + ሜስ = እሜስ (from mead), and by ሞ: እም + ሞገስ = እሞገስ (from grace).

The other thing is that, Ge’ez verbs those contain the phones ሀ/ሐ/ኀ፣ አ፣ ወ and የ change their regular phonological structures. For instance:

ሰሐበ (draw, pull) [past] then, ይስሕብ [future]፣ ይስሐብ [past participle] instead of ይሰሕብ ይስሕብ in the production of words.

ሰአለ (ask) then, ይስእል፣ ይስአል instead of ይሰእል ይስእል in the production of words.

ወረደ () then, ይወርድ፣ ይረድ (ው is avoided) instead of ይውረድ, ኖለወ (kept) then, ይኖሉ (ynollu)፣ ይኖሉ (ynolu) instead or ይኖልው, ሰመየ (give a name) ይሰሚ፣ ይስሚ instead of ይሰምይ፣ ይስምይ.

Also, verbs that hold swallower phones (ወኃጥያን [ቀ፣ ነ፣ ከ and ገ including their 7 orders without 5th order of them]) at the end, the assimilation is occurred (i.e. avoid the phoneme ከ፣ ኩ፣ ኪ፣ ካ፣ ከ፣ ኮ and ን from the word). As an example, table 4 depicts the influence of those phones on verbs.

Table 4: Examples of phonological influence of swallowers phones (ወኃጥያን)

Verbs with correct phone structure	Person	Wrong form	Description
አጥመቀ [‘äTmäqqä] (you baptized)	2^{nd} masculine	አጥመቅ**ከ**	**ከ** is discarded

አጥመቁ [‘äTmäqqu] I baptized)	1st masc. & fem.	አጥመቅ**ኩ**	**ኩ** is discarded
አጥመቂ [‘äTəmäqqi] (you baptized)	2nd feminine	አጥመቅ**ኪ**	**ኪ** is discarded
አጥመቃ [‘äTəmäqqa] (you baptized her)	2nd masculine	አጥመቅ**ካ**	**ካ** is discarded
ድኅነ [dəxənnä] (we be saved)	1st masc. & fem.	ድኅ**ን**ነ	**ን** is discarded
ሰበክን [säbäkkən] (you preached)	2nd feminine plural	ሰበክ**ክ**ን	**ክ** is discarded
ሰበኮ [säbäkko] (you preached him)	2nd masculine	ሰበክ**ኮ**	**ኮ** is discarded
ኀደግሙ [xädäggəmu] (you leaved)	2nd masculine plural	ኀደግ**ክ**ሙ	**ክ** is discarded
ኀደጋ [xädägga] (you leaved her)	2nd masculine	ኀደግ**ካ**	**ካ** is discarded

The other consonant assimilation is that happened when the phone ተ is followed by phones [ሠ፣ ሰ፣ ተ፣ ደ፣ ዘ፣ ጠ፣ ጸ፣ ፀ including their 4th and 6th order] (usually for passive form verbs during their perfect tense). Verbs followed by ተ[tä] are like ተሣረረ፣ ተሰብሐ፣ ተተክለ፣ ተደመረ፣ ተዘከረ፣ ተጠየረ፣ ተጸውዐ፣ ተፀወነ. With an example:

ተስሕበ [täsəHəbä] (he drawn) → ይትሰሐብ [yətsäHäb] → ይትሰሐብ[yətsäHäb] X

ተስሕበ [täsəHəbä] (he drawn) → ይሰሐብ [yəssäHäb] → ይሰሐብ[yəssäHäb] √ ት is removed

Furthermore, when verbs try to show the subject marker and object marker using subject marker phones (ከ፣ ኩ፣ ኪ፣ ክሙ፣ ክን፣ ነ፣ ት), the vowel changing is occurred. The following table 5 shows a typical example based on the verb ‘አእመረ’ (he knows).

Table 5: Example of subject and object maker phones

Subject marker	Object marker	Imagination form	Changed & correct verb form	Options
ከ	ኮ =ከኣ	ኣእመርከ + ኣ	ኣእመርኮ (you know him)	ኣእመርካሁ
	ካ= ከኣ	ኣእመርከ + ኣ	ኣእመርካ (you know her)	ኣእመርካሃ
	ሆሙ	ኣእመርከ + ሆሙ	ኣእመርኮሙ (you know them/for mas.)	ኣእመርካሆሙ
	ን	ኣእመርከ + ኣ+ን	ኣእመርኮን (you know them/for fem.)	ኣእመርካሆን
ኩ	ዎ	ኣእመርኩ+ዎ	ኣእመርክዎ (I known him)	
	ዋ	ኣእመርኩ+ዋ	ኣእመርክዋ (I known her)	
	ዎሙ	ኣእመርኩ+ዎሙ	ኣእመርክዎሙ (I known them/for mas.)	
	ዎን	ኣእመርኩ+ዎን	ኣእመርክዎን (I know them/for fem.)	
ኪ	ኒ	ኣእመርኪ + ኒ	ኣእመርክኒ (you(fem.) know me)	
	ነ	ኣእመርኪ + ነ	ኣእመርክነ (you(fem.) know us)	
ክሙ	ዎ	ኣእመርክሙ + ዎ	ኣእመርክምዎ (you(mas.) know him)	
	ዋ	ኣእመርክሙ +ዋ	ኣእመርክምዋ (you(mas.) know her)	
	ዎሙ	ኣእመርክሙ+ዎሙ	ኣእመርክምዎሙ (you(mas.) know them/for mas.)	
	ዎን	ኣእመርክሙ+ ዎን	ኣእመርክምዎን (you(mas.) know them/for fem.)	
ክን	ኣ+ሁ	ኣእመርክን + ኣ+ሁ	ኣእመርክናሁ (you(fem.) know him)	
	ኣ+ሃ	ኣእመርክን + ኣ+ሃ	ኣእመርክናሃ (you(fem.) know her)	
	ኣ+ሆሙ	ኣእመርክን+ኣ+ሆሙ	ኣእመርክናሆሙ (you(fem.) know them/for male)	
	ኣ+ሆን	ኣእመርክን + ኣ+ሆን	ኣእመርክናሆን (you(fem.) know them/for fem.)	
	ኣ+ኒ	ኣእመርክን + ኣ+ኒ	ኣእመርክናኒ (you(fem.) know me)	
	ኣ+ነ	ኣእመርክን + ኣ+ነ	ኣእመርክናነ (you(fem.) know us)	
ነ	ኣ+ሁ	ኣእመርነ + ኣ+ሁ	ኣእመርናሁ (we know him)	ኣእመርኖ
	ኣ+ሃ	ኣእመርነ + ኣ+ሃ	ኣእመርናሃ (we know her)	
	ኣ+ሆሙ	ኣእመርነ + ኣ+ሆሙ	ኣእመርናሆሙ (we know them/for masculine)	ኣእመርኖሙ
	ኣ+ ሆን	ኣእመርነ + ኣ+ ሆን	ኣእመርናሆን (we know them/for feminine)	ኣእመርኖን
	ኣ+ኪ	ኣእመርነ + ኣ+ኪ	ኣእመርናኪ (we know you/for single feminine)	
	ኣ+ ከ	ኣእመርነ + ኣ+ ከ	ኣእመርናከ (we know you/for single mas.)	
	ኣ+ክሙ	ኣእመርነ + ኣ+ክሙ	ኣእመርናክሙ (we know you/for masculine)	
	ኣ+ክን	ኣእመርነ + ኣ+ ክን	ኣእመርናክን (we know you/for feminine)	
ት	ቶ= ትኣ	ኣእመረት+ ኣ	ኣእመረቶ (she knows him)	

3.1.5. The Ge’ez morphology

Morphology is the study of the internal structure of words by answering the question how morphemes are combined to form new words. Meaning it is the investigation how the way in which words are developed from smaller meaning units called morphemes. A morpheme is often defined as the least possible meaningful morphological unit or building blocks of words in a language. It used to differentiate stems and affixes (prefixes, suffixes, circumfixes and infixes) of the given word (Carol & Adelman, 2014), (Jurafsky & Martin, 2007). In Ge'ez words, all affixes have been exercised. For example, in a verb አማሰነ (he corrupted) then አ is a prefix and ማሰነ is stem, in a verb ባረከቶሙ (she blessed them [masculine]) ቶሙ is a suffix and ባረከ is stem, in a word ክሳውድ (necks), ው is the plural maker and is an infix; ክሳድ (neck) is a stem. The word ትእምርት (sign/mark) came from the verb አመረ ['ämärä]; both ት- and -ት are prefix and suffix for እምር respectively. Hence this shows as Ge'ez supports the circumfix since circumfix is the combination of prefix and suffix (Jurafsky & Martin, 2007). Another example for circumfix is አጠየቀነ ['äTäyyäqännä] (he understood to us) [i.e. አ- prefix, -ነ suffix and ጠየቀ [Täyyäqä] is stem]. There are numerous approaches to consolidate morphemes to make words; among those are inflection, derivation, compounding, and cliticization (Jurafsky & Martin, 2007).

Inflection is the blend of a word stem with a linguistic morpheme, often the new word has the same class to the original word. For example, verb በላዕክሙ (you ate) is derived from verb በልዐ; hence the classes of both original and derived words are the same. Ge'ez has many inflectional morphemes: plural formers (አብዠ ቀለማት) [ል፣ ሙ፣ ት፣ ን፣ አ፣ ው፣ ይ] for marking the plural form of nouns, basic out breeding phones (አሥራው ቀለማት) [ይ፣ ት፣ ን፣ እ] for making the verbs to be past (present/future tense, past participle), negation makers (አሉታ) [አል፣ ኢ] to show the negativity of verbs and nouns, out subject markers (ባለቤት አመልካቾች) [ክሙ፣ ክን፣ ት፣ ነ፣ ኒ፣ ከ፣ ኩ፣ ኪ] for indicating the subject on verbs and out object markers (ተሳቢ አመልካቾች) [ሁ፣ ሃ፣ ሙ፣ ሆሙ፣ ሆን፣ ን፣ ካ፣ ኮ፣ ቶ፣ ዋ፣ ዎ፣ ዎሙ፣ ዎን፣ ያ፣ ዮ፣ ዮሙ፣ ዮን] for marking the object on verbs. A typical example for subject and object markers see Table 5 above and the following Table 6 summarizes the inflectional morphemes for Ge'ez. As it can see from the summarized table, the class of all new words is the same as to original word (stem).

Table 6: Inflectional morphemes of Ge’ez

<table>
<tr><td colspan="2">Plural formers [ል፣ ሙ፣ ት፣ ን፣ አ፣ ው፣ ይ]</td><td colspan="2">Negation makers (አሉታ) [አል፣ ኢ]</td></tr>
<tr><td>Singular</td><td>Plurals</td><td>positive</td><td>negative</td></tr>
<tr><td>ኪሩብ (Cherub)</td><td>ኪሩቤል (Cherubs)</td><td>ቦ (present)</td><td>አልቦ (absent)</td></tr>
<tr><td>ኵሉ (everything)</td><td>ኵሎሙ (all)</td><td rowspan="2">ሖረ (he went)</td><td rowspan="2">ኢሖረ (he not went)</td></tr>
<tr><td>ካህን (priest)</td><td>ካህናት (priests)</td></tr>
<tr><td>ኄር (kind)</td><td>ኄራን (kinds)</td><td rowspan="2">ይመጽእ (he will come)</td><td rowspan="2">ኢይመጽእ (he doesn’t come)</td></tr>
<tr><td>ደብር (mountain)
ወግር (hill)</td><td>አድባር (mountains)
አውግር (hills)</td></tr>
<tr><td>አብ (father)</td><td>አበው (fathers)</td><td rowspan="2">ንጉሥ (king)</td><td rowspan="2">ኢንጉሥ (without king)</td></tr>
<tr><td>ሌሊት (night)</td><td>ለያልይ (nights)</td></tr>
<tr><td colspan="4">Basic out breeding phones (አሥራው ቀለማት) [ይ፣ ት፣ ን፣ እ]</td></tr>
<tr><td>past</td><td>future</td><td colspan="2">Past participle</td></tr>
<tr><td>ቀደሰ (he sanctified)</td><td>ይቄድስ (he will sanctify)</td><td colspan="2">ይቀድስ (he has sanctified)</td></tr>
<tr><td>ቀደሰት (she sanctified)</td><td>ትቄድስ (she will sanctify)</td><td colspan="2">ትቀድስ (she has sanctified)</td></tr>
<tr><td>ቀደስነ (we sanctified)</td><td>ንቄድስ (we will sanctify)</td><td colspan="2">ንቀድስ (we have blessed)</td></tr>
<tr><td>ቀደስኩ (I sanctified)</td><td>እቄድስ (I will sanctify)</td><td colspan="2">እቀድስ (I have sanctified)</td></tr>
</table>

The second approach to combine or collect the morphemes is known as derivational. Derivational is the morphological process which is used to create new basic unit of meaning called lexeme. It can change the new word class (derived word) from the original word class (stem). Ge’ez also has derivational morphemes: out adjectives (ባዕድ ቅጽሎች) [መ፣ ሙ፣ ማ፣ ም፣ ሞ፣ ት፣ ታ፣ ቶ፣ ና፣ ን፣ አ፣ ያ], and out seeds (ባዕድ ዘሮች) [ም፣ እ፣ ት]. It might be good to recapitulate with examples using the following table 7.

Table 7: Ge'ez derivational morphemes

Original word	Word category	Affixes	Derived word	Word category
ነበረ (he sat)	verb	መ	መንበር	Noun
ወረደ (he down)	verb	ሙ	ሙራድ	Noun

ኀደረ (he lodged)	verb	ማ	ማኀደር	Noun
ገብረ (he made)	verb	ም	ምግባር	Noun
ወጸፈ (he thrown by sling)	verb	ሞ	ሞጸፍ	Noun
መለከ (he possessed)	verb	አ	አምላክ	Noun
ገሠጸ (he warned)	verb	ተ	ተግሣጽ	Noun
ቤተ (he lodged)	verb	ታ	ታቦት	Noun
ሠለሰ (he made 3)	verb	ት	ትሥልስት	Noun
ቀደሰ (he sanctified)	verb	ና	ቅድስና	Noun
በርሀ (be lighted)	verb	ን	ብርሃን	Noun
ሖረ (he went)	verb	ያ	ሐዋርያ	Noun
ገዝአ (he dominated)	verb	እ	እግዚእ	Noun

In the processes of inflectional and derivational, gemination is usually occurred (Kaye & Daniels, 1997). For instance, during the production of ይሰብር [yəsäbbər] (he will break) from ሰበረ [säbärä] (he broke), ብ[bbə] in ይሰብር is geminated as well as ቶ is geminated in ሰበረቶ [säbärätto] (she broke him). This shows that gemination is subjected to inflectional process. On the other hand, ወላድ [wällad (fecund) is derived from ወለደ [wälädä] (gave birth/born) and መካን [mäkkan] (barren) produced from መከነ [mäkänä] (he be childless). As a result, ላ[lla] and ካ[kka] are geminated and so gemination is employed the derivational morphological process.

Compounding is the other morphological process approach which is the combination of multiple word stems together (Carol & Adelman, 2014). In Ge'ez usually compounding morphemes are created by combining two or more nouns together. For instance, the word ቤተ መንግሥት (palace) is derived from ቤት and መንግሥት, ዐዘቅተ ማይ (spring water/burrow of water), ሳእረማይ (small water container), ቤተ ክርስቲያን (church), ቤተ ሞቅሕ (prison), ዜና መዋዕል (book of chronicles) and like those are compounding morphemes.

The least but not the last approach for morphological procedure is a cliticization. Defined the cliticization is the mixture of a word stem with a clitic. And a clitic is a morpheme or part of a word that is linguistically dependent on an adjacent word. For example, 'm is a clitic in English word 'I'm' to represent the phrase 'I am'. According to Jurafsky and

Martin, the syntactic conduct of clitic is more like words, usually acting as conjunctions, pronouns, verbs, or articles as well as the phonological characteristics of clitics is similar affixes; they have a habit to be short and unstressed. Clitic is categorized as proclitic that appears before a word and enclitic occurs on the right edge of a word that it is bound to (Jurafsky & Martin, 2007), (Carol & Adelman, 2014). Like Hebrew and Arabic Ge’ez also has clitics. Clitics in Ge’ez are frequently enclitics. According to (Dillmann, 1899) enclitic morphemes are ሁ፣ ሂ፣ ሃ፣ ሄ፣ መ፣ ሰ፣ ሶ፣ ነ፣ ኑ፣ ኒ፣ ኔ፣ አ፣ ኬ and they often attached externally. The enclitics in Ge’ez may conjunctions, pronouns, interrogatives, articles, or exclamatory and they do not make a change in the phonetic situations of the word to which they are functional. They append to the end of words like verbs, nouns, prepositions, adjectives, adverbs; for example, in those words: ውእቱኬ [wə’ətuke] (that is of course!), ተአምኑሁ [tä’ämənuhu] (did you believe?), ዳኅንኑ [daxnənu] (is he fine?), እስከነ [’ əskänä] (up to), ኬ፣ ሁ፣ ኑ፣ ነ are enclitics. In addition to this, Ge’ez also has proclitic morphemes for instance በ፣ ዘ፣ ኢ፣ እንተ፣ እለ and ወ. In contrast, (Weldegiorgis, 2010) points out Ge’ez language does not have a clitic morpheme. According to the researchers’ opinion, clitics may have different forms; for example, English and Tigrigna clitics have the apostrophe and others not. As an evidence, Arabic has both proclitic (e.g. the preposition b for ‘by’ and the conjunction w for ‘and’) and enclitic (e.g. the definite article Al for ‘the’) (Jurafsky & Martin, 2007). In addition to this (Weninger et al., 2011) said that all old Semitic languages have enclitic pronouns that can append to nouns, prepositions, verbs as well as some particles for example in Ge’ez ሰ [sä] (but), ሂ [hi] or ኒ[ni] (even). (Hetzron, 1997) also suggests that monosyllabic prepositions of Ge’ez are proclitic like በ[bä] (in) and ለ [lä] (to) as well as the characteristic highlight of Ge'ez linguistic structure is the utilization of enclitic planning and foregrounding particles which, habitually in conjunction. Hence, using those evidences and such ones, the investigators of this study can have concluded that Ge’ez language has clitics morpheme.

3.1.6. The Ge’ez syntax

Syntax is the study of grammatical constructions that are used for the sequencing of words into different levels called phrases and sentences (Carol & Adelman, 2014). syntax is an essential and critical component for linguistic communication. It is the knowledge about structural relationships between words; or it is the way in which all words can arrange together in a sentence (Jurafsky & Martin, 2007). As it is known as phonemes build morphemes, morphemes build words, words also form phrases, phrases combine into clauses, and clauses form a complex structure called sentences. So, any human language has a syntax or grammar in order to build a meaning full sentence. But the sentence structure may vary from one language to other languages based on their syntax.

Ge'ez language has its own set of sentence patterns (syntax). Unlike Amharic sentence structure (word order) SOV (Tefera, 2005), the most frequently used word order in Ge'ez is VSO. However, its word order in sentence is flexible or large production of word order possibilities (Hetzron, 1997); it might be SOV, SVO, OVS, VOS, OSV and so, where S= subject, O= object and V=verb. It attempts to illustrate it by way of example.

With VSO order:

a) አድኅነኒ (V) እግዚኦ (S) እስመ ኅልቀ ኄር (O) [object clause) → (Help, LORD; for the godly man ceased.),
b) ተወልደ(V) ክርስቶስ(S) እምድንግል(O) → (Christ born from virgin Mary.)

With VOS order:

a) ወይጸግቡ (V) ጠላተ(O) ገዳም (S) → (And the wilderness will be saturated the fatness)
b) ወሖረ (V) በፍኖተ አቡሁ (O) አሳ (S) → (And Asa went in the way of his father)

With OSV order:

a) ወይእዜኒ (O) ነገሥት (S) ለብዉ (V)→(Be wise now therefore, O ye kings)
b) ዮም ነፍሰከ (O) መላእክት (S) አዕረጉ (V)→ (Angels have raised your soul today)

With SOV order: አዕርክትየኒ ወቢጽየኒ (S clause) ዕድወ (O) ኮኑኒ ሮዱኒ ወደበዩኒ (V clause) i.e.

a) አዕርክትየኒ ወቢጽየኒ ዕድወ ኮኑኒ ሮዱኒ ወደበዩኒ→ ("My lovers and friends stand aloof from my sore")
b) ወአእይንቲሁኒ (S) ኀበ ነዳይ (O) ይኔጽራ (V) → (His eyes are going down to the poor)
c) ወናሁ ዛቲ ሀገር (S clause) ቅርብት (O) ይእቲ (V) → (And now that country is near)

With SVO order: ወአንትሙኒ (S clause), ኢትትካየዱ ኪዳነ (V clause), and ምስለ እለ ይነብሩ ውስተ ዛቲ ምድር (O clause) i.e.

a) ወአንትሙኒ ኢትትካየዱ ኪዳነ ምስለ እለ ይነብሩ ውስተ ዛቲ ምድር → ("And ye shall make no league with the inhabitants of this land.")

b) እግዚአብሔር (S) ይመይጥ (V) ምክሮሙ ለአሕዛብ (O clause) → ("The LORD brength the counsel of the heathen to naught")
c) ወምክሩሰ ለእግዚአብሔር (S) ይሄሉ (V) ለዓለም (O) → ("The counsel of the LORD stands for ever")
d) ወንጉሥ ሰሎሞን (S) ሰከበ (V) ምስለ አበዊሁ (O)→ ("So King Solomon slept with his forefathers")

With OVS order:

a) ወበይእቲ ዕለት (O) ኀልቀ (V) መና (S) → (And the manna ceased on that day)
b) እስከ እልህቅ ወእረሥእ (O) ኢትኅድገኒ (V) አምላኪየ (S) → (Do not forsake me, O God, until long and full of days)
c) ከመ ሕንጻ ደቂቅ (O) ኮነ (V) መቅሠፍቶሙ (S)→ ("And the arrow hurts them suddenly")

Ge'ez is inhabited to construct sentences without the subject (OV/VO). At this condition, the subject be known from the verb; like this: ወትእዛዛተ ጽድቅ በወንጌል(O) መጠወ (V) → ("He gave the commands of the righteousness in the gospel.") or መሀረ (V) ፊደለ ሕፃናተ (O) → (He taught the letter for children).

Another issue is that, in Ge'ez sentence the verb is usually has not seen clearly; it shows only S and O only. At this point it contains the auxiliary verbs like ውእቱ፣ ይእቲ፣ ይደሉ. For instance, in this sentence: ቀዳሚሃ ለጥበብ (S clause) ፈሪሀ እግዚአብሔር (O clause) → ("The fear of the LORD is the beginning of knowledge."). Now the word ፈሪሀ contains the auxiliary verb ውእቱ.

3.2. Challenges of Ge'ez language for designing ASR

The main challenge of Ge'ez language for developing automatic speech recognition is the presence of homophones and hetero-phones. The reason for homophones also the redundant of same sound letters. In Geez language there are four pair of alphabets or letters (ሀ፣ ሐ፣ ኀ፣ ሠ፣ ሰ፣ አ፣ ዐ፣ ጸ and ፀ including their seven orders) that represent the same sound. As any other Ge'ez letters, those letters have a great role in the creation of words. And we cannot use by exchanging or replacing one with another. Because the semantic and grammar error is created. For example, words 'ርሕበ' equivalent to 'be wide' and 'ርኀበ' to 'hungry' have different meaning. In sentence 'ዛቲ ባሕር ዐባይ ወረሓብ[15]' equivalent to 'this sea

[15] The word 'ረሓብ' produced from the verb 'ርሕበ'

is great and wide' and 'አጽገበ ነፍሰ ርኅብተ[16]' to 'He nourished the hungry soul'. This shows that the presence of a big difference in semantic of Ge'ez language by the insertion of two letters 'ሐ' and 'ኀ' between 'ር' and 'በ' to create a word 'ርሐበ' or 'ርኀበ' having different meaning but the same pronunciation. We cannot write the above sentence like 'ዛቲ ባሕር ዐባይ ወረኃብ' and 'አጽገበ ነፍሰ ርሕብተ' (ኃ and ሕ misused their place), the meaning is completely becoming wrong. They make confusion to the recognizer. So, the automatic speech recognin can be affected to differentiate homophone words. As well as the hetero-phones are challenges through training of the speech recognition systems because they involve ambiguity in the pronunciation of an orthographic representation of a word.

The other challenge is that speaking style of Ge'ez language. Unlike Amharic, Ge'ez language has its own speaking styles; those are ተነሽ [tenesh], ተጣይ [teTay], ወዳቂ [wedaki] ሰያፍ [seyaf]. They have presented on the same or different words. For example, ቀደሳ = 'wedaqi' for masculine ad ቀደሳ = [tenesh] for feminine. Those reading styles contain different information of speech signal. Finally, the syntax structure of Ge'ez language is very flexible and the developing of n-gram language model (n>= 2) be affected.

3.3. Amharic language versus Ge'ez language

Amharic and Ge'ez languages are the same family of Semitic languages. However, they have their individual characteristics. For example, the following points show Ge'ez is differ from Amharic in:

- Syllable: Syllable structure of Ge'ez is CV(C)(C) = CV, CVC, CVCC; but the Amharic syllable pattern is (C)V(C)(C) = V, CV, CVC, VC, CVCC, VCC.
- Phone arrangement to create words; for example, በልዐ in Ge'ez, በላ in Amharic.
- Reading mechanism
- Syntax structure: Ge'ez has at least six-word arrangements
- Alphabet: Ge'ez letters are only 26 while Amharic letters are 33

[16] The word 'ርኅብተ' produced from the verb 'ርኀበ'

CHAPTER FOUR

4. DESIGNING AND MODEL GEEZ SPEECH RECOGNITION SYSTEM

This chapter gives the description of how the automatic speech recognition of Ge’ez language was developed. It means that, it explains how data (text and speech corpus) was collected and analyzed. It also describes the proposed ASR system for Ge’ez language using HMM modeling technique and how training has been accomplished and also the testing & evaluation techniques.

4.1. Developing text corpus for Ge’ez language

It is clear that both text and speech corpora are needed in the development of ASR for any language. Those corpora can be developed either by collecting the previous recorded audio speech first and then transcribe it in to text manually or by collecting and designing the text corpus first then record it in the form of audio speech by reading the collected text. Since the developing corpus in this study is read corpus, the researchers followed the second one during the development of Ge’ez corpus. The texts were collected from the Ge’ez bible sections (Genesis, Exodus, Deuteronomy, Joshua, kings I, Samuel I, Psalms, Ruth) which is available at ‘Amharic-bible-books’ website and from other hard copy books (‘Wudasie-Mariam’[‘segno’ – ‘Arb’], ‘Melka-Sellasie’, ‘Melka-Gebriel’, ‘Melka-Mariam’,‘Melka-Eyesus’,‘yesene-golgota’, ‘Seqoqawe-Dingl’, ‘Sirate-Kidasie’, ‘Timhrte-Hibuat’) in order to include all 202 Ge'ez CV syllables and 20 Ge'ez number characters in the text corpus and to make the corpus is phonetically rich. Meaning phonetically rich, all the phonemes in Ge’ez are involved in this corpus. The researchers checked all 202 phones and 20 Ge'ez number characters in the text corpus by searching each phone from text corpus. After the text is collected, the next work was checking spellings of words in each sentences and grammars of the compiled Ge’ez script. Because the first collected text was not written well and carefully. Since, the Ge’ez script has the same sounds letters (described at problem statement) but their usage is different, all those letters should put in their appropriate words. This was done by cross-checking each and all words in text corpus with three dictionaries of (Kidanewold, 1948), (Leslau, 1991) and

(Leslau, 1989). At some point those dictionaries was made the difference on graphemes of words and in this time the researchers asked the Ge’ez scholars. But Kidanewold’s dictionary has taken the lion’s share for making the decision by the researchers. Corpus preparation was the main challenge and it took about 6 months. The words, sentences and grammars were checked by the professionals/experts of Ge’ez language. The minimum and maximum number of words in a sentence was 2 and 47 respectively and the total number of sentences was 5251. The sample text corpus is attached at appendix 1 Figure 17.

4.2. Developing the speech corpus for Ge’ez language

After the organization of text corpus, the next corpus development was recording the collected text by the speakers. As described above and since there is no commercial or free database for Ge’ez language, the researchers have selected 83 speakers (72 males and 11 females) to read the prepared Geez sentences (to prepare the read-speech corpus). The male experts have given 53 to 80 sentences; whereas the female experts have given 43 to 121 sentences to record while reading the sentences. From this read speech; 5251 utterances (audio files) were recorded.

The prepared text has been printed for delivering to each speaker to be recorded. The readers have selected by the researchers from different places and their Ge’ez knowledge is considered (Ge’ez scholars, priests, leader-lord ‘Meri-Geta’, and students). The criteria for selecting the readers was the ability to read any Ge’ez text properly. The age of the speakers was from 14 to 51.

Table 8: Age and gender coverage of speakers

Age boundary	No of females	No of males	Total	No of utterances
[14-20]	2	20	22	1364
[21-34]	7	29	36	2259
[35-51]	2	23	25	1628
Total	11	72	83	5251

Their profile like name, gender and age is listed at appendix 1 table 12. The devices for recording were laptop computer and mobile phones. The recorders were only the researchers and the reading level was at sentence level. At a time one sentence only read and save separately. If the reader made a word mistake, cough or sneezing during the recording time, it will be deleted and recorded again for each sentence. Hence for recording one reader it has taken 50 minutes to 1 hour approximately. The file name for each record includes the first three letters of the reader followed by gender indicators M for male or F for female (i.e. the speaker id) then the underscore followed by sequential numbers. For example, if the reader name is 'Kehali', then his first record file name looks like 'KehM_1'. After the completing of recording, all recorded data converted to wave file using Audacity. During the process of conversion, the following parameters were included:

Channel: single channel (mono)

Sample size: 16 bits

Bit rate: 256kbps and

Sample rate: 16kHz

The reason for using the above parameters is that sphinx 4 tool supports only those parameters if the proposed system is for desktop application (normal speech). The silences in speech corpus were removed from the beginning and ending of each recorded utterance using 'Audio Silence Trimmer Pro software' by allowing a minimum silence of 1 second at the start and end of all utterances. Those silences at the starting and finishing of the utterance were created during the recording time. Because we have given a starting gap until the reader starts speaking and finishing gap after the reader finishes the speaking in order to avoid losing of spoken words.

4.3. Automatic speech recognition model for Ge'ez language

The goal of speech recognition is that converting the speech that produced by humans' speech body to its graphical or symbolic representation called text (as it is discussed on section 1.1). Hence, the proposed automatic speech recognition model for Ge'ez language has shown on figure 3 below.

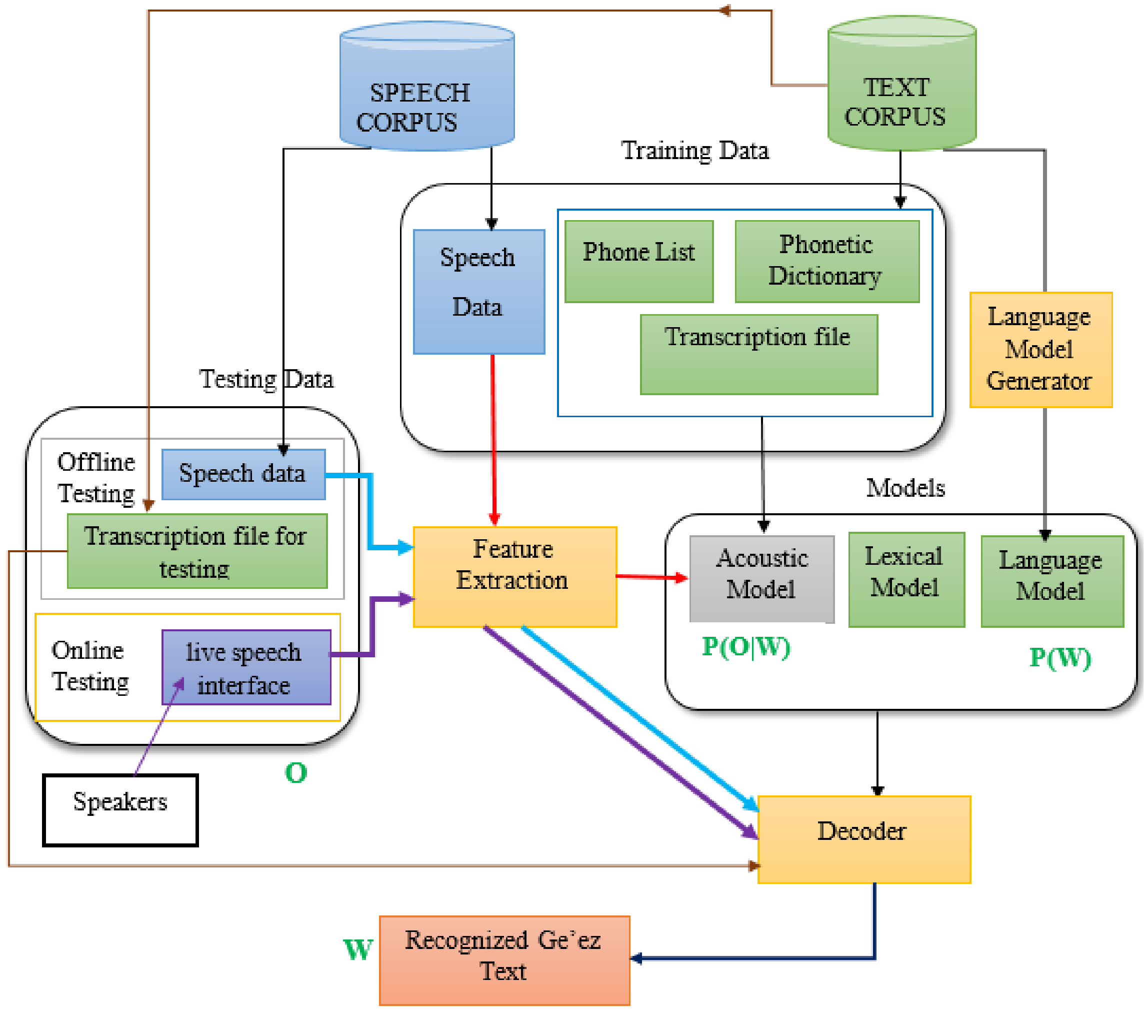


Figure 3: The proposed ASR system for Ge'ez language

Based on the discussion in Section 2.1.3, there are different methods to develop speech recognition system. In this study, the statistical or stochastic method is applied based on the hidden Markov model.

Hidden Markov model

Hidden Markov model is stochastic finite state automata (characterized by set of states) that can produce a sequence of visible states. The arrangement of states is a Markov chain which implies the changes between states has a related likelihood called progress likelihood. And a Markov chain [17] is weighted automation in which the information

[17] "Markov chain, sometimes called the observed Markov model" (Jurafsky & Martin, 2007)

arrangement interestingly figures out that states the automation will go through (Jurafsky & Martin, 2007). It is helpful for allotting probabilities to unambiguous sequences. In other words, an HMM is a Markov chain with the ability to contain further information, either related with its states or its transitions (Beigi, 2011). A Markov chain is only helpful to assign probabilities to unambiguous arrangements. According to Jurafsky, a Markov chain can be seen as a sort of probabilistic graphical model which is a technique of representing probabilistic expectations in a graph.

An HMM comprises of an arrangement of states and transitions set between specific states and each state has its individual probability function that is utilized to control the probability that a given speech frame is produced by a state; this probability function is described by vectors called a variance and means. The HMM technique gives a characteristic and very dependable method for perceiving speech for an extensive variety of applications (uses). To conclude, the observable state sequences in which the state is known from the data leads to Markov chain model while the non-observable states lead to a Hidden Markov Model. For example during Speech recognition process using HMM, acoustic events are the observed layers and texts are the hidden layers (Jurafsky & Martin, 2007). The HMMs can be categorized into discrete model and continuous models based on their observations distributions. Discrete model is a type of HMM model in which the state variables change one at a countable number of facts in time. And these facts in time are ones at that the event happens or changes in state. While in the continuous model, state variables change in continuous way, also not abruptly from one state to other (unlimited number of states).

HMM topology

In fact, an HMM topology is the statistical conduct of an observable node sequence in terms of a network states, which signifies the general movement behavior regarding to the movement between states of the procedure and describes the characteristic varieties in the behavior of the observable nodes inside a state. There are two types of HMM topology based on the its structure: ergodic or fully connected and Bakis or left to right (Sadaoki Furui, 2001). Ergodic HMM is an HMM topology in which each condition or state of the model could be come from each and every other state of the model. In other words, it can generate any sequence from the given topology; meaning, the zero probability of transitions is not occurred among any two states in a sequence. In contrast, Bakis demonstrates on the grounds that the fundamental state sequence related with the model has the property that, as time expands, the state list builds, that is, the framework states continue from left to right. It means that in such HMM type, the zero probability is occurred in many of transitions between states. In this study we used 3 state Bakis HMM topology with non-emitting (non-outgoing transition) terminating state. Because the left to right HMM type has the required property that it is able to model audio signals whose properties vary over time for instance speech (Rabiner Lawrence & Juang, 1993). And (Mittal, 2016) pointed out as ergodic topology not work for speech recognition since speech signals can follow only a specific sequence of sounds and procedures. The motive also to use 3 states is that, the 3-state HMM is a basic sub-word unit model; because the initial state represents the statistical features at starting of a sound, the middle state represents the core of a sound, and the final state represents the spectral features at the end of a sound. Hence, the word based model is created by concatenating those sub-word HMM models (R Lawrence & Ronald, 2007). In addition to that, the sound of a phone or sub word unit is affected through neighbors (predecessor and successor) phonemes. So, to handle the impact of neighbor phones over the change of sound of other phones, the three states HMM is applicable.

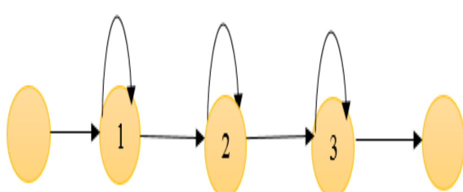


Figure 4: Example of Bakis HMM topology

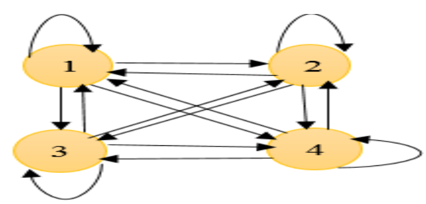


Figure 5: Example of Ergodic HMM topology

HMM components

HMM has the following components:

The first element is the state denoted by **Q** = {q1, q2, q3, … q_N} where N is number of states in a model.

The second element is that a transition probability matrix denoted by **A** = {a_{ij}, a_{ij+1}, … a_{i+1j-n} …$a_{ij=n}$} where $_{ij}$ =1 → n, and a_{ij} represents the moving probability from state *i* to state *j*.

The third component is number of distinct observation symbols per state denoted by **O** = {o_1, o_2, … o_T}, each one is drawn from the vocabulary V= v_1, v_2, … V_v, where T is number of observations.

The fourth component is that observation likelihoods or emission probabilities denoted by **B** = $b_i(o_t)$ generated from state i. And the last element of HMM is that the initial state (q_0) and final state (q_f) in a sequence.

Acoustic model

Acoustic model is a type of model that has a capable to represent the information and knowledge about acoustics, phonetics, variability in environment, gender, pronunciation styles and dialect differences among speakers and so on. The acoustic model has six elements namely model definitions, means, variances, mixture weights, noise dictionary and transition matrix. Those components are generated by the trainer from a given input data called audio file, transcription file and dictionaries (phonetic and filler). After the training is processed, all the existences of phoneme are mapped to the acoustic set of phones. As it is known Sphinx supports two methods (continuous and semi-continuous) for parametrizing the probability distributions of the state observation likelihoods. Here in continuous HMM, using Gaussian mixture density, for example, the probabilities of observation O_t in a given state *i*, $P(O_t|i)$ is computed as: $b_i(o_t) = P(O_t|q_t = i) = \sum_{i=1}^{L} w_{i,j} N_{i,j}(O_t)$, where $N_{i,j}$ is j^{th} Gaussian distribution as well as $w_{i,j}$ is the mixture weight and $\sum_i w_{i,j} = 1$. As a general the emission probabilities (observation likelihoods) are calculated by the acoustic model; meaning the acoustic model [P(W|O)] computes the likelihoods (Jurafsky & Martin, 2007).

Language model

The language model has a great role during the decoding process in a speech recognition system. Because it provides the grammar or the N-gram word order from a given sentence and their probabilities to be selected by the decoder. In the stochastic outline, the word sequence is selected by the decoder therefore the language model increases the product amongst probabilities of observed acoustic signal (input speech) O. When the speaker speaks to the system; it will be estimated by acoustic model $P(O|W)$, and words sequence W that will be estimated by $P(W)$ in a task of recognizing words.

Hence the language model P(W) is generalized as follows: Assume that we have words $w_{1,}$ $w_{2,}$ $w_{3,}$ …w_n in a given sentence. Then the N-gram model P(W) is formulated as:

$$P(W) = P(w_{1,} w_{2,} w_{3,} \ldots, w_n) \quad (3)$$

$$P(W) = P(w_1)\, P(w_2|w_1)\, P(w_3|\, w_{1,} w_2)\, P(w_n \mid w_{1,} w_{2,} w_{3,} \ldots, w_{n-1}) \quad (4)$$

$$P(W) = \prod_{i=1}^{n} P(w_i \mid w_1, w_2, w_3, \ldots, w_{i-1}) \quad (5)$$

Here lastly the mathematical representation of ASR architecture can be expressed as the following.

The acoustic model P(W|O) and the language model P(W) support the ASR system during the conversion of input acoustic signal to a string of words. It means that there is an acoustic observation O. Where, O is a sequence of specific observations gained from the input wave by segmenting it with a particular duration: $O = o_1, o_2, \ldots, o_t$ (6)

And a sentence W, as a sequence of words (w): $W= w_1, w_2, \ldots w_n$ (7)

Hence: $\hat{W} = \text{argmax } P(W|O).$ (8)

where Ŵ the new sequence of words.

$$\hat{W} = \text{argmax } P(W|O) \rightarrow \hat{W} = \frac{P(O|W)P(W)}{P(O)} \quad (9)$$

Finally, we get the following equation by removing the denominator.

$$\boxed{\hat{W} = \text{argmax } P(O|W)\, P(W)} \quad (10)$$

Where P(O|W) is the likelihood (acoustic observation of the word string), P(O) is acoustic observation of test speech and P(W) is prior (probability of word string predicted by language model). As it shown on equation (10), the numerator P(O) is eliminated from the equation (9) to find the unknown sequence words. Because, according to the Bayes rule, amplifying the likelihood of P (X|Y) is identical to expanding the posterior likelihood P(Y|X) P(X)/P (Y). The other reason is that, the denominator (P(O)) is fixed for all possible sequences and its removal does not change the order of competing sequences for the test sentence; meaning that, P(O) is not dependent on the sequence of words W.

4.4. Data preparation

Setting up the dataset plainly is a responsibility duty including sub-undertakings like determination of phonetically rich and phonetically adjusted sentences, choice of fitting members, editing information which is the most tedious parts, recording and transcribing information. The training data were utilized in the process of system development while test data gives the reference interpretations against which performance of decoder can be estimated. On account of the training data the provoke contents were utilized as a part of conjunction with an articulation word reference to give the underlying phone level translations expected to begin the process of HMM preparing.

4.4.1. Building the database for training and testing

The database is the source file (for training and testing) which is a collection of different extension files namely, transcription file, pronunciation dictionary, list of phones, filler file, fileids (file identifications/control files), language model and wave files. The training and testing data were should put separately using two folders namely training and testing for both speech and transcript data. Among the developed corpora, 4818 sentences with their audio file for training, the rest 433 sentences including their audio data were used for testing purpose and those testing data selected randomly using seven speakers.

Transcription file: is a text file used to represent what the speaker said in the audio file. Each sentence in the transcript file is tagged from the beginning and ending with <s> and </s> tags respectively. After the ending tag the corresponding audio file name is followed. Example:

<s> ወይእቲሰ አዕረገቶሙ ውስተ ናሕስ ወኀብአቶሙ ማእከለ ዕፀው ውስተ ሖለት ዘውጡሕ </s> (AtsM_43)

<s> ወዴገንዎሙ ውእቶሙ ዕደው እንበለ ይቢቱ ወይእቲሰ ዐርገት ኀቤሆሙ ውስተ ናሕስ </s> (AtsM_44)

The (AtsM_43) represents the audio file for that sentence. For this study 4818 tagged Ge’ez sentences were used.

Dictionaries: The dictionaries are important files because the trainer lookups to them for deriving sequence of sound units those have related with each transcription and signal. There are two dictionaries included in the database; one is the phonetic dictionary which made up of all unique words and their pronunciation for mapping words in the transcription file or in the language to sequences of sound units. The other is the filler dictionary that used to map non-speech sounds to corresponding non-speech sounds or speech-like sound units.

Sample of pronunciation dictionary

ሀለወ ሀ ለ ወ
ሀለወተነ ሀ ለ ወ ተ ነ
ሀለወት ሀ ለ ወ ት

Example of filler dictionary

<s> SIL
<sil> SIL
</s> SIL

We have developed two dictionaries the first one was included the testing data and size of the vocabulary (number of unique words in a dictionary) is 18973, and the second dictionary was not included the testing data and its size is 18100.

Phone set file: this file is list of all Ge’ez phones including silence; in which one phone listed per line. Hence all listed phones here are 203 (202 Ge’ez phones + 1 silence). A phone file tells a trainer what phonemes are included in the training set. To mean that, phones are means to represent the pronunciation of words in terms of sound units. Here all phones were prepared manually. Phone list examples shown below;

SIL
ሀ
ሁ
ሂ

Control files: are text files that contain all paths and names of audio recording file without file extension and can put one name only per line for example;

training/Abakmariam/AkmM_1

training/Abakmariam/AkmM_2

Wave files: are audio files that contain all audio recording files with '.wav' file extension.

language model: is a plain text which designates the likelihood, probability taken when a sequence of words has seen. In this study, for the development of the language model, the CMU-Cambridge Statistical Language Modeling Toolkit (version-2) was used. For this research, the commands of the CMU-SLM tool were executed on the Ubuntu 16.04 operating system; and the tool generated the word frequency in a corpus, vocabulary and N-gram language model. In our case the value of N was 3 (i.e. tri-gram language model). As it is known, when the value of N is increased the performance of the LM is also increased; because the increasing of N-gram order allows the sequence of words to be long in LM. This also indicates that the probability of the occurrence of hypothesis sentence using correct word sequence is high. However, to design the large N-gram language model, it needs large amount of training corpus since the N-gram language is very dependent on the training data. Because, the availability of long word sequence in small amount of corpus is very rare. In addition to this, it required large amount of memory and the computation time to search all word-based probabilities be long (O'Shaughnessy, 2003), (Jurafsky & Martin, 2007). Figure 5 below depicts all processes of generating a language model.

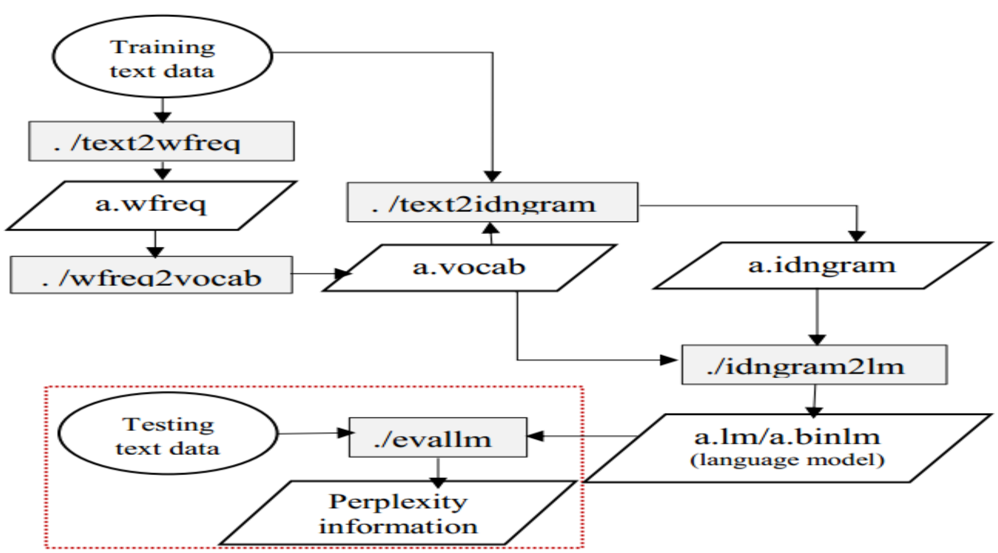


Figure 6: Process of creating and evaluating a language model

Finally, the performance of language model is measured using the evaluation metric called perplexity which answers the question how well the specified statistical model ties the test data by computing the OOV. The perplexity (PP) of a given language model on a test dataset is a probability function that a language model allocates to that test set and it is the most common intrinsic evaluation metric. It can be figured as: let W $=w_1, w_2, \ldots, w_N$ be a testing dataset. Then,

$$\mathrm{PP(W)} = P(\mathrm{w1, w2, \ldots wN})^{-1/N} \quad (1)$$

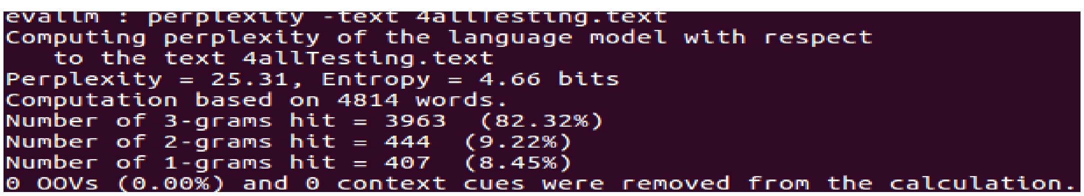
$$\mathrm{PP(W)} = \sqrt[N]{\frac{1}{\mathrm{P(w1,w2,..,wN)}}} \quad (2)$$

Using chain rule for expanding the probability of W, equation (13) would be written as:

$$\mathrm{PP(W)} = \sqrt[N]{\prod_{i=1}^{N} \frac{1}{\mathrm{P(wi|w1,..,wi-1)}}} \quad (3)$$

$$\mathrm{PP(W)} = \sqrt[N]{\prod_{i=1}^{N} \frac{1}{\mathrm{P(wi|wi-1)}}} \text{ (but this is for bigram LM calculation)} \quad (4)$$

In this study, two language models were developed. The first tri-gram language model was generated by including training and testing data. The output contained n-gram 1= 18494, n-gram 2= 58321 and n-gram 3= 73307. As a result, the perplexity value of this designed language model is 24.98 with 0 out of vocabulary (OOV) as shown below figure 7.



```
evallm : perplexity -text 4allTesting.text
Computing perplexity of the language model with respect
   to the text 4allTesting.text
Perplexity = 25.31, Entropy = 4.66 bits
Computation based on 4814 words.
Number of 3-grams hit = 3963  (82.32%)
Number of 2-grams hit = 444  (9.22%)
Number of 1-grams hit = 407  (8.45%)
0 OOVs (0.00%) and 0 context cues were removed from the calculation.
```

Figure 7: Language model Evaluation-1

The second tri-gram language model was developed using training data by excluding the testing data. Hence, the output included n-gram 1=18103, n-gram 2=57105 and n-gram 3=71682. Finally, the number of perplexity of this language model is 557.86 and the out of vocabulary is 830 as it is visible below on figure 8.

```
evallm : perplexity -text 4allTesting.text
Computing perplexity of the language model with respect
   to the text 4allTesting.text
Perplexity = 557.86, Entropy = 9.12 bits
Computation based on 3984 words.
Number of 3-grams hit = 979  (24.57%)
Number of 2-grams hit = 925  (23.22%)
Number of 1-grams hit = 2080  (52.21%)
830 OOVs (17.24%) and 0 context cues were removed from the calculation.
```

Figure 8:Language model Evaluation-2

4.5. Feature extraction

The feature extraction is the process of conversion the speech waveform into a sequence of acoustic feature vectors in the form of frames (most of the time in 10, 15, 20 milliseconds). In other words, it is the process of reduction of dimension or feature since in this process the irrelevant data present in the given input be eliminated whereas important information about the given data will be maintained. The feature vectors represent the evidence of audio signal in a minimum time window of the signal. Each time window or feature vector is represented by 39 MFCC features denoting this spectral information and information about energy as well as spectral change (Jurafsky & Martin, 2007), (Saksamudre et al., 2015). There are different feature extraction methods available such as Mel-Frequency Cepstral Coefficient (MFCC), Linear Predictive Cepstral Coefficient (LPCC), and so; but here the MFCC was applied. Because, the MFCC approaches human framework reply more nearly than any other systems or frameworks and mostly it used for ASR (Vimala, 2012), (Shikha et al., 2013). The method of processing MFCC depends on short-term investigation, and therefore from individual frame a MFCC vector is registered. In general, there are number of steps in the stage of feature extraction. Figure 8 below which is adapted from (Jurafsky & Martin, 2007) has shown all processes performed in feature extraction technique.

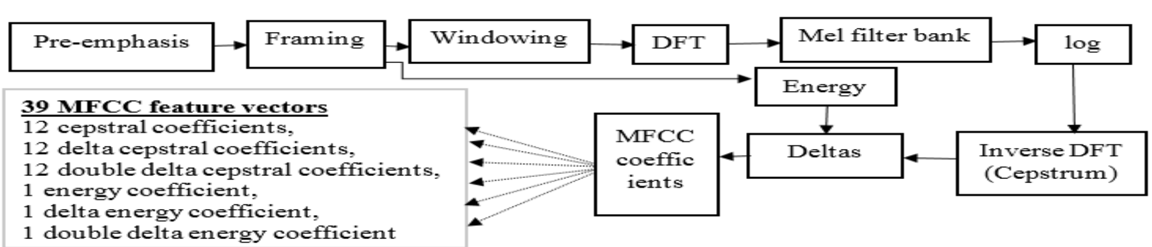


Figure 9: Process of MFCC feature extraction

In feature extraction the first step is pre-emphasis which is used to amplify the quantity of energy at high frequencies. Because, increasing high frequency energy creates information from those higher formants/peaks more accessible for an acoustic model and enhances the accuracy of phone detection.

The second stage is 'Framing' which is used for breaking up the input signals in to small frames with a short time. The motivation behind for this step is that the changing of frequencies in a signal over time; in other words, the statistical properties of speech are a non- stationary signal which means that they are not constant over time. As a result, this leads to uncomfortable to do the Fourier transform over the whole signal; in that the frequency shapes of the signal over the long period will lost. In order to avoid this, it is expected that frequencies in a signal to be stationary over a brief time frame. Consequently, for this reason hamming window is applied (Jurafsky & Martin, 2007). After the finishing of framing step windowing process is performed. It used for reducing or removing the discontinuities (gaps) with hamming window at the starting and end of each frame. Hamming window is utilized as window shape by thinking about the following portion in feature extraction procedure chain and coordinates all the nearest frequency lines; meaning it can shrink signal values near to zero at the window borders. After windowing step, the DFT process is done using the Fast Fourier Transform (FFT) algorithm. DFT is the tool used to extract spectral information for the windowed signal. The output of FFT is the information about the amount of energy at each frequency band.

The next step is the Mel[18] filter-bank. Filter bank is an arrangement of band pass channels having separating alongside data transfer capacity chose by fixed Mel frequency time. The job of Mel filter bank is that to model the auditory system and the auditory system model is used to warp the output of frequencies by the DFT onto the Mel scale. Then the bank of filters (collectors of energy from each frequency band) is achieved regarding to Mel scale. Since the response of human ear to signal level is logarithmic (i.e. when the amplitudes are high, then human's hearings are less sensitive and vice versa), the log of each Mel spectrum

[18] Mel is the abbreviation for the word melody which is the unit of pitch of sounds.

values has taken and the logarithm of the size of the discrete Fourier Transform (DFT) for every signal frame is estimated.

The following stage in MFCC feature extraction is the calculation of the Cepstrum, additionally called as “the spectrum of the log of the spectrum”. The Cepstrum can be viewed as the opposite DFT of the log magnitude of the DFT of a signal. In an extraction of the Cepstrum, 12 cepstral coefficients (for every frame) are produced from earlier steps with the help of inverse DFT. After that other features (delta or double delta) are added for thirteen features (12 coefficient features and one energy which is added from a frame produced by framing) to compensate the variation in Cepstral features through time because the speech signal is not constant from frame to frame. Finally, the delta value estimates the slope using a wider context of frames. The delta features denote the changes between outlines/frames in the corresponding cepstral or energy feature and in addition double delta features denote the change between frames in the matching delta features.

4.6. Training

After the feature extraction processes are completed, the next task in ASR is training the system. The training is done via the Baum welch (forward-backward) algorithm using sphinx 4 trainer tool. The input data for the training is consisted of dictionaries, audio files and the corresponding transcription files and the essential files are listed below in more detail. The output of the training was the acoustic model of Ge’ez language. The amount of the training speech corpus (without testing) in terms of hours was 13.31 as well as the number of sentences was 4818. As it stated above (in Section 4.1.2) the CMU sphinx 4 tool which is available at https://cmusphinx.github.io/wiki/download/ was used for the training process. After downloading the tool, it is needed to configure and setting up. Subsequently the data base is configured with required files listed below.

G_asr: is the name of database folder contains all data

etc: is the sub folder of the G_asr folder and holds the following 8 files

G_asr.dic is the Phonetic dictionary

G_asr.phone is the Phone-set file

G_asr.lm is a language model

G_asr.filler is list of fillers

G_asr_train.fileids is list of files' id and path for training

G_asr_test.fileids is list of files' id and path for testing

G_asr_train.transcription is the transcription file for training

G_asr_test.transcription is the transcription file for testing

wav: is sub folder of the G_asr folder and contains all speech data

training: is sub folder of the wav folder and contains all training speech data

testing: is the sub folder under the wav folder and contains all testing speech data.

Afterward the setting up of database is finished the next process was extracting the features from audio files and the feature vectors were extracted. Finally, after the acoustic model training is done, the acoustic model was generated. The sample for the training process is presented at appendix 1 figure18 and below figure 10 depicted all processes of the training tasks.

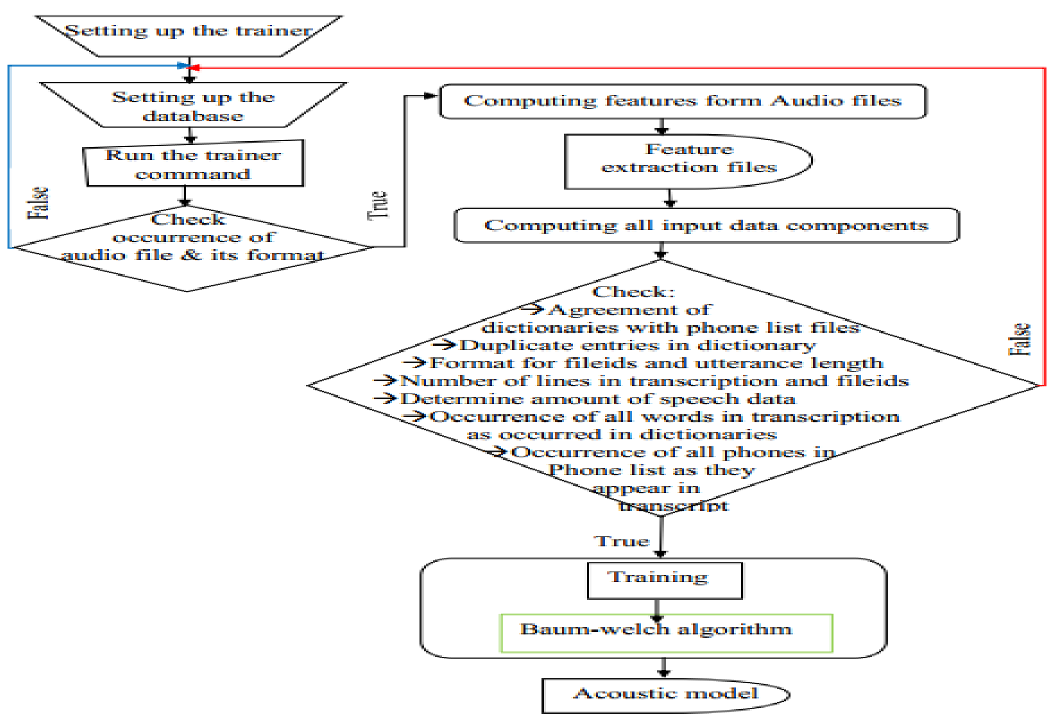


Figure 10: Steps for Training using sphinx-4

4.7. Testing

Testing procedure is an essential task to evaluate and know the accuracy or performance of the proposed system. As it is described in Section 4.1.5 above, the researchers used two methods online and offline methods. During offline test after the process of recognizing is completed, it has shown the recognized Ge'ez text, accuracy of recognizer, and word error rate. The sample for testing is presented at appendix 1 figure 18 and the process of testing using sphinx-4 decoder is shown on figure 11 below.

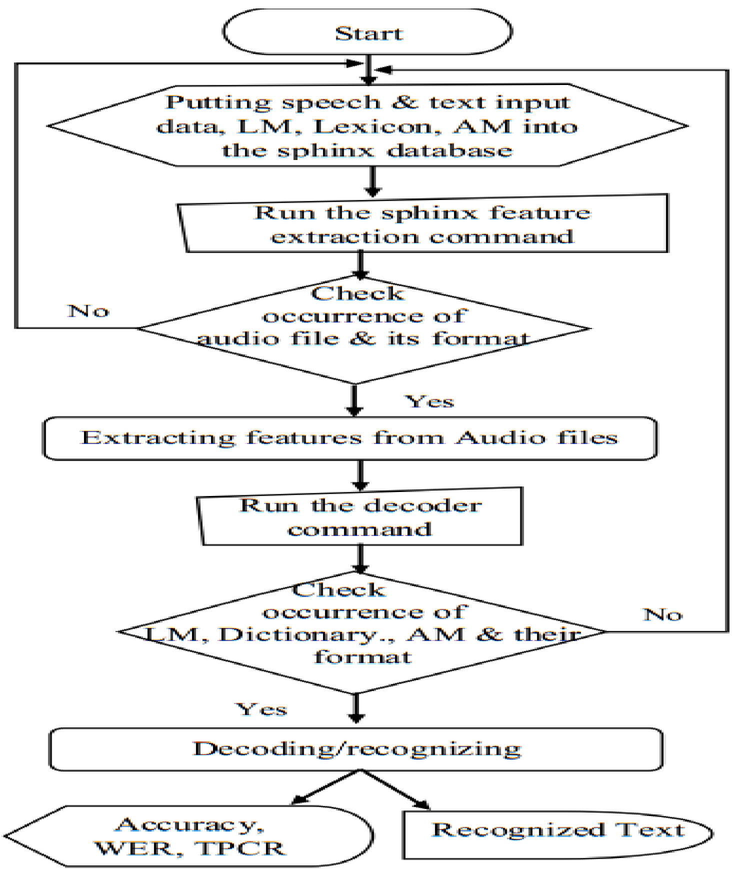


Figure 11: Flowchart for Testing process

4.8. Developing an interface of Ge'ez ASR

To test the system online, we develop a user interface. The implementation is done using java with NetBeans IDE 8.1 based on the CMU-Sphinx guide available at https://cmusphinx.github.io/wiki/tutorialsphinx4/. According to this CMU guide, there are three major high-level interfaces for speech recognition in sphinx 4 namely Live Speech

Recognizer, Stream Speech Recognizer and Speech Aligner. In this study, we have tried to show both the Live and Stream speech recognizers. In Live Speech Recognizer the speakers speak to the system interface using a microphone (for desktop) or with (out) microphone (for laptops) whereas in Stream Speech Recognizer, the input is given to the system by a user from pre- collected audio data by uploading from a file directory. The graphical user interface is developed by taking the language model, dictionaries and the trained acoustic model. Hence, after developing the speech interface, the online testing was performed using the following ASR interface.

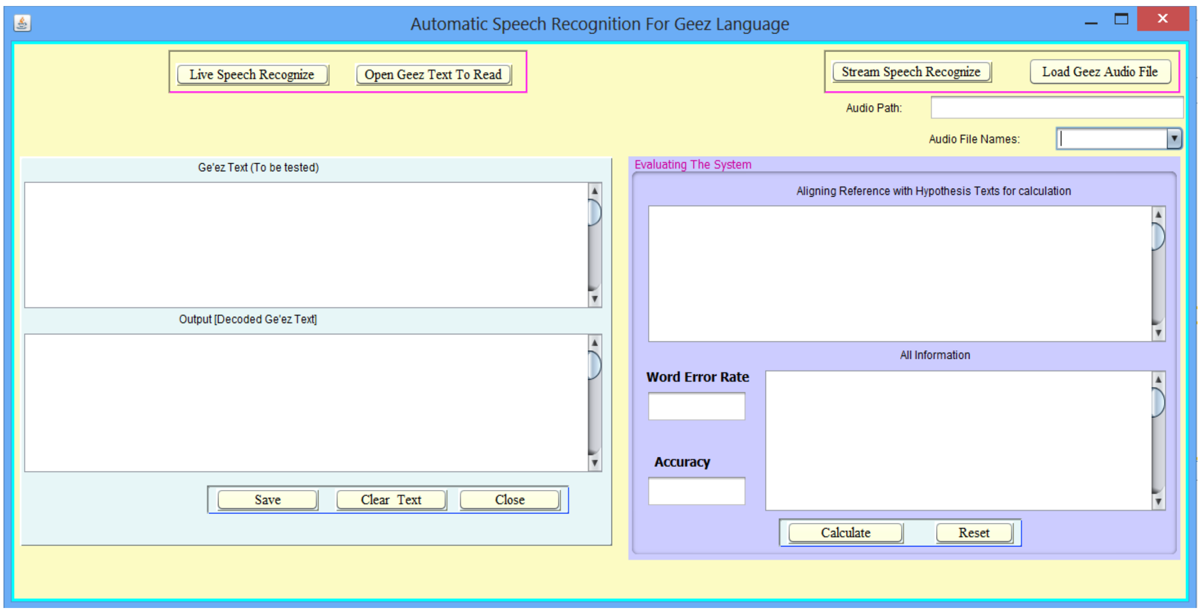


Figure 12: An interface for ASR of Ge'ez language

CHAPTER FIVE

5. RESULTS AND DISCUSSION

This chapter presents the results obtained in the experiments and discusses the results.

5.1. Results for offline testing

In the offline testing phase two experiments were done namely "Experiment 1" and "Experiment 2" using the sphinx tool. In the experiment one, we have used 4818 sentences of speech files for training and 433 sentences of speech files for testing. The experiment 1 was accomplished for evaluating how many the trained data was correctly achieved. During this experiment the language model and dictionary were contained the testing texts. The total number of words presented in 433 testing sentences was 4815. Among those, the 4408 words were recognized correctly and the other 439 words were not recognized correctly. Which means that total insertions = 32, deletions = 69 and substitutions = 338. As a result, the word accuracy rate and word error rate were 90.88% and 9.12% respectively.
The experiment 2 achieved by using the language model and dictionary those were designed without *the testing file*. But the testing data were the same as the experiment 1. The total correct recognized words were 3381 while the unrecognized words were 1517. Hence, word accuracy rate =68.49% and word error rate = 31.51%. And the word insertions = 83, word deletions = 150, and word substitutions = 1284.

In all recognized Ge'ez texts, the miss spelled or miss-used of Ge'ez letters is avoided. In other words, during the formation of words in Ge'ez script using letters was an ambiguity task since there were some Ge'ez graphemes that have been lost their sounds. However, in the output or the decoded Ge'ez script which is result of the ASR, those letters are gained their correct usage in Ge'ez words and the Ge'ez numerals were recognized. The results obtained using sphinx-4 tool for both experiment 1 and 2 presented Appendix 2. Moreover, the following table 9 shows the summary of offline testing results.

Table 9: result summary for offline testing

N<u>o</u> of sentences	Used for	N<u>o</u> of words	Experiments	Types of LM and Dictionary used	Results in 100%	
					WAR	WER
4818	training	71890	Experiment 1	Included testing file	90.88	9.12
4818	training	71890	Experiment 2	Excluded testing file	68.49	31.51
433	testing	4815	For both experiments			
				Average results	79.70	20.3

Here, the results of experiment 1and 2 have shown different accuracy with same testing data. The reason for this variation has come from the distinction of language models. As it discussed in Section 4.2.1, the perplexity in the first language model is less than the perplexity in the second language model. Because, the first language model is constructed with the knowledge of test data and this leads to minimize the perplexity artificially. As the result of this, the performance of language model is also increased as the same point. Because, minimizing the perplexity is identical to maximizing the probability of test set as per language model.

5.2. Results for online testing

In order to test the proposed system using live speech recognizer, you must press the “Live Speech Recognizer” button and wait until the message “Now it is ready for Listening and speak Ge'ez text” shows on the interface. After that you can speak Ge’ez words. The figure 13 below shown the interface for live speech recognizer.

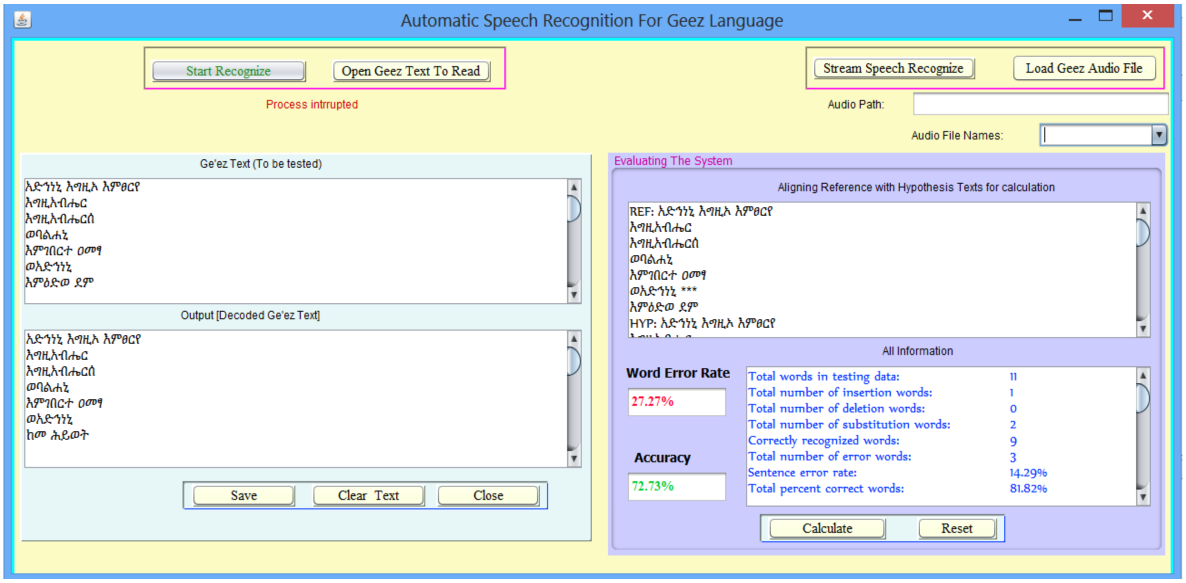


Figure 13: Testing output using live speech recognizer

In order to test the system using the stream speech recognizer, first it needed to load the Ge’ez audio file by pressing the “Load Geez File” button. After that the processing is started by pressing the “Stream Speech Recognizer” button. Finally, to compute the accuracy of the system, it needed to put the reference Ge’ez text in the “Ge’ez Text (to be tested)” text area. You can load the reference text from a file using “Open Geez Text To Read” button and then press the “Calculate” button. The following figure 14 shows an output interface for sample online testing result using stream speech recognizer.

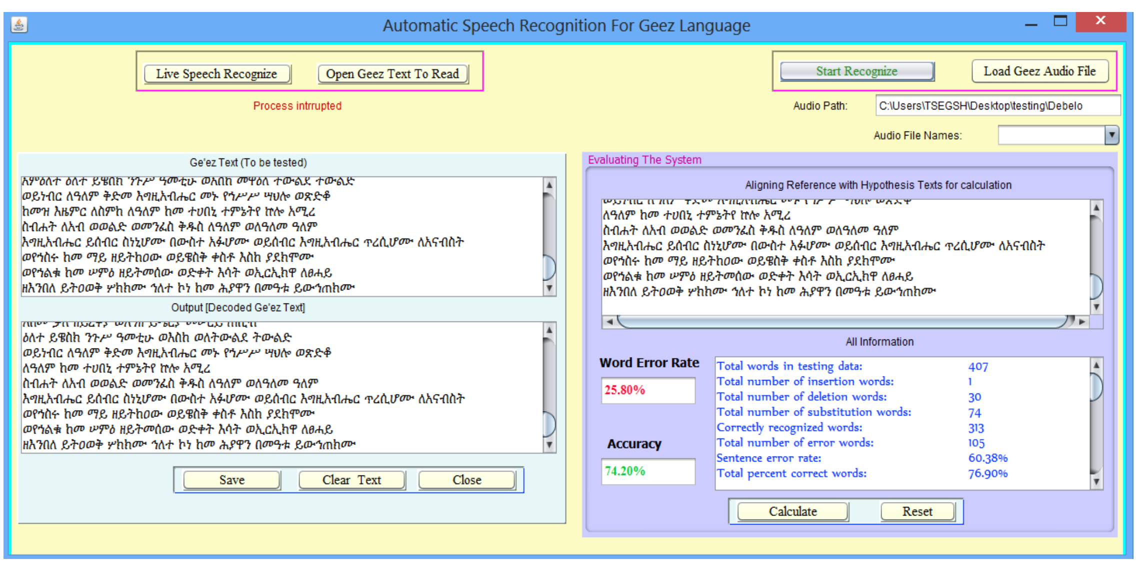


Figure 14: Testing output using Stream speech recognizer

We can calculate the accuracy and word error rate with their insertions, deletions, and substitutions using the above interface of the proposed ASR system. In this testing, seven speakers' audio files (the same as to offline testing data) were used. As a result, the average word accuracy and word error rate were 67.79% and 32.21 respectively.

When we compare the two speech interfaces (online/live modes) with the sphinx decoder (offline/batch mode[19]) they have not functioned at the same performance of the decoder. The reason is that, the live mode recognizer would be affected by the Cepstral Mean Normalization (CMN)[20]. It means that, in batch mode, the cepstral mean is computing using all frames to convey the estimation and the estimation turned out to be good. However, in live mode the cepstral mean is evaluated from previously seen utterance and it needs to approximate the CMN procedure with further means. As a result, some degradations presented regardless of those approximations. In addition to that, the live speech recognizer would be affected by different factors for example noise of external environment & the computer itself, performance of CPU, memory size and types of microphone. The online testing result is summarized below using table 10.

Table 10: Result summary for online testing

Speakers id	Gender	Age	Recording environment	No of sentence	No of words	Results in 100%		
						WAR	WER	SER
MigM	M	19	Open venue	70	571	67.03	32.97	80.52
AmsM	M	22	Open venue	60	963	69.95	30.05	81.47
SenF	F	25	Verandah	59	541	57.71	42.29	82.21
EyeF	F	26	Open venue	64	596	58.65	41.35	84.35
AbeM	M	30	House	63	1011	72.38	27.62	94.41
Eo2M	M	34	Class room	64	882	74.63	25.37	90.63
DebM	M	36	House	53	407	74.20	25.80	69.04
Total→				433	4815			
Average →						67.79	32.21	83.23

[19] "In batch mode the waveform is recorded first, then decoding is done for the whole waveform whereas, in live mode, the recognition is done on the fly when some speech samples was captured by the audio device" (Chan et al., 2007).

[20] CMN is used to estimate all Cepstrum means and utilized it to normalize all cepstral vectors and it is applied on utterance level.

The gender, age and recording environment are affected the performance of recognizer. As we have seen from the above table, the first two results are affected by the environmental situation (nose). The next two results (results of female speakers) have shown the effect of gender. Because, in the training data the participation of female speakers is less than male speakers (14% females, 86% males). And the last three results have shown better performance than the others. The reason is that they were recorded without environmental effect.

5.3. General discussion

As discussed above, the word accuracy rate is 90.88 using first language model (built by including test data). Similarly, using the language model which is constructed without the testing data, the word accuracy rate is 68.49% by performing the offline testing procedure with seven speakers' audio data. Actually, there are different factors that can affect the performance of Ge'ez speech recognizer. One of the factors for the degradation of this accuracy is that the occurrence of 830 OOVs among 4 testing words. Because, when a new word is spoken by a reader or it funded from the test audio file, the recognizer will endeavor to discover one or more words that best matched acoustic signals as the output. Hence, word errors will happen during the current selection procedure and also since our language model is created only from 4815 sentences (i.e. the training corpus), this language model is not enough for representing the all-natural grammatical structures of Ge'ez. To conclude the above idea, the dictionary and unigram language model have the ability to decide the existence of the word in the searching process. when the recognizer failed to find the pronunciation for a given word, then this word be out of the search (Chan et al., 2007).

On the other side, Ge'ez language by its nature, has different word arrangement in sentence level as pointed out in Section 3.1.6. (Ge'ez syntax sub topic). As a result, the presence of different syntax in a given language may have, an impact on language model. Because, the probability of sequence of words will be minimized; means that it makes a confusion or ambiguity during the computation of language model probability (specially for bi-gram and above). For example, the sentence 'ቀተለ ብዙኃነ አሕዛበ'→ 'he killed so many people' has three-word arrangements: ‘ቀተለ ብዙኃን አሕዛበ’, ‘ቀተለ አሕዛበ ብዙኃን’ and ‘ብዙኃን አሕዛበ ቀተለ’. Then assume those three sentences occurred at three different places in a given text corpus. Therefore, it is difficult to decide the correct word sequence and to get the highest probability of sequence of those words.

Ge’ez language also has phonological influence on speech recognition process with different grapheme representation for the same pronunciation (homophones) as stated above (in Section 3.1.4) like ሐረሰ [Häräsä] (cultivated) and ኀረሰ [xäräsä] (born). The other feature or characteristics of Ge’ez language related to phonology impact is that its reading style. Unlike Amharic Ge’ez has different stresses for the same words depending on the context of the semantics of a phrase of sentence. For instance, the word ‘አፍቀራ’ has two stresses based on the predecessor and successor words; in ‘አንስት አፍቀራ አምታቲሆን’ (women loved their husbands) the word ‘አፍቀራ’ has high sound stress (ተነሽ) while in ‘ብእሲ አፍቀራ ለብእሲቱ’ (a man loved his wife), ‘አፍቀራ’ has low sound stress (ወዳቂ). Those features can show the change on the pitch or the tone of speech and the energy that are carried by the signal of the speech. As a result, the mis-recognition or word error rate occurred in the experiment might be the above reasons, therefore, a future research is required to identify the semantic and context arrangement of geez words in a sentence. Based on those and others, in general, after the system is evaluated, the following three kinds of errors categories namely insertion, deletion and substitution were produced. The following table 11 shown summary of all errors with their number of occurrences.

Table 11: Summary of all error categories

Testing procedure	Experiments	Types of errors		
		Insertions	Deletions	Substitutions
Offline	Experiment 1	32	69	338
	Experiment 2	83	150	1284
Online	Experiment 1	87	370	1003

As the table above shown, in all testing, the number of substitution errors were more than insertions and deletions. As a result, this has pointed out the occurrence of confusability or ambiguity between words was high for the proposed system. Hence the reasons for the ambiguity between words be: set of phones in a word and their arrangements to create a word, homophones, word boundary problem during segmentation, acoustic and probability of language model similarity. The following sample output figure 15 shows all types of errors.

```
REF: ወለእመሰ ሰሚዐ ሰማዕከ ቃለ  እግዚአብሔር አምላክከሙ ወዐቀብከሙ ወገበርከሙ ኵሎ  ዘንተ ትእዛዘ ዘአነ እኤዝዘከ ዮም   (Abebe-ABEM_3)
HYP: ወለእመሰ ሰሚዐ ሰማዕከ ቃለ  እግዚአብሔር አምላክከ  ወዐቀብከሙ ወገበርከሙ ኵሎ  ዘንተ ትእዛዘ ዘአነ እኤዝዘከ ዮም   (Abebe-ABEM_3)
Words: 14 Correct: 13 Errors: 1 Percent correct = 92.86% Error = 7.14% Accuracy = 92.86%
Insertions: 0 Deletions: 0 Substitutions: 1

ወእንተ ውስተ ውእቱ መካን ዘኀርዮ እግዚአብሔር አምላክከ ከመ  ይሰመይ ስሙ  በህየ  (Abebe-ABEM_30)
ወአንተ ውስተ ውእቱ መካን ዘኀርዮ እግዚአብሔር አምላክከ ከመ  ይሰመይ ስሙ  በህየ  (Abebe-ABEM_30)

REF: *** ከፈለነ ንብላዕ እምዕፀ  ሕይወት ዘውእቱ ሥጋሁ ለክርስቶስ ወደሙ  ክቡር በእንተ ፍቅረ ዚአነ መጽአ ወአድኀነነ  (Eotc2-EO2M_31)
HYP: ከመ  ለነ   ለብዕላ እምጽዮን እድ   ዘውእቱ ሥጋሁ ለክርስቶስ ወለደሙ ክቡር ***  *** እዴሁ ዐጸዋ ለለነፍስ   (Eotc2-EO2M_31)
Words: 14 Correct: 4 Errors: 11 Percent correct = 28.57% Error = 78.57% Accuracy = 21.43%
Insertions: 1 Deletions: 2 Substitutions: 8

ሰበ  እንሥእ  እደውየ በጽርሐ መቅደስከ  (Abiyu-ABIM_24)
*** ወአነሥእ እደውየ ጽርሐ  መቅደስከ  (Abiyu-ABIM_24)
Words: 5 Correct: 2 Errors: 3 Percent correct = 40.00% Error = 60.00% Accuracy = 40.00%
Insertions: 0 Deletions: 1 Substitutions: 2

እግዚኦ በሥምረትከ *** *** ሀበ  ኀይለ  ለሕይወትየ ሜጥከሰ ገጸከ ወኮንኩ ድንጉፀ   (Abiyu-ABIM_48)
እግዚኦ በሥምረትከ አብ  ላዕለ እለ  ወትሬኢ ጽድቀ    እስመ  ገጸከ ወሖሩ  ወድንጉፀ  (Abiyu-ABIM_48)
Words: 9 Correct: 3 Errors: 8 Percent correct = 33.33% Error = 88.89% Accuracy = 11.11%
Insertions: 2 Deletions: 0 Substitutions: 6

***   ***   ወኢትክል ጠቢሖቶ ለፋሲካ ወኢበውስተ አሐቲ እምነ አህጉሪከ ዘወሀበከ እግዚአብሔር አምላክከ እንበለ ውስተ መካን ዘኀርየ እግዚአብሔር አምላክከ
ወይእዜኒ ሕዝቢሆሙ ወቶሬ   እስከ  ያውዓ  በውስተ   አሐቲ እምነ አህጉሪከ ዘወሀበከ እግዚአብሔር አምላክከ እንበለ ውስተ መካን ዘኀርየ እግዚአብሔር አምላክከ
Words: 31 Correct: 22 Errors: 11 Percent correct = 70.97% Error = 35.48% Accuracy = 64.52%
Insertions: 2 Deletions: 0 Substitutions: 9
```

Figure 15: Sample output with three error types

In figure 15 as shown, the misrecognition was occurred by the substitution of the original word 'አምላክከሙ' with 'አምላክከ' and 'ወእንተ' with 'ወአንተ'. The cause for the occurrence of this substitution is that an ambiguity among two words. Because the difference between set of phones of REF (reference) and HYP (hypothesis) words is very close (i.e. ከሙ and ከ, እ and አ respectively).

The other issue related to cause of error is that, the problem of word boundaries is occurred; for example, from the above figure 15, the word ‘ከፊለነ’ is divided in to two ‘ከመ’ and ‘ለነ’ words and this shows that the occurrence of insertion and substitution errors. The phrase ‘ሶበ አነሥእ’ is merged to word ‘ወአነሥእ’ and this causes to deletion error. The phrase ‘ሀባ ኀይለ ለሕይወትየ’ is shown another way of phone arrangement into words ‘አብ ላዕለ እለ ወትሬኢ’ and causes to substitution error. The word ‘ወኢትክል’ substituted with ‘ወቶራ’ (see figure 15 above); because they have equal LM probability (0.0799 from LM) and their acoustic is more proximity and this also causes for substitution error.

The other cause for ambiguity as described above were homophone words; for example, ‘እደው’ and ‘ዕደው’ are presented in the LM with their probability 0.0799 and 0.1669 respectively. In our testing, when we speak those words without preceding or succeeding words, the output was only ‘ዕደው’. It is difficult to decide the needed word to be recognized. Consequently, the substitution error will be occurred by replacing either the word ‘እደው’ with ‘ዕደው’ or ‘ዕደው’ with ‘እደው’.

CHAPTER SIX

6. CONCLUSIONS AND RECOMMENDATIONS

6.1. Conclusions

In this study, the researchers attempted to show the possibility of developing automatic speech recognition for Ge'ez language using hidden Markov model. Under the process of the study, many tasks were performed by the researchers. The Ge'ez corpora both text and speech were developed at the type of read speech from the ground by collecting from different resources and Ge'ez speakers. The age of the speakers of our speech corpus was between [14 and 51]. The output of text corpus was consisted of 5251 sentences and the speech corpus is 13.31 hours long (without including the testing data).

After that the model of the proposed system ("automatic speech recognition for Ge'ez language") is designed. And the tri-gram language models and dictionaries were developed from the text corpus. In this study, two experiments were implemented using two different language models. In this study, we used 4818 sentences for training data and 433 sentences for testing purpose and the training was done using sphinx-4 trainer as well as sphinx-4 decoder for offline testing. For online or real time testing a graphical user interface was developed using java programming language.

The results for experiment 1 were: word accuracy rate = 90.88%, word error rate = 9.12% and sentence error rate = 37.4%. In the same way, the results for experiment 2 were: word accuracy rate = 68.49%, word error rate = 31.51% and sentence error rate = 82.9%. on the other side, the results for stream speech recognizer were: average word accuracy rate = 67.79%, average word error rate = 32.21% and average sentence error = 83.23%.

The main challenge of Ge'ez language in the development of automatic speech recognition is the occurrence of homophones and hetero-phones words with the reason of redundant letters. So, in this study we have tackled to get the correct recognized words that are built with redundant letters by correcting our text corpus manually. Hence, as we have seen from testing result, the recognizer is displayed the exact and correct homophone and hetero-phone words. So, this is the core strength of this study. However, we have not used any other mechanism (for example rule based approach by developing algorithm) to handle the above challenge and this is counted as a weakness of this study. Therefore, this is an open issue for future research.

6.2. Contributions

In this study the following contributions are figured out.

- The researchers have developed corpora (text and speech) for Ge'ez language. So, the main contribution of this work is the prepared corpora. Because, researches in NLP need corpora either text or speech or both to conduct the study as well as it is a complex task, cost and time consuming specially for languages far from ICT technology like Ge'ez language.
- We developed the Ge'ez language model and dictionary. Those models will use for other NLP researchers.
- We have studied the main features of Ge'ez language and documented them. The document will use as a reference to other investigators.
- We attempted to show the feasibility of developing speech recognition for Ge'ez language since there is no attempt on the area of Ge'ez speech recognition.
- We have put our fingerprint in a little bit for the promotion and development of Ge'ez language with the help of speech technology.

6.3. Recommendations

Research of automatic speech recognition for Ge'ez language is now at an infant age. So, the investigators of this study tried to put the following future directions for the other researchers who have an interest in this area with respect to the Ge'ez language.

In statistical model, the value of accuracy is increased when the size of corpora is increased. As described above, the size of the speech and text corpus were 13.31 hours long and 5251 utterances respectively. So, it is possible to improve the speech recognizer accuracy by expanding the size of corpora using the same procedure.

Since, the developed ASR for Ge'ez language is at the level of word-based speech recognition; but it consumes more memory; the study can extend to phone based or syllable-based speech recognition. Since Ge’ez is morphologically, phonologically, and grammatically (syntax) rich, result of ASR performance might be affected.

As well as this study is done with the scope of speaker independent using read speech corpus. Henceforth, the extending to other types of speech recognition will possible by developing different types of corpora.

In Ge’ez language the gemination is occurred on the consonants. In this study we did not include how to represent the gemination in the dictionary or the text corpus for the processing of speech recognition.

As mentioned above (on section 4.2.5), the researchers have tried to show the possibilities of two speech interfaces (live and stream speech recognizer) for Ge’ez speech recognition. However, the live speech recognizer is not worked as expected like stream recognizer. Hence, it is needed other investigation to develop a good live speech recognizer to evaluate by different customers.

In addition to that, there is an issue related to gender of Ge’ez speakers. As it is known as, there is a shortage of female Ge’ez speakers and consequently the coverage ratio of speakers’ gender is not equal; the number of females was only four. It might be difficult to accept the female’s speech after the development of the speech recognizer system. Hence, in order to compensate the unbalanced ratio of male and female readers, it is needed to develop the balanced speech corpus.

REFERENCES

Abate, S. T., Tachbelie, M. Y., & Menzel, W. (2009). Amharic Speech Recognition : Past , Present and Future, In 16th Int. Conf. Ethiopian studies. 1391–1402.

Abate, Y. (2014). *Morphological Analysis of Ge'ez Verbs Using Memory Based Learning*.

Agrawal, M., & Raikwar, T. (2016). Speech Recognition Using Signal Processing Techniques, *5*(8), 65–68.

Alkhairy, A., & Jafri, A. (2016). Heterophonic speech recognition using composite phones. *SpringerPlus*. https://doi.org/10.1186/s40064-016-3332-9

Ann Copestake. Natural Language Processing : part 1 of lecture notes (2003).

Arora, S. J., & Singh, R. P. (2012). Automatic Speech Recognition : A Review, *60*(9), 34–44.

Beigi, H. (2011). *Fundamentals of Speaker Recognition*.

Belay, A. (2010). *Designing a stemmer for ge ' ez text using rule based approach school of graduate studies*.

Bengio, S., & Keshet, J. (2009). *Automatic speech and speaker recognition*.

Carol, G., & Adelman, A. (2014). *How languages work: An introduction to language and linguistics*.

Chan, A., Gouvea, E., Singh, R., Mosur, R., Rosenfeld, R., Sun, Y., & Seltzer, M. (2007). *The Hieroglyphs*.

Coe, M. D. (2012). Scripts and Languages. In *Breaking the Maya Code* (pp. 1–10).

Cole, R., Mariani, J., Uszkoreit, H., Zaenen, A., Zampolli, A., & Zue, V. (1997). *Survey of the State of the Art in Human Language Technology*.

Davis, K. ., Biddulph, R., & Balashe, S. (1952). Automatic Recognition of Spoken Digits, *24*.

Deng, D. Y. . L. (2015). *Automatic Speech Recognition, A Deep Learning Approach*.

Dewi, I. K. A. N., Firdausillah, F., & Supriyanto, C. (2013). Sphinx-4 indonesian isolated digit speech recognition, *53*(1).

Diakon Kibret, D. (2006). *ሃይማኖታውያን ያልሆኑ የግእዝ ጽሑፎችልሆኑ የግእዝ ጽሑፎች*.

Dillmann, A. (1899). *Ethiopic Grammar*.

Fukada, T., Yoshimura, T., & Sagisaka, Y. (1999). Automatic generation of multiple pronunciations based on neural networks, *27*.

Gebremedhin, Y. B., Duckhorn, F., Hoffmann, R., & Kraljevski, I. (2013). A new approach to develop a syllable based , continuous amharic speech recognizer, (July), 1684–1689.

Gelana, K. (2010). *A continuous, speaker independent speech recognizer for Afaan Oromoo*.

Ghai, W., & Singh, N. (2012). Literature Review on Automatic Speech Recognition.

Gold, B., Morgan, N., & Ellis, D. (2011). *Speech and Audio Signal processing: Processing and Perception of Speech and Music*. https://doi.org/10.1002/9781118142882

Goudi, M., & Nocera, P. (2014). Sounds and symbols: an overview of different types of methods dealing with letters-to-sounds relationships in a wide range of languages in automatic speech recognition.

Gupta, S., Pathak, A., & Saraf, A. (2014). A study on speech recognition system : a literature review, *3*(8), 2192–2196.

Hetzron, R. (1997). *The semitic languages*.

Himanshu, M., & Kaur, S. (2014). Literature Survey on Automatic Speech Recognition System, *4*(7), 398–402.

Ivanov, A. V, Lange, P. L., Suendermann-Oeft, D., Ramanarayanan, V., Qian, Y., Yu, Z., & Tao, J. (2016). Speed vs. Accuracy: Designing an Optimal ASR System for Spontaneous Non-Native Speech in a Real-Time Application, 1–12.

Jurafsky, D., & Martin, J. H. (2007). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. *Computational Linguistics* (Vol. 26). https://doi.org/10.1162/089120100750105975

Kaye, A. S., & Daniels, P. T. (1997). *Phonologies of Asia and Africa* (Vol. 1).

Kibble, R. Introduction to natural language processing Undergraduate study in Computing and related programmes (2013).

Kidanewold, K. (1948). *መጽሐፈ ሰዋስው ወግስ ወመዝገበ ቃላት ሐዲስ (Metshefe Sewasew wegiss wemezgebe qalat hedis)*.

Kothari, C. R. (2004). *Research methodology, methods and techniques*.

Lawrence, R., & Juang, B.-H. (1993). Fundamentals of speech recognition.

Lawrence, R., & Ronald, W. (2007). *Introduction to Digital Speech Processing* (Vol. 2).

Lee, K.-F., Reddy, R., & Allen, J. (1989). *Automatic speech recognition*.

Leslau, W. (1989). Leslau Wolf.-Concise Dictionary of Ge’ez.pdf.

Leslau, W. (1991). *Comparative Dictionary of Ge’ ez*.

Liddy, E. D. (2001). Natural Language Processing. *Annual Review of Applied Linguistics*, *37*(1), 51–89. https://doi.org/10.1017/S0267190500001446

Meyer, R. (2016). the ethiopic script : linguistic features and socio-cultural connotations. *Oslo Studies in Language*, *8*(1), 137–172.

Mittal, S. (2016). *Implementation of Statistical Speech to Text Recognition System for Punjabi Language Master of Engineering Shama Mittal Ms . Rupinderdeep Kaur*.

Mostafa, M., Tolba, H., Mahdy, S., & Fashal, M. (2008). Syllable-based automatic Arabic speech recognition Speech, 246–250.

O'Shaughnessy, D. (2003). Interacting with computers by voice: Automatic speech recognition and synthesis. *Proceedings of the IEEE*, *91*(9), 1272–1305. https://doi.org/10.1109/JPROC.2003.817117

Patnaik, D. B. B. (2016). IJMSRR E- ISSN - 2349-6746 International Journal of Management and Social Science Research Review , Vol . 1 , Issue . 11 , May - 2015 . Page 113 Research Paper E- ISSN - 2349-6746 International Journal of Management and Social Science Research Review , Vo, *1*(11), 113–116.

Rajasekar, S., & Philominathan, P. (2013). Research methodology, 1–53.

Reshamwala, A., Mishra, D., & Pawar, P. (2013). Review on natural language processing, *3*(1), 113–116.

Sadaoki Furui. (2001). *Digital speech processing, synthesis, and recognition*. *Speech Communication* (Vol. 8). https://doi.org/10.1016/0167-6393(89)90009-5

Saksamudre, S. K., Shrishrimal, P. P., & Deshmukh, R. R. (2015). A Review on Different Approaches for Speech Recognition System, *115*(22), 23–28.

Shikha, G., Jafreezal, J., Fatimah wan, A. W., & Bansal, A. (2013). Feature extraction using mfcc, *4*(4), 101–108.

Singh, Y. K. (2006). *Fundamentals of research methodology and statistics*.

Solomon, T., Yifiru, M., & Abebe, E. (2015). *Final Report*.

Study.com. (2016). Speech Recognition_ History & Fundamentals _ Study. Retrieved January 4, 2017, from http://study.com/academy/lesson/speech-recognition-history-fundamentals.html

Tefera, S. (2005). *Automatic Speech Recognition for Amharic*.

Teferra, S., & Menzel, W. (2007). Syllable based speech recognition for Amharic, (June), 33–40.

Trivedi, V. (2013). A Survey on English Digit Speech Recognition using HMM. *International Journal of Scientific Engineering and Research (IJSER)*, *1*(1–3), 247–253.

Ullendorff, E. (1951). Studies in the Ethiopic Syllabary, *21*(3), 207–217.

Vimala, C. (2012). A Review on Speech Recognition Challenges and Approaches, *2*(1), 1–7.

Weldegiorgis, D. (2010). *Design and implementation of automatic morphological analyzer for ge ' ez verbs design and implementation of automatic morphological analyzer for ge ' ez verbs*.

Weninger, S. (2010). Sounds of Ge'ez: How to Study the Phonetics and Phonology of an Ancient Language, *13*.

Weninger, S., Khan, G., Streck, M. P., & Watson, J. C. E. (2011). *The Semitic Languages: an International Handbook.*

ለይኩን፡ብርሃኑ ዶ/ር. (2006). የግዕዝ መማሪያ. Retrieved from www.debelo.org

APPENDIXES

Appendix 1: Documents related to corpus


BiT

Bahir Dar Institute of Technology - Bahir Dar University

Faculty of Computing

Ref.No:- FC/305/2009

Date :- 18/07/2009

Figure 16: Cooperation letter

ጸላእኩ ማኅበረ እኩያን ወኢይነብር ምስለ ጽልሕዋን
ወአኀፅብ በንጹሕ እደውየ ወአዐውድ ምሥዋዒከ እግዚኦ ከመ እስማዕ ቃለ ስብሐቲከ
ወከመ እንግር ኵሎ መንክረከ
እግዚኦ አፍቀርኩ ሥነ ቤትከ ወመካነ ማኅደረ ስብሐቲከ
ኢትግድፋ ምስለ ኃጥኣን ለነፍስየ ወኢምስለ ዕድወ ደም ለሕይወትየ
እለ ዐመፃ ውስተ እደዊሆሙ ወምልእት ሕልያነ የማኖሙ
ወአንሰ በየውሀትየ አሐውር አድኅነኒ እግዚኦ ወተሣሃለኒ
እስመ በርቱዕ ቆማ እገርየ በማኅበር እባርከከ እግዚኦ
እግዚአብሔር ያበርህ ሊተ ወያድኅነኒ ምንትኑ ያፈርሀኒ እግዚአብሔር ምእመና ለሕይወትየ ምንትኑ ያደነግፀኒ
ሶበ ይቀርቡኒ እኩያን ይብልዑኒ ሥጋየ ጸላእትየሰ እለ ይሣቅዩኒ እሙንቱ ደክሙ ወወድቁ
እመኒ ጸብአኒ ተዓይን ኢይፈርሀኒ ልብየ ወእመኒ ሮዱኒ ጸባኢት አንሰ ቦቱ ተወከልኩ
አሐተ ሰአልክዎ ለእግዚአብሔር ወኪያሃ አኀሥሥ ከመ እኅድር ቤቶ ለእግዚአብሔር በኵሉ መዋዕለ ሕይወትየ ወከመ ያርእየኒ ዘያሠምሮ ለእግዚአብሔር ወከመ እፀመድ
እስመ ኀብአኒ ውስተ ጽላሎቱ በዕለተ ምንዳቤየ ወሰወረኒ በምኅባአ ጽላሎቱ ወዲበ ኰኵሕ አልዐለኒ
ናሁ ይእዜ አልዐለ እግዚአብሔር ርእስየ ዲበ ጸላእትየ ዖድኩ ወሦዕኩ ውስተ ደብተራሁ መሥዋዕተ ወየበብኩ ሎቱ እሴብሕ ወእዜምር ለእግዚአብሔር
ስምዐኒ እግዚኦ ቃልየ ዘሰአልኩ ኅቤከ ተሣሃለኒ ወስምዐኒ ለከ ይብለከ ልብየ
ወኀሠሥኩ ገጸከ ገጸ ዚአከ አኀሥሥ እግዚኦ
ወኢትሚጥ ገጸከ እምኔየ ወኢትትገሐሥ እምገብርከ ተምዒዐከ ረዳኤ ኩነኒ ወኢትግድፈኒ ወኢትትሀየየኒ አምላኪየ ወመድኅኒየ

Figure 17: Sample text corpus

```
Running the training
MODULE: 000 Computing feature from audio files
Extracting features from  segments starting at  (part 1 of 1)
Extracting features from  segments starting at  (part 1 of 1)
Feature extraction is done
MODULE: 00 verify training files
    Phase 1: Checking to see if the dict and filler dict agrees with the phonelist file.
        Found 18976 words using 203 phones
    Phase 2: Checking to make sure there are not duplicate entries in the dictionary
    Phase 3: Check general format for the fileids file; utterance length (must be positive); files
exist
    Phase 4: Checking number of lines in the transcript file should match lines in fileids file
    Phase 5: Determine amount of training data, see if n_tied_states seems reasonable.
        Estimated Total Hours Training: 13.3143944444444
        Rule of thumb suggests 3000, however there is no correct answer
    Phase 6: Checking that all the words in the transcript are in the dictionary
        Words in dictionary: 18973
        Words in filler dictionary: 3
    Phase 7: Checking that all the phones in the transcript are in the phonelist, and all phones in
 the phonelist appear at least once
MODULE: 0000 train grapheme-to-phoneme model
```

Figure 18: Sample training process

```
ወበሳብዕ ዓም  ትገብር ኅድገተ ወከመዝ ውእቱ ትእዛዝ ለኅድገት ዘይትኀደግ ኵሎ  ንዋየከ ዘይፈድየከ ካልእከ ወእኁከ ኢትትፈደይ እስመ ኅድገት ተሰምየት ለእግዚአብሔ
ወበሳብዕ ዓም  ትገብር ኅድገተ ወከመዝ ውእቱ ትእዛዝ ለኅድገት ዘይትኀደግ ኵሎ  ንዋየከ ዘይፈድየከ ካልእከ ወእኁከ ኢትትፈደይ እስመ ኅድገት ተሰምየት ለእግዚአብሔ
Words: 20 Correct: 20 Errors: 0 Percent correct = 100.00% Error = 0.00% Accuracy = 100.00%
Insertions: 0 Deletions: 0 Substitutions: 0
ወዘኀቤከ ነኪር ትትፈደይ ኵሎ  ***  ***   ዘብከ  ኀቤሁ ወለእኁከሰ ኅድገተ ትገብር ዘይፈድየከ እስመ አልቦ ነዳየ እምኔከ እስመ ባርኮ  ይባርከከ እግዚአብ
ወዘኀቤከ ነኪር ትትፈደይ ኵሎ  ዘገብረ ተአምሪሁ ወበጽሑ ነፍሰ ሕይወት   እንተ  ትገብር ዘይፈድየከ እስመ አልቦ ነዳይ እምኔከ እስመ ኀባእኩ ይባርከከ እግዚአብሔ
Words: 27 Correct: 20 Errors: 9 Percent correct = 74.07% Error = 33.33% Accuracy = 66.67%
Insertions: 2 Deletions: 0 Substitutions: 7
ወለእመሰ ሰሚዐ ሰማዕከ ቃለ  እግዚአብሔር አምላክክሙ ወዐቀብክሙ ወገበርክሙ ኵሎ  ዘንተ ትእዛዘ ዘአነ እኤዝዘከ ዮም   (Abebe-ABEM_3)
ወለእመሰ ሰሚዐ ሰማዕከ ቃለ  እግዚአብሔር አምላክከ  ወዐቀብክሙ ወገበርክሙ ኵሎ  ዘንተ ትእዛዘ ዘአነ እኤዝዘከ ዮም   (Abebe-ABEM_3)
Words: 14 Correct: 13 Errors: 1 Percent correct = 92.86% Error = 7.14% Accuracy = 92.86%
Insertions: 0 Deletions: 0 Substitutions: 1
እስመ ባረከከ እግዚአብሔር አምላክከ በከመ ይቤለከ ወትሌቅሕ ለአሕዛብ ብዙኀ ወአንተሰ ኢትትሌቃሕ ወትኴንኖሙ ለአሕዛብ ብዙኃን አንተ ወለከሰ ኢይኴንኑከ  (Abe
መኑ  ባረከከ እግዚአብሔር አምላክከ በከመ ይቤለከ ወትሌቅሕ ለአሕዛብ ብዙኀ ወአንሰ  እትፌሣሕ  ወትኴንኖሙ ለአሕዛብ ብዙኃን አንተ ወረሰየ ኢይኴንኑከ  (Abe
Words: 17 Correct: 13 Errors: 4 Percent correct = 76.47% Error = 23.53% Accuracy = 76.47%
Insertions: 0 Deletions: 0 Substitutions: 4
```

Figure 19: Sample testing process

Table 12: speakers profile

No	Speakers ID	Age	Gender	Place	Record environment	Noise
1	AbeM	30	Male	A/Ababa	House	No
2	AkmM	30	Male	A/ Ababa	House	No
3	AtsM	45	Male	A/ Ababa	House	No
4	AbiM	20	Male	Bahir Dar	Venue	Yes
5	AduM	33	Male	Bahir Dar	Venue	Yes
6	AemM	17	Male	Bahir Dar	Venue	Yes
7	AleM	32	Male	Bahir Dar	Class room	No
8	AlmF	27	Female	Bahir Dar	Verandah	Yes
9	AmhM	35	Male	Bahir Dar	House	No
10	AmsM	22	Male	Bahir Dar	Open Venue	Yes
11	AseM	37	Male	A/Ababa	House	No
12	AtsF	20	Female	Bahir Dar	office	No
13	BekM	51	Male	A/ Ababa	Outside	Yes

14	BelM	42	Male	A/Ababa	Verandah	Yes
15	BetM	15	Male	Bahir Dar	Venue	No
16	BezF	22	Female	Bahir Dar	Verandah	Yes
17	BirM	28	Male	A/Ababa	Verandah	Yes
18	BiZM	24	Male	Bahir Dar	Open Venue	Yes
19	BzuM	32	Male	A/Ababa	Verandah	Yes
20	DawM	18	Male	Bahir Dar	Open Venue	Yes
21	DebM	36	Male	A/ Ababa	Open Venue	Yes
22	DibM	34	Male	A/ Ababa	Closed Venue	No
23	EmaF	45	Female	Bahir Dar	Open Venue	Yes
24	EseF	35	Female	chegodie	Open Venue	Yes
25	EwmF	31	Female	chegodie	Open Venue	Yes
26	EphM	29	Male	A/Ababa	House	No
27	ErmM	33	Male	Bahir Dar	House	No
28	EtsM	18	Male	Bahir Dar	Venue	No
29	EyaM	19	Male	A/Ababa	Class room	No
30	EyeF	26	Female	Bahir Dar	House	No
31	FseM	34	Male	A/Ababa	House	No
32	FirM	26	Male	Bahir Dar	Venue	No
33	FkaM	27	Male	A/Ababa	House	No
34	FmaM	14	Male	Bahir Dar	Venue	No
35	Eo1M	35	Male	A/Ababa	Class room	Yes
36	Eo2M	34	Male	A/Ababa	Class room	Yes
37	GirM	25	Male	A/Ababa	House	No
38	GkiM	20	Male	Bahir Dar	Open Venue	Yes
39	GmaM	17	Male	Bahir Dar	Open Venue	Yes
40	GmdM	34	Male	Bahir Dar	Open Venue	Yes
41	GmhM	36	Male	A/Ababa	House	No
42	GmeM	19	Male	Bahir Dar	Open Venue	Yes
43	GmiM	20	Male	Bahir Dar	Open Venue	Yes
44	GmsM	22	Male	Bahir Dar	Open Venue	Yes
45	HaiM	29	Male	A/Ababa	House	No
46	HawM	18	Male	Bahir Dar	Open Venue	Yes
47	HenM	21	Male	A/ Ababa	Outside	Yes
48	HerM	26	Male	A/Ababa	House	No
49	HelF	22	Female	Bahir Dar	office	No
50	HmaM	24	Male	Bahir Dar	Open Venue	Yes
51	HmmM	22	Male	A/Ababa	House	No
52	HmrM	33	Male	A/Ababa	House	No
53	HweM	41	Male	Bahir Dar	Open Venue	No
54	KehM	18	Male	Bahir Dar	Open Venue	No
55	KelM	36	Male	A/Ababa	House	No
56	KibM	14	Male	Bahir Dar	Open Venue	No
57	KidM	39	Male	A/Ababa	House	No
58	KirM	14	Male	Bahir Dar	Open Venue	No

59	LemM	35	Male	A/ Ababa	House	No
60	MasM	50	Male	A/ Ababa	House	No
61	MehM	35	Male	A/Ababa	House	No
62	MelM	48	Male	A/ Ababa	House	No
63	MegM	32	Male	A/ Ababa	House	No
64	MenM	20	Male	Bahir Dar	Open Venue	No
65	MigM	19	Male	Bahir Dar	Open Venue	No
66	RomF	20	Female	Bahir Dar	Office	No
67	SelM	44	Male	A/Ababa	House	No
68	SenF	25	Female	Bahir Dar	Verandah	Yes
69	ShiM	38	Male	A/Ababa	House	No
70	SirM	27	Male	A/ Ababa	Verandah	Yes
71	TarM	34	Male	A/ Ababa	Verandah	Yes
72	TseM	36	Male	A/Ababa	House	No
73	TegM	34	Male	Bahir Dar	House	No
74	TgaM	37	Male	A/Ababa	House	No
75	WweM	31	Male	A/Ababa	Close venue	No
76	YemM	38	Male	A/ Ababa	House	No
77	YohM	19	Male	Bahir Dar	Open Venue	No
78	YhaM	35	Male	A/Ababa	House	No
79	YosM	38	Male	A/ Ababa	House	No
80	YtbM	36	Male	A/ Ababa	Verandah	Yes
81	ZekM	19	Male	Bahir Dar	Open Venue	No
82	ZebF	27	Female	Bahir Dar	Verandah	Yes
83	ZelM	19	Male	Bahir Dar	Open Venue	No

Appendix 2: Result related files

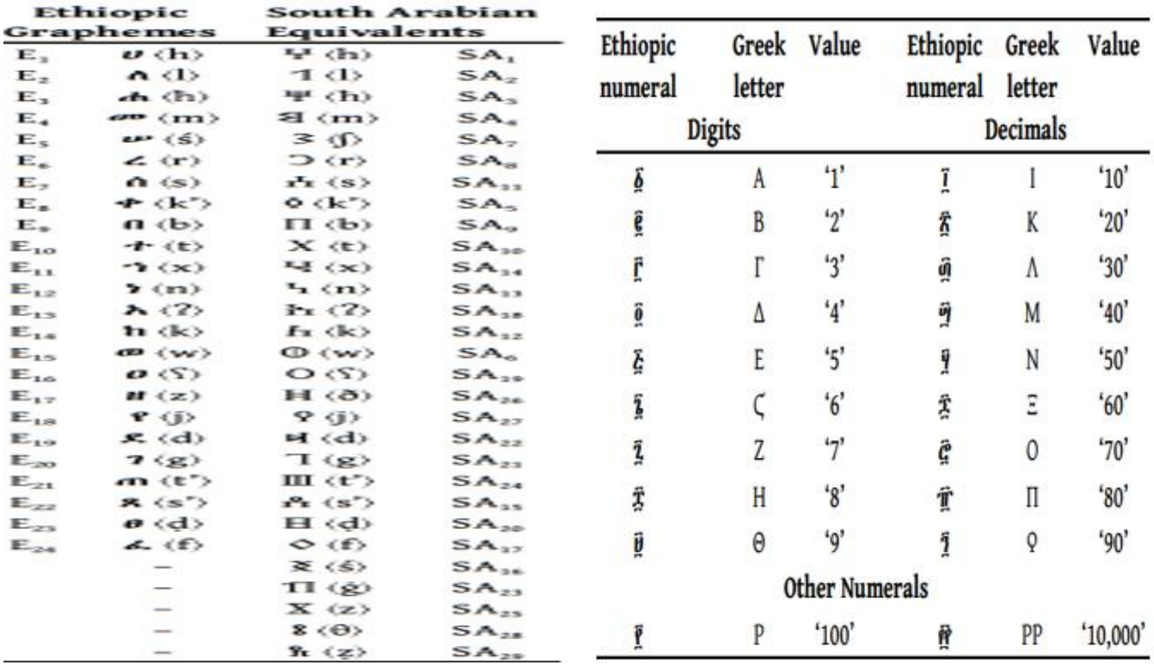
```
MODULE: DECODE Decoding using models previously trained
        Decoding 433 segments starting at 0 (part 1 of 1)
        0%
        Aligning results to find error rate
        SENTENCE ERROR: 37.4% (162/433)   WORD ERROR RATE: 9.1% (439/4815)
```

Figure 20: Experiment 1 result

```
MODULE: DECODE Decoding using models previously trained
        Decoding 433 segments starting at 0 (part 1 of 1)
        0%
        Aligning results to find error rate
        SENTENCE ERROR: 82.9% (359/433)   WORD ERROR RATE: 31.5% (1517/4815)
```

Figure 21: Experiment 2 result

Appendix 3: Derived Ethiopic scripts and numerals from South Arabian & Greek



Ethiopic Graphemes		South Arabian Equivalents	
E_1	ሀ ⟨h⟩	𐩠 ⟨h⟩	SA_1
E_2	ለ ⟨l⟩	𐩡 ⟨l⟩	SA_2
E_3	ሐ ⟨ḥ⟩	𐩢 ⟨ḥ⟩	SA_3
E_4	መ ⟨m⟩	𐩣 ⟨m⟩	SA_4
E_5	ሠ ⟨ś⟩	𐩦 ⟨ʃ⟩	SA_7
E_6	ረ ⟨r⟩	𐩧 ⟨r⟩	SA_8
E_7	ሰ ⟨s⟩	𐩪 ⟨s⟩	SA_{11}
E_8	ቀ ⟨kʼ⟩	𐩤 ⟨kʼ⟩	SA_5
E_9	በ ⟨b⟩	𐩨 ⟨b⟩	SA_9
E_{10}	ተ ⟨t⟩	𐩩 ⟨t⟩	SA_{10}
E_{11}	ኀ ⟨x⟩	𐩭 ⟨x⟩	SA_{14}
E_{12}	ነ ⟨n⟩	𐩬 ⟨n⟩	SA_{13}
E_{13}	አ ⟨ʔ⟩	𐩱 ⟨ʔ⟩	SA_{18}
E_{14}	ከ ⟨k⟩	𐩫 ⟨k⟩	SA_{12}
E_{15}	ወ ⟨w⟩	𐩥 ⟨w⟩	SA_6
E_{16}	ዐ ⟨ʕ⟩	𐩲 ⟨ʕ⟩	SA_{19}
E_{17}	ዘ ⟨z⟩	𐩸 ⟨ð⟩	SA_{26}
E_{18}	የ ⟨j⟩	𐩺 ⟨j⟩	SA_{27}
E_{19}	ደ ⟨d⟩	𐩵 ⟨d⟩	SA_{22}
E_{20}	ገ ⟨g⟩	𐩴 ⟨g⟩	SA_{21}
E_{21}	ጠ ⟨tʼ⟩	𐩷 ⟨tʼ⟩	SA_{24}
E_{22}	ጸ ⟨sʼ⟩	𐩮 ⟨sʼ⟩	SA_{15}
E_{23}	ፀ ⟨ḍ⟩	𐩳 ⟨ḍ⟩	SA_{20}
E_{24}	ፈ ⟨f⟩	𐩰 ⟨f⟩	SA_{17}
	—	𐩯 ⟨ś⟩	SA_{16}
	—	𐩶 ⟨ġ⟩	SA_{23}
	—	𐩹 ⟨z⟩	SA_{25}
	—	𐩻 ⟨θ⟩	SA_{28}
	—	𐩼 ⟨ẓ⟩	SA_{29}

Ethiopic numeral	Greek letter	Value	Ethiopic numeral	Greek letter	Value
Digits			Decimals		
፩	A	'1'	፲	I	'10'
፪	B	'2'	፳	K	'20'
፫	Γ	'3'	፴	Λ	'30'
፬	Δ	'4'	፵	M	'40'
፭	E	'5'	፶	N	'50'
፮	Ϛ	'6'	፷	Ξ	'60'
፯	Z	'7'	፸	O	'70'
፰	H	'8'	፹	Π	'80'
፱	Θ	'9'	፺	Ϙ	'90'
Other Numerals					
፻	P	'100'	፼	PP	'10,000'

Figure 22: **Ethiopic letters and numerals from South Arabian scripts and Greek letters**
Adopted from (Meyer, 2016)

Appendix 3. 1 Ge'ez alphabets, numerals and punctuations

Table 13: Ge'ez alphabets

	Vowel orders						
	ä	u	i	a	e	ə	O
h	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ
l	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
H	ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ
m	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ
s	ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ
r	ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ
S	ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ
q	ቀ	ቁ	ቂ	ቃ	ቄ	ቅ	ቆ
b	በ	ቡ	ቢ	ባ	ቤ	ብ	ቦ
t	ተ	ቱ	ቲ	ታ	ቴ	ት	ቶ
x	ኀ	ኁ	ኂ	ኃ	ኄ	ኅ	ኆ
n	ነ	ኑ	ኒ	ና	ኔ	ን	ኖ

'	አ	ኡ	ኢ	ኣ	ኤ	እ	ኦ
k	ከ	ኩ	ኪ	ካ	ኬ	ክ	ኮ
w	ወ	ዉ	ዊ	ዋ	ዌ	ው	ዎ
'	ዐ	ዑ	ዒ	ዓ	ዔ	ዕ	ዖ
z	ዘ	ዙ	ዚ	ዛ	ዜ	ዝ	ዞ
y	የ	ዩ	ዪ	ያ	ዬ	ይ	ዮ
d	ደ	ዱ	ዲ	ዳ	ዴ	ድ	ዶ
g	ገ	ጉ	ጊ	ጋ	ጌ	ግ	ጎ
T	ጠ	ጡ	ጢ	ጣ	ጤ	ጥ	ጦ
P	ጰ	ጱ	ጲ	ጳ	ጴ	ጵ	ጶ
s'	ጸ	ጹ	ጺ	ጻ	ጼ	ጽ	ጾ
ḍ	ፀ	ፁ	ፂ	ፃ	ፄ	ፅ	ፆ
f	ፈ	ፉ	ፊ	ፋ	ፌ	ፍ	ፎ
p	ፐ	ፑ	ፒ	ፓ	ፔ	ፕ	ፖ

Ge’ez labiovelars

	ä	i	a	e	ə
q^w	ቈ	ቊ	ቋ	ቌ	ቍ
x^w	ኈ	ኊ	ኋ	ኌ	ኍ
k^w	ኰ	ኲ	ኳ	ኴ	ኵ
g^w	ጐ	ጒ	ጓ	ጔ	ጕ

Table 14: Ge’ez numerals

፩	፪	፫	፬	፭	፮	፯	፰	፱	፲
1	2	3	4	5	6	7	8	9	10
፳	፴	፵	፶	፷	፸	፹	፺	፻	፼
20	30	40	50	60	70	80	90	100	10,000

adapted from (Meyer, 2016) with little modification

Table 15: Ge'ez punctuation marks

No	Punctuation marks & their names	Descriptions
1	፡ ንኡስ ነጥብ/ ንጽል	used between two words to separate them
2	፣ or ፥ ንኡስ ሥረዝ /ንጽል	used to isolate names, phrases and minor sentences those have not related and tied up themselves
3	፤ ዐቢይ ሥረዝ/ ክዑብ	used to separate small sentences those tied up themselves and may have the same or different idea
4	። ዐቢይ ነጥብ	can put at the end of sentence to show the idea is ended
5	…. ነጠብጣብ	can place between words or phrases to indicate words or phrases instead of writing them. The other usage is that it used to avoid repeated words for example, ኤርምያስ፥ ቤተ፡ . . . . ., ሰብአ፡ . . . . (which is equal to ኤርምያስ፥ ቤተ፡ ኤርምያስ, ሰብአ፡ ኤርምያስ)
6	‘’ ቀርን	used to indicate same thing or other option
7	“ ” አቅርንት	to indicate other people speech or word
8	፨ ምዕራፍ	Used to indicate the ending of chapter(s)

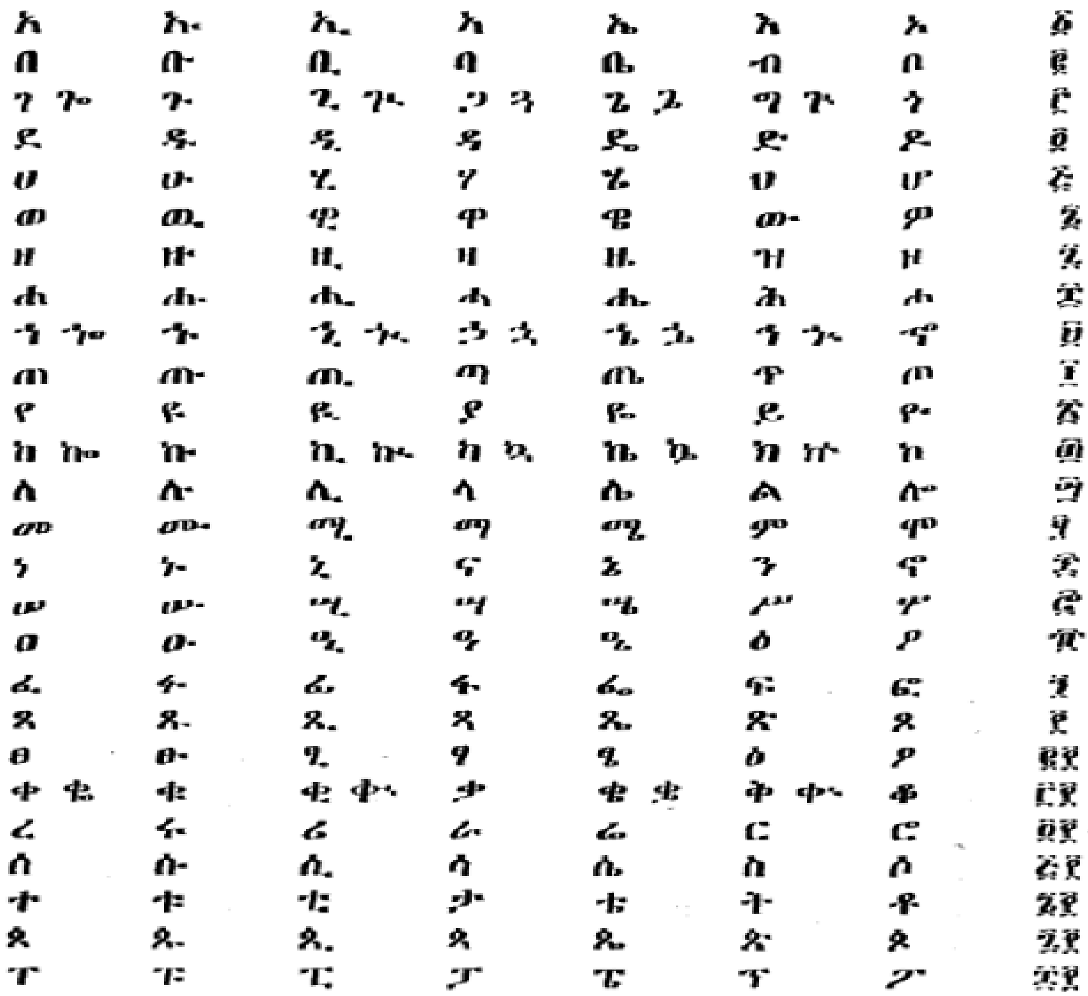


Figure 23: The Abegede Fidel

Abegede fidel/አበገደ ፊደል (Fidele-Hawaria/ፊደለ ሐዋርያ) (Kidanewold, 1948)